The Annals of Mathematical Statistics
1971, Vol. 42, No. 5, 1521–1539

CONVERGENCE RATES FOR EMPIRICAL BAYES TWO-ACTION PROBLEMS I. DISCRETE CASE

BY M. V. JOHNS, JR.[1] AND J. VAN RYZIN[2]

Stanford University and The University of Wisconsin

1. Introduction and summary. Situations involving sequences of similar but independent statistical decision problems arise in many areas of application. Routine bioassay (Chase (1966)) and lot by lot acceptance sampling are typical examples of such situations. In many instances it is reasonable to formulate the independent component problems of such a sequence as Bayes statistical decision problems involving a common, but completely unknown, prior probability distribution over the state space. Robbins (1955) has shown for certain estimation problems that the accumulated information acquired as the sequence of problems progresses may be used to improve the decision rule at each stage. Such "empirical Bayes" procedures may be asymptotically optimal in the sense that the risk for the nth decision problem converges to the Bayes optimal risk which would have been obtained if the prior distribution were *known* and the best decision rule based on this knowledge were used.

Johns (1957) exhibits asymptotically optimal empirical Bayes procedures for certain two-action (hypothesis testing) problems as well as for estimation problems in a nonparametric context. Robbins (1963) and Samuel (1963) consider parametric two-action problems where the distributions of the observations are members of a specified exponential family, and where the special loss functions of Johns (1957) are used. Robbins and Samuel each exhibit asymptotically optimal empirical Bayes procedures for both discrete and continuous observations.

The usefulness of empirical Bayes procedures in practical statistical applications clearly depends on the rapidity with which the risks incurred for the successive decision problems approach the optimal limit. The purpose of this paper and its sequel (Johns and Van Ryzin (1967)) is to investigate rates of convergence to optimality of empirical Bayes procedures for two-action decision problems when the distributions of the observations are of exponential type. The present paper considers discrete exponential families which include for example the geometric, the negative binomial, and the Poisson distributions. The sequel (Johns and Van Ryzin (1967)) considers continuous exponential families with particular emphasis on the normal and the negative exponential distributions.

Each component problem in the sequence of decision problems for which an empirical Bayes procedure is to be defined is assumed to have the following

Received October 23, 1967; revised August 24, 1970.

[1] Supported by Public Health Service Grant USPHS-GM-14554-01, Stanford University.

[2] Supported in part by Office of Naval Research Contract Nonr 225(52) (NR 342–022), Stanford University, and in part by National Science Foundation Grant NSF GP9324, University of Wisconsin.

structure: An observation X is obtained having a distribution with probability mass function

$$p_\lambda(x) = h(x)\lambda^x\beta(\lambda), \qquad x = 0, 1, \cdots; 0 \leqq \lambda < d, \tag{1}$$

where $h(x) > 0$ for all x and where d may be finite or infinite. The observation X may be thought of as the value of a sufficient statistic based on several i.i.d. observations. The hypothesis H_1: $\lambda \leqq c$, $c > 0$ is to be tested against H_2: $\lambda > c$ with loss function

$$\begin{aligned} L_1(\lambda) &= 0 && \text{if } \lambda \leqq c \\ &= b(\lambda - c) && \text{if } \lambda > c, b > 0, \\ L_2(\lambda) &= b(c-\lambda) && \text{if } \lambda \leqq c \\ &= 0 && \text{if } \lambda > c, \end{aligned}$$

where $L_i(\lambda)$ indicates the loss when action i (deciding in favor of H_i) is taken, $i = 1, 2$, and λ is the true value of the parameter. It is assumed that λ may be regarded as the value of a random variable Λ having prior distribution function $G(\lambda)$. If the randomized decision rule $\delta(x) = \Pr\{\text{Accepting } H_1 \text{ given } X = x\}$ is used, then the risk incurred is

$$\begin{aligned} r(\delta, G) &= \int \textstyle\sum_x \{L_1(\lambda)p_\lambda(x)\delta(x) + L_2(\lambda)p_\lambda(x)(1-\delta(x))\}\, dG(\lambda) \\ &= b \textstyle\sum_x \alpha(x)\delta(x) + C_G, \end{aligned} \tag{2}$$

where $\qquad C_G = \int L_2(\lambda)\, dG(\lambda)$ and

$$\alpha(x) = \int \lambda p_\lambda(x)\, dG(\lambda) - cp(x), \tag{3}$$

where $p(x)$ is the unconditional probability mass function for X and is given by

$$p(x) = \int p_\lambda(x)\, dG(\lambda), \qquad x = 0, 1, \cdots. \tag{4}$$

We consider only priors G such that $E\Lambda < \infty$ to insure that the risk is always finite.

A Bayes rule (i.e., a minimizer of (2) based on knowledge of G) is clearly given by

$$\begin{aligned} \delta_G(x) &= 1, && \alpha(x) \leqq 0 \\ &= 0, && \alpha(x) > 0. \end{aligned} \tag{5}$$

The resulting (minimal) Bayes risk is

$$r^*(G) = \inf_\delta r(\delta, G) = r(\delta_G, G). \tag{6}$$

For the case where $p_\lambda(x)$ is given by (1) it is easily verified that

$$\alpha(x) = w(x)p(x+1) - cp(x), \tag{7}$$

where $p(x)$ is given by (4) and

$$w(x) = \frac{h(x)}{h(x+1)}. \tag{8}$$

In the empirical Bayes context, a sequence of problems having the above structure occurs but $G(\lambda)$ is *not* assumed to be known. However, for the $(n+1)$st problem additional information in the form of the observations $X_1, X_2, \cdots, X_n$ obtained in the previous problems is available. The empirical Bayes procedures considered here involve the construction of a sequence of estimates $\alpha_n(x)$, $n = 1, 2, \cdots$ of the function $\alpha(x)$ where $\alpha_n(x)$ is based on the observations $X_1, X_2, \cdots, X_n$. The decision rule used for the $(n+1)$st decision problem is then

$$\begin{aligned} \delta_n(x) &= 1, & \alpha_n(x) &\leqq 0 \\ &= 0, & \alpha_n(x) &> 0. \end{aligned} \tag{9}$$

This rule imitates (5) but does not require knowledge of the prior G as long as $\alpha_n(x)$ does not depend on G. Specifically, if $p_\lambda(x)$ is given by (1) so that $\alpha(x)$ is of the form (7), we let

$$\alpha_n(x) = n^{-1} \textstyle\sum_{j=1}^n Z_j(x), \tag{10}$$

where for each x, j,

$$Z_j(x) = w(x)U_j(x+1) - cU_j(x) \tag{11}$$

where

$$\begin{aligned} U_j(x) &= 1, & X_j &= x \\ &= 0, & X_j &\neq x. \end{aligned}$$

For a given x the $Z_j(x)$'s are i.i.d. and $EZ_j(x) = \alpha(x)$. Letting $r_n =$ the risk in the $(n+1)$st problem using the decision rule (9) it is clear that $r_n - r^*(G)$ is nonnegative and it is easily shown (Robbins (1963), Samuel (1963)) that $r_n - r^*(G) \to 0$, as $n \to \infty$, i.e., that δ_n is asymptotically optimal, provided only that $E\Lambda < \infty$.

In Section 2 we give a very simple argument which provides an upper bound on the rate at which r_n approaches $r^*(G)$ as n becomes large. This result involves conditions only on the unconditional probabilities $p(x)$, which may be rephrased in terms of the existence of moments of $G(\lambda)$. Although more refined results are obtained in later sections the argument used in Section 2 forms the basis of the methods used in Johns and Van Ryzin (1967) to treat the continuous case where the technical difficulties are considerably more formidable. Thus a comparison of the results of this section with the more precise results of Section 4 provides an indication of the relative precision of the convergence rates obtained in Johns and Van Ryzin (1967).

