The Dempster-Shafer Theory for Single Parameter Univariate Distributions

by

C. Liu
Purdue University

Technical Report #07-12

Department of Statistics
Purdue University

October 2007

Technical Report

THE DEMPSTER-SHAFER THEORY FOR SINGLE PARAMETER UNIVARIATE DISTRIBUTIONS

BY CHUANHAI LIU

Department of Statistics, Purdue University

The Dempster-Shafer (DS) theory is attractive because the goal resembles the Bayesian goal, while the model assumptions mimic the weaker assumptions of sampling models. Following Dempster's 1966 idea remarked in his first paper [1] on DS, we are concerned in this article with DS inference for single parameter univariate distributions. It is shown that the DS posterior has a simple analytic form in terms of commonality for a class of univariate distributions with a single continuous parameters, which essentially includes exponential families as special cases. Asymptotic results of DS posteriors are also obtained, and investigated in detail for the univariate normal distribution $N(M, 1)$ with unknown mean M to be estimated.

1. Introduction. The Dempster-Shafer (DS) theory, or the theory of belief functions, has been known and attractive to many applied scientists needing situation-specific assessments of uncertainty. Dempster's original papers in 1960s [1, 2] were aimed at parametric statistical inference, and were designed to bridge a gap that continues to exist between methods based on sampling distributions and methods based on Bayesian posteriors. Soon thereafter, Glenn Shafer's 1976 book [7] clarified the mathematical basis for wider readership. Arthur Dempster's recent paper [4] provides an excellent resource for understanding of DS methodology such as the new terms, operations, and concepts introduced for DS analysis. For example, the theory leads to computed inferences of the form (p, q, r), with $p + q + r = 1$, where p expresses probability for the truth of an assertion, q expresses probability against the truth of the assertion, and r represents a residual probability of the new category of "don't know". On the contrary, the Bayesian theory does not allow for "don't know" and therefore simply reduces DS output (p, q, r) for an assertion to (p, q) with $p + q = 1$.

Statistical research on DS in the last 40 years has been somewhat slow due to the fact that the computations required to carry out the methods have been recognized to demand more than frequentist and Bayesian

AMS 2000 subject classifications: Primary 62F99; secondary 62F15

Keywords and phrases: Exponential families, Bayesian inference, Multinomial distribution, Poisson processes, Significance test

methods for widespread practical use. It can be expected that the required computing power for DS becomes affordable by major and continuing advances that have been accompanied by growing knowledge and experience among statistical researchers with Monte Carlo methods for approximate high-dimensional numerical integration. Associated theoretical studies of novel DS procedures can thus be expected to shed new light on unresolved differences between "Bayesian" and "frequentist" approaches to inference, including asymptotic properties.

It has long been believed that DS inference for continuous distributions, for example, $N(M, 1)$ with unknown mean M, would require computationally intensive large-scale simulations. As an integral part of DS, Dempster's original idea was to approximate the state space of a continuous random variable by a large number of tiny cells. That is, for a sample of size n from a standard parametric population model, the observed data are in effect multinomial, with n cells having one observation each and the remaining large number of cells having zero observations each. Consequently, DS inference for multinomial sampling is crucial and is the focus of Dempster's original paper [1]. Moreover, it was remarked at the conclusion of [1] that as long as certain discrete sums from approximating a continuous sample space converge to integrals in an expected way, the multinomial inferences converge to a well-defined unique DS model (DSM) for the parameters.

This article presents analytic results in DS inference for continuous distributions. The results show that DSMs for a class of univariate parametric models, including exponential families, have simple closed-form expressions in terms of *commonality*. The results are illustrated with familiar parametric models: the normal distribution $N(M, 1)$ with unknown mean M, the Poisson distribution Poisson (L) with unknown rate L, the Gamma distribution Gamma(α, B) with unknown rate B, and the negative binomial distribution NB (n_0, P) with unknown probability P. Asymptotic results of DS posteriors are also obtained, and investigated in detail for the univariate normal distribution $N(M, 1)$ with unknown mean M to be estimated.

Section 2 provides a brief review of needed DS models and calculus for carrying through above limiting posterior DSM inferences for univariate distributions with single parameters. An alternative to the DSM [1] for the multinomial model is formulated in terms of first arrival times of Poisson processes. Sections 3 and 4 establish DSMs for discrete and continuous distributions based on single observations. Section 5 discusses combined DSMs given a sample. Section 6 characterizes the large sample behavior of DS parametric inferences, where agreement in a sense with limiting Bayesian inferences is shown and the degree of "don't know" going away for large sample size is investigated.

2. The Dempster-Shafer theory: relevant basics. Dempster [1] proposed to use associated- (a-) random sets for making inference about multinomial distribution. Suppose that $X \sim \text{Multinomial}\,(P_1, ..., P_k)$, where

$$P \in \mathcal{S}_k \equiv \{(s_1, ..., s_k) : s_1 \geq 0, ..., s_k \geq 0 \text{ and } \textstyle\sum_{j=1}^{k} s_j = 1\}, \tag{2.1}$$

that is, $\Pr(X = j|P) = P_j$ for $j = 1, ...k$. The set $\mathcal{S}_k$ is known as the $(k-1)$-simplex. Dempster [1] employed the *barycentric system* as a convenient mathematical tool. Let $V_1 = (1, 0, 0, ..., 0), V_2 = (0, 1, 0, ..., 0), \ldots, V_k = (0, 0, 0, ..., 1)$ be the k vertices of the simplex. Any point $P \in \mathcal{S}$ can be written as

$$P = P_1V_1 + P_2V_2 + \ldots + P_kV_k,$$

i.e., the coefficients $(P_1, ..., P_k)$ are *barycentric coordinates* of P with respect to $V_1, ..., V_k$. If we imagine masses equal to $P_1, ..., P_k$ attached to the vertices of the simplex, the center of the mass (the barycenter) is then P. We refer to [1] for geometric explanation of the DS model for estimation of the multinomial distribution.

The a-random variable in the DSM [1] is a random point $U = (U_1, ..., U_k)$ following the uniform distribution on the $(k-1)$ simplex $\mathcal{S}_k$, *i.e.*, the Dirichlet distribution $\text{Dirichlet}_k\,(1, ..., 1)$. Given $P = (P_1, ..., P_k)$, the $(k-1)$-simplex is partitioned into k simplex regions:

$$\mathcal{U}_i(P) = \left\{U : U \in \mathcal{S}_k \text{ and } \tfrac{U_i}{P_i} \leq \tfrac{U_j}{P_j} \text{ for } j \neq i\right\} \tag{2.2}$$

Here we provide an alternative argument showing that $\Pr(U \in \mathcal{U}_i(P)|P) = P_i$ for $i = 1, ..., k$. It is well known that the $\text{Dirichlet}_k\,(1, ..., 1)$ random variable $U = (U_1, ..., U_k)$ can be represented by k *iid* standard exponential random variables $E_1, ..., E_k$ as follows:

$$U_j = \frac{E_j}{E_1 + \ldots + E_k} \qquad (j = 1, ..., k)$$

Note that the inequality $\frac{U_i}{P_i} \leq \frac{U_j}{P_j}$ in (2.2) is equivalent to the inequality $\frac{E_i}{P_i} \leq \frac{E_j}{P_j}$. Therefore, for given P we have

$$\Pr(X = i) = \Pr\left(\tfrac{E_i}{P_i} \leq \tfrac{E_j}{P_j} \text{ for } j \neq i\right) \overset{z=E_i}{=} \int_0^\infty e^{-\frac{z}{P_i}\left[P_i + \sum_{j\neq i} P_j\right]} dz = P_i,$$

that is, $X|P \sim \text{Multinomial}\,(P_1, ..., P_k)$.

The above argument suggests that the DS model corresponding to (2.2) can be interpreted in terms of first arrival times of k independent Poisson processes with rates proportional to $P_1, ..., P_k$. In terms of *iid* a-random exponential variables, the multinomial DS model (2.2) for inference about unknown P has the following a-random region

$$R_X = \left\{P : P \in \mathcal{S}_k \text{ and } \tfrac{E_X}{P_X} \leq \tfrac{E_j}{P_j} \text{ for } j \neq X\right\} \tag{2.3}$$

where $E_i \overset{iid}{\sim} \mathrm{Exp}\,(1)$ for $i = 1, ..., k$. Since U in DSM [1] is a many-to-one function of $E = (E_1, ..., E_k)'$, we refer to this multinomial DS model as the Augmented DS model (ADSM).

