AN ENQUIRY INTO THE LOGICAL FOUNDATIONS OF PROFILING

by

Anirban DasGupta
Purdue University

Technical Report #02-05

Department of Statistics
Purdue University
West Lafayette, IN USA

July 2002

*

AN ENQUIRY INTO THE LOGICAL FOUNDATIONS OF PROFILING

Anirban DasGupta

Purdue University, West Lafayette, IN

JUNE 25,2002

ABSTRACT

Profiling of specific groups has become a contentious issue in the aftermath of the tragic events of September 11,2001. We present a formal mathematical enquiry into the pragmatism of profiling. We do so by presenting two mathematical models, and also by a decision theory formulation. The first approach is less formal. We find several potentially controversial results from this mathematical analysis. Three main conclusions are drawn. The seriousness of a particular unlawful intent is a more important factor than the prevalence of the intent. If a certain intent has serious negative consequences, then a significant amount of screening appears to be justified and furthermore there is an optimal amount of screening that is surprisingly stable with respect to the other parameters of the model. And, when the consequences are very serious, and the screening technology is reliable, a complete 100% screening is optimal.

Certain classic mathematical facts on the theory of polynomials are used in deriving some of the results. The results are also complemented by graphs and tables for quick comprehension. Although the analysis is a formalism, it is intended to serve as a guide in this controversial issue in our public policy.

1 Introduction

Profiling of specific groups of individuals has been done by law enforcement agencies for a long time, although its morality and effectiveness have become very much a part of the public debate since the tragic events of September 11, 2001. Profiling used to be a tool principally for enforcement of drug laws, but is now being used by law enforcement for protection against terrorist activities as well. The sentiments for and against the practice of profiling are stubbornly strong. Individuals and groups, typically minorities in some sense, affected by profiling are very understandably against profiling. It obviously causes many innocent individuals not just an inconvenience, but anguish, and loss of dignity. On the other hand, law enforcement feels that profiling is a useful tool , a necessary evil, that they should not be asked to dispense with, for it saves the society from risks of greater evils. Profiling, clearly, has become a contentious issue, perhaps a terrible choice between morality and pragmatism. Morality, by and large, is a question outside the domain of science. But is it possible to enquire into the pragmatism of profiling in a scientific way ? Does mathematics have something useful to say about the pragmatism of profiling ? This is the question we address in this article.

A moment's reflection shows that it really is a question about proper interpretation of conditional probabilities. Let us try to understand through an example. Suppose presence of a certain intent (typically a criminal intent) X is called event A; an example would be the intention to commit an act of violence. And suppose satisfying a certain profile is called event B; an example would be being an individual of a specific ethnic group. From our life experience, as a matter of fact, we *know* that presence of intent X is more common among individuals of profile Y; i.e., the conditional probability

$P(B|A)$ is believed to be high. However, singling out individuals of profile Y is pragmatic or justified if the reverse conditional probability $P(A|B)$ is also adequately high. Unconsciously, we sometimes use one as a proxy for the other. This is obviously an error of interpretation *unless* we can demonstrate that in certain situations, a large value of $P(B|A)$ would imply a correspondingly large value for $P(A|B)$ as well. Thus, the quantity to be analyzed is the ratio $P(A|B)/P(B|A)$. One purpose of this article is to analyze this ratio through a mathematical model, and by use of Bayes' theorem.

But let us probe this a bit deeper. Common sense says that profiling could be pragmatic *even if* $P(A|B)$ is small, if the consequence of not detecting an individual with the stated intent is serious. On the other hand, screening does take resources, manpower, and time. So reckless profiling causing loss of resources, without a corresponding reduction in risk, cannot be very pragmatic either. So as a matter of public policy, we cannot look into even the pragmatism of profiling, leaving morality alone, without consideration of the risks associated with not detecting a person of an unsocial intent, and the resources needed for screening and profiling.

And we have to consider something else. Let us take an example. Suppose, for argument's sake, that terrorist groups got so sophisticated that available screening methods could not detect terrorist paraphernalia, for example, bombs or other explosive devices. Then, as a matter of pragmatism, profile based screening would evidently be useless. Screening would not detect a person intending to commit a crime. Then what is the use of screening ? Thus, we not only need to consider the question of risks and costs, but also the efficiency of the screening procedures, i.e., the probability that our screening methods will actually detect an individual with a criminal intent. To put it all together, whether profiling is a pragmatic thing to do is a decision problem. Indeed, we will approach it as a decision problem. This is a second purpose of this article.

For the benefit of the reader more interested in the conclusions rather than the mathematics, we give a summary of our findings. The findings are quite interesting. We find that

(1) If the cost (consequence) of failing to detect a person of criminal intent is much higher than the cost (say in dollars) of screening an innocent individual, then under reasonable variations of the mathematical model, *significant screening is pragmatically justified* (see Table 2 following Theorem 3).

(2) Moreover, there is something like an *optimum* screening fraction. For example, the optimum screening fraction could be 40%. The meaning is that 40% of individuals satisfying the stated profile Y should be screened, and quite interestingly, this optimum screening fraction is rather *robust* with respect to the parameters of the mathematical model.

(3) A main conclusion of our analysis is that the seriousness of the stated intent is a more important factor in screening decisions than how prevalent the intent is (see Theorem 3 and the discussion following Example 1).

(4) Treating the two conditional probabilities $P(B|A)$ and $P(A|B)$ as random variables, the distribution of the ratio $P(A|B)/P(B|A)$ typically has a very long right tail. The implication is that if indeed $P(B|A)$ is large, as our life experience might try to tell us, then a large value of $P(A|B)$ cannot be ruled out, and in fact $P(B|A)$ and $P(A|B)$ would be of *comparable magnitude*, under reasonable variations of our mathematical model (see Corollary 2, Table 1 and the discussion following Table 1).

(5) If the consequences of the particular intent are very serious, and if screening technology can be trusted, then 100% screening is optimal (see Theorem 4).

