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

BY
M.V. JOHNS, JR. and J. VAN RYZIN

TECHNICAL REPORT NO. 131
NOVEMBER 13, 1967

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE,
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
Nonr-225(52) (NR-342-022)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
CONVERGENCE RATES FOR EMPIRICAL BAYES
TWO-ACTION PROBLEMS I. DISCRETE CASE

by

M. V. Johns, Jr. and J. Van Ryzin

TECHNICAL REPORT NO. 131

NOVEMBER 13, 1967

PREPARED UNDER CONTRACT Nonr-225(52)
(NR-342-022) FOR
OFFICE OF NAVAL RESEARCH

and

SUPPORTED BY PUBLIC HEALTH SERVICE
GRANT USPHS-GM-14554-01

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
CONVERGENCE RATES FOR EMPIRICAL BAYES
TWO-ACTION PROBLEMS I. DISCRETE CASE

by

M.V. Johns, Jr. and J. Van Ryzin

1. **Introduction and Summary.**

Situations involving sequences of similar but independent statistical decision problems arise in many areas of application. Routine bioassay (Chase [1]) 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 [5] 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 $n$th 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 [3] exhibits asymptotically optimal empirical Bayes procedures for certain two-action (hypothesis testing) problems as well as for estimation problems in a non-parametric context. Robbins [7] and Samuel [8] 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 [3] 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 [4] 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 [4] 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 structure: An observation $X$ is obtained having a distribution with probability mass function

$$p_{\lambda}(x) = h(x)\lambda^x e^{-\lambda}, \quad x = 0, 1, \ldots; \quad 0 < \lambda < d,$$

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 $H_1: \lambda \leq c, \quad c > 0$ is to be tested against $H_2: \lambda > c$ with loss function
\begin{align*}
L_1(\lambda) &= \begin{cases} 
0 & \text{if } \lambda \leq c \\
b(\lambda - c) & \text{if } \lambda > c, \ b > 0,
\end{cases} \\
L_2(\lambda) &= \begin{cases} 
b(c - \lambda) & \text{if } \lambda \leq c \\
0 & \text{if } \lambda > c,
\end{cases}
\end{align*}

where \( L_i(\lambda) \) indicates the loss when action \( i \) (deciding in favor of \( H_i \)) is taken, \( i = 1, 2 \), and \( \lambda \) is the true value of the parameter.

It is assumed that \( \lambda \) may be regarded as the value of a random variable \( \Lambda \) 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{equation}
(2) \quad r(\delta, G) = \int_X \left\{ L_1(\lambda)p_\lambda(x)\delta(x) + L_2(x)p_\lambda(x)(1-\delta(x)) \right\}dG(\lambda)
\end{equation}

\begin{equation}
= b\sum_x \alpha(x)\delta(x) + C_G,
\end{equation}

where \( C_G = \int L_2(\lambda)dG(\lambda) \) and

\begin{equation}
(3) \quad \alpha(x) = \int \lambda p_\lambda(x)dG(\lambda) - cp(x),
\end{equation}

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

\begin{equation}
(4) \quad p(x) = \int p_\lambda(x)dG(\lambda), \ x = 0, 1, \ldots.
\end{equation}
A Bayes rule (i.e., a minimizer of (2) based on knowledge of \( G \)) is clearly given by

\[
\delta_G(x) = \begin{cases} 
1, & \alpha(x) \leq 0 \\
0, & \alpha(x) > 0.
\end{cases}
\]  

(5)

The resulting (minimal) Bayes risk is

\[
\mathbf{r}^*(G) = \inf_\delta \ r(\delta,G) = r(\delta_G,G).
\]  

(6)

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

\[
\alpha(x) = w(x)p(x+1) - cp(x),
\]  

(7)

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

\[
w(x) = \frac{h(x)}{h(x+1)}.
\]  

