The Annals of Statistics
2003, Vol. 31, No. 2, 379–390

INVITED PAPER

COMPOUND DECISION THEORY AND EMPIRICAL BAYES METHODS

BY CUN-HUI ZHANG

Rutgers University

1. Introduction. Compound decision theory and empirical Bayes methodology, acclaimed as "two breakthroughs" by Neyman (1962), are the most important contributions of Herbert Robbins to statistics. The purpose of this paper is to provide a brief description of his work in these two intimately connected fields, its impact and a number of important related developments.

Robbins introduced compound decision theory in 1950 at the Second Berkeley Symposium on Mathematical Statistics and Probability. Compound decision theory concerns a sequence of independent statistical decision problems of the same form. Its basic thrust is the possibility of gaining substantial reduction of total risk by allowing statistical procedures for the individual component problems to depend on the observations in the entire sequence. It demonstrates, against naive intuition, that stochastically independent experiments are not necessarily "noninformative" to each other in statistical decision making.

Five years later, at the Third Berkeley Symposium, Robbins developed empirical Bayes (EB) theory. EB concerns experiments in which the unknown parameters are i.i.d. random variables with an unknown common prior distribution. EB methodologies provide statistical procedures which approximate the ideal Bayes rule for the true model, so that the goal of the Bayesian inference is nearly achieved without specifying a prior. EB procedures usually perform well conditionally on the unknown parameters and thus provide solutions to compound decision problems. EB methods also find applications in problems with more complex structures and for inference about multivariate and infinite-dimensional parameters in a single experiment.

Compound decision theory and EB have had great influence on modern statistical thinking and practice. Since Robbins' pioneering papers, EB methods have been applied in a wide range of paradigms and to numerous real-life problems; cf. Neyman (1962), Cover (1968), Copas (1972), Carter and Rolph (1974), Simar (1976), Efron and Morris (1977), Van Ryzin and Susarla (1977), Susarla and Van Ryzin (1978), Rubin (1980), Hoadley (1981), Morris (1983),

Received December 2001; revised June 2002.

AMS 2000 subject classifications. Primary 62C12, 62C25; secondary 62G08, 62G05.

Key words and phrases. Empirical Bayes, compound decision theory, nonparametric inference, prediction.

Cover (1991), Zaslavsky (1993), van Houwelingen and Thorogood (1995), Carlin and Louis (1996), Efron, Storey and Tibshirani (2001) and Efron, Tibshirani, Storey and Tusher (2001).

2. Compound decision problems. Let $f(x;\theta)\nu(dx)$ be a family of probability measures with a parameter θ. Consider a sequence of independent experiments with observations $X_i \sim f(x;\theta_i)$, $i = 1,\dots,n$, where θ_i are deterministic unknown parameters. Suppose that we are interested in making statistical decisions δ_i about θ_i with a loss function $L(a,\theta)$. Robbins (1951) formulated the compound decision problem, in which δ_i are allowed to depend on the observations $X_{(n)}$ from all n experiments, under the compound risk

$$R_n\big(\delta_{(n)},\theta_{(n)}\big) \equiv \frac{1}{n}\sum_{i=1}^{n} E_{\theta_{(n)}} L\big(\delta_i(X_{(n)}),\theta_i\big). \tag{1}$$

Here and elsewhere, $h_{(n)} \equiv (h_1,\dots,h_n)^{\mathrm{tr}}$ for all sequences $\{h_i\}$ and the superscript "tr" indicates transposition. For separable decision rules of the form $\delta_i(X_{(n)}) = t(X_i)$, that is, the ith decision being a fixed deterministic function of X_i, the compound risk (1) is equal to the Bayes risk

$$R(t,G) \equiv \int\left[\int L\big(t(x),\theta\big) f(x;\theta)\nu(dx)\right] G(d\theta) \tag{2}$$

for a single decision problem under the unknown prior $G(A) = G_n(A) \equiv n^{-1}\sum_{i=1}^{n} I\{\theta_i \in A\}$. Robbins' proposal is to seek asymptotically optimal procedures satisfying

$$R_n\big(\delta_{(n)},\theta_{(n)}\big) = R^*(G_n) + o(1) \qquad \text{for large } n, \tag{3}$$

where $R^*(G) \equiv \min_t R(t,G)$ is the minimum Bayes risk given a prior G.

