TOWARDS AGREEMENT: BAYESIAN EXPERIMENTAL DESIGN

by

Jameson Burt
Purdue University

Technical Report #90-41

Department of Statistics
Purdue University

Thesis: May 1989
August 1990

*

TOWARDS AGREEMENT: BAYESIAN EXPERIMENTAL DESIGN

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jameson Burt

In partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 1989

When reporters asked one of the county commissioners her response to the mayor's objection, she said in a voice of pained innocence, as though her statement would clearly prove how illogical the opposition was,

> "Their claim is that they don't have all the facts, and therefore are opposing this. They have all the facts that we have!"
>
> ——Missoula, Montana

ACKNOWLEDGEMENTS

I thank Professor Leon Gleser for suggesting this worthy problem. His enthusiasm and criticism kept this project moving on a straighter course.

TABLE OF CONTENTS

Page

LIST OF TABLES

GLOSSARY

		Page
N_ϵ^+	the smallest sample size n for which $\rho_m \geq \epsilon$ for all $m \geq n$	19
H_i for $i = 0, 1$	the hypotheses	22
Θ	the parameter space	22
λ	a dominating measure on the sample densities $f(x\|\theta)$	22
$\mathcal{X}$	the sample space	22
audience		22
Γ	the audience of observers γ	22
observer		22
γ	the index for an observer	22
π_γ	the prior of observer γ	22
μ	a dominating measure for the prior densities π_γ	22
$L_\gamma(a, \theta)$	the loss function for observer γ	22
$\mathcal{A}$	the action space	22
a_i for $i = 0, 1$	actions in hypothesis testing	22
$m_\gamma(\underset{\sim}{x}_n)$	the marginal density $\int_\Theta f(\underset{\sim}{x}_n\|\theta)\pi_\gamma(\theta)\, d\mu(\theta)$	23
A_i	those $\underset{\sim}{x}_n$ for which all observers γ make the decision i	23
π_*	the experimenter's prior on Θ	23
ρ_n	experimenter's probability that all observers make the correct decision	23
N_ϵ	the smallest sample size for which $\rho_n \geq \epsilon$	24
T_n	a p-dimensional sufficient statistic for $\underset{\sim}{X}_n$	26
$Q_{i\gamma}(T_n)$	the posterior $\pi_\gamma(H_i\|\underset{\sim}{x}_n)$ when T_n exists	26
$\mathcal{T}$	the range of the p-dimensional sufficient statistic T_n	27
V_i for $i = 0, 1$	the probability $\inf_{\gamma\in\Gamma} \pi_\gamma(H_i\|\underset{\sim}{x}_n)$	29
$\tilde{A}_i$	the subspace of $\mathcal{X}^n$: $\{V_i > 0.5\}$	29

ABSTRACT

Burt, Jameson. Ph.D. Purdue University, May 1989. Towards Agreement: Bayesian Experimental Design. Major Professor: Leon J. Gleser.

An experimenter wishes to design an experiment to settle an inferential question about the value of a parameter θ. The data $X_1, \ldots, X_n$ from such an experiment will be viewed by a class Γ of Bayesians, where each such Bayesian γ has a prior distribution $\pi_\gamma(\theta)$ for θ. Denote by A_θ the event: "the collection of all samples $X_1, \ldots, X_n$ for which all Bayesians in Γ agree to the correct decision concerning θ." Using his own prior distribution $\pi_*(\theta)$, the experimenter wishes the preposterior probability $P(A_\theta)$ to be at least as large as a prespecified constant ϵ $(0 < \epsilon < 1)$.

In the case of hypothesis testing, this paper gives necessary conditions for the existence of a sample size N_ϵ achieving these goals, and also gives some sufficient conditions for N_ϵ to exist. Interestingly, $P(A_\theta)$ need not be monotone increasing in n, so that observing data additional to the experiment can cause $P(A_\theta)$ to decrease from above ϵ to below ϵ. Consequently, to better settle the correct decision concerning θ, the smallest value of N_ϵ such that $P(A_\theta) \geq \epsilon$ for all $n \geq N_\epsilon$ is sought. Bounds and numerical algorithms for N_ϵ are given. Some results extending the theory to estimation problems involving θ are also presented.

Restrict the event A_θ so that each Bayesian, in addition to choosing the correct decision, also satisfies his own goal for a low posterior expected loss using that correct decision. This definition of A_θ extends the theory, reducing to the original theory through an induced set of new priors in the case of hypothesis testing.

1. INTRODUCTION

One goal of any inquiry—and certainly of an experiment—is that truth be found and, usually, that the truth found out be agreed upon by others. An early adherent of such ideas was C.S. Peirce (Dewey(1938), page 490), a mathematician who wrote to the layman.

> C.S. Peirce is notable among writers on logical theory for his explicit recognition of the necessity of the social factor in the determination of evidence and its probative force. The following representative passage is cited: "The next most vital factor of the method of modern science is that it has been made social. On the one hand, what a scientific man recognizes as a fact of science must be something open to anybody to observe, provided he fulfills the necessary conditions, external and internal. As long as only one man has been able to see a marking upon the planet Venus, it is not an established fact. ... On the other hand, the method of modern science is social in respect to the solidarity of its efforts. The scientific world is like a colony of insects, in that the individual strives to produce that which he himself cannot hope to enjoy."

Peirce (Murphee(1964)) emphasized the use of a consensus in a scientific community,

> Given Peirce's definitions of truth—namely, that to which a community of investigators would give assent, based upon the results of their

> cooperative inquiry—it clearly follows that ... "a claim to truth is a public claim which only a public can verify."

He also emphasized that the appropriate consensus is that resulting from observations in the long run, Peirce(1878),

> "The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real."

The last two decades have seen a burgeoning of interest in the "social" aspect of inferences. What inferences represent a consensus for several people? How can the degree of consensus be measured? How can the opinions of others be used for the inference of one? The ideas of some researchers about these questions are given briefly in this introduction. Here are mentioned classical approaches to choosing an experiment, Bayes approaches to choosing an experiment, and approaches to reaching a consensus with or without an experiment. The last section of this introduction gives an overview concerning the choice of a sample size for an experiment so that a unanimous consensus results: the interest of this paper.

To facilitate this introduction, the following conventions will be used when appropriate. The notation introduced here will usually differ from that of attributed papers. Inferences or decisions are to be made concerning the value of a parameter θ in the parameter space Θ. When this introduction discusses consensus, the m members of a group Γ are to reach a consensus concerning θ—a consensus represented by a probability distribution π_G or a decision a_G, possibly a randomized rule (though our notation will not account for this). The group members have prior densities

$$\pi_1(\theta), \pi_2(\theta), \ldots, \pi_m(\theta)$$

with respect to some dominating measure $\mu(\theta)$ on Θ. The members may also have the loss functions, for some decision a in the action space $\mathcal{A}$,

$$L_1(a,\theta), L_2(a,\theta), \ldots L_m(a,\theta).$$

When an experiment will be performed, the data arise out of the sample space $\mathcal{X}$ through the probability density $f(x|\theta)$ with respect to the measure $\lambda(x)$. A sample of size n uses the corresponding notation $\mathcal{X}^n$, $f(\underset{\sim}{x}_n|\theta)$ and $\lambda(\underset{\sim}{x}_n)$. An action "$a$" might then be denoted $a(\underset{\sim}{x}_n)$. An *external observer*—a *decision maker*, an *arbitrator*, or a fictitious though altruistic *supra-Bayesian*—may oversee the inference for a consensus. Denote the external observer's prior density by $\pi_*(\theta)$. He may be concerned only about his own decision, not some consensus. He then has his own loss function $L_*(a,\theta)$. When the external observer considers the group members' probability distributions as data, albeit of an unusual nature, call the group members *experts*. When the external observer considers the group members' welfare, call the group an *audience* (or a community). The notation $P_\zeta(\cdot)$ indicates that the considered probability uses an implicit parameter with the value ζ or uses the distribution ζ. For example, $P_{n_0}(\cdot)$ indicates that $n = n_0$, and $P_\nu(\cdot)$ indicates that the probability density ν is used.

1.1 Classical sample size.

A frequentist decision approach, in the absence of an experimental cost assessment, chooses a sample size n giving some small risk K. Denote the risk by

$$R(a,\theta,n) = \int L\Big(a(\underset{\sim}{x}_n),\theta\Big) f(\underset{\sim}{x}_n|\theta) d\lambda(\underset{\sim}{x}_n).$$

A minimax approach seeks the sample size

$$n_0 = \inf\left\{n \geq 0 : \inf_{a\in\mathcal{A}} \sup_{\theta\in\Theta} R(a,\theta,n) \leq K\right\}$$

should such an n_0 exist.

Hypothesis testing

When the parameter space Θ is viewed as two sub-spaces Θ_0 and Θ_1 through an action space $\mathcal{A}$ containing two corresponding actions a_0 and a_1, then the decision problem is a hypothesis testing problem. When the subspaces Θ_0 and Θ_1 have disjoint convex hulls, one version of this problem chooses a sample size n_0 through a consideration of two points $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ which are each near to the other subspace. If the loss for the wrong decision $a(\underline{x}_n) = i$ is C_i, 0 otherwise, the loss function may be written

$$L(a(\underline{x}_n), \theta) = \begin{cases} 0 & \text{if } a(\underline{x}_n) = i \text{ and } \theta \in \Theta_i \\ C_i & \text{if } a(\underline{x}_n) = i \text{ and } \theta \notin \Theta_i \quad \text{where } i = 0, 1\,. \end{cases}$$

For Θ restricted to the two points θ_0 and θ_1, the sample size for the risk bound K is

$$n_0 = \inf\left\{n \geq 0 : \inf_{a \in \mathcal{A}} \left\{\max_{i=0,1} R(a, \theta_i, n)\right\} \leq K\right\}. \tag{1.1}$$

For some numbers $0 < \alpha < 1$ and $0 < \beta < 1$, let $C_1 = 1/\alpha$, $C_0 = 1/(1-\beta)$ and $K = 1$. Then (1.1) can be rewritten

$$n_0 = \inf\Big\{n \geq 0 : P_{\theta_0,n}\left(a(\underline{x}_n) = 1\right) \leq \alpha \quad \text{and} \quad P_{\theta_1,n}\left(a(\underline{x}_n) = 1\right) \geq \beta \quad \text{for some action } a\Big\}.$$

In this form, we see that n_0 is the smallest sample size that can be used for an α-level test having power β.

Estimation

When $L(a(\underline{x}_n), \theta) = (a(\underline{x}_n) - \theta)^2$, then $\inf_{a \in \mathcal{A}} R(a, \theta, n)$ is the expected mean square. For some problems (eg, for the Gaussian distribution),

$$\inf_{a \in \mathcal{A}} R(a, \theta, n) = \inf_{a \in \mathcal{A}} \sup_{\theta \in \Theta} R(a, \theta, n) \quad \text{for all } \theta \in \Theta\,.$$

A small expected mean square K is sought through the sample size n_0.

Let $I(\cdot)$ denote the indicator function. When $L(a(\underset{\sim}{x}_n), \theta) = I\left\{|a(\underset{\sim}{x}_n) - \theta| > d\right\}$, then $R(a, \theta, n) = P_\theta\left(|a(\underset{\sim}{x}_n) - \theta| > d\right)$. This problem seeks that the width of a $100(1 - K)$ percent confidence interval be no larger than $2d$. The minimax approach chooses the sample size

$$n_0 = \inf\left\{n \geq 0 : \inf_{a \in \mathcal{A}} \sup_{\theta \in \Theta} P_{\theta,n}\left(|a(\underset{\sim}{x}_n) - \theta| > d\right) \leq K\right\}.$$

1.2 Single Bayesian sample size.

Design for linear models

The primary problem of Bayesian experimental design in linear regression models is not so much the sample size as the design matrix $X_{n \times r}$ to use (this subsection uses notation in fidelity with the literature). A set of parameters $\theta_{k \times 1}$ is estimated by way of an experiment, having the design matrix $X_{k \times n}$, modeled as

$$Y_{n \times 1} = X_{n \times k}^T \theta_{k \times 1} + e_{n \times 1}$$

where $X_{k \times n} = (x_1, x_2, \ldots x_n)$ is the design matrix, $E(e_{n \times 1}) = 0_{n \times 1}$, $Cov(e_{n \times 1}) = \sigma^2 I_{n \times n}$, $E(\sigma^2) = \sigma_0^2$, $E(\theta_{k \times 1}|\sigma) = E(\theta_{k \times 1}) = \mu$, and $Cov(\theta_{k \times 1}) = \Lambda_{k \times k}$.

Interest in estimating $c^T\theta$ with interest (in $c_{k \times 1}$) expressed through some measure $\nu(c_{k \times 1})$ leads to minimizing (for least squares linear estimators):

$$\text{tr}\left[(\psi)\left(R + XX^T\right)^{-1}\right], \tag{1.2}$$

where $\psi_{k \times k} = \int (cc^T)\, d\nu(c)$ and $R_{k \times k} = \sigma_0^2 \Lambda^{-1}$. Minimizing (1.2) is equivalent to minimizing the Bayes risk when the priors and likelihoods are normal. The optimality criterion to minimize (1.2) is called variously *ψ-optimality*, Bayes L-optimality (L_B-optimality), and Bayes A-optimality (A_B-optimality; particularly when $\psi = I_{k \times k}$). Bandemer, Näther and Pilz (1987) survey Bayes experimental design for linear regression models.

The ψ-optimal design points x_i minimizing (1.2) depend upon the sample size n, unlike classical designs. However, Chaloner (1984, Theorem 2) showed (for continuous designs) that the number of distinct design points x_i constituting a ψ-optimal design matrix need be no more than $r(2k - r + 1)/2$, where $r = \text{rank}(\psi_{k \times k})$; no more than $k(k+1)/2 + 1$ for actual discrete designs.

Sample size

Bayesian decision theory, in the absence of a cost of experimentation, chooses a sample size giving some small Bayes risk K. Denote the Bayes risk of the decision "a" by

$$R(a, n) = \int_\Theta \int_{\mathcal{X}^n} L(a(\underset{\sim}{x}_n), \theta)\, f(\underset{\sim}{x}_n | \theta)\, \pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta)\,.$$

Then n_0 is that sample size for which

$$\inf_{a \in \mathcal{A}} R(a, n_0) = \min_{n \geq 0} \inf_{a \in \mathcal{A}} R(a, n) \leq K\,.$$

Denoting the indicator function by $I(\cdot)$, let g be some metric on $a(x) - \theta$ and let $d > 0$. One common loss function for estimation problems is

$$L(a(x), \theta) = I\{g(a(x) - \theta) \leq d\}\,.$$

Adcock (1987) considers a multinomial distribution $f(x|\underset{\sim}{\theta})$ with k classes (ie, $\underset{\sim}{\theta} = \theta_{k \times 1}$), using the conjugate Dirichlet prior density for $\pi_*(\underset{\sim}{\theta})$. He uses metrics like $g(\underset{\sim}{z}) = \underset{\sim}{z}' M \underset{\sim}{z}$ for some positive definite matrix M, and like $g(\underset{\sim}{z}) = \max_{i \leq k} |r_i z_i|$ for some $r_i > 0$. He calls $C(\underset{\sim}{x}_n) = \left\{\theta : g(a(\underset{\sim}{x}_n) - \theta) \leq d\right\}$ a *tolerance region*, either ellipsoidal or hyper-cubic with his metrics. Letting $0 < \epsilon < 1$ and $m_*(\cdot)$ be the marginal distribution of $\underset{\sim}{x}_n$, then Adcock seeks a sample size n for which

$$\int_{\mathcal{X}^n} P\left(\underset{\sim}{\theta} \in C(\underset{\sim}{x}_n) | \underset{\sim}{x}_n\right) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \geq \epsilon\,.$$

Reworded, Adcock seeks a Bayesian confidence interval that is of fixed width and, while not of some minimum confidence level, that has on average a 100ϵ-percent confidence level.

1.3 Opinion-Preference pools (no experiment problems).

Suppose that an opinion or preference must be made with the information at hand, though a sample may have already been collected. Although some of the research in this area uses odds ratios and preference relations, this review considers only group members fitting the usual Bayesian paradigm with prior probability distributions and possibly with loss functions. The priors and loss functions of the observers represent information that plausibly can form/improve some opinion or some decision. Four aspects delineate the research into opinion-preference pools. First, the point of view may be categorized as follows.

I-1. *external observer*- An external observer uses his own prior (and possibly loss function) to pool the group's opinions, to pool the group's preferences, or to make his own decision.

I-2. *group only*- The group's opinions/preferences are pooled without some single guiding Bayesian prior. Often, an "axiomatic" (see below) approach is used for the group to assure that its actions are rational: Bayesian. Without an external observer, this problem has some deficiencies. If the axiomatic approach, having its own deficiencies, is not used, then only approaches even more ad hoc remain.

Second, whether decisions are to be made may be exhibited as follows.

II-1. *aggregate probabilities*- Only an opinion pool, sometimes called a consensus of opinion, is to be made.

II-2. *make decisions*- A preference pool is to be made, usually by the use of both the loss functions and the Bayes priors.

Third, the approach used to solve the problem is one of the following.

III-1. *axiomatic*- A set of reasonable (though on retrospect often unreasonable) assumptions are used to deduce a formula for pooling.

III-2. *modeling*- In the most common models, the distribution of the group members' priors are modeled by some probability distribution $\psi(\pi_1, \pi_2, \ldots, \pi_m|\theta)$, possibly accounting for priors that are not independent of one another. Modeling generally treats the group members' priors as just data, rather than as probability distributions.

Sometimes, both the axiomatic and the modeling approaches lead to the same pooling form, eg, the linear opinion pool $\pi_G = \sum_{i=1}^{m} \alpha_i \pi_i$ for some nonnegative α_i summing to 1. The modeling approach may then provide values for the parameters, parameters that the axiomatic approach only requires to exist.

Fourth, whether data have already been observed may be exhibited as follows.

IV-1. *entirely a priori*- No data can be separated from the group members' priors.

IV-2. *data*- Some data may already have been observed by the group members. The likelihood function for the data may be different for each group member. Or, with the same likelihood function, the data observed may have been different for each group member. Or, with the same likelihood function the same data may have been observed by all the group members. Some

> researchers try to extract the likelihood from a group member's prior. When group members differ only because of the different data that they have seen, not because of intrinsically different priors, this extraction of the likelihoods results in unanimous agreement.

Simon French (1985) calls the "group" "aggregation of probabilities" problem the text-book problem. For the text-book problem, a summary of the group must be made for unknown other(s) in unknown circumstances. French also considered the "group" "decision" problem.

deFinetti (Genest and Zidek (1986), page 130) showed that if $L_1 = L_2 = \cdots = L_m$, then a "decision" based on an average opinion of the "group" members is better than a decision based on an average of the individual group members' decisions. Consequent separate pooling of the group members' priors and of the group members' loss functions have resulted from many "axiomatic" approaches to a group's decision. Simon French (1985) critically surveys many of the opinion pools, especially those arising from the axiomatic approach.

Considering separate pooling of priors and losses in the group decision problem, Hylland and Zeckhauser (1979) investigated the following fundamental axiom.

Weak Pareto Principle: *If*

$$\int L_i(a_1, \theta)\pi_i(\theta)d\mu(\theta) > \int L_i(a_2, \theta)\pi_i(\theta)d\mu(\theta)$$

for all group members $i = 1, 2, \ldots, m$, then action a_2 is preferred to action a_1 (a_G would not be a_1).

When the group is to behave rationally—ie, as some Bayesian would—Hylland and Zeckhauser show (under mild conditions) that the *Weak Pareto Principle* leads to an undesirable rule: a dictatorial decision which ignores the members' priors π_i. Raiffa (1968) reasons that the Pareto Principle may not be fundamental, especially when the group members agree for disparate reasons. Keeping the Pareto Principle, Weerahandi and Zidek (1981, page 88) allow the group to act irrationally, arriving at the non-Bayesian group decision rule a_G which minimizes

$$(1.3) \qquad \prod_{i=1}^{m} \left[\int L(a,\theta)\pi_i(\theta)d\mu(\theta) \right]^{\alpha_i}, \qquad \text{where } \sum_{i=1}^{m} \alpha_i = 1,$$

for some nonnegative α_i of any origin. When $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 1/m$, (1.3) is called the Nash product.

Many workers in this field have concluded that an "axiomatic" approach is best replaced by a "modeling" approach, usually necessitating an "external observer". Simon French (1985) calls this the expert problem. The opinion-preference pool resulting from a modeling approach is then considered to validate or invalidate pooling axioms in retrospect. A modeling approach typically updates an external observer's prior probability on the parameter space to

$$\pi_G = \pi_*(\theta|\pi_1, \pi_2, \ldots \pi_m) \propto \pi_*(\theta)\psi(\pi_1, \pi_2, \ldots \pi_m|\theta) \ ,$$

as in Genest and Zidek (1986, page 120). Genest and Schervish (1985) do this for an external observer who knows only the moments (at least one moment) of the group members' distributions.

One approach to a "group" "aggregation of probabilities" has the group members engage in dialogue, called the Delphi technique when the group does not physically meet. A formalized variant of this is the DeGroot-Lehrer "model", Lehrer and Wagner (1981). Here, each group member i elicits a weight $w_{ij} \geq 0$ representing how much i would follow the opinion of j, where $\sum_{j=1}^{m} w_{ij} = 1$. Group member i also has, for some single event, the probability $\pi_i^{(0)}$ which he will update to $\pi_i^{(k)}$ on the kth iteration of dialogue between the group members. On the kth iteration of dialogue, i's probability for the event is formalized to be

$$\pi_i^{(k)} = \underset{\sim i}{w}'\, \underset{\sim}{\pi}^{(k-1)} = \sum_{j=1}^{m} w_{ij} \pi_j^{(k-1)} \, .$$

Under mild conditions, there is one π_G to which $\lim_{k \to +\infty} \pi_i^{(k)} = \pi_G$ for every group member i. In a variation of this problem, an opinion pool ("consensual probability") like π_G results from similar "dialogues" in which the weights w_{ij} are now allowed to vary at each stage k. Lehrer argues that any vagueness in a group member's prior is represented in the credence, through $\underset{\sim i}{w}$, that he gives to others' opinions. For an example, Lehrer considers "the definition of some word." For a proper definition of a word, person i defers to some person j, who himself defers to some person j' who is unknown to i. Consequently, some expert who may have put a definition in a dictionary is largely deferred to. The DeGroot-Lehrer scheme iterates, adding information to the individuals but not to the group. The DeGroot-Lehrer model demonstrates that just—without an experiment—group members' judgements of each other lead to agreement, though not necessarily to a correct agreement.

Besides what actual examples would indicate, this discussion indicates that no single formula for opinion-preference pooling is universally suitable. Some hopes for opinion-preference pools are expressed by Genest and Zidek (1986):

> Ignoring practical problems of implementation which are the object of current research, the Bayesian program would seem to be entirely satisfactory as a normative theory for the individual. However, groups of individuals are left stranded; no concept equivalent to the classical notion of objectivity is available to them.

> Weerahandi and Zidek (1981,1983) propose such a concept. Their idea is related to what Dawid (1982a) defines and calls "intersubjectivity." According to this definition, the opinion or conclusion reached by an individual from the results of an experiment would be called "objective" or perhaps "intersubjective," if the same conclusion were reached by a succession of individuals faced with the same results. But just as the classical notion of objectivity is challenged by inevitable variations in the results of repeated experiments, so intersubjectivity needs to contend with variations in the conclusions derived by the succession of individuals viewing the evidence. This calls for an analogue of the law of averages, that is, a method of "averaging" the possibly diverging opinions of a group of analysts and a limit theory for the long run.

Simon French (Genest and Zidek (1986, page 138) responded,

> Intersubjectivity is about consensus in the *strict* sense of that word, that of unanimous agreement.

That is what this thesis will address.

1.4 Experiment induced consensus.

<u>Consensus viewed as persuasion</u>

Jackson, Novick and DeKeyrel (1980) consider an external observer ("advocate" of his own position, with prior π_*) who wishes to convince a single group member ("adversary" with prior π_1; the number of group members $m = 1$) of the advocate's opinion (or position) through an experiment. For example, the advocate may wish that his level of achievement θ be conveyed to a teacher (adversary) through an exam of sufficient size n to convince the teacher that $\theta > \theta_0$. The advocate would then want $\pi_1(\theta > \theta_0 | \underset{\sim}{x}_n)$ to be large. As the advocate is assessing probabilities preposterior and as $\underset{\sim}{x}_n$ is a random sample, then the advocate might assess

$$\int_{\mathcal{X}^n} \pi_1\left(\theta > \theta_0 | \underset{\sim}{x}_n\right) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n)\,, \tag{1.4}$$

where

$$m_*(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\,.$$

Jackson, Novick and DeKeyrel call the marginal probability density

$$\pi_{*\cdot 1}(\theta) = \int_{\mathcal{X}^n} \pi_1(\theta|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n)\,, \tag{1.5}$$

the advocate's "preposterior density" for the adversary's posterior density. The probability (1.4) re-formulates to

$$\pi_{*\cdot 1}\Big((\theta_0, +\infty)\Big) = \int_{\theta_0}^{+\infty} \pi_{*\cdot 1}(\theta)\, d\mu(\theta)\,. \tag{1.6}$$

Generally, the authors consider the rate at which the advocate's preposterior density for the adversary, $\pi_{*\cdot 1}(\theta)$, converges to the advocate's prior density $\pi_*(\theta)$. They measure this rate of convergence through the corresponding rates of convergence of the mean and variance of $\pi_{*\cdot 1}(\theta)$ to the mean and variance of $\pi_*(\theta)$. The mean and variance are themselves the best measures of the convergence rate when appropriate loss functions are used. Notice that the probability $\pi_{*\cdot 1}\Big((\theta_0, +\infty)\Big)$ converges to $\pi_*\Big((\theta_0, +\infty)\Big)$, not 1 since the advocate is not sure of θ.

<u>Consensus measured by "votes"</u>

> "A group of statisticians or experts making inference from a common source of data would normally be expected to approach a consensus in their inference, even without communicating among themselves, as the amount of data increases indefinitely. This obviously assumes not only that no members of the group make mistakes, but also that none

have adopted initial beliefs so prejudiced that they preclude a sound conclusion. In terms of formal Bayesian inference one would expect consensus to form if the members of the group make inference from a common data set, if they have a common model or likelihood function, and if none of their prior beliefs totally excludes any possible values of the parameters,"

Owen (1985, page 1036). Owen restricts his attention to consensus without the group members ("experts" in Owen (1985)) necessarily reaching correct decisions. Owen's comment above is true for the approach to a unanimously correct decision—see Theorem 5.2 on page 109 of this paper. Considering h finite hypotheses, Owen considers group member γ's *inaccuracy* to be measured by the Bayes risk:

$$\inf_{a\in\mathcal{A}} \int_{\Theta} \int_{\mathcal{X}^n} L(a(\underset{\sim}{x}_n), \theta) f(\underset{\sim}{x}_n | \theta) \pi_\gamma(\theta) \, d\lambda(\underset{\sim}{x}_n) \, d\mu(\theta) \,,$$

where Θ and $\mathcal{A}$ contain $h < \infty$ corresponding elements and L is 0 for $a(\underset{\sim}{x}_n) = \theta$. Owen considers group member γ's *personal confidence* in his decision to be measured by

$$\max_{\theta\in\Theta} \pi_\gamma(\theta | \underset{\sim}{x}_n) \qquad \text{for } \gamma = 1, 2, \ldots m$$

(Owen actually puts no finite limits on the size of the group Γ). Let $M_\theta(\underset{\sim}{x}_n)$ be the number of members making the decision θ. Owen measures the *amount of consensus* by

$$\max_{\theta\in\Theta} \frac{M_\theta(\underset{\sim}{x}_n)}{m} \,, \tag{1.7}$$

the proportion of the group (at time n) who have chosen the majority decision a_G—in contradistinction to the probability that all group members make the same decision,

$$\mathrm{P}\left(\max_{\theta\in\Theta} \frac{M_\theta(\underset{\sim}{x}_n)}{m} = 1 \right)$$

relative to some measure on $\mathcal{X}^n$. Owen shows that the "amount of consensus" converges to 1, the "personal confidence" converges to 1, and "inaccuracy" converges to 0—all at the same (suitably defined) rate as n increases. Owen (1985) explains both the main interest and the conclusions of his paper,

> "Since accuracy implies consensus, but not vice-versa, one would expect that consensus forms at a rate at least as fast as accuracy, and probably faster. This partly explains the extent to which these results [in Owen's paper] are counter-intuitive. However, a more important reason for the counter-intuitive nature of the results is that experts are likely to communicate, and this would accelerate consensus with or without 'political' forces coming into play."

Dickey and Freeman (1975) was seminal to Owen (1985). Dickey and Freeman consider a similar problem with the finite parameter space $\Theta = (\theta_1, \theta_2, \ldots, \theta_h)$. However, the breadth of prior distributions (distribution vectors $(p_1, p_2, \ldots p_h)$) of the members in the group Γ are accounted for in an unusual way—indeed, Γ cannot be finite (m cannot be finite). The priors of the group Γ are inventoried by the Dirichlet distribution (not a "hierarchical" prior). The posteriors (vectors) of the group Γ have a corresponding inventorying distribution. This inventorying distribution facilitates finding the proportion of the group Γ who choose the same action (ie, $\theta \in \Theta$). This proportion is Dickey and Freeman's measure of consensus [The sheer number or breadth of posteriors (vectors) prevents a unanimous agreement, whatever the sample size]. Dickey and Freeman (1975) comment,

> "We do not discuss the data-sampling process, but it should be clear that the theory, if so enriched, would offer a possible model for the evolution of knowledge in a community of scientists ... The concept,

introduced here, of the coherent transformation of a population of prior probabilities may have uses in experimental design when a scientist wishes to perform an experiment that will have a high chance of bringing members of the population into close agreement. He might, for example, choose a sample size large enough to make the variance of a posterior probability in the modeled population less than a predetermined value. Alternatively, if the experimenter is himself convinced that $\theta = 1$, he may choose to continue observation until at least $100(1-\delta)$ percent of the modeled population have posterior probabilities, q_1, within the range from $1 - \epsilon$ to 1."

1.5 Towards agreement: Bayesian experimental design.

In this paper, an experimenter wishes to design an experiment for the unanimous agreement of a community of Bayesians (or *audience*) Γ about the value of the parameter θ. The observers γ in Γ will make inferences about the value of θ through their prior densities $\pi_\gamma(\theta)$ on Θ and data $\underset{\sim}{x}_n$ from the experiment. Denote by A_θ the event

$$A_\theta = \left\{ \underset{\sim}{x}_n : \text{all observers in } \Gamma \text{ choose the correct decision} \middle| \, \theta \right\}$$

that the data $\underset{\sim}{x}_n$ from the experiment results in a unanimous and a correct decision corresponding to the parameter θ. Using his own prior distribution $\pi_*(\theta)$, the experimenter wishes the preposterior probability of correct agreement,

$$\rho_n = \int_\Theta \int_{A_\theta} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta)\,,$$

to be at least as large as some prespecified constant ϵ $(0 < \epsilon < 1)$.

The goal of the experimenter can be restated in terms of the aspects delineating opinion-preference pools earlier.

I. The point of view is that of an "external observer," here called the experimenter, who evaluates the observers' decisions. The point of view is also that of the observers in the audience—the individual "members," not the "group"—who can make their own separate decisions, possibly spoiling the experimenter's wish for unanimity. In one sense, the observers' decisions are only facilitated by the experimenter, himself only providing odds for their decisions. By allowing unanimity to fail (with probability $1 - \rho_n$), whose point of view is taken becomes muddled. This is elaborated upon later in this introduction.

II. While he doesn't really make a decision himself (unless he is a member of the audience), the experimenter wants the observers in the audience ("group members") to "make decisions" for themselves. Choosing an experiment for unanimous agreement, the experimenter does not intend to compromise any observer's decision.

III. Since the experimenter anticipates unanimous agreement, the approach used to solve his consensus problem is one of experimental design, not an "axiomatic" or a "modeling" approach.

IV. Whether or not "data" have already been observed by some group members, the experimenter intends that more "data" will be collected through some experimental design.

Viewed another way, the experimenter wants to satisfy the Weak Pareto Principle: by satisfying its antecedent with a correct decision, and by satisfying its conclusion (consequent), with a high probability ϵ anyway.

Here is another perspective when Γ includes the experimenter. The experimenter seeks that a correct posterior decision be made using his prior π which can only be refined to a class of priors $\{\pi_\gamma, \gamma \in \Gamma\}$. This is a *robustness* interpretation: a planned posterior Bayes robustness via experiment. As the robustness will only occur with a probability of ρ_n, the decision $a(\underset{\sim}{x}_n)$ is robust with respect to the sample $\underset{\sim}{x}_n$ as much as it is with respect to the prior π.

As mentioned, the experimenter could wish to settle an inferential question for a community of observers. Equivalently, changing only the evocative words when the experimenter is also a member of the audience Γ, the experimenter could wish that the Bayesian employer (customers, clients, boss, or board of directors) make the same decision that the experimenter makes—rather like the persuasion of Jackson, Novick and DeKeyrel (1980). Or symmetrically, the experimenter could wish that he conform his own decision to his employer's decision through an appropriate experiment. This symmetry in the experimenter's perception of the decisions that will be reached (with a probability of ρ_n) shows that the perspective of the experimenter is one of "agreement," not one of "persuasion."

A strong interpretation ensues when the parameter space is binary, $\Theta = \{\theta_0, \theta_1\}$: when there are two hypotheses, both simple. For exposition, let the audience Γ comprise two observers with the priors $\pi_0 = \pi(\theta_0) = 0.95$ and $\pi_1 = \pi(\theta_0) = 0.05$, and let the experimenter's preposterior probability ρ_n be at least $\epsilon = 0.99$. The experimenter's aim may be rendered:

> The experimenter wishes that the correct decision—whichever it is—be made when the odds are 19 to 1 against that decision.

For the experimenter, this wish will occur with a probability of at least .99.

Many correct agreement problems can be reduced to equivalent problems with just two—though often extreme—observers, as above. This is the case for all simple-simple hypothesis problems.

Even in simple-simple hypothesis problems, ρ_n behaves unexpectedly. When the density $f(x|\theta)$ is Gaussian, the fifth example in Table 3.1 on page 50 exhibits such behavior. There, the experimenter would prefer taking no sample, for which the probability of agreement $\rho_n = .80$, than to chance his audience seeing the outcome of a sample of size $n = 30$, for which $\rho_n = .79$. Observing data additional to the experiment can cause ρ_n to decrease from above ϵ to below ϵ.

This behavior of ρ_n leads to the following judgment about the appropriate sample size to lead the audience to a correct agreement. When the audience will see no data—from whatever likelihood function—that

(i) has not been seen a priori,

that

(ii) is in addition to the experimenter's data,

and that

(iii) impinges upon the audience's inferential question,

then the appropriate sample size is the smallest one giving correct agreement with an adequately high probability:

$$N_\epsilon = \min\left\{n : n \geq 0,\ \rho_n \geq \epsilon\right\}.$$

Alternatively, when the above conditions are not met, for example, when other experimenters will be performing similar experiments, then the appropriate sample size should give a correct agreement no matter how much data will be seen by the audience:

$$N_\epsilon^+ = \min\left\{n : n \geq 0, \rho_m \geq \epsilon \text{ for all } m \geq n\right\}.$$

This gradation of the experimenter's goal can be expounded upon. In a diluted form, the experimenter could wish just that the observers agree. It is this weak agreement that Owen (1985) approached while also using a weakened measure of consensus that did not demand unanimous agreement. He measured consensus by the proportion of observers agreeing (see (1.7) on page 14 of this thesis), thus avoiding the need for an experimenter's evaluation of any agreement. Unifying some possible experimenter's goals: just "agreement" graduates to "correct agreement," agreement to the correct decision (the experimenter uses the sample size N_ϵ); which itself graduates to "correct agreement at an arbitrarily large sample size" (the experimenter uses the sample size N_ϵ^+). In the last example of Table 4.1 on page 93, the observers unanimously agree for every sample. However, they may agree to the wrong decision for some samples. Thus, $\rho_n = 1.000$ for $n = 1, 2$ and 3, but $\rho_n = .84$ for $n = 100$, while $\rho_n > .95$ beyond $n = 1000$. Consequently, $N_{.95} = 0$ but $N_{.95}^+ \approx 1000$.

Section 4.6 considers the sample sizes that the experimenter would choose for sub-audiences Γ_0 of the audience Γ. The sample size N_ϵ for the audience Γ may be larger than the maximum of the corresponding sample sizes for the sub-audiences Γ_0 of pairs of observers. In other words, there may be some sample size $\grave{n}$ at which the probability of agreement is large for each pair of observers in Γ. Yet, that same sample size $\grave{n}$ may not yield a large probability of agreement for all of the observers in Γ.

Bounds and numerical algorithms for the experimenter's sample size are given for hypothesis problems. Simple-simple hypothesis problems yield bounds on the sample size in closed mathematical form. For a composite hypothesis problem with a Gaussian likelihood, the formula (4.72) on page 92 gives a crude closed form bound on the sample size. For simple-simple hypothesis problems, when the likelihood is Gaussian, a necessary condition that ρ_n not be monotone increasing in n is that the audience agree a priori: $\pi_\gamma > 0.5$ for all $\gamma \in \Gamma$, or else $\pi_\gamma < 0.5$ for all $\gamma \in \Gamma$. In some other problems, ρ_n can be constant on an interval of possible sample sizes n, ρ_n can lack a certain "continuity" (at $n = 0$), and ρ_n can change its monotonicity several times: not just two times.

The experimenter's goal of correct agreement is extended. The experimenter wishes that each observer, in addition to making a correct decision, also satisfies his own (observer's) goal for a low posterior expected loss using that correct decision. Chapter 5 is devoted to this extension in the case of hypothesis testing. There, the extension is reduced to the experimenter's original problem—by replacing each prior with three induced priors, doppelgängers if you will. The audience Γ seemingly increases threefold (page 105). In a special case of this extension, the experimenter wishes that each observer have a large posterior probability:

$$\pi_\gamma(\theta | \underset{\sim}{x}_n) > \zeta \quad \text{for some } \zeta > 0.5 \quad \text{and for all } \gamma \in \Gamma .$$

Some extensions to estimation problems are also made in Chapter 6.

2. TWO-ACTION PROBLEMS

2.1 Formulation of the problem.

We begin our study with two-action problems. Here, two hypotheses H_0, H_1 are under consideration. About these hypotheses, the experimenter can obtain the statistical information of n independent observations $X_1, X_2, \cdots, X_n$. The common (marginal) distribution of these observations is a member of a parametric family of distributions indexed by a parameter θ, $\theta \in \Theta \subseteq \mathbf{R}^1$, and having densities $f(x|\theta)$ with respect to a dominating σ-finite measure $\lambda(x)$ on a sample space $\mathcal{X}$. If hypothesis H_0 is true, the parameter θ belongs to a subset Θ_o of Θ. If hypothesis H_1 is true, θ belongs to $\Theta_1 = \Theta \sim \Theta_o$.

The choice of experimental design in this context reduces to that of choosing the sample size n, $n = 0, 1, 2, \cdots$.