Section 3 contains the development of the asymptotic tools necessary to obtain the exact rates of convergence to optimality presented in Section 4.

Theorem 3 of Section 4 deals with the case where the natural parameter space is compact. Without loss of generality this space is taken to be the unit interval. The function $h(x)$ appearing in (1) is taken to be of the form $h(x) = h_1(x)x^\gamma$, where $h_1(x)$ is a positive slowly varying function (i.e., $h_1(cx)/h_1(x) \to 1$ as $x \to \infty$ for any $c > 0$). This theorem shows that rates of convergence as bad as $n^{-\varepsilon}$ for ε arbitrarily small may in principle be obtained when $G'(\lambda)$ behaves like $(1-\lambda)^\delta$ for λ close to one, for sufficiently small δ. However, if $G'(\lambda)$ decreases exponentially as $\lambda \to 1$, the rate becomes n^{-1} multiplied by a slowly varying function. If $G'(\lambda)$ is zero in some interval $(\lambda_0, 1)$, $\lambda_0 < 1$, then the rate n^{-1} is attained and this is best possible for procedures of the form (9), (10).

Theorem 4 deals with the Poisson case and similar results are obtained for the cases where $G'(\lambda)$ behaves like λ^δ or $\lambda^\delta e^{-\sigma\lambda}$, $\sigma > 0$, as $\lambda \to \infty$. For the Poisson case and for other cases where the parameter space is not compact the rate n^{-1} is unattainable and a slowly varying factor (such as log n) always appears.

A comparison of these results with those of Section 2 shows that the more elementary bounds are not far from being ratewise sharp. It should perhaps be emphasized that without the exact rates obtained in Section 4 it would be impossible to evaluate the degree of precision of the elementary rate results of Section 2 and of the analogous but less elementary results for the continuous case given in Johns and Van Ryzin (1967).

2. Preliminary results. In order to establish a simple upper bound for the rate of approach of r_n to $r^*(G)$ we first observe that, since r_n is the risk associated with the decision rule δ_n given by (9), we have, recalling (2),

$$\begin{aligned} r_n &= b\textstyle\sum_x \alpha(x)E\delta_n(x) + C_G \\ &= b\textstyle\sum_x \alpha(x)\Pr\{\alpha_n(x) \leqq 0\} + C_G. \end{aligned}$$

We now state a lemma, useful here and in Johns and Van Ryzin (1967).

LEMMA 1. *If $r^*(G)$ is given by* (6), *then*

$$0 \leqq r_n - r^*(G) \leqq b\textstyle\sum_x |\alpha(x)| \Pr\{|\alpha_n(x)-\alpha(x)| \geqq |\alpha(x)|\}.$$

PROOF. Recalling (5) we have

$$\begin{aligned} r_n - r^*(G) &= b\textstyle\sum_x \alpha(x)(\Pr\{\alpha_n(x) \leqq 0\} - \delta_G(x)) \\ &= b\textstyle\sum_x |\alpha(x)|\Delta_n(x), \end{aligned}$$

where

$$\begin{aligned} \Delta_n(x) &= \Pr\{\alpha_n(x) > 0\} \qquad \text{if} \quad \alpha(x) \leqq 0 \\ &= \Pr\{\alpha_n(x) \leqq 0\} \qquad \text{if} \quad \alpha(x) > 0. \end{aligned}$$

The desired result follows from the fact that the event $\{|\alpha_n(x)-\alpha(x)| \geqq |\alpha(x)|\}$ is implied by $\{\alpha_n(x) > 0\}$ when $\alpha(x) \leqq 0$, and by $\{\alpha_n(x) \leqq 0\}$ when $\alpha(x) > 0$.

The following theorem and its corollary provide simple conditions guaranteeing specified rates of convergence. Let

$$\sigma^2(x) = \operatorname{Var}(Z_j(x)). \tag{12}$$

THEOREM 1. *If for some* $\delta, 0 < \delta < 2$, *there exists a constant* $K > 0$ *such that*

$$\textstyle\sum_{x \in S} |\alpha(x)|^{1-\delta}[\sigma(x)]^{\delta} < K < \infty,$$

where $S = \{x : \alpha(x) \neq 0\}$, *then*

$$0 \leqq r_n - r^*(G) \leqq Kbn^{-\frac{1}{2}\delta}.$$

PROOF. By Lemma 1 and (10) we have

$$\begin{aligned} r_n - r^*(G) &\leqq b \sum_{x \in S} |\alpha(x)| \Pr\left\{\left|\frac{1}{n}\sum_j Z_j(x) - \alpha(x)\right|^{\delta} \geqq |\alpha(x)|^{\delta}\right\} \\ &\leqq b \sum_{x \in S} |\alpha(x)|^{1-\delta} E\left|\frac{1}{n}\sum_j (Z_j(x) - \alpha(x))\right|^{\delta} \\ &\leqq b \sum_{x \in S} |\alpha(x)|^{1-\delta}\left(\frac{\sigma^2(x)}{n}\right)^{\frac{1}{2}\delta}, \end{aligned}$$

which yields the desired result.

COROLLARY 1. *If* $p_\lambda(x)$ *and* $w(x)$ *are given by* (1) *and* (8) *and if for some* $\delta, 0 < \delta < 2$, *either there exist constants* c_0 *and* c_1 *such that*

(i) $\max[w(x), p(x+1)/p(x)] < c_0 < \infty$,

and

(ii) $|w(x)p(x+1)/p(x) - c| > c_1 > 0$

for all sufficiently large x, *and furthermore*

(iii) $\sum_x [p(x)]^{1-\frac{1}{2}\delta} < \infty$,

or alternatively, there exists a constant c_0 *such that*

(iv) $p(x)/p(x+1) < c_0 < \infty$,

and furthermore

(v) $w(x) \to \infty$, *as* $x \to \infty$,

and

(vi) $\sum_x w(x)[p(x+1)]^{1-\frac{1}{2}\delta} < \infty$,

then there exists a $K > 0$ *such that*

$$0 \leqq r_n - r^*(G) \leqq Kn^{-\frac{1}{2}\delta}.$$

PROOF. By (7), if (i) and (ii) hold then $|\alpha(x)| = e^{O(1)}p(x)$ as $x \to \infty$, and by (11) and (12)

$$\sigma^2(x) \leqq EZ_j^2(x) = w^2(x)p(x+1)+c^2p(x) \leqq K_1p(x)$$

for some $K_1 > 0$ for sufficiently large x. Thus, the desired conclusion follows by Theorem 1 if (i), (ii) and (iii) hold. Alternatively, if (iv) and (v) hold then $|\alpha(x)| \sim w(x)p(x+1)$, and $\sigma^2(x) \sim w^2(x)p(x+1)$ as $x \to \infty$, and if (vi) holds the conclusion follows.

We now discuss two illustrative examples. These examples are also considered again later for comparison purposes using the results of Section 4.

EXAMPLE 1. (The geometric distribution). Suppose that

$$p_\lambda(x) = \lambda^x(1-\lambda), \qquad x = 0, 1, \cdots; 0 \leqq \lambda < 1,$$

and that the prior distribution has probability density function

$$G'(\lambda) = (\gamma+1)(1-\lambda)^\gamma, \qquad 0 \leqq \lambda < 1, \gamma > -1.$$

then

$$p(x) = (\gamma+1)\int_0^1 \lambda^x(1-\lambda)^{\gamma+1}\,d\lambda = \frac{(\gamma+1)\Gamma(x+1)\Gamma(\gamma+2)}{\Gamma(x+\gamma+3)}$$

$$\sim (\gamma+1)\Gamma(\gamma+2)x^{-(\gamma+2)}, \qquad \text{as } x \to \infty.$$

Taking $0 < c < 1$ and noting that $w(x) \equiv 1$ for this case, we see that (i), (ii) and (iii) of the corollary are satisfied for $\delta < 2(\gamma+1)/(\gamma+2)$. Thus, for a given value of γ we are assured of a convergence rate faster than $n^{-(\gamma+1-\varepsilon)/(\gamma+2)}$ for any $\varepsilon > 0$. If γ is taken sufficiently large, the rate becomes arbitrarily close to n^{-1}.