In the simple example of testing $\theta_i = 1$ against $\theta_i = -1$ based on $X_i \sim N(\theta_i,1)$, with loss $L(a,\theta) = I\{a \neq \theta\}$, Robbins constructed decision rules $\delta_{(n)}(X_{(n)})$, based on suitable estimates of G_n, such that (3) holds uniformly in $\theta_{(n)}$. This exhibits the benefits of utilizing "noninformative" observations $\{X_j,\ j \neq i\}$ in the ith decision problem and the possibility of aggressive adaptation in the sense of (3) without any knowledge of G_n.

The compound decision theory has been further developed by Robbins (1955 with Hannan, 1962, 1962 with Samuel) and others, including many students of his; cf. Hannan (1957), Samuel (1963a, 1964, 1965), Van Ryzin (1966a, b), Gilliland (1968) and Vardeman (1978) on sequential compound decision problems, Gilliland and Hannan (1969), Gilliland, Hannan and Huang (1976) and Datta (1991) on Bayes methods, Fox (1970) on estimating unknown priors, Oaten (1972) on problems with compact decision spaces, Hannan and Van Ryzin (1965), Susarla (1974) and Zhang (1997) on convergence rates and asymptotic minimaxity, and references in Section 4 for linear/parametric EB methods.

3. Empirical Bayes methods. Robbins (1956) introduced the EB approach to statistical decision problems. In the EB setting, the unknown parameters θ_i are considered as independent random variables with an unknown common prior G, and the aim is to find decision rules performing nearly as well as the ideal Bayes rule. EB procedures can be applied in two different scenarios in practice. The first one, on which Robbins (1956) focused, is the sequential EB problem, in which only observations $X_1, \ldots, X_j$ are available for the jth decision problem. The second one is the compound problem of Robbins (1951), in which the whole vector $X_{(n)}$ can be used in the inference about all θ_i. The risk for the compound version is

$$R_n(\delta_{(n)}, G) \equiv n^{-1} \sum_{i=1}^{n} E_G L\big(\delta_i(X_{(n)}), \theta_i\big) = E_G R_n(\delta_{(n)}, \theta_{(n)}).$$

Here the compound risk $R_n(\delta_{(n)}, \theta_{(n)})$ in (1) becomes the conditional risk given $\theta_{(n)}$. Furthermore, the asymptotic optimality criterion (3) is naturally replaced by $R_n(\delta_{(n)}, G) = R^*(G) + o(1)$. If we are confined to symmetric procedures under permutations of decision problems in the compound case, the two versions of the EB decision problems are mathematically equivalent.

Robbins' (1951, 1956) solutions to the compound and EB decision problems are essentially the same. Let $R(t, G)$ be the Bayes risk in (2) and let

$$t_G^*(x) \equiv t_G^* \equiv \arg\min_{t \in \mathcal{D}} R(t, G) \tag{4}$$

be the ideal Bayes rule, where G is the unknown prior in EB problems and $G = G_n$ is the empirical distribution of unknown parameters, as in (3), in compound decision problems. Robbins' procedures can be written as

$$\delta_i(X_{(n)}) = \widehat{t}_n(X_i), \tag{5}$$

where $\widehat{t}_n(\cdot) \equiv \widehat{t}_n(\cdot; X_{(n)})$ is an estimate of t_G^* based on the whole vector $X_{(n)}$. As a general solution, he suggested using $\widehat{t}_n = t^*_{\widehat{G}_n}$ with a suitable estimate $\widehat{G}_n$ of G and formulated the problem of estimating the prior G based on $X_{(n)}$. Interestingly, Robbins (1951) described this process as an attempt to "lift ourselves by our own bootstraps."

For point estimation of θ_i with several discrete and Laplace-type [i.e., $\nu(dx) = dx$] exponential families $f(x; \theta)$, Robbins (1956) proposed simpler methods. He expressed Bayes estimators $t_G^*(x) = E_G[\theta_i \mid X_i = x]$ as functionals of the mixture density $f_G(x) \equiv \int f(x; \theta)\, dG(\theta)$ of X_i and suggested deriving $\widehat{t}_n(x)$ directly from density estimators. For example, if $X_i \sim N(\theta_i, \sigma_0^2)$ given θ_i with a known $\sigma_0^2 > 0$, a special case of the Laplace-type kernel, the Bayes estimator of θ_i under the squared-error loss is

