The Annals of Statistics
1982, Vol. 10, No. 2, 625–629

A USEFUL EMPIRICAL BAYES IDENTITY

BY NOEL CRESSIE

Visiting Research Scientist, Educational Testing Service, Princeton

For any decision problem, one wishes to find that estimator which minimizes the expected loss. If the loss function is squared error, then the estimator is the mean of the Bayes posterior distribution. Unfortunately the prior distribution may be unknown, but in certain situations empirical Bayes methods can circumvent this problem by using past observations to estimate either the prior or the Bayes estimate directly. Empirical Bayes methods are particularly appealing when the Bayes estimate depends only on the marginal distribution of the observed variable, yielding what is known as a simple empirical Bayes estimate. The paper looks at the underlying circumstance of when a simple empirical Bayes estimator is available, and shows its occurrence not to be happenstance.

1. Introduction. A statistician is often faced with the following problem: Observations $X_1, \cdots, X_n$ from some probability distribution have been collected and the statistical model has been chosen as $h(x \mid \zeta)$. The parameter ζ is interpretable with regard to the phenomenon under study, and so some inference is needed from data $(X_1, \cdots, X_n)$ to parameter ζ. Sometimes however, the interpretation that ζ is a *fixed* but unknown parameter, is unrealistic. For example, latent trait models in mental testing (Lord and Novick, 1968) must realistically model a sequence of bivariate random vectors (X_1, Z_1), $\cdots, (X_n, Z_n)$, where Z_i is the unobserved trait (or true score) of the ith examinee, and X_i is that examinee's observed score on a test. The population from which the examinee comes has a certain distribution of true scores, call it G. So if Z is the true score of a randomly chosen member of that population, then $G(\zeta) = \Pr(Z \leq \zeta)$. Thus there are contexts where it makes sense to model the unknown parameter ζ as also being random. A sensible question then to ask is: Suppose I observe $X = x$, what then can I say about the associated unobserved Z?

The approach we have just been describing would be called empirical Bayes (Robbins, 1956, 1964; Maritz, 1970) if we used other observed variables $X_1, \cdots, X_n$ to make an inference from observed $X = x$ to unobserved Z. Section 2 briefly sets out the general ideas behind the empirical Bayes approach and considers the very important notion of a simple empirical Bayes estimator. Up to now these have been discovered in a very haphazard way. However, the identity derived in this section shows their appearance not to be happenstance. The binomial model is used throughout to illustrate the approach and new results for it are derived. Concluding remarks are made in Section 3.

2. Empirical Bayes and the underlying relation. Consider a probability space generated by the bivariate random vector (X, Z), where the first member is observable but the second is not. Let Z have an unknown distribution function G and, given $Z = \zeta$, X is modeled to have probability distribution or density $h(x \mid \zeta)$, where h is known. Hence the marginal distribution of X is

$$\phi_G(x) = \int h(x \mid \zeta)\, dG(\zeta). \tag{2.1}$$

Received January 1981; revised October 1981.

AMS 1980 *subject classification*. Primary 62F10, secondary 62P15.

Key words and phrases. Bayes estimator, binomial model, exponential families, linear functionals, power series distribution.

Given $X = x$, we wish to make a decision $\delta(x)$ about the associated value of Z. Let a loss $L(\delta(x), \zeta) \equiv \{\delta(x) - \zeta\}^2$ be incurred when the parameter value is ζ and the decision $\delta(x)$ is made. The overall expected loss is then $\mathscr{E}\{\delta(X) - Z\}^2$, and the resulting δ which minimizes this, called the Bayes estimator, is the posterior mean $\delta_G(x) = \int \zeta \, dP(\zeta \mid x)$, written as $\mathscr{E}_{Z|x}(Z)$; i.e.,

$$(2.2) \qquad \mathscr{E}_{Z|x}(Z) = \{\phi_G(x)\}^{-1} \int \zeta h(x \mid \zeta) \, dG(\zeta).$$

If G were known, the decision δ_G, a function of observed $X = x$, would be easy to calculate. The "empirical" part of empirical Bayes involves using an already observed sample $X_1, \cdots, X_n$ (n independent and identically distributed observations from ϕ_G) independent from X, to estimate δ_G and possibly G.