DS inference based on a sample or even a single observation relies on Dempster's rule of combination [1, 4, 7]. Here we consider the simple situation where the a-random set about a single parameter $T \in \Theta \subseteq \mathcal{R}$ in an a-random interval $[V_L, V_U] \subseteq \Theta$. The a-random interval $[V_L, V_U]$ can be characterized by the joint distribution of (V_L, V_U) or equivalently the commonality, the probability that the a-random interval $[V_L, V_U]$ covers the fixed interval $[a, b]$:

$$c([a,b]) = \Pr(V_L \leq a, V_U \geq b) \qquad (a \leq b \in \Theta)$$

Commonality provides a simple way of combining independent pieces of information because it multiplies under Dempster's rule of combination. More specifically, suppose that there are n independent DSMs about a common single parameter $T \in \Theta$ with commonalities $c_i([a, b])$ for $i = 1, ..., n$. Then the combined DSM has commonality

$$c([a,b]) \propto \prod_{i=1}^{n} c_i([a,b])$$

In addition, if the a-joint distribution of a-random interval (V_L, V_U) is continuous, its pdf, $m(a, b)$, can be obtained as follows:

$$m(a,b) \propto -\frac{\partial^2 c([a,b])}{\partial a \partial b}.$$

Furthermore, the a-pdf of V_L, referred to as the lower a-pdf, is given by

$$\underline{h}(\theta) \propto \left.\frac{\partial c([a,b])}{\partial a}\right|_{a=b=\theta} \tag{2.4}$$

and the a-cdf of V_U, referred to as the upper a-cdf, is

$$\overline{h}(\theta) \propto -\left.\frac{\partial c([a,b])}{\partial b}\right|_{a=b=\theta}. \tag{2.5}$$

Let $\underline{H}(\theta)$ and $\overline{H}(\theta)$ be the cdfs corresponding to $\underline{h}(\theta)$ and $\overline{h}(\theta)$. We call $\underline{H}(\theta)$ and $\overline{H}(\theta)$ the lower and upper a-cdfs, respectively. The DS output (p, q, r) for the assertion $\{T \leq T_0\}$ about the parameter T is obtained as follows:

$$p = \overline{H}(T_0), \quad q = 1 - \underline{H}(T_0), \quad \text{and} \quad r = 1 - p - q = \underline{H}(T_0) - \overline{H}(T_0), \tag{2.6}$$

where T_0 is a known constant in Θ,

3. Discrete distributions. Suppose that a single observation X is to be considered from a discrete distribution $F(x|T)$ with unknown parameter $T \in \Theta \subseteq \mathcal{R}^1$, where the sample space is $\mathcal{X} = \{0, 1, 2, \ldots\}$. We denote by $f(x|T)$ the probability density function. For a multinomial approximation to $F(x|T)$, we consider the DS estimation of a truncated version of $F(x|T)$ for $x \leq N$ from a single observation $X \leq N$:

$$\frac{E_X}{f(X|T)} \leq \frac{E_y}{f(y|T)} \qquad (y \neq X, y = 0, 1, ..., N)$$

where $E_0, ..., E_N$ are *iid* with the standard exponential distribution. DS estimation of $F(X|T)$ is obtained by letting N go to infinity:

$$\frac{E_X}{f(X|T)} \leq \frac{E_y}{f(y|T)} \qquad (y \neq X, y = 0, 1, 2, ...,) \tag{3.1}$$

where $E_0, E_1, E_2, ...,$ are *iid* with the standard exponential distribution. Here we consider the class of discrete distributions satisfying the following condition.

ASSUMPTION 3.1. *As a function of $T \in \Theta \subseteq \mathcal{R}$, $\ell(T|y, X) \equiv \frac{f(y|T)}{f(X|T)}$ is increasing for all $y > X$ and decreasing for all $y < X$.*

We note that the case where $\ell(T|y, X)$ is decreasing for all $y > X$ and increasing for all $y < X$ can be transformed to satisfy Assumption 3.1 via simple reparameterization. This is illustrated below with the negative binomial distribution. It should be remarked that Assumption 3.1 is satisfied by exponential families. DS inference about discrete distributions satisfying Assumption 3.1 has a simple form, which is given in the following theorem.

THEOREM 3.1. *Suppose that $X \sim F(X|T)$ on the sample space $\mathcal{X} = \{0, 1, 2, ...\}$ with unknown parameter $T \in \Theta \subseteq \mathcal{R}^1$ and that the parameter space $\Theta \subseteq \mathcal{R}$ is an interval. Then under Assumption 3.1, after ruling out conflict cases* *(i) the a-random set for T is a random interval of the form*

$$\left[\max_{y<X} T_y,\ \min_{y>X} T_y\right] \tag{3.2}$$

where $T_y = \max\{T : \ell(T|y,X) \le E_y/E_X, T \in \Theta\}$ for $y > X$ and $T_y = \min\{T : \ell(T|y,X) \le E_y/E_X, T \in \Theta\}$ for $y < X$ with $E_y \overset{iid}{\sim} Exp(1)$ for $y = 0, 1, \ldots,$ and

(ii) the DS posterior has commonality

$$c([a,b]|X) \propto \left[1 + \frac{F(X-1|a)}{f(X|a)} + \frac{1-F(X|b)}{f(X|b)}\right]^{-1} \tag{3.3}$$

where $a \le b$ $(a, b \in \Theta)$.

PROOF. A realization of $E_0, E_1, \ldots$ is conflict *iff* $\max_{y<X} T_y > \min_{y>X} T_y$. It is easy to see that (*i*) follows Assumption 3.1. The commonality is the probability that the a-random interval (3.2) covers the given interval $[a, b]$. Thus, Assumption 3.1 leads to

$$\begin{aligned} c([a,b]|X) &\propto \Pr\left(\ell(a|y,X) \le \frac{E_y}{E_X} \text{ for } y < X, \ell(b|y,X) \le \frac{E_y}{E_X} \text{ for } y > X\right) \\ &\overset{z=E_X}{=} \int_0^\infty e^{-z\left[1+\sum_{y<X}\ell(a|y,X)+\sum_{y>X}\ell(b|y,X)\right]} dz \\ &= \left[1 + \sum_{y<X} \ell(a|y,X) + \sum_{y>X} \ell(b|y,X)\right]^{-1}, \end{aligned}$$

which results in (*ii*) according to the definition $\ell(T|y,X) = \frac{f(y|T)}{f(X|T)}$. □

We conclude this section with two illustrative examples.

EXAMPLE 3.1. (The Poisson distribution) For the Poisson distribution, the pdf is

$$f(x|L) = \frac{L^x}{x!} e^{-L} \qquad (x \in \mathcal{X} = \{0, 1, \ldots\}).$$

For a given single observation X from Poisson(L), the function

$$\ell(L|y,X) = \frac{y!}{X!} L^{y-X}$$

is increasing in L for $y > X$ and decreasing for $y < X$. Thus, the commonality for inference about L is

$$c_X([a,b]) \propto \frac{1}{1 + \frac{\Pr(Y<X|a)}{f(X|a)} + \frac{\Pr(Y>X|b)}{f(X|b)}} \tag{3.4}$$

for $0 \le a \le b < \infty$, where $Y|\lambda \sim \text{Poisson}(\lambda)$.

We show that the lower a-cdf and upper a-cdf are

(3.5)
$$\underline{H}(\lambda|X) = \begin{cases} \text{pG}(2\lambda, 2X) + \frac{2^{2X-1}\Gamma(X+1)\Gamma(X)}{\Gamma(2X)}\text{Pr}(Y < X|\lambda) f(X|\lambda) & \text{if } X > 0; \\ 1 & \text{if } X = 0, \end{cases}$$

and

(3.6) $$\overline{H}(\lambda|X) = \text{pG}(2\lambda, 2X+1) + \frac{2^{2X}\Gamma^2(X+1)}{\Gamma(2X+1)}\text{Pr}(Y > X|\lambda) f(X|\lambda)$$

for $\lambda \in [0, \infty)$, where $\text{pG}(., \alpha)$ denotes the cdf of the Gamma distribution with shape parameter α and $Y|\lambda \sim \text{Poisson}(\lambda)$. The result is obvious for $\underline{H}(\lambda|X=0)$. The lower a-pdf is

$$\underline{h}(\lambda|X) \propto \left.\frac{\partial c_X([a,b])}{\partial a}\right|_{a=b=\lambda} = \text{Pr}(Y < X|\lambda)\frac{\partial f(X|\lambda)}{\partial \lambda} - f(X|\lambda)\frac{\partial \text{Pr}(Y < X|\lambda)}{\partial \lambda}.$$

Note that

$$\frac{\partial \text{Pr}(Y < X|\lambda)}{\partial \lambda} = \frac{\partial \sum_{i=0}^{X-1} \frac{\lambda^i}{i!}e^{-\lambda}}{\partial \lambda} = -f(X-1|\lambda).$$

Hence,

$$\begin{aligned} \underline{H}(\lambda|X) &\propto \text{Pr}(Y < X|\lambda) f(X|\lambda) - 2\int_0^\lambda f(X|u) d\text{Pr}(Y < X|u) \\ &= 2\int_0^\lambda f(X|u) f(X-1|u) du + \text{Pr}(Y < X|\lambda) f(X|\lambda) \\ &= 2\int_0^\lambda \frac{u^{2X-1}}{X!(X-1)!}e^{-2u} du + \text{Pr}(Y < X|\lambda) f(X|\lambda) \\ &= \frac{\Gamma(2X)}{2^{2X-1}\Gamma(X+1)\Gamma(X)} \int_0^{2\lambda} \frac{z^{2X-1}}{\Gamma(2X)} e^{-z} dz + \text{Pr}(Y < X|\lambda) f(X|\lambda). \end{aligned}$$

This completes the proof of (3.5). (3.6) can be proved similarly.