(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 \( X_1, X_2, \ldots, X_n \) obtained in the previous problems is available. The empirical Bayes procedures considered here involve the construction of a sequence of estimates \( \alpha_n(x), n = 1, 2, \ldots \) of the function \( \alpha(x) \) where \( \alpha_n(x) \) is based on the observations \( X_1, X_2, \ldots, X_n \). The
decision rule used for the \((n+1)^{st}\) decision problem is then

\[
\delta_n(x) = \begin{cases} 
1, & \alpha_n(x) \leq 0 \\
0, & \alpha_n(x) > 0.
\end{cases}
\]

(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

\[
\alpha_n(x) = \frac{1}{n} \sum_{j=1}^{n} Z_j(x),
\]

(10)

where for each \(x, j,\)

\[
Z_j(x) = w(x)U_j(x+1) - cU_j(x)
\]

(11)

where

\[
U_j(x) = \begin{cases} 
1, & X_j = x \\
0, & X_j \neq x.
\end{cases}
\]

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 non-negative and it is easily shown ([7], [8]) that \(r_n - r^*(G) \to 0\), as \(n \to \infty\), i.e., that \(\delta_n\) is asymptotically optimal, provided only that \(EA < \infty\).
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 [4] 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 [4].

Section 3 contains the development of the asymptotic tools necessary to obtain the exact rates of convergence to optimality discussed 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 non-negative 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 \( \varepsilon \) arbitrarily small may in principle be obtained when \( G'(\lambda) \) behaves like \( (1-\lambda)^\sigma \) for \( \lambda \) close to 1, for sufficiently small \( \sigma \). 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 $\lambda^{\delta}$ or $\lambda^{\delta} e^{-\sigma \lambda}$, $\sigma \geq 0$, 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 sharp.
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 $\delta_n$ given by (9), we have, recalling (2),

$$r_n = \sum_{x} \alpha(x) E \delta_n(x) + C_G$$

$$= \sum_{x} \alpha(x) \Pr\{\alpha_n(x) \leq 0\} + C_G.$$

We now state a lemma, useful here and in [4].

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

$$0 \leq r_n - r^{*}(G) \leq \sum_{x} |\alpha(x)| \Pr\{|\alpha_n(x) - \alpha(x)| \geq |\alpha(x)|\}.$$

Proof. Recalling (5) we have

$$r_n - r^{*}(G) = \sum_{x} \alpha(x) (\Pr\{\alpha_n(x) \leq 0\} - \delta_G(x))$$

$$= \sum_{x} |\alpha(x)| \Delta_n(x),$$

where

$$\Delta_n(x) = \begin{cases} 
\Pr\{\alpha_n(x) > 0\} & \text{if } \alpha(x) \leq 0 \\
\Pr\{\alpha_n(x) \leq 0\} & \text{if } \alpha(x) > 0.
\end{cases}$$
The desired result follows from the fact that the event
\( \{|a_n(x) - a(x)| \geq |a(x)|\} \) is implied by \( \{a_n(x) > 0\} \) when \( a(x) \leq 0 \),
and by \( \{a_n(x) \leq 0\} \) when \( a(x) > 0 \).

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

\[(12) \quad \sigma^2(x) = \text{Var}(Z_j(x))\]

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

\[ \sum_x |a(x)|^{1-\delta} \sigma(x)^{\delta} < K < \infty, \]

then

\[ 0 \leq r_n - r^*(G) \leq \frac{\delta}{2}. \]

**Proof.** By Lemma 1 and (10) we have

\[ r_n - r^*(G) \leq b \sum_x |a(x)| \Pr\left( \left| \frac{1}{n} \sum_j Z_j(x) - a(x) \right| \geq |a(x)|^{\delta} \right) \]

\[ \leq b \sum_x |a(x)|^{1-\delta} E \left| \frac{1}{n} \sum_j (Z_j(x) - a(x)) \right|^{\delta} \]

\[ \leq b \left( \frac{\sigma^2(x)}{n} \right)^{\frac{\delta}{2}}, \]

which yields the desired result.