Suppose we rewrite (2.1) as

$$(2.3) \qquad \phi_G(\cdot) = \mathscr{E}_Z\{h(\cdot \mid Z)\},$$

where $\mathscr{E}_Z$ is the expectation operator (linear), with expectations being taken over Z. Now suppose L is a linear functional with domain contained in the space of all real-valued functions; then, provided the range of Z does not depend upon x,

$$(2.4) \qquad L(\phi_G(\cdot)) = \mathscr{E}_Z\{L(h(\cdot \mid Z))\},$$

when both sides are well defined. Examples of L are the differential operator $D\{f(x)\} = df(x)/dx$, ($x \in \mathbb{R}$), and the difference operator $\Delta^{\ell}\{f(x)\} = f(x + \ell) - f(x)$, ($x \in \mathbb{R}$).

Now consider only those $x \in \mathbb{R}$ for which $\phi_G(x) > 0$, $S_\phi \equiv \{x : \phi_G(x) > 0\}$. Then from (2.1), for $x \in S_\phi$

$$\frac{L(\phi_G(x))}{\phi_G(x)} = \frac{\int L(h(x \mid \zeta)) \, dG(\zeta)}{\int h(x \mid \alpha) \, dG(\alpha)} = \int \frac{L(h(x \mid \zeta))}{h(x \mid \zeta)} \times \frac{h(x \mid \zeta) \, dG(\zeta)}{\int h(x \mid \alpha) \, dG(\alpha)} = \int \frac{L(h(x \mid \zeta))}{h(x \mid \zeta)} \, dP(\zeta \mid x),$$

where $P(\zeta \mid x)$ is the posterior distribution function of Z, given $X = x$. Thus for $x \in S_\phi$

$$(2.5) \qquad L(\phi_G(x))/\phi_G(x) = \mathscr{E}_{Z|x}\{L(h(x \mid Z))/h(x \mid Z)\},$$

where $\mathscr{E}_{Z|x}$ is the conditional expectation operator (linear), with expectations being taken over Z conditional upon $X = x$.

Equation (2.5) is a surprising consequence of (2.3). Define

$$(2.6) \qquad R : f(\cdot) \rightarrow L(f(\cdot))/f(\cdot),$$

a functional whose domain is the same as that of L. Then (2.3) to (2.6) together imply that for $x \in S_\phi$,

$$(2.7) \qquad R(\phi_G(x)) = \mathscr{E}_{Z|x}\{R(h(x \mid Z))\}.$$

So although R is nonlinear, it exhibits a "quasi" linear property; it can be taken under the expectation operator but in so doing it modifies the operator to a conditional expectation operator. This underlying relation (2.7) was presented in terms of distribution functions for the case where X is a continuous variable by Maritz and Lwin (1975); see also Rutherford and Krutchkoff (1969) and Nichols and Tsokos (1972) where (2.7) is used for particular choices of R.

The potential use of (2.7) is apparent if we can find a functional R_0 for which

$$(2.8) \qquad R_0(h(x \mid \zeta)) = \zeta u(x) + v(x),$$

where u is some function such that $u(x) > 0$ for $x \in S_\phi$, and v is a real-valued function defined on S_ϕ. Then (2.7) becomes

$$\{R_0(\phi_G(x)) - v(x)\}/u(x) = \mathscr{E}_{Z|x}\{Z\},$$

and the Bayes estimate of Z given by (2.2) can be written

$$\delta_G(x) = \{R_0(\phi_G(x)) - v(x)\}/u(x), \tag{2.9}$$

where R_0 satisfies (2.8). The important thing to notice about (2.9) is that the right-hand side depends only on x and the marginal distribution of the observed variable X. Therefore an observed sample $X_1, \cdots, X_n$ can be used to estimate it, and by implication $\delta_G(x)$. The result: a simple empirical Bayes estimator of Z given $X = x$.

Now (2.7) also suggests another approach to the estimation of Z, given $X = x$. If we can find functions u, v and w that satisfy the decomposition

$$R_1(h(x \mid \zeta)) = u(x)w(\zeta) + v(x), \tag{2.10}$$

for the particular R_1 we have chosen, then (2.7) becomes

$$\{R_1(\phi_G(x)) - v(x)\}/u(x) = \mathscr{E}_{Z|x}\{w(Z)\}.$$

So in this context, it makes more sense to estimate the "natural" parameter $T \equiv w(Z)$, whose Bayes estimate is $\{R_1(\phi_G(x)) - v(x)\}/u(x)$.