The a-cdfs for various values of X are displayed in Figure 3.1. For a comparison, the a-cdfs based on the Poisson DSM (pDSM) [4] are also shown. The simple test of significance provides an interesting case to compare the two DSMs. Figure 3.2 shows the (p, q, r)s for the simple test of significance: $H_0: L = L_0$ vs. $H_a: L > L_0$. It is interesting that the r-curve, the degree of "don't know", of ADSM dominates the frequentist p-value, providing an attractive solution to resolve the logical disjunction problem in the simple test of significance [4, 5]. On the contrary, the advantage of using "don't know" is disappearing in pDSM [4] for large values of X (or equivalently L).

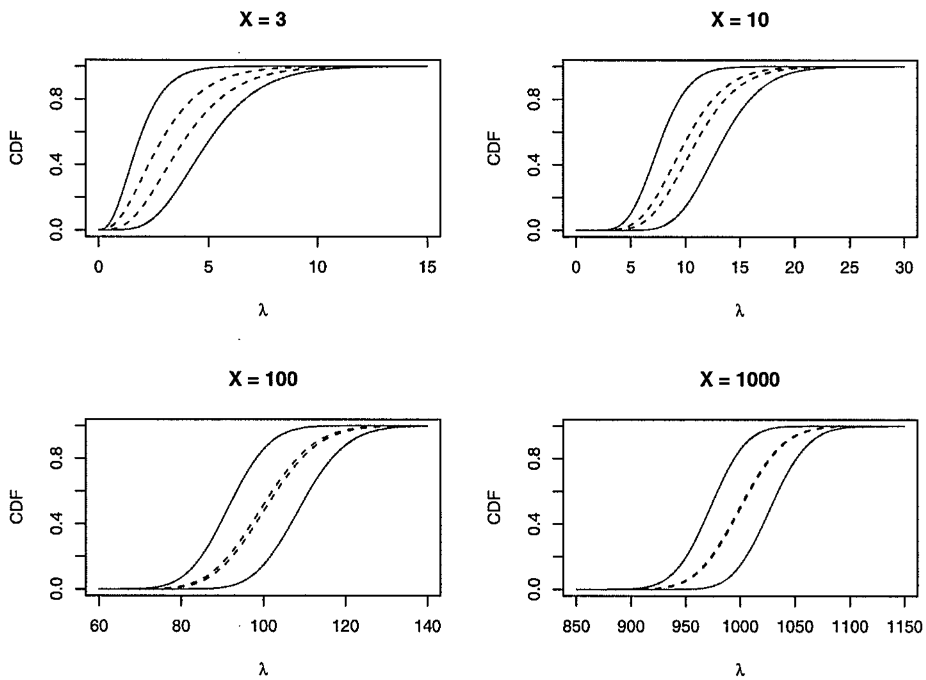

FIG 3.1. *The solid lines are the a-cdfs obtained by ADSM. The dashed lines are the a-cdfs obtained by pDSM [4].*

EXAMPLE 3.2. (The negative binomial distribution) Suppose that X is a single observation from the negative binomial distribution NB (n_0, P) with known size n_0 (> 0) and unknown probability, P $(\in [0,1])$, of success in each trial. That is, X represents the number of failures which occur in a sequence of Bernoulli trials before the target number of successes (n_0) is reached.

The negative binomial distribution NB (n_0, P) has density

$$f_{n_0}(x|P) = \frac{(n_0 - 1 + x)!}{(n_0 - 1)!x!} P^{n_0}(1-P)^x \qquad (x = 0, 1, 2, ...)$$

DSM about $Q = 1 - P$ for given X can be obtained straightforwardly by applying Theorem 3.1. Hence, the DS posterior of P has commonality

$$c_P([a,b]|X) = C_Q([1-b, 1-a]|X) \propto \left[1 + \frac{F_{n_0}(X-1|b)}{f_{n_0}(X|b)} + \frac{1 - F_{n_0}(X|a)}{f_{n_0}(X|a)}\right]^{-1}$$

for $0 \le a \le b \le 1$. The lower and upper a-pdfs can be obtained from (2.4)

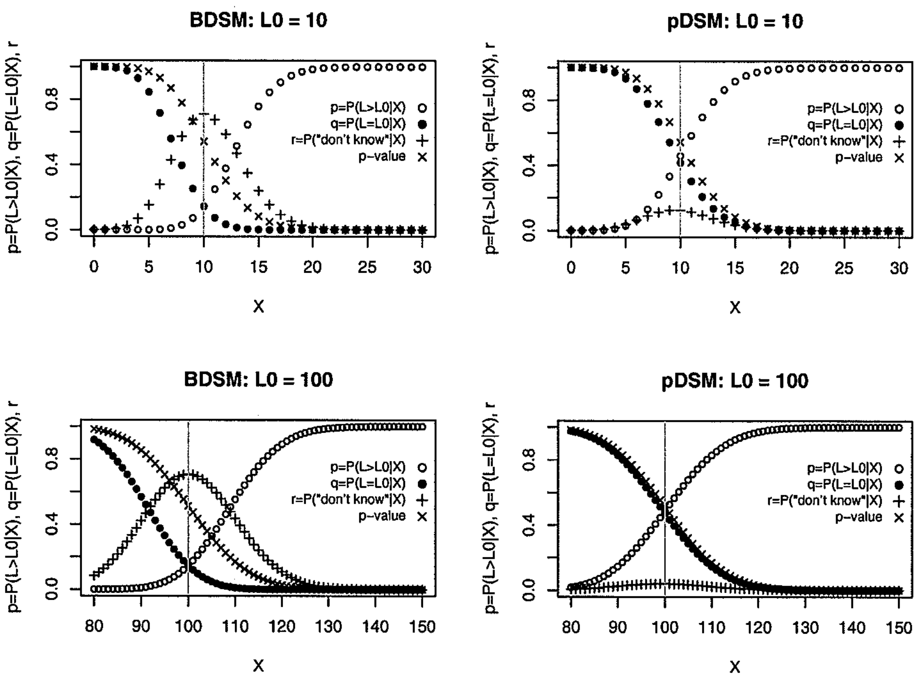

FIG 3.2. *The DS (p, q, r)s for the simple test of significance: $H_0: L = L_0$ vs. $H_a: L > L_0$. The pDSM results were computed based on the a-cdfs in [4].*

and (2.5) as follows:

$$\underline{h}(p|X) \propto [1 - F_{n_0}(X|p)]\frac{\partial f_{n_0}(X|p)}{\partial p} - f_{n_0}(X|p)\frac{\partial[1 - F_{n_0}(X|p)]}{\partial p}$$

and

$$\overline{h}(p|X) \propto f_{n_0}(X|p)\frac{\partial F_{n_0}(X-1|p)}{\partial p} - F_{n_0}(X-1|p)\frac{\partial f_{n_0}(X|p)}{\partial p}$$

for $p \in [0, 1]$. The corresponding lower and upper a-cdfs are

$$\underline{H}(p|X) \propto f_{n_0}(X|p)[1 + F_{n_0}(X|p)] - 2\int_0^p F_{n_0}(X|t)\frac{\partial f_{n_0}(X|t)}{\partial t}dt$$

and

$$\overline{H}(p|X) \propto f_{n_0}(X|p)F_{n_0}(X-1|p) - 2\int_0^p F_{n_0}(X-1|t)\frac{\partial f_{n_0}(X|t)}{\partial t}dt$$

for $p \in [0,1]$. Let

$$C_0 = 1 \quad \text{and} \quad C_x = \sum_{y=0}^{x-1} \frac{(x-y)(n_0-1+y)!}{(n_0-1)!y!} \frac{(2n_0)!(y+x-1)!}{(2n_0+y+x)!} \qquad (x>0)$$

Then $\overline{H}(p=1|X)=1$,

$$\overline{H}(p|X) = \frac{1}{C_X} \sum_{y=0}^{X-1} \frac{(X-y)(n_0-1+y)!}{(n_0-1)!y!} \frac{(2n_0)!(y+X-1)!}{(2n_0+y+X)!} \mathrm{pB}\,(p, 2n_0+1, y+X)$$

for $p \in [0,1)$, and

$$\underline{H}(p|X) = \frac{1}{C_X} p^{n_0}(1-p)^X + \overline{H}(p|X)$$

for $p \in [0,1]$, where $\mathrm{pB}\,(., \alpha, \beta)$ denotes the cdf of the Beta distribution Beta(α,β). Numerical results for $n_0 = 10$ and $X = 30$ is shown in Figure 3.3.