$$t_G^*(x) = x + \sigma_0^2 f_G'(x)/f_G(x). \tag{6}$$

Since Robbins' methodologies aim at the ideal Bayes rule (4) with no restriction on $\mathcal{D}$ or G, they are referred to as general EB [Robbins (1980b, 1983)] or nonparametric EB [Morris (1983)]. Further developments of general EB methods were undertaken by Robbins' students Johns (1957, 1961) and Samuel (1963b) and Robbins (1963, 1964, 1983) himself. Important contributions have been made by Miyasawa (1961) and Kagan (1962) on estimation problems, Cogburn (1965, 1967) on stringent EB methods, Rutherford and Krutchkoff (1967) on methods based on estimates of the prior, Meeden (1972) on admissibility, Martz and Krutchkoff (1969) and Wind (1973) on regression, Johns and Van Ryzin (1971, 1972), Singh (1979) and Zhang (1997) on rates of convergence and asymptotic minimaxity, O'Bryan (1976) on problems with nonidentical components and van Houwelingen (1977) on monotone EB estimators. For the closely related demixing problem, that is, estimation of the unknown prior G in the EB setting, see Kiefer and Wolfowitz (1956), Dempster, Laird and Rubin (1977) and Lindsay (1983) for nonparametric maximum likelihood methods, Teicher (1961, 1963) for identifiability, Deely and Kruse (1968) for a minimum distance method and Carroll and Hall (1988), Zhang (1990, 1995) and Fan (1991) for optimal rates of convergence. We further refer to the survey paper by Copas (1969) and the books by Maritz and Lwin (1989) and Carlin and Louis (1996).

4. Parametric and restricted EB methods. An early development of major importance was Stein's (1956) proof of the inadmissibility of the maximum likelihood (best equivariant) estimator of a multivariate normal mean and the subsequent, more precise results of James and Stein (1961). These papers led to the parametric EB approach and a tremendous amount of research in the fields of multivariate estimation and admissibility.

In the framework of compound decision theory, Stein's (1956) problem is the estimation of θ_i with squared-error loss based on $X_i \sim N(\theta_i, \sigma_0^2)$. James and Stein (1961) proposed $\widehat{\theta}_i = (1 - B_n)X_i$, with $B_n = \sigma_0^2(n-2)/\sum_{i=1}^n X_i^2$, and proved $R_n(\widehat{\theta}_{(n)}, \theta_{(n)}) < R_n(X_{(n)}, \theta_{(n)})$ for all $\theta_{(n)} \in \mathbb{R}^n$ and $n \geq 3$, where $R_n(\delta_{(n)}, \theta_{(n)})$ is the compound risk in (1). This also demonstrated the advantage of using the whole sequence $X_{(n)}$ to estimate θ_i versus using X_i alone. Stein's inadmissibility result is of nonasymptotic nature, and (3) holds for the James–Stein estimator if and only if $G_n \to N(0, \tau^2)$ or $R^*(G_n) \to 1$.

In a series of seminal papers, Efron and Morris (1972a, b, 1973a, b) provided an EB interpretation of the James–Stein estimator and proposed a more parsimonious EB approach. We now describe this development. For $G \sim N(0, \tau^2)$, the Bayes estimator is $t_G^*(x) = (1 - B)x$ with shrinkage factor $B = \sigma_0^2/(\sigma_0^2 + \tau^2)$. Thus, under the working assumption of the normality of the unknown prior G, the James–Stein estimator is EB in the sense of (5), with $\widehat{t}_n(x) = (1 - B_n)x$, since $B_n \to B$. This is what Morris (1983) called the parametric EB, since the working assumption is characterized by a regular finite-dimensional model for the unknown

prior. As in James and Stein (1961), the performance of parametric EB procedures is often evaluated in the more general compound and nonparametric EB settings.

For related contributions in multivariate influence, we refer to Brown (1966) on general loss and location parameters, Strawderman (1971) on proper Bayes minimax shrinkage, Berger (1980) on robust methods, George (1986) on minimax multiple shrinkage, Efron (1996) on combining likelihoods, George and Foster (2000) on variable selection and Berger's (1985) book.

The James–Stein estimator can be also viewed as linear EB [Robbins (1983)], since the normality assumption on G in (4) could be replaced by the linearity restriction on $\mathcal{D}$ to produce the same type of ideal Bayes rules. EB methods with restricted $\mathcal{D}$ in (4), called restricted EB [Robbins (1980b, 1983, 1985)], could aim at certain simple non-Bayesian rules, for example, the class of all thresholding estimators [Donoho and Johnstone (1995)].

