Fourier Analytic Characterizations of Bayes Rules

by

Anirban DasGupta*
Purdue University

Technical Report #93-45

Department of Statistics
Purdue University

August 1993

*Research supported by NSF grant DMS 93–077–27.

△

FOURIER ANALYTIC CHARACTERIZATIONS OF BAYES RULES

by

Anirban DasGupta
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

ABSTRACT

For the problem of estimating the mean of the $N(\theta, 1)$ distribution using squared error loss, we give some necessary conditions and one necessary and sufficient condition for a given (measurable) function $\delta(x)$ to be Bayes. It is shown that there are connections to some results of G.H. Hardy and M. Mathias on Fourier analysis. We illustrate the use of these conditions by a number of examples. In particular, a Fourier analytic proof of the fact that only linear polynomials can be Bayes is given. Other examples include specific rational functions.

Key Words: Bayes, Fourier transform, Fourier cosine transform, Hermite polynomial, polynomial, rational function.

1. Introduction. The purpose of this note is to point out some interesting connections between some very deep results in the theory of Fourier transforms and some apparently basic results in Bayesian statistics. For estimating the mean θ of the $N(\theta, 1)$ distribution using the standard squared error loss, we give four necessary conditions and one necessary and sufficient condition for a measurable function $\delta(x)$ of the data x to be a proper Bayes estimate for θ. The necessary conditions will be used to give a Fourier analytic proof of the well known fact that linear estimates of the form $a_0 + a_1 x, -\infty < a_0 < \infty, 0 \leq a_1 < 1$ are the only polynomials that are proper Bayes rules for estimating θ; they are also used to demonstrate that certain rational functions of the form $\frac{a_0+a_1 x}{1+a_2 x}$ cannot be proper Bayes estimates in this problem. The necessary and sufficient condition is used to prove that certain other rational functions are in fact Bayes estimates for the mean θ. There are very interesting possibilities for numerical application of this characterization result for proving that a given function suspected of not being Bayes in fact is not Bayes. The reader is to be reminded that characterization of Bayes estimates has been an area of active research in the past; particularly important are the works of Berger and Srinivasan (1978), Brown (1971), Diaconis and Ylvisaker (1978), Farrell (1966), Sacks (1964) and Strawderman and Cohen (1971); also see Berger (1985). Rational functions of X as admissible estimates of θ were studied in Ralescu and Ralescu (1981). Clearly at least some of our results are reinventions of known results by use of alternative mathematical tools. It is the connection to the deep results of Hardy and Mathias that is of greater scientific importance and our prime motive is to point out these surprising connections to Fourier analysis.

2. Main Result. The following notation will be used for the rest of this article: $\delta(x)$ will denote a generic Bayes rule; we will write

$$\delta(x) = x + \frac{h'(x)}{h(x)}; \tag{2.1}$$

it is well known that such a representation is always possible for the Bayes rule in this problem. $H_n(x)$ will denote the nth Hermite polynomial, i.e.,

$$\frac{d^n}{dx^n} e^{-\frac{x^2}{2}} = H_n(x) e^{-\frac{x^2}{2}} \tag{2.2}$$

$$\text{We will denote } H^*_{2n}(x) = 2^{\frac{n}{2}} H_{2n}(x\sqrt{2}); \tag{2.3}$$

thus,

$$\frac{d^n}{dx^n} e^{-x^2} = H^*_{2n}(x) e^{-x^2} \tag{2.4}$$

For future use, note that H^*_{2n} is defined as the $2n$th Hermite Polynomial and denoted as H_{2n} in Gradshteyn and Ryzhik (1980). Finally, for any L^1 function $h(x), C_h(y)$ will denote the Fourier cosine transform of h, i.e.,

$$C_h(y) = \int_{-\infty}^{\infty} (\cos(xy))h(x)dx \tag{2.5}$$

We will now state the results in the form of a single theorem.