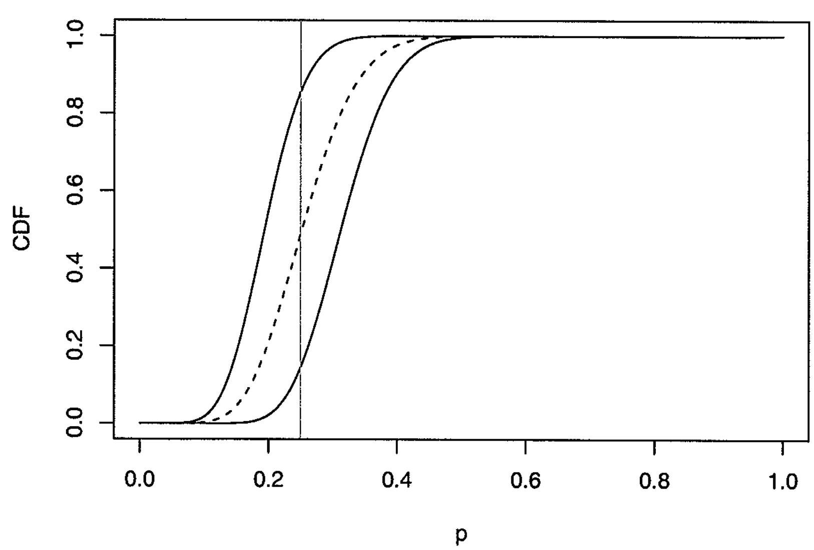

FIG 3.3. *The two a-cdfs for inference about P (solid curves) in $NB(n_0 = 10, P)$ for a given single observation $X = 30$, the Bayesian posterior (dashed curve) with the non-informative prior Beta$(0.5, 0.5)$, and the maximum likelihood estimate of P (vertical line).*

4. Continuous distributions. Suppose that a single observation X was taken from $F(x|T)$, a univariate continuous distribution on the sample space $\mathcal{X} = \mathcal{R}$ with the parameter $T \in \Theta \subset \mathcal{R}$. Following Dempster's (1966) suggestion, we partition the sample space $\mathcal{X} = \mathcal{R}$ into a sequence of small intervals $\mathcal{R} = \cup_{i=-\infty}^{\infty}[x_i, x_{i+1})$. Let $Z = i$ if $X \in [x_i, x_{i+1})$. Then

$$\Pr(Z = i|T) = F(x_{i+1}|T) - F(x_i|T) \qquad (i = \dots, -1, 0, 1, \dots)$$

Thus, according to ADSM(2.3) DS inference about T based on $Z = i$ is obtained by

$$\frac{E_i}{F(x_{i+1}|T) - F(x_i|T)} \le \frac{E_j}{F(x_{j+1}|T) - F(x_j|T)} \qquad (j \neq i) \tag{4.1}$$

where $E_j \overset{iid}{\sim} \text{Exp}(1)$ for $j = 0, \pm 1, \dots$, subject to the non-conflict constraint. DS inference about T based on X itself is obtained by letting $x_{j+1} - x_i$ go to zero for all j. The following theorem provides a useful tool for DS inference on a class of continuous distributions.

THEOREM 4.1. *Suppose that $F(x|T)$ is a univariate continuous distribution on $\mathcal{X} = \mathcal{R}$ with density $f(x|T)$ and parameter $T \in \Theta \subset \mathcal{R}$, where Θ is an interval. If $f(x|T)$ as a function of x is continuous for any $T \in \Theta$ and the function $\ell(T|y,x) = \frac{f(y|T)}{f(x|T)}$ is increasing in T for any $y > x$ and decreasing for any $y < x$, then the a-random sets of DS inference about T for given X are intervals and the commonality is*

$$c([a,b]|X) \propto \left[\frac{F(X|a)}{f(X|a)} + \frac{1 - F(X|b)}{f(X|b)}\right]^{-1} \qquad (a \le b) \tag{4.2}$$

PROOF. We note that a proof can be established by making use of Eqn. (5.18) of Dempster (1966). Here we provide a proof using the a-random variables in ADSM. First, we show that as a function of T

$$L(T|x_i, x_{i+1}, x_j, x_{j+1}) \equiv \frac{F(x_{j+1}|T) - F(x_j|T)}{F(x_{i+1}|T) - F(x_i|T)} \qquad (T \in \Theta) \tag{4.3}$$

is increasing, where $x_i < x_{i+1} \le x_j < x_{j+1}$. From the assumptions we have for $T_2 > T_1 \in \Theta$

$$\begin{aligned}
\frac{F(x_{j+1}|T_2) - F(x_j|T_2)}{F(x_{i+1}|T_2) - F(x_i|T_2)} &= \lim_{n\to\infty} \frac{(x_{j+1} - x_j)\sum_{k=0}^{n-1} f(x_j + k[x_{j+1} - x_j]/n|T_2)}{(x_{i+1} - x_i)\sum_{k=0}^{n-1} f(x_i + k[x_{i+1} - x_i]/n|T_2)} \\
&\ge \lim_{n\to\infty} \frac{(x_{j+1} - x_j)\sum_{k=0}^{n-1} f(x_j + k[x_{j+1} - x_j]/n|T_1)}{(x_{i+1} - x_i)\sum_{k=0}^{n-1} f(x_i + k[x_{i+1} - x_i]/n|T_1)} \\
&= \frac{F(x_{j+1}|T_1) - F(x_j|T_2)}{F(x_{i+1}|T_1) - F(x_i|T_1)}.
\end{aligned}$$

Similarly, we can show that

$$L(T|x_i, x_{i+1}, x_j, x_{j+1}) \equiv \frac{F(x_{j+1}|T) - F(x_j|T)}{F(x_{i+1}|T) - F(x_i|T)} \qquad (T \in \Theta) \tag{4.4}$$

is decreasing in T where $x_j < x_{j+1} \leq x_i < x_{i+1}$. For any partition of $\mathcal{X}$ described above, denote by δ_X the length of interval $[x_i, x_{i+1})$ containing X. For given E_i, according to (4.3) and (4.4) the random set (4.1) is

$$(-\infty, T_j(E_i, E_j)] \cap \Theta \qquad (j > i)$$

and

$$[T_j(E_i, E_j), \infty) \cap \Theta \qquad (j < i)$$

where $T_j(E_i, E_j)$ is given by

$$\frac{E_i}{f(X|T_j(E_i, E_j))} = \frac{E_j}{f(x_j|T_j(E_i, E_j))}.$$

Thus the a-random set is the interval

$$[\max_{j<i} T_j(E_i, E_j),\ \min_{j>i} T_j(E_i, E_j)],$$

subject to the non-conflict constraint $\max_{j<i} T_j(E_i, E_j) \leq \min_{j>i} T_j(E_i, E_j)$.

For any interval $[a, b]$, $a \leq b \in \Theta$, the commonality is

$$\begin{aligned}
c([a,b]|X) &\propto \lim \Pr\left(\max_{j<i} T_j(E_i, E_j) \leq a, \min_{j>i} T_j(E_i, E_j) \geq b\right) \\
&= \lim \mathrm{p}\left(E_j \geq \frac{F(x_{j+1}|b) - F(x_j|b)}{F(x_{i+1}|b) - F(x_i|b)} E_i \text{ for } j > i \text{ and} \right. \\
&\qquad \left. E_j \geq \frac{F(x_{j+1}|a) - F(x_j|a)}{F(x_{i+1}|a) - F(x_i|a)} E_i \text{ for } j < i\right) \\
&\overset{z=E_i}{=} \lim \int_0^\infty e^{-z\left[1 + \frac{F(x_i|a)}{F(x_{i+1}|a) - F(x_i|a)} + \frac{1 - F(x_{i+1}|b)}{F(x_{i+1}|b) - F(x_i|b)}\right]} dz \\
&= \lim \delta_X \left[\delta_X + \frac{F(X|a)}{f(X|a)} + \frac{1 - F(X|b)}{f(X|b)}\right]^{-1} \\
&\propto \lim \left[\delta_X + \frac{F(X|a)}{f(X|a)} + \frac{1 - F(X|b)}{f(X|b)}\right]^{-1} \\
&= \left[\frac{F(X|a)}{f(X|a)} + \frac{1 - F(X|b)}{f(X|b)}\right]^{-1}
\end{aligned}$$

as $x_{k+1} - x_k \to 0$ for all $k = 0, \pm 1, \ldots$. Hence, Theorem 4.1 is proved. □

Applications of Theorem 4.1 are illustrated with following two examples.