An example is provided by the binomial model, which we write as

$$h_b(x \mid \zeta) = \binom{k}{x} (1 - \zeta)^k \{\zeta/(1 - \zeta)\}^x, \quad x = 0, 1, \cdots, k. \tag{2.11}$$

Then

$$R_1(h_b(x \mid \zeta)) \equiv h_b(x + 1 \mid \zeta)/h_b(x \mid \zeta) = \{\zeta/(1 - \zeta)\}(k - x)/(x + 1), \quad x = 0, 1, \cdots, k - 1.$$

Hence from (2.10) the Bayes estimate of $T = Z/(1 - Z)$, given $X = x$, is

$$\theta(x) \equiv (x + 1)\phi_G(x + 1)/\{(k - x)\phi_G(x)\}, \quad x = 0, 1, \cdots, k - 1. \tag{2.12}$$

Equally, the Bayes estimate of $U = (1 - Z)/Z$ is

$$\nu(x) \equiv (k - x + 1)\phi_G(x - 1)/\{x\, \phi_G(x)\}, \quad x = 1, 2, \cdots k. \tag{2.13}$$

Use of Jensen's inequality on $\mathscr{E}_{Z|x}\{Z/(1 - Z)\}$ and $\mathscr{E}_{Z|x}\{(1 - Z)/Z\}$ yields

$$1/\{1 + \nu(x)\} \le \mathscr{E}_{Z|x}\{Z\} \le 1/\{1 + \theta(x)^{-1}\}, \quad x = 0, 1, \cdots, k, \tag{2.14}$$

where $\nu(0) \equiv \infty$ and $\theta(k) \equiv \infty$. It is fortuitous that $\theta(x)/\{1 + \theta(x)\} = 1/\{1 + \nu(x + 1)\}$ $(x = 0, 1, \cdots, k)$, where $\nu(k + 1) \equiv 0$. Therefore from (2.13) and (2.14), the quantities

$$\theta(x - 1)/\{1 + \theta(x - 1)\}, \quad x = 0, 1, \cdots, k + 1, \tag{2.15}$$

where $\theta(-1) \equiv 0$, $\theta(k) \equiv \infty$ are increasing in x and partition the interval [0, 1] into $k + 1$ disjoint intervals such that

$$\theta(x - 1)/\{1 + \theta(x - 1)\} \le \mathscr{E}_{Z|x}\{Z\} \le \theta(x)/\{1 + \theta(x)\}, \quad x = 0, \cdots, k. \tag{2.16}$$

In a personal communication, Professor H. Robbins has informed us of the appearance of these intervals in a series of lectures given by him at SUNY, Stony Brook in 1979–1980.

Empirical Bayes estimation for the binomial parameter has been studied by a number of authors; e.g. Martz and Lian (1974), Lord and Cressie (1975), Vardeman (1978), Berry and Christensen (1979) and Cressie (1979). To solve (2.1) for G in this case is an ill-posed problem since two G's with the same first k moments will result in the same ϕ_G. It is not surprising that a simple Bayes estimator of Z has eluded researchers to date, since vastly different priors G could support the same marginal ϕ_G. What is surprising is that no such difficulty is encountered for a suitably chosen function $Z/(1 - Z)$ or $(1 - Z)/Z$. Further-more, it has been long known and makes good intuitive sense that $\mathscr{E}_{Z|x}\{Z\}$ should be increasing in x, but now we can do much better since (2.15) defines increasing intervals given by (2.16) that tie these quantities down even further.

Sometimes it is very difficult in a particular problem to find functions u, v and w, and a functional R_1, to satisfy (2.10). Cressie (1982) shows that for most discrete distributions such a functional is available. See also Cressie and Holland (1981) for an application to the Rasch model. Typically, though, we will have $R_1(h(x|\zeta)) = a_x(\zeta)$, and so $R_1(\phi_G(x)) = \mathscr{E}_{Z|x}\{a_x(Z)\}$. It is still possible to find an empirical Bayes estimate of Z, given $X = x$, if some way of undoing the x-dependent transformation can be found. Obviously the naive choice $a_x^{-1}(R_1(\phi_G(x)))$ will be biased. Even in the case where there is not x-dependence, care is needed.