Theorem 2.1. Let $\delta(x)$ as in (2.1) be a Bayes rule for estimating the mean θ of the $N(\theta, 1)$ distribution under the squared error loss. Then,

(i) $\lim_{|x|\to\infty} \int_0^x (\delta(t) - t)dt$ exists and equals $-\infty$;

(ii) $\lim_{|x|\to\infty} \int_0^x \delta(t)dt$ exists and equals $+\infty$, unless $\delta(t) \equiv 0$;

(iii) The function $f(x) = e^{\int_0^x \delta(t)dt}$ is the Bilateral Laplace transform of a probability measure μ on the Real line;

(iv) If $\delta(\cdot)$ is skew-symmetric, i.e., $\delta(-x) = -\delta(x)$ for every x, then

$$(-1)^n \int_0^{\infty} H_{2n}(cy)e^{-\frac{y^2}{2}(c^2-1)}C_h(y)dy \geq 0 \ \ \forall\, n \geq 0, c \geq 0; \tag{2.6}$$

in particular, $\int_0^{\infty} h(y)H_{2n}(y)dy \geq 0 \ \ \forall n \geq 0$, if $h(\cdot)$ also satisfies the integrability condition $\int_0^{\infty} h(y)e^{y^2/2}dy < \infty$. Conversely, if (2.6) holds for every $n, c \geq 0$, then $\delta(x) = x + \frac{h'(x)}{h(x)}$ is a proper Bayes rule in the problem.

The following two results will be used for the proof of Theorem 2.1; we state them for ease of reading.

Lemma 2.2 (Hardy). Let $f = \mathbf{R}^1 \to \mathbf{R}^1$ be any measurable function such that $f(x) = O(e^{-x^2/2})$ as $|x| \to \infty$ and such that $\hat{f}(y) = O(e^{-y^2/2})$ as $|y| \to \infty$, where $\hat{f}$ denotes the Fourier transform of f. Then $f(x) = ce^{-x^2/2}$ for some real c.

Remark. Lemma 2.2 asserts that a probability density function f and its Fourier transform $\hat{f}$ cannot both converge to zero at a standard Gaussian rate unless f is the standard normal density itself.

See Chandrasekharan (1987) (page 112) for a proof.

Lemma 2.3 (Mathias). A symmetric and continuous function $\psi(y)$ such that $\psi(0) = 1$ is the Fourier transform of a probability measure if and only if $(-1)^n \int_{-\infty}^{\infty} \psi(py)H_{2n}(y)e^{-y^2/2}dy \geq 0$ for all integer $n \geq 0$ and all $p \geq 0$.

See Lukacs (1970) (page 79) for a proof.

Proof of Theorem 2.1. It is well known that if $\delta(x)$ as in (2.1) is a Bayes rule with respect to the prior G (say), then $h(x)$ is proportional to the marginal density

$$m(x) = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} dG(\theta). \tag{2.7}$$

Therefore, we can assume for all results of this article that $h(x) = m(x)$. Part (i) of the Theorem now follows on simply using

$$\begin{aligned} & e^{\int_0^x (\delta(t)-t)dt} \\ &= e^{\int_0^x \frac{m'(t)}{m(t)} dt} \\ &= \frac{m(x)}{m(0)} \end{aligned} \tag{2.8}$$

and on observing that $\infty > m(0) > 0$ and $m(x) \to 0$ as $|x| \to \infty$ by a standard application of the Dominated convergence theorem.

For part (ii), note that

$$\begin{aligned} m(x)e^{\frac{x^2}{2}} &= m(0) \cdot e^{\int_0^x (\delta(t)-t)dt} \cdot e^{\frac{x^2}{2}} \\ &= m(0) \cdot e^{\int_0^x \delta(t)dt} \end{aligned} \tag{2.9}$$

Now notice that $m(x)$ is the density of the convolution $Z + Y$ where $Z \sim N(0,1)$ and $Y \sim G$, and hence the Fourier transform of m is $O(e^{-y^2/2})$. Consequently, by use of Lemma 2.2, $\limsup_{|x|\to\infty} m(x)e^{x^2/2} = \infty$, from which part (ii) follows immediately on use of (2.9).

For part (iii), on using (2.9) again,

$$\begin{aligned} & e^{\int_0^x \delta(t)dt} \\ &= \frac{e^{\frac{x^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} e^{\theta x} e^{-\frac{1}{2}\theta^2} dG(\theta)}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2} dG(\theta)} \\ &= \frac{\int_{-\infty}^{\infty} e^{\theta x} d\mu(\theta)}{\int_{-\infty}^{\infty} d\mu(\theta)}, \end{aligned}$$

where $d\mu = e^{-\frac{1}{2}\theta^2} dG$, and hence the result follows.