EXAMPLE 4.1. (The normal distribution) Suppose that $X \sim N(M, 1)$ with unknown $M \in (-\infty, \infty)$. It is easy to verify that the density function

$$\phi(x|M) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2}} \qquad x \in (-\infty, \infty)$$

satisfies the condition of Theorem 4.1. Thus the commonality of the DSM for M for given X is

$$c([a,b]|X) \propto \left[\frac{\Phi(X-a)}{\phi(X-a)} + \frac{1-\Phi(X-b)}{\phi(X-b)}\right]^{-1} \qquad (-\infty < a \leq b < \infty)$$

where $\Phi(.)$ stands for the cdf of the standard normal distribution. The lower a-pdf and a-cdf are obtained as follows:

$$\underline{h}(\mu|X) = \sqrt{\pi}\phi(\mu - X)\left[\phi(\mu - X) - (\mu - X)\Phi(X - \mu)\right]$$

and

$$\underline{H}(\mu|X) = \Phi(\sqrt{2}[\mu - X]) + \sqrt{\pi}\Phi(X - \mu)\phi(\mu - X)$$

for $\mu \in (-\infty, \infty)$. Similarly, the upper a-pdf and a-cdf are given by

$$\overline{h}(\mu|X) = \sqrt{\pi}\phi(\mu - X)\left[\phi(\mu - X) + (\mu - X)\Phi(\mu - X)\right]$$

and

$$\overline{H}(\mu|X) = \Phi(\sqrt{2}[\mu - X]) - \sqrt{\pi}\Phi(\mu - X)\phi(\mu - X).$$

Numerical results for given $X = 0$ are shown in Figure 4.1, where the Bayesian posterior with the commonly used flat prior is also shown for a comparison. As for the Poisson model in Section 3, ADSMfor $N(M, 1)$ resolves nicely the logical disjunction in the simple test of significance such as $H_0 : M = M_0$ vs. $H_a : M > M_0$. Here we omit the details.

EXAMPLE 4.2. (The gamma distribution) Suppose that $X \sim \Gamma(\alpha, B)$ with known shape parameter α and unknown rate (1/scale) parameter B. The density function of the Gamma distribution $\Gamma(\alpha, B)$ is

$$f(x|B) = \frac{B^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-Bx} \qquad (x > 0; B > 0)$$

Let $S = 1/B$. Then

$$\ell(S|y,x) \equiv \frac{f(y|B = 1/S)}{f(x|B = 1/S)} \propto e^{-\frac{y-x}{S}}$$

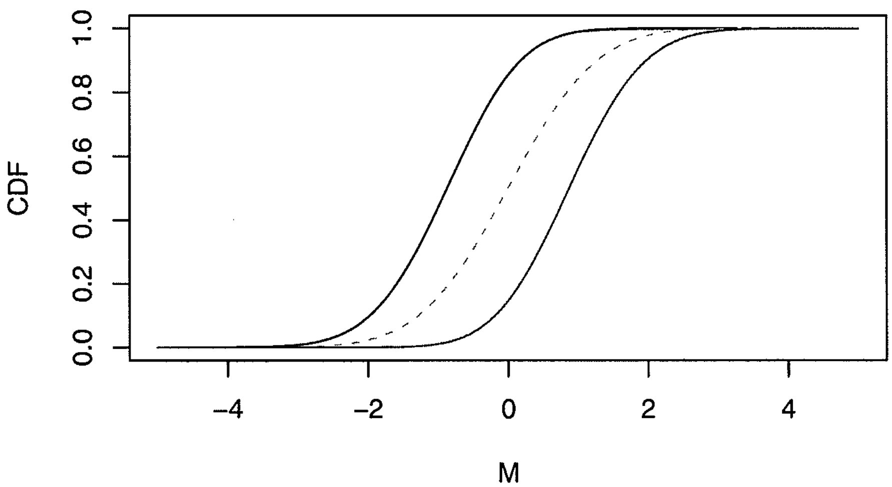

FIG 4.1. *The a-cdfs of DSM for $N(M, 1)$ for given $X = 0$. The solid lines are the a-cdfs obtained by ADSM. The dashed line is the fiducial or the Bayesian posterior corresponding to the flat prior on M.*

is increase in S for $y > x$ and decreasing in S for $y < x$. Thus, Theorem 4.1 can be applied directly to making DS inference about S for given X. This yields

$$c_B([a,b]|X) \propto c_S([1/b, 1/a]|X) \propto \left[\frac{F(X|b)}{f(X|b)} + \frac{1 - F(X|a)}{f(X|a)}\right]^{-1}$$

where $F(.|\beta) = \text{pG}\,(., \alpha, \beta)$ denotes the cdf of the Gamma distribution Gamma(α, β) with shape α and rate β. Furthermore, the lower and upper a-cdfs are

$$\underline{H}(\beta|X) \propto [1 - F(X|\beta)]f(X|\beta) + \frac{\Gamma(2\alpha)}{2^{2\alpha-1}X\Gamma^2(\alpha)}\text{pG}\,(2X\beta, 2\alpha)$$

and

$$\underline{H}(\beta|X) \propto -F(X|\beta)f(X|\beta) + \frac{\Gamma(2\alpha)}{2^{2\alpha-1}X\Gamma^2(\alpha)}\text{pG}\,(2X\beta, 2\alpha)\,,$$

where $\text{pG}\,(., \alpha)$ denotes the cdf of the Gamma distribution Gamma$(\alpha, 1)$.

5. Estimation based on a sample. Consider a discrete univariate distribution satisfying the conditions in Theorem 3.1. Suppose that $\{X_1, ..., X_n\}$ is a sample of size n from a univariate discrete distribution $F(x|T)$. Applying Dempster's rule of combination leads to the DSM for T. That is, the DSM has commonality

$$c([a,b]|X_1,...,X_n) \propto \prod_{i=1}^{n} \left[1 + \frac{F(X_i - 1|a)}{f(X_i|a)} + \frac{1 - F(X_i|b)}{f(X_i|b)}\right]^{-1} \tag{5.1}$$

where $a \leq b$. Let

$$\ell(\theta|X_1, ..., X_n) = \prod_{i=1}^{n} f(X_i|\theta)$$

be the likelihood function of $T = \theta$ given $X_1, ..., X_n$. Then, the two a-pdfs are

$$\begin{aligned} &\underline{h}(\theta|X_1,...,X_n) \\ \propto\ &\ell(\theta|X_1,...,X_n) \sum_{i=1}^{n} \left[F(X_i - 1|\theta)\frac{\partial \ln f(X_i|\theta)}{\partial\theta} - \frac{\partial F(X_i - 1|\theta)}{\partial\theta}\right] \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} &\overline{h}(\theta|X_1,...,X_n) \\ \propto\ &\ell(\theta|X_1,...,X_n) \sum_{i=1}^{n} \left[\frac{\partial(1 - F(X_i|\theta))}{\partial\theta} - (1 - F(X_i|\theta))\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right] \end{aligned} \tag{5.3}$$

For a univariate continuous distribution $F(x|T)$, the results are similar to (5.1), (5.2), and (5.3). More specifically, suppose that $F(x|T)$ satisfies the conditions of Proposition 2, we have

$$c([a,b]|X_1,...,X_n) \propto \prod_{i=1}^{n} \left[\frac{F(X_i|a)}{f(X_i|a)} + \frac{1 - F(X_i|b)}{f(X_i|b)}\right]^{-1},$$

$$\begin{aligned} &\underline{h}(\theta|X_1,...,X_n) \\ \propto\ &\ell(\theta|X_1,...,X_n) \sum_{i=1}^{n} \left[F(X_i|\theta)\frac{\partial \ln f(X_i|\theta)}{\partial\theta} - \frac{\partial F(X_i|\theta)}{\partial\theta}\right], \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} &\overline{h}(\theta|X_1,...,X_n) \\ \propto\ &\ell(\theta|X_1,...,X_n) \sum_{i=1}^{n} \left[\frac{\partial(1 - F(X_i|\theta))}{\partial\theta} - (1 - F(X_i|\theta))\frac{\partial \ln f(X_i|\theta)}{\partial\theta}\right], \end{aligned} \tag{5.5}$$

where $\ell(\theta|X_1, ..., X_n) = \prod_{i=1}^n f(X_i|\theta)$ denotes the be the likelihood function of $T = \theta$ given the sample $X_1, ..., X_n$ from $F(x|T)$. This is illustrated by the following example with the univariate normal distribution $N(M, 1)$.