The fact that, for this example, the value of γ also determines which moments of $(1-\Lambda)^{-1}$ exist, suggests that moment conditions can be given to assure any specified rate of convergence. It can in fact be shown that, for the geometric case, a convergence rate of at least $n^{-\frac{1}{2}\delta}$ is achieved provided only that $p(x)/p(x+1) \to 1$ as $x \to \infty$, and $E(1-\Lambda)^{-t} < \infty$, where $t = \delta(1+\varepsilon)/(2-\delta)$, $\varepsilon > 0$. The argument which is based on Corollary 1 is similar to that of Corollary 3.1 of Johns and Van Ryzin (1967) and will not be reproduced here.

EXAMPLE 2. (The Poisson distribution). Let

$$p_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{\Gamma(x+1)}, \qquad x = 0, 1, \cdots; \lambda > 0.$$

Then letting the prior probability density be $G'(\lambda) = e^{-\lambda}$, $\lambda > 0$, we have

$$p(x) = \frac{1}{\Gamma(x+1)}\int_0^\infty \lambda^x e^{-2\lambda}\,d\lambda = \left(\frac{1}{2}\right)^{x+1}$$

Thus, for this case $w(x) = x+1$ and conditions (iv), (v) and (vi) of the corollary are satisfied for every $0 < \delta < 2$. The rate of convergence to optimality is therefore faster than $n^{-1+\varepsilon}$ for any $\varepsilon > 0$. In Section 4 it will be shown that this is really a consequence of the exponentially decreasing tail of $G'(\lambda)$.

3. Bounds and asymptotic propositions. In order to determine exact rates of convergence it is necessary first to develop certain asymptotic results. The main tool used in Section 4 is Theorem 2 given below which provides upper and lower bounds for $r_n - r^*(G)$. For this result we do not assume that $p_\lambda(x)$, $\alpha(x)$ and $Z_j(x)$ are necessarily of the forms given in (1), (7) and (11) respectively. We assume only that for each x the $Z_j(x)$'s are i.i.d. with means $\alpha(x)$ and variances $\sigma^2(x)$.

Let

$$m_x(t) = E\exp\{tZ_1(x)\}, \qquad \ddot{m}_x(t) = \frac{\partial^2}{\partial t^2} m_x(t),$$
$$\pi^+(x) = \Pr\{Z_1(x) > 0\},$$
$$\pi^-(x) = \Pr\{Z_1(x) < 0\}.$$

CONDITION 1. $\alpha(x) = 0(\sigma^2(x))$, as $x \to \infty$.

CONDITION 2. For some $\delta > 0$, $m_x(t)$ exists for each x for all $t \in (-\delta, \delta)$.

CONDITION 3. For all real τ in some open interval containing zero, there exists a positive constant K, independent of x and τ, such that $\ddot{m}_x(\tau|\alpha(x)|/\sigma^2(x)) \leqq K\sigma^2(x)$, for all x.

THEOREM 2. *If $\alpha_n(x)$ is given by* (10) *and if S^+ and S^- are the x sets where $\alpha(x) > 0$ and $\alpha(x) < 0$ respectively, then*

$$(13)\quad r_n - r^*(G) \geqq n\textstyle\sum_{x\in S^+} \alpha(x)\pi^-(x)[1-\pi^+(x)-\pi^-(x)]^{n-1} - n\textstyle\sum_{x\in S^-} \alpha(x)\pi^+(x)[1-\pi^+(x)-\pi^-(x)]^{n-1}.$$

If Conditions 1, 2, and 3 are satisfied, then there exists a positive constant c_0 such that

$$(14)\quad r_n - r^*(G) \leqq b\sum_x |\alpha(x)|\left[1-c_0\frac{\alpha^2(x)}{\sigma^2(x)}\right]^n.$$

PROOF. We first note that, as in Lemma 1,

$$(15)\quad r_n - r^*(G) = b\textstyle\sum_{x\in S^+} \alpha(x)\Pr\{\alpha_n(x) \leqq 0\} - b\textstyle\sum_{x\in S^-}\alpha(x)\Pr\{\alpha_n(x) > 0\}.$$

But

$$\Pr\{\alpha_n(x) \leqq 0\} \geqq \textstyle\sum_{k=1}^n \Pr\{Z_j(x) = 0, j = 1, 2, \cdots, k-1, k+1, \cdots, n; Z_k(x) < 0\}$$
$$= n\pi^-(x)[1-\pi^+(x)-\pi^-(x)]^{n-1}.$$

Similarly,

$$\Pr\{\alpha_n(x) > 0\} \geqq n\pi^+(x)[1-\pi^+(x)-\pi^-(x)]^{n-1},$$

and (13) follows immediately from (15).

Now by Condition 2, for $0 < t < \delta$ we have

$$\Pr\{\alpha_n(x) > 0\} \leqq [m_x(t)]^n, \tag{16}$$

$$\Pr\{\alpha_n(x) \leqq 0\} \leqq [m_x(-t)]^n. \tag{17}$$

For $t \in (-\delta, \delta)$ we may write

$$m_x(t) = 1 + \alpha(x)t + \ddot{m}_x(\theta(x,t))t^2/2, \tag{18}$$

where $\theta(x, t)$ lies between 0 and t. Now $\ddot{m}_x(t) = E\{Z_1{}^2(x)\exp\{tZ_1(x)\}\}$ so that for $t \in (-\delta, \delta)$

$$\begin{aligned} \ddot{m}_x(\theta(x,t)) &\leqq EZ_1{}^2(x) + E\{Z_1{}^2(x)\exp\{tZ_1(x)\}\} \\ &= EZ_1{}^2(x) + \ddot{m}_x(t). \end{aligned} \tag{19}$$

Hence by (18) and (19), for $t \in (-\delta, \delta)$

$$m_x(t) \leqq 1 + \alpha(x)t + (EZ_1{}^2(x) + \ddot{m}_x(t))t^2/2. \tag{20}$$

Now by Condition 1 and the summability of $\alpha(x)$ we have $\alpha^2(x) = 0(\sigma^2(x))$ so that for some $c_1 > 0$, for all x

$$EZ_1{}^2(x) \leqq c_1\sigma^2(x). \tag{21}$$

Hence, by (20), (21) and Condition 3, taking $t_x = \tau|\alpha(x)|/\sigma^2(x)$ for sufficiently small $\tau > 0$, there exists a constant $c_0 > 0$ such that for $x \in S^-$,

$$\begin{aligned} m_x(t_x) &\leqq 1 - \frac{\alpha^2(x)}{\sigma^2(x)}(\tau - \tfrac{1}{2}\tau^2(c_1 + K)) \\ &\leqq 1 - c_0\frac{\alpha^2(x)}{\sigma^2(x)}. \end{aligned} \tag{22}$$

Similarly for $x \in S^+$,

$$m_x(-t_x) \leqq 1 - c_0\frac{\alpha^2(x)}{\sigma^2(x)}. \tag{23}$$

The desired result (14) then follows from (15)–(17), (22) and (23), and the proof of the theorem is complete.