Once the sample size n has been chosen, and n observations $x_1, x_2, \ldots, x_n$ obtained, the likelihood function for θ given the sample $\underset{\sim}{x}_n = (x_1, x_2, \ldots, x_n)$ is

$$f(\underset{\sim}{x}_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \qquad \text{for } \theta \in \Theta.$$

This data is presented to an *audience* (collection) Γ of Bayesian *observers*. Any particular observer γ in Γ is assumed to have a prior density $\pi_\gamma(\theta)$ for θ over Θ relative to a dominating σ-finite measure $\mu(\theta)$. Each such observer may also have a loss function $L_\gamma(a, \theta)$ defined on the action space

$$\mathcal{A} \equiv \{a_0,\ a_1\} = \{\text{choose } H_0, \text{ choose } H_1\}$$

and Θ. However, it is shown in Appendix A that under reasonable assumptions this added structure is unnecessary, and that we can assume without loss of generality that the loss function for all observers is the 0−1 loss function:

$$L(a,\theta) = \begin{cases} 1 & \text{if } a = a_0,\ \theta \in \Theta_1 \ \underline{\text{or}}\ a = a_1,\ \theta \in \Theta_0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

For this loss, it is well known that the Bayes decision for observer γ is to select hypothesis H_i if the posterior probability of H_i exceeds 0.5. That is, if

$$\pi_\gamma\left(H_i|\underset{\sim}{x}_n\right) = \left[m_\gamma(\underset{\sim}{x}_n)\right]^{-1} \int_{\Theta_i} f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\,d\mu(\theta)\,, \qquad \text{for } i = 0, 1, \tag{2.2}$$

where

$$m_\gamma(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\,d\mu(\theta)\,,$$

then observer γ chooses hypothesis H_0 if $\pi_\gamma(H_0|\underset{\sim}{x}_n) > 0.5$, and chooses hypothesis H_1 if $\pi_\gamma(H_1|\underset{\sim}{x}_n) > 0.5$. If $\pi_\gamma(H_0|\underset{\sim}{x}_n) = \pi_\gamma(H_1|\underset{\sim}{x}_n) = 0.5$, observer γ can randomize arbitrarily over the actions a_i = "choose H_i", $i = 0, 1$.

The experimenter wants all observers in his audience Γ to choose the correct hypothesis. However, since the decisions of Bayesians who randomize between a_0 and a_1 are unpredictable, we assume that the experimenter is conservative and excludes cases where an observer reaches a correct decision by randomization. Thus, the experimenter wants the data to belong to one of the following two sets:

$$\begin{aligned} A_i &= \left\{\underset{\sim}{x}_n : \text{all observers in } \Gamma \text{ choose } H_i\right\} \\ &= \left\{\underset{\sim}{x}_n : \pi_\gamma(H_i|\underset{\sim}{x}_n) > 0.5,\ \text{all } \gamma \text{ in } \Gamma\right\}, \qquad \text{for } i = 0, 1. \end{aligned} \tag{2.3}$$

The experimenter has his or her personal prior density $\pi_*(\theta)$ for θ. Thus, the experimenter's probability of obtaining data $\underset{\sim}{x}_n$ leading to agreement of all observers at the correct decision is

$$\rho_n = \sum_{i=0}^{1} \int_{\Theta_i} \int_{A_i} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\,d\lambda(\underset{\sim}{x}_n)\,d\mu(\theta), \tag{2.4}$$

where $\lambda(\underset{\sim}{x}_n) = \prod_{i=1}^n \lambda(x_i)$ is the product measure on $\mathcal{X}^n$ obtained from $\lambda(x)$. For a given probability ϵ, $0 < \epsilon < 1$, the experimenter wishes to choose n such that

$$\rho_n \geq \epsilon. \tag{2.5}$$

When there is more than one sample size n for which (2.5) holds, the experimenter will use the smallest one,

$$N_\epsilon = \min_{n \geq 0} \{n : \rho_n \geq \epsilon\}. \tag{2.6}$$

<u>The Experimenter as an Observer</u>

The experimenter may want to be included as a member of the class Γ of observers. To do this, and yet keep the distinction between experimenter and observer, we can assume that there exists γ in Γ such that $\pi_\gamma(\theta) = \pi_*(\theta)$, all $\theta \in \Theta$.

<u>The Choice $n = 0$</u>

The experimenter has the option of taking no observations. In this case, each observer if pressed to choose would select the hypothesis H_i for which his or her prior odds

$$\pi_\gamma(H_i) = \int_{\Theta_i} \pi_\gamma(\theta)\, d\mu(\theta), \qquad i = 0, 1, \tag{2.7}$$

exceeds 0.5. Thus, these observers agree on a single hypothesis H_i only if (again excluding randomizers) $\pi_\gamma(H_i) > 0.5$ for all $\gamma \in \Gamma$. However, even if all observers agree on hypothesis H_i, the goal (2.5) cannot be achieved by the experimenter for $n = 0$ unless the experimenter's own prior probability

$$\pi_*(H_i) = \int_{\Theta_i} \pi_*(\theta)\, d\mu(\theta) \tag{2.8}$$

exceeds ϵ. Consequently, a necessary and sufficient condition that (2.5) holds for $n = 0$ is that

(2.9) $$\pi_\gamma(H_i) > 0.5\,, \text{ all } \gamma \in \Gamma \quad \text{and} \quad \pi_*(H_i) > \epsilon.$$

for some i, $i = 0, 1$. When this condition is met, the experimenter lets the a priori agreement of his audience stand, and takes no data ($N_\epsilon = 0$).

Obdurate Bayesians

If any observer γ in Γ assigns prior probability 1 to one of the hypotheses, and if the experimenter's prior probability for this hypothesis is less than ϵ, then $\rho_n < \epsilon$ for all $n = 0, 1, \cdots$. To see this, suppose for example that $\pi_{\gamma_0}(H_1) = 1$ for observer γ_0. In this case, observer γ_0 always chooses H_1, regardless of the sample size, and consequently $A_0 = \phi$. Thus, from (2.4),

$$\begin{aligned} \rho_n &= \int_{\Theta_1}\int_{A_1} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \\ &\leq \int_{\Theta_1}\left[\int_{\mathcal{X}^n} f(\underset{\sim}{x}_n|\theta) d\lambda(\underset{\sim}{x}_n)\right] \pi_*(\theta)\, d\mu(\theta) \\ &= \int_{\Theta_1} \pi_*(\theta)\, d\mu(\theta) \;=\; \pi_*(H_1) \end{aligned}$$

and if $\pi_*(H_1) < \epsilon$, it is impossible to have $\rho_n \geq \epsilon$, regardless of the sample size n.

In consequence, it is assumed that no observer in Γ is *obdurate.* That is,

ASSUMPTION 2.1.

(2.10) $$0 < \pi_\gamma(H_0) < 1$$

for all $\gamma \in \Gamma$

(and hence also $0 < \pi(H_1) < 1$, all $\gamma \in \Gamma$). Otherwise, it is impossible for the experimenter to find a design (sample size) such that (2.5) holds.

Although Assumption 2.1 is necessary for the existence of an n such that (2.5) holds, this condition is not sufficient in general. An illustration of this assertion will be given in Chapter 3. See also Section 2.3 of this chapter.

2.2 Some simplifications.

The major difficulties in determining a sample size n satisfying (2.5) are

(i) The integral (2.4) defining ρ_n is multidimensional, with the dimension increasing with n,

(ii) the events A_i defined by (2.3) are possibly infinite intersections

$$A_i = \bigcap_{\gamma \in \Gamma} \left\{ \underset{\sim}{x}_n : \pi_\gamma(H_i | \underset{\sim}{x}_n) > 0.5 \right\}$$

and thus can be irregular in form. (Indeed, A_i may not even be measurable.)

However, in special cases, considerable simplification is possible.

<u>Reduction of Dimensionality</u>

For example, the problem of having the dimensionality of the integral (2.4) increasing with n does not occur if all the functions $\pi_\gamma(H_i|\underset{\sim}{x}_n)$ depend on $\underset{\sim}{x}_n$ only through a p-dimensional vector function $T_n = T_n(\underset{\sim}{x}_n)$ of $\underset{\sim}{x}_n$. In this case

(2.11) $$\pi_\gamma(H_i|\underset{\sim}{x}_n) = Q_{i\gamma}(T_n), \qquad \text{for } i = 0, 1, \quad \gamma \in \Gamma,$$

and

(2.12) $$A_i = \left\{ t : Q_{i\gamma}(t) > 0.5, \text{ all } \gamma \in \Gamma \right\}, \qquad \text{for } i = 0, 1.$$

Thus,

(2.13) $$\rho_n = \sum_{i=0}^{1} \int_{\Theta_i} \int_{A_i} f_n(t|\theta) \pi_*(\theta) \, d\lambda(t) d\mu(\theta) \, ,$$

where the events A_i are now subsets of the p-dimensional range $\mathcal{T}$ ($\mathcal{T}$ is a function of n when X is discrete) of T_n, and where $f_n(t|\theta)$ is the density function (and $\lambda(t)$ is a dominating measure on $\mathcal{T}$) for $T_n = T_n(\underset{\sim}{x}_n)$ obtained from the density $f(\underset{\sim}{x}_n|\theta)$ of the sample $(X_1, \ldots, X_n)$. The integral over t in (2.11) is now p-dimensional regardless of the value of n, $n = 1, 2, \cdots$.

Often, (2.11) holds for some functions $Q_{i\gamma}(\cdot)$ because a p-dimensional sufficient (or Bayesian sufficient) statistic T_n exists for the family $\{f(x|\theta) : \theta \in \Theta\}$. For example, if Θ_0 and Θ_1 consist of single points, $\Theta_i = \{\theta_i\}$ $i = 0, 1$, then

$$T_n = \sum_{i=1}^{n} \log \frac{f(x_i|\theta_1)}{f(x_i|\theta_0)}$$

is a one-dimensional sufficient (Bayesian sufficient) statistic — see Chapter 3. For another example, suppose that $\{f(x|\theta),\ \theta \in \Theta\}$ is a p-parameter exponential family:

$$f(x|\theta) = \exp\{T(x) \cdot c(\theta) + d(\theta) + S(x)\} I_B(x) \,, \tag{2.14}$$

where $a \cdot b$ denotes the inner product of the vectors a, b and $I_B(\cdot)$ denotes the indicator function of the event B. In this case

$$T_n = \sum_{i=1}^{n} T(X_i) \tag{2.15}$$

is a p-dimensional sufficient statistic for θ, and $\eta = c(\theta)$ is the p-dimensional natural parameter of the exponential family.

The following is a special, but interesting, case where (2.11) holds, and yet no sufficient statistic of dimension less than n need exist. First, in general, we may define the conditional prior density of θ given H_i for observer γ:

$$\pi_{i\gamma} = \begin{cases} [\pi_\gamma(H_i)]^{-1}\pi_\gamma(\theta) & , \text{ if } \theta \in \Theta_i \\ 0 & , \text{ if } \theta \notin \Theta_i , \end{cases} \tag{2.16}$$

where $\pi_\gamma(H_i)$ is defined by (2.7), $i = 0, 1$. Thus,

$$\pi_\gamma(\theta) = \pi_\gamma(H_0)\pi_{0\gamma}(\theta) + \pi_\gamma(H_1)\pi_{1\gamma}(\theta)\,. \tag{2.17}$$

Now consider a class of k_0 density functions $u_j^{(0)}(\theta)$, $j = 1, 2, \ldots, k_0$ on Θ_0, and a class of k_1 density functions $u_j^{(1)}(\theta)$, $j = 1, 2, \ldots, k_1$ on Θ_1. Suppose that

$$\pi_{i\gamma}(\theta) = \sum_{j=1}^{k_i} d_{j\gamma}^{(i)} u_j^{(i)}(\theta)$$

where $k_{j\gamma}^{(i)} \geq 0$ all $j = 1, \ldots, k_i$, and $\sum_{j=1}^{k_i} d_{j\gamma} = 1$, $i = 0, 1$. That is, for each observer γ, the conditional density $\pi_{i\gamma}(\theta)$ of θ given that H_i is true is a finite mixture of $u_1^{(i)}(\theta), \ldots, u_{k_i}^{(i)}(\theta)$, $i = 0, 1$. It then follows that

$$\pi_\gamma(H_i|\underset{\sim}{x}_n) = \frac{\pi_\gamma(H_i)\sum_{j=1}^{k_i} d_{j\gamma}^{(i)} m_j^{(i)}(\underset{\sim}{x}_n)}{\sum_{i=0}^{1}\pi_\gamma(H_i)\sum_{j=1}^{k_i} d_{j\gamma}^{(i)} m_j^{(i)}(\underset{\sim}{x}_n)} \tag{2.18}$$

where

$$m_j^{(i)}(\underset{\sim}{x}_n) = \int_{\Theta_i} f(\underset{\sim}{x}_n|\theta)u_j^{(i)}(\theta)\,d\mu(\theta)\,.$$

Let

$$T_{nj} \quad = \quad T_{nj}(\underset{\sim}{x}_n) \quad = \quad \begin{cases} \dfrac{m_j^{(0)}(\underset{\sim}{x}_n)}{m_1^{(1)}(\underset{\sim}{x}_n)} & \text{when } j = 1, \ldots, k_0, \\[2ex] \dfrac{m_{j-(k_0-1)}^{(1)}(\underset{\sim}{x}_n)}{m_1^{(1)}(\underset{\sim}{x}_n)} & \text{when } j = k_0+1, \ldots, k_0+k_1-1. \end{cases}$$

Then it is straightforward to show that $\pi_\gamma(H_i|\underset{\sim}{x}_n)$ satisfies (2.11) with

$$T_n = (T_{n1}, \ldots, T_{n,k_0+k_1-1})'\,.$$

However, $\{f(x|\theta),\ \theta \in \Theta\}$ can be the family of Cauchy densities with location parameter θ; that is,

$$f(x|\theta) = \frac{1}{\pi[1 + (x-\theta)^2]} \qquad \text{on } -\infty < x < \infty.$$

In this case, the order statistics based on the sample $\underset{\sim}{x}_n = (x_1, \ldots, x_n)$ are known to be minimal sufficient. Thus, we have an example of (2.11) being satisfied even when no sufficient statistic of dimension less than n exists.

Further Simplification

Let

$$V_i = V_i(\underset{\sim}{x}_n) = \inf_{\gamma\in\Gamma} \pi_\gamma(H_i|\underset{\sim}{x}_n)\,, \qquad \text{for } i = 0, 1. \tag{2.19}$$

It is easily shown that

$$\tilde{A}_i \equiv \{V_i > 0.5\} \subseteq A_i \subseteq \{V_i \geq 0.5\} \tag{2.20}$$

Assume that V_0, V_1 are measurable functions of $\underset{\sim}{x}_n$, $n = 1, 2, \cdots$. If Γ is a finite set, this assumption is always correct (indeed, in this case $\{V_i > 0.5\} = A_i$). In general, V_0 and V_1 are measurable under relatively weak regularity conditions. However, exposition of such conditions takes us too far from our main theme. It is usually easier to verify measurability directly in each specific application.

If V_0, V_1 are measurable, let

$$G_i^{(n)}(v|\theta) = \mathrm{P}\{V_i \leq v\}\,,$$

be the cumulative distribution function of V_i, $i = 1, 2$, $n = 1, 2, \cdots$, all $\theta \in \Theta$. Corresponding to the set relation (2.20) is the probability inequality

$$\sum_{i=0}^{1} \int_{\Theta_i} \left[1 - G_i^{(n)}(0.5|\theta)\right] \pi_*(\theta)\, d\mu(\theta) \leq \rho_n \leq \sum_{i=0}^{1} \int_{\Theta_i} \left[1 - G_i^{(n)}(0.5-|\theta)\right] \pi_*(\theta)\, d\mu(\theta)\,, \tag{2.21}$$

where

$$G_i^{(n)}(v-|\theta) = \lim_{w\uparrow v} G_i^{(n)}(w|\theta)\,.$$

Notice that V_0, V_1 are scalar random variables. Indeed, since $\pi_\gamma(H_i|\underset{\sim}{x}_n)$ is between 0 and 1 for all $\gamma \in \Gamma$, it follows that

$$0 \leq V_i \leq 1\,, \qquad \text{for } i = 0, 1, \quad n = 1, 2, \cdots.$$

If we can find $G_i^{(n)}(v)$, $i = 0,1$, for all $n = 1,2,\cdots$, (2.21) gives us a computable (by computer, if necessary) way to bound ρ_n. Indeed, finding n such that

$$\text{(2.22)} \qquad \tilde{\rho}_n \equiv \sum_{i=0}^{1} \int_{\Theta_i} P_\theta(\tilde{A}_i)\pi_*(\theta)\, d\mu(\theta) = \sum_{i=0}^{1} \int_{\Theta_i} \left[1 - G_i^{(n)}(0.5|\theta)\right] \pi_*(\theta)\, d\mu(\theta) \geq \epsilon,$$

gives us an upper bound for the value of n needed to achieve (2.5) — that is, to make $\rho_n \geq \epsilon$. When Γ is "closed," $\tilde{\rho}_n = \rho_n$.

The above results look simpler than the original problem, but obscure the fundamental difficulty of actually finding V_0, V_1 for each n and obtaining the cumulative distribution functions of these random variables. Further, as we will see in Chapters 3 and 4, even when both of these tasks can be done, there are still complications in computing $\tilde{\rho}_n$. Further, there are some unexpected and subtle philosophical problems that arise in determining a sample size n.

2.3 A theoretical point.

It is worth noting that in some special cases of two-action problems (notably the case $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$) it is possible to determine prior densities $\delta_0(\theta)$, $\delta_1(\theta)$ on Θ such that

$$\text{(2.23)} \qquad V_i = \frac{\int_{\Theta_i} f(\underset{\sim}{x}_n|\theta)\delta_i(\theta)\, d\mu(\theta)}{\int_{\Theta} f(\underset{\sim}{x}_n|\theta)\delta_i(\theta)\, d\mu(\theta)}, \qquad \text{for } i = 0,1.$$

If there exist γ_0, γ_1 in Γ such that

$$\pi_{\gamma_i}(\theta) = \delta_i(\theta), \qquad \text{for both } i = 0,1,$$

then we can characterize observer γ_0 as being the most difficult observer in Γ to convince about the truth of H_0, and observer γ_1 as being the most difficult observer in Γ to convince about the truth of H_1. If no such observers γ_0, γ_1 exist, we can "close" Γ by adding observers with priors $\delta_0(\theta)$, $\delta_1(\theta)$ to Γ. If Γ is "closed" ($\tilde{A}_i = A_i$; or can be "closed," $\tilde{A}_i \subseteq A_i$) this way,

$$\tilde{A}_i = \left\{V_i > 0.5\right\} = \left\{\delta_i(H_i|\underset{\sim}{x}_n) > 0.5\right\}, \qquad \text{for } i = 0, 1, \tag{2.24}$$

so that the experimenter can ignore all observers in Γ except for the extreme observers with the priors $\delta_0(\theta)$ and $\delta_1(\theta)$. That is, Γ can be treated as if it contains only two members.

Sometimes (again see Chapter 3), no member of Γ is obdurate, but $\delta_0(\theta)$ or $\delta_1(\theta)$ correspond to obdurate observers. That is, we can have $0 < \pi_\gamma(H_0) < 1$, all $\gamma \in \Gamma$, and yet

$$\delta_0(H_0) = 0 \text{ and/or } \delta_1(H_0) = 1\,.$$

If this is the case, the discussion in Section 2.1 shows that it may be impossible for the experimenter to find a sample size n such that $\rho_n \geq \epsilon$.

A generalization of the situation described above is as follows. Assume that for each $n = 1, 2, \cdots$, there exists a finite partition $B_1^{(n)}, \ldots, B_M^{(n)}$ of $\mathcal{X}^n$ and prior densities $\delta_{ij}^{(n)}(\theta)$ on Θ, $i = 0, 1$, $j = 1, \ldots, M$, such that

$$V_i = \frac{\int_{\Theta_i} f(\underset{\sim}{x}_n|\theta)\delta_{ij}(\theta)\, d\mu(\theta)}{\int_{\Theta} f(\underset{\sim}{x}_n|\theta)\delta_{ij}(\theta)\, d\mu(\theta)}, \qquad \text{on } \underset{\sim}{x}_n \in B_j^{(n)}. \tag{2.25}$$

Then, observers with priors $\delta_{ij}(\theta)$ are most difficult to convince about the truth of H_i when data $\underset{\sim}{x}_n$ in $B_j^{(n)}$ are observed. (Note: We can allow M to also depend on n.) If for all $n = 1, 2, \cdots$, all $j = 1, \ldots, M$ and $i = 0, 1$, there exists an observer $\gamma_{ij}^{(n)}$ such that

$$\pi_{\gamma_{ij}^{(n)}}(\theta) = \delta_{ij}^{(n)}(\theta)\,, \qquad \text{all } \theta \in \Theta_i\,,$$

we say that Γ is "compact." Otherwise, we can "compactify" Γ by adding observers with priors $\delta_{ij}^{(n)}$, $i = 0, 1$, $j = 1, \cdots, M$, $n = 1, 2, \cdots$ to Γ. Consequently, if we know that the desired sample size n is no greater than some N, $1 \leq N < \infty$, we can treat Γ as if it were a finite collection of observers, even if originally Γ was assumed to be uncountable.

Although such "compactification" of Γ is of some theoretical interest, it does not really simplify the problem of choosing n, since to verify the existence of priors $\delta_{ij}^{(n)}(\theta)$ satisfying (2.25), we need to either obtain V_i, $i = 0, 1$, or at least know some properties of V_i. (In this sense, our argument is somewhat circular in nature.) However, our discussion does suggest the potential for simplifying two-action problems in which a large collection Γ of observers is hypothesized. We have also indicated in passing that (2.10) is not, in general, enough to insure the existence of an n such that $\rho_n \geq \epsilon$.

2.4 Priors in monotone likelihood ratio families have two extreme observers.

We present a class of problems for which there are two extreme observers with priors $\delta_0(\theta)$ and $\delta_1(\theta)$. We make two assumptions,

ASSUMPTION 2.2 . *The parameter space Θ_0 is to the left of the parameter space Θ_1. That is, $\theta_0 < \theta_1$ when $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$.*

We introduce the notation "$\prec$" to denote a monotone likelihood relation between two densities. That is, for two densities g_1 and g_2,

$$g_1 \prec g_2 \qquad \text{if } \tfrac{g_2(y)}{g_1(y)} \text{ is non-decreasing in y.} \tag{2.26}$$

Similarly, define $g_2 \succ g_1$ when $g_1 \prec g_2$. For an index $\gamma \in B \subseteq \mathbf{R}^1$, we say that $\{g_\gamma, \gamma \in B\}$ forms a *monotone likelihood ratio family* when

$$g_{\gamma_1} \prec g_{\gamma_2} \text{ or } g_{\gamma_1} \succ g_{\gamma_2} \qquad \text{for all } \gamma_1, \gamma_2 \in B. \tag{2.27}$$

We say that this is a <u>*non-decreasing*</u> *monotone likelihood ratio family* when

(2.28) $$g_{\gamma_1} \prec g_{\gamma_2} \qquad \text{for all } \gamma_1 < \gamma_2 \in B.$$

Equivalently,

$$\frac{g_\gamma(y_3)}{g_\gamma(y_2)}$$

is a non-decreasing function of γ whenever $y_2 < y_3$. To ensure that extreme priors exist, we make

ASSUMPTION 2.3.

(a) ***The audience space Γ can be mapped one-to-one into a subset of $\mathbf{R}^1$.***

(b) ***The audience's priors $\{\pi_\gamma,\ \gamma \in \Gamma\}$ form a non-decreasing monotone likelihood ratio family.***

With this assumption, whenever $\theta_2 < \theta_3$,

$$\frac{\pi_\gamma(\theta_3|\underset{\sim}{x}_n)}{\pi_\gamma(\theta_2|\underset{\sim}{x}_n)} = \frac{f(\underset{\sim}{x}_n|\theta_3)}{f(\underset{\sim}{x}_n|\theta_2)}\frac{\pi_\gamma(\theta_3)}{\pi_\gamma(\theta_2)}$$

is a non-decreasing function of γ. So $\{\pi_\gamma(\theta|\underset{\sim}{x}_n),\ \gamma \in \Gamma\}$ forms a non-decreasing monotone likelihood ratio family. [Note: Let $\{f(x|\theta), \theta \in \Theta\}$ form a non-decreasing monotone likelihood ratio family and consider that "$\underset{\sim}{x}_n \le \underset{\sim}{y}_n$" means $x_i \le y_i$ for every $i = 1, 2, \ldots, n$. With Assumption 2.3, the distributions $\pi_\gamma(\theta|\underset{\sim}{x}_n)$ satisfy the formal relation (2.27) for a non-decreasing monotone likelihood ratio family in the three senses involving any pair of the three variables γ, θ, or $\underset{\sim}{x}_n$]. Consequently, for $\gamma_1 < \gamma_2$,

(2.29) $$\pi_{\gamma_2}(\Theta_0) \le \pi_{\gamma_1}(\Theta_0)$$

and

(2.30) $$\pi_{\gamma_2}(\Theta_0|\underset{\sim}{x}_n) \quad \le \quad \pi_{\gamma_1}(\Theta_0|\underset{\sim}{x}_n).$$

Let

$$\gamma_- \equiv \inf_{\gamma \in \Gamma} \gamma \quad \text{and} \quad \gamma_+ \equiv \sup_{\gamma \in \Gamma} \gamma.$$

When Γ is "closed,"

$$\delta_0(\theta) = \pi_{\gamma_+}(\theta) \quad \text{and} \quad \delta_1(\theta) = \pi_{\gamma_-}(\theta) \tag{2.31}$$

and

$$\delta_1 \prec \pi_\gamma \prec \delta_0. \tag{2.32}$$

Whether or not Γ can be "closed,"

$$\lim_{\gamma \to \gamma_+} \pi_\gamma(\Theta_0) \;\le\; \pi_\gamma(\Theta_0) \;\le\; \lim_{\gamma \to \gamma_-} \pi_\gamma(\Theta_0) \qquad \text{for every } \gamma \in \Gamma \tag{2.33}$$

and

$$V_0 \;=\; \lim_{\gamma \to \gamma_+} \pi_\gamma(\Theta_0 | \underset{\sim}{x}_n) \;\le\; \pi_\gamma(\Theta_0 | \underset{\sim}{x}_n) \;\le\; \lim_{\gamma \to \gamma_-} \pi_\gamma(\Theta_0 | \underset{\sim}{x}_n) \;=\; 1 - V_1. \tag{2.34}$$

2.5 An interpretation of the experimenter's goal.

Above, the experimenter sought that the correct hypothesis H_0 or H_1 be chosen for any π_γ, $\gamma \in \Gamma$. Equivalently, he sought that the correct hypothesis H_0 or H_1 be chosen *robustly* — robustly with respect to the prior π. That is, the experimenter can view π as belonging to a class $\{\pi_\gamma,\ \gamma \in \Gamma\}$ of possible priors whether or not such a class derives from an audience. With this interpretation, the sample size N_ϵ will produce a robust sample — not a robust decision procedure per se — with a probability ρ_n of at least ϵ. The experimenter plans for posterior Bayes robustness via experiment.

We now turn to some special cases of two action problems. In Chapter 3, we treat the case where Θ_0 and Θ_1 are both simple ($\Theta_i = \{\theta_i\}$, $i = 0, 1$). In Chapter 4, we consider one-sided testing problems with special attention to the case where $\{f(x|\theta),\ \theta \in \Theta\}$ is a one-parameter exponential family.

3. SIMPLE VS. SIMPLE HYPOTHESIS TESTING

3.1 Introduction.

In the context of the two-action problems considered in Chapter 2, it may be the case that $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$, so that the parameter spaces corresponding respectively to hypotheses H_0, H_1 each contain one point. In this simple vs. simple hypothesis testing situation, each observer's prior distribution $\pi_\gamma(\theta)$ on Θ is determined by a single number

$$\pi_\gamma = \pi_\gamma(H_0) = 1 - \pi_\gamma(H_1), \qquad \text{for } \gamma \in \Gamma.$$

We assume that no observer is obdurate; thus,

$$0 < \pi_\gamma < 1 \, , \qquad \text{for all } \gamma \in \Gamma \, . \tag{3.1}$$

If a sample $\underset{\sim}{x}_n = (x_1, \ldots, x_n)$ is presented to observer γ, that observer calculates his or her posterior probability

$$\pi_\gamma(H_0 | \underset{\sim}{x}_n) \;=\; \frac{\pi_\gamma f(\underset{\sim}{x}_n | \theta_0)}{\pi_\gamma f(\underset{\sim}{x}_n | \theta_0) + (1 - \pi_\gamma) f(\underset{\sim}{x}_n | \theta_1)} \, ,$$

where

$$f(\underset{\sim}{x}_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta) \, , \qquad \text{with } \theta = \theta_0, \, \theta_1 \, .$$

Observer γ decides in favor of action $a_0 =$ "decide H_0 is true" if $\pi_\gamma(H_0 | \underset{\sim}{x}_n) > 0.5$, and decides in favor of action $a_1 =$ "decide H_1 is true" if $\pi_\gamma(H_0 | \underset{\sim}{x}_n) < 0.5$. If $\pi_\gamma(H_0 | \underset{\sim}{x}_n) = 0.5$, the observer can arbitrarily randomize between the two actions.

The experimenter's prior distribution on Θ is determined by $\pi_* = \pi_*(H_0) = 1-\pi_*(H_1)$. From the experimenter's perspective, the probability that all observers in Γ arrive at the correct decision is

$$\rho_n = \pi_* \int_{A_0} f(\underset{\sim}{x}_n|\theta_0)d\lambda(\underset{\sim}{x}_n) + (1-\pi_*)\int_{A_1} f(\underset{\sim}{x}_n|\theta_1)d\lambda(\underset{\sim}{x}_n), \tag{3.2}$$

where

$$\begin{aligned} A_0 &= \left\{\underset{\sim}{x}_n : \pi_\gamma(H_0|\underset{\sim}{x}_n) > 0.5, \quad \text{all } \gamma \in \Gamma\right\}, \\ A_1 &= \left\{\underset{\sim}{x}_n : \pi_\gamma(H_0|\underset{\sim}{x}_n) < 0.5, \quad \text{all } \gamma \in \Gamma\right\}. \end{aligned} \tag{3.3}$$

The simple vs. simple two-action problem permits all the reductions mentioned in Chapter 2, Section 2.2. For example, for $n \geq 1$, a sufficient statistic for $\{f(\underset{\sim}{x}_n|\theta),\ \theta = \theta_0, \theta_1\}$ is

$$T_n = T_n(\underset{\sim}{X}_n) = \sum_{i=1}^{n} \ln\left[\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)}\right] \tag{3.4}$$

and

$$\pi_\gamma(H_0|\underset{\sim}{X}_n) = \frac{1}{1+\left(\frac{1-\pi_\gamma}{\pi_\gamma}\right)e^{T_n}}. \tag{3.5}$$

Also,

$$V_0 = \inf_{\gamma\in\Gamma} \pi_\gamma(H_0|\underset{\sim}{x}_n) = \frac{1}{1+l_0 e^{T_n}}, \tag{3.6}$$

where

$$l_0 = \sup_{\gamma\in\Gamma}\left[\frac{1-\pi_\gamma}{\pi_\gamma}\right] = \frac{1-\inf\limits_{\gamma\in\Gamma}\pi_\gamma}{\inf\limits_{\gamma\in\Gamma}\pi_\gamma},$$

while

$$V_1 = \inf_{\gamma\in\Gamma}\pi_\gamma(H_1|\underset{\sim}{x}_n) = 1-\sup_{\gamma\in\Gamma}\pi_\gamma(H_0|\underset{\sim}{x}_n) = \frac{1}{1+l_1 e^{T_n}}, \tag{3.7}$$

where

$$l_1 = \inf_{\gamma\in\Gamma}\left[\frac{1-\pi_\gamma}{\pi_\gamma}\right] = \frac{1-\sup\limits_{\gamma\in\Gamma}\pi_\gamma}{\sup\limits_{\gamma\in\Gamma}\pi_\gamma}.$$

The random variables V_0, V_1 are the posterior probabilities of H_0, H_1 for two Bayesian observers having prior distributions determined by

(3.8) $$\pi_L = \inf_{\gamma \in \Gamma} \pi_\gamma = \pi_L(H_0) = 1 - \pi_L(H_1)$$

and

(3.9) $$\pi_U = \sup_{\gamma \in \Gamma} \pi_\gamma = \pi_U(H_0) = 1 - \pi_U(H_1),$$

respectively. As discussed in Chapter 2, Section 2.3, we can "compactify" Γ, if necessary, by adding two observers to Γ with prior probabilities π_L, π_U, respectively, for H_0. These two observers are the most extremely opinionated, and the experimenter need only concentrate on these two observers in order to design the experiment. [Note: It is possible that one of these two observers has the same prior probability for H_0 as the experimenter. That is, the experimenter may be a member of Γ, and have one of the two extremes of prior opinion in Γ. As mentioned in Chapter 2, this possibility is easily accommodated by the theory.]

Note that although $0 < \pi_\gamma < 1$ for all $\gamma \in \Gamma$, the collection Γ can be large enough so that there exists a sequence $\{\gamma_i : i = 1, 2, \cdots\}$ of observers for which either

$$\lim_{i \to \infty} \pi_{\gamma_i} = 0\,,$$

or

$$\lim_{i \to \infty} \pi_{\gamma_i} = 1\,,$$

or both. In this case, as discussed in Chapter 2, if the experimenter's own prior probability for H_0 satisfies $\pi_* < \epsilon$ or $\pi_* > 1 - \epsilon$, the experimenter cannot find a sample size n such that $\rho_n \geq \epsilon$. Consequently, we make the following assumption.

ASSUMPTION 3.1 . $0 < \pi_L \leq \pi_U < 1$.

From Assumption 3.1, it follows that

$$0 < l_1 \leq l_0 < \infty. \tag{3.10}$$

It now follows from (3.2), (3.3), (3.6) and (3.7) that

$$\rho_n = \pi_* P_{\theta_0}\{T_n < c_0\} + (1-\pi_*) P_{\theta_1}\{T_n > c_1\}, \tag{3.11}$$

where

$$c_0 = -\ln(l_0) = \ln\left[\frac{\pi_L}{1-\pi_L}\right], \qquad c_1 = -\ln(l_1) = \ln\left[\frac{\pi_U}{1-\pi_U}\right]. \tag{3.12}$$

Note that it follows from (3.10) that

$$-\infty < c_0 \leq c_1 < \infty.$$

3.2 Existence of n such that $\rho_n \geq \epsilon$.

It is easily seen that no observations need to be taken if $\pi_U < \frac{1}{2}$ and $\pi_* \leq 1-\epsilon$, or if $\pi_L > \frac{1}{2}$ and $\pi_* \geq \epsilon$. In the former case, all observers in Γ will choose action a_1 in the absence of data, and the experimenter's probability that this action is the correct one is $\rho_0 = \pi_*(H_1) = 1 - \pi_* \geq \epsilon$. In the latter case, all observers in Γ will choose action a_0 in the absence of data, and the experimenter's probability that this action is the correct one is $\rho_0 = \pi_*(H_0) = \pi_* \geq \epsilon$.

In order that observations provide information about the truth of hypotheses H_0 and H_1, we make the following assumption.

ASSUMPTION 3.2. *The parameterization of $f(x|\theta)$ is identifiable.*

That is, for some measurable set $\mathcal{A} \subset \mathcal{X}$ for which $\lambda(A) > 0$, we have $f(x|\theta_0) \neq f(x|\theta_1)$ for x in $\mathcal{A}$.

As already shown, we can assume without loss of generality that there are only two observers γ_L and γ_U in Γ with respective prior probabilities π_L and π_U

for H_0 (or, equivalently, for θ_0). The experimenter may have prior probability π_* for H_0 included within the interval $[\pi_L, \pi_U]$, or π_* may lie outside of this interval. The latter case has some resemblance to Jackson, Novick and Dekeyrel's (1980) adversarial setting, but the goals are different.

A Frequentist Approach

One way to approach the problem of determining a sample size n such that $\rho_n \geq \epsilon$ is to solve a frequentist problem. That is, find a sample size n as small as possible such that both of the following inequalities hold:

$$(3.13)\qquad \begin{aligned} P_{\theta_0}(A_0) = P_{\theta_0}\{T_n < c_0\} &\geq \epsilon, \\ P_{\theta_1}(A_1) = P_{\theta_1}\{T_n > c_1\} &\geq \epsilon. \end{aligned}$$

It will then immediately follow from (3.11) that $\rho_n \geq \epsilon$.

The above approach is clearly conservative, but has the merit of providing a convenient algorithm for finding n. Suppose that the distribution (particularly the cumulative distribution function) of T_n is known under θ_0, θ_1, for all $n \geq 1$. In this case, one can simply start calculating

$$P_{\theta_0}(A_0) = F_n^{(0)}(c_0-), \qquad P_{\theta_1}(A_1) = 1 - F_n^{(1)}(c_1)$$

for $n = 1, 2, \cdots$, increasing n in steps of 1 until (3.13) is satisfied for the first time (Here, $F_n^{(0)}(\cdot)$ is the c.d.f. of T_n under θ_0, $F_n^{(1)}(\cdot)$ is the c.d.f. of T_n under θ_1). The existence of such an n is guaranteed by the following theorem, which also provides a way of obtaining a value for n when the distribution of T_n is unknown, or difficult to calculate.

THEOREM 3.1. *Let $M_i(t)$ be the moment generating function of*

$$T_1 = \ln[f(x|\theta_1)/f(x|\theta_0)]$$

under θ_i, $i = 0, 1$. *That is,*

$$M_i(t) = \int_{\mathcal{X}} \left[\frac{f(x|\theta_1)}{f(x|\theta_0)}\right]^t f(x|\theta_i) d\lambda(x), \qquad \textit{for } i = 0,1. \tag{3.14}$$

Let

$$M = \inf_{t \geq 0} M_0(t). \tag{3.15}$$

Then $M = 0$ iff $f(x|\theta_0)$ *and* $f(x|\theta_1)$ *have disjoint supports. Otherwise, there is a unique* t_0, $0 < t_0 < 1$, *for which*

$$0 < M = M_0(t_0) = M_1(t_0 - 1) < 1. \tag{3.16}$$

Moreover, if we let

$$N = \begin{cases} 1 & \textit{for } M = 0 \\ (\ln M)^{-1}[\ln(1-\epsilon) + \min\{c_0 t_0, c_1(t_0 - 1)\}] & \textit{otherwise}, \end{cases} \tag{3.17}$$

then $N \geq N_\epsilon$ *and the inequalities in (3.13) hold for all* $n \geq N$.

PROOF OF THEOREM 3.1.