**Corollary 1.** If \( p_\lambda(x) \) and \( w(x) \) are given by (1) and (8) and

(i) \( p(x)/p(x+1) \rightarrow 1 \), as \( x \rightarrow \infty \),
and if for $0 < \delta < 2$, either 

(ii) $w(x) \to 0$, or $w(x) \sim c_o \neq c$, as $x \to \infty$, 

and 

(iii) $\sum_{1}^{L} p(x) \frac{1 - \delta}{2} < \infty$, 

or alternatively, if 

(iv) $p(x)/p(x+1) < c_1 < \infty$, 

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

and 

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

then there exists a $K > 0$ such that 

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

Proof. By (7), if (i) and (ii) hold 

$$|\alpha(x)| \sim K_1 p(x)$$

for some $K_1 > 0$ as $x \to \infty$, 

and by (11) and (12) 

$$\sigma^2(x) \leq E\sigma^2_j(x) = w^2(x)p(x+1) + c^2p(x) \leq K_2 p(x)$$

for some $K_2 > 0$ 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

\[
 p_\lambda(x) = \lambda^x (1-\lambda), \ x = 0,1,\ldots; \ 0 \leq \lambda < 1,
\]

and that the prior distribution has probability density function

\[
 G'(\lambda) = (\gamma+1)(1-\lambda)^\gamma, \ 0 \leq \lambda < 1, \ \gamma > -1.
\]

then

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

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

Taking \( 0 < c < 1 \) and noting that \( w(x) = 1 \) for this case, we see that (i), (ii) and (iii) of the corollary are satisfied for

\[
 \delta < 2(\gamma+1)/(\gamma+2). \text{ Thus, for a given value of } \gamma \text{ we are assured of a convergence rate faster than } n^{-(\gamma+1-\varepsilon)/(\gamma+2)} \text{ for any } \varepsilon > 0. \text{ If } \gamma \text{ may be taken arbitrarily large, the rate becomes arbitrarily close to } n^{-1}.
\]

The fact that, for this example, the value of \( \gamma \) 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 $n^{-\frac{\delta}{2}}$ is achieved provided only that condition (i) of the corollary is satisfied and $E(1-\lambda)^{-t} < \infty$, where $t = \delta(1+\epsilon)/(2-\delta)$, $\epsilon > 0$. The argument based on corollary 1 is similar to that of corollary 3.1 of [4] and will not be reproduced here.

Example 2. (The Poisson distribution). Let

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

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

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

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 $n^{-1+\epsilon}$ for any $\epsilon > 0$. In section 4 it will be shown that this is really a consequence of the exponentially decreasing tail of $G'(\lambda)$. 

12
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 \( r_n - r^*(G) \). Let

\[
m_x(t) = \mathbb{E} \exp\{tZ_1(x)\}, \quad m_x'(t) = \frac{\partial^2}{\partial t^2} m_x(t)
\]

\[
\pi^+(x) = \Pr\{Z_1(x) > 0\},
\]

\[
\pi^-(x) = \Pr\{Z_1(x) < 0\}.
\]

**Condition 1.** \( \alpha(x) = O(\sigma^2(x)) \), 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 \( \tau \) in some open interval containing zero, there exists a positive constant \( K \), independent of \( x \) and \( \tau \), such that

\[
\frac{m_x(\tau | \alpha(x)|/\sigma^2(x))}{\sigma^2(x)} \leq K\sigma^2(x), \text{ for all } x.
\]

**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

\begin{align*}
(13) \quad r_n - r^*(G) & \geq \sum_{x \in S^+} \alpha(x)\pi^-(x)[1-\pi^+(x)\pi^-(x)] - \prod_{x \in S^-} \alpha(x)\pi^+(x)[1-\pi^+(x)\pi^-(x)]^{n-1}.
\end{align*}
If conditions 1, 2, and 3 are satisfied, then there exists a positive constant $c_0$ such that

$$r_n - r^*(G) \leq \sum_{x} |a(x)| \left[ 1 - c_0 \frac{a^2(x)}{\sigma^2(x)} \right]^n. \tag{14}$$