A point of technical interest is worth mentioning. In the final section we have stated the sufficient conditions under which 100% screening is recommended. The point of technical interest is that certain classical facts about characterization of nonnegative polynomials and a classic inequality of Markov on the supnorm of derivatives of polynomials are used in deriving these sufficient conditions. Considering the current public appeal of the issue, we have deliberately attempted to present our findings with many graphs and tables. We hope we have conducted an imprejudiced and useful enquiry into this important issue of our public policy.

2 Notation and Mathematical Model

First we describe the notation used in the rest of the article. The models would be described next. The following notation would be used :

$A =$ Presence of a specific unlawful intent X;

$B =$ Having a specific profile Y;

$P(A|B) =$ The conditional probability of A given B; $P(B|A) =$ The conditional probability of B given A;

$\lambda = E(P(A|B))$;

(NOTE : The meaning is that we are going to have a formal distribution for $P(A|B)$ to reflect our beliefs, and λ is the expected value under that distribution (the prior).)

p = Fraction of individuals of profile Y to get screened;

$f(p)$ = Probability that screening fails to detect an individual with an unlawful intent;

(NOTE: this probablity is allowed to depend on the screening fraction p because the screeners are likely to make more errors if more individuals are screened. Thus,$f(p)$ is likely to be an increasing function of p; see the subsequent sections for further details.)

N_1 = Cost of screening an innocent individual;

N_2 = Cost of failing to detect an individual with an unlawful intent;

(NOTE : In practice, assigning a real dollar value to N_2 would be difficult. Thus we can only make qualitative judgements, rather than precise judgements.)

2.1 The Mathematical Model

Two different probability models will be presented and analyzed. The basis for both models is *Bayes' Theorem* which says that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$$

$$\Leftrightarrow \frac{P(A|B)}{P(B|A)} = \frac{P(A)}{P(B)}.$$

Under one model, we study the ratio $\frac{P(A|B)}{P(B|A)}$ by assigning a joint probability distribution to $(P(A), P(B))$; in the other model, we study $P(A|B)$

directly by assigning a joint probability distribution to $(P(A), P(B|A^c))$. The details are given a bit later. But Bayes' theorem is the basis for the analysis under each model.

2.1.1 Model I

First, we should mention that distributions for $P(A), P(B)$ are at all needed because we will not have precise information about what percentage of the population have the specific criminal intent or satisfy the specific profile. We will have to build a distribution from imprecise knowledge. Note, however, that certain intents may be very very rare, and others more common. For example, drug law violation is certainly much more common than wanting to commit terrorist acts. Thus, the distributions for $P(A), P(B)$ would have to be selected with care and caution.

Under our first model, we denote $P(A) = X, P(B) = Y$, and let (X, Y) have a joint *bivariate Beta density* in the unit square $[0,1] \times [0,1]$. The bivariate beta density is defined as follows :

$$f(x,y) = cx^{\alpha_1-1}(1-x)^{\beta_1-1}y^{\alpha_2-1}(1-y)^{\beta_2-1}$$
$$(1+\lambda(x-\tfrac{\alpha_1}{\alpha_1+\beta_1})(y-\tfrac{\alpha_2}{\alpha_2+\beta_2})), \tag{2.1}$$

where λ is any number in the interval $[-1,1]$, and c is the constant

$$c = \Gamma(\alpha_1+\beta_1)\Gamma(\alpha_2+\beta_2)/(\Gamma(\alpha_1)\Gamma(\beta_1)\Gamma(\alpha_2)\Gamma(\beta_2)). \tag{2.2}$$

This is indeed a density function, and furthermore the marginal densities are $Beta(\alpha_1, \beta_1)$ and $Beta(\alpha_2, \beta_2)$, respectively. However, X and Y are not independent. It was important that we did not use a model in which $P(A), P(B)$ are independent because it is likely that there is some correlation between the two. The correlation can be adjusted by playing with the parameter λ. Indeed, the correlation is equal to $\lambda\sigma_1\sigma_2$, where σ_1, σ_2 are the standard deviations of X, Y. See Sarmanov(1966) for further information on this version of the bivariate beta density.

2.1.2 Model II

Under our second model, we assume that $P(B|A)$ is known to us, and assume a joint distribution on $(P(A), P(B|A^c))$. Denoting $P(A^c) = X$, and $P(B|A^c) = Y$, Bayes' theorem produces a distribution for $P(A|B)$ by the identity

$$P(A|B) = \tfrac{\theta(1-X)}{\theta(1-X)+XY},$$

where θ denotes the known value for $P(B|A)$. Note that there are probably certain problems in which $P(B|A)$ is indeed known in the sense certain activities are believed to be committed by a specific subgroup of the population which would be the profile Y.

It seems reasonable to assume independent *Beta* densities for X and Y (because one is a marginal and the other a conditional, and there is no obvious reason that one should affect the other). Thus, under model II,

$$X \sim Beta(\alpha_1, \beta_1) \text{ and } Y \sim Beta(\alpha_2, \beta_2), X, Y \text{ independent.}$$

3 Analysis under Model I

Theorem 1 Let $(X, Y) \triangleq (P(A), P(B))$ have the bivariate Beta density defined above. Let $F(a, b; c; x)$ denote the Hypergeometric ${}_2F_1$ function, and $\mu_1 = \alpha_1/(\alpha_1 + \beta_1), \mu_2 = \alpha_2/(\alpha_2 + \beta_2)$. Then the density $h(u)$ of $U \triangleq \frac{P(A|B)}{P(B|A)}$ is given by the following (complicated) formula :