5. Applications to high-dimensional problems. EB methodologies apply naturally to problems with high- and infinite-dimensional data. An important example, discussed in Efron, Storey and Tibshirani (2001), Efron, Tibshirani, Storey and Tusher (2001) and Efron (2003), is the analysis of microarrays. These papers also provided a connection between general EB and Benjamini and Hochberg's (1995) method of controlling the false discovery rate in multiple comparisons. Here we describe how EB estimators work, and in certain senses work much better than classical smoothing methods, in nonparametric regression.

Suppose we observe (x_i, y_i) with $y_i = f(x_i) + \varepsilon_i$, $i = 1, \ldots, n$, where $x_i = i/n$ and ε_i are i.i.d. $N(0, \sigma_0^2)$ errors with a known σ_0. Consider the estimation of the unknown f under risk

$$R_n(\widehat{f}_n, f) \equiv \frac{1}{n}\sum_{i=1}^{n} E\big(\widehat{f}_n(x_i) - f(x_i)\big)^2. \tag{7}$$

Although (7) is of the same form as (1), smoothness properties of the unknown f are best exploited under different coordinate systems. Let $\{\psi_{k,n}\}$ be suitable orthonormal (e.g., discrete Fourier or wavelet) bases in $\mathbb{R}^n$. Define $z_{k,n} \equiv \psi_{k,n}^{\mathrm{tr}} y_{(n)}$ and $\theta_{k,n} \equiv \psi_{k,n}^{\mathrm{tr}} f_{(n)}$, $k = 1, \ldots, n$, where $f_{(n)} \equiv (f(x_1), \ldots, f(x_n))^{\mathrm{tr}}$. Estimation of $f_{(n)}$ based on $y_{(n)}$ is equivalent to estimation of $\{\theta_{k,n}\}$ based on $\{z_{k,n}\}$, since $R_n(\widehat{f}_n, f) = n^{-1}\sum_{k=1}^{n} E(\widehat{\theta}_{k,n} - \theta_{k,n})^2$ with $\widehat{f}_{(n)} = \sum_{k=1}^{n} \widehat{\theta}_{k,n}\psi_{k,n}$. The advantage of the z–θ representation is that smoothness conditions on f often imply $\theta_{k,n} \sim \sqrt{n}k^{-(\alpha+1/2)}$ in a certain sense with a smoothness index α. It follows that $R_n(\widehat{f}_n, f) = O(1)n^{-\alpha/(\alpha+1/2)}$ if we essentially estimate $n - O(n^{1/(2\alpha+1)})$ of those in $\{\theta_{k,n} : |\theta_{k,n}| \le \sigma_0\}$ by zero and estimate the remaining, mostly larger $\theta_{k,n}$ by $z_{k,n}$. The problem is adaptation to different types of smoothness conditions and smoothness indices α.

Block EB estimators of f are constructed by dividing $\{1, \ldots, n\}$ into a number of blocks and then applying EB methods in individual blocks. More precisely, let

$k_0 < \cdots < k_m = n$ and let $[j] \equiv (k_{j-1}, k_j]$ be m blocks. Block EB estimators are of the form $\widehat{\theta}_{k,n} = \widehat{t}_{[j]}(z_{k,n})$, where $\widehat{t}_{[j]}(x)$ are estimates of certain ideal Bayes rules $t^*_{[j]}(x)$ for the jth block. Direct application of the James–Stein (JS) estimator yields block JS estimators $\widehat{\theta}_{k,n} = (1 - B_{[j]})z_{k,n}$ for $k \in [j]$, where $B_{[j]} = \sigma_0^2(\#[j]-2)/\sum_{k\in[j]} z_{k,n}^2$. For block general EB estimators, $\widehat{t}_{[j]}(x)$ approximate (6) with $G(A) = \sum_{k\in[j]} I\{\theta_{k,n} \in A\}/(\#[j])$, since $z_{k,n} \sim N(\theta_{k,n}, \sigma_0^2)$. From this point of view, Efromovich and Pinsker's (1984) estimators are block linear EB, and Donoho and Johnstone's (1995) are block threshold EB. Block general EB methods were developed in Zhang (2000).