The case for which $f(x|\theta_0)$ and $f(x|\theta_1)$ have disjoint supports is immediate. For $n = 1$, $P_{\theta_0}(A_0) = 1 = P_{\theta_1}(A_1)$ so that (3.13) holds. We suppose for the rest of this proof that the supports of $f(x|\theta_0)$ and $f(x|\theta_1)$ overlap. The function $g(x) = x^t$, $x \geq 0$, is strictly concave on $x > 0$ if $0 < t < 1$. It is strictly convex if $t < 0$ or $t > 1$. Under Assumption 3.2, Jensen's inequality—as in Marshall and Olkin (1979, pg 454)—implies that

$$M_0(t) < M_0^t(1) = 1, \qquad \text{when } 0 < t < 1, \tag{3.18}$$

and that

$$M_0(t) > M_0^t(1) = 1, \qquad \text{when } t < 0 \text{ or } t > 1. \tag{3.19}$$

An argument like Bahadur's (1971, pg 3) shows that $M_0(t)$ is strictly convex for $0 \leq t \leq 1$. As $M_0(0) = 1$, there is a unique

$$0 < t_0 < 1 \tag{3.20}$$

for which $M = M_0(t_0) < 1$. Since we presuppose that $f(x|\theta_0)$ and $f(x|\theta_1)$ have disjoint supports, then $M_0(t) \neq 0$ for finite t. Thus,

$$0 < M = M_0(t_0) < 1 . \tag{3.21}$$

Also,

$$\inf_{t \leq 0} M_1(t) = \inf_{t \leq 0} M_0(t+1) .$$

Considering (3.18), (3.19) and (3.20),

$$\inf_{t \leq 0} M_0(t+1) = M_0[(t_0 - 1) + 1] ,$$

so that

$$\inf_{t \leq 0} M_1(t) = M_1(t_0 - 1) = M_0(t_0) = M . \tag{3.22}$$

Let

$$Y_i = \ln \left[\frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \right] , \qquad \text{for } i = 1, 2, \cdots .$$

Then the $Y_i^{'s}$ are iid with common moment generating function $M_0(t)$ under θ_0 and $M_1(t)$ under θ_1. Further, $T_n = \sum_{i=1}^{n} Y_i$ has moment generating function $M_0^n(t)$ under θ_0 and $M_1^n(t)$ under θ_1. A well known inequality (see Chernoff, 1952) states that if Z has moment generating function $M(t)$ and c is any real number, then for $t \geq 0$

$$\mathrm{P}\{Z \geq c\} \leq e^{-ct} M(t) .$$

Consequently,

$$P_{\theta_0}\{T_n \geq c_0\} \leq \inf_{t>0} e^{-c_0 t} M_0^n(t) . \tag{3.23}$$

Similarly, letting $Z = -T_n$,

$$P_{\theta_1}\{T_n \leq c_1\} = P_{\theta_1}\{-T_n \geq -c_1\} \leq \inf_{t>0} e^{tc_1} M_1^n(-t) ,$$

or equivalently,

$$P_{\theta_1}\{T_n \leq c_1\} \leq \inf_{t<0} e^{-tc_1} M_1^n(t) . \tag{3.24}$$

From (3.21) and (3.23) it follows that

$$P_{\theta_0}\{T_n < c_0\} \;\geq\; 1 - e^{-c_0 t_0} M^n . \tag{3.25}$$

Similarly, from (3.22) and (3.24) it follows that

$$P_{\theta_1}\{T_n > c_1\} \;\geq\; 1 - e^{-c_1 (t_0 - 1)} M^n . \tag{3.26}$$

Since we showed in (3.21) that $0 < M < 1$, then

$$1 - e^{-c_0 t_0} M^n \;\geq\; \epsilon \qquad \text{when } n \geq N$$

and

$$1 - e^{-c_1 (t_0 - 1)} M^n \;\geq\; \epsilon \qquad \text{when } n \geq N .$$

It follows from (3.25) and (3.26) that the inequalities of (3.13) hold when $n \geq N$. □

In Theorem 3.1, M measures the "similarity" of the densities $f(x|\theta_0)$ and $f(x|\theta_1)$. We mentioned in Theorem 3.1 that M attains its minimum value $M = 0$ when and only when these two densities are so different that they have disjoint support. At the other extreme, M reaches its maximum value $M = 1$ when and only when the two densities are not identifiable (are violating Assumption 3.2, which itself led to (3.16) when $M \neq 0$).

Two Direct Approaches

The frequentist approach did not make full use of (3.23) and (3.24) to bound ρ_n in Theorem 3.1. Through (3.11), we could have bounded ρ_n as follows:

$$\rho_n \;\geq\; 1 - \left(\pi_* \inf_{t>0}[e^{-c_0 t} M_0^n(t)] + (1 - \pi_*) \inf_{t<0}[e^{-c_1 t} M_1^n(t)]\right) . \tag{3.27}$$

However, this bound is difficult to compute and the resulting bound on N_ϵ cannot be put in closed form. Further, for large enough n, the magnitude of the lower bound in (3.2.15) is determined by $M_0^n(t)$ and $M_1^n(t)$. These are the terms minimized in Theorem 3.1.

The frequentist approach did not even make full use of the weaker (weaker than (3.23) and (3.24)) bounds (3.25) and (3.26) in Theorem 3.1. Using (3.25) and (3.26) in (3.11), we get

$$\rho_n \geq 1 - \left[\pi_* e^{-c_0 t_0} + (1-\pi_*)e^{-c_1(t_0-1)}\right] M^n . \tag{3.28}$$

Consequently, for

$$N^* \equiv (\ln M)^{-1}\Big(\ln(1-\epsilon) - \ln\left[\pi_* e^{-c_0 t_0} + (1-\pi_*)e^{-c_1(t_0-1)}\right]\Big) \tag{3.29}$$

we have $N_\epsilon \leq N^*$.

For its theoretical interest, and its occasional utility, we present one further way to obtain an upper bound on N_ϵ. Observe that

$$M_0(\tfrac{1}{2}) = M_1(-\tfrac{1}{2}) = \int \left[f(x|\theta_0)f(x|\theta_1)\right]^{1/2} d\lambda(x) = 1 - \tfrac{1}{2}H , \tag{3.30}$$

where

$$H = \int_{\mathcal{X}} \left([f(x|\theta_0)]^{1/2} - [f(x|\theta_1)]^{1/2}\right)^2 d\lambda(x) \tag{3.31}$$

is the Hellinger distance between $f(x|\theta_0)$ and $f(x|\theta_1)$. Using $t = 1/2$ and $t = -1/2$ in the first and second infinums of (3.27), respectively, we get

$$\rho_n \geq 1 - \left[\pi_* e^{-\frac{1}{2}c_0} + (1-\pi_*)e^{\frac{1}{2}c_1}\right](1 - \tfrac{1}{2}H)^n . \tag{3.32}$$

Consequently, for

$$\tilde{N} \equiv [\ln(1-\tfrac{1}{2}H)]^{-1}\Big(\ln(1-\epsilon) - \ln[\pi_* e^{-\frac{1}{2}c_0} + (1-\pi_*)e^{\frac{1}{2}c_1}]\Big) \tag{3.33}$$

we have $N_\epsilon \leq \tilde{N}$.

As M measured the similarity between the densities $f(x|\theta_0)$ and $f(x|\theta_1)$, here the Hellinger distance H measures the dissimilarity. H is 1 when and only when the two densities have disjoint support, and H is 0 when and only when the two densities are not identifiable (Assumption 3.2 is violated).

It is straightforward to show that

$$N^* \leq N\,. \tag{3.34}$$

Since H is more easily calculated than M, $\tilde{N}$ is more easily calculated than N or N^*. Any of these bounds N, N^* or $\tilde{N}$ can serve as a starting point from which a backwards search can be made for N_ϵ.

The next section presents the Gaussian case where $X \sim \mathcal{N}(\theta, 1)$. We show there that $M = 1 - \frac{1}{2}H$, so that

$$N \geq N^* = \tilde{N}$$

in this case.

3.3 An exponential family reduction.

Any simple-simple hypothesis problem is an exponential family problem since the two densities can be written

$$f(y|\eta) \;=\; \exp[C(\eta)T(y) + D(\eta) + G(y)]\,, \tag{3.35}$$

where

$$\begin{aligned} C(\eta) &= \begin{cases} 0 & \text{if } \eta = 0 \\ 1 & \text{if } \eta = 1\,, \end{cases} \\ T(y) &= \ln\left[\frac{f(y|1)}{f(y|0)}\right], \\ D(\eta) &= 0\,, \end{aligned}$$

and

$$G(y) \;=\; \ln[f(y|0)]\,.$$

Whether the exponential family form (3.35) is thus concocted or arises naturally for some other functions C, T, D and G, we can make the following reductions to canonical form. For fixed G and T, $C(\eta)$ determines the density's scalar $D(\eta)$. Since Assumption 3.2 states that the parameterization of $f(y|\eta)$ is identifiable, then $C(0) \neq C(1)$. Denoting $C(0)$ by θ_0 and $C(1)$ by θ_1, w.l.o.g. we may assume that $\theta_0 < \theta_1$. The density now takes the form

$$\exp[\theta T(y) + d(\theta) + G(y)],$$

where $d(\theta)$ is the density's scalar. Denoting $T(y)$ by x, we get the canonical exponential family form

$$\exp[\theta x + d(\theta) + S(x)], \tag{3.36}$$

where $d(\theta)$ determines $S(x)$. Our spaces become

$$\Theta = \{C(\eta_0), C(\eta_1)\} \qquad \text{and} \qquad \mathcal{X} = \{T(y)\}.$$

It is well known for exponential families that our identifiability Assumption 3.2 requires that $\mathcal{X}$ contains at least two elements.

For this canonical form (3.36), we can write the essentials (3.14) and (3.16) for our bounds N and N^*:

$$\text{(3.37)}\ M_i(t) = \exp\Big\{t\big[d(\theta_1) - d(\theta_0)\big] + d(\theta_i) - d\big[\theta_i + t(\theta_1 - \theta_0)\big]\Big\}, \qquad i = 0, 1,$$

and

$$t_0 = [\theta_1 - \theta_0]^{-1}\left\{d'^{-1}\left[\frac{d(\theta_1) - d(\theta_0)}{\theta_1 - \theta_0}\right] - \theta_0\right\}, \tag{3.38}$$

where

$$d'^{-1} = \left\{\frac{d}{d\theta}[d(\theta)]\right\}^{-1}$$

is the inverse function of $d'(\theta) = \frac{d}{d\theta}[d(\theta)]$. Using (3.37) and (3.38), we get M in (3.16):

$$M = \exp\left\{r[d'^{-1}(r) - \theta_0] + d(\theta_0) - d[d'^{-1}(r)]\right\}, \tag{3.39}$$

where

$$r = \frac{d(\theta_1) - d(\theta_0)}{\theta_1 - \theta_0}.$$

Using (3.37), we get in (3.30):

$$(3.40) \qquad 1 - \tfrac{1}{2}H \;=\; M_0(\tfrac{1}{2}) \;=\; \exp\left[\frac{d(\theta_0) + d(\theta_1)}{2} - d\left(\frac{\theta_0 + \theta_1}{2}\right)\right].$$

With these simplifications, we now present the Gaussian example.

3.4 The Gaussian distribution example.

Here we consider observations X_i from a Gaussian distribution, with canonical density

$$(3.41) \qquad f(x|\theta) \;=\; \exp\left\{\theta x - \theta^2/2 - \left[x^2/2 + (\ln\sqrt{2\pi})\right]\right\},$$

where $x \in \mathbf{R}^1$, $\theta = \theta_0$ or θ_1. The standard normal density $f(x|0)$ we denote $\phi(x)$, and its cumulative distribution $\int_{-\infty}^{x} \phi(t)\, dt$ we denote $\Phi(x)$. As a density of the exponential form (3.36), the density (3.41) has

$$(3.42) \qquad d(\theta) = -\theta^2/2.$$

So

$$(3.43) \qquad d'^{-1}(\delta) = d'(\delta) = -\delta.$$

For the derivation of a closed formula for ρ_n, we introduce

$$(3.44) \qquad a \;=\; \frac{\theta_1 - \theta_0}{2},$$

$$(3.45) \qquad b \;=\; -\frac{c_1}{\theta_1 - \theta_0},$$

and

$$(3.46) \qquad c \;=\; -\frac{c_0}{\theta_1 - \theta_0},$$

where

$$c_0 = \ln\left(\frac{\pi_L}{1-\pi_L}\right)$$

and

$$c_1 = \ln\left(\frac{\pi_U}{1-\pi_U}\right)$$

as in (3.12). From their definitions,

(3.47) $$a > 0 \quad \text{and} \quad b \leq c.$$

Let us now find ρ_n. From (2.4),

$$T_n = n(\theta_1 - \theta_0)\left[\overline{x} - \frac{\theta_0+\theta_1}{2}\right].$$

From (3.11),

$$\rho_n = \pi_* P_{\theta_0}\left(\overline{X} < \frac{1}{n}\frac{c_0}{\theta_1-\theta_0} + \frac{\theta_0+\theta_1}{2}\right) + (1-\pi_*)P_{\theta_1}\left(\overline{X} > \frac{1}{n}\frac{c_1}{\theta_1-\theta_0} + \frac{\theta_0+\theta_1}{2}\right).$$

Hence

(3.48) $$\begin{aligned} \rho_n &= \rho_n(\pi_*, a, b, c, n) \\ &= \pi_*\Phi(a\sqrt{n} - c/\sqrt{n}) + (1-\pi_*)[1 - \Phi(-a\sqrt{n} - b/\sqrt{n})]. \end{aligned}$$

Using the symmetry of ϕ, we rewrite ρ_n as

$$\rho_n = (1-\pi_*)\Phi(a\sqrt{n} + b/\sqrt{n}) + (1-\pi_*)[1 - \Phi(-a\sqrt{n} + c/\sqrt{n})].$$

Since ρ_n has the same value for the ordered quintuple (π_*, a, b, c, n) as for $(1-\pi_*, a, -c, -b, n)$, we may assume w.l.o.g. that $c \geq 0$.

We now determine the bounds N, N^*, and $\tilde{N}$. From (3.38), (3.39), (3.40), (3.42) and (3.43),

$$M = 1 - \frac{1}{2}H = \exp\left[-\frac{(\theta_1-\theta_0)^2}{8}\right],$$

and

$$t_0 = \frac{1}{2}.$$

Consequently, from (3.17),

$$N = \frac{-8}{(\theta_1 - \theta_0)^2}\left[\ln(1-\epsilon) + \frac{1}{2}\min\{c_0, -c_1\}\right],$$

while from (3.29) and (3.33),

$$N^* = \tilde{N} = \frac{-8}{(\theta_1 - \theta_0)^2}\left[\ln(1-\epsilon) - \ln(\pi_* e^{-\frac{1}{2}c_0} + (1-\pi_*)e^{\frac{1}{2}c_1})\right].$$

Table 3.1 presents several Gaussian examples. In the body of this table, ρ_n is computed using (3.48). Consider the first example, where the experimenter is a member of the audience. When $\epsilon = 0.95$, the experimenter should choose a sample of size $n = 5$. Table 3.2 presents this sample size N_ϵ and its bounds N, N^*, $\tilde{N}$. In the second example, the priors π_L and π_U are closer together, while the parameters θ_0 and θ_1 are farther apart. Both of these changes contribute to larger ρ_n values. In the third example, the priors are further from each other than in either of the first two examples, and these priors have values symmetrical about 0.5. With the parameters θ_0 and θ_1 closer to each other than in either of the first two examples, the value of ρ_n is smaller for any sample size n. A larger difference between the parameters θ_0 and θ_1 in the fourth example again results in larger ρ_n for any sample size n. Just one datum contributes a great amount to the audience's agreement here.

The only difference between the second and the fifth examples is a smaller difference between the parameters in the fifth example. As expected, at each sample size n, ρ_n is larger in the fifth than in the second example. However, a larger sample size need not give a larger probability that all observers will correctly agree: it need not give a larger ρ_n! The experimenter would rather take no sample, $\rho_n = 0.800$, than to let his audience see the data from a sample of size $n = 40$, $\rho_n = 0.793$. When $\epsilon = 0.95$, the experimenter should sample $n = 266$ observations so that ρ_n is at least 0.95.

In the sixth example, the parameters θ_0 and θ_1 are far enough apart that any small sample makes little contribution to the audience's agreement. The experimenter must plan to sample all of $n = 29,028$ observations so that his audience will correctly agree with a probability as high as 0.95. With the priors π_L and π_U further apart in the seventh example, small samples contribute even less to the audience's agreement than in the sixth example. But with the parameters θ_0 and θ_1 further apart in the seventh example, large samples contribute more to the audience's agreement than in the sixth example. Thus, a sample of but 974 gets $\rho_n \geq 0.95$. The eighth example, like the fourth, presents an audience whose priors have values symmetrical about 0.5.

The experimenter uses $N_\epsilon = 0$ in the last example. In Table 3.2, the large bounds N^* and N for this last example are not germane since the experimenter always considers whether $\rho_0 \geq \epsilon$ before considering sampling data, as discussed in Chapter 2. Also observe that while $\rho_0 > \epsilon$, the probability $\rho_n < \epsilon$ for $n = 6$ through $n = 10,000$ (when $\epsilon = 0.90$)!

This non-monotonicity of ρ_n is reflected more simply in the probability that a single Bayesian correctly chooses H_1. By an argument like that which led to ρ_n in (3.48), if a Bayesian γ_1 has prior π_{γ_1}, then

$$(3.49) \qquad \begin{aligned} P_{\theta_1}(\gamma_1 \text{ chooses } H_1) &= P_{\theta_1}\left[T_n > \ln\left(\tfrac{\pi_{\gamma_1}}{1-\pi_{\gamma_1}}\right)\right] \\ &= 1 - \Phi(-a\sqrt{n} - b/\sqrt{n}), \end{aligned}$$

where a and b are as in (3.44) and (3.45), respectively, with

$$(3.50) \qquad c_1 = \ln[\pi_{\gamma_1}/(1-\pi_{\gamma_1})]$$

in the expression for b. This probability (3.49) is smaller for a sample of size $n = 1$ than for $n = 0$ whenever $b > a$: whenever

$$(3.51) \qquad \pi_{\gamma_1} < \left\{1 + \exp\left[\frac{(\theta_1 - \theta_0)^2}{2}\right]\right\}^{-1}$$

Table 3.1 ρ_n for simple hypotheses, where X has a Gaussian distribution

π_L	.1(.9)	.1(.9)	.1	.1	.1	.45000	.0001	.3775	.48
π_U	.8(.2)	.2(.8)	.9	.9	.2	.50025	.5500	.6225	.49
π_*	.8(.2)	.2(.8)	.1	.1	.2	1.00000	1.0000	any	.01
$\lvert\theta_1 - \theta_0\rvert$	2.0(2.0)	19.0(19.0)	.6	3.0	.2	.02000	.2000	2.0000	.01
n									
0	.000	.8	.000	.000	.800	.000	.000	.000	.990
1	.493	>.999	.000	.779	.800	.000	.000	.773	.990
2	.755		.015	.946	.800	.000	.000	.892	.988
3	.873	.	.055	.985	.800	.000	.000	.944	.980
4	.931		.109	.996	.800	.000	.000	.970	.968
5	.962	.	.167	.999	.800	.000	.000	.983	.954
6	.979		.224	>.999	.799	.000	.000	.991	.941
7	.988	.	.277		.798	.000	.000	.995	.927
8	.993		.328	.	.798	.000	.000	.997	.914
9	.996		.374		.796	.000	.000	.998	.902
10	.998		.417	.	.795	.001	.000	.999	.891
20	>.999		.699		.786	.014	.000	>.999	.813
30			.835	.	.787	.038	.000		.769
40	.		.906		.793	.064	.000	.	.740
50			.946		.803	.089	.000		.720
100	.		.996		.856	.183	.000	.	.669
200		.	>.999	.	.925	.285	.033		.635
300	.				.960	.342	.177	.	.622
400			.		.978	.381	.381		.616
500					.988	.411	.570		.612
1000			.		.999	.500	.956		.611
5000					>.999	.714	>.999		.659
10,000			.			.816			.705
100,000						.999			.944

Table 3.2 N_ϵ and its bounds, where X has a Gaussian distribution

π_L	.1(.9)	.1(.9)	.1	.1	.1	.45000	.0001	.3775	.48
π_U	.8(.2)	.2(.8)	.9	.9	.2	.50025	.5500	.6225	.49
π_*	.8(.2)	.2(.8)	.1	.1	.2	1.00000	1.0000	any	.01
$\lvert\theta_1 - \theta_0\rvert$	2.0(2.0)	19.0(19.0)	.6	3.0	.2	0.02000	.2000	2.0000	.01
ϵ	.95	.95	.95	.95	.95	.95	.95	.90	.90
N_ϵ	5	1	52	3	266	29,028	974	4	0
$N^* = \tilde{N}$	9	1	91	4	600	61,922	1521	6	182,657
N	9	1	91	4	819	61,922	1521	6	187,409

(using (3.44), (3.45), and (3.50)). A final comment on this single observer example relates back to our experimenter's problem. Should $\pi_* = 1$ and the audience Γ contain just one observer γ_1 whose prior satisfies (3.51), then $\rho_0 > \rho_1$. While ρ_n need not be an increasing function of n for a singleton audience Γ, Theorem B.1 in Appendix B says that ρ_n must increase when that single observer is the experimenter himself—if he has 0-1 loss.

To better investigate the monotonicity of ρ_n, the use of (3.48) leads to

$$(3.52)\quad \frac{d\rho_n}{dn} = \pi_*(an+c)\exp[ac - c^2/2n] + (1-\pi_*)(an-b)\exp[-ab - b^2/2n]\,.$$

The following theorem indicates when ρ_n is monotone.

THEOREM 3.2. *For Gaussian X, should $\rho_n < \rho_{n+1}$ then $\rho_{n+m} < \rho_{n+m+1}$ for every $m \geq 0$.*

PROOF OF THEOREM 3.2.

Consider $c \geq 0$. The derivative (3.52) is positive iff

$$\exp\left[a(b+c) - \frac{c^2-b^2}{2n} > \frac{-an+b}{an+c}\left(\frac{1-\pi_*}{\pi_*}\right)\right],$$

for which the left side is strictly increasing and the right side is strictly decreasing. Consequently, once $d\rho_n/dn$ is positive it remains positive: ρ_n always increases after it first increases. Recall that ρ_n in (3.48) has the same value if the ordered quintuple $(1-\pi_*, a, -c, -b, n)$ replaces (π_*, a, b, c, n). Thus, our argument holds for $c < 0$ also. □

Theorem 3.2 shows that the function ρ_n can dip below ϵ, $\rho_{n-1} \geq \epsilon > \rho_n$, but once as n increases. Consequently, either $N_\epsilon = 0$ or else $\rho_n \geq \epsilon$ for any $n \geq N_\epsilon > 0$. Accordingly, if $\rho_0 < \epsilon$ then the experimenter can find N_ϵ by decreasing n from N_ϵ until he finds a sample size m such that $\rho_m < \epsilon$. The experimenter's choice for the sample size is then $N_\epsilon = m + 1$. Any of N, N^*, or $\tilde{N}$ could be used to start this backward search for N_ϵ.

Theorem 3.2 told when ρ_n is monotone. The following theorem tells when ρ_n is not monotone.

THEOREM 3.3. *For Gaussian X, (i), (ii), and (iii) below are necessary conditions that* $\rho_n > \rho_{n+1}$ *for some* $n \geq 0$.

(i)
$$(\pi_U - 0.5)(\pi_L - 0.5) \; > \; 0\,,$$
so that the audience agrees a priori $(b \cdot c > 0)$;

(ii)
$$\pi_* \begin{cases} < \; \frac{\Phi(-a-b)}{\Phi(a-c)+\Phi(-a-b)} & \text{if } \pi_L < 0.5 \text{ (with (i): } b,c > 0\text{)} \\ > \; \frac{1-\Phi(-a-b)}{2-\Phi(a-c)-\Phi(-a-b)} & \text{if } \pi_L > 0.5 \text{ (with (i): } b,c < 0\text{)}; \end{cases}$$
and

(iii)
$$n \; \leq \; \begin{cases} 1 + \min\{\frac{b}{a}, e_0\} & \text{if } \pi_L < 0.5 \\ 1 + \min\{-\frac{c}{a}, e_1\} & \text{if } \pi_L > 0.5, \end{cases}$$
where
$$e_j \; = \; \begin{cases} \frac{(-1)^j(c^2-b^2)}{2a(b+c)+2\ln[\pi_*/(1-\pi_*)]} & \text{if this is positive} \\ \left(\frac{b}{a}\right)^{1-j}\left(-\frac{c}{a}\right)^j & \text{otherwise,} \end{cases}$$
for $j = 0, 1$. *Moreover, when condition (i) holds,* $\rho_n > \rho_{n+1}$ *for some nonnegative integer* n iff *(ii) holds.*

PROOF OF THEOREM 3.3.

(i) From (3.47), $b \leq c$. Should (i) fail then $b \leq 0 \leq c$. Accordingly, $d\rho_n/dn > 0$ in (3.52).

(ii) Assume that (i) holds. Suppose first that $c > 0$, then $b > 0$. As a result, $\rho_0 = 1 - \pi_*$. Because ρ_n increases once it first does (Theorem 3.2), then ρ_n decreases for some n iff $\rho_1 - \rho_0 < 0$. From (3.48),

$$\begin{aligned} \rho_1 - \rho_0 &= \pi_*\Phi(a-c) + (1-\pi_*)[1-\Phi(-a-b)] - (1-\pi_*) \\ &= \pi_*\Phi(a-c) - (1-\pi_*)\Phi(-a-b). \end{aligned}$$

Thus, ρ_n decreases for some n iff

$$\pi_* < \frac{\Phi(-a-b)}{\Phi(a-c)+\Phi(-a-b)}.$$

Suppose second that $c < 0$, then $b < 0$. As a result, $\rho_0 = \pi_*$. From (3.48),

$$\begin{aligned}\rho_1 - \rho_0 &= \pi_*\Phi(a-c) + (1-\pi_*)[1-\Phi(-a-b)] - \pi_* \\ &= -\pi_*[1-\Phi(a-c)] + (1-\pi_*)[1-\Phi(-a-b)].\end{aligned}$$

Thus, ρ_n decreases for some n iff

$$\pi_* > \frac{1-\Phi(-a-b)}{2-\Phi(a-c)-\Phi(-a-b)}.$$

(iii) Suppose that $c > 0$. From the necessary condition (i), $0 \leq b \leq c$. Since the first term of (3.52) is positive, ρ_n can decrease only if the second term is negative: $an - b < 0$ or $n < b/a$. Also, ρ_n can decrease only if the magnitude of the second term is larger than that of the first term. As $0 \leq b \leq c$ implies that $|an + c| > |an - b|$, necessarily

$$\pi_* \exp(ac - c^2/2n) < (1-\pi_*)\exp(-ab - b^2/2n). \tag{3.53}$$

That is,

$$n < (c^2 - b^2)\left[2a(b+c) + 2\ln\left(\frac{\pi_*}{1-\pi_*}\right)\right]^{-1}$$

if this is positive. If it is negative, then $\pi_* e^{ac} < (1-\pi_*)e^{-ab}$ so that (3.52) holds for all n. The case $c < 0$ is handled similarly. □

In (3.48), consider ρ_n defined for real $n \in [0, +\infty)$ instead of for the integer sample sizes. Viewed this way, ρ_n in (3.48) is continuous for $n \in [0, +\infty)$. Whenever (i) of Theorem 3.3 holds, the proof of that theorem shows that the function ρ_n always decreases for some real n—possibly $n \in (0, 1)$.

The experimenter can have a smaller probability at a larger sample size that his audience will correctly agree ($\rho_m > \rho_n$, for $m < n$). Curiously, condition (i) of Theorem 3.3 implies that this can occur when the audience agrees a priori. And never when the audience disagrees a priori! Moreover, should one more observer, with $\pi_\gamma = 0.5$, be in the experimenter's audience, then larger sample sizes give only larger probabilities of correct agreement—ρ_n only increases. On the other hand, at each sample size n the augmented audience Γ (with a $\pi_\gamma = 0.5$) has a smaller probability of a correct agreement ρ_n than the unaugmented audience. When ρ_n is not monotone, the audience Γ must agree a priori (Theorem 3.3(i)), yet ρ_0 can still be less than ϵ. This occurs because

(a) the audience can agree to an incorrect hypothesis a priori.

Specifically, it occurs because

(b) the experimenter has an a priori probability less than ϵ for the hypothesis agreed upon a priori by the audience.

3.5 The exponential distribution example.

This example will bring three new aspects to ρ_n:

(a) As a function on the nonnegative reals, ρ_n need not be continuous at $n = 0$,

(b) ρ_n can be constant for several n,

(c) ρ_n can change its monotonicity twice.

Here, we consider observations X_i having exponential distributions $f(x|\theta)$. In canonical form,

$$f(x|\theta) = \exp[x\theta + \ln(-\theta)]\,, \tag{3.54}$$

for $x > 0$, $\theta = \theta_0$ or θ_1, and $\theta < 0$. Again, we may assume without loss of generality that $\theta_0 < \theta_1$, so that $0 < \theta_1/\theta_0 < 1$. From the canonical form (3.54), we infer through (3.36) that

$$d(\theta) = \ln(-\theta) \tag{3.55}$$

and

$$d'^{-1}(\delta) = 1/\delta\,. \tag{3.56}$$

We now find ρ_n. Let

$$z_{0,n} = \max\Big\{0,\ \ (1-\theta_1/\theta_0)^{-1}[c_0 - n\ln(\theta_1/\theta_0)]\Big\}\,, \tag{3.57}$$

$$z_{1,n} = \max\Big\{0,\ \ [(\theta_1/\theta_0)^{-1} - 1]^{-1}[c_1 - n\ln(\theta_1/\theta_0)]\Big\}\,, \tag{3.58}$$

and

$$\Gamma_z(n) = \int_0^Z \frac{t^{n-1}e^{-t}}{\Gamma(n)}\,dt\,, \qquad \text{for } Z \geq 0\,, \tag{3.59}$$

which is the incomplete gamma function. We say that $Y \sim Gamma(\alpha,\beta)$ when the density of Y has the general form

$$g(y|\alpha,\beta) = \frac{\beta^\alpha y^{\alpha-1}e^{-\beta y}}{\Gamma(\alpha)}\,.$$

Since $X_i \sim Exponential(\theta)$, then

$$X_i \sim Gamma(1,|\theta|) \qquad \text{and} \qquad |\theta|\,n\overline{X} \sim Gamma(n,1)\,. \tag{3.60}$$

From (3.4) and (3.54),

$$T_n = n\Big[(\theta_1-\theta_0)\overline{X} + \ln(\theta_1/\theta_0)\Big]\,. \tag{3.61}$$

From (3.11), (3.57), (3.58) and (3.61),

$$\rho_n = \pi_* P_{\theta_0}\Big(n\,|\theta_0|\,\overline{X} < z_{0,n}\Big) + (1-\pi_*)P_{\theta_1}\Big(n\,|\theta_1|\,\overline{X} > z_{1,n}\Big)\,, \qquad \text{for } n = 1,2,\cdots\,.$$

Because of (3.59) and (3.60), we can rewrite this

$$\rho_n = \pi_*\Gamma_{z_{0,n}}(n) + (1-\pi_*)[1-\Gamma_{z_{1,n}}(n)]\,, \qquad \text{for } n = 1,2,\cdots\,. \tag{3.62}$$

Table 3.3 ρ_n for simple hypotheses, where X has an Exponential distribution

π_L	.2	.1	.8	.3775	.51	.4999995	.5	.502	.5	.45
π_U	.9	.9	.9	.6225	.52	.9	.7	.7	.7	.49
π_*	.2	.9	.8	.6	.99	1	1	1	.7	.50
θ_1/θ_0	.5	.6	.2	.5	.01	.999	.99	.999	.999	.99
n										
0	0	0	.800	0	.990	0	0	1	0	.5
1	.044	.002	.859	.313	.990	.632	.634	>.999	.443	.5
2	.102	.005	.904	.492	.999	.594	.597	>.999	.416	.5
3	.193	.008	.935	.576	>.999	.577	.580	.999	.404	.5
4	.276	.012	.956	.634		.567	.570	.998	.397	.500
5	.347	.018	.970	.678	.	.560	.564	.996	.392	.498
6	.408	.041	.979	.713		.554	.559	.994	.388	.492
7	.462	.079	.985	.743	.	.551	.556	.992	.386	.483
8	.510	.123	.990	.768		.547	.553	.990	.383	.474
9	.552	.169	.993	.790	.	.545	.550	.987	.381	.466
10	.590	.213	.995	.809		.542	.548	.985	.380	.458
20	.819	.540	>.999	.918		.530	.539	.952	.371	.408
30	.915	.715		.961		.525	.535	.921	.368	.383
40	.958	.817	.	.981		.522	.534	.893	.366	.368
50	.979	.880		.990		.520	**.533**	.870	.364	.357
100	>.999	.983	.	> .999		.515	.533	.793	.361	**.341**
200		>.999				.512	.538	.722	.359	.357
300	.		.	.		.511	.542	.686	.358	.379
400		.				.511	.547	.664	.357	.398
500	.			.		**.510**	.551	.649	**.357**	.413
1000		.				.510	.567	.610	.357	.464
5000	.			.		.516	.641	.561	.361	.594
10,000		.				.521	.694	**.553**	.365	.663
100,000						.563	.944	.573	.396	.941
1,000,000						.692	>.999	.694	.594	>.999
10,000,000						.943		.943	.932	
ϵ	.95	.95	.95	.95	.95	.95	.95	.95	.95	.95
N_ϵ	38	73	4	27	0	10,810,034	0	0	11,945,995	110,439
N^*	66	129	9	55	2	23,941,903	237,265	23,909,922	25,115,831	240,598
$\tilde{N}$	69	127	11	56	2	23,941,903	237,265	23,909,926	25,116,054	240,590
N	67	129	13	55	2	32,720,540	270,762	27,327,132	27,327,132	245,225

While ρ_n in (3.48) depended on θ_0 and θ_1 only through the difference of location parameters, $|\theta_1 - \theta_0|$, in the Gaussian case, ρ_n in (3.62) depends on θ_0 and θ_1 only through the ratio of the scale parameters, θ_1/θ_0, in this case.

We now seek the bounds N, N^* and $\tilde{N}$ on N_ϵ. From (3.37), (3.38), (3.55) and (3.56),

$$M = \left(\frac{\theta_1}{\theta_0} - 1\right)^{-1} \ln\left(\frac{\theta_1}{\theta_0}\right) \exp\left[1 - \left(\frac{\theta_1}{\theta_0} - 1\right)^{-1} \ln\left(\frac{\theta_1}{\theta_0}\right)\right]$$

and

$$t_0 = \left[\ln\left(\frac{\theta_1}{\theta_0}\right)\right]^{-1} - \left[\frac{\theta_1}{\theta_0} - 1\right]^{-1}.$$

Using both M and t_0 in (3.17) and in (3.29), we get N and N^*, respectively. From (3.40) and (3.55),

$$1 - \frac{1}{2}H \;=\; -\frac{2\sqrt{\theta_0\theta_1}}{\theta_0 + \theta_1} \;=\; 2\left(1 + \frac{\theta_1}{\theta_0}\right)^{-1}\sqrt{\frac{\theta_1}{\theta_0}}.$$

Using $1 - \frac{1}{2}H$ in (3.33), we get the bound $\tilde{N}$.

We present ρ_n and the bounds on N_ϵ in Table 3.3. In the first five examples, ρ_n is strictly increasing. With $\pi_* = 1$ in the sixth, seventh, and eighth examples of Table 3.3, ρ_n in (3.62) and thus in Table 3.3 has the same value if π_U has any value $\pi_U \geq \pi_L$. Although, in the sixth example $\rho_0 = 1$ if $\pi_L \leq \pi_U < 0.5$. In this sixth example, whatever be the functional form of ρ_n on $(0, +\infty)$, ρ_0 can separately be 0 or 1 as $\pi_U \geq 0.5$ or $\pi_L \leq \pi_U < 0.5$, respectively. A forteriori, as a function on the reals $[0, +\infty)$, ρ_n is not continuous at 0 from the right. Because of this, $n = 0$ begins one monotonicity change of ρ_n. As a result, this sixth example presents a ρ_n with two changes of monotonicity. While the last section's Gaussian example had a ρ_n both right continuous at $n = 0$ and limited to one monotonicity change, our exponential distribution example need have neither property.

The sixth and seventh examples present two changes of monotonicity when π_L is near 0.5. The eighth example is identical to the sixth excepting that π_L is a little larger than 0.5. But one change of monotonicity results. The ninth example is similar to the previous three, having $\pi_* < 1$ though. It presents two changes of monotonicity.

While the Gaussian example was strictly monotone wherever it was monotone, this exponential distribution example need not be. Consider the last example of Table 3.3. In (3.57) and (3.58), $z_{0,n}$ and $z_{1,n}$ are both 0 for $n = 1, 2, 3$. Consequently, from (3.62), $\rho_0 = \rho_1 = \rho_2 = \rho_3 = 0.5$ exactly. For $n \leq 3$, whatever might be the sample $\underset{\sim}{x}_n$, every observer $\gamma \in \Gamma$ would choose but one hypothesis, H_0.

Examples like the sixth show that backstepping n from bounds on N_ϵ until $\rho_n < \epsilon$ may not produce N_ϵ. While this method produces a smaller bound on N_ϵ than N^*, $\tilde{N}$ or N, ρ_n must be compared with ϵ—for every positive integer n less than an N_ϵ bound—to be sure that N_ϵ has been found.

This exponential distribution example presented three features of ρ_n not seen in the Gaussian example. We now present a discrete example for which ρ_n has more monotonicity changes than seen in either the Gaussian or the exponential distribution example.

3.6 A discrete distribution example.

Suppose that the sample space $\mathcal{X}$ of observations has but three members: $\mathcal{X} = \{b_1, b_2, b_3\}$. Also, for $\theta = \theta_0$ suppose that the sample density is specified by

$$f(b_1|\theta_0) = f(b_2|\theta_0) = f(b_3|\theta_0) = 1/3, \tag{3.63}$$

and for $\theta = \theta_1$

$$f(b_1|\theta_1) = 0.8, \qquad f(b_2|\theta_1) = 0.1, \qquad f(b_3|\theta_1) = 0.1. \tag{3.64}$$

Consider

$$\begin{aligned} & P_{\theta_0}\{T_n < c_0\} \\ = \; & P_{\theta_0}\left\{\sum_{i=1}^{n} \ln\left[\frac{f(x_i|\theta_1)}{f(x_i|\theta_0)}\right] < c_0\right\} \\ \doteq \; & P_{\theta_0}\left\{\sum_{i=1}^{n} \ln\left[3f(x_i|\theta_1)\right] < c_0\right\}. \end{aligned} \tag{3.65}$$

We seek to write this probability as an easily recognized distribution. Let

$$y_1 = \ln[3f(b_1|\theta_1)] = \ln(2.4)$$

and

$$y_2 = \ln[3f(b_2|\theta_1)] = \ln[3f(b_3|\theta_1)] = \ln(0.3).$$

Let Y be a binomial random variable for a sample of size n of Bernoulli random variables, each with the distribution

$$\begin{cases} 1 & p = P_{\theta_0}(y_1) = P_{\theta_0}(X = b_1) = 1/3 \\ 0 & 1 - p = P_{\theta_0}(y_2) = P_{\theta_0}(X = b_2 \text{ or } b_3) = 2/3 . \end{cases}$$

Since $y_2 < 0 < y_1$, (3.65) for $\underset{\sim}{X}$ is equivalent to the binomial probability for Y that Y has any value k, $0 \leq k \leq n$, for which

(3.66) $$ky_1 + (n-k)y_2 < c_0 .$$

Letting

$$[\![r]\!]$$

denote the largest integer strictly smaller than r, (3.66) can be written

(3.67) $$k \leq d_0 \equiv \left[\!\!\left[\frac{c_0 - ny_2}{y_1 - y_2} \right]\!\!\right] .$$

Thus,

(3.68) $$P_{\theta_0}(T_n < c_0) = \sum_{k=0}^{d_0} \binom{n}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{n-k} .$$

Similarly,

(3.69) $$P_{\theta_1}(T_n > c_1) = \sum_{k=d_1}^{n} \binom{n}{k} (0.8)^k (0.2)^{n-k} ,$$

where

(3.70) $$d_1 = -\left[\!\!\left[-\frac{c_1 - ny_2}{y_1 - y_2} \right]\!\!\right] .$$

With (3.11), (3.68) and (3.69), we arrive at

(3.71) $$\rho_n = \pi_* \sum_{k=0}^{d_0} \binom{n}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{n-k} + (1-\pi_*) \sum_{k=d_1}^{n} \binom{n}{k} (0.8)^k (0.2)^{n-k} .$$

Now to get the bounds for N_ϵ. From (3.14),

$$M_0(t) = \left[\frac{0.8}{1/3}\right]^t\left(\frac{1}{3}\right) + \left[\frac{0.1}{1/3}\right]^t\left(\frac{1}{3}\right) + \left[\frac{0.1}{1/3}\right]^t\left(\frac{1}{3}\right),$$

that is,

$$M_0(t) = \frac{1}{3}\left[(2.4)^t + 2(0.3)^t\right]. \tag{3.72}$$

Through its derivative, we can show that $M_0(t)$ is minimized at

$$t_0 = (\ln 8)^{-1}\ln\left[-2(\ln 0.3)\big/(\ln 2.4)\right]. \tag{3.73}$$

From (3.30),

$$1 - \tfrac{1}{2}H = M_0\left(\tfrac{1}{2}\right) = (\sqrt{1.2} + \sqrt{2.4})/3. \tag{3.74}$$

Equations (3.16), (3.17), (3.72) and (3.73) specify the bound N. Equations (3.16), (3.29), (3.72) and (3.73) specify the bound N^*. Equations (3.33) and (3.74) specify the bound $\tilde{N}$.