For $u \leq 1, h(u) = c[(1+\lambda\mu_1\mu_2)u^{\alpha_1-1}\Gamma(\alpha_1+\alpha_2)\Gamma(\beta_2)/\Gamma(\alpha_1+\alpha_2+\beta_2)F(\alpha_1+\alpha_2, 1-\beta_1; \alpha_1+\alpha_2+\beta_2; u) + \lambda u^{\alpha_1-1}\Gamma(\alpha_1+\alpha_2+2)\Gamma(\beta_2)/\Gamma(\alpha_1+\alpha_2+\beta_2+2)F(\alpha_1+\alpha_2+2, 1-\beta_1; \alpha_1+\alpha_2+\beta_2+2; u) - \lambda\mu_2 u^{\alpha_1}\Gamma(\alpha_1+\alpha_2+1)\Gamma(\beta_2)/\Gamma(\alpha_1+\alpha_2+\beta_2+1)F(\alpha_1+\alpha_2+1, 1-\beta_1; \alpha_1+\alpha_2+\beta_2+1; u) - \lambda\mu_1 u^{\alpha_1-1}\Gamma(\alpha_1+\alpha_2+1)\Gamma(\beta_2)/\Gamma(\alpha_1+\alpha_2+\beta_2+1)F(\alpha_1+\alpha_2+1, 1-\beta_1; \alpha_1+\alpha_2+\beta_2+1; u)];$

(3.1)

For $u \geq 1, h(u) = c[(1+\lambda\mu_1\mu_2)u^{-\alpha_2-1}\Gamma(\alpha_1+\alpha_2)\Gamma(\beta_1)/\Gamma(\alpha_1+\alpha_2+\beta_1)F(\alpha_1+\alpha_2, 1-\beta_2; \alpha_1+\alpha_2+\beta_1; 1/u) + \lambda u^{-\alpha_2-3}\Gamma(\alpha_1+\alpha_2+2)\Gamma(\beta_1)/\Gamma(\alpha_1+\alpha_2+\beta_1+2)F(\alpha_1+\alpha_2+2, 1-\beta_2; \alpha_1+\alpha_2+\beta_1+2; 1/u) - \lambda\mu_2 u^{-\alpha_2-1}\Gamma(\alpha_1+\alpha_2+1)\Gamma(\beta_1)/\Gamma(\alpha_1+\alpha_2+\beta_1+1)F(\alpha_1+\alpha_2+1, 1-\beta_2; \alpha_1+\alpha_2+\beta_1+1; 1/u) - \lambda\mu_1 u^{-\alpha_2-2}\Gamma(\alpha_1+\alpha_2+1)\Gamma(\beta_1)/\Gamma(\alpha_1+\alpha_2+\beta_1+1)F(\alpha_1+\alpha_2+1, 1-\beta_2; \alpha_1+\alpha_2+\beta_1+1; 1/u)].$

(3.2)

Proof : There is not much point in showing the calculation in detail. The main steps are as follows :

(i) For a given joint density $f(x, y)$ in the unit square $[0, 1] \times [0, 1]$, by a standard jacobian calculation, the density of $U = \frac{X}{Y}$ is given by $h(u) = \int_0^{min(1/u,1)} v f(uv, v)dv$.

(ii) Use now the explicit form of the bivariate Beta density $f(x,y)$ and break up $(x-\mu_1)(y-\mu_2)$ as $xy - \mu_2 x - \mu_1 y + \mu_1\mu_2$.

(iii) Use the following two integration formulas :

(a) $\int_0^1 v^{a-1}(1-v)^{b-1}(1-uv)^{c-1}dv$

$= \Gamma(a)\Gamma(b)/\Gamma(a+b)F(a, 1-c; a+b; u);$

$$(3.3)$$

(b) $\int_0^{1/u} v^{a-1}(1-v)^{b-1}(1-uv)^{c-1}dv$

$= \Gamma(a)\Gamma(c)/\Gamma(a+c)F(a, 1-b; a+c; 1/u)u^{-a};$

$$(3.4)$$

(see Gradshteyn and Ryzhik(1980)).

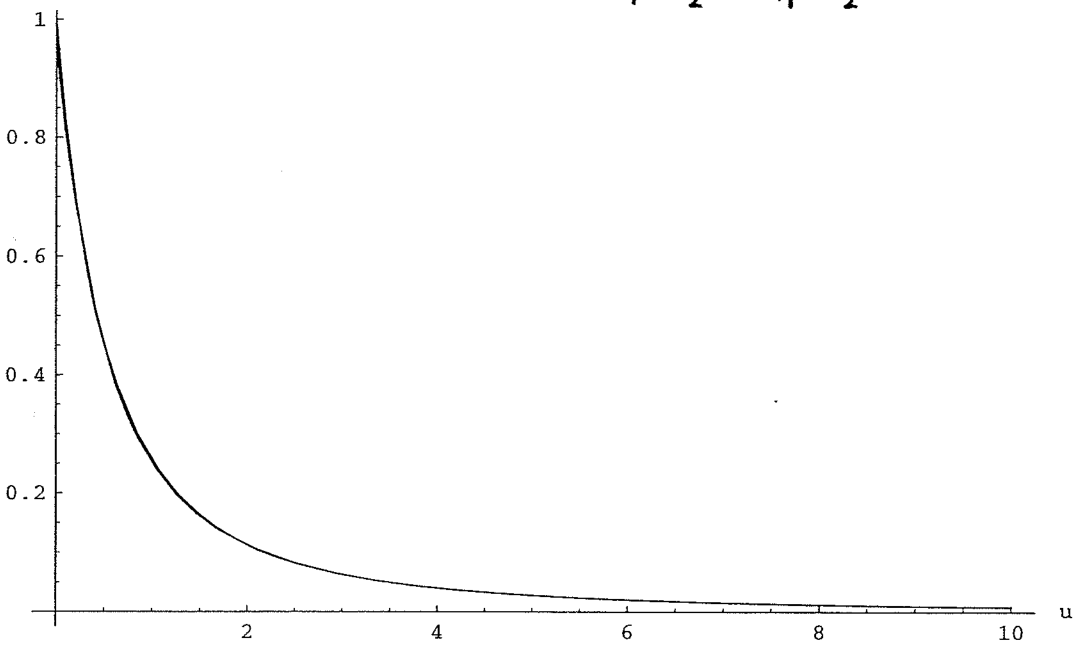
(iv) Integrate each of the four terms obtained from splitting apart $vf(uv,v)$ as indicated in (ii) by using the integration formulas in (iii), separately for $u \le 1$ and $u > 1$, and then put the four terms back together again.