Let us describe the adaptive minimaxity and superefficiency of block EB estimators. Consider linear EB methods for simplicity. Let $\{\psi_{k,n}\}$ be the discrete Fourier bases and let the block sizes satisfy $\max_{2\le j\le m} k_j/k_{j-1} \to 1$ and $m/n^\varepsilon \to 0$ for all $\varepsilon > 0$. The block JS estimators are exactly adaptive minimax in the sense that

$$\frac{\sup\{R_n(\widehat{f}_n, f) : f \in \mathcal{F}_{\alpha,c}\}}{\inf_{\tilde{f}_n} \sup\{R_n(\tilde{f}_n, f) : f \in \mathcal{F}_{\alpha,c}\}} = 1 + o(1) \tag{8}$$

simultaneously for all Sobolev balls $\mathcal{F}_{\alpha,c} \equiv \{f : \|f\|_{(\alpha)} \le c\}$, $\alpha > 0$, $c > 0$, where $\|f\|_{(\alpha)} \equiv \limsup_n \sqrt{\sum_{k=1}^n k^{2\alpha}|\theta_{k,n}|^2/n}$. This is the main result of Efromovich and Pinsker (1984). In fact, they considered modified JS estimators to cover non-Gaussian $z_{k,n}$. For positive integers α, $\|f\|^2_{(\alpha)} = \int_0^1 |(d/dx)^\alpha f|^2\,dx$ for functions with period 1. Compared to the minimax rate $\inf_{\tilde{f}_n} \sup\{R_n(\tilde{f}_n, f) : f \in \mathcal{F}_{\alpha,c}\} \sim n^{\alpha/(\alpha+1/2)}$, the block JS estimators are also superefficient in the sense that

$$\lim_{n\to\infty} n^{\alpha/(\alpha+1/2)} \sup\bigl\{R_n(\widehat{f}_n, f) : f \in \mathcal{C}\bigr\} = 0 \tag{9}$$

for all compact sets $\mathcal{C}$ under the $\|f\|_{(\alpha)}$ norm. These properties are much stronger than the standard results, with $O(1)$ on the right-hand side of (8) and $M(\mathcal{C}) > 0$ on the right-hand side of (9), for classical smoothing methods with optimal bandwidth and penalty parameters.

The literature in nonparametric estimation is one of the richest in statistics; cf. Ibragimov and Khasminskii (1981), Breiman, Friedman, Olshen and Stone (1984), Wahba (1990), Friedman (1991), Stone (1994), Donoho, Johnstone, Kerkyacharian and Picard (1995), Fan and Gijbels (1996), Barron, Birgé and Massart (1999), Efromovich (1999), Hastie, Tibshirani and Friedman (2001) and the references therein.

6. Prediction and related problems. Robbins (1977, 1980a, b) extended the EB methodology to the prediction of sums of the form $S_n \equiv \sum_{i=1}^n Y_i u(X_i)$ based on $\{X_i\}$ and statistical inference based on biased allocation schemes.

For example, given a pool of motorists, how do we predict the total number of traffic accidents next year for those in the pool with a prespecified number of accidents this year, for example, those with clean records? If the overall traffic condition is unchanged, it is appropriate to assume that, for the ith individual in the pool, the number of accidents X_i for the current year and Y_i for the next are independent Poisson variables with a common mean θ_i. The problem is the prediction of S_n with $u(x) = I\{x = a\}$ for a prespecified integer $a \geq 0$.

In the EB setting, the θ_i are assumed to be independent with an unknown distribution G and the ideal Bayes predictor under squared-error loss is $T_n^* = \sum_{i=1}^n t_G^*(X_i)u(X_i)$, where $t_G^*(x) = E_G[\theta_i | X_i = x]$. General and restricted EB predictors of the form $T_n = \sum_{i=1}^n \widehat{t}_n(X_i)u(X_i)$ were considered by Robbins (1980a, b). In the above example, the general EB estimator $\widehat{t}_n(x) = (x+1) \times f_n(x+1)/f_n(x)$ of Robbins (1956) provides the predictor $T_n = \sum_{i=1}^n v(X_i)$ with $v(x) = xu(x-1)$, where $f_n(x)$ is the observed frequency of $\{X_i\}$ at x. The predictor T_n is unbiased for deterministic θ_i and asymptotically efficient for i.i.d. θ_i [Robbins and Zhang (2000)]. The variance $E(T_n - S_n)^2$ can be estimated by EB methods to produce prediction intervals for S_n [Robbins (1977)].

In related biased-allocation problems, treatment groups are assigned according to the pre-treatment X_i, and the distributions of post-treatment Y_i, given (X_i, θ_i), depend on both the treatment and θ_i. For example, if only those motorists with a accidents are treated, for example, with a reduction in insurance premium for $a = 0$, and the treatment effect is a multiplicative factor λ in the Poisson model, that is, $E[Y_i | X_i = a, \theta_i] = \lambda\theta_i$, the general EB method gives $\widehat{\lambda} = S_n/T_n$, based only on those Y_i in the treatment group. Methodologies for problems related to this type of biased allocation have been further developed by Robbins (1982, 1988), Robbins and Zhang (1988, 1989, 1991), Finkelstein, Levin and Robbins (1996a, b) and Levin, Robbins and Zhang (2002).