Table 3.4 presents several computations of ρ_n. N_ϵ and its bounds are given at the bottom of this table.

An explanation about Table 3.4: when π_L and π_U are both on the same side of 0.5, the first three examples have $N_\epsilon = 0$. In these three examples, the probabilities ρ_n of Table 3.4 are still correct when $\pi_* = \pi_L = 0$ or $\pi_* = \pi_U = 1$ (violating Assumption 3.1) while the restrictions on π_L and π_U, given in Table 3.4, are met.

The reader might have gathered from the Gaussian and the exponential distribution examples that, starting from $n = 1$, ρ_n can change its monotonicity but once. This conception is quickly routed by a glance at any Table 3.4 example. A forteriori, in the third example ρ_n changes its monotonicity for every n from $n = 1$ to $n = 8$.

If the sample space $\mathcal{X}$ had contained three distinct elements, $b_1 \neq b_2 \neq b_3 \neq b_1$, the computations of ρ_n would have been much more difficult: the computation

Table 3.4 ρ_n for simple hypotheses, where X has a discrete distribution

π_L	$\leq .55$	$\leq .99$	.1	.49	.05	.9	.3
π_U	.55	.99	$\geq .1$	.51	.95	.99	.9
π_*	0	0	1	.5	.5	.5	.9
ϵ	.95	.95	.95	.95	.95	.95	.95
n							
0	0 or 1	0 or 1	0 or 1	0	0	.5	0
1	.800	0	0	.733	0	.500	.600
2	.640	0	.444	.764	0	.500	.400
3	.896	0	.296	.818	.148	.481	.718
4	.819	0	.593	.854	.304	.494	.574
5	.942	0	.461	.866	.394	.477	.785
6	.901	.262	.680	.900	.668	.622	.875
7	.852	.210	.571	.903	.574	.601	.802
8	.944	.503	.741	.928	.622	.742	.901
9	.914	.436	.855	.936	.694	.714	.843
10	.967	.678	.787	.945	.732	.829	.919
11	.950	.617	.878	.955	.775	.804	.874
12	.927	.558	.822	.957	.808	.777	.920
13	.970	.747	.896	.968	.899	.869	.959
14	.956	.698	.942	.969	.860	.847	.935
15	.982	.836	.912	.976	.874	.914	.966
16	.973	.798	.950	.979	.896	.897	.947
17	.989	.758	.925	.981	.909	.875	.972
18	.984	.867	.957	.985	.953	.932	.982
19	.977	.837	.935	.976	.934	.918	.972
20	.990	.913	.962	.989	.965	.955	.985
21	.986	.891	.979	.989	.951	.945	.977
22	.994	.944	.967	.991	.956	.970	.988
23	.991	.928	.981	.992	.962	.963	.980
24	.996	.911	.972	.993	.968	.954	.989
25	.994	.953	.984	.994	.983	.976	.993
26	.992	.941	.975	.995	.976	.970	.989
27	.997	.970	.986	.996	.978	.984	.994
28	.995	.961	.992	.996	.982	.980	.991
29	.998	.951	.987	.997	.984	.975	.995
30	.997	.974	.993	.997	.986	.987	.993
35	.998	.986	.997	.999	.993	.993	.997
40	.999	.992	.999	.999	.997	.996	.998
45	.999	.996	.998	1.000	.998	.998	.999
50	1.000	.997	.999	1.000	.999	.999	1.000
N_ϵ	10 or 0	25 or 0	16 or 0	11	18	20	13
N	42	43	33	24	36	43	33
N_*	25	43	33	24	36	38	28
$\tilde{N}$	25	42	33	24	36	37	28

of N_ϵ would have been difficult. Yet, the bound $\tilde{N}$ on N_ϵ could easily be obtained and be used by the experimenter as a sample size satisfying his correct agreement goals, $\rho_n \geq \epsilon$.

3.7 A remark about the experimenter's goal.

In its generality, our experimenter's goal has been to have all observers choose the same and the correct hypothesis. A special case arises when the parameterization satisfies $\pi_U = 1 - \pi_L$, so that $c_0 = c_1$ in (3.11). Consider the specific parameterization $\pi_U = 1 - \pi_L = 0.95$ and $\epsilon = 0.99$ for exposition. We may then provide this specific interpretation of the experimenter's goal:

> The experimenter seeks that an observer a priori choosing the correct hypothesis with a probability of but 0.05 would a posteriori choose the correct hypothesis.

The sample size N_ϵ attains this goal with a probability of at least 0.99. A posteriori, when this goal is not attained for some sample $\underset{\sim}{x}_n$, then $\underset{\sim}{x}_n$ leads to one of three consequences:

(i) the same incorrect decision is made whether the prior is $\pi_L = 0.05$ or $\pi_U = 0.95$,

(ii) the decision is randomized when the prior is $\pi_L = 0.05$ or when the prior is $\pi_U = 0.95$,

or else

(iii) the decision is H_0 for the prior $\pi_U = 0.95$ but is H_1 for the prior $\pi_L = 0.05$.

Both the case (ii) and the case (iii) reveal a posteriori (through $\underset{\sim}{x}_n$) that "the" decision is not H_0 or H_1, but an "equivocal" decision.

An imperfect analogy can be made to classical hypothesis testing where a sample size N'_ϵ is chosen so that

$$\alpha = P_{\theta_0}(\text{choose } H_1) = 0.01 \quad \text{and} \quad \beta = P_{\theta_1}(\text{choose } H_0) = 0.01\,.$$

After $\underset{\sim}{x}_n$ is observed, a classical decision does not reveal its inconclusive nature as clearly as the experimenter's class of Bayesian decisions ($\pi_U = 1 - \pi_L = 0.95$) should $\underset{\sim}{x}_n$ satisfy (ii) or (iii) above.

The interpretation in Section 2.5 that the experiment gives posterior Bayes robustness with respect to the prior extends to composite hypotheses in the next chapter. The interpretation in this section that the experiment can be viewed through contrary-priors, π_U and π_L here, does not extend.

4. ONE-SIDED HYPOTHESIS TESTING

4.1 Introduction.

We now extend our hypotheses to the one-sided class of composite hypotheses. Specifically, we consider the two-action problem of Chapter 2 when the parameter spaces associated with H_0 and H_1 are the composite parameter spaces

$$\Theta_0 = \{\theta \leq b\} \quad \text{and} \quad \Theta_1 = \{\theta > b\}, \qquad \text{some } b \in \mathbf{R}^1, \tag{4.1}$$

respectively. The case where $\pi_\gamma(\Theta_0) = 0$ for all $\gamma \in \Gamma$, or $\pi_\gamma(\Theta_1) = 1$ for all $\gamma \in \Gamma$, was precluded by Assumption 2.1 of Chapter 2. From (2.4) of Chapter 2, the experimenter views that the Bayesian observers in his audience will choose the same correct hypothesis with a probability of

$$\rho_n \;=\; \sum_{i=0}^{1} \int_{\Theta_i} \int_{A_i} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta) d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta), \tag{4.2}$$

where

$$A_i = \left\{ \underset{\sim}{x}_n : \int_{\Theta_i} f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta) d\mu(\theta) > \int_{\Theta_j} f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta) d\mu(\theta),\ \text{all } \gamma \text{ in } \Gamma \right\}, \tag{4.3}$$

for $i \neq j = 0, 1$. As in (2.6) there, the experimenter wants to take a sample of size

$$N_\epsilon \;=\; \min_{n \geq 0} \{n : \rho_n \geq \epsilon\}. \tag{4.4}$$

In Chapter 5, we present an experimenter's more general goal that the observers will all choose the same correct hypothesis—and, in addition—that each observer will meet his own posterior expected loss goals. In that chapter, Assumption 5.4

is analogous to the obdurate assumption, (2.10) of Chapter 2, and is used to guarantee a finite N_ϵ. That assumption reduces to Assumption 4.1 below, in this chapter. It precludes priors too different, even allowing all parties to choose one hypothesis a.s. if the experimenter chooses the same hypothesis (although this trivial case itself is precluded through Assumption 2.1 so that conclusions will hold for any prior π_*).

ASSUMPTION 4.1. *For each $\delta > 0$ there is a $k_\delta > 0$ and a Borel set $G \subset \Theta$, $\pi_*(G) < \delta$, for which*

$$k_\delta < \frac{\pi_\gamma(\theta)}{\pi_{\gamma'}(\theta)} \qquad \text{for } \theta \in \Theta - G \text{ and all } \gamma, \gamma' \in \Gamma,$$

and

$$\int_G \sup_{\gamma \in \Gamma} \pi_\gamma(\theta)\, d\mu(\theta) < \infty .$$

Two other assumptions made in Chapter 5, Assumption 5.5 and Assumption 5.6, reduce to these:

ASSUMPTION 4.2. *For each $a \in \mathcal{X}$,*

$$\int_{-\infty}^{a} f(x|\theta)\, d\lambda(x)$$

is a Baire function of θ.

ASSUMPTION 4.3. *The parameterization is identifiable on Θ.*

These three assumptions guarantee, through Theorem 5.2, a finite N_ϵ in this chapter's inelaborate composite hypothesis problem. The Assumption 4.1 is on the priors, while the Assumptions 4.2 and 4.3 are on the likelihood function $f(x|\theta)$. Assumption 4.1 implies that the experimenter can have non-zero mass $\pi_*\{b\}$ on the parameter value b when the observers also have non-zero masses $\pi_\gamma\{b\}$ for $\gamma \in \Gamma$ on b $\left(\pi_\gamma\{b\} = 1 \text{ if } \pi_*\{b\} = 1\right)$.

For its amenability as a whole family of distributions, we now consider the one-sided hypothesis problem when sampling from a distribution in the exponential family. For this family, Assumption 4.2 always holds, while Assumption 4.3 holds whenever the sample space $\mathcal{X}$ contains at least two points. Consequently, the experimenter need only check the validity of Assumption 4.1 on the priors when the likelihood is from an exponential family.

4.2 The exponential family.

In this section, we consider data X_k, $\quad k = 1, \cdots, n$, arising from a distribution in the exponential family. As in Section 3.3, the random variable X and the parameter θ can be transformed so that the sample density of X has the canonical form

$$(4.5) \qquad f(x|\theta) = \exp[\theta x + d(\theta) + S(x)]\,, \qquad \text{for } x \in \mathcal{X}, \quad \theta \in \Theta\,.$$

As the discussion leading to (2.15) of Chapter 2 mentioned, a sample of size n allows the reduction of the density in $\underset{\sim}{x}_n$ to a density in the one-dimensional sufficient statistic $\overline{x}_n$:

$$(4.6) \qquad f_n(y|\theta) = \exp[n(\theta y + d(\theta)) + S_n(y)]\,, \qquad \text{for } y \in \mathcal{X}_n, \quad \theta \in \Theta\,,$$

where $\mathcal{X}_n$ denotes the space of sample mean values. Write the corresponding cumulative distribution function for $\overline{X}_n$ as

$$F_n(z|\theta) = \int_{(-\infty,z]} f_n(y|\theta)\, d\lambda_n(y)\,,$$

where $\lambda_n(y)$ is the dominating measure for $f_n(y|\theta)$ corresponding to the dominating measure $\lambda(\underset{\sim}{x}_n)$.

Define for $\gamma \in \Gamma$

$$(4.7) \qquad I_1(y; n, \gamma) = \int_{(-\infty,b]} e^{n[\theta y + d(\theta)]} \pi_\gamma(\theta)\, d\mu(\theta)\,,$$

(4.8) $$I_2(y;n,\gamma) = \int_{(b,+\infty)} e^{n[\theta y+d(\theta)]}\pi_\gamma(\theta)\,d\mu(\theta)\,,$$

(4.9) $$\tilde{I}_1(y;n,\gamma) = \int_{(-\infty,b]} e^{n[(\theta-b)y+d(\theta)]}\pi_\gamma(\theta)\,d\mu(\theta)\,,$$

and

(4.10) $$\tilde{I}_2(y;n,\gamma) = \int_{(b,+\infty)} e^{n[(\theta-b)y+d(\theta)]}\pi_\gamma(\theta)\,d\mu(\theta)\,.$$

With these definitions, we can represent the posterior probability of H_0 by

(4.11) $$\pi_\gamma(H_0|\overline{x}_n) = I_1(y;n,\gamma)\big/[I_1(y;n,\gamma)+I_2(y;n,\gamma)]$$

or

(4.12) $$\pi_\gamma(H_0|\overline{x}_n) = \tilde{I}_1(y;n,\gamma)\big/[\tilde{I}_1(y;n,\gamma)+\tilde{I}_2(y;n,\gamma)]\,.$$

The integral $\tilde{I}_1$ decreases strictly from $+\infty$ to 0 as y goes from $-\infty$ to $+\infty$. The integral $\tilde{I}_2$ increases strictly from 0 to $+\infty$ as y goes from $-\infty$ to $+\infty$. Accordingly, $\pi_\gamma(H_0|\overline{x}_n)$ in (4.12) goes from 1 to 0 as y goes from $-\infty$ to $+\infty$. Consequently, there is a unique $z_{n\gamma}$ for which

(4.13) $$\pi_\gamma(H_0|z_{n\gamma}) = 0.5\,.$$

When $z_{n\gamma}$ is in $\mathcal{X}_n$, the range of $\overline{X}_n$, then $z_{n\gamma}$ can be interpreted as the sample mean that makes b the posterior median for observer γ.

Let

(4.14) $$z_{nL} = \inf_{\gamma\in\Gamma} z_{n\gamma}$$

and

(4.15) $$z_{nU} = \sup_{\gamma\in\Gamma} z_{n\gamma}\,.$$

Equation (4.13) allows us to reduce A_0 and A_1 in (4.3) to functions of the sample mean:

(4.16) $$A_0 = (-\infty, z_{nL})$$

and

$$A_1 = (z_{nU}, +\infty)\,. \tag{4.17}$$

Furthermore, we can write $\tilde{\rho}_n$ in (2.22) as

$$\begin{aligned} \tilde{\rho}_n \;=\; & \int_{(-\infty,b]} \int_{-\infty}^{z_{nL}} f_n(\overline{x}_n|\theta)\pi_*(\theta)\, d\lambda_n(\overline{x}_n)\, d\mu(\theta) \\ & + \int_{(b,+\infty)} \int_{z_{nU}}^{+\infty} f_n(\overline{x}_n|\theta)\pi_*(\theta)\, d\lambda_n(\overline{x}_n)\, d\mu(\theta) \end{aligned} \tag{4.18}$$

—ρ_n in (4.2) when Γ is "closed."

With these simplifications of Θ_0, Θ_1, A_0 and A_1, we now outline a numerical program to find ρ_n and N_ϵ. Assume that Γ is finite, possibly by reducing Γ to two extreme observers as in Section 2.3. Only $d(\theta)$, $\pi_\gamma(\theta)$ with $\gamma \in \Gamma$, and $\pi_*(\theta)$ need be known, but we also assume that the cumulative distributions $F(a|\theta)$, $\pi_\gamma(-\infty, a]$ with $\gamma \in \Gamma$, and $\pi_*(-\infty, a]$, for "a" in the appropriate domains, are at hand. Here is the numerical program.

(i) *With an algorithm for univariate numerical integration, construct a procedure for finding I_1 and I_2 in (4.7) and (4.8).*

(ii) *With an algorithm for the zero of a function, construct a procedure for the zero of "$\pi_\gamma(H_0|\overline{x}_n) - 0.5$" in $\overline{x}_n$.*

Here $\pi_\gamma(H_0|\overline{x}_n)$ uses step (i) by way of equation (4.11), and n is fixed. Since $\pi_\gamma(H_0|\overline{x}_n)$ is strictly monotone in $\overline{x}_n$, an efficient algorithm can be used. This step (ii) will produce $z_{n\gamma}$ for all $\gamma \in \Gamma$.

(iii) *With an algorithm for univariate numerical integration, construct a procedure for*

$$\tilde{\rho}_n \;=\; \int_{(-\infty,b]} F(z_{nL}|\theta)\pi_*(\theta)\, d\mu(\theta) + \int_{(b,+\infty)} \left[1 - F(z_{nU}|\theta)\right]\pi_*(\theta)\, d\mu(\theta)\,. \tag{4.19}$$

(iv) *With an algorithm for the zero of a function, construct a procedure for a zero of* $\tilde{\rho}_n - \epsilon$ *in* n.

Since $\tilde{\rho}_n - \epsilon$ may have several zeros, a less efficient algorithm must be used here than what can be used in step (ii).

Each step of this program relies on the previous step. The resulting sample size in step (iv) is a bound on N_ϵ. A smallest bound on N_ϵ can be found by using steps (i) through (iii) to compute $\tilde{\rho}_n$ from $n = 0$ until some $n = m$ for which $\tilde{\rho}_m \geq \epsilon$——the resulting m is N_ϵ when Γ is closed (m is another bound on N_ϵ when Γ is not closed).

A computer program in Appendix C implements this numerical program when there are two extreme priors. That computer program requires that the following functions be given as subroutines:

1. $d(\theta)$, the scaling term for the exponential family density.

2. $F_n(y|\theta)$, the cumulative distribution function for the exponential family.

3. $\pi_\gamma(\theta)$ for all $\gamma \in \Gamma$, and $\pi_*(\theta)$, the prior densities.

4. $\pi_\gamma((-\infty, \theta])$ for all $\gamma \in \Gamma$, and $\pi_*((-\infty, \theta])$, the cumulative distribution functions for the priors.

The posterior distribution is not required.

When the priors π_γ, $\gamma \in \Gamma$, are conjugate priors to the likelihood $f(x|\theta)$, then steps (i) and (ii) of the numerical program on page 68 can be replaced by the single step:

(i-ii) *With the known posterior distribution* $\pi_\gamma(\theta|\overline{x}_n)$, *construct a procedure for the "inverse" of the conditional cumulative distribution function* $\pi_\gamma(H_0|z_{n\gamma}) = 0.5$.

Done for each $\gamma \in \Gamma$, this step produces the $z_{n\gamma}$ which make b a posterior median. When $f(x|\theta)$ is the Gaussian density, this modified step simply states the $z_{n\gamma} = \overline{x}_n$ for which the posterior mean is b.

4.3 Bounds on N_ϵ.

Still considering sample data X from a member of the exponential family, we present two theorems which give bounds on N_ϵ. These bounds, N, will satisfy the stronger result

$$\rho_n \geq \epsilon \qquad \text{when } n \geq N.$$

We will make use of the scaling factor $d(\theta)$ in (4.5) and its derivatives $d'(\theta)$, $d''(\theta)$. It is well known that $-d'(\theta)$ is the mean and $-d''(\theta)$ is the variance.

THEOREM 4.1. *Assume that $\pi_*\{b\} < \epsilon$. Then there exists a $\delta_0 > 0$ and a $\delta_1 > 0$ for which $\pi_*(b-\delta_0, b+\delta_1) < 1-\epsilon$ and $(b-\delta_0, b+\delta_1) \subset \Theta$. Let*

$$C_0 = (-\infty, -d'(b-\delta_0)) \tag{4.20}$$

and

$$C_1 = (-d'(b+\delta_1), +\infty)\,. \tag{4.21}$$

For each $\gamma \in \Gamma$, define (there exist) three sample sizes $N_{0\delta\gamma}$, $N_{1\delta\gamma}$ and N_2 for which

(a) $\int_{\Theta_0} f_n[-d'(b-\delta_0) \mid \theta]\pi_\gamma(\theta)\, d\mu(\theta) > \int_{\Theta_1} f_n[-d'(b-\delta_0) \mid \theta]\pi_\gamma(\theta)\, d\mu(\theta)$ *for $n \geq N_{0\delta\gamma}$,*

(b) $\int_{\Theta_0} f_n[-d'(b+\delta_1) \mid \theta]\pi_\gamma(\theta)\, d\mu(\theta) < \int_{\Theta_1} f_n[-d'(b+\delta_1) \mid \theta]\pi_\gamma(\theta)\, d\mu(\theta)$ *for $n \geq N_{1\delta\gamma}$,*

and

(c) $$\int_{\Theta_0}\int_{C_0} f_n(y|\theta)\pi_*(\theta)\, d\lambda_n(y)\, d\mu(\theta) + \int_{\Theta_1}\int_{C_1} f_n(y|\theta)\pi_*(\theta)\, d\lambda_n(y)\, d\mu(\theta) \geq \epsilon \quad \textit{for } n \geq N_2.$$

Define

$$N_{0\delta} = \sup_\gamma N_{0\delta\gamma}, \qquad N_{1\delta} = \sup_\gamma N_{1\delta\gamma}, \qquad \textit{and } N_{\epsilon\delta} = \max\{N_{0\delta}, N_{1\delta}, N_2\}\,.$$

Assume that $N_{0\delta}$ and $N_{1\delta}$ are finite—they are when Γ is a finite audience. Then

$$\rho_n \geq \epsilon \qquad \textit{for } n \geq N_{\epsilon\delta}.$$

Steps (i) and (ii) of the numerical program on page 68 to find $z_{n\gamma}$ to a high precision are replaced in Theorem 4.1 by steps (a) and (b) to find $N_{0\delta\gamma}$ and $N_{1\delta\gamma}$ to a low precision (rounded up to an integer). However, this theorem does demand that a δ_0 and a δ_1 be supplied, somewhat arbitrarily. The larger δ_0 and δ_1, the smaller $N_{0\delta}$ and $N_{1\delta}$, respectively. At the same time, the larger is N_2. $N_{\epsilon\delta}$ could be minimized over δ_0 and δ_1. In many problems, Theorem 4.1 allows that $\delta_0 = \delta_1 = \delta$, from which $N_{\epsilon\delta}$ could be minimized at a slightly larger minimum.

Below, Corollary 4.2 simplifies conditions (a) and (b). This corollary gives conditions under which the inequalities in (a) and (b) of Theorem 4.1 need only be met at $n = N_{0\delta\gamma}$ and $n = N_{1\delta\gamma}$, respectively. Those inequalities are then met for all greater n. Corollary 4.3 simplifies condition (c).

For the following Corollary 4.2, we introduce some notation. Denote by $g(\theta)$ the exponential family likelihood at $x = -d'(b - \delta_0)$ without the datum term $S(x)$:

$$g(\theta) \;=\; \exp\{\theta[-d'(b-\delta_0)] + d(\theta)\}\,.$$

With this, define

$$G_0(n) \;\equiv\; \int_{\Theta_0} \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta)\, d\mu(\theta)$$

and

$$G_1(n) \;\equiv\; \int_{\Theta_1} \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta)\, d\mu(\theta)\,.$$

Notice that G_0 and G_1 are implicitly functions of the observer γ. For an observer $\gamma \in \Gamma$, the inequality in condition (a) of Theorem 4.1 may be written

$$\int_{\Theta_0} [g(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta) > \int_{\Theta_1} [g(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta), \tag{4.22}$$

or equivalently

$$G_0(n) > G_1(n). \tag{4.23}$$

COROLLARY 4.2. *Suppose that the inequality in condition (a) of Theorem 4.1 holds at* $n = N_{0\delta\gamma}$. *Then it holds for all* $n \geq N_{0\delta\gamma}$ *if any of the following three conditions holds.*

(a1) The measure $\mu(\theta)$ *is uniform on its support (eg, Lebesgue measure or counting measure) while both the exponential family likelihood function* $f_1(x|\theta)$ *and the prior* $\pi_\gamma(\theta)$ *are symmetrical in* θ *about some points.*

(a2) Either $G_0(n) < G_0(n+1)$ *for some* $n \leq N_{0\delta\gamma}$, *or else* $G_0(n) > 1$ *for some* $n \leq N_{0\delta\gamma}$.

(a3) When there is a $\theta < b - \delta_0$ *for which* $g(\theta) = g(b)$, *it is unique. Label this* θ *as* b_s *and define* $\Theta_s = (b_s, b]$. *The condition is as follows. The value* b_s *exists, and for* $n = N_{0\delta\gamma}$,

$$\int_{\Theta_s} [g(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta) > \int_{\Theta_1} [g(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta).$$

A similar set of conditions implies the inequality (b) of Theorem 4.1 for all $n \geq N_{1\delta\gamma}$.

In the condition (a3), usually b_s exists for all $\delta_0 > 0$ for which $b - \delta_0 \in \Theta$. For δ_0 set small enough, b_s always exists.

The following Corollary 4.3 simplifies condition (c) of Theorem 4.1. It simplifies a two-dimensional integral problem on $\Theta \times \mathcal{X}$ into a one-dimensional problem on $\mathcal{X}$.

This is done at the cost of a possibly larger N_2. Plus coincidentally larger $N_{0\delta}$ or $N_{1\delta}$ should δ_0 or δ_1 be smaller in Corollary 4.3 than Theorem 4.1 demands.

COROLLARY 4.3. *Let β (one exists) be any real number for which*

$$\frac{\pi_*\{b\}}{1-\epsilon} < \beta < 1 .$$

Let ζ_0 and ζ_1 (they exist) be any real numbers for which $\zeta_0 > 0$, $\zeta_1 > 0$, and

$$\pi_*(\Theta_b) \leq \beta(1-\epsilon)\,, \tag{4.24}$$

where

$$\Theta_b = (b-\zeta_0, b+\zeta_1)\,.$$

Let δ_0 and δ_1 be any real numbers for which $0 < \delta_0 < \zeta_0$ and $0 < \delta_1 < \zeta_1$. Then, letting $\bar{A} = \mathcal{X} - A$, there is an N_2 for which

$$P_n\left(\bar{C}_0 \,\middle|\, \theta = b - \zeta_0\right) \;\leq\; (1-\beta)(1-\epsilon) \qquad \textit{when } n \geq N_2, \tag{4.25}$$

and

$$P_n\left(\bar{C}_1 \,\middle|\, \theta = b + \zeta_1\right) \;\leq\; (1-\beta)(1-\epsilon) \qquad \textit{when } n \geq N_2. \tag{4.26}$$

Furthermore, (c) of Theorem 4.1 holds for this N_2.

PROOF OF THEOREM 4.1.

First, we prove that $N_{0\delta\gamma}$, $N_{1\delta\gamma}$, and N_2 exist.

(a) Let

$$g(\theta) \;=\; \exp\{(\theta-b)[-d'(b-\delta_0)] + d(\theta)\}$$

and

$$h(\theta) \;=\; g(\theta) \Big/ \int_{(b-\delta_0,b]} g(\theta)\pi_\gamma(\theta)\,d\mu(\theta)\,.$$

The inequality in (a) is equivalent to

$$\int_{\Theta_0} [h(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta) \; > \; \int_{\Theta_1} [h(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta)\,.$$

Since $-d''(\theta) > 0$ for every $\theta \in \Theta$, then $g(\theta)$ is unimodal with its mode at $\theta = b + \delta_0$. Since $\pi_*(b-\delta_0, b) > 0$, there is a δ_0', $0 < \delta_0' < \delta_0$, for which $\pi_*(b-\delta_0, b-\delta_0') > 0$. This together with the unimodality of $g(\theta)$ implies that

$$\inf_{\theta \in (b-\delta_0, b-\delta_0')} h(\theta) \; > \; \sup_{\theta \in \Theta_1} h(\theta)\,.$$

Consequently,

$$\int_{\Theta_0} [h(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta) \; \geq \int_{(b-\delta_0, b-\delta_0')} [h(\theta)]^n \pi_\gamma(\theta)\, d\mu(\theta)$$

increases in n to $+\infty$ while

$$\int_{\Theta_1} [h(\theta)]^n \, \pi_\gamma(\theta)\, d\mu(\theta)$$

decreases in n to 0. So, the inequality in (a) holds for $n \geq N_{0\delta\gamma}$, some $N_{0\delta\gamma} \geq 1$.

(b) An argument just like the one above shows the existence of $N_{1\delta\gamma}$.

(c) Corollary 4.3 implies the existence of N_2.

Next, we prove that $\rho_n \geq \epsilon$ for $n \geq N_{\epsilon\delta}$.

(1) For the moment, assume that $\overline{x}_n \in C_0$: $\overline{x}_n < -d'(b-\delta_0)$. When $\theta \leq (\geq) b$, then

$$b\Big[-d'(b-\delta_0) - \overline{x}_n\Big] \quad \geq (\leq) \quad \theta\Big[-d'(b-\delta_0) - \overline{x}_n\Big]\,,$$

or equivalently,

$$\theta\overline{x}_n + d(\theta) + b\Big[-d'(b-\delta_0) - \overline{x}_n\Big] \quad \geq (\leq) \quad \theta[-d'(b-\delta_0)] + d(\theta)\,. \tag{4.27}$$

Consequently,

$$\begin{aligned} \int_{\Theta_0} \exp\Big\{n\Big[\theta\overline{x}_n + d(\theta)\Big] + nb\Big[-d'(b-\delta_0) - \overline{x}_n\Big]\Big\}\pi_\gamma(\theta)\, d\mu(\theta) \\ \geq \quad \int_{\Theta_0} \exp\Big\{n(\theta\Big[-d'(b-\delta_0)\Big] + d(\theta))\Big\}\pi_\gamma(\theta)\, d\mu(\theta)\,. \end{aligned} \tag{4.28}$$

Assumption (a) implies that this is larger than

$$\int_{\Theta_1} \exp\Big\{n\Big(\theta[-d'(b-\delta_0)] + d(\theta)\Big)\Big\}\pi_\gamma(\theta)\,d\mu(\theta) \qquad \text{for } n \geq N_{\epsilon\delta}.$$

Using (4.27) with $\theta \geq b$ instead of $\theta \leq b$, this is no smaller than

$$\text{(4.29)}\quad \int_{\Theta_1} \exp\Big\{n\Big[\theta\overline{x}_n + d(\theta)\Big] + nb\Big[-d'(b-\delta_0) - \overline{x}_n\Big]\Big\}\pi_\gamma(\theta)\,d\mu(\theta) \quad \text{for } n \geq N_{\epsilon\delta}.$$

Our resulting inequality between (4.28) and (4.29) reduces to

$$\int_{\Theta_0} e^{n[\theta\overline{x}_n + d(\theta)]}\pi_\gamma(\theta)\,d\mu(\theta) \;>\; \int_{\Theta_1} e^{n[\theta\overline{x}_n + d(\theta)]}\pi_\gamma(\theta)\,d\mu(\theta) \quad \text{for } n \geq N_{\epsilon\delta}.$$

Thus, if $\overline{x}_n \in C_0$, then $\overline{x}_n \in A_0$ for $n \geq N_{\epsilon\delta}$.

(2) The same argument shows that if $\overline{x}_n \in C_1$, then $\overline{x}_n \in A_1$ for $n \geq N_{\epsilon\delta}$ (the inequalities in (4.27) are reversed and assumption (b) is used).

Summarizing, for $n \geq N_{\epsilon\delta}$,

$$\rho_n \geq \tilde{\rho}_n \geq \int_{\Theta_0}\int_{C_0} f_n(y|\theta)\pi_*(\theta)\,d\lambda_n(y)\,d\mu(\theta) + \int_{\Theta_1}\int_{C_1} f_n(y|\theta)\pi_*(\theta)\,d\lambda_n(y)\,d\mu(\theta) \;\geq\; \epsilon$$

by assumption (c). □

PROOF OF COROLLARY 4.2.

<u>(a1)</u>. Let $\Theta_{0L} = (-\infty, b-2\delta_0)$ and $\Theta_{0R} = [b-2\delta_0, b]$. Let the symmetry of π_γ be about θ_γ. Define

$$G_{0L}(n) \;\equiv\; \int_{\Theta_{0L}} \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta)\,d\mu(\theta)\,,$$
$$G_{0R}(n) \;\equiv\; \int_{\Theta_{0R}} \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta)\,d\mu(\theta)\,,$$

for which $G_{0L}(n) + G_{0R}(n) = G_0(n)$. Define

$$\ddot{G}_{0L}(n) \;=\; G_{0L}(n) - G_{0L}(n-1)$$

and

$$\ddot{G}_1(n) = G_1(n) - G_1(n-1).$$

With these definitions, we may write the condition (4.22) as

$$\frac{G_{0L}(n) + G_{0R}(n)}{G_1(n)} > 1. \tag{4.30}$$

Since the mode of $\theta[-d'(b-\delta_0)] + d(\theta)$ is at $\theta = b - \delta_0$, the symmetry of $f_1(x|\theta)$, ergo $g(\theta)$, is about $\theta = b - \delta_0$. The log concavity of $g(\theta)$ implies that

$$\begin{aligned} g[(b-\delta_0) - \xi_3] &= g[(b-\delta_0) + \xi_3] \\ &< g[(b-\delta_0) - \xi_2] = g[(b-\delta_0) + \xi_2] < g(b-\delta_0) \end{aligned} \tag{4.31}$$

if $0 < \xi_2 < \xi_3$, a forteriori that

$$g(\theta) \geq g(b) \qquad \text{on } \Theta_{0R}$$

and

$$g(\theta) < g(b) \qquad \text{on } \Theta_{0L} \text{ and } \Theta_1. \tag{4.32}$$

Consequently, $G_{0R}(n)$ is strictly increasing in n.

Case 1. $\theta_\gamma > b$; ie, $\theta_\gamma \in \Theta_1$.

Consider $\theta < b - 2\delta_0$, ie, $\theta \in \Theta_{0L}$. The symmetry of $g(\theta)$ and $\pi_\gamma(\theta)$ imply that

$$g^n(\theta) = g^n\Big\{(b-\delta_0) + [(b-\delta_0) - \theta]\Big\} = g^n\Big\{b + [(b-2\delta_0) - \theta]\Big\} \tag{4.33}$$

and

$$\pi_\gamma(\theta) = \pi_\gamma[\theta_\gamma + (\theta_\gamma - \theta)]. \tag{4.34}$$

As $\theta_\gamma > b > b - \delta_0$, then $2\theta_\gamma - \theta > 2b - 2\delta_0 - \theta$. Moreover, with the symmetry of π_γ about θ_γ,

$$\pi_\gamma(\theta) = \pi_\gamma[\theta_\gamma + (\theta_\gamma - \theta)] \leq \pi_\gamma\{b + [(b-2\delta_0) - \theta]\} \qquad \text{if } b + [(b-2\delta_0) - \theta] > \theta_\gamma$$

and

$$\pi_\gamma(\theta) \leq \pi_\gamma\{b + [(b - 2\delta_0) - \theta]\} \qquad \text{if } b + [(b - 2\delta_0) - \theta] \leq \theta_\gamma.$$

Thus,

$$\pi_\gamma(\theta) \leq \pi_\gamma\{b + [(b - 2\delta_0) - \theta]\}. \tag{4.35}$$

For each $\theta < b - 2\delta_0$, (4.32) with the equality (4.33), and the inequality (4.35) imply that

$$\begin{aligned} 0 &\leq \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta) - \left[\frac{g(\theta)}{g(b)}\right]^{n+1} \pi_\gamma(\theta) \\ &< \left[\frac{g\{b + [(b - 2\delta_0) - \theta]\}}{g(b)}\right]^n \pi_\gamma\{b + [(b - 2\delta_0) - \theta]\} \\ &\qquad - \left[\frac{g\{b + [(b - 2\delta_0) - \theta]\}}{g(b)}\right]^{n+1} \pi_\gamma\{b + [(b - 2\delta_0) - \theta]\}. \end{aligned}$$

Since $b + [(b - 2\delta_0) - \theta]$ is a translation mapping $\theta < b - 2\delta_0$ one-to-one and onto $\theta > b$, then

$$0 \leq -\ddot{G}_{0L}(n+1) = G_{0L}(n) - G_{0L}(n+1) < G_1(n) - G_1(n+1) = -\ddot{G}_1(n+1).$$

We assume in Corollary 4.2 that (4.30) holds for $n = N_{0\delta\gamma}$. Inductively, assume it holds for some $n \geq N_{0\delta\gamma}$. Then

$$\begin{aligned} 1 &< \frac{G_{0L}(n) + \ddot{G}_{0L}(n+1) + G_{0R}(n)}{G_1(n) + \ddot{G}_{0L}(n+1)} \\ &< \frac{G_{0L}(n) + \ddot{G}_{0L}(n+1) + G_{0R}(n)}{G_1(n) + \ddot{G}_1(n+1)} = \frac{G_{0L}(n+1) + G_{0R}(n)}{G_1(n+1)}. \end{aligned}$$

With this and the increasing character of $G_{0R}(n)$,

$$1 < \frac{G_{0L}(n+1) + G_{0R}(n+1)}{G_1(n+1)}.$$

That is, the condition (4.23) is met for every $n \geq N_{0\delta\gamma}$.

Case 2. $\theta_\gamma \leq b$; ie, $\theta_\gamma \in \Theta_0$.

For $\theta \leq b$, the symmetry of $g(\theta)$ about $b - \delta_0$ implies that $g(\theta) > g[b + (b - \theta)]$, and the symmetry of π_γ about θ_γ implies that $\pi_\gamma(\theta) > \pi_\gamma[b + (b - \theta)]$. Since the translation $b + (b - \theta)$ maps Θ_0 onto Θ_1, then (4.22) holds for every $n \geq 0$.

(a2). Let H_n be any nonnegative function and m any nonnegative measure. The following difference, when finite, may be expressed

$$
\begin{aligned}
(4.36) \qquad & \int H^{n+1}(z)\,dm(z) - \int H^n(z)\,dm(z) \\
&= \int H^n(z)[H(z) - 1]\,dm(z) \\
&= \int_{\{H \leq 1\}} H^n(z)[H(z) - 1]\,dm(z) + \int_{\{H > 1\}} H^n(z)[H(z) - 1]\,dm(z)\,.
\end{aligned}
$$

Here, the first term is strictly increasing (decreasing in absolute value) in n, and the second term is strictly increasing in n. As a result, once the value in (4.36) is positive, it remains positive for all larger n. Reworded, once

$$\int H^n(z)\,dm(z)$$

increases, it increases for all larger n (being infinite once it first is, also). As a special case, if $G_0(n) < G_0(n + 1)$ for some $n \leq N_{2\delta\gamma}$, then $G_0(n) < G_0(n + 1)$ for every $n \geq N_{2\delta\gamma}$. Also, since $g(\theta)$ is log-concave with its mode at $\theta = b - \delta_0$, then $G_1(n)$ is decreasing in n. Consequently, if $G_0(n) < G_0(n + 1)$ for some $n \leq N_{0\delta\gamma}$, then (4.23) holds for all $n \geq N_{0\delta\gamma}$.