Discussion It could be of interest to see the form of the density function $h(u)$ of the ratio $\frac{P(A|B)}{P(B|A)}$, and to observe the behavior of the expected value of $\frac{P(A|B)}{P(B|A)}$ under our bivariate Beta model.

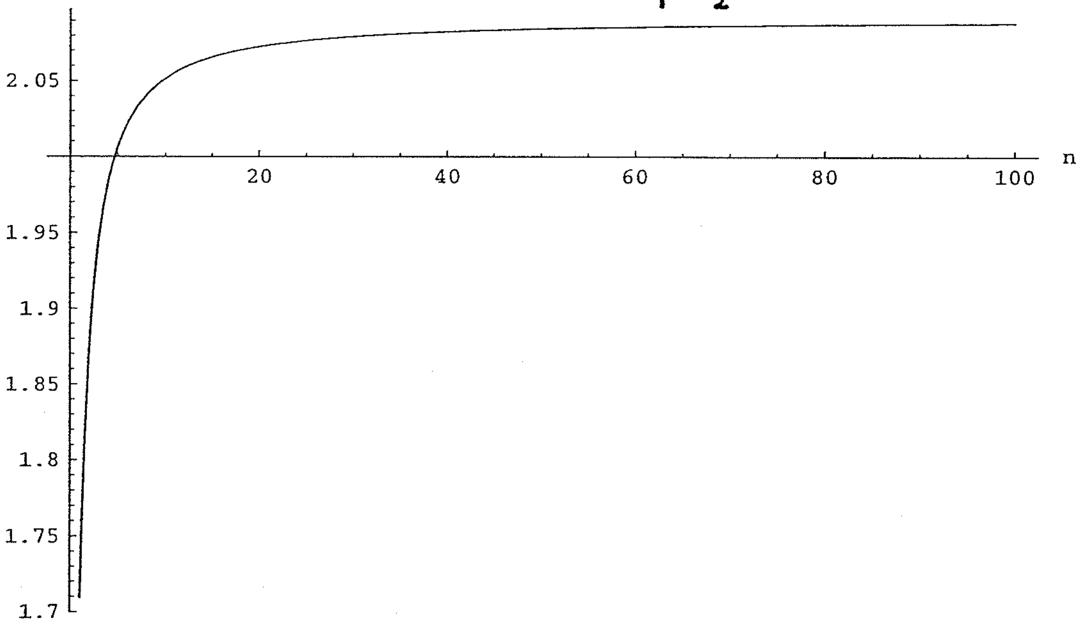
The next two graphs plot the density function of $U = \frac{P(A|B)}{P(B|A)}$ and its expected value when the parameters α_1, α_2 are taken to be 1, and the parameters β_1, β_2 are allowed to vary. The reason for taking α_1, α_2 to be 1 is that then the marginal densities of $P(A), P(B)$ are monotone decreasing. This is quite important in some applications in our context. If the specific criminal intent X as well as the specific profile Y are *RARE*, then we should reasonably formulate the densities of $P(A), P(B)$ to be decreasing. The parameters β_1, β_2 should depend on how rare the traits are. For example, if they occur one in hundred times, then β_1, β_2 should be roughly 100 each.

The most striking observations from these two pictures are the following. The density function of $\frac{P(A|B)}{P(B|A)}$ has a long right tail but is insensitive to the common value of the β parameters as we can see from the superimposed plots. They look virtually identical. The other obvious feature from the plot of the expected value of $\frac{P(A|B)}{P(B|A)}$ is that it settles down to an asymptote, a bit larger than 2, rather quickly, and like the density function, the effect of the β parameters is limited. Note also that the asymptote is just about 2. Loosely speaking, this means that $P(A|B)$ would not differ from its dual, namely, $P(B|A)$, by a factor too much larger than 2. The plot would thus suggest that, on the average, the two conditionals are of roughly the same magnitude, a positive and pleasant property. Recall that this is exactly the kind of information we want in making the inference that $P(A|B)$ is large if $P(B|A)$ is large. Model I and its analysis have provided us with some tangible and useful conclusions.

Density of U = P(A|B)/P(B|A) when $\lambda = 1$, $\alpha_1 = \alpha_2 = 1$, $\beta_1 = \beta_2 = 20$ (100)



Plot of E[P(A|B)/P(B|A)] when $\lambda = 1$, $\alpha_1 = \alpha_2 = 1$, and m = n



4 Analysis under Model II

In the second model, we take $P(B|A)$ to be known. We assign a probability distribution to $(P(A), P(B|A^c))$, and use Bayes' theorem to produce a distribution for $P(A|B)$.

Theorem 2 Let $P(B|A) = \theta$ (known) and let $(X, Y) \triangleq (P(A^c), P(B|A^c))$ be independently distributed as $Beta(\alpha_1, \beta_1)$, $Beta(\alpha_2, \beta_2)$, respectively. Then the density of $Z = P(A|B)$ is given by

$$h(z) = \eta^{\beta_1}\Gamma(\alpha_1+\beta_1)\Gamma(\alpha_2+\beta_1)\Gamma(\alpha_2+\beta_2)/(\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\beta_1)\Gamma(\alpha_2+\beta_1+\beta_2))\times$$

$$z^{\beta_1-1}F(\alpha_2+\beta_1, \alpha_1+\beta_1; \alpha_2+\beta_1+\beta_2; -\eta\tfrac{z}{1-z})/(1-z)^{\beta_1+1}, \tag{4.1}$$

where $\eta = \frac{1}{\theta}$.