EXAMPLE 5.1. (The normal distribution) Suppose that a sample of n observations, $X_1, ..., X_n$ is considered from $N(M, 1)$. Let $\ell(\mu|X_1, ..., X_n)$ denote the likelihood, *i.e.*,

$$\ell(\mu|X_1, ..., X_n) = \prod_{i=1}^n \phi(X_i - \mu) \qquad (-\infty < \mu < \infty)$$

Then routine algebraic operations lead to that the combined DSM for inference about M from general case with n observations has

(i) commonality

$$c([a, b]|X_1, ..., X_n) \propto \prod_{i=1}^n \left[\frac{\Phi(X_i - a)}{\phi(X_i - a)} + \frac{1 - \Phi(X_i - b)}{\phi(X_i - b)}\right]^{-1} \tag{5.6}$$

for $-\infty < a \leq b < \infty$, which for singletons $a = b = \mu$ is proportional to the likelihood $\ell(\mu|X_1, ..., X_n)$,

(ii) two marginal pdfs for the a-random lower and end variables

$$\begin{aligned} &\underline{h}(\mu|X_1, ..., X_n) \\ \propto\ &\frac{\sum_{i=1}^n [\phi(X_i - \mu) + (X_i - \mu)\Phi(X_i - \mu)]}{n} \cdot \ell(\mu|X_1, ..., X_n) \end{aligned} \tag{5.7}$$

and

$$\begin{aligned} &\overline{h}(\mu|X_1, ..., X_n) \\ \propto\ &\frac{\sum_{i=1}^n [\phi(X_i - \mu) + (\mu - X_i)\Phi(\mu - X_i)]}{n} \cdot L(\mu|X_1, ..., X_n), \end{aligned} \tag{5.8}$$

and

(iii) the two marginal cdfs for the a-random lower and end variables

$$\begin{aligned} \underline{H}(\mu|X_1, ..., X_n) &= \frac{C_n}{n} \sum_{i=1}^n \left[\int_{-\infty}^{\mu} (X_i - u)\Phi(X_i - u)\phi(\sqrt{n}(u - \bar{X}))du \right. \\ &\quad \left. + \int_{-\infty}^{\mu} \phi(u - x_i)\phi(\sqrt{n}(u - \bar{X}))du\right] \end{aligned}$$

and

$$\begin{aligned} \overline{H}(\mu|X_1, ..., X_n) &= \frac{C_n}{n} \sum_{i=1}^n \left[\int_{-\infty}^{\mu} (u - X_i)\Phi(u - X_i)\phi(\sqrt{n}(u - \bar{X}))du \right. \\ &\quad \left. + \int_{-\infty}^{\mu} \phi(u - X_i)\phi(\sqrt{n}(u - \bar{X}))du\right] \end{aligned}$$

where $\bar{X}$ denotes the sample mean $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and C_n is the normalizing constant.

Here we take a look at the terms

$$(5.9)\quad \phi(X_i-\mu)+(X_i-\mu)\Phi(X_i-\mu) \quad \text{and} \quad \phi(X_i-\mu)+(\mu-X_i)\Phi(\mu-X_i)$$

in (5.7) and (5.8). The lower data-adaptive "prior"

$$\underline{\pi}(\mu|X_1,...,X_n) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\phi(X_i-\mu)+(X_i-\mu)\Phi(X_i-\mu)\right]$$

is heavily skewed to the left and the upper "prior"

$$\overline{\pi}(\mu|X_1,...,X_n) \equiv \frac{1}{n}\sum_{i=1}^{n}\left[\phi(X_i-\mu)+(\mu-X_i)\Phi(\mu-X_i)\right]$$

to the right. As a result, the two marginal a-cdfs are *extreme-value* "skew-" normal distributions. We note that

$$\begin{aligned}&\int_{\mu}^{\infty}\left[\phi(X_i-u)+(X_i-u)\Phi(X_i-u)\right]du\\ =\;& \Phi(X_i-\mu)+(X_i-\mu)^2\Phi(X_i-\mu)-\int_{-\infty}^{X_i-\mu}u^2\phi(u)du\end{aligned}$$

goes to ∞ as $\mu \to -\infty$ and that

$$\int_{-\infty}^{\mu}\left[\phi(X_i-u)+(u-X_i)\Phi(u-X_i)\right]du$$

goes to ∞ as $\mu \to \infty$. Incidentally, we also note that $\pi(\mu|X_1,...,X_n) \equiv= \underline{\pi}(\mu|X_1,...,X_n) + \overline{\pi}(\mu|X_1,...,X_n)$ could be served as a data-adaptive prior for Bayesian inference.

6. Asymptotic results. The main asymptotic results on DS posteriors are (i) like Bayesian inference, DS inference is efficient in the sense that a-cdfs are asymptotically normal and converge in the order of $n^{-1/2}$, and (ii) the probability of "don't know" about the assertion $\{T \le T_0\}$ with fixed $T_0 \in \Theta$ converges to zero in the order of $n^{-1/2}$ if the maximum likelihood estimate of T is close to T_0 and goes away exponentially otherwise.

To save space, here we consider the limiting case of (5.4). Corresponding results can be obtained similarly for (5.2), and (5.3), and (5.5). We write (5.4) as

$$(6.1)\qquad \underline{h}(\theta|X_1,...,X_n) \propto \ell(\theta|X_1,...,X_n)\pi(\theta|X_1,...,X_n),$$

where

$$\pi(\theta|X_1,...,X_n) \equiv \sum_{i=1}^{n}\left[F(X_i|\theta)\frac{\partial \ln f(X_i|\theta)}{\partial\theta} - \frac{\partial F(X_i|\theta)}{\partial\theta}\right] \tag{6.2}$$

In the analogy with the Bayes theory, the factor $\pi(\theta|X_1,...,X_n)$ in (6.1) plays the role of the prior distribution and $\underline{h}(\theta|X_1,...,X_n)$ the posterior. Denote by

$$L(\theta|X_1,...,X_n) = \ln \ell(\theta|X_1,...,X_n) \qquad (\theta \in \Theta)$$

the log-likelihood function of T given $X_1,...,X_n$. Let

$$\pi(\theta|T) \equiv \int_{-\infty}^{\infty}\left[F(x|\theta)\frac{\partial \ln f(x|\theta)}{\partial\theta} - \frac{\partial F(x|\theta)}{\partial\theta}\right] f(x|T)dx \tag{6.3}$$

for $T,\theta \in \Theta$. We require here the following assumption.

ASSUMPTION 6.1. *For any fixed $T \in \Theta$, $\pi(\theta|T)$ defined in (6.3) exists (i.e., $\pi(\theta|T) < \infty$) and is positive and continuous for all $\theta \in \Theta$.*

We prove that under mild "regularity" conditions for large values of n the a-posterior distribution $\underline{h}(\theta|X_1,...,X_n)$ is approximately normal with

$$\text{mean} = T + \frac{1}{nI(T)}\frac{\partial L(T|X_1,...,X_n)}{\partial T} \quad \text{and} \quad \text{variance} = \frac{1}{nI(T)}, \tag{6.4}$$

where $I(T)$ stands for the Fisher information

$$I(T) \equiv \int_{-\infty}^{\infty}\left[\frac{\partial \ln f(x|T)}{\partial T}\right]^2 f(x|T)dx = -\int_{-\infty}^{\infty}\frac{\partial^2 \ln f(x|T)}{\partial T^2} f(x|T)dx.$$

First, we restate the well-known asymptotic result on Bayesian posterior distributions in, for example, [6]:

THEOREM 6.1. *Let $\pi^*(t|X_1,...,X_n)$ be the posterior distribution of $\sqrt{n}(\theta - \theta_n)$ with the "prior" $\pi(\theta|T)$ satisfying Assumption 6.1, where*

$$T_n \equiv T + \frac{1}{nI(T)}\frac{\partial L(T|X_1,...,X_n)}{\partial T}. \tag{6.5}$$

If the regularity conditions (B1)-(B3) of Lehmann ([6], pp. 454-455) hold for the likelihood function $\ell(\theta|X_1,...,X_n)$, then

$$\int_{-\infty}^{\infty}\left|\pi^*(t|X_1,...,X_n) - \sqrt{I(T)}\phi\left(t\sqrt{I(T)}\right)\right| dt \xrightarrow{p} 0. \tag{6.6}$$

We now extend Theorem 6.1 to the following corresponding result for DS posteriors:

THEOREM 6.2. *Suppose that Assumption 2 for $\pi(\theta|T)$ defined in (6.3) holds. If the regularity conditions (B1)-(B3) of Lehmann ([6], pp. 454-455) for the likelihood function $\ell(\theta|X_1, ..., X_n)$ hold, then*

$$\int_{-\infty}^{\infty} \left| \underline{h}(\theta|X_1, ..., X_n) - \sqrt{nI(T)}\phi\left(\sqrt{nI(T)}[\theta - T_n]\right) \right| d\theta \xrightarrow{p} 0 \tag{6.7}$$

where T_n is given in (6.5).