Suppose that $G_0(n) > 1$ for some $n \leq N_{0\delta\gamma}$. Then $G_0(n) > 1$ for all $n \geq N_{0\delta\gamma}$ ——an application of Liapounov's inequality. As $G_1(n) < 1$ for all n, if (4.23) holds at $n = N_{0\delta\gamma}$, then it holds for all $n \geq N_{0\delta\gamma}$.

(a3). By its definition, b_s is the θ satisfying

$$\theta[-d'(b - \delta_0)] + d(\theta) \;=\; b[-d'(b - \delta_0)] + d(b)\,.$$

Since $b_s \neq b$, b_s equivalently satisfies

$$d'(b - \delta_0) \;=\; \frac{d(b) - d(\theta)}{b - \theta}\,. \tag{4.37}$$

That is, the points $(b, d(b))$ and $(\theta, d(\theta))$ form a secant line whose slope is the same as that of the tangent line at $b-\delta_0$, the derivative $d'(b-\delta_0)$. Since $d(\theta)$ is concave, the right side of (4.37) is increasing in θ, making b_s unique if it exists.

From the definition of $G_0(n)$ and condition (a3),

$$(4.38) \qquad G_0(n) \ge \int_{\Theta_s} \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta)\, d\mu(\theta) > \int_{\Theta_1} \left[\frac{g(\theta)}{g(b)}\right]^n \pi_\gamma(\theta)\, d\mu(\theta)$$

at $n = N_{0\delta\gamma}$. Since the middle integral here increases in n and the last integral decreases in n, the inequality (4.38) holds for all $n \ge N_{0\delta\gamma}$. Consequently, so does the inequality (4.23). □

PROOF OF COROLLARY 4.3.

First, we prove the existence of β, ζ_0 and ζ_1. Theorem 4.1 assumed that $\pi_*\{b\} < 1-\epsilon$. Consequently, there is a β for which $0 < \beta < 1$ and $\pi_*\{b\} < \beta(1-\epsilon)$. Furthermore, there exist ζ_0 and ζ_1 for which

$$\pi_*(\Theta_b) = \pi_*(b-\zeta_0, b+\zeta_1) \le \beta(1-\epsilon).$$

Also, for any δ_0 satisfying $0 < \delta_0 < \zeta_0$,

$$\begin{aligned} & P_n\left(\bar{C}_0 \,\middle|\, \theta = b-\zeta_0\right) \\ = \; & P_n\left(\overline{X}_n \ge -d'(b-\delta_0) \,\middle|\, \theta = b-\zeta_0\right) \\ = \; & P_n\left(\overline{X}_n + d'(b-\zeta_0) \ge -d'(b-\delta_0) + d'(b-\zeta_0) \,\middle|\, \theta = b-\zeta_0\right). \end{aligned}$$

Since $-d'(\theta)$ is strictly increasing in θ, then $-d'(b-\delta_0) + d'(b-\zeta_0) > 0$. An application of Tchebyshev's inequality leads to

$$P_n(\bar{C}_0 | \theta = b-\zeta_0) \le \frac{-d''(b-\zeta_0)}{n[-d'(b-\delta_0) + d'(b-\zeta_0)]^2}.$$

Thus, for some $N_0 \ge 1$, (4.25) holds when $n \ge N_0$. Similarly, for some $N_1 \ge 1$, (4.26) holds when $n \ge N_1$. Let $N_2 = \max\{N_0, N_1\}$.

Now to show that (c) of Theorem 4.1 holds. For any random variable Y whose distribution has a monotone increasing likelihood ratio—as random variables from a canonical exponential family do—

$$\mathrm{P}(Y > a|\theta_2) \;\leq\; \mathrm{P}(Y > a|\theta_3)$$

for any $a \in \mathbf{R}^1$ and $\theta_2 < \theta_3 \in \Theta$. Consequently,

$$\begin{aligned}
&\int_{\Theta_0}\int_{C_0} f_n(y|\theta)\pi_*(\theta)\, d\lambda_n(y)\, d\mu(\theta) + \int_{\Theta_1}\int_{C_1} f_n(y|\theta)\pi_*(\theta)\, d\lambda_n(y)\, d\mu(\theta) \\
\geq &\int_{\Theta_0-\Theta_b} P_n\Big(C_0 \;\Big|\; \theta = b-\zeta_0\Big)\pi_*(\theta)\, d\mu(\theta) + \int_{\Theta_1-\Theta_b} P_n\Big(C_1 \;\Big|\; \theta = b+\zeta_1\Big)\pi_*(\theta)\, d\mu(\theta) \\
&\quad + \Bigg[\int_{\Theta_0\cap\Theta_b} P_n\Big(C_0 \;\Big|\; \theta = b-\zeta_0\Big)\pi_*(\theta)\, d\mu(\theta) \\
&\qquad + \int_{\Theta_1\cap\Theta_b} P_n\Big(C_1 \;\Big|\; \theta = b+\zeta_1\Big)\pi_*(\theta)\, d\mu(\theta) \;-\; \pi_*(\Theta_b)\Bigg].
\end{aligned}$$

Using (4.25) and (4.26), this is no smaller than

$$\int_\Theta \Big[1-(1-\beta)(1-\epsilon)\Big]\pi_*(\theta)\, d\mu(\theta) \;-\; \pi_*(\Theta_b) \qquad \text{when } n \geq N_2.$$

Using (4.24), this is no smaller than

$$\Big[1-(1-\beta)(1-\epsilon)\Big] - \beta(1-\epsilon) \;=\; \epsilon \qquad \text{when } n \geq N_2. \qquad \square$$

The following theorem gives an alternative to Theorem 4.1 for bounding N_ϵ. This theorem uses a Tchebyshev type bound to reduce the integral on $\Theta \times \mathcal{X}$ in (4.18) to an integral on Θ.

For the following theorem, we introduce some new notation. With Assumption 4.1 guaranteeing that π_γ, $\gamma \in \Gamma$, have the same support, define

$$b_- \;\equiv\; \sup\Big\{\theta\colon\; \pi_\gamma(\theta, b] > 0\,, \quad \text{for any } \gamma \in \Gamma\Big\}$$

and

$$b_+ \equiv \inf\Big\{\theta\colon \pi_\gamma[b,\theta) > 0\,, \quad \text{for any } \gamma \in \Gamma\Big\}.$$

Often, as when π_* is continuous, $b_- = b_+ = b$. Assumption 2.1 of Chapter 2 guarantees that b_- and b_+ exist. Define

$$\theta_- \equiv \inf\{\theta\colon \theta \in \Theta\}$$

and

$$\theta_+ \equiv \sup\{\theta\colon \theta \in \Theta\}.$$

Usually Θ is open, so usually θ_- and θ_+ are not in Θ. These definitions satisfy

$$\theta_- \le b_- \le b_+ \le \theta_+.$$

Define

$$x_- = \inf\{x\colon x \in \mathcal{X}\}$$

and

$$x_+ = \sup\{x\colon x \in \mathcal{X}\}.$$

The function $-d'(\theta)$ is a continuous 1-1 function mapping (θ_-, θ_+) onto (x_-, x_+), see Brown (1986, pg 74). So, $x_- \le -d'(\theta) \le x_+$. Define

$$D(y) \equiv \begin{cases} -d'^{-1}(y) & \text{if } y \in (x_-, x_+) \\ \theta_- & \text{if } y \le x_- \\ \theta_+ & \text{if } y \ge x_+, \end{cases} \tag{4.39}$$

where, as before, $d'^{-1}(y) = \left\{\frac{d}{d\theta} d(\theta)\right\}^{-1}(y)$. $D(y)$ is a non-decreasing continuous function of y—an increasing function for $y \in (x_-, x_+)$. Define

$$\Theta_n \equiv [D(z_{nL}), D(z_{nU})], \tag{4.40}$$

where z_{nL} is defined in (4.14), and z_{nU} is defined in (4.15).

THEOREM 4.4. *Let $\epsilon > 0$. Assume that $\pi_\gamma(\theta_-, b] > 0$, $\gamma \in \Gamma$, and that $\pi_\gamma(b, \theta_+) > 0$, $\gamma \in \Gamma$ [This assumption holds when the exponential family is regular—Θ is open—because of the obdurate Assumption 2.1 in Chapter 2]. Assume that*

$$\pi_*[b_-, b_+] < 1 - \epsilon .$$

Then

$$(4.41)\quad 1 - \rho_n \leq \int_{\Theta_0 - \Theta_n} \frac{-d''(\theta)}{-d''(\theta) + n[z_{nL} + d'(\theta)]^2} \pi_*(\theta)\, d\mu(\theta) + \int_{\Theta_1 - \Theta_n} \frac{-d''(\theta)}{-d''(\theta) + n[z_{nU} + d'(\theta)]^2} \pi_*(\theta)\, d\mu(\theta) + \pi_*(\Theta_n) .$$

Furthermore, there is a positive integer N for which the bound in (4.41) is less than $1 - \epsilon$ at $n = N$, ie, for which $\rho_N \geq \epsilon$.

PROOF OF THEOREM 4.4.

First, we show that

$$(4.42)\qquad -d'(b_-) \leq \lim_{n \to +\infty} z_{n\gamma} \leq -d'(b_+) ,$$

where the $z_{n\gamma}$ are as defined in (4.13). Define

$$(4.43)\qquad H_{n\gamma}(\theta) \equiv \exp\{\theta z_{n\gamma} + d(\theta)\} .$$

Suppose that (4.42) is not true. Then for some $\delta > 0$ and some countably infinite set M of positive integers,

$$z_{n\gamma} < -d'(b_-) - \delta , \qquad \text{for } n \in M,$$

or else

$$z_{n\gamma} > -d'(b_+) + \delta , \qquad \text{for } n \in M.$$

<u>case 1</u>. $z_{n\gamma} < -d'(b_-) - \delta, \qquad \text{for } n \in M.$

Let $r = -d'(b_-) - \delta$, so that this case assumes $z_{n\gamma} < r$, $n \in M$. If $z_{n\gamma} < x_-$, then

$$\frac{d}{d\theta}\Big[\log H_{n\gamma}(\theta)\Big] \;=\; z_{n\gamma} + d'(\theta) \;\le\; z_{n\gamma} - x_- \;<\; 0\,,$$

so the mode of $H_{n\gamma}(\theta)$ is "at"

$$\theta = \theta_- = D(z_{n\gamma})\,. \tag{4.44}$$

If $z_{n\gamma} > x_+$, then

$$\frac{d}{d\theta}\Big[\log H_{n\gamma}(\theta)\Big] \;=\; z_{n\gamma} + d'(\theta) \;\ge\; z_{n\gamma} - x_+ \;>\; 0\,,$$

so the mode of $H_{n\gamma}(\theta)$ is "at"

$$\theta = \theta_+ = D(z_{n\gamma})\,. \tag{4.45}$$

These two cases, (4.44) and (4.45), plus the concavity of $d(\theta)$ imply that the mode of $H_{n\gamma}(\theta)$ is "at" $\theta = D(z_{n\gamma})$. Since $D(y)$ is increasing on (x_-, x_+) and $x_- < -d'(b_-) < x_+$, then $D(r) < b_- \le b$. Since the mode of $H_{n\gamma}$ is at $\theta = D(z_{n\gamma})$, for those θ such that

$$D(z_{n\gamma}) \le D(r) < \theta \le b\,,$$

the function $H_{n\gamma}(\theta)$ is no smaller than $H_{n\gamma}(b)$:

$$\frac{H_{n\gamma}(\theta)}{H_{n\gamma}(b)} \ge 1\,, \qquad \text{for } D(r) < \theta \le b, \quad n \in M\,.$$

As an immediate consequence,

$$\int_{\Theta_0} \left[\frac{H_{n\gamma}(\theta)}{H_{n\gamma}(b)}\right]^n \pi_\gamma(\theta)\, d\mu(\theta) \ge \pi_\gamma(D(r), b] > 0\,, \qquad \text{for } n \in M. \tag{4.46}$$

Since the mode of $H_{n\gamma}(\theta)$ is at $\theta = D(z_{n\gamma})$, for those θ such that

$$D(z_{n\gamma}) \le D(r) < b < \theta\,,$$

the function $H_{n\gamma}(b)$ is smaller than $H_{n\gamma}(\theta)$:

$$\frac{H_{n\gamma}(\theta)}{H_{n\gamma}(b)} < 1 , \qquad \text{for } \theta \in \Theta_1, \quad n \in M .$$

As an immediate consequence, for n sufficiently large

$$(4.47) \qquad \pi_\gamma\{(D(r), b]\} > \int_{\Theta_1} \left[\frac{H_{n\gamma}(\theta)}{H_{n\gamma}(b)}\right]^n \pi_\gamma(\theta)\, d\mu(\theta) , \qquad \text{for } n \in M.$$

Together, (4.46) and (4.47) imply that for n sufficiently large,

$$\int_{\Theta_0} [H_{n\gamma}]^n \pi_\gamma(\theta)\, d\mu(\theta) \neq \int_{\Theta_1} [H_{n\gamma}]^n \pi_\gamma(\theta)\, d\mu(\theta) , \qquad n \in M .$$

This contradicts the definition of $z_{n\gamma}$ in (4.13). Thus, this case 1 is not possible.

case 2. $z_{n\gamma} > -d'(b_+) + \delta, \qquad$ for $n \in M$.

The argument of case 1 is easily adapted to show that this case is not possible either. Together, case 1 and case 2 imply that the bounds in (4.42) must be correct.

Now to find a lower bound on ρ_n. From (4.18),

$$(4.48) \quad 1 - \tilde{\rho}_n = \int_{\Theta_0} \mathrm{P}\Big(\overline{X}_n + d'(\theta) \geq z_{nL} + d'(\theta) \,\Big|\, \theta\Big) \pi_*(\theta)\, d\mu(\theta) + \int_{\Theta_1} \mathrm{P}\Big(\overline{X}_n + d'(\theta) \leq z_{nU} + d'(\theta) \,\Big|\, \theta\Big) \pi_*(\theta)\, d\mu(\theta) .$$

A version of the Tchebyshev inequality (Cramer(1946, pg 256)) states that for any random variable Y with first moment μ and standard deviation σ,

$$(4.49) \qquad \mathrm{P}(Y - \mu \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \qquad \text{when } a > 0.$$

Equivalently,

$$(4.50) \qquad \mathrm{P}(Y - \mu \leq a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \qquad \text{when } a < 0.$$

The value $z_{nL} + d'(\theta)$ in (4.48) is positive on the set

$$\left\{\theta_- < \theta < -d'^{-1}(\theta)\right\} = \Theta_0 - \Theta_n .$$

The value $z_{n\mu} + d'(\theta)$ in (4.48) is negative on the set

$$\left\{-d'^{-1}(z_{nL}) < \theta < \theta_+\right\} \;=\; \Theta_1 - \Theta_n\,.$$

Applying the Tchebyshev inequality to (4.48),

$$(4.51)\quad 1 - \tilde{\rho}_n \;\leq\; \int_{\Theta_0 - \Theta_n} \frac{-d''(\theta)}{-d''(\theta) + n[z_{nL} + d'(\theta)]^2}\pi_*(\theta)\,d\mu(\theta) + \int_{\Theta_1 - \Theta_n} \frac{-d''(\theta)}{-d''(\theta) + n[z_{nU} + d'(\theta)]^2}\pi_*(\theta)\,d\mu(\theta) \;+\; \pi_*(\Theta_n)\,.$$

From (4.42) and the continuity of $D(\cdot)$,

$$\lim_{n\to+\infty} \pi_*(\Theta_n) \;\leq\; \pi_*[b_-, b_+] \;<\; 1 - \epsilon\,.$$

Consequently, there is an N for which

$$1 - \rho_N \;\leq\; 1 - \tilde{\rho}_N \;<\; 1 - \epsilon$$

in (4.51). □

4.4 The Gaussian distribution example.

Here we consider sample data X from a Gaussian distribution. Transformed to canonical form, X has the density $f(x|\theta, 1)$, where

$$(4.52)\quad f(x|\theta,\sigma^2) \;=\; \exp\left\{\frac{x\theta}{\sigma^2} - \frac{\sigma^2}{2\sigma^2} + \left[-\frac{x^2}{2\sigma^2} - \ln(\sqrt{2\pi}\sigma)\right]\right\}, \quad \text{on } x \in \mathbf{R}^1, \quad \text{for } \theta \in \mathbf{R}^1, \quad \sigma > 0\,.$$

Denote the cumulative distribution function of f by $F(x|\theta,\sigma^2)$. The density $f(x|\theta,1)$ has the scaling term

$$(4.53)\qquad d(\theta) = -\frac{1}{2}\theta^2\,.$$

The sample mean $\overline{X}_n$ has the density

$$(4.54)\qquad f(x|\theta, 1/n)\,.$$

Assume the observers' priors π_γ, $\gamma \in \Gamma$, are the conjugate priors, with parameters μ_γ and τ_γ, having the Gaussian densities

$$\pi_\gamma(\theta|\mu_\gamma, \tau_\gamma) = f(\theta|\mu_\gamma, \tau_\gamma^2)\,, \tag{4.55}$$

where f is as defined in (4.52). The experimenter's prior is $\pi_*(\theta|\mu_*, \tau_*)$, where "$*$" replaces "$\gamma$" in (4.55). These functions $d(\theta)$, $F(x|\theta, 1/n)$, $f(\theta|\mu_\gamma, \tau_\gamma^2)$, and $F(\theta|\mu_\gamma, \tau_\gamma^2)$ enable the use of the computer program for ρ_n in Appendix C.

As mentioned on page 69, with conjugate priors, $z_{n\gamma}$ can be found more efficiently through another method than through steps (i) and (ii) of the numerical program for ρ_n. We use the more efficient method now. The posterior density is

$$\pi_\gamma(\theta|\overline{x}_n) = f\Big(\theta \;\Big|\; \mu_\gamma(\overline{x}_n), \sigma_\gamma^2(\overline{x}_n)\Big)\,, \tag{4.56}$$

where

$$\mu_\gamma(\overline{x}_n) \;\equiv\; \frac{1}{1+n\tau_\gamma^2}\mu_\gamma + \frac{n\tau_\gamma^2}{1+n\tau_\gamma^2}\overline{x}_n\,,$$

and

$$\sigma_\gamma(\overline{x}_n) \equiv \frac{\tau_\gamma^2}{1+n\tau_\gamma^2}\,.$$

The specification for $z_{n\gamma}$ is

$$\pi_\gamma\Big(H_0 \;\Big|\; \mu_\gamma(z_{n\gamma}), \sigma_\gamma(z_{n\gamma})\Big) \;=\; F\Big(b \;\Big|\; \mu_\gamma(z_{n\gamma}), \sigma_\gamma^2(z_{n\gamma})\Big) \;=\; 0.5\,.$$

Transforming X into

$$Y \;=\; [X - \mu_\gamma(z_{n\gamma})]/\sigma_\gamma(z_{n\gamma})\,,$$

$z_{n\gamma}$ is specified by

$$F([b - \mu_\gamma(z_{n\gamma})]/\sigma_\gamma(z_{n\gamma})|0,1) \;=\; 0.5\,.$$

Since $F(\cdot|0,1)$ is the standard normal distribution function, its median occurs at 0. Consequently, $z_{n\gamma}$ is specified by

$$[b - \mu_\gamma(z_{n\gamma})]/\sigma_\gamma(z_{n\gamma}) \;=\; 0$$

or

$$z_{n\gamma} = b\left(\frac{1+n\tau_\gamma^2}{n\tau_\gamma^2}\right) - \frac{\mu_\gamma}{n\tau_\gamma^2}\,. \tag{4.57}$$

Using (4.18), we can write

$$\rho_n = \int_{-\infty}^{b}\int_{-\infty}^{z_{nL}} f(t|\theta,1/n)\;dt\;f(\theta|\mu_*,\tau_*^2)\;d\theta + \int_{b}^{+\infty}\int_{z_{nU}}^{+\infty} f(t|\theta,1/n)\;dt\;f(\theta|\mu_*,\tau_*^2)\;d\theta\,.$$

Changing the variables through the transformations

$$u = \sqrt{n}(t-\theta) \qquad \text{and} \qquad \eta = (\theta-\mu_*)/\tau_*\,,$$

we get

$$\begin{aligned}\rho_n \;=\; & \int_{-\infty}^{L_1}\int_{-\infty}^{L_2} f(u|0,1)\;du\;f(\eta|0,1)\;d\eta \\ & + \int_{L_1}^{+\infty}\int_{L_3}^{+\infty} f(u|0,1)\;du\;f(\eta|0,1)\;d\eta\,,\end{aligned} \tag{4.58}$$

where

$$\begin{aligned} L_1 &= (b-\mu_*)/\tau_*\,, \\ L_2 &= \sqrt{n}[z_{nL} - (\tau_*\eta + \mu_*)]\,, \end{aligned}$$

and

$$L_3 = \sqrt{n}[z_{nU} - (\tau_*\eta + \mu_*)]\,.$$

Consider the following reparameterization:

$$\bar{b} = 0, \quad \bar{\mu}_* = \mu_* - b, \quad \bar{\mu}_\gamma = \frac{\mu_\gamma - b}{\tau_\gamma^2}, \quad \bar{\tau}_\gamma = 1\,.$$

The two ordered sets, $(b, \mu_*, \tau_*, \mu_\gamma, \tau_\gamma)$ and $(\bar{b}, \bar{\mu}_*, \tau_*, \bar{\mu}_\gamma, \bar{\tau}_\gamma)$, $\gamma \in \Gamma$, give the same values for L_1, L_2, and L_3 (z_{nL} and z_{nU} are functions of these ordered sets). Without losing generality, we assume that $b = 0$ and $\tau_\gamma = 1$, $\gamma \in \Gamma$. From (4.57), this assumption makes

$$z_{nL} = -\mu_U/n \tag{4.59}$$

and

$$z_{nU} = -\mu_L/n\,, \tag{4.60}$$

where

$$\mu_L = \inf_{\gamma\in\Gamma} \mu_\gamma \qquad \text{and} \qquad \mu_U = \sup_{\gamma\in\Gamma} \mu_\gamma\,.$$

This gives ρ_n the same form as the general form (4.58) but using

$$L_1 = -\mu_*/\tau_*\,, \tag{4.61}$$

$$L_2 = -\left[(\mu_* + \tau_*\eta)\sqrt{n} + \frac{\mu_U}{\sqrt{n}}\right], \tag{4.62}$$

and

$$L_3 = -\left[(\mu_* + \tau_*\eta)\sqrt{n} + \frac{\mu_L}{\sqrt{n}}\right]. \tag{4.63}$$

We now find a bound on N_ϵ through Theorem 4.1. From (4.53), $-d'(\theta) = \theta$. Let $\delta_0 = \delta_1 = \delta$ in Theorem 4.20, so that $\delta > 0$ must satisfy $\pi_*(-\delta,\delta) < 1-\epsilon$. We will now show that one such δ is

$$\delta = \min_{j=1,2}\left|F^{-1}\left[F(0|\mu_*,\tau_*^2) + (-1)^j\frac{1-\epsilon}{2K}\,\Big|\,\mu_*,\tau_*^2\right]\right|, \qquad \text{any } K \geq 1 \tag{4.64}$$

($K > 1$ if $\mu_* = 0$). If $\mu_* < 0$, then

$$\begin{aligned}
\pi_*(-\delta,\delta) &< 2\pi_*(-\delta,0)\\
&= 2\left[F(0|\mu_*,\tau_*^2) - F(-\delta|\mu_*,\tau_*^2)\right]\\
&\leq 2\left[F(0|\mu_*,\tau_*^2) - F\left\{F^{-1}\left[F(0|\mu_*,\tau_*^2) - \frac{1-\epsilon}{2K}\,\Big|\,\mu_*,\tau_*^2\right]\,\Big|\,\mu_*,\tau_*^2\right\}\right]\\
&= \frac{1-\epsilon}{K}.
\end{aligned}$$

So, when $\mu_* < 0$, $\pi_*(-\delta,\delta) < 1-\epsilon$. A similar argument shows this same inequality when $\mu_* > 0$.

Here are the three needed component sample sizes for Theorem 4.1.

(a) Part (a) of Theorem 4.1 requires that

$$\int_{-\infty}^{0} f(-\delta|\theta,1/n)f(\theta|\mu_\gamma,1)\,d\theta > \int_{0}^{+\infty} f(-\delta|\theta,1/n)f(\theta|\mu_\gamma,1)\,d\theta\,.$$

Writing this in the posterior distribution,

$$F\left(0 \,\middle|\, \frac{1}{1+n}\mu_\gamma - \frac{n}{1+n}\delta, \frac{1}{1+n}\right) \;>\; 0.5\,,$$

which occurs iff

$$\frac{1}{1+n}\mu_\gamma - \frac{n}{1+n}\delta \;<\; 0\,,$$

or

$$n \;>\; \frac{\mu_\gamma}{\delta} = N_{0\delta\gamma}\,. \tag{4.65}$$

With inequality ($n > \mu_\gamma/\delta$) instead of equality in (4.65), our conclusion from Theorem 4.1 will be $n > N_{\epsilon\delta}$ instead of $n \geq N_{\epsilon\delta}$.

(b) By a similar argument, part (b) of Theorem 4.1 requires that

$$n \;>\; -\frac{\mu_\gamma}{\delta} = N_{1\delta\gamma}\,.$$

(c) Let

$$\begin{aligned} h_n(a1, a2) \;=\; & \int_{-\infty}^{a1} F\Big[\sqrt{n}(-\delta - \mu_* - \tau_*\theta) \;\Big|\; 0,1\Big] f(\theta|0,1)\; d\theta \\ & + \;\int_{a2}^{+\infty} \Big\{1 - F\Big[\sqrt{n}(\delta - \mu_* - \tau_*\theta) \;\Big|\; 0,1\Big]\Big\} f(\theta|0,1)\; d\theta\,. \end{aligned}$$

The condition (c) of Theorem 4.1,

$$\begin{aligned} & \int_{-\infty}^{0}\int_{-\infty}^{-\delta} f(y|\theta, 1/n) f(\theta|\mu_*, \tau_*^2)\; dy\; d\theta \\ & + \;\int_{0}^{+\infty}\int_{\delta}^{+\infty} f(y|\theta, 1/n) f(\theta|\mu_*, \tau_*^2)\; dy\; d\theta \;\geq\; \epsilon \qquad \text{for } n \geq N_2, \end{aligned}$$

may be written

$$h_n(-\mu_*/\tau_*, -\mu_*/\tau_*) \geq \epsilon \qquad \text{for } n \geq N_2.$$

Since

$$h_n(-\mu_*/\tau_*, -\mu_*/\tau_*) \;>\; h_n\left(\frac{-\mu_* - \delta}{\tau_*}, \frac{-\mu_* + \delta}{\tau_*}\right)$$

and $h_n(\frac{-\mu_*-\delta}{\tau_*}, \frac{-\mu_*+\delta}{\tau_*})$ is monotone increasing in n, the condition

$$(4.66) \qquad h_n\left(\frac{-\mu_* - \delta}{\tau_*}, \frac{-\mu_* + \delta}{\tau_*}\right) \geq \epsilon \qquad \text{at } n = N_2$$

suffices for condition (c) of Theorem 4.1.

Theorem 4.1 concludes that any n larger than

$$N_{\epsilon\delta} = \max\left\{\frac{\mu_U}{\delta}, -\frac{\mu_L}{\delta}, N_2\right\}$$

is a bound on N_ϵ.

Recall that δ depends on a pre-specified K in (4.64). With this δ determined in (4.64) and N_2 determined in (4.66), we designate the above bound $N_{\epsilon\delta}$ on N_ϵ by

$$(4.67) \qquad N_{\epsilon,\delta,K}\,.$$

Here, ϵ and K receive explicit values, but "δ" is included only as a mnemonic that K determines δ. For example, $N_{.95,\delta,1.5}$.

We now find another bound on N_ϵ through Theorem 4.4. From (4.53), $d'(\theta) = -\theta$ and $d''(\theta) = -1$. From (4.39) and (4.40), $D(y) = y$ and $\Theta_n = (z_{nL}, z_{nU})$. Using these items in Theorem 4.4,

$$\begin{aligned} 1 - \rho_n \;\leq\; & \int_{-\infty}^{-\mu_U/n} \frac{1}{1 + n(-\mu_U/n - \theta)^2} f(\theta|\mu_*, \tau_*^2)\, d\theta \\ & + \int_{-\mu_L/n}^{+\infty} \frac{1}{1 + n(-\mu_L/n - \theta)^2} f(\theta|\mu_*, \tau_*^2)\, d\theta + \int_{-\mu_U/n}^{-\mu_L/n} f(\theta|\mu_*, \tau_*^2)\, d\theta\,. \end{aligned}$$

Substituting $t = -\theta - \mu_U/n$ in the first integral and $t = \theta + \mu_L/n$ in the second integral,

$$\begin{aligned} (4.68)\quad 1 - \rho_n \;\leq\; & \int_0^{+\infty} \frac{1}{1 + nt^2} f\left(t \,\middle|\, -\mu_* - \mu_U/n, \tau_*^2\right) dt \\ & + \int_0^{+\infty} \frac{1}{1 + nt^2} f\left(t \,\middle|\, \mu_* + \mu_L/n, \tau_*^2\right) dt \;+ \int_{-\mu_U/n}^{-\mu_L/n} f(\theta|\mu_*, \tau_*^2)\, d\theta\,. \end{aligned}$$

Let

$$\mu_1 = \mu_* + \frac{\mu_L}{n}$$

and

$$\mu_2 = \mu_* + \frac{\mu_U}{n}.$$

Bound each of the first two integrals in (4.68) by $2K$, $K \geq 1$, integrals: integrals integrated on the $2K$ equal lengthed subintervals that constitute the interval $(\mu_i - 3\tau_*, \mu_i + 3\tau_*)$, $i = 2, 1$, respectively. Plus a small probability for the rest of the positive reals (the tails of the experimenter's prior in (4.68)). Since

$$\int \frac{1}{1+nt^2}\, dt \;=\; \frac{1}{\sqrt{n}} \tan^{-1}(\sqrt{n}t), \tag{4.69}$$

the following bound results from (4.68).

$$\begin{aligned} (4.70) \qquad & 1 - \rho_n \\ \leq \; & \sum_{i=1}^{2} \sum_{j=-K}^{K-1} \Bigg\{ \frac{1}{\sqrt{n}} \Big[\tan^{-1}(\sqrt{n}V_{ij}) - \tan^{-1}(\sqrt{n}U_{ij})\Big] \frac{1}{\sqrt{2\pi}\tau_*} I\{U_{ij}V_{ij} \leq 0\} \\ & + \Big\{ F\Big[\tfrac{U_{ij}+(-1)^i\mu_i}{\tau_*} \Big| 0,1\Big] - F\Big[\tfrac{V_{ij}+(-1)^i\mu_i}{\tau_*} \Big| 0,1\Big] \Big\} \Big[1 + n(\min\{|U_{ij}|, |V_{ij}|\})^2\Big] \Bigg\} I\{U_{ij}V_{ij} > 0\} \\ & + 2F\Big(-3 \,\Big|\, 0,1\Big) \;+\; F\Big(-\tfrac{\mu_1}{\tau_*} \,\Big|\, 0,1\Big) \;-\; F\Big(-\tfrac{\mu_2}{\tau_*} \,\Big|\, 0,1\Big); \end{aligned}$$

where

$$U_{ij} \;=\; (-1)^{i+1}\mu_i + \frac{3j}{K}\tau_*$$

and

$$V_{ij} \;=\; (-1)^{i+1}\mu_i + \frac{3(j+1)}{K}\tau_* \qquad \text{for } i = 1, 2; \quad j = -2, -1, 0, 1.$$

We will use $K = 2$. A computer algorithm for the zero of a function can find an n for which the bound on $1 - \rho_n$ in (4.70) is $1 - \epsilon$. We will use N to denote the smallest such solution, a bound on N_ϵ.

An explicit, though crude, bound on $1-\rho_n$ uses the indefinite integral (4.69) on all of $(0,\infty)$ in (4.68). A resulting bound is

$$1-\rho_n \le \left[\frac{1}{\sqrt{n}}\pi + \frac{1}{n}(\mu_U - \mu_L)\right]\frac{1}{\sqrt{2\pi}\tau_*}. \tag{4.71}$$

From this, a crude bound on N_ϵ, using the quadratic formula, is

$$N_\epsilon \le \frac{2rs+1+\sqrt{4rs+1}}{2r^2}, \tag{4.72}$$

where

$$r = (1-\epsilon)\tau_*\sqrt{2/\pi}$$

and

$$s = \frac{\mu_U - \mu}{\pi}.$$

Table 4.1 presents ρ_n and bounds on N_ϵ in six examples. In the first example the experimenter's prior is disparate from any observer's prior. The second example's more diverse audience requires a sample size with two extra data to achieve the same level of agreement, ρ_n. The first datum is not nearly as useful as in the first example. The third example presents an experimenter with his prior mean in the middle of the audience's prior means. In the fourth example, the experimenter would use $N_{.95} = 0$, and would not want some larger sample sizes (eg, $n = 4$). The fifth example presents an experimenter having a prior with small variance. The additive contribution to ρ_n of an extra datum is much greater for n between 1000 and 5000 than for n less than 1000. The same could not have been said if $\mu_* = 0$. In the sixth example, the audience agrees for every n. The sample size affects only the correctness of the audience's decision. The bottom of this table presents N_ϵ and three bounds on it. Notice that $N_{\epsilon,\delta,K}$ generally performs best when μ_*/τ_* is far from $b = 0$. The bound $N = 1$ in the sixth example is the smallest solution using (4.70)—the next solution is much larger.

Table 4.1 ρ_n for composite hypotheses, where X has a Gaussian distribution

μ_L	2	2	-2	1.7	-150	-5
μ_U	4	10	2	10	2	-5
μ_*	-6	-6	0	1.7	.1	-.05
τ_*	1	1	1	1	.03	.01
n						
0	.000	.000	.000	.955	.000	1.000
1	.921	.002	.151	.952	.000	1.000
2	.999	.793	.395	.949	.000	1.000
3	1.000	.990	.537	.948	.000	.999
4	1.000	.999	.623	.948	.000	.995
5	1.000	1.000	.680	.948	.000	.991
6	1.000	1.000	.721	.949	.000	.985
7	1.000	1.000	.751	.949	.000	.978
8	1.000	1.000	.775	.951	.000	.972
9	1.000	1.000	.794	.952	.000	.965
10	1.000	1.000	.809	.953	.000	.959
20	1.000	1.000	.884	.965	.000	.910
30	1.000	1.000	.913	.973	.000	.882
40	1.000	1.000	.928	.977	.000	.865
50	1.000	1.000	.938	.980	.000	.855
100	1.000	1.000	.960	.988	.000	.840
200	1.000	1.000	.973	.992	.000	.853
300	1.000	1.000	.979	.994	.000	.872
400	1.000	1.000	.982	.995	.000	.890
500	1.000	1.000	.984	.996	.000	.905
1000	1.000	1.000	.989	.997	.126	.951
5000	1.000	1.000	.995	.999	.983	.998
10,000	1.000	1.000	1.000	.999	.997	1.000
100,000	1.000	1.000	1.000	1.000	1.000	1.000
$N_{.95}$	2	3	71	0	3459	0
$N_{.95,\delta,1.5}$	2	3	366	65	4114	17,720
$N_{.95,\delta,3}$	2	3	96	122	5221	9620
N	3	5	749	1051	28,914	1

4.5 The gamma distribution example.

Here we consider sample data X from a $Gamma(\alpha, \beta)$ distribution. In canonical form, this designates the density

$$(4.73) \qquad f(x|\alpha,\theta) = \frac{(-\theta)^{\alpha} x^{\alpha-1} e^{\theta x}}{\Gamma(\alpha)}, \qquad \text{on } x > 0, \quad \text{for } \alpha > 0, \quad \theta < 0.$$

Since $\theta < 0$ in (4.73), we will assume that $b < 0$ in (4.1)—otherwise samples from the gamma distribution will not help to decide between H_0 and H_1 (Assumption 4.3 would not hold). We can rewrite (4.73)

$$(4.74) \qquad f(x|\alpha,\theta) = \exp\Big\{\theta x + [\alpha \ln(-\theta)] + \Big[(\alpha-1)\ln x - \ln\big(\Gamma(\alpha)\big)\Big]\Big\}, \qquad \text{on } x > 0, \quad \text{for } \alpha > 0, \quad \theta < 0.$$

For this gamma density, the scaling term is

$$(4.75) \qquad d(\theta) = \alpha \ln(-\theta).$$

From (4.74), the sample mean $\bar{X}_n$ has the density

$$(4.76) \qquad f(x \mid n\alpha, n\theta) = \frac{(-n\theta)^{n\alpha} x^{n\alpha-1} e^{n\theta x}}{\Gamma(n\alpha)}$$

For the numerical program on page 68, we convert the gamma distribution to the chi-square distribution, which has readily available computer algorithms. Signify the chi-square distribution by χ_k^2 when it has k degrees-of-freedom. Denote by $\ast_k^2(\cdot)$ the cumulative distribution function of the χ_k^2. Letting $Y/2 = -n\theta\bar{X}_n$, the density of Y is

$$g_1(y \mid \alpha) \quad \propto \quad y^{n\alpha-1} e^{-y/2},$$

the $\chi_{2n\alpha}^2$ distribution. We see that the cumulative distribution for $\bar{X}_n$ is

$$P(\bar{X}_n \le A) = P(Y \le -2n\theta A), \qquad \text{for } A \ge 0,$$

or

$$(4.77) \qquad F_n(A \mid n\alpha, n\theta) \;=\; *^2_{2n\alpha}(-2n\theta A)\,, \qquad \text{for } A \geq 0.$$

Assume the prior π_γ is the conjugate prior, with parameters $\zeta_\gamma > 0$ and $\delta_\gamma > 0$, having the gamma density (replace "γ" by "$*$" for the experimenter)

$$(4.78) \quad \pi_\gamma(\theta \mid \zeta_\gamma, \delta_\gamma) \;=\; \frac{\delta_\gamma^{\zeta_\gamma}(-\theta)^{(\zeta_\gamma - 1)} e^{\delta_\gamma \theta}}{\Gamma(\zeta_\gamma)}\,, \qquad \text{on } \theta < 0, \quad \text{for } \zeta_\gamma > 0, \quad \delta_\gamma > 0.$$

Letting $-\xi/2 = \delta_\gamma\theta$, the density of ξ is

$$(4.79) \qquad g_2(\xi \mid \zeta_\gamma, \delta_\gamma) \;=\; \frac{\xi^{(\zeta_\gamma - 1)} e^{-\xi/2}}{\Gamma(\zeta_\gamma) 2^{\zeta_\gamma}}\,,$$

the $\chi^2_{2\zeta_\gamma}$ density. From this we can write the cumulative prior distribution of θ as

$$P(\theta \leq B) = P(\xi \geq -2\delta_\gamma B)\,, \qquad \text{for } B \leq 0$$

or

$$(4.80) \qquad \pi_\gamma\Big((-\infty, B]\Big|\zeta_\gamma, \delta_\gamma\Big) \;=\; 1 - *^2_{2\zeta_\gamma}(-2\delta_\gamma B)\,,$$
$$\text{for } B \leq 0\,, \quad \zeta_\gamma > 0\,, \quad \delta_\gamma > 0.$$

With (4.76) and (4.78), the posterior distribution is

$$\pi_\gamma\Big(\theta \,\Big|\, \bar{x}_n;\; \alpha, \zeta_\gamma, \delta_\gamma\Big) \quad \propto \quad \Big[(-\theta)^{n\alpha} e^{n\theta\bar{x}_n}\Big] \Big[(-\theta)^{(\zeta_\gamma - 1)} e^{\delta_\gamma\theta}\Big]\,,$$

the *Gamma*$(n\alpha + \zeta_\gamma, n\bar{x}_n + \delta_\gamma)$ in canonical form. Letting $\xi/2 = -(n\bar{x}_n + \delta_\gamma)\theta$, the density of ξ is

$$g_3(\xi \mid \alpha, \zeta_\gamma) \;\propto\; (-\xi)^{(n\alpha + \zeta_\gamma - 1)} e^{-\xi/2}\,,$$

the $\chi^2_{2(n\alpha + \zeta_\gamma)}$ distribution. From this we can write the cumulative posterior distribution of θ as

$$P(\theta \leq B) \;=\; P(\xi \geq -2[n\bar{x}_n + \delta_\gamma]B)\,,$$

or

$$(4.81) \qquad \pi_\gamma\Big((-\infty, B] \;\Big|\; \bar{x}_n;\; \alpha,\zeta_\gamma,\delta_\gamma\Big) \;=\; 1 - \text{\ding{83}}^2_{2(n\alpha+\zeta_\gamma)}\Big(-2[n\bar{x}_n + \delta_\gamma]B\Big)\,.$$

The point $\bar{x}_n = z_{n\gamma}$——for which the posterior distribution of θ has median b ——is that $\bar{x}_n$ for which $\pi_\gamma\Big((-\infty, b] \;\Big|\; \bar{x}_n;\; \alpha,\zeta_\gamma,\delta_\gamma\Big) = 0.5$. Denoting the median of the $\chi^2_{2(n\alpha+\zeta_\gamma)}$ distribution by $b_{0.5}$, then

$$(4.82) \qquad z_{n\gamma} = -\frac{1}{n}\left(\frac{b_{0.5}}{2b} + \delta_\gamma\right)\,.$$

This equation (4.82) satisfies (i-ii) in the numerical integration program on page 69. Together, (4.77), (4.78) for "γ"="$*$", and (4.82) can be used in steps (i-ii), (iii) and (iv) to find ρ_n and N_ϵ. Less efficiently, (4.75), (4.77), (4.78) for "γ"="$*$", and (4.80) can be used in steps (i) through (iv) on page 68 of the numerical program to find ρ_n and N_ϵ. Because of its generality, this last program is included in Appendix C as a computer program for the gamma distribution example.