The binomial model given by (2.11) once again provides a good illustrative example. If $T = Z/(1 - Z)$, then $Z = T/(1 + T) \equiv b(T)$. Hence using the bias correction given by Kendall and Stuart (1963, page 231) we see that

$$(2.17) \qquad \mathscr{E}_{Z|x}\{Z\} \simeq \theta(x)/\{1 + \theta(x)\} - \mathrm{Var}_{T|x}(T)/\{1 + \theta(x)\}^3, \quad x = 0, \cdots, k - 1.$$

When we recognize the first term of (2.17) as the upper bound of $\mathscr{E}_{Z|x}(Z)$, then clearly the second term is a compensating correction. Another application of the useful identity (2.7) yields $\mathrm{Var}_{T|x}(T) = \theta(x + 1)\theta(x) - \theta^2(x)$, $x = 0, \cdots, k - 2$. Care is needed here because we know that $\mathscr{E}_{Z|x}\{Z\}$ must lie in the interval given by (2.16). Some slight modification to $\mathrm{Var}_{T|x}\{T\}/(1 + \theta(x))^3$ will guarantee this; instead of correcting $\theta(x)/(1 + \theta(x))$ with $-\{\theta(x + 1)\theta(x) - \theta^2(x)\}/\{1 + \theta(x)\}^3$, we propose the following simple approximate Bayes estimator

$$(2.18) \qquad d(x) = \frac{\theta(x)}{1 + \theta(x)} - \frac{\{\theta(x) - \theta(x - 1)\}\theta(x)}{[1 + \theta(x) + \theta(x)\{1 + \theta(x - 1)\}]\{1 + \theta(x)\}}, \quad x = 0, \cdots, k,$$

where $\theta(\cdot)$ is given by (2.12) and $\theta(-1) = 0$, $\theta(k) = \infty$. We are led to (2.18) for the following reasons: it satisfies the important interval constraint (2.16); similar considerations given to $\mathscr{E}_{Z|x}\{(1 - Z)/Z\}$ and a correction from below yields the identical estimator; when G is assumed to be a beta distribution (as it often is for binomial h) with parameter r, s then $\mathscr{E}_{Z|x}\{Z\}$ can be calculated to be $(x + r)/(r + s + k)$, which is exactly $d(x)$ since $\theta(x) = (x + r)/(x + 1 + s + k)$. Hence there are sensible ways to control for bias.

3. Conclusions. The useful identity presented in Section 2, namely (2.7), explains why certain families of models are always chosen in the empirical Bayes context. For discrete distributions, see the family considered by Nichols and Tsokos (1972). This family is perfectly constructed for the linear operator in (2.6) to be $L = F^\ell$, the ℓth order forward shift. Nichols and Tsokos also consider the continuous exponential family of distributions, which is perfectly constructed for the linear operator in (2.6) to be $L = D^\ell$, the ℓth derivative.

However, we are then faced with the problem of estimating density derivatives $\phi_G^{(\ell)}(x)$, $\ell = 0, 1, 2, \cdots$, from marginal observations $X_1, \cdots, X_n$. This is a non-trivial task as the number of papers devoted to smooth estimation of a density and its derivatives will testify. Some of these include Johns and Van Ryzin (1972), Lin (1975), and Kim and Van Ryzin (1980). The problem is basically that derivatives of density functions are numerically and statistically very unstable quantities to estimate. Usually an approximation is made by substituting some sort of smooth differencing for differentiation. This rather circumvent and approximate approach can be avoided by once again realizing the potential of (2.7). Instead of choosing D for the linear operator L in (2.6), go immediately to choosing F^ℓ: $f(\cdot) \to f(\cdot + \ell)$. So if $h(x|\zeta) = \exp\{x\zeta + a(x) + b(\zeta)\}$; i.e. the continuous exponential family, then from (2.7),

$$R(\phi_G(x)) = \phi_G(x + \ell)/\phi_G(x) = \mathscr{E}_{Z|x}[\exp(\ell Z)\cdot\exp\{a(x + \ell) - a(x)\}].$$