**Proof.** We first note that, as in Lemma 1,

$$r_n - r^*(G) = \sum_{x \in S^+} a(x) \Pr\{a_n(x) \leq 0\} \tag{15}$$

$$- \sum_{x \in S^-} a(x) \Pr\{a_n(x) > 0\}.$$

But

$$\Pr\{a_n(x) \leq 0\} \geq \frac{1}{n} \sum_{k=1}^{n} \Pr\{Z_j(x)=0, j=1,2,\ldots,k-1, k+1,\ldots,n; Z_k(x) < 0\}$$

$$= n \pi^-(x) \left[ 1 - \pi^+(x) - \pi^-(x) \right]^{n-1}.$$

Similarly,

$$\Pr\{a_n(x) > 0\} \geq n \pi^+(x) \left[ 1 - \pi^+(x) - \pi^-(x) \right]^{n-1},$$

and (13) follows immediately from (15).

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

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

$$\Pr\{a_n(x) \leq 0\} \leq [m_x(-t)]^n. \tag{17}$$
For $0 < t < \delta$, $x \in S^-$ we may write

$$(18) \quad m_x(t) = 1 - |a(x)|t + \theta(x,t) \frac{t^2}{2},$$

where

$$\theta(x,t) = E\{Z^2(x) \exp\{\xi t Z(x)\}\},$$

and where $\xi$ is a quantity depending on $tZ(x)$ and such that $0 < \xi < 1$. Thus,

$$(19) \quad \theta(x,t) \leq EZ^2(x) + E\{Z^2(x) \exp\{tZ(x)\}\}$$

$$= EZ^2(x) + m_x(t).$$

Similarly, for $0 < t < \delta$, $x \in S^+$,

$$(20) \quad m_x(-t) = 1 - |a(x)|t + \theta(x,-t) \frac{t^2}{2},$$

where

$$(21) \quad \theta(x,-t) \leq EZ^2(x) + m_x(-t).$$

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

$$(22) \quad EZ^2(x) \leq c_1 \sigma^2(x).$$
Hence, by (18), (19), (22), and condition 3, taking $t_x = \tau |a(x)|/\sigma^2(x)$ for sufficiently small $\tau > 0$, there exists a constant $c_o > 0$ such that for $x \in S^-$,

$$m_x(t_x) \leq 1 - \frac{\sigma^2(x)}{\sigma^2(x)} \left( \tau - \frac{1}{2}\tau^2(c_1+K) \right)$$

$$\leq 1 - c_o \frac{\sigma^2(x)}{\sigma^2(x)}.$$

Similarly, by (20), (21), (22) we have for $x \in S^+$,

$$m_x(-t_x) \leq 1 - c_o \frac{\sigma^2(x)}{\sigma^2(x)}.$$

The desired result (14) then follows from (15) - (17), (23) and (24), 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

$$\max \left[ \pi^+(x_n), \pi^-(x_n) \right] = n^{-1} = 0(1)$$

then by (13) we have

$$r_n - r_n^* \geq |a(x_n)|e^0(1).$$

In the particular case where $p_\lambda(x)$ is given by (1) and $a(x)$ by (7) we have $\pi^+(x) = p(x+1)$ and $\pi^-(x) = p(x)$, and typically
\( \alpha(x) \geq p(x+1)e^{0(1)} = p(x)e^{0(1)} \). Under these circumstances (25) yields

\[
\frac{r_n - r_n^*(G)}{n^{-1} e^{0(1)}} = n^{-\frac{1}{2}},
\]

so that a rate of convergence to optimality of \( n^{-\frac{1}{2}} \) is best possible. 

Instances where this rate is actually achieved will be discussed in section 4.