Table 4.2 presents several examples with $\Gamma = \{\pi_1, \pi_2\}$. In the first example, π_1 concentrates on Θ_0, π_2 on Θ_1, while π_* gives moderate probability to both Θ_0 and Θ_1. Notice the large absolute contribution to ρ_n by including a fifth datum in the sample. In the second example, π_1 is the same as the first example π_1, π_2 concentrates on Θ_0 instead of Θ_1, and π_* shifts some probability from Θ_0 to Θ_1. Thus, ρ_n is larger in the second example than in the first example for small n. Since π_1 and π_2 concentrate on Θ_0 in the second example while π_* puts much of its mass on Θ_1, ρ_n is smaller for many larger samples in the second example than in the first example. In the third example, π_1 gives a little over half its probability to Θ_1, π_2 concentrates on Θ_0, and π_* is the prior of the audience member with prior π_1. A sample of but $n = 1$ contributes substantially to the audience's agreement. The probability of correct agreement, ρ_n, is not monotone in the next three examples. In the fourth example, π_1 and π_* concentrate on Θ_1

Table 4.2 ρ_n for composite hypotheses, where X has a Gamma distribution

b	-1	-1	-1	-1	-1	-1
α	0.25	1	1	0.25	0.25	1
δ_1	1	1	0.8	0.8	8.8	5
δ_2	1	1	0.3	0.3	0.9	2
ζ_1	50	50	1	0.3	0.5	1
ζ_2	0.3	13	1	0.3	0.5	1
δ_*	1	3	0.8	2.5	2.5	0.5
$\zeta*$	2	2	1	0.5	0.5	1
n						
0	.000	.199	.000	.975	.975	.393
1	.000	.202	.686	.975	.975	.393
2	.000	.208	.773	.812	.975	.374
3	.000	.216	.815	.850	.954	.367
4	.000	.225	.840	.881	.930	.366
5	.232	.236	.858	.901	.930	.428
6	.345	.247	.871	.914	.933	.574
7	.413	.259	.881	.924	.936	.667
8	.459	.271	.889	.932	.938	.726
9	.492	.284	.896	.937	.941	.766
10	.517	.296	.902	.942	.943	.795
20	.622	.418	.932	.965	.955	.901
30	.658	.515	.945	.973	.959	.930
40	.680	.588	.952	.978	.962	.944
50	.695	.644	.958	.981	.964	.952
100	.745	.793	.970	.987	.975	.970
200	.804	.888	.979	.991	.985	.980
300	.841	.923	.983	.993	.988	.984
400	.866	.941	.985	.994	.991	.987
500	.884	.952	.987	.994	.992	.988
1000	.930	.974	.991	.996	.995	.992
5000	.982	.993	.996	.998	.998	.996
10,000	.989	.995	.997	.999	.999	.997
$N_{.95}$	1504	479	37	0	0	48

while π_2 gives a little over half its probability to Θ_1. Similar to the fourth example, the fifth example has the same π_*, but π_1 concentrates even more on Θ_1 while π_2 shifts some probability to Θ_1. Thus, the non-monotonicity of ρ_n is not as severe. In the sixth example, π_1 and π_2 concentrate on Θ_1 while π_* gives a little more than half its probability to Θ_0. Thus, ρ_n is smaller in this example than the previous two examples.

4.6 High ρ_n for each pair of observers does not imply high ρ_n for all of Γ.

In the simple hypotheses problem of Chapter 3, every audience Γ had two extreme observers π_L and π_U. This need not be true for the composite hypotheses of this chapter. The following is a possible consequence.

Let the audience Γ have m members with priors π_γ, $\gamma = 1, 2, \ldots,$ m. Let $\underset{\sim}{\gamma}_k$ denote one of the $\binom{m}{k}$ combinations of k members, $1 \leq k \leq m$, from among all the members $\{1, 2, \ldots, m\}$. Specifically, $\underset{\sim}{\gamma}_k$ represents the vector of selected integers in increasing order: $\left(\gamma_{(1)}, \gamma_{(2)}, \cdots, \gamma_{\binom{m}{k}}\right)$ where $\gamma_{(1)} < \gamma_{(2)} < \cdots < \gamma_{\binom{m}{k}}$; eg, $\underset{\sim}{\gamma}_4$ could be $(2, 5, 15, m-1)$ or $(1, 2, 5, m)$. Let $\Gamma_{\underset{\sim}{\gamma}_k}$ denote the sub-audience with the priors $\left\{\pi_{\gamma_{(i)}}, \quad i = 1, 2, \ldots, \binom{m}{k}\right\}$, depending on the specific combination $\underset{\sim}{\gamma}_k$. Let $N_{\underset{\sim}{\gamma}_k}$ denote N_ϵ for the combination $\underset{\sim}{\gamma}_k$. Then, for $k < m$, it can be the case that

$$N_\epsilon > \max\left\{N_{\underset{\sim}{\gamma}_k} : \underset{\sim}{\gamma}_k \text{ is one of the } \binom{m}{k} \text{ combinations}\right\}.$$

For example, when $k = 2$ and $m = 3$, it can be the case that

$$N_\epsilon \; > \; \max\left\{N_{(1,2)}, N_{(1,3)}, N_{(2,3)}\right\}.$$

So, finding N_ϵ cannot always be reduced to finding $N_{\underset{\sim}{\gamma}_k}$ for sub-audiences $\Gamma_{\underset{\sim}{\gamma}_k}$. Unless, as mentioned in Section 2.3, there are two extreme observers δ_0 and δ_1. There were two such observers in the Gaussian example above, but not in the Gamma example.

5. SATISFYING ADDITIONAL GOALS INVOLVING OBSERVERS' POSTERIOR LOSSES

5.1 Composite hypotheses.

We now consider an experimenter who wants all observers to choose the correct hypothesis, plus each observer to have some low posterior expected loss for that hypothesis. Our context is largely that of Chapter 2 for two-action problems, and our notation often the same.

An experimenter has an audience Γ of observers γ. Each observer will choose either the hypothesis H_0 or the hypothesis H_1, denoted by action a, $a = a_0$ or $a = a_1$, respectively (a_0 and a_1 are implicitly functions of γ). Only one of H_0 and H_1 is in fact true. When H_0 is true, a parameter θ is from the set Θ_0. When H_1 is true, this parameter θ is from the set Θ_1. The parameter space is designated by $\Theta = \Theta_0 \cup \Theta_1$, assumed a subset of $\mathbf{R}^1$. Each observer γ has a prior probability on Θ, $\pi_\gamma(\theta)$, with respect to a dominating σ-finite measure $\mu(\theta)$. Additionally, he has a loss for his choice of action a_j, $\quad j = 0,\ 1$, when θ is the true parameter: $L_\gamma(a_j, \theta)$.

To aid the observers in their choices, the experimenter provides his audience with data $\underset{\sim}{x}_n = (x_1, x_2, \ldots, x_n)$ from a sample of size n. Each datum x_i comes from the density $f(x|\theta)$ with the same, though unknown, parameter θ. These densities are defined with respect to a dominating σ-finite measure $\lambda(x)$ on the sample space $\mathcal{X} \subseteq \mathbf{R}^1$. For the whole sample of size n, we designate the likelihood function

$$f(\underset{\sim}{x}_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

defined on the product space $\mathcal{X}^n$ with respect to the dominating product measure $\lambda(\underset{\sim}{x}_n) = \prod_{i=1}^{n} \lambda_i(x_i)$.

From the sample $\underset{\sim}{x}_n$, each observer γ forms his posterior density for θ,

$$\pi_\gamma(\theta|\underset{\sim}{x}_n) = [f_\gamma(\underset{\sim}{x}_n)]^{-1} f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\,, \tag{5.1}$$

where

$$f_\gamma(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\,d\mu(\theta)\,. \tag{5.2}$$

With this posterior density dominated by $\mu(\theta)$, we designate the posterior probability of $\Delta \subseteq \Theta$ by

$$\pi_\gamma(\Delta|\underset{\sim}{x}_n) = \int_\Delta \pi_\gamma(\theta|\underset{\sim}{x}_n)\,d\mu(\theta)\,. \tag{5.3}$$

With the posterior density (5.1), observer γ also forms the posterior expected loss for the choice of action a_j:

$$l_\gamma(a_j|\underset{\sim}{x}_n) = \int_\Theta L_\gamma(a_j,\theta)\pi_\gamma(\theta|\underset{\sim}{x}_n)\,d\mu(\theta) \qquad \text{for } j = 0,\ 1. \tag{5.4}$$

If H_0 is true, observer γ will decide correctly when

$$l_\gamma(a_0|\underset{\sim}{x}_n) < l_\gamma(a_1|\underset{\sim}{x}_n)\,. \tag{5.5}$$

If H_1 is true, observer γ will decide correctly when this inequality is reversed. For our experimenter's planning, samples for which randomization—equality in (5.5)—does not occur are sought.

Observer γ wishes that his posterior expected loss be small if he chooses action a_j:

$$l_\gamma(a_j|\underset{\sim}{x}_n) < R_\gamma\,, \tag{5.6}$$

for $R_\gamma \in \mathbf{R}^1$. Our experimenter wishes not just that each observer γ choose the correct action, but also that each observer γ "satisfy," (5.6), the observer's own goals. Combining (5.5) and (5.6), when H_0 is true the experimenter wishes that

$$l_\gamma(a_0|\underset{\sim}{x}_n) < l_\gamma(a_1|\underset{\sim}{x}_n) \quad \text{and} \quad l_\gamma(a_0|\underset{\sim}{x}_n) < R_\gamma . \tag{5.7}$$

When H_1 is true the experimenter wishes that

$$l_\gamma(a_1|\underset{\sim}{x}_n) < l_\gamma(a_0|\underset{\sim}{x}_n) \quad \text{and} \quad l_\gamma(a_1|\underset{\sim}{x}_n) < R_\gamma . \tag{5.8}$$

Let

$$B_{0\gamma} = \{\underset{\sim}{x}_n : l_\gamma(a_0|\underset{\sim}{x}_n) < l_\gamma(a_1|\underset{\sim}{x}_n)\}, \tag{5.9}$$

$$B_{1\gamma} = \{\underset{\sim}{x}_n : l_\gamma(a_1|\underset{\sim}{x}_n) < l_\gamma(a_0|\underset{\sim}{x}_n)\}, \tag{5.10}$$

$$D_{0\gamma} = \{\underset{\sim}{x}_n : l_\gamma(a_0|\underset{\sim}{x}_n) < R_\gamma\}, \tag{5.11}$$

$$D_{1\gamma} = \{\underset{\sim}{x}_n : l_\gamma(a_1|\underset{\sim}{x}_n) < R_\gamma\}, \tag{5.12}$$

$$E_0 = \bigcap_{\gamma \in \Gamma} B_{0\gamma} \bigcap_{\gamma \in \Gamma} D_{0\gamma}, \tag{5.13}$$

and

$$E_1 = \bigcap_{\gamma \in \Gamma} B_{1\gamma} \bigcap_{\gamma \in \Gamma} D_{1\gamma}. \tag{5.14}$$

When H_0 is true the experimenter wishes that $\underset{\sim}{x}_n \in E_0$. When H_1 is true the experimenter wishes that $\underset{\sim}{x}_n \in E_1$.

As for the experimenter, we denote the experimenter's prior by $\pi_*(\theta)$, and his posterior by $\pi_*(\theta|\underset{\sim}{x}_n)$. Thus, that each observer γ be satisfied and choose the correct hypothesis, the experimenter assesses the probability

$$\psi_n \equiv \sum_{j=0}^{1} \int_{\Theta_j} \int_{E_j} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta). \tag{5.15}$$

The experimenter wants to choose a sample size n for which

$$\psi_n \geq \epsilon \tag{5.16}$$

for some $0 < \epsilon < 1$ specified by him.

So that every observer might make the correct decision—(5.16) is not precluded—we assume, as in Appendix A,

ASSUMPTION 5.1. *Excepting a set $B \subset \Theta$, $\pi_*(B) = 0$, for all $\gamma \in \Gamma$:*

$$L_\gamma(a_1, \theta) - L_\gamma(a_0, \theta) > 0 \qquad \text{if } \theta \in \Theta_0$$

and

$$L_\gamma(a_0, \theta) - L_\gamma(a_1, \theta) > 0 \qquad \text{if } \theta \in \Theta_1.$$

So that every observer γ can be satisfied, (5.6), with the correct decision—(5.16) is not precluded—we assume

ASSUMPTION 5.2. *Excepting a set $B \subset \Theta$, $\pi_*(B) = 0$, for $j = 0, 1$ and all $\gamma \in \Gamma$:*

$$L_\gamma(a_j, \theta) - R_\gamma < 0 \qquad \text{if } \theta \in \Theta_j.$$

We now make some definitions which will facilitate expressing the posterior loss criteria for an observer $\gamma \in \Gamma$, (5.7) and (5.8), as posterior probability criteria. The result will be a problem having the same form as the problem of Chapter 2. Let

$$\Theta_{0\gamma} = \{\theta : L_\gamma(a_0, \theta) < R_\gamma\} \tag{5.17}$$

and

$$\Theta_{1\gamma} = \{\theta : L_\gamma(a_1, \theta) < R_\gamma\}. \tag{5.18}$$

By Assumption 5.2,

$$\Theta_0 \subseteq \Theta_{0\gamma} \tag{5.19}$$

and

$$(5.20) \qquad \Theta_1 \subseteq \Theta_{1\gamma}.$$

For each $\gamma \in \Gamma$, define three new priors on Θ:

$$(5.21) \qquad \pi_{0\gamma}(\theta) = c_{0\gamma}^{-1} \, |L_\gamma(a_0,\theta) - R_\gamma| \, \pi_\gamma(\theta)$$

where

$$c_{0\gamma} = \int_\Theta |L_\gamma(a_0,\theta) - R_\gamma| \pi_\gamma(\theta) \, d\mu(\theta),$$

$$(5.22) \qquad \pi_{1\gamma}(\theta) = c_{1\gamma}^{-1} \, |L_\gamma(a_1,\theta) - R_\gamma| \, \pi_\gamma(\theta)$$

where

$$c_{1\gamma} = \int_\Theta |L_\gamma(a_1,\theta) - R_\gamma| \pi_\gamma(\theta) \, d\mu(\theta),$$

and

$$(5.23) \qquad \tilde{\pi}_\gamma(\theta) = \begin{cases} \tilde{c}_\gamma^{-1} \pi_\gamma(\theta) \big[L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta) \big] & \text{if } \theta \in \Theta_0 \\ \tilde{c}_\gamma^{-1} \pi_\gamma(\theta) \big[L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta) \big] & \text{if } \theta \in \Theta_1 \end{cases}$$

where

$$\tilde{c}_\gamma = \int_{\Theta_0} \pi_\gamma(\theta) \big[L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta) \big] \, d\mu(\theta) + \int_{\Theta_1} \pi_\gamma(\theta) \big[L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta) \big] \, d\mu(\theta).$$

From (5.17), (5.21) and Assumption 5.2, we may write the condition

$$\pi_{0\gamma}(\Theta_{0\gamma} | \underset{\sim}{x}_n) > 0.5$$

as the condition

$$\int_{\Theta_{0\gamma}} \big[R_\gamma - L_\gamma(a_0,\theta) \big] \pi_\gamma(\theta) f(\underset{\sim}{x}_n|\theta) \, d\mu(\theta) > \int_{\Theta - \Theta_{0\gamma}} \big[L_\gamma(a_0,\theta) - R_\gamma \big] \pi_\gamma(\theta) f(\underset{\sim}{x}_n|\theta) \, d\mu(\theta).$$

Equivalently,

$$\int_\Theta L_\gamma(a_0,\theta) \pi_\gamma(\theta) f(\underset{\sim}{x}_n|\theta) \, d\mu(\theta) < R_\gamma \int_\Theta \pi_\gamma(\theta) f(\underset{\sim}{x}_n|\theta) \, d\mu(\theta);$$

that is,

$$l_\gamma(a_0 | \underset{\sim}{x}_n) < R_\gamma.$$

Thus, from (5.11),

$$D_{0\gamma} = \{\underset{\sim}{x}_n : \pi_{0\gamma}(\Theta_{0\gamma}|\underset{\sim}{x}_n) > 0.5\}. \tag{5.24}$$

Similarly, from (5.12), (5.19), (5.22) and Assumption 5.2

$$D_{1\gamma} = \{\underset{\sim}{x}_n : \pi_{1\gamma}(\Theta_{1\gamma}|\underset{\sim}{x}_n) > 0.5\}, \tag{5.25}$$

from (5.9), (5.23) and Assumption 5.1

$$B_{0\gamma} = \{\underset{\sim}{x}_n : \tilde{\pi}_{\gamma}(\Theta_0|\underset{\sim}{x}_n) > 0.5\}, \tag{5.26}$$

and from (5.10), (5.23) and Assumption 5.1

$$B_{1\gamma} = \{\underset{\sim}{x}_n : \tilde{\pi}_{\gamma}(\Theta_1|\underset{\sim}{x}_n) > 0.5\}. \tag{5.27}$$

Define

$$V_{0\gamma} = V_{0\gamma}(H_0|\underset{\sim}{x}_n) = \min\Big\{\pi_{0\gamma}(\Theta_{0\gamma}|\underset{\sim}{x}_n), \quad \tilde{\pi}_{\gamma}(\Theta_0|\underset{\sim}{x}_n)\Big\},$$

$$V_{1\gamma} = V_{1\gamma}(H_1|\underset{\sim}{x}_n) = \min\Big\{\pi_{1\gamma}(\Theta_{1\gamma}|\underset{\sim}{x}_n), \quad \tilde{\pi}_{\gamma}(\Theta_1|\underset{\sim}{x}_n)\Big\},$$

$$V_0 = V_0(H_0|\underset{\sim}{x}_n) = \inf_{\gamma\in\Gamma} V_{0\gamma}, \tag{5.28}$$

and

$$V_1 = V_1(H_1|\underset{\sim}{x}_n) = \inf_{\gamma\in\Gamma} V_{1\gamma}. \tag{5.29}$$

From the definitions of E_0 and E_1 in (5.13) and (5.14); and from (5.24), (5.25), (5.26) and (5.27), then

$$E_j = \Big\{\underset{\sim}{x}_n \colon V_{j\gamma} > 0.5, \quad \text{all } \gamma \in \Gamma\Big\} \tag{5.30}$$

and

$$E_j \supseteq \tilde{E}_j \equiv \Big\{\underset{\sim}{x}_n \colon V_j > 0.5\Big\} \qquad \text{for both } j = 0, 1. \tag{5.31}$$

Defining

$$\tilde{\psi}_n = \sum_{j=0}^{1} \int_{\Theta_j} \int_{\tilde{E}_j} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta), \tag{5.32}$$

then

$$\tilde{\psi}_n \le \psi_n\,.$$

If Γ is "closed," then

$$E_j = \tilde{E}_j$$

and

$$\tilde{\psi}_n = \psi_n\,.$$

Thus, through the simplification of E_j in (5.30), ψ_n in (5.15) is a function of the probabilities $V_{0\gamma}$ and $V_{1\gamma}$, themselves functions of $\pi_{0\gamma}$, $\pi_{1\gamma}$ and $\tilde{\pi}_\gamma$. Consequently, these simplifications of ψ_n allow the calculation of ψ_n to be calculated instead as ρ_n in (2.4) of Chapter 2, with E_i as A_i in (2.3), $i = 1, 2$. In addition, V_0 and V_1 in (5.28) and (5.29) allow the calculation of $\tilde{\psi}_n$ in (5.32) to be calculated instead as $\tilde{\rho}_n$ in (2.22). This reformulation of (5.16) as (2.5) has tripled the number of priors; ie, tripled the size of Γ. At the same time, this reformulation allows the use of 0-1 losses, and it subsumes the observers' posterior loss goals.

With this reformulation, we now add a few more assumptions which will guarantee a finite n satisfying (5.16). So that no observer precludes the correct decision through his prior, we assume

ASSUMPTION 5.3. *The support of every $\pi_\gamma(\theta)$, $\gamma \in \Gamma$, contains the support of $\pi_*(\theta)$.*

That is, for every $\gamma \in \Gamma$,

$$\Big\{\theta\colon\ \pi_\gamma(\theta)\,d\mu(\theta) > 0\Big\} \quad \supseteq \quad \Big\{\theta\colon\ \pi_*(\theta)\,d\mu(\theta) > 0\Big\}.$$

So that all the observers are correctly satisfied at a finite n—(5.16) is attainable for the whole of Γ at once—we make the following assumption which disallows a factor of the prior and a factor of the loss to multiply to an extreme value for too many θ.

ASSUMPTION 5.4. *For each $\delta > 0$ there is a $k_\delta > 0$ and a Borel set $G_\delta \subset \Theta$ for which $\pi_*(G_\delta) < \delta$ and*

(i) $k_\delta < \pi_{0\gamma}(\theta) \big/ \pi_{0\gamma'}(\theta)$ *for $\theta \in \Theta - G_\delta$ and $\gamma, \gamma' \in \Gamma$,*

and

$$\int_{G_\delta} \sup_{\gamma \in \Gamma} \pi_{0\gamma}(\theta)\, d\mu(\theta) < \infty$$

(ii) $k_\delta < \pi_{1\gamma}(\theta) \big/ \pi_{1\gamma'}(\theta)$ *for $\theta \in \Theta - G_\delta$ and $\gamma, \gamma' \in \Gamma$,*

and

$$\int_{G_\delta} \sup_{\gamma \in \Gamma} \pi_{1\gamma}(\theta)\, d\mu(\theta) < \infty$$

(iii) $k_\delta < \tilde{\pi}_{\gamma}(\theta) \big/ \tilde{\pi}_{\gamma'}(\theta)$ *for $\theta \in \Theta - G_\delta$ and $\gamma, \gamma' \in \Gamma$,*

and

$$\int_{G_\delta} \sup_{\gamma \in \Gamma} \tilde{\pi}_{\gamma}(\theta)\, d\mu(\theta) < \infty .$$

For this assumption, we make the following observations. First, since we allow $\gamma = \gamma'$ in (i), (ii), and (iii), then $0 < k_\delta \leq 1$. Second, this assumption does not disallow that $\pi_{\gamma'}(\Theta_0) = 0$ or 1 for some $\gamma' \in \Gamma$. But it does demand that $\pi_\gamma(\Theta_0) = 0$ or 1, respectively, for all $\gamma \in \Gamma$ and for $\pi_*(\Theta_0)$ then. Accordingly, $n = 0$ satisfies (5.16) then. Third, for some problems (notably, when Θ has finitely many elements) the integrals on G_δ in (i), (ii), and (iii) are necessarily finite for any G. Fourth, from assumptions 5.1 and 5.2, and from definitions (5.21), (5.22), and (5.23); if Γ is finite and Assumption 5.3 holds, then Assumption 5.4 holds perfunctorily. Fifth, since $\pi_*(G_\delta)$ can be made arbitrarily small, Assumption 5.4 implies Assumption 5.3. In order that a Bayes rule will be consistent in its choice of hypothesis, we make the following assumptions about the sampling distributions:

ASSUMPTION 5.5. *$P_\theta(X < x)$ is a Baire function of θ for each fixed $x \in \mathcal{X}$.*

ASSUMPTION 5.6 . *The parameterization is identifiable; ie, for each pair of parameters $\theta \neq \theta'$ in Θ there exists a set $A \subseteq \mathcal{X}$ for which $P_\theta(A) \neq P_{\theta'}(A)$.*

We now prove that the posterior distribution asymptotically concentrates around the parameter. The proof follows a proof by Schwartz (1965, Theorem 3.2). We require only assumptions 5.5 and 5.6—no assumptions are made about the moments of the prior. We use the notation $\mathcal{X}^\infty$ for the Cartesian product of countably infinite $\mathcal{X}$ spaces, and we denote an element of $\mathcal{X}^\infty$ by $\underset{\sim}{x}_\infty$.

LEMMA 5.1 . *Make assumptions 5.5 and 5.6. Let $\dot{\pi}$ be any prior measure on Θ, and let M be any Borel set for which $M \subseteq \Theta$. Then there is a set $A \subset \Theta$, $\dot{\pi}(A) = 0$, for which*

$$P_\theta\left\{\underset{\sim}{x}_\infty : \lim_{n\to+\infty} \dot{\pi}(M|\underset{\sim}{x}_n) = I(\theta \in M)\right\} = 1\,, \qquad \textit{when } \theta \in \Theta - A\,.$$

PROOF OF LEMMA 5.1.

Let the space $\Omega = \Theta \times \mathcal{X}^\infty$, let $\mathcal{B}$ be the σ-field generated by the Borel sets of Θ, and let $\mathcal{U}$ be the σ-field generated by the m-rectangles of $\{\mathcal{X}^m, \quad m = 1, 2, \ldots\}$. Let ξ be the measure on $\mathcal{B} \times \mathcal{U}$ determined by $\dot{\pi}$ and $\{P_\theta, \quad \text{all } \theta \in \Theta\}$. For $\omega \in \Omega$, define

$$\zeta \;=\; \zeta(\omega) \;=\; \zeta(\theta, \underset{\sim}{x}_\infty) \;=\; \zeta(\theta) \;=\; I(\theta \in M)\,,$$

and define

$$\beta_n \;=\; \beta_n(\omega) \;=\; \beta_n(\underset{\sim}{x}_\infty) \;=\; E(\zeta|\underset{\sim}{x}_\infty)\,.$$

Since $E|\zeta| \leq 1$, then $\{\beta_n\}$ forms a martingale sequence. By the martingale convergence theorem,

$$\beta_n \longrightarrow E(\zeta|\underset{\sim}{x}_\infty) \qquad \text{a.s.}(\xi)\,.$$

Assumptions 5.5 and 5.6 imply (see Theorem 3.1 in Schwartz) the existence of some $\mathcal{U}$-measurable function h on $\mathcal{X}$ such that

$$\theta = h(\underline{x}_\infty) \qquad \text{a.s.}(P_\theta)$$

for each $\theta \in \Theta$.

Let

$$C \;=\; \left\{\omega = (\theta, \underline{x}_\infty)\colon\ h(\underline{x}_\infty) = \theta\right\}$$

and

$$D_\theta \;=\; \left\{\underline{x}_\infty\colon\ h(\underline{x}_\infty) = \theta\right\}.$$

Then

$$1 \;=\; \int_\Theta P_\theta(D_\theta)\, d\dot{\pi}(\theta) \;=\; \int_\Theta\int_{D_\theta} dP_\theta(\underline{x}_\infty)\, d\dot{\pi}(\theta) \;=\; \xi(C).$$

That is,

$$\theta = h(\underline{x}_\infty) \qquad \text{a.s.}(\xi).$$

So,

$$\begin{aligned} E\Big(\zeta(\omega)\;\Big|\;\underline{x}_\infty\Big) \;=\; E\Big(\zeta(\omega)\;\Big|\;h(\underline{x}_\infty), \underline{x}_\infty\Big) \;&=\; E\Big(\zeta(\omega)\;\Big|\;\theta, \underline{x}_\infty\Big) \\ &=\; \zeta(\theta, \underline{x}_\infty) \qquad \text{a.s.}(\xi), \end{aligned}$$

implying that

$$\lim_{n\to+\infty} \beta_n(\omega) = \zeta(\omega) \qquad \text{a.s.}(\xi).$$

Let

$$\hat{C} \;=\; \left\{\omega\colon \beta_n(\omega) \to \zeta(\omega)\right\}$$

and

$$\hat{D}_\theta \;=\; \left\{\underline{x}_\infty\colon \beta_n(\underline{x}_\infty) \to I(\theta \in M)\right\}.$$

Then

$$1 \;=\; \xi(\hat{C}) \;=\; \int_{\Theta}\int_{\hat{D}_\theta} dP_\theta(\underset{\sim}{x}_\infty)\, d\dot{\pi}(\theta) \;=\; \int_{\Theta} P_\theta(\hat{D}_\theta)\, d\dot{\pi}(\theta)\,,$$

so that

$$P_\theta(\hat{D}_\theta) = 1 \qquad \text{a.s.}(\dot{\pi}).$$

That is,

$$P_\theta\Big\{\underset{\sim}{x}_\infty\colon\ \dot{\pi}(M|\underset{\sim}{x}_n) \to I(\theta \in M)\Big\} = 1 \qquad \text{a.s.}(\dot{\pi}). \qquad \square$$

We now show, for some sample size n, that all observers can satisfy their goals while choosing the correct hypothesis, with a high probability ϵ anyway—(5.16) is attainable.

THEOREM 5.2. *Under assumptions 5.1, 5.2, 5.4, 5.5 and 5.6,*

$$\lim_{n\to+\infty} \psi_n = 1\,.$$

PROOF OF THEOREM 5.2.

Let $\delta > 0$. From Assumption 5.4, there is a set $G_\delta \subset \Theta$ for which $\pi_*(G_\delta) < \delta$, and there is a $k_\delta > 0$ satisfying (i), (ii) and (iii) there. Define, for some specific $\gamma' \in \Gamma$,

$$\dot{\pi}(\theta) = \begin{cases} g^{-1}k_\delta^2\pi_{0\gamma'}(\theta) & \text{on } \Theta_0 - G_\delta \\ g^{-1}\pi_{0\gamma'}(\theta) & \text{on } \Theta_1 - G_\delta \\ g^{-1}k_\delta \sup\limits_{\gamma\in\Gamma} \pi_{0\gamma}(\theta) & \text{on } G_\delta, \end{cases}$$

where

$$g \;=\; k_\delta^2 \int\limits_{\Theta_0 - G_\delta} \pi_{0\gamma'}(\theta)\, d\mu(\theta) \;+\; \int\limits_{\Theta_1 - G_\delta} \pi_{0\gamma'}(\theta)\, d\mu(\theta) \;+\; \int\limits_{G_\delta} \sup_{\gamma\in\Gamma} \pi_{0\gamma}(\theta)\, d\mu(\theta)\,,$$

which is finite by Assumption 5.4(i). For any $\gamma \in \Gamma$,

$$\pi_{0\gamma}(\Theta_0 \mid \underset{\sim}{x}_n) \;\geq\; \pi_{0\gamma}(\Theta_0 - G_\delta|\underset{\sim}{x}_n)\,.$$

From Assumption 5.4(i),

$$(5.33)\qquad \pi_{0\gamma}(\Theta_0 - G_\delta|\underset{\sim}{x}_n) \;\geq\; g_1^{-1}k_\delta \int_{\Theta_0 - G_\delta} \pi_{0\gamma'}(\theta) f(\underset{\sim}{x}_n|\theta)\, d\mu(\theta)\,,$$

where

$$g_1 \;=\; k_\delta \int_{\Theta_0 - G_\delta} \pi_{0\gamma'}(\theta) f(\underset{\sim}{x}_n|\theta)\, d\mu(\theta) \;+\; k_\delta^{-1} \int_{\Theta_1 - G_\delta} \pi_{0\gamma'}(\theta) f(\underset{\sim}{x}_n|\theta)\, d\mu(\theta) \\ +\int_{G_\delta} \sup_{\gamma\in\Gamma} \pi_{0\gamma}(\theta) f(\underset{\sim}{x}_n|\theta)\, d\mu(\theta)\,.$$

Since the right hand term of (5.33) is just $\dot{\pi}(\Theta_0 - G_\delta|\underset{\sim}{x}_n)$, we now have

$$(5.34)\qquad \pi_{0\gamma}(\Theta_0|\underset{\sim}{x}_n) \;\geq\; \dot{\pi}(\Theta_0 - G_\delta|\underset{\sim}{x}_n)\,.$$

By Lemma 5.1 (here we use assumptions 5.5 and 5.6), if $\theta_0 \in \Theta_0 - G_\delta$ then

$$(5.35)\qquad \lim_{n\to+\infty} \dot{\pi}(\Theta_0 - G_\delta|\underset{\sim}{x}_n) \;=\; 1$$

a.s.$(P_{\theta_0} \times \dot{\pi})$, a forteriori a.s.$(P_{\theta_0} \times \pi_*)$ since Assumption 5.4 implies Assumption 5.3. Similarly, there is a $\ddot{\pi}$ for which

$$(5.36)\qquad \tilde{\pi}_\gamma(\Theta_0|\underset{\sim}{x}_n) \;\geq\; \ddot{\pi}(\Theta_0 - G_\delta|\underset{\sim}{x}_n) \overset{n\to+\infty}{\longrightarrow} 1 \qquad \text{a.s. } (P_{\theta_0} \times \pi_*).$$

So, $\tilde{E}_0$ in (5.31) satisfies (implicitly using Assumptions 5.1 and 5.2 for the formation of (5.31))

$$(5.37)\qquad E_0 \supseteq \tilde{E}_0 \supseteq \dot{E}_0\,,$$

where

$$\dot{E}_0 \;=\; \left\{\underset{\sim}{x}_n\colon \min\left\{\dot{\pi}(\Theta_0 - G_\delta|\underset{\sim}{x}_n),\; \ddot{\pi}(\Theta_0 - G_\delta|\underset{\sim}{x}_n)\right\} > 0.5\right\}.$$

Combining (5.35), (5.36) and (5.37), we have that

$$(5.38)\qquad \lim_{n\to+\infty} \int_{\Theta_0 - G_\delta} \int_{E_0} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \;=\; \pi_*(\Theta_0 - G_\delta)\,.$$

A parallel argument shows that

$$(5.39) \qquad \lim_{n\to+\infty} \int_{\Theta_1 - G_\delta} \int_{E_1} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \;=\; \pi_*(\Theta_1 - G_\delta)\,.$$

Together with (5.38) and (5.39), $\pi_*(G_\delta) < \delta$ in Assumption 5.4 implies that

$$\psi_n \geq 1 - \delta \qquad \text{if } n > {}_\delta N,$$

for some ${}_\delta N > 0$. Since δ was arbitrary, our theorem holds:

$$\lim_{n\to+\infty} \psi_n = 1\,. \quad \square$$

5.2 Simple hypotheses.

Here we consider the special case of simple hypotheses, $H_0\colon \theta = \theta_0$ and $H_1\colon \theta = \theta_1$, as in Chapter 3. Let

$$T_n \;=\; T_n(\underset{\sim}{x}_n) \;=\; \ln\left[\frac{f(\underset{\sim}{x}_n|\theta_1)}{f(\underset{\sim}{x}_n|\theta_0)}\right]$$

as in (3.4). We will use the simplified notation

$$L_{\gamma ij} = L_\gamma(a_i, \theta_j) \qquad \text{with } i,j = 0,1$$

for the losses. And we will use the simplified notation

$$\pi_\gamma \;=\; \pi_\gamma(H_0) \;=\; 1 - \pi_\gamma(H_1)$$

and

$$\pi_* \;=\; \pi_*(H_0) \;=\; 1 - \pi_*(H_1)$$

for the priors.

Assumption 5.1 implies that for all $\gamma \in \Gamma$

$$L_{\gamma 10} > L_{\gamma 00} \qquad \text{and} \qquad L_{\gamma 01} > L_{\gamma 11}\,.$$

Assumption 5.2 implies that for all $\gamma \in \Gamma$

$$L_{\gamma 00} < R_\gamma \qquad \text{and} \qquad L_{\gamma 11} < R_\gamma .$$

Occasionally we will indicate that one of the following two conditions holds:

Condition $C_\gamma 1$. $\Theta_{0\gamma} = \Theta_0$; *ie,* $L_{\gamma 01} > R_\gamma$.

Condition $C_\gamma 2$. $\Theta_{1\gamma} = \Theta_1$; *ie,* $L_{\gamma 10} > R_\gamma$.