Proof : It is more convenient to outline the steps for the case of general densities$f(x), g(y)$ for X, Y. Suppose then X, Y are independent random variables taking values in $[0, 1]$, with densities $f(x), g(y)$. Denote $U = \frac{X}{1-X}$, $W = UY$, and $Z = \frac{1}{1+\eta W}$. With this notation, the conditional probability $P(A|B)$, from Bayes' theorem, is just Z. The steps are as follows. We omit the details of these steps.

(i) The joint density of (U, Y) is $f(\frac{u}{1+u})g(y)/(1+u)^2, u > 0, 0 < y < 1$.

(ii) Therefore, the joint density of (W, Y) is $yg(y)f(\frac{w}{w+y})/(w+y)^2, w > 0, 0 < y < 1$.

(iii) On integrating y out, the marginal density of W is $\int_0^1 yg(y)f(\frac{w}{w+y})/(w+y)^2dy$.

(iv) Hence, the density of $Z = \frac{1}{1+\eta W}$ is

$\eta \int_0^1 yg(y)f(\frac{1-z}{1-z+\eta zy})/(1-z+\eta zy)^2dy$.

(v) This is the general case. In the special case when f and g are densities of $Beta(\alpha_1, \beta_1), Beta(\alpha_2, \beta_2)$, respectively, the aforementioned integral in step (iv), fortunately, can be done in closed form, and that closed form expression is the formula stated in the theorem.

Corollary 1 Suppose $P(A^c), P(B|A^c)$ are independent $Beta(m, 1), Beta(1, n)$ distributed. Then the density of $Z = P(A|B)$ is

$$h(z) = \frac{m\eta}{n+1}F(2, m+1; n+2; -\eta\frac{z}{1-z})/(1-z)^2.$$

Remark The densities in Corollary 1 have the following motivation. Consider the case when having the intent X as well as the profile Y are rare. Then $P(A)$ is going to be small, and so it is prudent to assign it a decreasing density, or equivalently to assign $P(A^c)$ an increasing density. The particular $Beta(m, 1)$ density does that. Likewise, if intent X is rare, then $P(B|A^c)$ should be more or less the same as the unconditional probability $P(B)$. But if profile Y is also rare, then $P(B)$ would be small, and hence the decreasing $Beta(1, n)$ density is a good one to use in such a case.

The next two results give the expectation of Z and the effect of the value of θ on the expectation of Z.

Corollary 2 Under the density in Corollary 1, (a)$E(Z) = \frac{m(n+1)!}{(n+1)m!} G^{3,1}_{2,3}(\theta \mid \{1, n+1\}, \{1, 1, m\})$, where $G^{m,n}_{p,q}$ denotes the Meijer-G function;

(b) $\frac{d}{d\theta} E(Z) = \frac{1}{\theta} G^{3,2}_{3,4}(\theta \mid \{0, 1, n+1\}, \{1, 1, m, 1\})$.

Proof of Corollary 2: From direct integration of $zh(z)$, one has that $E(Z) = \frac{m(n+1)!}{(n+1)m!} G^{3,2}_{3,4}(\theta \mid \{0, 1, n+1\}, \{1, 1, m, 0\}) = \frac{m(n+1)!}{(n+1)m!} G^{3,1}_{2,3}(\theta \mid \{1, n+1\}, \{1, 1, m\})$, by an identity on the Meijer-G function (see, e.g., Mathai(1993)). The stated form is a bit simpler computationally than the direct form from integration of $zh(z)$. For part(b), use the identity $\frac{d}{d\theta} G^{m,n}_{p,q}(\theta \mid \{a_1, a_2, \ldots, a_p\}, \{b_1, b_2, \ldots, b_q\}) = \frac{1}{\theta} G^{m,n+1}_{p+1,q+1}(\theta \mid \{0, a_1, a_2, \ldots, a_p\}, \{b_1, b_2, \ldots, b_q, 1\})$ (again, see Mathai(1993)).

Let us use Corollary 2 to see some values of $E(P(A|B))$ for selected values of θ, m, and n.

Table 1 : $E(P(A|B)) : m = n = 20$

$\theta = P(B\|A)$	.1	.25	.5	.75	.9
$E(P(A\|B))$	.18	.29	.39	.46	.49

Thus, for *smaller* values of $\theta, E(P(A|B))$ is larger than $P(B|A)$, while the contrary is true for larger values of θ. But, in all cases, they are within a factor of 2 of each other. They are of *comparable magnitude.* This is an important finding.

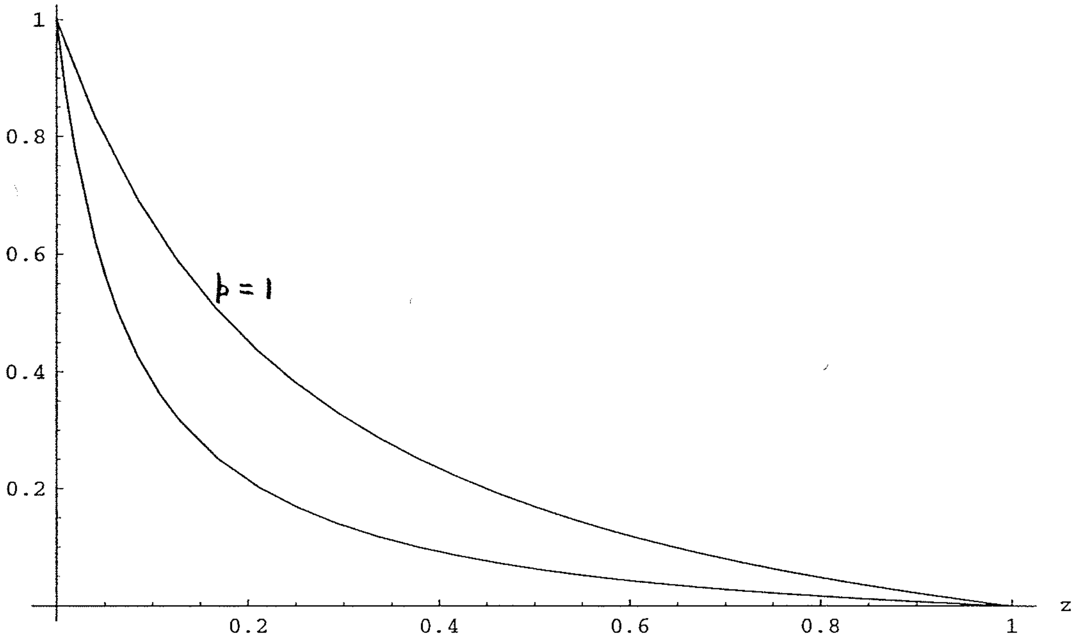
From part (b) of Corollary 2, the effect of θ on $E(P(A|B))$ can be worked out. We found that the derivatives are small for large θ, but large for small θ. Thus, the effect of θ is variable.