REMARK 1. If $\pi^+(x)$ and $\pi^-(x)$ approach zero as x becomes large it is always possible to choose a sequence of integers x_n such that

$$\max[\pi^+(x_n), \pi^-(x_n)] = n^{-1}e^{O(1)}, \tag{24}$$

provided G is not degenerate at 0 in which case (24) cannot be achieved. Then by (13) we have

$$r_n - r^*(G) \geqq |\alpha(x_n)|\,e^{O(1)}. \tag{25}$$

In the particular case where $p_\lambda(x)$ is given by (1) and $\alpha(x)$ by (7) we have $\pi^+(x) = p(x+1)$ and $\pi^-(x) = p(x)$, and typically $|\alpha(x)| \geqq p(x+1)e^{O(1)} = p(x)e^{O(1)}$. Under these circumstances (25) yields

$$r_n - r^*(G) \geqq n^{-1} e^{O(1)},$$

so that the rate of convergence to optimality of n^{-1} is best possible for the decision procedures under consideration. Instances where this rate is actually achieved are discussed in Section 4.

Inspection of the form of the conclusions of Theorem 2 shows that application of this theorem to specific cases will require results about the asymptotic behavior for large n of quantities of the form

$$\phi(n) = \sum_{x \geqq 0} f(x)[1-g(x)]^n, \tag{26}$$

where $f(x) \geqq 0$ and $0 \leqq g(x) \leqq 1$ for all $x \geqq 0$. We shall assume throughout that $f(x)$ is summable so that $\phi(n)$ always exists. Our investigation of this subject will involve "slowly varying" and "regularly varying" functions and we recall the following standard definition:

DEFINITION 1. A positive function $k(\cdot)$ defined on $(0, \infty)$ is said to be a *slowly varying* (s.v.) function if for any $c > 0$, $k(ct)/k(t) \to 1$ as $t \to 0$ (or alternatively as $t \to \infty$).

We remark that a function of the form $k(t)t^\sigma$ with $-\infty < \sigma < \infty$ is said to be a *regularly varying* function with exponent σ if $k(t)$ is a s.v. function. We shall also need the notion of a function which is slowly varying with respect to another function as detailed in the following:

DEFINITION 2. A positive function $k_1(\cdot)$ on $(0, \infty)$ is said to be *slowly varying with respect to* the positive function $k_2(\cdot)$ if $k_1(tk_2(t))/k_1(t) \to 1$ as $t \to 0$ (or alternatively as $t \to \infty$).

It is easy to construct examples of s.v. functions which are not s.v. with respect to one another. Typical examples of functions which satisfy both Definitions 1 and 2 are logarithms, iterated logarithms and their powers and roots.

Throughout the remainder of this section we shall make use of certain standard results concerning s.v. functions and a fundamental Abelian theorem. These results are now readily available in admirably complete and concise form in Chapter VIII, Sections 8 and 9, and Chapter XIII, Section 5 of Feller (1966). We shall refer directly to the appropriate propositions in Feller (1966) without reproducing them here. In the remainder of this section and in Section 4 we shall make considerable use of the symbol $e^{O(1)}$ indicating a positive function asymptotically bounded away from zero and infinity (the argument of the function being understood from the context). This will enable us to achieve a degree of notational simplification without important loss of precision. In many (but not all) instances, as will be noted in the appropriate contexts, these symbols could be replaced by specific constants.

The general asymptotic properties of $\phi(n)$ needed for our applications are described in Lemmas 2 and 3 below.

For given positive functions $f^*(x)$ and $g^*(x)$ defined for nonnegative integer values of x, let

$$S^*(t) = \{x : g^*(x) \leqq t, x \quad \text{an integer}\} \tag{27}$$

and

$$v^*(t) = \sum_{x \in S^*(t)} f^*(x), \qquad \text{for } t \geqq 0. \tag{28}$$

Then for $\phi(n)$ given by (26) we have

LEMMA 2. *If* $f(x) \sim f^*(x)$ *and* $g(x) \sim g^*(x) \to 0$, *as* $x \to \infty$, *and if* $v^*(t) = e^{O(1)}k(t)t^\beta$, *where* $\beta > 0$ *and* $k(t)$ *is* s.v. *as* $t \to 0$, *then*

$$\phi(n) \sim e^{O(1)}k(n^{-1})n^{-\beta} \qquad \text{as } n \to \infty. \tag{29}$$

Furthermore, if $v^*(t) \sim k(t)t^\beta$ *as* $t \to 0$, *then*

$$\phi(n) \sim \Gamma(\beta+1)k(n^{-1})n^{-\beta} \qquad \text{as } n \to \infty. \tag{30}$$

PROOF. For $t \geqq 0$ let

$$S(t) = \{x : -\log[1-g(x)] \leqq t, x \quad \text{an integer}\}$$

and

$$v(t) = \sum_{x \in S(t)} f(x).$$

Then from (26) we may write

$$\phi(n) = \int_0^\infty e^{-nt}\, dv(t) = n \int_0^\infty e^{-nt} v(t)\, dt. \tag{31}$$

For (29) it suffices to show that $v(t) = e^{O(1)}v^*(t)$, as $t \to 0$, since the desired result then follows immediately from the Abelian conclusion of Theorem 2, page 421 of Feller (1966). (The symbol $e^{O(1)}$ appearing in the asymptotic expression for $v(t)$ can be "brought through the integral sign" in the second integral of (30).) Now for arbitrary fixed $\varepsilon > 0$, there exists an x_ε such that $x > x_\varepsilon$ implies

$$(1-\varepsilon)f^*(x) < f(x) < (1+\varepsilon)f^*(x),$$

and

$$(1-\varepsilon)g^*(x) < -\log[1-g(x)] < (1+\varepsilon)g^*(x).$$

Furthermore $S(t)$ and $S^*(t)$ will contain only x's greater than x_ε for all sufficiently small values of t, so that for such values

$$S^*(t/(1-\varepsilon)) \supset S(t) \supset S^*(t/(1+\varepsilon)),$$

and hence

$$(1-\varepsilon)v^*(t/(1+\varepsilon)) < v(t) < (1+\varepsilon)v^*(t/(1-\varepsilon)),$$

which completes the proof of (29) since ε is arbitrary. The proof of (30) is essentially the same.

The next lemma applies Lemma 2 to the two specific cases of interest in our investigation.

LEMMA 3. *If* (i) $f(x) \sim f_1(x)x^{-s-1}$ *and* $g(x) \sim g_1(x)x^{-r}$ *as* $x \to \infty$, *or* (ii) $f(x) \sim f_1(e^{x^\beta})e^{-sx^\beta}$ *and* $g(x) \sim g_1(e^{x^\beta})e^{-rx^\beta}$ *as* $x \to \infty$, *where* $0 < \beta \leqq 1$, *and where for both cases* $r, s > 0$ *and* f_1, g_1 *are* s.v. *functions with* f_1 *and* g_1 *each* s.v. *with respect to* $[e^{O(1)}g_1]^{1/r}$, *then for case* (i)

$$\phi(n) \sim e^{O(1)}f_1(n^{1/r})[g_1(n^{1/r})]^{-s/r}n^{-s/r}, \tag{32}$$

and for case (ii)

$$\phi(n) \sim e^{O(1)}f_1(n^{1/r})[g_1(n^{1/r})]^{-s/r}(\log n)^{1/\beta-1}n^{-s/r}, \tag{33}$$

as $n \to \infty$.