We can write the following sets more simply as

$$\begin{aligned}
B_{0\gamma} &= \left\{ \underline{x}_n \colon \pi_\gamma(\theta_0|\underline{x}_n) > \frac{L_{\gamma 01} - L_{\gamma 11}}{(L_{\gamma 01} - L_{\gamma 11}) + (L_{\gamma 10} - L_{\gamma 00})} \right\} \\
&= \left\{ \underline{x}_n \colon T_n < \ln\left[\frac{L_{\gamma 10} - L_{\gamma 00}}{L_{\gamma 01} - L_{\gamma 11}} \frac{\pi_\gamma}{1 - \pi_\gamma} \right] \right\}, \\
B_{1\gamma} &= \left\{ \underline{x}_n \colon \pi_\gamma(\theta_0|\underline{x}_n) < \frac{L_{\gamma 01} - L_{\gamma 11}}{(L_{\gamma 01} - L_{\gamma 11}) + (L_{\gamma 10} - L_{\gamma 00})} \right\} \\
&= \left\{ \underline{x}_n \colon T_n > \ln\left[\frac{L_{\gamma 10} - L_{\gamma 00}}{L_{\gamma 01} - L_{\gamma 11}} \frac{\pi_\gamma}{1 - \pi_\gamma} \right] \right\}, \\
D_{0\gamma} &= \left\{ \underline{x}_n \colon \pi_\gamma(\theta_0|\underline{x}_n) > \frac{L_{\gamma 01} - R_\gamma}{L_{\gamma 01} - L_{\gamma 00}} \right\} \\
&= \begin{cases} \left\{ \underline{x}_n \colon T_n < \ln\left[\frac{R_\gamma - L_{\gamma 00}}{L_{\gamma 01} - R_\gamma} \frac{\pi_\gamma}{1-\pi_\gamma} \right] \right\} & \text{if Condition } C_\gamma 1 \text{ holds} \\ \mathcal{X}^n & \text{otherwise} \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
D_{1\gamma} &= \left\{ \underline{x}_n \colon \pi_\gamma(\theta_0|\underline{x}_n) < \frac{R_\gamma - L_{\gamma 11}}{L_{\gamma 10} - L_{\gamma 11}} \right\} \\
&= \begin{cases} \left\{ \underline{x}_n \colon T_n > \ln\left[\frac{L_{\gamma 10} - R_\gamma}{R_\gamma - L_{\gamma 11}} \frac{\pi_\gamma}{1-\pi_\gamma} \right] \right\} & \text{if Condition } C_\gamma 2 \text{ holds} \\ \mathcal{X}^n & \text{otherwise.} \end{cases}
\end{aligned}$$

With these sets, the experimenter's probability ψ_n in (5.15) may be written for simple hypotheses as

$$\psi_n = \pi_* P_{\theta_0}(E_0) + (1 - \pi_*) P_{\theta_1}(E_1), \tag{5.40}$$

where

$$E_0 = \bigcap_{\gamma \in \Gamma} B_{0\gamma} \bigcap_{\gamma \in \Gamma} D_{0\gamma}$$

and

$$E_1 = \bigcap_{\gamma\in\Gamma} B_{1\gamma} \bigcap_{\gamma\in\Gamma} D_{1\gamma}\,.$$

With E_j in (5.30), the simplification of ψ_n in (5.15) to a form of ρ_n in Chapter 2 occurred through (5.21), (5.22) and (5.23). We now give this simplification for the simple hypotheses case. If Condition $C_\gamma 1$ holds, (5.21) gives

$$\begin{aligned}
\pi_{0\gamma} &= \pi_{0\gamma}(\Theta_{0\gamma}) \\
&= \frac{(R_\gamma - L_{\gamma 00})\pi_\gamma}{(R_\gamma - L_{\gamma 00})\pi_\gamma + |L_{\gamma 01} - R_\gamma|(1-\pi_\gamma)} && (5.41)\\
&= \left[1 + \frac{1-\pi_\gamma}{\pi_\gamma}\frac{|L_{\gamma 01} - R_\gamma|}{R_\gamma - L_{\gamma 00}}\right]^{-1} && (5.42)\\
&= \pi_\gamma + \pi_\gamma(1-\pi_\gamma)\frac{(R_\gamma - L_{\gamma 00}) - |L_{\gamma 01} - R_\gamma|}{(R_\gamma - L_{\gamma 00})\pi_\gamma + |L_{\gamma 01} - R_\gamma|(1-\pi_\gamma)}\,, && (5.43)
\end{aligned}$$

otherwise $\pi_{0\gamma} = 1$. If Condition $C_\gamma 2$ holds, (5.22) gives

$$\begin{aligned}
\pi_{1\gamma} &= \pi_{1\gamma}(\bar{\Theta}_{1\gamma}) \\
&= \frac{|L_{\gamma 10} - R_\gamma|\pi_\gamma}{|L_{\gamma 10} - R_\gamma|\pi_\gamma + (R_\gamma - L_{\gamma 11})(1-\pi_\gamma)} && (5.44)\\
&= \left[1 + \frac{1-\pi_\gamma}{\pi_\gamma}\frac{R_\gamma - L_{\gamma 11}}{|L_{\gamma 10} - R_\gamma|}\right]^{-1} && (5.45)\\
&= \pi_\gamma + \pi_\gamma(1-\pi_\gamma)\frac{|L_{\gamma 10} - R_\gamma| - (R_\gamma - L_{\gamma 11})}{|L_{\gamma 10} - R_\gamma|\pi_\gamma - (R_\gamma - L_{\gamma 11})(1-\pi_\gamma)}\,, && (5.46)
\end{aligned}$$

otherwise $\pi_{1\gamma} = 0$. From (5.23)

$$\begin{aligned}
\tilde{\pi}_\gamma &= \tilde{\pi}_\gamma(H_0) \\
&= \frac{(L_{\gamma 10} - L_{\gamma 00})\pi_\gamma}{(L_{\gamma 10} - L_{\gamma 00})\pi_\gamma + (L_{\gamma 01} - L_{\gamma 11})(1-\pi_\gamma)} && (5.47)\\
&= \left[1 + \frac{1-\pi_\gamma}{\pi_\gamma}\frac{L_{\gamma 01} - L_{\gamma 11}}{L_{\gamma 10} - L_{\gamma 00}}\right]^{-1} && (5.48)\\
&= \pi_\gamma + \pi_\gamma(1-\pi_\gamma)\frac{(L_{\gamma 10} - L_{\gamma 00}) - (L_{\gamma 01} - L_{\gamma 11})}{(L_{\gamma 10} - L_{\gamma 00})\pi_\gamma + (L_{\gamma 01} - L_{\gamma 11})(1-\pi_\gamma)}\,. && (5.49)
\end{aligned}$$

Used in E_0 and E_1 of (5.30),

$$E_j = \left\{ \underset{\sim}{x}_n \colon V_{j\gamma} > 0.5\,, \quad \text{all } \gamma \in \Gamma \right\}, \tag{5.50}$$

where

(5.51) $$V_{0\gamma} = \min\left\{\pi_{0\gamma}(\theta_0|\underset{\sim}{x}_n),\ \tilde{\pi}_\gamma(\theta_0|\underset{\sim}{x}_n)\right\}$$

and

(5.52) $$V_{1\gamma} = \min\left\{1-\pi_{1\gamma}(\theta_0|\underset{\sim}{x}_n),\ 1-\tilde{\pi}_\gamma(\theta_0|\underset{\sim}{x}_n)\right\}.$$

Let

$$\pi_L = \inf_{\gamma\in\Gamma}\left\{\min\left\{\pi_{0\gamma},\tilde{\pi}_\gamma\right\}\right\}$$

and

$$\pi_U = \sup_{\gamma\in\Gamma}\left\{\max\left\{\pi_{1\gamma},\tilde{\pi}_\gamma\right\}\right\}.$$

Then

$$V_0 = \pi_L(\theta_0|\underset{\sim}{x}_n)$$

and

$$V_1 = 1-\pi_U(\theta_0|\underset{\sim}{x}_n).$$

Observe that the inclusion of $\tilde{\pi}_\gamma$ in the definition of π_L and π_U makes $\pi_L \le \pi_U$. V_0 and V_1 define $\tilde{E}_j = \{\underset{\sim}{x}_n : V_j > 0.5\}$ in (5.31).

Define an audience $\dddot{\Gamma}$ indexed by $\dddot{\gamma}$. Let each $\dddot{\gamma}$ have but one of the priors $\pi_{0\gamma}$, $\pi_{1\gamma}$ or $\tilde{\pi}_\gamma$ and let

$$\bigcup_{\gamma\in\Gamma}\left\{\pi_{0\gamma},\pi_{1\gamma},\tilde{\pi}_\gamma\right\} = \bigcup_{\dddot{\gamma}\in\dddot{\Gamma}}\pi_{\dddot{\gamma}}.$$

Then

$$\pi_L = \inf_{\dddot{\gamma}\in\dddot{\Gamma}}\left\{\pi_{\dddot{\gamma}}\right\} \quad \text{and} \quad \pi_U = \sup_{\dddot{\gamma}\in\dddot{\Gamma}}\left\{\pi_{\dddot{\gamma}}\right\}.$$

As mentioned on page 105, our reformulation tripled the size of the audience Γ to $\dddot{\Gamma}$. When $\pi_{0\gamma} < \pi_{1\gamma}$, $\pi_{0\gamma} < \tilde{\pi}_\gamma$, or $\tilde{\pi}_\gamma < \pi_{1\gamma}$, then the single observer γ, with prior π_γ, in a singleton audience Γ for ψ_n is represented by many observers $\dddot{\gamma}$, with priors $\pi_{0\gamma}$, $\pi_{1\gamma}$ and $\tilde{\pi}_\gamma$. These priors span an interval $[\pi_L,\pi_U]$ with $\pi_L < \pi_U$, in a multitudinous audience $\dddot{\Gamma}$ for the simpler ρ_n.

The above has reduced the experimenter's probability ψ_n in (5.40) to

$$\rho_n = \psi_n$$

using

$$A_0 = E_0 = \left\{V_{0\gamma} > 0.5, \quad \text{all } \gamma \in \Gamma\right\}$$

and

$$A_1 = E_1 = \left\{V_{1\gamma} > 0.5, \quad \text{all } \gamma \in \Gamma\right\}.$$

The above has also reduced $\tilde{\psi}_n$ to

$$\tilde{\rho}_n = \tilde{\psi}_n$$

using

$$\tilde{A}_0 = \tilde{E}_0 = \left\{V_0 > 0.5\right\} = \left\{T_n < \ln\left(\frac{\pi_L}{1-\pi_L}\right)\right\}$$

and

$$\tilde{A}_1 = \tilde{E}_1 = \left\{V_1 > 0.5\right\} = \left\{T_n > \ln\left(\frac{\pi_U}{1-\pi_U}\right)\right\}$$

as in Chapter 3, with the audience $\ddot{\Gamma}$. The audience has been reduced to two members, π_L and π_U.

When conditions $C_\gamma 1$ and $C_\gamma 2$ fail, then $\pi_{0\gamma} = 1$ and $\pi_{1\gamma} = 0$, respectively, by definition. When these two conditions fail for every γ, then V_0 and V_1 of (5.51) and (5.52) are the same as in Chapter 3. Thus

$$\tilde{\rho}_n = \tilde{\psi}_n$$

in both formula and meaning. The observers have satisfied their posterior expected loss requirements a.s. for any sample size. Now, $\tilde{\psi}_n$ is the probability that all observers γ in Γ and the induced extreme observers of Γ choose the correct hypothesis.

<u>High posterior probability for observers</u>

A special case of the goal $\psi_n \geq \epsilon$ in (5.16) asks that

$$\text{(5.53)} \qquad \pi_\gamma(H_i|\underset{\sim}{x}_n) \geq \delta\,, \qquad \text{for } i = 0,1 \quad \text{and all } \gamma \in \Gamma,$$

with some $\delta > 0$, usually $\delta \geq 0.5$. That is, the observers believe the decision (the correct one) with a high probability. One condition of the sets E_j in (5.50) was that

$$\pi_{0\gamma}(\theta_0|\underset{\sim}{x}_n) > 0.5 \qquad \text{and} \qquad \pi_{1\gamma}(\theta_0|\underset{\sim}{x}_n) < 0.5\,.$$

Through (5.42) and (5.45), these two inequalities can be written

$$\pi_\gamma(\theta_0|\underset{\sim}{x}_n) \;>\; \frac{L_{\gamma01} - L_{\gamma00}}{L_{\gamma01} - R_\gamma}$$

and

$$\pi_\gamma(\theta_0|\underset{\sim}{x}_n) \;<\; \frac{L_{\gamma10} - L_{\gamma11}}{R_\gamma - L_{\gamma11}}$$

when conditions $C_\gamma 1$ and $C_\gamma 2$ hold. Consequently, the goal $\psi_n \geq \epsilon$ then includes (5.53) (plus the choice of a correct decision) when

$$\frac{L_{\gamma01} - L_{\gamma00}}{L_{\gamma01} - R_\gamma} \;=\; 1 - \frac{L_{\gamma10} - L_{\gamma11}}{R_\gamma - L_{\gamma11}} \;=\; \delta \qquad \text{for all } \gamma \in \Gamma.$$

That is, when

$$R_\gamma \;=\; L_{\gamma01} - \frac{L_{\gamma01} - L_{\gamma00}}{\delta} \;=\; \frac{L_{\gamma10} - \delta L_{\gamma11}}{1 + \delta} \qquad \text{for all } \gamma \in \Gamma$$

while conditions $C_\gamma 1$ and $C_\gamma 2$ hold.

6. ESTIMATION

When a statistical inference is to be an estimate, the observers $\gamma \in \Gamma$ cannot agree about the exact value of the parameter θ. An alternative goal does result in what can be called "correct agreement." The notation in this chapter will be largely the same as that in Chapter 2. An audience Γ of observers γ each have priors $\pi_\gamma(\theta)$ and loss functions $L_\gamma(a,\theta)$. Of particular interest will be the observers' actions "a" which estimate θ by using the posterior expected value $\int \theta\, \pi_\gamma(\theta | \underset{\sim}{x}_n)\, d\mu(\theta)$; eg, when the observers have the squared error loss function. As before, an experimenter "$*$" with his own prior $\pi_*(\theta)$ will present to the audience the results $\underset{\sim}{x}_n$ of an experiment from the likelihood function $f(x|\theta)$.

So that the observers do not prevent themselves from agreeing, assume that their priors are mutually absolutely continuous:

$$\pi_\gamma \ll \pi_{\gamma'} \qquad \text{for all } \gamma, \gamma' \in \Gamma\,.$$

Denote by $\mathrm{E}_\zeta(\cdot)$ an expectation with respect to the constant ζ or the distribution ζ; eg,

$$\mathrm{E}_{\pi_*,f,n}\big(L(a,\theta)\big) = \int_\Theta \int_{\mathcal{X}} L\big(a(\underset{\sim}{x}_n),\theta\big) f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\, d\lambda(\underset{\sim}{x}_n)\,.$$

Denote by $m_\zeta(\underset{\sim}{x}_n)$ the marginal distribution with parameters or distributions ζ. For example,

$$m_*(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\,.$$

The observers will be considered to *agree* when their estimates $\hat\theta_\gamma$ are all within some prespecified $\delta > 0$ of each other: $\left|\hat\theta_\gamma - \hat\theta_{\gamma'}\right| < \delta$ for all $\gamma, \gamma' \in \Gamma$. That is, when the observers' confidence intervals, of width $\delta/2$ on either side of their posterior means $\hat{\theta}_\gamma$, intersect. Consider that the observers *correctly agree* when their estimates $\hat{\theta}_\gamma$ are all within δ of the parameter value θ; specifically, when the experimental results $\underset{\sim}{x}_n$ are in the set

$$A_\theta = \left\{ \underset{\sim}{x}_n : \left|\hat{\theta}_\gamma - \theta\right| < \delta \quad \text{for all } \gamma \in \Gamma \right\} . \tag{6.1}$$

The experimenter chooses the sample size n for ϵ-correct agreement, using the above $\delta > 0$ and some prespecified ϵ, $0 < \epsilon < 1$, so that the probability of correct agreement

$$\rho_n = \int_\Theta \int_{A_\theta} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \; > \; \epsilon . \tag{6.2}$$

Tchebyshev's inequality implies that

$$\underset{\pi_*,f}{P} \left(\left|\hat{\theta}_\gamma - \theta\right| > \delta\right) \leq \frac{E_{\pi_*,f}[(\hat{\theta}_\gamma - \theta)^2]}{\delta^2}$$

for one observer. Although

$$1 - \rho_n = \underset{\pi_*,f}{P} \left(\sup_{\gamma \in \Gamma} |\hat{\theta}_\gamma - \theta| > \delta \right) ,$$

the last term is not necessarily bounded by

$$D_n = \sup_{\gamma \in \Gamma} \frac{E_{\pi_*,f}[(\hat{\theta}_\gamma - \theta)^2]}{\delta^2} . \tag{6.3}$$

Still, D_n in (6.3) is often more tractable than ρ_n in (6.2). Jackson, Novick and DeKeyrel (1980) considered quantities like D_n but with $\hat{\theta}_\gamma$ replaced by the mean of the density

$$\int_{\mathcal{X}} \pi_\gamma(\theta|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) .$$

They labeled the mean "$\mu_{1\cdot 2}$" for this density "$b_{1\cdot 2}(\theta)$" when Γ has but one member. We show in the following Theorem 6.1 conditions which lead D_n to converge to 0 for large n.

THEOREM 6.1. *Assume*

(i) $\int_\Theta |\theta|^3 \pi_*(\theta)\, d\mu(\theta) < \infty$,

(ii) $\sup_{n\geq 0} \sup_{\gamma\in\Gamma} \int_{\mathcal{X}} \int_\Theta |\hat{\theta}_\gamma|^3 f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\, d\lambda(\underset{\sim}{x}_n) < \infty$

(elaborated upon in Corollary 6.2), and

(iii) $\hat{\theta}_\gamma$ *is a consistent estimator of* θ.

Then $D_n \overset{n\to+\infty}{\longrightarrow} 0$.

PROOF OF THEOREM 6.1.

$$\begin{aligned}
&\sup_{n\geq 0} \sup_{\gamma\in\Gamma} \int_{\mathcal{X}} \int_\Theta |\theta - \hat{\theta}_\gamma|^3 f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\, d\lambda(\underset{\sim}{x}_n) \\
\leq\ & 8\int_\Theta |\theta|^3\pi_*(\theta)\, d\mu(\theta) + 8 \sup_{n\geq 0} \sup_{\gamma\in\Gamma} \int_{\mathcal{X}} \int_\Theta |\hat{\theta}_\gamma|^3 f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\, d\lambda(\underset{\sim}{x}_n) \\
<\ & \infty
\end{aligned}$$

by conditions (i) and (ii). Consequently, $(\theta - \hat{\theta}_\gamma)^2$ is uniformly integrable for γ and n. Uniform integrability and condition (iii) imply [essentially, Chow and Teicher (1978, page 98)] that

$$D_n = \sup_{\gamma\in\Gamma} \int_{\mathcal{X}} \int_\Theta (\theta - \hat{\theta}_\gamma)^2 f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\, d\mu(\theta)\, d\lambda(\underset{\sim}{x}_n) \overset{n\to+\infty}{\longrightarrow} 0\,. \qquad \square$$

Schwartz (1965) gives several conditions under which Bayes estimators are consistent. In particular, under weak conditions (see Lemma 5.1 in this thesis), the posterior mean estimator

$$\hat{\theta}_\gamma = \int_\Theta \theta\, \pi_\gamma(\theta|\underset{\sim}{x}_n)\, d\mu(\theta)$$

is consistent.

The following Corollary 6.2 gives criteria under which condition (ii) of Theorem 6.1 holds. For this corollary, make the usual definition that a probability distribution π_2 is *stochastically larger* than (or equal to) π_1, written $\pi_1 \leq^{st} \pi_2$, when

$$P_{\pi_1}\left(\theta > t\right) \leq P_{\pi_2}\left(\theta > t\right) \qquad \text{for all } t \in \mathbf{R}^1 .$$

Say that $\underset{\sim}{x}_n$ is no larger than $\underset{\sim}{y}_n$ when

$$\underset{\sim}{x}_n \leq \underset{\sim}{y}_n \qquad \text{which signifies } x_i \leq y_i \text{ for all } i = 1, 2, \ldots n .$$

Also, say that π_2 forms a *non-decreasing monotone likelihood ratio* (see page 33) with π_1 when $\pi_1 \prec \pi_2$. Similarly, say that $m_2(\underset{\sim}{x}_n)$ forms a nondecreasing monotone likelihood ratio with $m_1(\underset{\sim}{x}_n)$ when $m_1(\underset{\sim}{x}_n) \prec m_2(\underset{\sim}{x}_n)$; ie, $\frac{m_2(\underset{\sim}{x}_n)}{m_1(\underset{\sim}{x}_n)}$ is nondecreasing in $\underset{\sim}{x}_n$. The following facts will be used without elaboration.

FACT 6.1. *If $\pi_1 \leq^{st} \pi_2$ and $\phi(\theta) \geq 0$ is a non-decreasing function, then*

$$\int_\Theta \phi(\theta)\pi_1(\theta)\, d\mu(\theta) < \int_\Theta \phi(\theta)\pi_2(\theta)\, d\mu(\theta) .$$

Similarly, if $m_1(\underset{\sim}{x}_n) \prec m_2(\underset{\sim}{x}_n)$ and $\phi(\underset{\sim}{x}_n) \geq 0$ is a nondecreasing function, then

$$\int_{\mathcal{X}^n} \phi(\underset{\sim}{x}_n) m_1(\underset{\sim}{x}_n) d\lambda(\underset{\sim}{x}_n) < \int_{\mathcal{X}^n} \phi(\underset{\sim}{x}_n) m_2(\underset{\sim}{x}_n) d\lambda(\underset{\sim}{x}_n) .$$

FACT 6.2. *If $f(x|\theta)$ forms a non-decreasing monotone likelihood ratio family in x and θ, and $\pi_1 \leq^{st} \pi_2$, then*

$$m_1(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_1(\theta)\, d\mu(\theta) \prec \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_2(\theta)\, d\mu(\theta) = m_2(\underset{\sim}{x}_n) .$$

FACT 6.3. *If $\pi_1 \prec \pi_2$, then $\pi(\theta|\underset{\sim}{x}_n) \prec \pi_2(\theta|\underset{\sim}{x}_n)$ for any $\underset{\sim}{x}_n \in \mathcal{X}^n$.*

FACT 6.4. *If $f(x|\theta)$ forms a non-decreasing monotone likelihood ratio family in x and θ, and $\phi(\theta) \geq 0$ is a non-decreasing function, then (for any prior π)*

$$\int_\Theta \phi(\theta)\pi(\theta|\underset{\sim}{x}_n)\, d\mu(\theta)$$

is non-decreasing in $\underset{\sim}{x}_n$.

Similar relations hold when $\prec$ is replaced with $\succ$, $\overset{\text{st}}{\leq}$ is replaced with $\overset{\text{st}}{\geq}$, or "non-decreasing" is replaced with "non-increasing."

COROLLARY 6.2. *Assume that*

(iv) *there exist densities $\pi_{\gamma\uparrow}$, relative to $\mu(\theta)$ for $\gamma \in \Gamma$, for which*

(a) *$\pi_{\gamma\uparrow}$ forms a non-decreasing monotone likelihood ratio with π_γ: $\pi_\gamma \prec \pi_{\gamma\uparrow}$,*

(b) *$\pi_{\gamma\uparrow}$ is stochastically larger than (or equal to) π_*: $\pi_* \overset{\text{st}}{\leq} \pi_{\gamma\uparrow}$,*

that

(v) *there exist densities $\pi_{\gamma\downarrow}$, relative to $\mu(\theta)$ for $\gamma \in \Gamma$, for which*

(a) *$\pi_{\gamma\downarrow}$ forms a non-increasing monotone likelihood ratio with π_γ: $\pi_\gamma \succ \pi_{\gamma\downarrow}$,*

(b) *$\pi_{\gamma\downarrow}$ is stochastically smaller than (or equal to) π_*: $\pi_{\gamma\downarrow} \overset{\text{st}}{\leq} \pi_*$,*

that

(vi) $\sup_{\gamma\in\Gamma} \int_\Theta |\theta|^3 \pi_{\gamma\uparrow}\, d\mu(\theta) < \infty$ *and* $\sup_{\gamma\in\Gamma} \int_\Theta |\theta|^3 \pi_{\gamma\downarrow}\, d\mu(\theta) < \infty$,

and that

(vii) *$f(x|\theta)$ forms either a non-increasing or a non-decreasing monotone likelihood ratio family.*

Then, when the observers' estimates of θ are the posterior means

$$\hat{\theta}_\gamma = \int_\Theta \theta\, \pi_\gamma(\theta|\underset{\sim}{x}_n)\, d\mu(\theta) \qquad \textit{for } \gamma \in \Gamma,$$

condition (ii) of Theorem 6.1 holds:

$$\sup_{n\geq 0} \sup_{\gamma\in\Gamma} \int_{\mathcal{X}} \int_\Theta |\hat{\theta}_\gamma|^3 f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\, d\lambda(\underset{\sim}{x}_n) < \infty .$$

Condition (iv) is met for an observer γ in Γ when there is some $M > 0$ for which the following forms a density:

$$\pi_{\gamma\uparrow}(\theta) = \begin{cases} 0 & \text{for } \theta < M \\ K_{\gamma\uparrow}\left[\sup_{M\leq\theta'\leq\theta} \frac{\pi_*}{\pi_\gamma}(\theta')\right]\pi_\gamma(\theta) & \text{for } \theta \geq M\,, \end{cases}$$

where $K_{\gamma\uparrow}$ is a normalizing constant.

Condition (v) is met for an observer γ in Γ when there is some $M > 0$ for which the following forms a density:

$$\pi_{\gamma\downarrow}(\theta) = \begin{cases} 0 & \text{for } \theta > -M \\ K_{\gamma\downarrow}\left[\inf_{\theta\leq\theta'\leq -M} \frac{\pi_*}{\pi_\gamma}(\theta')\right]\pi_\gamma(\theta) & \text{for } \theta \leq -M\,, \end{cases}$$

where $K_{\gamma\downarrow}$ is a normalizing constant.

In many problems, a $\pi_\gamma(\theta)$ and π_* form monotone likelihood ratios in the tails of Θ. Condition (iv) holds for an observer γ when either $\frac{\pi_\gamma}{\pi_*}(\theta)$ increases, or exclusively $\frac{\pi_*}{\pi_\gamma}(\theta)$ increases, for all $\theta > M$ and some $M > 0$. Just let

$$\pi_{\gamma\uparrow}(\theta) = \begin{cases} 0 & \text{when } \theta < M \\ K_{\gamma\uparrow}\pi_\gamma(\theta) & \text{when } \theta \geq M\,, \end{cases} \qquad \text{for some normalizing constant } K_{\gamma\uparrow}$$

if $\frac{\pi_\gamma}{\pi_*}(\theta)$ increases, or

$$\pi_{\gamma\uparrow}(\theta) = \begin{cases} 0 & \text{when } \theta < M \\ K_{\gamma\uparrow}\pi_*(\theta) & \text{when } \theta \geq M\,, \end{cases} \qquad \text{for some normalizing constant } K_{\gamma\uparrow}$$

if $\frac{\pi_*}{\pi_\gamma}(\theta)$ increases. Similarly, condition (v) holds when either $\frac{\pi_\gamma}{\pi_*}(\theta)$ increases, or exclusively $\frac{\pi_*}{\pi_\gamma}(\theta)$ increases, for all $\theta < -M$ and some $M > 0$. When (iv) and (v) are both satisfied this way for all $\gamma \in \Gamma$, condition (vi) simplifies to

$$\sup_{\gamma\in\Gamma}\int_\Theta |\theta|^3\pi_\gamma(\theta)\, d\mu(\theta) < \infty \qquad \text{and} \qquad \int_\Theta |\theta|^3\pi_*(\theta)\, d\mu(\theta) < \infty\,.$$

PROOF OF COROLLARY 6.2.

For this proof consider that $f(x|\theta)$ forms a non-decreasing monotone likelihood ratio family: the non-increasing case having a parallel proof. Since $\hat{\theta}_\gamma$ is the posterior mean,

$$\begin{aligned}
\mathrm{E}\left(|\hat{\theta}_\gamma|^3\right) &= \int_{\mathcal{X}}\int_{\Theta} |\hat{\theta}_\gamma|^3 f(\underline{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\, d\lambda(\underline{x}_n) \\
&\leq \mathrm{E}\left(\int_{\Theta} |\eta|^3 \pi_\gamma(\eta|\underline{x}_n)\, d\mu(\eta)\right) \\
&= \mathrm{E}\left(\int_0^{+\infty} |\eta|^3 \pi_\gamma(\eta|\underline{x}_n)\, d\mu(\eta)\right) + \mathrm{E}\left(\int_{-\infty}^{0} |\eta|^3 \pi_\gamma(\eta|\underline{x}_n)\, d\mu(\eta)\right) \\
&\leq \mathrm{E}\left(\int_0^{+\infty} |\eta|^3 \pi_{\gamma\uparrow}(\eta|\underline{x}_n)\, d\mu(\eta)\right) + \mathrm{E}\left(\int_{-\infty}^{0} |\eta|^3 \pi_{\gamma\downarrow}(\eta|\underline{x}_n)\, d\mu(\eta)\right), \\
&\qquad \text{using Fact 6.3 and Fact 6.1; iv(a), v(a) and vii} \\
&\leq \int_{\mathcal{X}^n} \left[\int_0^{+\infty} |\eta|^3 \pi_{\gamma\uparrow}(\eta|\underline{x}_n)\, d\mu(\eta)\right] m_{\gamma\uparrow}(\underline{x}_n)\, d\lambda(\underline{x}_n) \\
&\qquad + \int_{\mathcal{X}^n} \left[\int_{-\infty}^{0} |\eta|^3 \pi_{\gamma\downarrow}(\eta|\underline{x}_n)\, d\mu(\eta)\right] m_{\gamma\downarrow}(\underline{x}_n)\, d\lambda(\underline{x}_n), \\
&\qquad \text{using Fact 6.2, Fact 6.4 and Fact 6.1; iv(b), v(b) and vii} \\
&= \int_0^{+\infty} |\eta|^3 \pi_{\gamma\uparrow}(\eta)\, d\mu(\eta) + \int_{-\infty}^{0} |\eta|^3 \pi_{\gamma\downarrow}(\eta)\, d\mu(\eta).
\end{aligned}$$

Taking the supremum over n and γ in the above inequality, condition (vi) implies that

$$\sup_{n\geq 0}\sup_{\gamma\in\Gamma} \mathrm{E}|\hat{\theta}_\gamma|^3 \leq \sup_{\gamma\in\Gamma}\int_{\Theta} |\theta|^3 \pi_{\gamma\uparrow}\, d\mu(\theta) + \sup_{\gamma\in\Gamma}\int_{\Theta} |\theta|^3 \pi_{\gamma\downarrow}\, d\mu(\theta) < \infty. \quad \square$$

6.1 Gaussian estimation example.

Consider that each datum X_i, $i = 1, 2, \ldots n$, comes from a Gaussian density with mean θ and variance σ^2: $X_i \sim \mathcal{N}(\theta, \sigma^2)$. The observers γ in Γ are presumed to have the conjugate prior densities $\pi_\gamma(\theta) \sim \mathcal{N}(\mu_\gamma, \tau_\gamma^2)$. Similarly, the experimenter's prior is presumed to be $\pi_*(\theta) \sim \mathcal{N}(\mu_*, \tau_*^2)$. The observers' posteriors have the consequent posterior densities

$$\pi_\gamma(\theta|\overline{x}) \sim \mathcal{N}\left(\frac{\sigma^2}{\sigma^2+n\tau_\gamma^2}\mu_\gamma + \frac{n\tau_\gamma^2}{\sigma^2+n\tau_\gamma^2}\overline{x}, \frac{\sigma^2\tau_\gamma^2}{\sigma^2+n\tau_\gamma^2}\right), \qquad \text{for } \gamma \in \Gamma.$$

The subset of $\mathcal{X}^n$, in terms of $\overline{x}$, which would give correct agreement is then

$$\begin{aligned} &A_\theta \\ &= \left\{\overline{x} : \left|\frac{\sigma^2}{\sigma^2+n\tau_\gamma^2}\mu_\gamma + \frac{n\tau_\gamma^2}{\sigma^2+n\tau_\gamma^2}\overline{x} - \theta\right| < \delta \quad \text{for all } \gamma \in \Gamma\right\} \\ &= \begin{cases} \left(\sup_{\gamma\in\Gamma}\left[(\theta-\mu_\gamma-\delta)\frac{\sigma^2}{n\tau_\gamma^2} + (\theta-\delta)\right], \inf_{\gamma\in\Gamma}\left[(\theta-\mu_\gamma+\delta)\frac{\sigma^2}{n\tau_\gamma^2} + (\theta+\delta)\right]\right) \\ \qquad\qquad \text{if the ``sup'' is smaller than the ``inf'' endpoint} \\ \emptyset \qquad\qquad \text{otherwise} \end{cases} \\ (6.4) \quad &= \begin{cases} \left(\sup_{\gamma\in\Gamma}\left[(\theta-\mu_\gamma-\delta)\frac{\sigma}{\sqrt{n\tau_\gamma^2}} - \delta\frac{\sqrt{n}}{\sigma}\right], \inf_{\gamma\in\Gamma}\left[(\theta-\mu_\gamma+\delta)\frac{\sigma}{\sqrt{n\tau_\gamma^2}} + \delta\frac{\sqrt{n}}{\sigma}\right]\right) \\ \qquad\qquad \text{if the ``sup'' is smaller than the ``inf'' endpoint} \\ \emptyset \qquad\qquad \text{otherwise}, \end{cases} \end{aligned}$$

if we change variables by the transformation $y = \frac{x-\theta}{(\sigma/\sqrt{n})}$ (simultaneously changing the likelihood function to the standard normal). Now,

$$\rho_n = \int_{-\infty}^{+\infty}\int_{A_\theta} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\frac{1}{\sqrt{2\pi}\tau_*}e^{-\frac{1}{2\tau_*}(\theta-\mu_*)^2}\,dt\,d\theta\,.$$

While the width of A_θ does not depend on θ, it does depend on n. Also, the γ or sequence of γ determining the "sup" or "inf" endpoints in (6.4) do change with θ. Consequently, Γ cannot be reduced (compactified) to an audience of but two observers.

We may assume that $\sigma = 1$. When $\sigma \neq 1$, transform the parameters to $\tau_\gamma^2 = \frac{\bar{\tau}_\gamma^2}{\bar{\sigma}}$ and $\delta = \frac{\bar{\delta}}{\bar{\sigma}}$ where the actual parameter values are the barred values $\bar{\sigma}$, $\bar{\delta}$ and $\bar{\tau}_\gamma$. Table 6.1 presents some values of ρ_n for various values of the parameters when the audience contains two observers, $\Gamma = \{1, 2\}$.

Table 6.1 ρ_n for estimation, where X is a Gaussian random variable

δ	0.10	.05	.10	.50	.10	.05	.05
μ_1	10	-2	-2	-2	-2	0	0
μ_2	1.7	2	2	2	2	0	0
μ_*	1.7	0	0	0	0	0	0
τ_1	1	1	1	1	1	.1	.1
τ_2	1	1	1	1	1	.1	.1
τ_*	1	1	1	1	10	.00316	.1
n							
1	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	.3847
2	0.0000	0.0000	0.0000	0.0000	0.0000	.9996	.3864
3	0.0000	0.0000	0.0000	0.0000	0.0000	.9966	.3882
4	0.0000	0.0000	0.0000	.1769	0.0000	.9898	.3899
5	0.0000	0.0000	0.0000	.3169	0.0000	.9799	.3916
6	0.0000	0.0000	0.0000	.4292	0.0000	.9681	.3933
7	0.0000	0.0000	0.0000	.5205	0.0000	.9553	.3950
8	0.0000	0.0000	0.0000	.5953	0.0000	.9422	.3967
9	0.0000	0.0000	0.0000	.6572	0.0000	.9292	.3983
10	0.0000	0.0000	0.0000	.7087	0.0000	.9165	.4000
20	0.0000	0.0000	.0174	.9364	.0073	.8192	.4161
30	0.0000	0.0000	.1566	.9847	.0769	.7639	.4314
40	0.0000	.0062	.2571	.9961	.1409	.7310	.4459
50	.0895	.0614	.3358	.9990	.1998	.7107	.4597
100	.4136	.2385	.5797	1.0000	.4332	**.6824**	.5205
200	.7192	.4298	.7983	1.0000	.7040	.7110	.6135
300	.8542	.5481	.8947	1.0000	.8400	.7517	.6827
400	.9212	.6326	.9429	1.0000	.9116	.7886	.7364
500	.9565	.6969	.9684	1.0000	.9504	.8202	.7793
1000	.9974	.8712	.9981	1.0000	.9969	.9180	.9027
5000	1.0000	.9995	1.0000	1.0000	1.0000	.9997	.9996
10,000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100,000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Adcock, C. J. A Bayesian approach to calculating sample sizes for multinomial sampling. *The Statistician*, 36:155–159, 1987.

[2] Bahadur, R. R. *Some Limit Theorems in Statistics.* SIAM, Philadelphia, Pennsylvania, 1971.

[3] Bandemer, Hans, Wolfgang Näther, and Jürgen Pilz. Once more: optimal experimental design for regression models. *Statistics*, 18(2):171–217, 1987.

[4] Berger, James. The robust Bayesian viewpoint. In Joseph B. Kadane, editor, *Robustness of Bayesian Analyses*, pages 63–124, Elsevier Science Publishers B.V., Amsterdam, 1984.

[5] Brown, Lawrence D. *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory.* Institute of Mathematical Statistics, Hayward, California, 1986.

[6] Chaloner, Kathryn. Optimal Bayesian experimental design for linear models. *The Annals of Statistics*, 12(1):283–300, 1984.

[7] Chernoff, Herman. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493–507, 1952.

[8] Chow, Yuan Shih and Henry Teicher. *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, New York, 1978.

[9] Chung, Kai Lai. *A Course in Probability Theory.* Academic Press, New York, second edition, 1974.

[10] Cramer, Harald. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, 1946.

[11] Dewey, John. *Logic: The Theory of Inquiry.* Henry Holt and Co., New York, 1938.

[12] Dickey, James and Peter Freeman. Population-distributed personal probabilities. *Journal of the American Statistical Association*, 70(350):362–364, June 1975.

[13] French, Simon. Group consensus probability distributions: a critical survey. In J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, editors, *Bayesian Statistics 2*, pages 183–202, Elsevier Science Publishers B.V., Amsterdam, 1985.

[14] Genest, Christian and Mark J. Schervish. Modeling expert judgments for Bayesian updating. *The Annals of Statistics*, 13(3):1198–1212, 1985.

[15] Genest, Christian and James V. Zidek. Combining probability distributions: a critique and an annotated bibliography. *Statistical Science*, 1(1):114–148, 1986.

[16] Hylland, Aanund and Richard Zeckhauser. The impossibility of Bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47(6):1321–1336, November 1979.

[17] Jackson, Paul H., Melvin R. Novick, and Dennis F. DeKeyrel. Adversary preposterior analysis for simple parametric models. In Arnold Zellner and Harold Jeffreys, editors, *Bayesian Analysis in Econometrics and Statistics*, chapter 10, pages 113–132, North-Holland, Amsterdam, 1980.

[18] Lehrer, Keith and Carl Wagner. *Rational Consensus in Science and Society: A Philosophical and Mathematical Study.* Dordrecht, Holland, Boston, 1981.

[19] Marshall, Albert W. and Ingram Olkin. *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York, 1979.

[20] Murphee, Idus. The theme of positivism in Peirce's pragmatism. In Edward C. Moore and Richard S. Robin, editors, *Studies in the Philosophy of Charles Sanders Peirce, Second Series*, pages 226–241, University of Massachussetts Press, Amherst, 1964.

[21] Owen, Roger J. Accuracy, confidence and consensus in Bayesian hypothesis inference. *Communications in Statistics—Theory and Methods*, 14(5):1035–1052, 1985.

[22] Peirce, Charles Sanders. How to make our ideas clear. *Popular Science Monthly*, January 1878.

[23] Raiffa, Howard. *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Addison-Wesley, Reading, Massachussetts, 1968.

[24] Savage, L. J. *The Foundations of Statistics.* Wiley, New York, 1954.

[25] Schwartz, Lorrain. On Bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4(1):10–26, 1965.