PROOF. From the inequality

$$\begin{aligned} &\int_{-\infty}^{\infty} \left| \underline{h}(\theta|X_1, ..., X_n) - \sqrt{nI(T)}\phi\left(\sqrt{nI(T)}[\theta - T_n]\right) \right| d\theta \\ \leq &\int_{-\infty}^{\infty} \left| \underline{h}(\theta|X_1, ..., X_n) - \sqrt{n}\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n) \right| d\theta \\ &+ \int_{-\infty}^{\infty} \left| \sqrt{n}\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n) - \sqrt{nI(T)}\phi\left(\sqrt{nI(T)}[\theta - T_n]\right) \right| d\theta \end{aligned}$$

and that the last integral converges to zero in probability (Theorem 6.1), it remains to show that

$$\int_{-\infty}^{\infty} \left| \underline{h}(\theta|X_1, ..., X_n) - \sqrt{n}\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n) \right| d\theta \xrightarrow{p} 0 \tag{6.8}$$

where $\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n)$ is defined in Theorem 6.1. Now

$$\begin{aligned} &\int_{-\infty}^{\infty} \left| \underline{h}(\theta|X_1, ..., X_n) - \sqrt{n}\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n) \right| d\theta \\ \leq &\int_{-\infty}^{\infty} \underline{h}(\theta|X_1, ..., X_n) d\theta + \int_{-\infty}^{\infty} \sqrt{n}\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n) d\theta = 2 \end{aligned}$$

and

$$\pi(\theta|X_1, ..., X_n) \xrightarrow{wp1} \pi(\theta|M).$$

Thus

$$\int_{-\infty}^{\infty} \left| \underline{h}(\theta|X_1, ..., X_n) - \sqrt{n}\pi^*(\sqrt{n}[\theta - T_n]|X_1, ..., X_n) \right| d\theta \xrightarrow{wp1} 0$$

and hence (6.8) is proved. □

Compared to Bayesian inference, the extra component r for "don't know" about any assertion is of interest. Here we consider the assertion $\{T \leq T_0\}$ for fixed $T_0 \in \Theta$. Under the assumptions of Theorem 6.2, the following result based on (5.4) and (5.5) follows. Let $\hat{T}_n$ denote the maximum likelihood estimate of T given $X_1, ..., X_n$. Then for large values of n, we have

$$\begin{aligned} r_n(\{T \leq T_0\}) &\equiv \int_{-\infty}^{T_0} \left[\underline{h}(\theta|X_1, ..., X_n) - \overline{h}(\theta|X_1, ..., X_n)\right] d\theta \\ &\doteq \frac{\int_{-\infty}^{T_0} \ell(\theta|X_1, ..., X_n) \sum_{i=1}^{n} \frac{\partial \ln f(X_i|\theta)}{\partial \theta} d\theta}{n \int_{-\infty}^{\infty} \ell(\theta|X_1, ..., X_n)\pi(\theta|T)d\theta} \\ &= \frac{\ell(T_0|X_1, ..., X_n)}{n \int_{-\infty}^{\infty} \ell(\theta|X_1, ..., X_n)\pi(\theta|T)d\theta}. \end{aligned}$$

Thus,

$$\begin{aligned} \max_{T_0 \in \Theta} r_n(\{T \leq T_0\}) &= \frac{\ell(\hat{T}_n|X_1, ..., X_n)}{n \int_{-\infty}^{\infty} \ell(\theta|X_1, ..., X_n)\pi(\theta|T)d\theta} \\ &= \frac{e^{\omega(\hat{t}_n)}}{n^{1/2} \int_{-\infty}^{\infty} e^{\omega(t)} \pi(T_n + n^{-1/2}t|T)dt}, \end{aligned}$$

where $\hat{t}_n = \sqrt{n}(\hat{T}_n - T_n)$ and

$$\begin{aligned} \omega(t) &= L(T_n + n^{-1/2}t|X_1, ..., X_n) - L(T|X_1, ..., X_n) \\ &\quad - \frac{1}{2nI(T)} \left[\frac{\partial L(T|X_1, ..., X_n)}{\partial T}\right]^2 \end{aligned}$$

with

$$L(\theta|X_1, ..., X_n) = \ln \ell(\theta|X_1, ..., X_n)$$

and T_n defined in (6.5). According to Eqn. (14) of Lehmann ([6], p. 456), for large values of n we have

$$\int_{-\infty}^{\infty} e^{\omega(t)} \pi(T_n + n^{-1/2}t|T)dt \doteq \pi(T|T)\sqrt{2\pi/I(T)}.$$

in probability and hence

$$\max_{T_0 \in \Theta} r_n(\{T \leq T_0\}) \doteq \frac{\sqrt{I(T)} e^{\omega(\hat{t}_n)}}{\pi(T|T)\sqrt{2\pi n}}.$$

From Lehmann ([6], pp.462-463), the function $\omega(t)$ can be written as

$$\omega(t) = -I(T)\frac{t^2}{2} - \frac{1}{2n} R_n\left(T_n + n^{-1/2}t\right)\left[t + \frac{1}{\sqrt{n}I(T)} \frac{\partial L(T|X_1, ..., X_n)}{\partial T}\right]^2$$

where

$$\sup\left\{\left|\frac{1}{n}R_n\left(T_n+n^{-1/2}t\right)\right|\left[t+\frac{1}{\sqrt{n}I(T)}\frac{\partial L(T|X_1,...,X_n)}{\partial T}\right]^2\right\}\xrightarrow{p}0$$

and

$$\frac{1}{I(T)\sqrt{n}}\frac{\partial L(T|X_1,...,X_n)}{\partial T}\quad\text{is bounded in probability.}$$

Thus,

$$\hat{t}_n=\sqrt{n}(\hat{T}_n-T)-\frac{1}{I(T)\sqrt{n}}\frac{\partial L(T|X_1,...,X_n)}{\partial T}$$

where $\sqrt{n}(\hat{T}_n-T)\dot{\sim}N(0,I^{-1}(T))$. It is easy to see that the above derivation applies to the general case with $T_0\in\Theta$. Thus, $r_n(T_0)$ *is expected to converge in the order of* $1/\sqrt{n}$ *for* T_0 *near the unknown true value* T *and exponentially otherwise.* Stronger results for $N(M,1)$ is obtained as follows.

THEOREM 6.3. *Suppose that* $X_1,...,X_n$ *are iid with* $N(M,1)$. *Let* $r_n(M_0)$ *be the* $r-$ *component of the DS output* (p,q,r) *about the assertion* $\{M\le M_0\}$ *for fixed* $M_0\in(-\infty,\infty)$. *Then*

(*i*) $\max_{M_0} r_n(M_0)$ *converges with probability one to zero in the order* $n^{-1/2}$; *and*

(*ii*) $\sqrt{n}r_n(M_0)$ *converges in distribution to* $\frac{1}{\sqrt{2}}e^{-\frac{\chi_1^2}{2}}$ *for* $M_0=M$ *and* $\sqrt{n}e^{\frac{(M_0-M)^2}{2}}r_n(M_0)\doteq\frac{1}{\sqrt{2}}$ *with probability one, where* χ_1^2 *is the chi-square random variable with one degree of freedom.*

PROOF. We write

$$\begin{aligned}\underline{H}_n(\mu) &= \frac{C}{n}\sum_{i=1}^n\left[\int_{-\infty}^{\mu}(X_i-u)\Phi(X_i-u)\phi(\sqrt{n}(u-\bar{X}))du\right.\\ &\qquad\left.+\int_{-\infty}^{\mu}\phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du\right]\end{aligned}$$

and

$$\begin{aligned}\overline{H}_n(\mu) &= \frac{C}{n}\sum_{i=1}^n\left[\int_{-\infty}^{\mu}(u-X_i)\Phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du\right.\\ &\qquad\left.+\int_{-\infty}^{\mu}\phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du\right],\end{aligned}$$

where C is the normalizing constant. Thus,

$$
\begin{aligned}
r_n(M_0) = \underline{H}_n(M_0) - \overline{H}_n(M_0) &= C\int_{-\infty}^{M_0}(\bar{X}-u)\phi(\sqrt{n}(u-\bar{X}))du \\
&= \frac{C}{n}\int_{-\infty}^{\sqrt{n}(M_0-\bar{X})}(-u)\phi(u)du = \frac{C}{n}\int_{-\infty}^{\sqrt{n}(M_0-\bar{X})} d\phi(u) \\
&= \frac{C}{n}\phi\left(\sqrt{n}(M_0-\bar{X})\right) \\
&= \frac{n^{-1}\phi\left(\sqrt{n}(M_0-\bar{X})\right)}{\frac{1}{n}\sum_{i=1}^{n}\left[\int\phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du + \int(u-X_i)\Phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du\right]}.
\end{aligned}
$$