The above Poisson prediction problem is also related to the estimation of the total probability of unobserved outcomes, for example, fish in a lake, bugs in a piece of software, in multinomial models. Let $\{X_j\}$ be a multinomial vector with parameters m and $\{p_j\}$. The total probability of unobserved outcomes is $\eta \equiv \sum_j p_j I\{X_j = 0\}$. If $\{X_j\}$ is viewed as a vector of independent Poisson variables conditionally on $\sum_j X_j = m$, then the general EB solution T_n for the Poisson prediction problem above leads to $\widehat{\eta} = \sum_j I\{X_j = 1\}/m$. See Good (1953) and Robbins (1968).

7. Final remark. Herbert Robbins' legendary career was characterized by his great originality and power, the elegance of his ideas and his way of communicating them through surprising but simple examples. His work in compound decision theory and empirical Bayes was no exception. His contributions, spirit and wisdom have benefited many tremendously, including myself, and will undoubtedly continue to do so for generations to come.

REFERENCES

BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300.

BERGER, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8** 716–761.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA.

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

CARLIN, B. P. and LOUIS, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, New York.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

CARTER, G. and ROLPH, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *J. Amer. Statist. Assoc.* **69** 880–885.

COGBURN, R. (1965). On the estimation of a multivariate location parameter with squared error loss. In *Bernoulli, Bayes, and Laplace Anniversary Volume* (J. Neyman and L. Le Cam, eds.) 24–29. Springer, Berlin.

COGBURN, R. (1967). Stringent solutions to statistical decision problems. *Ann. Math. Statist.* **38** 447–463.

COPAS, J. B. (1969). Compound decisions and empirical Bayes (with discussion). *J. Roy. Statist. Soc. Ser. B* **31** 397–425.

COPAS, J. B. (1972). Empirical Bayes methods and the repeated use of a standard. *Biometrika* **59** 349–360.

COVER, T. (1968). Learning in pattern recognition. In *Methodologies of Pattern Recognition* (S. Watanabe, ed.) 111–132. Academic Press, New York.

COVER, T. (1991). Universal portfolios. *Math. Finance* **1** 1–29.

DATTA, S. (1991). Asymptotic optimality of Bayes compound estimators in compact exponential families. *Ann. Statist.* **19** 354–365.

DEELY, J. J. and KRUSE, R. L. (1968). Construction of sequences estimating the mixing distribution. *Ann. Math. Statist.* **39** 286–288.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the *EM* algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57** 301–369.

EFROMOVICH, S. (1999). *Nonparametric Curve Estimation.* Springer, New York.

EFROMOVICH, S. and PINSKER, M. S. (1984). An adaptive algorithm of nonparametric filtering. *Autom. Remote Control* **11** 58–65.

EFRON, B. (1996). Empirical Bayes methods for combining likelihoods. *J. Amer. Statist. Assoc.* **91** 538–550.

EFRON, B. (2003). Robbins, empirical Bayes and microarrays. *Ann. Statist.* **31** 366–378.

EFRON, B. and MORRIS, C. (1972a). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika* **59** 335–347.

EFRON, B. and MORRIS, C. (1972b). Limiting the risk of Bayes and empirical Bayes estimators. II. The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

EFRON, B. and MORRIS, C. (1973a). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130.

EFRON, B. and MORRIS, C. (1973b). Combining possibly related estimation problems (with discussion). *J. Roy. Statist. Soc. Ser. B* **35** 379–421.

EFRON, B. and MORRIS, C. (1977). Stein's paradox in statistics. *Sci. Amer.* **236** 119–127.

EFRON, B., STOREY, J. and TIBSHIRANI, R. (2001). Microarrays, empirical Bayes methods, and false discovery rates. Technical report, Stanford Univ. Available online at http://www-stat.stanford.edu/~tibs/research.html.

EFRON, B., TIBSHIRANI, R., STOREY, J. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160.

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman and Hall, London.

FINKELSTEIN, M. O., LEVIN, B. and ROBBINS, H. (1996a). Clinical and prophylactic trials with assured new treatment for those at greater risk. I. A design proposal. *Amer. J. Public Health* **86** 691–695.

FINKELSTEIN, M. O., LEVIN, B. and ROBBINS, H. (1996b). Clinical and prophylactic trials with assured new treatment for those at greater risk. II. Examples. *Amer. J. Public Health* **86** 696–705.