For part (iv), $\delta(x) = x + \frac{h'(x)}{h(x)}$ is a Bayes estimate if and only if $h(\cdot)$ is a Gaussian convolution i.e., if and only if $\psi(y) \cdot e^{y^2/2}$ is (also) a Fourier transform of a Probability measure where $\psi(y)$ is the Fourier transform of h. Since $\delta(x)$ is assumed to be skew-symmetric for part (iv) (although it does not have to be; a parallel result is apparent for the general case), we have, using earlier notation, $\psi(y) = 2C_h(y)$ and hence, by Lemma 2.3,

$$(-1)^n \int_0^\infty H_{2n}(y) e^{-\frac{y^2}{2}} e^{\frac{p^2y^2}{2}} C_h(py)dy \geq 0 \quad \forall\, n \geq 0, p \geq 0 \tag{2.10}$$

(since the integrand in (2.10) is an even function).

Part (iv) will now follow on writing $c = \frac{1}{p}$; the converse follows on using the if part of Lemma 2.3. To observe that $\int_0^\infty h(y)H_{2n}(y) \geq 0$ for all $n \geq 0$, rewrite, for $0 \leq p < 1$, (2.10) as

$$\begin{aligned}
0 &\leq (-1)^n \int_0^\infty H_{2n}(y) e^{-\frac{y^2}{2}} e^{\frac{p^2y^2}{2}} C_h(py)dy \\
&= (-1)^n \int_0^\infty H_{2n}(y) e^{-\frac{y^2}{2}} e^{\frac{p^2y^2}{2}} \left(\int_0^\infty \cos(pxy) h(x) dx \right) dy \\
&= (-1)^n \int_0^\infty h(x) \left(\int_0^\infty \cos(pxy) e^{-\frac{y^2}{2}(1-p^2)} H_{2n}(y) dy \right) dx
\end{aligned} \tag{2.11}$$

(by Fubini's Theorem, which is valid here for $0 \leq p < 1$). In the inside integral in (2.11), now use the integral identity

$$\begin{aligned}
&\int_0^\infty e^{-x^2} \cos(\beta x\sqrt{2}) H^*_{2n}(\alpha x) dx \\
&= \frac{\sqrt{\pi}}{2}(1-\alpha^2)^n e^{-\frac{1}{2}\beta^2} H^*_{2n}\left(\frac{\alpha\beta}{\sqrt{2(\alpha^2-1)}} \right)
\end{aligned} \tag{2.12}$$

(see item 7.388.4, page 840 in Gradshteyn and Ryzhik (1980)).

On using (2.12) in (2.11) and rewriting H^*_{2n} in terms of H_{2n}, as in (2.3), one has

$$\begin{aligned}
&0 \leq (-1)^n \int_0^\infty H_{2n}(y) e^{-y^2/2} e^{\frac{p^2y^2}{2}} C_h(y) dy \\
&= \frac{\sqrt{2\pi}}{2\sqrt{1-p^2}} \left(\frac{p^2}{1-p^2} \right)^n \int_0^\infty h(x) H_{2n}\left(\frac{x}{\sqrt{1-p^2}} \right) e^{-\frac{p^2x^2}{2(1-p^2)}} dx \quad \forall\, n, p \geq 0 \qquad (2.13) \\
&\Rightarrow \int_0^\infty h(x) H_{2n}\left(\frac{x}{\sqrt{1-p^2}} \right) e^{-\frac{p^2x^2}{2(1-p^2)}} dx \geq 0 \quad \forall\, n, p \geq 0,
\end{aligned}$$

which gives the required inequality $\int_0^\infty h(x)H_{2n}(x) \geq 0$ by letting $p \to 0$; it is at this penultimate step that the condition $\int_0^\infty h(x)e^{\frac{1}{2}x^2}dx < \infty$ is required for validity of Dominated convergence.

3. Applications.

Example 1. In this example, we use parts (i) and (ii) of Theorem 2.1 to prove that if $\delta(x) = \sum_{i=0}^{n} a_i x^i$ is a Bayes estimate for θ, then $a_i = 0$ for all $i > 1$. In combination with the well known fact that $a_0 + a_1 x$ is a Bayes rule whenever $0 \leq a_1 < 1$, this will then characterize those polynomials which can occur as Bayes estimates. The result is known, of course. According to usual convention, we will assume $a_n \neq 0$ and $n \geq 2$. Towards this end, first note that if $n \geq 2$, then part (i) of Theorem 2.1 implies that n cannot be even. So assume $n = 2k+1, k \geq 1$. However, part (ii) of Theorem 2.1 would force $a_{2k+1} > 0$, which would then contradict part (i). This proves $a_{2k+1} = 0$. The proof is now completed by standard induction arguments.