[26] Weerahandi, S. and J.V. Zidek. Multi-Bayesian statistical decision theory. *Journal of the Royal Statistical Society, series A*, 144(part 1):85–93, 1981.

APPENDICES

Appendix A: The zero-one loss provides a canonical form for two-action decision problems.

Suppose that each observer γ has a loss function $L_\gamma(a, \theta)$ defined over the action space $\mathcal{A} = \{a_0, a_1\}$ and the parameter space Θ. Recall that a_i is the action "decide that H_i is true", $i = 0, 1$. It is reasonable to assume that for each observer the loss for a correct action is no greater than the loss for an incorrect action. Thus, we make the following assumption.

ASSUMPTION A.1. *For all $\gamma \in \Gamma$,*

$$L_\gamma(a_1, \theta) - L_\gamma(a_0, \theta) \geq 0\,, \quad \textit{if } \theta \in \Theta_0,$$

$$L_\gamma(a_0, \theta) - L_\gamma(a_1, \theta) \geq 0\,, \quad \textit{if } \theta \in \Theta_1.$$

Since the observers are Bayesians, each observer chooses the action that minimizes his or her posterior expected loss given the data $\underset{\sim}{x}_n$. Let

$$\text{(A.1)} \quad l_\gamma(a|\underset{\sim}{x}_n) = \int_\Theta L_\gamma(a, \theta) m_\gamma^{-1}(\underset{\sim}{x}_n) f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\, d\mu(\theta)\,, \qquad \text{for } a = a_0, a_1,$$

where

$$m_\gamma(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\, d\mu(\theta)\,,$$

as in (2.2). Then, observer γ will choose action a_0 without randomization if and only if

$$\text{(A.2)} \qquad l_\gamma(a_0|\underset{\sim}{x}_n) < l_\gamma(a_1|\underset{\sim}{x}_n)\,,$$

and will choose action a_1 without randomization if and only if

$$\text{(A.3)} \qquad l_\gamma(a_0|\underset{\sim}{x}_n) > l_\gamma(a_1|\underset{\sim}{x}_n)\,.$$

If $l_\gamma(a_0|\underset{\sim}{x}_n) = l_\gamma(a_1|\underset{\sim}{x}_n)$, observer γ may randomize between the actions a_0, a_1 with arbitrary probabilities (possibly depending on $\underset{\sim}{x}_n$).

Note that

$$\text{(A.4)} \qquad l_\gamma(a|\underset{\sim}{x}_n) = \sum_{i=0}^{1} \int_{\Theta_i} L_\gamma(a,\theta) m_\gamma^{-1}(\underset{\sim}{x}_n) f(\underset{\sim}{x}_n|\theta)\pi_\gamma(\theta)\, d\mu(\theta)\,.$$

Hence, inequality (A.2) holds if and only if

$$\text{(A.5)} \qquad \int_{\Theta_0} [L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta)] f(\underset{\sim}{x}_n|\theta)\pi(\theta)\, d\mu(\theta) > \int_{\Theta_1} [L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta)] f(\underset{\sim}{x}_n|\theta)\pi(\theta)\, d\mu(\theta)\,,$$

and inequality (A.3) holds if and only if the inequality is reversed ("$>$" is replaced by "$<$" in (A.5)).

Define

$$\text{(A.6)} \qquad \tilde{\pi}_\gamma(\theta) = \begin{cases} c_\gamma \pi_\gamma(\theta)[L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta)]\,, & \text{if } \theta \in \Theta_0 \\ c_\gamma \pi_\gamma(\theta)[L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta)]\,, & \text{if } \theta \in \Theta_1\,, \end{cases}$$

where

$$\text{(A.7)} \qquad c_\gamma^{-1} = \int_{\Theta_0} \pi_\gamma(\theta)[L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta)]\, d\mu(\theta) + \int_{\Theta_1} \pi_\gamma(\theta)[L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta)]\, d\mu(\theta)\,.$$

For $\tilde{\pi}_\gamma(\theta)$ to be well-defined, we must make the assumption:

ASSUMPTION A.2. *For all* $\gamma \in \Gamma$, $0 < c_\gamma < \infty$.

If Assumption A.2 holds, then it follows from Assumption A.1 and (A.6) that $\tilde{\pi}_\gamma(\theta)$ defines a probability distribution over Θ, all $\gamma \in \Gamma$.

Define

$$\tilde{m}_\gamma(\underset{\sim}{x}_n) = \int_{\Theta} f(\underset{\sim}{x}_n|\theta)\tilde{\pi}_\gamma(\theta)\, d\mu(\theta)\,,$$

and

$$\begin{aligned} \tilde{\pi}_\gamma(H_0|\underset{\sim}{x}_n) &= \left[\tilde{m}_\gamma(\underset{\sim}{x}_n)\right]^{-1} \int_{\Theta_0} f(\underset{\sim}{x}_n|\theta)\tilde{\pi}_\gamma(\theta)\, d\mu(\theta)\,, \\ \tilde{\pi}_\gamma(H_1|\underset{\sim}{x}_n) &= \left[\tilde{m}_\gamma(\underset{\sim}{x}_n)\right]^{-1} \int_{\Theta_1} f(\underset{\sim}{x}_n|\theta)\tilde{\pi}_\gamma(\theta)\, d\mu(\theta) \\ &= 1 - \tilde{\pi}_\gamma(H_0|\underset{\sim}{x}_n)\,. \end{aligned}$$

It is easily seen that inequality (A.5) holds if and only if

$$\tilde{\pi}_\gamma(H_0|\underset{\sim}{x}_n) < \tilde{\pi}_\gamma(H_1|\underset{\sim}{x}_n) = 1 - \tilde{\pi}_\gamma(H_0|\underset{\sim}{x}_n),$$

which in turn holds if and only if

$$\tilde{\pi}_\gamma(H_0|\underset{\sim}{x}_n) > 0.5 .$$

Similarly, the reverse of inequality (A.5) holds if and only if $\tilde{\pi}_\gamma(H_1) > 0.5$. (Finally, the two sides of (A.5) are equal if and only if $\tilde{\pi}_\gamma(H_0) = \tilde{\pi}_\gamma(H_1) = 0.5$.) Thus, we have shown that each observer γ acts as if he or she had a prior distribution $\tilde{\pi}_\gamma(\theta)$ over Θ, and a zero-one loss function (2.1).

We have already indicated why Assumption A.1 is reasonable. Assumption A.2 is also reasonable since if this assumption fails to hold, observer γ will not be influenced by the data in making a decision. To see this, first note that if $c_\gamma = \infty$, then the right side of (A.7) equals 0. Thus, by Assumption A.1,

$$L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta) = 0, \qquad \text{if } \pi_\gamma(\theta) \neq 0 \text{ and } \theta \in \Theta_0 ,$$

$$L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta) = 0, \qquad \text{if } \pi_\gamma(\theta) \neq 0 \text{ and } \theta \in \Theta_1 ,$$

and (see (A.5)) regardless of the data $\underset{\sim}{x}_n$ it will be the case that $l_\gamma(a_0|\underset{\sim}{x}_n) = l_\gamma(a_1|\underset{\sim}{x}_n)$. Consequently, observer γ will always randomize between a_0 and a_1, no matter what sample $\underset{\sim}{x}_n$ is observed.

On the other hand, suppose that $c_\gamma = 0$. In this case, the right side of (A.7) is infinite. Thus, either

$$\text{(A.8)} \qquad \int_{\Theta_0} \pi_\gamma(\theta)[L_\gamma(a_1,\theta) - L_\gamma(a_0,\theta)]\, d\mu(\theta) = \infty$$

or

$$\text{(A.9)} \qquad \int_{\Theta_1} \pi_\gamma(\theta)[L_\gamma(a_0,\theta) - L_\gamma(a_1,\theta)]\, d\mu(\theta) = \infty$$

or both (A.8) and (A.9) hold. If (A.8) holds, but (A.9) does not hold, then it is easily seen that the Bayes risk of action a_1 for observer γ is infinite, $n = 0, 1, \cdots$, so that observer γ will always choose action a_0. Similarly, if (A.9) holds but not (A.8), observer γ always (for all n) chooses action a_1. Finally, if both (A.8) and (A.9) hold, the Bayes risks for both action a_0 and action a_1 are infinite, so that observer γ will always randomize between actions a_0 and a_1. In all these cases, observer γ can ignore any data that is obtained. Thus, we see that Assumption A.2 is necessary to justify the taking of data.

Appendix B: The decision of the experimenter himself in two-action decision problems.

Suppose that the experimenter is interested in his own decision, not just his audience's decisions. Then he would include himself in Γ. Here, we present a lower bound on the resulting ρ_n — as a function of Γ with the experimenter.

We assume — till later in this appendix — that the experimenter has 0−1 loss. For the experimenter, denote the marginal density for a sample $\underset{\sim}{x}_n$ by

$$m_*(\underset{\sim}{x}_n) = \int_\Theta f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\mu(\theta)\,,$$

as in (2.2). For a set $A \subseteq \mathcal{X}^n$,

$$\mathrm{P}_*(A) = \int_A m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \qquad \text{if } n \in \{1, 2, \ldots\}\,.$$

Let

$$\rho_n^* = \rho_n \qquad \text{when } \Gamma = \{\text{experimenter}\}.$$

That is, ρ_n^* is the experimenter's probability that he himself chooses the correct hypothesis. We first present for ρ_n^* a lower bound that follows from well known Bayes risk results.

THEOREM B.1. *For every $n \geq 0$,*

$$\rho_n^* \;\leq\; \rho_{n+1}^*$$

and

$$\rho_n^* \;\geq\; \rho_0^* > 0.5 \qquad \textit{if } \pi_* \neq 0.5$$

($\rho_0^* = 0$ if $\pi_* = 0.5$).

PROOF OF THEOREM B.1.

We can write

$$\rho_{n+1}^* = 1 - \int_{\mathcal{X}} \int_{\mathcal{X}^n} \int_\Theta L[a(\underset{\sim}{x}_{n+1}), \theta]\pi_*(\theta|\underset{\sim}{x}_{n+1})m_*(\underset{\sim}{x}_n)m_*(x_{n+1})d\mu(\theta)d\lambda(\underset{\sim}{x}_n)d\lambda(x_{n+1}),$$

where $a(\underset{\sim}{x}_{n+1})$ is the Bayes action minimizing the interior integral, and the loss L is 0−1 loss. Since $a(\underset{\sim}{x}_{n+1})$ uses $\underset{\sim}{x}_n$ to minimize the interior integral, if we replace it by the non-minimizing $a(\underset{\sim}{x}_n)$ then the interior integral is measurable $\underset{\sim}{x}_n$ — the value ρ_n^* results. Thus, $\rho_n^* \leq \rho_{n+1}^*$. This is just a statement that the Bayes risk decreases in n, as is well known.

For the second statement in this theorem,

$$\rho_0^* \;=\; \pi_* I(\pi_* > 0.5) + (1 - \pi_*) I(\pi_* < 0.5)\,,$$

From this follows

$$\rho_0^* \text{ is } \begin{cases} > 0.5 & \text{if } \pi_* \neq 0.5 \\ = 0 & \text{if } \pi_* = 0.5. \end{cases} \qquad \square$$

In order that the audience Γ has two extreme priors, we assume for the rest of this appendix the assumptions of Section 2.4 — Assumption 2.2 and Assumption 2.3 — and

ASSUMPTION B.1. *For the audience* $\Gamma^{+*} \equiv \Gamma \cup \{\text{experimenter}\}$,

$$\Pi^{+*} \quad \equiv \quad \left\{\pi_*,\, \pi_\gamma, \quad \textit{with } \gamma \in \Gamma\right\}$$

forms a monotone likelihood likelihood ratio family.

With Assumption 2.3, this implies that

$$\pi_* \prec \delta_1 \prec \delta_0\,, \qquad \delta_1 \prec \pi_* \prec \delta_0\,, \text{ or } \qquad \delta_1 \prec \delta_0 \prec \pi_*.$$

The next theorem presents a lower bound on ρ_n^* different from the bound in Theorem B.1. The class Γ, from which $\tilde{\rho}_n$ is determined, need not contain the experimenter.

THEOREM B.2. *For every sample size* $n \geq 0$,

$$\rho_n^* \quad \geq \quad \tilde{\rho}_n - P_*(E)\,,$$

where

$$E = \{\underset{\sim}{x}_n \colon \pi_*(\Theta_0|\underset{\sim}{x}_n) = 0.5\}.$$

When Γ is "closed," this becomes

$$\rho_n^* \geq \rho_n - P_*(E).$$

PROOF OF THEOREM B.2.

From (2.23),

$$\begin{aligned}\tilde{\rho}_n &= \sum_{i=0}^{1} \int_{\Theta_i} \int_{\tilde{A}_i} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \\ &= \sum_{i=0}^{1} \int_{\tilde{A}_i} \pi_*(\Theta_i|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n)\end{aligned}$$

for $n \in \{0, 1, 2, \ldots\}$. From (2.24),

$$\tilde{A}_0 = \{\delta_0(\Theta_0|\underset{\sim}{x}_n) > 0.5\}$$

$$\tilde{A}_1 = \{\delta_1(\Theta_0|\underset{\sim}{x}_n) < 0.5\}.$$

Let

$$\grave{A}_0 = \{\delta_0(\Theta_0|\underset{\sim}{x}_n) < 0.5\}.$$

Since $\delta_1(\Theta_0|\underset{\sim}{x}_n) \geq \delta_0(\Theta_0|\underset{\sim}{x}_n)$, then

$$\tilde{A}_1 \subseteq \grave{A}_0.$$

Consequently,

$$\text{(B.1)} \quad \tilde{\rho}_n \leq \int_{\tilde{A}_0} \pi_*(\Theta|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) + \int_{\grave{A}_0} \pi_*(\Theta_1|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n).$$

case 1. $\boldsymbol{\pi_* \prec \delta_o}$.

We can rewrite (B.1) as

$$\tilde{\rho}_n \leq \rho_n^* + \left[-\int_{B_0} \pi_*(\Theta_0|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) + \int_{B_1} \pi_*(\Theta_1|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \right],$$

where

$$B_0 = \left\{ \underset{\sim}{x}_n : \delta_0(\Theta_0|\underset{\sim}{x}_n) \leq 0.5 < \pi_*(\Theta_0|\underset{\sim}{x}_n) \right\}$$

and

$$B_1 = \left\{ \underset{\sim}{x}_n : \delta_0(\Theta_0|\underset{\sim}{x}_n) < 0.5 \leq \pi_*(\Theta_0|\underset{\sim}{x}_n) \right\}.$$

Since $\pi_*(\Theta_0|\underset{\sim}{x}_n) > \pi_*(\Theta_1|\underset{\sim}{x}_n)$ on B_0, then

$$\begin{aligned} \tilde{\rho}_n &\leq \rho_n^* + \Big[-\int_{B_0} \pi_*(\Theta_1|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) + \\ &\qquad \int_{B_0} \pi_*(\Theta_1|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \Big] + \int_E (0.5) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \\ &= \rho_n^* + 0.5 \int_E m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n). \end{aligned}$$

case 2. $\boldsymbol{\pi_* \succ \delta_o}$.

We can rewrite (B.1) as

$$\tilde{\rho}_n \leq \rho_n^* + \left[\int_{B_0} \pi_*(\Theta_0|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) - \int_{B_1} \pi_*(\Theta_1|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \right],$$

where

$$B_0 = \{ \underset{\sim}{x}_n : \delta_0(\Theta_0|\underset{\sim}{x}_n) > 0.5 \geq \pi_*(\Theta_0|\underset{\sim}{x}_n) \}$$

and

$$B_1 = \{ \underset{\sim}{x}_n : \delta_0(\Theta_0|\underset{\sim}{x}_n) \geq 0.5 > \pi_*(\Theta_0|\underset{\sim}{x}_n) \}.$$

Since $\pi_*(\Theta_1|\underset{\sim}{x}_n) > \pi_*(\Theta_0|\underset{\sim}{x}_n)$ on B_1, then

$$\begin{aligned} \tilde{\rho}_n &\leq \rho_n^* + \Big[\int_{B_1} \pi_*(\Theta_0|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) - \\ &\qquad \int_{B_1} \pi_*(\Theta_0|\underset{\sim}{x}_n) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \Big] + \int_E (0.5) m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \\ &= \rho_n^* + 0.5 \int_E m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n). \end{aligned}$$

Combining case 1 and case 2,

$$\begin{aligned} \tilde{\rho}_n &\leq \rho_n^* + \int_E m_*(\underset{\sim}{x}_n)\, d\lambda(\underset{\sim}{x}_n) \\ &= \rho_n^* + P_*(E). \quad \square \end{aligned}$$

Using Theorem B.1 and Theorem B.2, when $n \geq N_\epsilon$ (a forteriori, $\rho_n \geq \epsilon$) and Γ is "closed" then

$$\text{(B.2)} \qquad \rho_n^* \quad \geq \quad \max\Big\{0.5I(\pi_* \neq 0.5), \quad \epsilon - P_*(E)\Big\}.$$

Using that $P(B \cup C) \geq P(B) + P(C) - 1$ for any sets B and C,

$$\text{(B.3)} \qquad \rho_n \quad \geq \quad \max\Big\{\epsilon + 0.5I(\pi_* \neq 0.5) - 1, \quad 2\epsilon - [1 + P_*(E)]\Big\}$$

for the audience Γ^{+*} when $n \geq N_\epsilon$. If $f(x|\theta_0)$ and $f(x|\theta_1)$ are continuous, then for the audience Γ^{+*}, $\rho_n \geq 0.9$ when $n \geq N_{.95}$.

The Experimenter's Bayes Risk for the Audience's Actions

So far, we have assumed that the experimenter has 0−1 loss. This was necessary for both Theorem B.1 and Theorem B.2. The reductions of Appendix A do not extend to the experimenter for his decision. However, those reductions do extend to the experimenter for a goal somewhat different than $\rho_n \geq \epsilon$.

For the remainder of this appendix, let $\ddot{a}$ represent the "fictitious" action

$$\ddot{a} = \begin{cases} a_0 & \text{if } \underset{\sim}{x}_n \in A_0 \\ a_1 & \text{if } \underset{\sim}{x}_n \in A_1 \\ a_1 & \text{if } \underset{\sim}{x}_n \notin A_0, A_1 \text{ and } \theta \in \Theta_0 \\ a_0 & \text{if } \underset{\sim}{x}_n \notin A_0, A_1 \text{ and } \theta \in \Theta_1 . \end{cases}$$

This is the action (possibly the wrong action) of every observer when every observer makes the same decision. It is the wrong action when observers make different decisions. Let

$$L_*(a, \theta)$$

be the experimenter's loss for action "a" when the parameter is θ. There are two perspectives on L_*:

(i) $L_*(a,\theta)$ is the experimenter's loss of his personal decision, a_i if $\pi_*(H_i|\underset{\sim}{x}_n) > 0.5$ — viewed from the inside as an audience member.

(ii) $L_*(\ddot{a},\theta)$ is the experimenter's loss for "the" decision $\ddot{a}$ of the audience — viewed from the outside as a judge of the observers' decisions. While each observer naturally makes his own decision, often a decision in fact must be the same for all observers. For example, when a board must decide whether to build a factory. So, the action $\ddot{a}$ could be forced upon the experimenter, either as an observer himself, or while unrepresented in the audience.

Both of these perspectives (i) and (ii) are in use when the experimenter is a member of the audience Γ. The second perspective we use to define a new goal.

Consider the problem for which the experimenter's goal is not that $\rho_n \geq \epsilon$ but instead that the experimenter's Bayes risk for the audience's actions

$$\text{(B.4)} \qquad \Lambda_n \equiv \int_\Theta \int_{\mathcal{X}^n} L_*(\ddot{a},\theta) f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \quad \leq \quad \zeta$$

for some $\zeta \in \mathbf{R}^1$. We now assume the assumptions of the last appendix, ie, Assumption A.1 and Assumption A.2, for the observers $\gamma \in \Gamma$ and for the experimenter (ie, $\gamma = *$, while not implying that the experimenter is in the audience).

As we wrote ρ_n in (2.4), we can write

$$\begin{aligned}
\Lambda_n &= \Big\{ \int_{\Theta_0} \Big[\int_{A_0} L_*(a_0,\theta) + \int_{\bar{A}_0} L_*(a_1,\theta) \Big] + \\
&\qquad \int_{\Theta_1} \Big[\int_{A_1} L_*(a_1,\theta) + \int_{\bar{A}_1} L_*(a_0,\theta) \Big] \Big\} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) \\
&= \Big\{ \int_{\Theta_0} \int_{A_0} [L_*(a_0,\theta) - L_*(a_1,\theta)] + \\
&\qquad \int_{\Theta_1} \int_{A_1} [L_*(a_1,\theta) - L_*(a_0,\theta)] \Big\} f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta) + W\,,
\end{aligned}$$

where

$$W = \Big[\int_{\Theta_0} \int_{\mathcal{X}^n} L_*(a_1,\theta) + \int_{\Theta_1} \int_{\mathcal{X}^n} L_*(a_0,\theta) \Big] f(\underset{\sim}{x}_n|\theta)\pi_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta)$$

$$= \left[\int_{\Theta_0} L_*(a_1,\theta) + \int_{\Theta_1} L_*(a_0,\theta)\right] \pi_*(\theta)\, d\mu(\theta)\,.$$

With $\tilde{\pi}_*(\theta)$ defined in (A.6) and c_* defined in (A.7),

$$\Lambda_n = -c_*^{-1}\tilde{\rho}_n + W\,,$$

where

$$\tilde{\rho}_n = \sum_{i=0}^{1} \int_{\Theta_i} \int_{A_i} f(\underset{\sim}{x}_n|\theta)\tilde{\pi}_*(\theta)\, d\lambda(\underset{\sim}{x}_n)\, d\mu(\theta)\,.$$

So, the experimenter's goal (B.4) can be written

$$\tilde{\rho}_n \geq c^*(W-\zeta)\,. \tag{B.5}$$

The problem (B.4) has thus been reduced to 0−1 loss.

The goal (B.4) is feasible — the right term of (B.5) is less than 1 — when the a priori expected loss of the correct decision is less than ζ:

$$\left[\int_{\Theta_0} L_*(a_0,\theta) + \int_{\Theta_1} L_*(a_1,\theta)\right] \pi_*(\theta)\, d\mu(\theta) \quad < \quad \zeta\,.$$

The results of Theorem B.1 and Theorem B.2 apply to the experimenter's modified goal (B.4) through the reduction to $\tilde{\rho}_n$ in place of ρ_n for those theorems. This requires the assumptions of Section 2.4, the last appendix, and this appendix. (Assumption B.1 requires that Π^{+*} modified for the 0−1 loss reduction to $\{\tilde{\pi}_*, \tilde{\pi}_\gamma,\ \gamma \in \Gamma\}$ be a monotone likelihood ratio family.)

In summary, this appendix has found bounds on ρ_n^* in Theorem B.2. These bounds implied the bounds on ρ_n for $\gamma \in \Gamma^{+*}$ in (B.3). The goal that the experimenter's Bayes risk for the audience's actions be small, (B.4), reduced to (B.5) with 0−1 loss for all observers and the experimenter. This goal allowed the bounds on ρ_n^* and ρ_n with $\gamma \in \Gamma^{+*}$ to be applicable after reduction to 0−1 loss.

Appendix C: A computer program for one-sided hypothesis testing.

```
      program COMPOUN(input, output, tape5=input, tape6=output)
c     This program computes the probability that all parties choose the same,
c     correct, hypothesis for a list of sample sizes to be read in.
c     The hypotheses are composite hypotheses.
c     This program will accept any continuous prior distributions, not just
c     conjugate priors.  This program is not written for efficiency,
c     but for its possible readability.  Capital letters are used only for
c     readability -- most fortran compilers convert all small letters
c     to capital letters.  Four subroutines must be supplied by the user:
c     1. CONCAVE, the scaling function d(theta) for the canonical form
c     of the exponential family sample distribution.
c     2. LIKCUMU, the cumulative distribution of the sample mean.
c     3. LNPRIOR, the logarithm of the prior densities.
c     4. PRICUMU, the cumulative distribution of the prior distribution.
c          The parameters for 3 and 4 are supplied at the beginning of this
c     program, being sent to subroutines by common statements as needed.
c     Parameters  must be supplied for a first observer, a second observer,
c     and an experimenter.  As supplied below, these four subroutines solve
c     the gamma distribution example with conjugate priors.
c          Wherever the user might make changes is shown by a box of
c     "SSSSSSSSSS."  Except that the parameters themselves --- here, alpha,
c     delta, and zeta --- must be changed for different exponential family
c     distributions.
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
common common common common common common common common common common common   I
      common b,alpha,delta,zeta                                                  I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                     I
      real b,alpha,delta,zeta                                                    I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                       I
common common common common common common common common common common common   I
      real H,PRBJNT1,PRBJNT2                                                     I
      real delta1,delta2,zeta1,zeta2,deltast,zetast                              I
      real prior1L,prior1H,prior2L,prior2H,pristL,pristH                         I
      real pcntile,AA,BB,Z1,Z2,integr1,integr2,rho                               I
      integer itmax,ier                                                          I
      EXTERNAL H,PRBJNT1,PRBJNT2                                                 I
c                                                                                I
c     The following parameter values must be changed.                           I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS             I
      b       = -1.0                                                       S     I
      alpha   = 1.0                                                        S     I
      delta1  = 1.0                                                        S     I
      delta2  = 1.0                                                        S     I
      zeta1   = 50.0                                                       S     I
      zeta2   = 13.0                                                       S     I
      deltast = 3.0                                                        S     I
      zetast  = 2.0                                                        S     I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS             I
```

```
c                                                                              I
c     The following are the parameters of the first observer,                 I
c     used thru the common statement.                                          I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      delta = delta1                                                   S       I
      zeta  = zeta1                                                    S       I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      pcntile = 0.0001                                                         I
      CALL SUPPORT(pcntile,prior1L)                                            I
      pcntile = 0.9999                                                         I
      CALL SUPPORT(pcntile,prior1H)                                            I
c                                                                              I
c     The following are the parameters of the second observer,                 I
c     used thru the common statement.                                          I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      delta = delta2                                                   S       I
      zeta  = zeta2                                                    S       I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      pcntile = 0.0001                                                         I
      CALL SUPPORT(pcntile,prior2L)                                            I
      pcntile = 0.9999                                                         I
      CALL SUPPORT(pcntile,prior2H)                                            I
c                                                                              I
c     The following are the parameters of the experimenter,                   I
c     used thru the common statement.                                          I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      delta = deltast                                                  S       I
      zeta  = zetast                                                   S       I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      pcntile = 0.0001                                                         I
      CALL SUPPORT(pcntile,pristL)                                             I
      pcntile = 0.9999                                                         I
      CALL SUPPORT(pcntile,pristH)                                             I
      priorL = AMIN1(prior1L,prior2L,pristL)                                   I
      priorH = AMAX1(prior1H,prior2H,pristH)                                   I
      print*, 'prior1L = ',prior1L,  'prior1H = ',prior1H                      I
      print*, 'prior2L = ',prior2L,  'prior2H = ',prior2H                      I
      print*, 'pristL = ',pristL,  'pristH = ',pristH                          I
      print*, 'priorL = ',priorL,  'priorH = ',priorH                          I
c                                                                              I
    5 READ(5,100,end=30) n                                                     I
  100 FORMAT(F10.0)                                                            I
c     The following assumes that to give the posterior median of b, the        I
c     sample mean is between AA and BB.  Wider limits are more time consuming. I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      delta = delta1                                                   S       I
      zeta = zeta1                                                     S       I
c WARNING: the following are set to very narrow values                 S       I
      AA = 1.0E-1                                                      S       I
      BB = 1.0E2                                                       S       I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
```

```
      itmax = 300                                                          I
c     The zero of the function H is produced by ZBRENT as BB.              I
      CALL ZBRENT(H,0.0,6,AA,BB,itmax,ier)                                 I
      IF (ier .EQ. 130) THEN                                               I
      print*, 'H has the same value at A and at B ---- a larger            I
     c        range must be provided'                                      I
      ENDIF                                                                I
      Z1 = BB                                                              I
c                                                                          I
c                                                                          I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      delta = delta2                                                  S    I
      zeta = zeta2                                                    S    I
c WARNING: the following are set to very narrow values                S    I
      AA = 1.0E-1                                                     S    I
      BB = 1.0E2                                                      S    I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      itmax = 300                                                          I
      CALL ZBRENT(H,0.0,8,AA,BB,itmax,ier)                                 I
      IF (ier .EQ. 130) THEN                                               I
      print*, 'H has the same value at AA and at BB ---- a larger          I
     c        range must be provided'                                      I
      ENDIF                                                                I
      Z2 = BB                                                              I
      ZL = AMIN1(Z1,Z2)                                                    I
      ZH = AMAX1(Z1,Z2)                                                    I
      print*, 'ZL = ',ZL,   'ZH = ',ZH                                     I
c     The above values,ZL and ZH, are the sample means giving b as the median  I
c     of the posterior distributions for two different observers, at sample    I
c     size n.                                                              I
c                                                                          I
c                                                                          I
c     The following assumes a continuous sample density, or                I
c     that the Z's giving median b for the sample distribution             I
c     are not at the mass points of that distribution.                     I
c     The following DCADRE program is an IMSL integration routine.         I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      delta = deltast                                                 S    I
      zeta = zetast                                                   S    I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      integr1 = DCADRE(PRBJNT1,priorL,b,0.0001,0.0001,error,ier)           I
      integr2 = DCADRE(PRBJNT2,b,priorH,0.0001,0.0001,error,ier)           I
      rho = integr1 + integr2                                              I
      write(*,301) 'n = ',n, ' rho = ',rho                                 I
  301 format(A3,F11.1,A7,F10.6)                                            I
      GOTO 5                                                               I
   30 continue                                                             I
      stop                                                                 I
      end                                                                  I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
```

```
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION CONCAVE(theta)                                          I
c     This is the scaling function d(theta) for the exponential             I
c     family sampling distribution.                                         I
common common common common common common common common common common common I
      common b,alpha,delta,zeta                                             I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                I
      real b,alpha,delta,zeta                                               I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                  I
common common common common common common common common common common common I
      real theta                                                            I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS        I
      CONCAVE = alpha * alog(-theta)                                 S      I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS        I
      return                                                                I
      end                                                                   I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION LIKCUMU(xcumu,theta)                                    I
c     This function, corresponding to CONCAVE, gives the cumulative         I
c     distribution function for the sample distribution.                    I
common common common common common common common common common common common I
      common b,alpha,delta,zeta                                             I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                I
      real b,alpha,delta,zeta                                               I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                  I
common common common common common common common common common common common I
      real xcumu,theta                                                      I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS        I
      integer ier1                                                   S      I
      real df1,chicumu                                               S      I
      df1 = 2*n*alpha                                                S      I
      chicumu = -2*n*theta*xcumu                                     S      I
c     Unless 0.5 < df1 < 200,000 then the following cumulative       S      I
c     chi-square distribution program MDCH will give error 129       S      I
c     (error 129 can also occur if chicumu is less than zero).       S      I
      CALL MDCH(chicumu,df1,LIKCUMU,ier1)                            S      I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS        I
      return                                                                I
      end                                                                   I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
```

```
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION LNPRIOR(theta)                                       I
c     This function gives the logarithm of the density of the prior.     I
c     Note that the parameters delta and theta for this distribution are I
c     supplied to every subroutine through the common statements.        I
c     A prior different from the gamma prior would use other parameters, with I
c     their inclusion in the common statements in the place of delta and zeta. I
common common common common common common common common common common common I
      common b,alpha,delta,zeta                                          I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens             I
      real b,alpha,delta,zeta                                            I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens               I
common common common common common common common common common common common I
      real theta                                                         I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      real ALGAMA                                                        I
c     ALGAMA is an IMSL routine for the logarithm of the           S     I
c     gamma function.                                              S     I
      LNPRIOR = zeta*alog(delta) + (zeta-1)*alog(-theta) +         S     I
     c          delta*theta - ALGAMA(zeta)                         S     I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      return                                                             I
      end                                                                I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION PRICUMU(thacumu)                                     I
c     This function, corresponding to LNPRIOR, gives the cumulative      I
c     distribution function for the prior.                              I
c     It is implicitly a function of parameters appearing in the         I
c     common statements.                                                 I
common common common common common common common common common common common I
      common b,alpha,delta,zeta                                          I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens             I
      real b,alpha,delta,zeta                                            I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens               I
common common common common common common common common common common common I
      real thacumu                                                       I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
      integer ier2                                                       I
      real df2,chicumu,cumuchi                                     S     I
      df2 = 2*zeta                                                 S     I
      chicumu = -2 * delta*thacumu                                 S     I
c     Unless 0.5 < df2 < 200,000 then the following cumulative     S     I
c     chi-square distribution program MDCH will give error 129     S     I
c     (error 129 can also occur if chicumu is less than zero).     S     I
      CALL MDCH(chicumu,df2,cumuchi,ier2)                          S     I
      PRICUMU = 1 - cumuchi                                        S     I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS     I
```

```
      return                                                            I
      end                                                               I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION LNDENSJ(theta)                                      I
c     This function is the logarithm of the product of the likelihood   I
c     and the prior density, adjusted by the term factden.              I
common common common common common common common common common common common  I
      common b,alpha,delta,zeta                                         I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens            I
      real b,alpha,delta,zeta                                           I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens              I
common common common common common common common common common common common  I
      real CONCAVE,LNPRIOR                                              I
      real theta                                                        I
      real temp                                                         I
c     "factden" is a scaling factor (term) to prevent LNDENSJ from having too  I
c     small or too large values.  "factden" is determined via "maxdens."  I
      LNDENSJ = n*(theta*w + CONCAVE(theta))  + LNPRIOR(theta) - factden  I
      temp = amax1(maxdens,LNDENSJ)                                     I
      maxdens = temp                                                    I
      return                                                            I
      end                                                               I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION densjnt(theta)                                      I
c     This function is the product of the likelihood and the prior      I
c     density --- adjusted by the factor exp(factden).                  I
common common common common common common common common common common common  I
      common b,alpha,delta,zeta                                         I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens            I
      real b,alpha,delta,zeta                                           I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens              I
common common common common common common common common common common common  I
      real LNDENSJ                                                      I
      real theta                                                        I
      DENSJNT = exp(LNDENSJ(theta))                                     I
      return                                                            I
      end                                                               I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
```

```
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION MPRICUM(thacumu)                                      I
c     This function is needed for the following subroutine, SUPPORT,     I
c     which finds the  zero ---not of pricumu--- but of "pricumu - pcntile."  I
common common common common common common common common common common common  I
      common b,alpha,delta,zeta                                           I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens              I
      real b,alpha,delta,zeta                                             I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                I
common common common common common common common common common common common  I
      real PRICUMU                                                        I
      real thacumu                                                        I
      MPRICUM = PRICUMU(thacumu) - REpctle                                I
      return                                                              I
      end                                                                 I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      SUBROUTINE SUPPORT(pcntile,priored)                                 I
c     This subroutine finds the endpoints of a distribution having        I
c     probability 0.9998 between them.                                    I
common common common common common common common common common common common  I
      common b,alpha,delta,zeta                                           I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens              I
      real b,alpha,delta,zeta                                             I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                I
common common common common common common common common common common common  I
      real MPRICUM,PRICUMU                                                I
      real pcntile,priored                                                I
      real AA,BB                                                          I
      integer itmax,ier                                                   I
      EXTERNAL MPRICUM,PRICUMU                                            I
      REpctle = pcntile                                                   I
c     The following assumes that the parameter theta is between AA and BB.  I
c     The zero of the function Mpricum is output by ZBRENT as BB.         I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS       I
      AA = -1.0E15                                                 S      I
      BB = -1.0E-15                                                S      I
c SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS       I
      itmax = 300                                                         I
      CALL ZBRENT(MPRICUM,0.0,9,AA,BB,itmax,ier)                          I
      IF (ier .EQ. 130) THEN                                              I
      print*, 'MPRICUM has the same value at AA and at BB ---- a larger   I
     c         range must be provided'                                    I
      ENDIF                                                               I
      priored = BB                                                        I
      return                                                              I
      end                                                                 I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

```
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION H(y)                                                  I
common common common common common common common common common common common  I
      common b,alpha,delta,zeta                                           I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens              I
      real b,alpha,delta,zeta                                             I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                I
common common common common common common common common common common common  I
      real DCADRE,LNDENSJ,DENSJNT                                         I
      real y,integr1,integr2,error,temp                                   I
      EXTERNAL LNDENSJ,DENSJNT                                            I
      integer ier                                                         I
      w = y                                                               I
      factden = 0.0                                                       I
      maxdens = -1.0E30                                                   I
c                                                                         I
c     The next DCADRE program is used only to get the scaling factor      I
c     "factden."  DCADRE is an IMSL integration routine.                  I
      temp = DCADRE(LNDENSJ,priorL,priorH,1.0E-3,0.0,error,ier)           I
      factden = maxdens                                                   I
      ier = 0                                                             I
c                                                                         I
      integr1 = DCADRE(DENSJNT,priorL,b,1.0E-5,0.0,error,ier)             I
      IF  (ier .NE. 0) THEN                                               I
          print*, 'error on integral1---estimated absolute error is ',    I
     c            error                                                   I
      ENDIF                                                               I
      ier = 0                                                             I
c                                                                         I
      integr2 = DCADRE(DENSJNT,b,priorH,1.0E-5,0.0,error,ier)             I
      IF  (ier .NE. 0) THEN                                               I
          print*, 'error on integral2---estimated absolute error is ',    I
     c            error                                                   I
      ENDIF                                                               I
c                                                                         I
      H = integr1 / (integr1 +  integr2) - 0.5                            I
      return                                                              I
      end                                                                 I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
```

```
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION PRBJNT1(theta)
c     This function gives the product of the cumulative sample distribution   I
c     and the prior density.                                                  I
common common common common common common common common common common common I
      common b,alpha,delta,zeta                                                I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                   I
      real b,alpha,delta,zeta                                                  I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                     I
common common common common common common common common common common common I
      real LIKCUMU,LNPRIOR                                                     I
      real xcumu,theta                                                         I
      xcumu = ZL                                                               I
      PRBJNT1 = LIKCUMU(xcumu,theta) *                                         I
     c              exp(LNPRIOR(theta))                                        I
      return                                                                   I
      end                                                                      I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
c
c
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
      real FUNCTION PRBJNT2(theta)
c     This function gives the product of "1 - cumulative sample distribution" I
c     and the prior density.                                                  I
common common common common common common common common common common common I
      common b,alpha,delta,zeta                                                I
      common n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                   I
      real b,alpha,delta,zeta                                                  I
      real n,priorL,priorH,ZL,ZH,REpctle,w,factden,maxdens                     I
common common common common common common common common common common common I
      real LIKCUMU,LNPRIOR                                                     I
      real xcumu,theta                                                         I
      xcumu = ZH                                                               I
      PRBJNT2 = (1 - LIKCUMU(xcumu,theta)) *                                   I
     c          exp(LNPRIOR(theta))                                            I
      return                                                                   I
      end                                                                      I
c IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
```

VITA

VITA

Jameson Burt was born on October 11, 1953, in Bethesda, Maryland. He received his Bachelor of Science degree in Mathematics (statistics option) from Montana State University in August of 1975. He worked in Wyoming for the Statistical Reporting Service, USDA, from February of 1976 to June of 1979. In August of 1979 he went to Purdue University, graduating in May of 1989.