Again, it could be interesting to see the shape of the density of $P(A|B)$

in some trial cases. It would also be worth asking what is the influence of the assumption that $P(B|A)$ is known; i.e., how much is the density influenced by the exact value of $\theta = P(B|A)$. The graph below plots the *survival function* of $Z = P(A|B)$ for $\theta = 1$ and $\theta = \frac{1}{3}$, when $m = 100$ and $n = 20$. Thus, the particular criminal intent is supposed to be *more* rare than the particular profile, because m is much larger than n.

What do we see in the plot that is worth noting ? The value of θ has some effect, clearly. But the general shape of the two survival curves are very similar. Moreover, large values of $Z = P(A|B)$ are not unlikely. From the plot, we see that under each curve, $P(Z > .05)$ is between 50 to 90%, $P(Z > .2)$ is between 25 to 50%, and $P(Z > .5)$ is between 10 to 25%. What are we to make of these numbers ? If the particular criminal intent X is a dangerous one (rather than being just self-destructive), then these values for $P(Z > z)$ are probably to be regarded as *high*. In other words, going back to the issue of whether it is valid to conclude that $P(A|B)$ is seriously large by knowing that $P(B|A)$ is large, the conclusion could be that such a leap is at least partially justified. In this regard, what we find by analyzing Model II is consistent with our findings from Model I. We will give a more formal decision theoretic treatment of this question in the next section.

Plot of P(Z > z) when m = 100, n = 20, and p = 1, .33



5 The Optimal Screening Fraction

In this section, we intend to study the question of an optimal screening fraction, formulating it as a decision problem. We must mention at the outset that we surely do not mean that the optimal screening fraction is a number to be religiously used! Rather, the purpose is to study the question of how much screening to do from a broad qualitative viewpoint. The optimal screening fraction should provide some general and useful guidelines about the extent of screening that would be advised in a given context. The context would dictate the various factors that go into determining the optimal amount of screening. Putting the various relevant factors together, we have a decision problem. The calculation, although a formalism, is supposed to be a guide, rather than a strict prescription.

The various factors relevant in this calculation were introduced in section 2. Putting them together, we introduce the *loss function*

$$L = L(f, N_1, N_2, p) = N_1 P(A^c|B)p + N_2 P(A|B)(1-p) + N_2 P(A|B)pf(p). \tag{5.1}$$

Recalling that $E(P(A|B)) = \lambda$, we seek to minimize the risk :

$$R = R(f, N_1, N_2, p) = (N_1(1-\lambda) - N_2\lambda)p + N_2\lambda pf(p) + N_2\lambda. \tag{5.2}$$

The next result characterizes the optimal value of p, i.e., the optimal screening fraction. Some illustrative examples follow this result.

Before stating the theorem, we will give a brief motivation for the assumptions we make on the function $f(p)$, the rate of error in screening individuals. As the screening fraction goes up, we can expect that the error rate goes up too, due to human fatigue, and lack of sufficient time to properly screen someone. So we can expect f to be an increasing function of p. If f were a constant, $pf(p)$ would be linear and so convex; but if f is not a constant, and grows, say according to some positive power of p, then $pf(p)$ would be strictly convex. We assume f to be increasing and $pf(p)$ to be strictly convex in our next theorem.

Theorem 3 Suppose the function $f(.)$ is defined and nondecreasing ,and once continuously differentiable on all of $[0, \infty)$ and assume $xf(x)$ is strictly convex.

(a) The *optimal screening fraction* minimizing R is the unique root of the equation $f(p) + pf'(p) = 1 - \frac{N_1(1-\lambda)}{N_2\lambda}$, if a root p_0 exists and belongs to the interval $[0,1]$. If p_0 is outside of the interval $[0,1]$, the optimal screening fraction is one of the boundary values 0,1, whichever is closer to the root p_0. If a root of $f(p) + pf'(p) = 1 - \frac{N_1(1-\lambda)}{N_2\lambda}$ does not exist, then also the optimal screening fraction is one of 0,1, according as $\frac{N_1(1-\lambda)}{N_2\lambda} >$ or < 1;

(b) Let p^* denote the unique root of $f(p) + pf'(p) = 1$. Then $p_0 \to p^*$ if $\frac{N_1}{N_2} \to 0$, provided the roots exist;

(c) If f is twice differentiable, then

$$p_0 = p^* - \frac{N_1(1-\lambda)}{N_2\lambda} \frac{1}{2f'(p^*) + p^* f''(p^*)} + o(\frac{N_1}{N_2}).$$

Proof: For the proof of part (a), first note that the equation $f(p) + pf'(p) = 1 - \frac{N_1(1-\lambda)}{N_2\lambda}$ has at most one root due to the strict convexity of $pf(p)$. If a root p_0 exists and belongs to the interval $[0,1]$, it minimizes the risk R as $\frac{dR}{dp} = N_1(1-\lambda) - N_2\lambda + N_2\lambda(f(p) + pf'(p))$. If a root p_0 exists but lies outside [0,1], from convexity considerations, the fraction minimizing R is 0 or 1, whichever one is closer to the root p_0. If a root p_0 does not exist, then R is monotone, increasing or decreasing, according as $N_1(1-\lambda) >$ or $< N_2\lambda$, and hence the fraction minimizing R is again 0 or 1.