FOX, R. J. (1970). Estimating the empiric distribution function of certain parameter sequences. *Ann. Math. Statist.* **41** 1845–1852.

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–67.

GEORGE, E. I. (1986). Minimax multiple shrinkage estimation. *Ann. Statist.* **14** 188–205.

GEORGE, E. I. and FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747.

GILLILAND, D. C. (1968). Sequential compound estimation. *Ann. Math. Statist.* **39** 1890–1905.

GILLILAND, D. C. and HANNAN, J. F. (1969). On an extended compound decision problem. *Ann. Math. Statist.* **40** 1536–1541.

GILLILAND, D. C., HANNAN, J. F. and HUANG, J. S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. *Ann. Statist.* **4** 1101–1112.

GOOD, I. J. (1953). On the population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264.

HANNAN, J. F. (1957). Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games* (M. Dresher and A. W. Tucker, eds.) **3** 97–139. Princeton Univ. Press.

HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37–51.

HANNAN, J. F. and VAN RYZIN, J. R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743–1752.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. H. (2001). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*. Springer, New York.

HOADLEY, B. (1981). The quality management plan (QMP). *Bell System Tech. J.* **60** 215–273.

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory.* Springer, New York.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

JOHNS, M. V. (1957). Non-parametric empirical Bayes procedures. *Ann. Math. Statist.* **28** 649–669.

JOHNS, M. V. (1961). An empirical Bayes approach to non-parametric two-way classification. In *Studies in Item Analysis and Prediction* (H. Solomon, ed.) 221–232. Stanford Univ. Press.

JOHNS, M. V. and VAN RYZIN, J. R. (1971). Convergence rates for empirical Bayes two-action problems. I. Discrete case. *Ann. Math. Statist.* **42** 1521–1539.

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

KAGAN, A. M. (1962). An empirical Bayes approach to the estimation problem. *Dokl. Acad. Sci. U.S.S.R.* **147** 1020–1021.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

LEVIN, B., ROBBINS, H. and ZHANG, C.-H. (2002). Mathematical aspects of estimating two treatment effects and a common variance in an assured allocation design. *J. Statist. Plann. Inference* **108** 255–262.

LINDSAY, B. G. (1983). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94.

MARITZ, J. S. and LWIN, T. (1989). *Empirical Bayes Methods*, 2nd ed. Chapman and Hall, London.

MARTZ, H. and KRUTCHKOFF, R. G. (1969). Empirical Bayes estimators in a multiple linear regression model. *Biometrika* **56** 367–374.

MEEDEN, G. (1972). Some admissible empirical Bayes procedures. *Ann. Math. Statist.* **43** 96–101.

MIYASAWA, K. (1961). An empirical Bayes estimator of the mean of a normal population. *Bull. Internat. Statist. Inst.* **38** 181–188.

MORRIS, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–55.

NEYMAN, J. (1962). Two breakthroughs in the theory of statistical decision-making. *Rev. Internat. Statist. Inst.* **30** 11–27.

OATEN, A. (1972). Approximation to Bayes risk in compound decision problems. *Ann. Math. Statist.* **43** 1164–1184.

O'BRYAN, T. E. (1976). Some empirical Bayes results in the case of component problems with varying sample sizes for discrete exponential families. *Ann. Statist.* **4** 1290–1293.

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 131–148. Univ. California Press, Berkeley.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163. Univ. California Press, Berkeley.

ROBBINS, H. (1962). Some numerical results on a compound decision problem. In *Recent Developments in Information and Decision Processes* (R. E. Machol and P. Gray, eds.) 52–62. Macmillan, New York.

ROBBINS, H. (1963). The empirical Bayes approach to testing statistical hypotheses. *Rev. Internat. Statist. Inst.* **31** 195–208.

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

ROBBINS, H. (1968). Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Math. Statist.* **39** 256–257.

ROBBINS, H. (1977). Prediction and estimation for the compound Poisson distribution. *Proc. Nat. Acad. Sci. U.S.A.* **74** 2670–2671.

ROBBINS, H. (1980a). Estimation and prediction for mixtures of the exponential distribution. *Proc. Nat. Acad. Sci. U.S.A.* **77** 2382–2383.

ROBBINS, H. (1980b). An empirical Bayes estimation problem. *Proc. Nat. Acad. Sci. U.S.A.* **77** 6988–6989.