PROOF. We first treat the slightly more complicated case (ii) and identify the functions f^* and g^* in (28) and (27) with the asymptotic expressions for f and g for this case. Using the Karamata representation for s.v. functions (see the corollary on page 274 of Feller (1966)) we have for $y \geqq x$,

$$\frac{g^*(y)}{g^*(x)} = \frac{g_1(e^{y^\beta})}{g_1(e^{x^\beta})}\exp\{-r(y^\beta - x^\beta)\}$$
$$= \frac{a(e^{y^\beta})}{a(e^{x^\beta})}\exp\left\{\int_{e^{x^\beta}}^{e^{y^\beta}}\frac{\varepsilon(u)}{u}\,du - r(y^\beta - x^\beta)\right\},$$

where $a(x)$ and $\varepsilon(u)$ are functions such that $a(x)$ converges to a positive constant as $x \to \infty$ and $\varepsilon(u) \to 0$ as $u \to \infty$. Hence for arbitrary ε, $0 < \varepsilon < r$,

$$\frac{g^*(y)}{g^*(x)} < (1+\varepsilon)\exp\{(\varepsilon - r)(y^\beta - x^\beta)\} < 1+\varepsilon, \tag{34}$$

for all x large enough so that $|\varepsilon(u)| < \varepsilon$ for $u > e^{x^\beta}$ and $a(e^{y^\beta}) < (1+\varepsilon)a(e^{x^\beta})$ for $y > x$. Now for fixed δ let

$$x(t,\delta) = r^{-1/\beta}\{\log[g_1(t^{-1/r})t^{-1}] + \delta\}^{1/\beta}. \tag{35}$$

Then

$$g^*(x(t,\delta)) = g_1([g_1(t^{-1/r})]^{1/r}t^{-1/r}e^{\delta/r})[g_1(t^{-1/r})]^{-1}t\,e^{-\delta} \sim t\,e^{-\delta} \quad \text{as } t \to 0.$$

Hence, observing that $x(t, \delta) \to \infty$ as $t \to 0$ and choosing $\delta_0 > 0$ such that $e^{-\delta_0} < 1-\varepsilon$, we have $g^*(x(t, \delta_0)) < t(1-\varepsilon)$ for all sufficiently small t. Thus from (27) and (34), for all sufficiently small t, and for integer x,

$$x > x(t,\delta_0) \Rightarrow g^*(x) < (1-\varepsilon^2)t < t \Rightarrow x \in S^*(t). \tag{36}$$

Similarly, choosing $\delta_1 < 0$ such that $e^{-\delta_1} > 1+\varepsilon$ we have $x \leqq x(t, \delta_1) \Rightarrow g^*(x) > (1+\varepsilon)^{-1} g^*(x(t, \delta_1)) > t$ for all sufficiently small t. Hence for all such t

$$x \in S^*(t) \Rightarrow g^*(x) \leqq t \Rightarrow x > x(t, \delta_1). \tag{37}$$

From (28), (36) and (37) we have for all sufficiently small t,

$$\textstyle\sum_{x > x(t,\delta_1)} f^*(x) \leqq v^*(t) \leqq \sum_{x > x(t,\delta_0)} f^*(x). \tag{38}$$

The asymptotic behavior of $v^*(t)$ is now deduced by examining the sums appearing in (38). For fixed δ,

$$\textstyle\sum_{x > x(t,\delta)} f^*(x) = \sum_{x > x(t,\delta)} f_1(e^{x^\beta}) e^{-sx^\beta},$$

and since for any c, $f_1(e^{(x+c)^\beta}) e^{-s(x+c)^\beta} / f_1(x^\beta) e^{-sx^\beta} \sim e^{-sc\beta\, x^{\beta-1}}$, as $x \to \infty$, we may bound the left-hand sum by the corresponding integral multiplied by appropriate constants (and indeed these constants may be taken to be arbitrarily close to one for small t except when $\beta = 1$). Furthermore, letting $v = e^{x^\beta}$,

$$\int_{x(t,\delta)}^{\infty} f_1(e^{x^\beta}) e^{-sx^\beta}\, dx = \frac{1}{\beta} \int_{e^{x^\beta(t,\delta)}}^{\infty} f_1(v) (\log v)^{(1/\beta)-1} v^{-s-1}\, dv$$

$$\sim \frac{1}{\beta s} f_1(e^{x^\beta(t,\delta)}) [x(t,\delta)]^{1-\beta} e^{-sx^\beta(t,\delta)} \qquad \text{as } t \to 0,$$

by Theorem 1, page 273 of Feller (1966). Substituting $x(t, \delta)$ given by (35) in the expression above and making use of the properties of f_1 and g_1 we obtain

$$\textstyle\sum_{x > x(t,\delta)} f^*(x) = e^{O(1)} f_1(t^{-1/r}) [g_1(t^{-1/r})]^{-s/r} [-\log t]^{(1/\beta)-1} t^{s/r} \quad \text{as } t \to 0, \tag{39}$$

upon observing that $\log[g_1(t^{-1/r})t^{-1}] + \delta \sim -\log t$ as $t \to 0$. Now from (38) we see that $v^*(t)$ must have an asymptotic expression of the same form as the left-hand side of (39). The desired result (33) follows immediately by Lemma 2. The result for case (i) follows by a similar argument with $x(t, \delta) = (1+\delta)[g_1(t^{-1/r})t^{-1}]^{1/r}$. The integral approximation for $v^*(t)$ for this case has integrand $f_1(x)x^{-(s+1)}$ to which the theorem of Feller (1966) cited above may be applied directly. Thus we obtain $v^*(t) = e^{O(1)} f_1(t^{-1/r})[g_1(t^{-1/r})] t^{s/r}$ as $t \to 0$, and the desired result (32) follows by Lemma 2. This completes the proof of the lemma.

A considerably more delicate argument would be needed to replace the $e^{O(1)}$ symbols by specific constants and in fact it appears that for case (ii) with $\beta = 1$ an oscillating factor is unavoidable.

The next three lemmas relate the asymptotic behavior of the unconditional probabilities $p(x)$ to the tail properties of the prior c.d.f. $G(\lambda)$. Lemma 4 treats the case where the function $h(x)$ appearing in (1) behaves asymptotically like a power of x so that the natural parameter space is the unit interval. For this case we suppose that G possesses a density near $\lambda = 1$ which approaches zero either algebraically or exponentially as $\lambda \to 1$. Lemma 5 deals similarly with the Poisson case. Lemma 6 applies to both models for the case where G assigns probability one to an interval $[0, \lambda_0]$ which is a strict subset of the natural parameter set. In these lemmas the letter C will represent a generic constant.

LEMMA 4. *If $p_\lambda(x)$ is given by* (1) *with*

$$h(x) \sim h_1(x)x^\gamma \qquad \text{as } x \to \infty, \gamma > -1,$$

and if $G'(\lambda)$ exists for $1-\varepsilon < \lambda < 1$ for some $\varepsilon > 0$ and

$$G'(\lambda) \sim G_1(1-\lambda)(1-\lambda)^\delta \exp\{-\sigma(1-\lambda)^{-\tau}\} \qquad \text{as } \lambda \to 1,$$

where $\tau > 0$, and either $\sigma > 0$ or $\sigma = 0$ and $\delta > -1$, and where h_1 and G_1 are s.v. *functions, then when $\sigma > 0$*

$$\text{(i)} \quad p(x) \sim C\frac{h_1(x)}{h_1(x^{1/(\tau+1)})} G_1(x^{-1/(\tau+1)})x^{(\tau(\gamma-\frac{1}{2})-\delta-2)/(\tau+1)} \cdot \exp\{-(1+\tau^{-1})(\sigma\tau)^{1/(\tau+1)}x^{\tau/(\tau+1)}\}$$

as $x \to \infty$, and when $\sigma = 0$ and $\delta > -1$,

$$\text{(ii)} \quad p(x) \sim CG_1(x^{-1})x^{-\delta-2} \quad \text{as } x \to \infty.$$