Whence,

$$\mathscr{E}_{Z|x}\{\exp(Z)\} = \exp\{a(x) - a(x+1)\}\phi_G(x+1)/\phi_G(x). \tag{3.1}$$

Thus the Bayes estimator of $\exp(Z)$ is given by the right-hand side of (3.1), which can be estimated from observations on X alone more stably than can anything involving derivatives. But it might be the Bayes estimate of Z that we want, so an "undoing" transformation needs to be applied to yield $\mathscr{E}_{Z|x}\{Z\}$, such as for the binomial in the previous section. The result is an approximate simple empirical Bayes estimate of Z that only requires estimation of a density and not its derivatives.

Sometimes it is a consequence of the model that the Bayes estimator satisfies certain smoothness properties. For example, $\mathscr{E}_{Z|x}\{Z\}$ should be monotonic increasing in x for the binomial model (2.11). Hence any empirical Bayes estimator should reflect this; the smoothness aspect of empirical Bayes has not been pursued in this short note and the interested reader is referred to Van Houwelingen (1977).

Acknowledgements. The author would like to thank Dr. T. Lwin and Professor J. S. Maritz for helpful input, and in particular, Professor H. Robbins for many useful suggestions.

REFERENCES

BERRY, D. A., and CHRISTENSEN, R. (1979). Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes. *Ann. Statist.* **7** 558–568.

CRESSIE, N. (1979). A quick and easy empirical Bayes estimate of true scores. *Sankyhā B* **41** 101–108.

CRESSIE, N. (1982). Empirical Bayes estimation for discrete distributions. *South African Statist. J.* **16,** to appear.

CRESSIE, N., and HOLLAND, P. W. (1981). Characterising the manifest probabilities of latent trait models. Unpublished manuscript.

JOHNS, M. V. JR., and VAN RYZIN, J. (1972). Convergence rates in empirical Bayes two-action problems. II. Continuous case. *Ann. Math. Statist.* **43** 934–947.

KENDALL, M., and STUART, A. (1963). *Advanced Theory of Statistics, Vol. 1* (2nd edn.). Griffin, London.

KIM, B. K., and VAN RYZIN, J. (1980). Estimation of the derivative of a density and related functions. *Comm. Statist. A—Theory Methods* **9** 697–709.

LIN, P. E. (1975). Rates of convergence in empirical Bayes estimation problems: continuous case. *Ann. Statist.* **3** 155–164.

LORD, F. M. and CRESSIE, N. (1975). An empirical Bayes procedure for finding an interval estimate. *Sankhyā B* **37** 1–9.

LORD, F. M., and NOVICK, M. R. (1968). *Statistical Theories of Mental Test Scores.* Addison-Wesley, Reading.

MARITZ, J. S. (1970). *Empirical Bayes Methods.* Methuen, London.

MARITZ, J. S., and LWIN, T. (1975). Construction of simple empirical Bayes estimators. *J. Roy. Statist. Soc. B* **39** 421–425.

MARTZ, H. F., and LIAN, M. G. (1974). Empirical Bayes estimation of the binomial parameter. *Biometrika* **61** 517–523.

NICHOLS, W. G., and TSOKOS, C. P. (1972). Empirical Bayes point estimation in a family of probability distributions. *Internat. Statist. Rev.* **40** 147–151.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium* (ed. J. Neyman). 157–164. Univ. Calif. Press, Berkeley.

ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

RUTHERFORD, J. R., and KRUTCHKOFF, R. G. (1969). Some empirical Bayes techniques in point estimation. *Biometrika* **56** 133–137.

VAN HOUWELINGEN, J. C. (1977). Monotonizing empirical Bayes estimators for a class of discrete distributions with monotone likelihood ratio. *Statistica Neerlandica* **31** 95–104.

VARDEMAN, S. B. (1978). Bounds on the empirical Bayes and compound risks of truncated versions of Robbins' estimator of a binomial parameter. *J. Statist. Plan. Infer.* **2** 245–252.

SCHOOL OF MATHEMATICAL SCIENCES
THE FLINDERS UNIVERSITY OF SOUTH AUSTRALIA
BEDFORD PARK, S.A. 5042 AUSTRALIA