For part (b), use the fact that if a function g is one-to-one and continuous, if $g(x_n) = y_n, y_n \to y, g(x) = y$, then $x_n \to x$. Now identify g with the function $f(p) + pf'(p)$ and use the strict convexity of $pf(p)$.

Part (c) follows from a Taylor expansion of the function $f(p) + pf'(p)$ around p^* at $p = p_0$.

Example 1 We illustrate Theorem 3 by this example.

Let $f(p)$ be the slowly increasing function $log(1+p)$. For this choice of the error rate in screening, the error probability increases from zero at no screening to about 70% at 100% screening. Let us study the optimal screening fraction by varying the parameters N_1, N_2. We consider six different values of $\frac{N_2}{N_1}$, namely, $\frac{N_2}{N_1} = .25, 1, 10, 100, 1000, 10000$, and three different values of $\lambda = E(P(A|B))$, namely, $\lambda = .01, .1, .3.$. For example, the value $\frac{N_2}{N_1} =$ .25 represents a case where the criminal intent is supposedly benign, and the consequence of not detecting a person with the intent is small. On the other hand, the value 10,000 represents a case where the consequences are very serious. Similarly, the value of λ reflects the prevalentness of the said intent. By selecting a range of values for $\frac{N_2}{N_1}$ and λ, we wish to qualitatively study how the optimal screening fraction would vary with the seriousness and prevalence of the criminal intent.

Application of Theorem 3 results in the following table :

Table 2: Optimal Screening Fraction

$\frac{N_2}{N_1}$	$\lambda = .01$	$\lambda = .1$	$\lambda = .3$
.25	0	0	0
1	0	0	0
10	0	.052	.525
100	.005	.666	.737
1000	.656	.753	.761
10000	.752	.762	.763

Discussion The most interesting things to learn from this Table are the following :

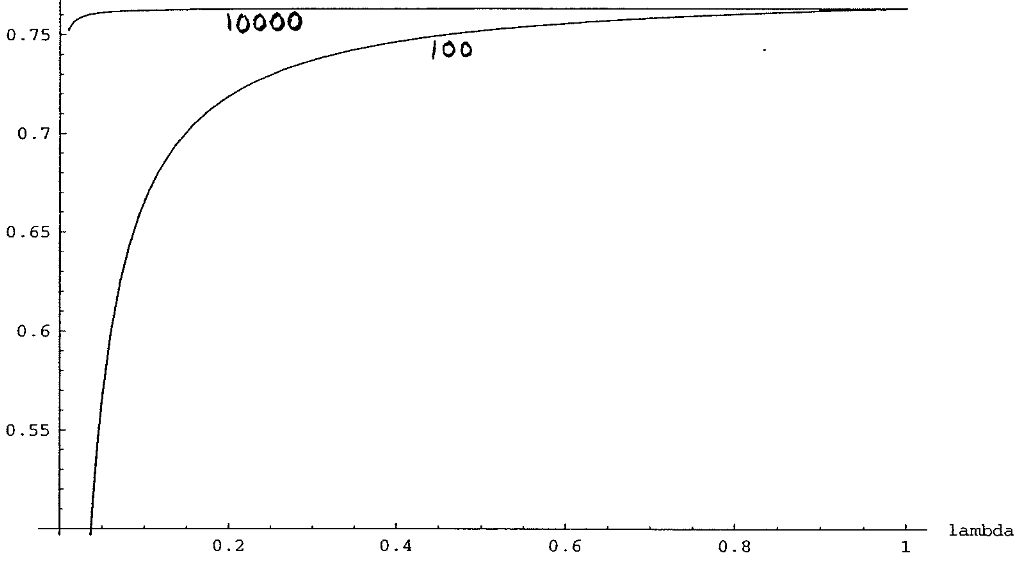
(i) The optimal screening proportion is very stable, almost independently of the value of λ, when $\frac{N_2}{N_1}$ is large. This *stable value* is just the manifestation of part (b) of our Theorem 3. It is the value p^* discussed in Theorem 3. Thus, *if the criminal intent is of such a type that failure to detect it has serious consequences, then significant amount of screening seems to be a good idea, regardless of how prevalent the intent is.* This is our finding, even if it is likely to be controversial, and open to questions, because we have certainly not done a comprehensive study. For example, the function $f(p)$ has *NOT* been varied in this example. But see the pictures below.

(ii) If the criminal intent is of a rather benign variety, then screening does not make much pragmatic sense as we see from the zero values in the above Table. But screening *can still make a lot of moral sense.* However, we do not go into that issue.

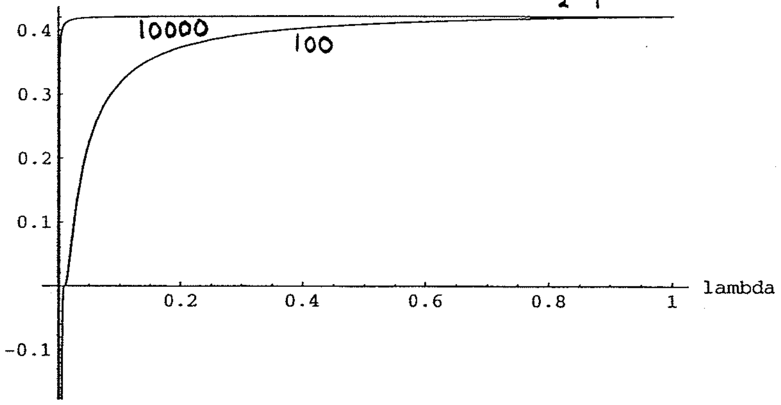
(iii) The only situation where λ, i.e., the prevalence of the intent makes a difference is when the consequences of the intent are neither too benign nor too serious, as we can see, e.g., when $\frac{N_2}{N_1} = 10$. It is thus very interesting to see that the seriousness of the intent is more important than prevalence of the intent in deciding how much to screen. We believe this is an important conclusion.