PROOF. We first consider case (i) where $\sigma > 0$ and recall

$$p(x) = h(x)\textstyle\int_0^1 \lambda^x\beta(\lambda)\,dG(\lambda), \qquad \text{where} \tag{40}$$

$$[\beta(\lambda)]^{-1} = \textstyle\int_0^\infty e^{x\log\lambda}\,h(x)\,d\mu(x),$$

where μ represents counting measure on the integers. Hence by the previously cited Abelian theorem, since $\log\lambda^{-1} \sim 1-\lambda$ as $\lambda \to 1$, we have

$$[\beta(\lambda)]^{-1} \sim \Gamma(\gamma+1)h_1\left(\frac{1}{1-\lambda}\right)(1-\lambda)^{-\gamma-1} \qquad \text{as } \lambda \to 1.$$

Hence $p(x)$ may be bounded above and below by expressions of the form

$$C\frac{h_1(x)x^\gamma}{\Gamma(\gamma+1)}\int_{1-\varepsilon}^1\left[h_1\left(\frac{1}{1-\lambda}\right)\right]^{-1} \cdot G_1(1-\lambda)(1-\lambda)^{\delta+\gamma+1}\exp\{x\log\lambda-\sigma(1-\lambda)^{-\tau}\}d\lambda, \tag{41}$$

where the constants C may be taken arbitrarily close to one for sufficiently small $\varepsilon > 0$, provided (41) is of larger order than $e^{-\varepsilon x}$ which bounds the order of the neglected portion of the integral in (40). Now let $\theta = (1-\lambda)x^{1/(\tau+1)}-(\sigma\tau)^{1/(\tau+1)}$. Then $\log\lambda \sim -[\theta+(\sigma\tau)^{1/(\tau+1)}]x^{-1/(\tau+1)}$ uniformly for $\theta \in (-\varepsilon, \varepsilon)$ as $x \to \infty$. Furthermore, $G_1(x^{-1/(\tau+1)})/G_1([\theta+(\sigma\tau)^{1/(\tau+1)}]x^{-1(\tau+1)} \sim 1$ uniformly for $\theta \in (-\varepsilon, \varepsilon)$ as $x \to \infty$, and the analogous result holds for h_1. Hence, neglecting contributions of exponentially smaller order, the integral in (41) is asymptotically equivalent to

$$\frac{G_1(x^{-1/(\tau+1)})}{h_1(x^{1/(\tau+1)})}x^{-(\delta+\gamma+2)/(\tau+1)}\exp\{-(1+\tau^{-1})(\sigma\tau)^{1/(\tau+1)}x^{\tau/(\tau+1)}\} \cdot \textstyle\int_{-\varepsilon}^{\varepsilon}[\theta+(\sigma\tau)^{1/(\tau+1)}]^{\delta+\gamma+1} \cdot \exp\{-x^{\tau/(\tau+1)}(\theta+\sigma[\theta+(\sigma\tau)^{1/(\tau+1)}]^{-\tau}-\sigma(\sigma\tau)^{-\tau/(\tau+1)})\}\,d\theta. \tag{42}$$

The coefficient of $-x^{\tau/(\tau+1)}$ in the exponential part of the integrand of (42) is of the form $\frac{1}{2}(\tau+1)\theta^2(1+o(1))$ as $\theta \to 0$, so that by the well-known asymptotic theorem for Laplace integrals (see e.g. De Bruijn (1961), pages 63–65) the integral in (42) is asymptotic to $(2\pi)^{\frac{1}{2}}(\sigma\tau)^{(\delta+\gamma+1)/(\tau+1)}[(\tau+1)x^{\tau/(\tau+1)}]^{-\frac{1}{2}}$ as $x \to \infty$, and the desired result follows.

For case (ii) where $\sigma = 0$ and $\delta > -1$, we let $\lambda = e^{-\theta}$ in (41) and extend the range of integration to $(0, \infty)$ resulting in a contribution which is exponentially small in x. Since $1-\lambda \sim \theta$ as $\lambda \to 1$, the desired result follows by the fundamental Abelian theorem and the proof of the lemma is complete.

REMARK 2. In Lemma 4 we require $\gamma > -1$ in the expression for $h(x)$. The case of $\gamma < -1$ is not substantially different. For this case $\beta(\lambda) \to \beta(1) > 0$ as $\lambda \to 1$, and in Lemma 4 the power of x is reduced by $(\gamma+1)/(\tau+1)$ and the factor $h_1(x^{1/(\tau+1)})^{-1}$ disappears in asymptotic expression (i) for $p(x)$. In expression (ii) of Lemma 4 the power of x becomes $-(\delta-\gamma+1)$ and $h_1(x)$ appears as a factor.

We now consider the Poisson distribution which illustrates the case where the natural parameter space is unbounded.

LEMMA 5. *If* $h(x) = \Gamma(x+1)^{-1}$ *and* $G'(\lambda)$ *exists for all sufficiently large* λ *with*

$$G'(\lambda) \sim C_0\lambda^\delta e^{-\sigma\lambda}, \quad \sigma > 0, \text{ or } \sigma = 0 \text{ and } \delta < -1,$$

as $\lambda \to \infty$, *then*

$$p(x) \sim C_0 x^\delta(\sigma+1)^{-(x+\delta+1)},$$

as $x \to \infty$.

PROOF. For this case $\beta(\lambda) = e^{-\lambda}$ and

$$p(x) = \frac{1}{\Gamma(x+1)}\int_0^\infty \lambda^x e^{-\lambda}\,dG(\lambda) \sim \frac{C_0}{\Gamma(x+1)}\int_T^\infty \lambda^{x+\delta}e^{-(\sigma+1)\lambda}\,d\lambda,$$

as $x \to \infty$, provided the integral on the right is asymptotically independent of T and of larger order than $T^{x+\delta}$ which bounds the order of the neglected portion. But if the range of integration is extended to $(0, \infty)$ the integral becomes $\Gamma(x+\delta+1)/(\sigma+1)^{x+\delta+1}$ and the desired result follows since

$$\frac{\Gamma(x+\delta+1)}{\Gamma(x+1)} \sim x^\delta, \qquad \text{as } x \to \infty,$$

by Stirling's formula.

The final lemma of this section is concerned with the case where $G(\lambda)$ assigns all its probability mass to an interval $[0, \lambda_0]$ which is strictly contained within the natural parameter range.

LEMMA 6. *If* $p_\lambda(x)$ *is given by* (1) *and if* $\lambda_0 > 0$ *is such that* $G(\lambda_0) = 1$, *and there exists a* $\lambda > \lambda_0$ *for which* $\beta(\lambda) > 0$, *and if for some* $\varepsilon > 0$, $G'(\lambda)$ *exists in* $(\lambda_0-\varepsilon, \lambda_0)$ *and*

$$G'(\lambda) \sim G_1(\lambda_0-\lambda)(\lambda_0-\lambda)^\delta, \qquad \delta > -1, \text{ as } \lambda \uparrow \lambda_0,$$

where G_1 *is* s.v., *then*

$$p(x) \sim \beta(\lambda_0)\Gamma(\delta+1)h(x)G_1\left(\frac{1}{x}\right)x^{-\delta-1}\lambda_0^{x+\delta+1},$$

as $x \to \infty$.

PROOF. We observe that $\beta(\lambda) \to \beta(\lambda_0) > 0$ as $\lambda \uparrow \lambda_0$ for this case, so that

$$p(x) = h(x)\textstyle\int_0^{\lambda_0} e^{x \log \lambda}\beta(\lambda)\,dG(\lambda)$$

$$\sim -\beta(\lambda_0)\lambda_0^{\,x}\textstyle\int_0^{\infty} e^{-x \log \lambda_0/\lambda}\,dG(\lambda) \qquad \text{as } \lambda \uparrow \lambda_0.$$

The desired result follows as in part (ii) of Lemma 4.

4. Exact rates of convergence. The first theorem of this section is concerned with the case where the natural parameter range is bounded above. This theorem connects rates of convergence of r_n to $r^*(G)$ with upper tail conditions on the prior distribution $G(\lambda)$.