Example 2. Let $\delta(x) = \frac{a_0+a_1x}{1+a_2x}$ for $x \geq 0$; we will assume $a_2 > 0$, since $a_2 \leq 0$ is clearly uninteresting. Notice that there is no loss of generality in keeping the leading term in the denominator as 1. In this example, we will use part (iii) of Theorem 2.1 to demonstrate that estimates of the above form cannot be Bayes estimates if $\gamma = a_0 a_2 - a_1 > 0$. To achieve this, first note the elementary fact that for $x \geq 0$,

$$\int_0^x \delta(t)dt = \left(\frac{a_0}{a_2} - \frac{a_1}{a_2^2}\right) \ell n(1 + a_2 x) + \frac{a_1 x}{a_2}. \tag{3.1}$$

By virtue of Theorem (2.1), $e^{\int_0^x \delta(t)dt}$ is a Bilateral Laplace transform corresponding to a probability measure, i.e., for $x \geq 0$,

$$e^{\int_0^x \delta(t)dt} = f(x) = \int_{-\infty}^{\infty} e^{xy} d\nu(y)$$

for some probability measure ν. It is an easy fact that therefore $f(x)$ must be log convex. However,

$$\log f(x) = \frac{\gamma}{a_2^2} \ell n(1 + a_2 x) + \frac{a_1 x}{a_2}$$

and is clearly not convex if $\gamma > 0$. Of course, there are more direct ways to obtain this example.

Example 3. This is a positive example in the sense that we will now use part (iv) of Theorem 2.1 to prove that for every $0 < r < 1$, the rational function

$$\delta(x) = \frac{3rx + r^2x^3}{1 + rx^2}$$

is a proper Bayes estimate in this problem. It follows that

$$h(x) = A_1 \cdot r(1 + rx^2)e^{-\frac{x^2}{2}(1-r)}, \tag{3.2}$$

where A_1 is a positive constant and will henceforth be justifiably ignored. Our intention is to show that

$$(-1)^n \int_0^\infty H_{2n}(cy)e^{-\frac{y^2}{2}(c^2-1)}C_h(y)dy \geq 0 \quad \forall\ n, c \geq 0.$$

The calculations in this example are complex, but the idea is to present an evidence that part (iv) of Theorem 2.1 is not an academic characterization. Towards this end, note the following series of facts:

a $C_h(y) = A_2 \cdot (1 - \frac{ry^2}{1-r})e^{-\frac{y^2}{2(1-r)}}$, (3.3)

where A_2 is a positive constant. (3.3) can be obtained by direct integration or by using the formula for the Fourier transform of the normal distribution $N(0, \sigma^2)$ and then taking two derivatives to obtain $\int_{-\infty}^{\infty} x^2(\cos xy)e^{-\frac{x^2}{2\sigma^2}}dx$, but for a sign.

b $\int_0^\infty e^{-y^2}H^*_{2n}(xy)dy = \frac{\sqrt{\pi}}{2}\frac{(2n)!}{n!}(x^2-1)^n$ (3.4)

(see 7.373.2, page 837, Gradshteyn and Ryzhik (1980)).

c $\int_0^\infty e^{-2\alpha x^2}x^2 H^*_{2n}(x)dx = \frac{(-1)^n x^{2n-\frac{5}{2}}\cdot\Gamma(n+\frac{1}{2})\cdot F(-n,\frac{3}{2};\frac{1}{2};\frac{1}{2\alpha})}{2\cdot\alpha^{\frac{3}{2}}}$, (3.5)

(see 7.376.2 in page 838 in Gradshteyn and Ryzhik (1980)) where $F(a, b; c; z)$ is the usual Hypergeometric series

d $F(-n, b; b; z) = (1-z)^n$ for any b

(see 9.121.1 in page 1040 in Gradshteyn and Ryzhik (1980)).

e $c(c+1)F(a, b; c, z) - c(c+1)F(a, b; c+1; z) - abzF(a+1, b+1; c+2; z) = 0$ (3.6)

(see 9.137.6 in page 1044 in Gradshteyn and Ryzhik (1980)).