ROBBINS, H. (1982). Estimating many variances. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta, ed.) **2** 251–261. Academic Press, New York.

ROBBINS, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11** 713–723.

ROBBINS, H. (1985). Linear empirical Bayes estimation of means and variances. *Proc. Nat. Acad. Sci. U.S.A.* **82** 1571–1574.

ROBBINS, H. (1988). The u, v method of estimation. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) **1** 265–270. Springer, New York.

ROBBINS, H. and SAMUEL, E. (1962). Testing statistical hypotheses; the compound approach. In *Recent Developments in Information and Decision Processes* (R. E. Machol and P. Gray, eds.) 63–70. Macmillan, New York.

ROBBINS, H. and ZHANG, C.-H. (1988). Estimating a treatment effect under biased sampling. *Proc. Nat. Acad. Sci. U.S.A.* **85** 3670–3672.

ROBBINS, H. and ZHANG, C.-H. (1989). Estimating the superiority of a drug to a placebo when all and only those patients at risk are treated with the drug. *Proc. Nat. Acad. Sci. U.S.A.* **86** 3003–3005.

ROBBINS, H. and ZHANG, C.-H. (1991). Estimating a multiplicative treatment effect under biased allocation. *Biometrika* **78** 349–354.

ROBBINS, H. and ZHANG, C.-H. (2000). Efficiency of the u, v method of estimation. *Proc. Nat. Acad. Sci. U.S.A.* **97** 12976–12979.

RUBIN, D. (1980). Using empirical Bayes techniques in the law school validity studies (with discussion). *J. Amer. Statist. Assoc.* **75** 801–827.

RUTHERFORD, J. R. and KRUTCHKOFF, R. G. (1967). The empirical Bayes approach: Estimating the prior distribution. *Biometrika* **54** 326–328.

SAMUEL, E. (1963a). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079–1094.

SAMUEL, E. (1963b). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Math. Statist.* **34** 1370–1385.

SAMUEL, E. (1964). Convergence of the losses of certain decision rules for the sequential compound decision problem. *Ann. Math. Statist.* **35** 1606–1621.

SAMUEL, E. (1965). Sequential compound estimators. *Ann. Math. Statist.* **36** 879–889.

SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209.

SINGH, R. S. (1979). Empirical Bayes estimation in Lebesgue exponential families with rates near the best possible rate. *Ann. Statist.* **7** 890–902.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–171.

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.

SUSARLA, V. (1974). Rate of convergence in the sequence-compound squared-distance loss estimation problem for a family of m-variate normal distributions. *Ann. Statist.* **2** 118–133.

SUSARLA, V. and VAN RYZIN, J. R. (1978). Empirical Bayes estimation of a distribution (survival) function from right-censored observations. *Ann. Statist.* **6** 740–754.

TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.

TEICHER, H. (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34** 1265–1269.

VAN HOUWELINGEN, H. C. and THOROGOOD, J. (1995). Construction, validation and updating of a prognostic model for kidney graft survival. *Statistics in Medicine* **14** 1999–2008.

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

VAN RYZIN, J. R. (1966a). The sequential compound decision problem with $m \times n$ finite loss matrix. *Ann. Math. Statist.* **37** 954–975.

VAN RYZIN, J. R. (1966b). Repetitive play in finite statistical games with unknown distributions. *Ann. Math. Statist.* **37** 976–994.

VAN RYZIN, J. R. and SUSARLA, V. (1977). On the empirical Bayes approach to multiple decision problems. *Ann. Statist.* **5** 172–181.

VARDEMAN, S. B. (1978). Admissible solutions of finite state sequence compound decision problems. *Ann. Statist.* **6** 673–679.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

WIND, S. L. (1973). An empirical Bayes approach to multiple linear regression. *Ann. Statist.* **1** 93–103.

ZASLAVSKY, A. M. (1993). Combining census, dual-system, and evaluation study data to estimate population shares. *J. Amer. Statist. Assoc.* **88** 1092–1105.

ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

ZHANG, C.-H. (1995). On estimating mixing densities in discrete exponential family models. *Ann. Statist.* **23** 929–945.

ZHANG, C.-H. (1997). Empirical Bayes and compound estimation of normal means. *Statist. Sinica* **7** 181–193.

ZHANG, C.-H. (2000). General empirical Bayes wavelet methods. Technical Report 2000-007, Dept. Statist., Rutgers Univ., Piscataway, NJ.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER, BUSCH CAMPUS
PISCATAWAY, NEW JERSEY 08854
E-MAIL: czhang@stat.rutgers.edu