Without loss of generality, we consider the parameter space to be the interval $[0, 1)$ since this may be obtained by a simple scale transformation of any parameter space bounded above. Also, to avoid trivialities, we take the constant c appearing in the loss functions and expression (7) for $\alpha(x)$ to lie strictly between zero and one.

THEOREM 3. *If* $p_\lambda(x)$ *is given by* (1) *with* $h(x) \sim h_1(x)x^\gamma$, $\gamma > -1$, *as* $x \to \infty$, *where* $h_1(x)$ *is* s.v., *and if* $\delta_n(x)$ *is given by* (9) *with* $\alpha_n(x)$ *given by* (10), *then if* $G'(\lambda)$ *exists for* $1-\varepsilon < \lambda < 1$ *for some* $\varepsilon > 0$, *and*

(a) $G'(\lambda) \sim G_1(1-\lambda)(1-\lambda)^\delta$ *as* $\lambda \to 1, \delta > -1$,

where G_1 *is* s.v., *and* G_1 *and* $G_1^{\,2}$ *are* s.v. *with respect to* $G_1^{\,1/(\delta+2)}$, *then as* $n \to \infty$

$$r_n - r^*(G) = e^{O(1)}[G_1(n^{-1/(\delta+2)})]^{1/(\delta+2)}n^{-(\delta+1)/(\delta+2)}.$$

Alternatively, if

(b) $G'(\lambda) \sim G_1(1-\lambda)(1-\lambda)^\delta \exp\{-\sigma(1-\lambda)^{-\tau}\}$ *as* $\lambda \to 1$, $\sigma, \tau > 0$, *and if* $[h_1(x)G_1(x^{-1/(\tau+1)})/h_1(x^{1/(\tau+1)})]^\xi$ *for* $\xi = 1, 2$ *is* s.v. *with respect to the same unction with* $\xi = (1+\tau^{-1})(\sigma\tau)^{-1/(\tau+1)}$, *then as* $n \to \infty$,

$$r_n - r^*(G) = e^{O(1)}[\log n]^{1/\tau}n^{-1}.$$

Finally, if for some λ_0, $c < \lambda_0 < 1$, *and some* $\varepsilon > 0$, $G(\lambda_0) = 1$ *and* $G'(\lambda)$ *exists in* $(\lambda_0-\varepsilon, \lambda_0)$ *with*

(c) $G'(\lambda) \sim G_1(\lambda_0-\lambda)(\lambda_0-\lambda)^\delta, \delta > -1$ *as* $\lambda \uparrow \lambda_0$,

where G_1 *is* s.v. *and* $[G_1(x^{-1})h_1(x)]^\xi$ *for* $\xi = 1, 2$ *is* s.v. *with respect to the same function with* $\xi = (-\log \lambda_0)^{-1}$, *then as* $n \to \infty$,

$$r_n - r^*(G) = e^{O(1)}n^{-1}.$$

PROOF. For this case $w(x) = h(x)/h(x+1) \to 1$ as $x \to \infty$. Furthermore, for cases (a) and (b) we have $p(x+1)/p(x) \to 1$ as $x \to \infty$ by Lemma 4, and for case (c), $p(x+1)/p(x) \to \lambda_0 > c$, as $x \to \infty$, by Lemma 6. Thus, by (7), (11) and (12) we see that for each case, for some $c_1, c_2 > 0$

$$\alpha(x) \sim c_1 p(x), \quad \text{and} \quad \sigma^2(x) \sim c_2 p(x), \qquad \text{as } x \to \infty. \tag{43}$$

Furthermore, by (11) the quantities $\pi^+(x)$ and $\pi^-(x)$ appearing in (13) of Theorem 2 are given by

$$\pi^+(x) = p(x+1), \pi^-(x) = p(x). \tag{44}$$

Thus, the lower bound on $r_n - r^*(G)$ given by (13) is of the form $n\phi(n-1)$ where $\phi(n)$ is given by (26) with

$$f(x) \sim c_1 p^2(x), g(x) \sim 2p(x), \qquad \text{as } x \to \infty.$$

If (a) holds then the desired conclusion follows for the lower bound by Lemma 3(i) and Lemma 4(ii) upon setting $g_1(x) = 2\Gamma(\gamma+\delta+2)\Gamma(\gamma+1)^{-1}G_1(x^{-1})$ and $f_1(x) = \frac{1}{4}c_1 g_1{}^2(x)$, and observing that $r = \delta+2$ and $s = 2(\delta+2)-1$ in Lemma 3 for this case. Similarly, the desired conclusions for the lower bound are obtained for the cases where (b) or (c) hold by applying Lemma 3(ii) to the results of Lemma 4(i) for case (b) and Lemma 6 for case (c).

To verify that the upper bounds given by (14) of Theorem 2 are also of the desired order we must first verify that Conditions 1, 2, and 3 are satisfied. Condition 1 is implied by (43), and by (11)

$$m_x(t) = p(x+1)\,e^{w(x)t} + p(x)\,e^{-ct}, \tag{45}$$

so that Condition 2 is also satisfied. Furthermore, differentiating (45) twice yields

$$\ddot{m}_x(t) = w^2(x)p(x+1)\,e^{w(x)t} + c^2 p(x)\,e^{-ct}, \tag{46}$$

and by (43), $\tau|\alpha(x)|/\sigma^2(x) \sim \tau c_1/c_2$ as $x \to \infty$, so that Condition 3 is satisfied.

The upper bound (14) of Theorem 2 is of the form $\phi(n)$ given by (26) with $f(x) \sim c_1 p(x)$ and $g(x) \sim c_0 c_1{}^2 c_2{}^{-1} p(x)$. Hence if (a) holds the desired conclusion for the upper bound follows by Lemmas 3(i) and 4(ii) upon setting $f_1(x) = c_0{}^{-1}c_1{}^{-1}c_2 g_1(x) = c_1\Gamma(\gamma+\delta+2)\Gamma(\gamma+1)^{-1}G_1(x^{-1})$ and $r = s = \delta+2$ in Lemma 3(i). The upper bounds for cases (b) and (c) are verified similarly using Lemmas 3, 4 and 6, and the proof of the theorem is complete.

REMARK 3. It is possible to construct functions $h_1(x)$ which do not satisfy the conditions of Theorem 3 and yet for which $h(x) = h_1(x)x^\gamma$ defines an exponential family whose natural parameter space is [0, 1). Nevertheless, the family of functions $h_1(x)$ which do satisfy these conditions is sufficiently broad to include not only the geometric and negative binomial distributions and their obvious modifications, but also all other distributions ever likely to be proposed as models for the naturally bounded parameter case. It should be noted that the theorem could be modified to include the case $\gamma < -1$ in accordance with Remark 2.

REMARK 4. The significance of Theorem 3 is that the rate of convergence to optimality is essentially n^{-1} in any situation likely to arise in actual applications of the empirical Bayes method to the cases under consideration. As noted in Remark 1, the rate n^{-1} which is actually attained under condition (c) of the theorem is best possible.

Although algebraically slower rates occur under hypothesis (a) of the theorem, the prior distributions satisfying this condition put excessive weight in the right tail (near one) to the extent that the unconditional distribution $p(x)$ lacks higher moments (by Lemma 4(ii)). Such priors are not likely to represent real world situations.

REMARK 5. Although Theorem 3 does not actually give the first term in an asymptotic expansion of $r_n - r^*(G)$, the two constants (one for each bound of Theorem 2) obscured by the factor $e^{O(1)}$ could be obtained explicitly from the lemmas and the proof of Theorem 2. This was not done, because the result would still be only asymptotic and would not shed much light on the behavior of $r_n - r^*(G)$ for small n. A more delicate analysis, replacing the lemmas of Section 3 by actual bounds on $\phi(n)$ and $p(x)$, would be required to obtain sharp bounds on $r_n - r^*(G)$ valid for all n.