The next two pictures plot the optimal screening fraction for two different choices of $f(p)$, namely, $f(p) = \log(1 + p)$, and $p^{1/3}$. Note that the first is a function of slow variation, while the second is a function of regular variation. The pictures support the main phenomena (i) - (iii) we discussed above. For example, take the case when $f(p) = p^{1/3}$ and $\frac{N_2}{N_1} = 10000$. Then, almost regardless of the value of λ, the optimal screening fraction is about 40%. Of course, the suggestion is not that we are to follow a 40% rule or a 75% rule strictly. The suggestion in broad terms is to conduct significant screening whenever the behavior in consideration has very serious consequences, without much regard to how rare the behavior is.

Optimal Screening Fraction as a function of E[P(A|B)] when N_2/N_1 = 100(10,000), f(p) = log(1+p)



Optimal Screening Fraction as a function of E[P(A|B)] when N_2/N_1 = 100(10,000), f(p) = $p^{1/3}$



6 Conditions when 100% Screening is Recommended

In this final section, we give a set of conditions under which 100% screening is recommended. We do need to again emphasize that this is a formalism in the sense most mathematical results are formalisms. But the conditions are intended to serve as a guide for situations when 100% or *close to* 100% screening would be recommended. The result, rather interestingly, uses an old classic inequality about variations of polynomials. The spirit of the result, as one may well anticipate, is that if the consequence of failing to detect an individual with the stated intent is serious, and if screening methods themselves are reliable, then one should go for 100% screening.

Theorem 4 Suppose the screening error rate $f(p)$ is a polynomial function of some degree n. Suppose moreover that f is nondecreasing and that $||f||_\infty = sup_{0 \leq p \leq 1} f(p) = f(1) \leq \frac{\delta}{2n^2+1}$, where $\delta = 1 - \frac{N_1(1-\lambda)}{N_2\lambda}$. Then the optimal screening fraction equals 1.

First let us illustrate this result for the case when $f(p)$ is a cubic polynomial.

Example 2 Let $f(p) = a_0 + a_1 p + a_2 p^2 + a_3 p^3$. Then $f(p)$ is increasing iff the derivative $f'(p) = a_1 + 2a_2 p + 3a_3 p^2 \geq 0 \forall p \in [0, 1]$. Now, another classic result about polynomials says that an even degree polynomial is nonnegative on $[0, 1]$ iff it is of the form $P^2(p) + p(1-p)Q^2(p)$(see, e.g., Sźego(1975)). Because the derivative $f'(p)$ is of degree 2 in this example, P is a linear polynomial, and Q is a constant. A little bit of manipulation then shows that $f'(p) \geq 0 \forall p \in [0, 1]$ iff :

$a_1 \geq 0, a_1 + 2a_2 + 3a_3 \geq 0, a_1 + a_2 \geq \sqrt{a_1(a_1 + 2a_2 + 3a_3)}$, and $a_1 + a_2 + \frac{3}{2}a_3 \geq \sqrt{a_1(a_1 + 2a_2 + 3a_3)}$.

(6.1)

Since $f(p)$ is nondecreasing, $||f||_\infty = f(1) = a_0 + a_1 + a_2 + a_3$. Therefore, Theorem 4 says that if the error rate $f(p)$ is a cubic polynomial $a_0 + a_1 p + a_2 p^2 + a_3 p^3$ with any coefficients satisfying these five conditions :

$a_1 \geq 0, a_1 + 2a_2 + 3a_3 \geq 0, a_1 + a_2 \geq \sqrt{a_1(a_1 + 2a_2 + 3a_3)}$,$a_1 + a_2 + \frac{3}{2}a_3 \geq \sqrt{a_1(a_1 + 2a_2 + 3a_3)}$, and $a_0 + a_1 + a_2 + a_3 \leq \frac{\delta}{19}$, then the optimal screening fraction is 100%.

Similar examples can be worked out when $f(p)$ is a quartic or a quadratic. We will now sketch the proof of Theorem 4.

Proof of Theorem 4: Denote $f(p) + pf'(p)$ by $g(p)$. The theorem will be proved if we show that $||g||_\infty \leq \delta$. The proof uses the classic Markov inequality that for a polynomial g of degree n on an interval $[a, b]$, $||g'||_\infty \leq \frac{2n^2}{b-a}||g||_\infty$ (see, e.g., Bullen(1998)). Therefore, if $||f||_\infty \leq \frac{\delta}{2n^2+1}$, then,

$||g||_\infty \leq ||f||_\infty + ||f'||_\infty$ (by the triangular inequality) $\leq ||f||_\infty + 2n^2||f||_\infty$ (by Markov's inequality) $= (2n^2 + 1)||f||_\infty \leq \delta$, proving the assertion made in the statement of Theorem 4.

An examination of the proof shows that we did not use the assumption that f is increasing. But in practice it should hold.

Bibliography

Bullen,P.(1998).A Dictionary of Inequalities, Addison-Wesley Longman, Edinburgh, England.

Gradshteyn,I. and Ryzhik,I.(1980).Table of Integrals, Series and Products, Academic Press, New York.

Mathai,A.M.(1993).A Handbook of Generalized Special Functions for Statistical and Physical Sciences, Oxford Science Publications, New York.

Sarmanov,O.V.(1966).Generalized normal correlation and two-dimensional Fréchet classes, Doklady Soviet Math, 7, 596-599.

Sźego,G.(1975).Orthogonal Polynomials, Amer. Math. Soc., Providence, Rhode Island.