Note that M denotes the (unknown) true mean. For the second integral in the denominator, we have

$$
\begin{aligned}
&\int_{-\infty}^{\infty}(u-X_i)\Phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du \\
= &\frac{1}{n}\int_{-\infty}^{\infty}\phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du + (\bar{X}-X_i)\int_{-\infty}^{\infty}\Phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du
\end{aligned}
$$

and thereby

$$
r_n(M_0) \doteq \frac{n^{-1}\phi\left(\sqrt{n}(M_0-\bar{X})\right)}{\frac{1}{n}\sum_{i=1}^{n}\left[\int\phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du + (\bar{X}-X_i)\int\Phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du\right]}.
$$

Simple algebraic operations lead to the following expression for the first integral in the denominator of the above formula:

$$
\int_{-\infty}^{\infty}\phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du = \frac{1}{\sqrt{2\pi(n+1)}}e^{-\frac{n(X_i-\bar{X})^2}{2(n+1)}}.
$$

For the second integral in the denominator, we have the following Taylor expansion for $\Phi(u-X_i)$ at $u=\bar{X}$:

$$
\Phi(u-X_i) = \Phi(\bar{X}-X_i) + \phi(\bar{X}-X_i)(u-\bar{X}) + \frac{-(\xi-X_i)\phi(\xi-X_i)}{2}(u-\bar{X})^2,
$$

where ξ is between $\bar{X}$ and u. Note that $|(\xi-X_i)\phi(\xi-X_i)|$ is bounded for all $\xi\in(-\infty,\infty)$. It follows that

$$
(\bar{X}-X_i)\int_{-\infty}^{\infty}\Phi(u-X_i)\phi(\sqrt{n}(u-\bar{X}))du = \frac{(\bar{X}-X_i)\Phi(\bar{X}-X_i)}{\sqrt{n}} + O(n^{-3/2}).
$$

Hence, we can write

$$
r_n(M_0) \doteq \frac{n^{-1}\phi\left(\sqrt{n}(M_0-\bar{X})\right)}{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi(n+1)}}e^{-\frac{n(X_i-\bar{X})^2}{2(n+1)}} + \frac{(\bar{X}-X_i)\Phi(\bar{X}-X_i)}{\sqrt{n}} + O(n^{-3/2})\right]}.
$$

Making use of the identity

$$\sum_{i=1}^{n}(\bar{X}-X_i)\Phi(\bar{X}-X_i) \;=\; \sum_{i=1}^{n}(X_i-\bar{X})\Phi(X_i-\bar{X})$$

we have

$$\begin{aligned}\max_{M_0}\sqrt{n}r_n(M_0) &= \sqrt{n}r_n(\bar{X}) \\ &\doteq \frac{1}{\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{n(X_i-\bar{X})^2}{2(n+1)}} + \frac{\sqrt{2\pi}}{n}\sum_{i=1}^{n}(X_i-\bar{X})\Phi(X_i-\bar{X})}\end{aligned}$$

Let $Z_i = X_i - M$ for $i = 1, ..., n$. Note that $\bar{Z} \overset{w.p.1}{\longrightarrow} 0$ and that the Taylor expansion of $\Phi(Z_i - \bar{Z})$ for $\bar{Z}$ at 0:

$$\Phi(Z_i-\bar{Z}) = \Phi(Z_i) - \phi(Z_i)\bar{Z} + \frac{\xi\phi(X_i)}{2}\bar{Z}^2,$$

where ξ is between 0 and $\bar{Z}$. Then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})\Phi(X_i-\bar{X}) &= \frac{1}{n}\sum_{i=1}^{n}(Z_i-\bar{Z})\Phi(Z_i-\bar{Z}) \\ &\doteq \frac{1}{n}\sum_{i=1}^{n} Z_i\Phi(Z_i-\bar{Z}) \;\doteq\; \frac{1}{n}\sum_{i=1}^{n} Z_i\Phi(Z_i)\end{aligned}$$

and hence

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})\Phi(X_i-\bar{X}) \overset{w.p.1}{\longrightarrow} \int_{-\infty}^{\infty} z\Phi(z)\phi(z)dz = \frac{1}{2\sqrt{\pi}}. \tag{6.9}$$

The Taylor expansion of $e^{-\frac{n(X_i-\bar{X})^2}{2(n+1)}} = e^{-\frac{n(Z_i-\bar{Z})^2}{2(n+1)}}$ for $\bar{Z}$ at 0 is given by

$$e^{-\frac{nZ_i^2}{2(n+1)}} + \frac{n}{n+1}Z_i e^{-\frac{nZ_i^2}{2(n+1)}}\bar{Z} + \frac{n}{2(n+1)}\left[\frac{n}{n+1}(\xi - Z_i)^2 - 1\right]e^{-\frac{n(Z_i-\Xi)^2}{2(n+1)}}\bar{Z}^2,$$

where ξ is between 0 and $\bar{Z}$. Since

$$\mathrm{E}\left(e^{-\frac{nZ_i^2}{2(n+1)}}\right) \;=\; \left(\frac{n+1}{2n+1}\right)^{1/2}$$

and $e^{-\frac{nZ_i^2}{2(n+1)}}$ has finite variance, we have

$$\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{n(X_i-\bar{X})^2}{2(n+1)}} \overset{w.p.1}{\longrightarrow} \frac{1}{\sqrt{2}}. \tag{6.10}$$

Combining (6.9) and (6.10) yield

$$\max_{M_0} \sqrt{n} r_n(M_0) \stackrel{w.p.1}{\longrightarrow} \frac{1}{\sqrt{2}}.$$

Hence, (i) is proved. In general, we can write

$$\sqrt{n} r_n(M_0) = Y_n \frac{\phi(\sqrt{n}(M_0 - \bar{X})}{\phi(0)} = Y_n e^{-\frac{n(\bar{X}-M_0)^2}{2}},$$

where $Y_n \stackrel{w.p.1}{\longrightarrow} \frac{1}{\sqrt{2}}$. Thus,

$$\sqrt{n} r_n(M_0) \stackrel{d}{\longrightarrow} \frac{1}{\sqrt{2}} e^{-\frac{\chi_1^2}{2}}$$

for $M_0 = M$, where χ_1^2 stands for the chi-square random variable with one degree of freedom, and

$$\sqrt{n} r_n(M_0) \doteq \frac{1}{\sqrt{2}} e^{-\frac{n(M-M_0)^2}{2}}$$

for $M_0 \neq M$. This completes the proof of (ii). □

Acknowledgements. The author is grateful to Professor Arthur P. Dempster for extensive helpful e-mail exchanges on inference. Professor Arthur P. Dempster also kindly suggested the author explain that the article represents a different approach to earlier similar work of A. P. Dempster (1969) and that neither the author nor A. P. Dempster had remembered the earlier paper as the author was doing this work in the framework of the modern DS calculus (Demspter, 2007).

References.

[1] Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Annals of Math. Statist.* **37** 355-374. MR0187357

[2] Dempster, A. P. (1968). A generalization of Bayesian inference (with discussion). *J. of the Royal Statist. Soc.* B **30** 205-247. MR0238428

[3] Dempster, A. P. (1969). Upper and lower probability inferences for families of hypotheses with monotone density ratios. *Annals of Math. Statist.* **40** 953-969. MR0246427

[4] Dempster, A. P. (2007). The Dempster-Shafer calculus for statisticians. To appear in *International Journal of Approximate Reasoning.* WWWhttp://www.stat.purdue.edu/ chuanhai/projects/DS/docs/06IJARdraft.pdf

[5] Fisher, R. A. (1959). *Statistical Methods for Scientific Induction*, 2nd Ed. Hafner Publishing Company: New York. MR0131909

[6] LEHMANN, E. L. (1991). *Theory of Point Estimation*, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California. MR1138212
[7] SHAFER, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press. MR0464340

CHUANHAI LIU
DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, IN 47907
E-MAIL: chuanhai@stat.purdue.edu