REMARK 6. The exact rates of Theorem 3 may easily be compared in particular cases with the upper bounds given by Corollary 1 of Section 2. In Example 1 of Section 2 it was shown that for the geometric distribution (i.e., $h(x) \equiv 1$) with prior density

$$G'(\lambda) = (\gamma+1)(1-\lambda)^{\gamma}, \qquad 0 \leqq \lambda < 1, \gamma > -1,$$

the rate of convergence of r_n to $r^*(G)$ was at least as fast as $n^{-(\gamma+1-\varepsilon)/(\gamma+2)}$ for any $\varepsilon > 0$. Theorem 3 shows that the exact rate for this case is $n^{-(\gamma+1)/(\gamma+2)}$. Thus, for this case the simple results based on Theorem 1 only miss the mark by an arbitrarily small power of n. As was noted before, this fact has implications for the sequel (Johns and Van Ryzin (1967)).

THEOREM 4. (Poisson case). *If $p(x)$ is given by* (1) *with $h(x) = \Gamma(x+1)^{-1}$, and if $\delta_n(x)$ is given by* (9) *with $\alpha_n(x)$ given by* (10), *then if $G'(\lambda)$ exists for all sufficiently large λ with*

(a) $G'(\lambda) \sim C_0 \lambda^{\delta}, \delta < -2,$

as $\lambda \to \infty$, then

$$r_n - r^*(G) = e^{O(1)} n^{-(1+2/\delta)}.$$

Alternatively, if

(b) $G'(\lambda) \sim C_0 \lambda^{\delta} e^{-\sigma\lambda}, \sigma > 0,$

as $\lambda \to \infty$, then

$$r_n - r^*(G) = e^{O(1)} (\log n) n^{-1}.$$

PROOF. For this case $w(x) = h(x)/h(x+1) = x+1$, so that by (7), (11), (12) and Lemma 5,

$$\alpha(x) \sim c_1 x^{\delta+1}(\sigma+1)^{-x}, \sigma^2(x) \sim c_1 x^{\delta+2}(\sigma+1)^{-x} \tag{47}$$

as $x \to \infty$, for a certain $c_1 > 0$.

Since (44) holds for this case also, the lower bound (13) of Theorem 2 is of the form $n\phi(n-1)$ where $\phi(n)$ is given by (26) with

$$f(x) \sim c_2 x^{2\delta+1}(\sigma+1)^{-2x}, \qquad g(x) \sim c_3 x^{\delta}(\sigma+1)^{-x},$$

as $x \to \infty$, for suitable c_2, $c_3 > 0$, by Lemma 5. The required conclusion for the lower bound then follows for case (*a*) ($\sigma = 0$) by Lemma 3(i) with $f_1(x) \equiv c_2$, $g_1(x) \equiv c_3$, $r = -\delta$ and $s = -2\delta-2$. The conclusion for case (b) follows by Lemma 3(ii) with $\beta = 1$, $2r = s = 2 \log(\sigma+1)$, $f_1(x) = c_2(\log x)^{2\delta+1}$ and $g_1(x) = c_3(\log x)^{\delta}$.

The upper bound (14) is of the form of $\phi(n)$ given by (26) with

$$f(x) \sim c_1 x^{\delta+1}(\sigma+1)^{-x}, \qquad g(x) \sim c_1 x^{\delta}(\sigma+1)^{-x},$$

as $x \to \infty$ by (47). Also, by (47) and (45) Conditions 1 and 2 are satisfied, and by (47) $\tau|\alpha(x)|/\sigma^2(x) \sim x^{-1}$ as $x \to \infty$, so that by (46) and Lemma 5, Condition 3 is satisfied. The desired result follows for case (a) by Lemma 3(i) with $f_1(x) = g_1(x) \equiv c_1$, $r = -\delta$ and $s = -\delta-2$. The conclusion for case (b) follows by Lemma 3(ii) with $\beta = 1$, $r = s = \log(\sigma+1)$, $f_1(x) = (\log x)^{\delta+1}$ and $g_1(x) = (\log x)^{\delta}$. This completes the proof of the theorem.

REMARK 7. The conclusions of Theorem 4 are completely analogous to those for the corresponding cases (a) and (b) of Theorem 3 and the relevant comments in Remarks 4 and 5 therefore apply also to the Poisson case. The requirement $\delta < -2$ in case (b) of Theorem 4 is related to the fact that we must have $E\Lambda < \infty$ for this case.

REMARK 8. The rate n^{-1} cannot actually be achieved for the Poisson case even when the support of G is bounded. To see this we note that the lower bound (13) of Theorem 2 behaves like $n\sum_x xp^2(x)[1-2p(x)]^{n-1}$ and if we choose x_n such that $p(x_n) \sim n^{-1}$ as $n \to \infty$, then the expression for the bound is itself bounded below by $e^{-2}x_n/n$, and $x_n \to \infty$ as $n \to \infty$.

This remark applies also to all cases where $h(x)$ becomes small at an even faster rate than $\Gamma(x+1)^{-1}$ (e.g., $h(x) = e^{-x^2}$) since $w(x) \to \infty$ as $x \to \infty$ for such examples. We do not discuss these cases further here because for technical reasons it appears that each example must be treated separately if precise results are to be obtained. It seems likely that the rates of convergence for such cases would be at least as good as those for the Poisson case under similar circumstances.

REMARK 9. The Poisson case was discussed in Example 2 of Section 2 where it was shown that the rate of convergence to optimality was faster than $n^{-1+\varepsilon}$ for any $\varepsilon > 0$ when $G'(\lambda) = e^{-\lambda}$. By Theorem 4, case (b) we see that the exact rate is $(\log n)n^{-1}$ for this case. Thus again, as noted in Remark 6, the results based on Theorem 1 only underestimate the correct rate by an arbitrarily small power of n.

REMARK 10. In all of the cases considered in detail in this paper, $p_\lambda(x)$ has been of the form (1) and $Z_j(x)$ has been a trinomial given by (11). It should be noted, however, that there exist other examples, both parametric and "non-parametric" in the sense of Johns (1957), to which the results of Sections 2 and 3 apply. Such an example is provided by the reparameterized negative binomial distribution,

$$p_\lambda(x) = h(x)\left(\frac{k}{k+\lambda}\right)^k\left(\frac{\lambda}{k+\lambda}\right)^x, \qquad \lambda > 0, x = 0, 1, \cdots$$

where k is a specified positive constant, $h(0) = 1$ and

$$h(x) = \frac{k(k+1)\cdots(k+x-1)}{x!}, \qquad x = 1, 2, \cdots.$$

The appropriate choice of $Z_j(x)$ in this case is

$$\begin{aligned} Z_j(x) &= \frac{kh(x)}{h(x+y)}, && X_j = x+y, y = 1, 2, \cdots \\ &= -c, && X_j = x \\ &= 0, && \text{otherwise.} \end{aligned}$$

Rate results analogous to those of Theorem 3 could reasonably be expected from a detailed analysis of this example using the results of Section 3.

REFERENCES

CHASE, G. R. (1966). An empirical Bayes approach in routine bioassay. Technical Report No. 13, Department of Statistics, Stanford Univ.

DE BRUIJN, N. G. (1961). *Asymptotic Methods in Analysis*, 2nd ed. North-Holland Publishing Co., Amsterdam.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications* **2**. Wiley, New York.

JOHNS, Jr., M. V. (1957). Nonparametric empirical Bayes procedures. *Ann. Math. Statist.* **28** 649–669.

JOHNS, Jr., M. V. and VAN RYZIN, J. R. (1967). Convergence rates for empirical Bayes two-action problems II. Continuous case. Technical Report No. 4, Department of Statistics, Stanford University.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** Univ. of California Press, 157–164.

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

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