At this stage, recall that the required inequality is

$$\begin{aligned}
& (-1)^n \int_0^\infty H_{2n}(cy)e^{-\frac{y^2}{2}(c^2-1+\frac{1}{1-r})}\left(1 - \frac{ry^2}{1-r}\right) \geq 0 \\
\iff\ & (-1)^n \int_0^\infty H_{2n}(cy)e^{-\frac{y^2}{2}(c^2+\frac{r}{1-r})}\left(1 - \frac{ry^2}{1-r}\right)dy \geq 0 \\
\iff\ & (-1)^n \left[\int_0^\infty H_{2n}(cy)e^{-\frac{y^2}{2}(c^2+\frac{r}{1-r})}dy - \frac{r}{1-r}\int_0^\infty H_{2n}(cy)e^{-\frac{y^2}{2}(c^2+\frac{r}{1-r})}y^2dy\right] \geq 0
\end{aligned} \tag{3.7}$$

To obtain the first integral in the right side of (3.7), we use fact **b** listed above, while fact **c** is used to obtain the second integral.

On using the notation

$$z = \frac{c^2(1-r)}{c^2(1-r)+r},$$

this reduces (3.7) on algebra and use of fact **e** to

$$\sqrt{\pi}\cdot\frac{(2n)!}{n!}(1-z)^n - \frac{1-z}{z}\cdot 2^{2n-1}z^{\frac{3}{2}}\cdot\Gamma(n+\frac{1}{2})((1-z)^n - 2nz(1-z)^{n-1}) \geq 0 \qquad (3.8)$$

This, however, will follow if one can show

$$\sqrt{\pi}\cdot\frac{(2n)!}{n!}(1-z)^n - \frac{1-z}{z}\cdot 2^{2n-1}z^{\frac{3}{2}}\Gamma(n+\frac{1}{2})(1-z)^n \geq 0 \qquad (3.9)$$

since $0 < z < 1$.

(3.9) is achieved on using Legendre's duplication formula

$$\Gamma\left(n+\frac{1}{2}\right) = \frac{\sqrt{\pi}\Gamma(2n)}{2^{2n-1}\Gamma(n)}$$

(see Dettman (1984), page 199) in (3.9); actually one proves a stronger inequality in the process. We skip these details for brevity. This shows $\delta(x) = \frac{3rx+r^2x^3}{1+rx^2}$ to be a Bayes estimate for every $0 < r < 1$.

4. Final Remarks and Summary. We have presented here some techniques for verifying if a given estimate is Bayes. Connections to Fourier analysis are derived. We also give significant evidence that the methods are useful in some explicit examples. The characterization result in Theorem 2.1 is of potential in numerical verification of the Bayes status of an estimate.

References

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis,* Springer Verlag, New York.

Berger, J. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems, *Ann. Statist.* **6**, 783–801.

Brown, L.D. (1971). Admissible estimators, recurrent diffusions and boundary value problems, *Ann. Math. Stat.* **42**, 855–903.

Chandrasekharan, S.(1987). *Classical Fourier Transforms*, Springer Verlag, New York.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for Exponential families, *Ann. Statist.* **7**, 269–281.

Dettman, J.W. (1984). *Applied Complex Variables*, Dover Publications, New York.

Farrell, R. (1966). Weak limits of sequences of Bayes procedures in estimation theory, *Proc. Fifth Berkeley Symposium*, **1**, 83-11, Univ. California Press, Berkeley.

Gradshteyn, I.S. and Ryzhik, I.M. (1980). *Tables of Integrals, Series, and Products*, Academic Press, New York.

Lukacs, E. (1970). *Characteristic Functions*, Hafner Publishing Company, New York.

Ralescu, D. and Ralescu, S. (1981). A class of nonlinear admissible estimators in the one parameter Exponential family, *Ann. Statist.* **9**, 177–183.

Sacks, J. (1963). Generalized Bayes solutions in estimation problems, *Ann. Math. Stat.* **34**, 751–768.

Strawderman, W.E. and Cohen, A. (1971). Admissibility of estimators of the mean vector of a multivariate normal distribution with quadratic loss, *Ann. Math. Stat.* **42**, 270–296.