In order to apply theorem 2 we must first investigate the behavior for large \( n \) of quantities of the form

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

where \( f(x) \geq 0 \) and \( 0 \leq g(x) \leq 1 \) for all \( x \geq 0 \). It will be convenient to introduce the concept of "slowly varying" functions, i.e., functions \( k(t) \) satisfying the following condition:

**Condition A.** For any \( c > 0 \), \( k(ct)/k(t) \to 1 \) as \( t \to 0 \), (or alternatively as \( t \to \infty \)).

We will also require an additional condition on ordered pairs of functions \( (k_1(t), k_2(t)) \).

**Condition B.** \( k_1(tk_2(t))/k_1(t) \to 1 \) as \( t \to 0 \) (or alternatively as \( t \to \infty \)).

It is easily shown by examples that there exist functions \( k \) satisfying condition A but where \( (k, k) \) does not satisfy condition B. Typical functions satisfying both conditions A and B are logarithms and iterated logarithms and their powers and roots.
For any \( t \geq 0 \) and \( f(x) \) as in (26) let

\[
S(t) = \{ x : -\log[1 - g(x)] \leq t \},
\]

and let

\[
u(t) = \sum_{x \in S(t)} f(x).
\]

Lemma 2. If \( \psi(n) \) is given by (26) and if \( \nu(t) \sim t^\beta k(t), \quad \beta > 0, \)

as \( t \to 0 \), where \( k(t) \) is a positive function satisfying Condition A,

then

\[
\psi(n) \sim \Gamma(\beta+1) k \left( \frac{n-1}{n} \right) n^{-\beta}
\]

Proof. We may write

\[
\psi(n) = \int_0^\infty e^{nt\log[1 - g(x)]} d\mu(x)
\]

where for any \( y \)

\[
u(y) = \sum_{x \leq y} f(x).
\]

Equivalently,

\[
\psi(n) = \int_0^\infty e^{-nt\nu(t)} d\nu(t).
\]

The desired result is now equivalent to a standard Abelian proposition for Laplace transforms (see e.g., Doetsch [2] page 460) and the proof is
complete.

Now for $f^*(x) \geq 0$, $g^*(x) \geq 0$ let

\begin{equation}
S^*(t) = \{x: g^*(x) \leq t\},
\end{equation}

\begin{equation}
u^*(t) = \sum_{x \in S^*(t)} f^*(x).
\end{equation}

Lemma 3. If $f(x) \sim f^*(x)$, $g(x) \sim g^*(x) \to 0$, as $x \to \infty$, and if $\nu^*(t+o(t)) \sim \nu^*(t)$ as $t \to 0$, then

\[ \nu(t) \sim \nu^*(t), \quad t \to 0. \]

Proof: For arbitrary fixed $\varepsilon > 0$, there exists an $x_{\varepsilon}$ such that $x > x_{\varepsilon}$ implies

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

and

\[ (1-\varepsilon)f^*(x) < f(x) < (1+\varepsilon)f^*(x). \]

Furthermore, for all sufficiently small $t$, $S(t)$ and $S^*(t)$ will contain only $x$'s greater than $x_{\varepsilon}$, so that for all such $t$

\[ S^*(\frac{t}{1-\varepsilon}) \supset S(t) \supset S^*(\frac{t}{1+\varepsilon}), \]

and hence
\[(1-\varepsilon)\nu^*\left(\frac{t}{1+\varepsilon}\right) < \nu(t) < (1+\varepsilon)\nu^*\left(\frac{t}{1+\varepsilon}\right).\]

The desired result follows since \(\varepsilon\) is arbitrary and \(\nu^*(t) \sim \nu^*(t+o(t))\) as \(t \to 0\).

**Lemma 4.** If \(\phi(n)\) is given by (26), and if \(f(x) \sim f_1(x)x^{-s-1}\) and
\(g(x) \sim g_1(x)x^{-r}\) as \(x \to \infty\), where \(r, s > 0\), and where \(f_1(x)\) and \(g_1(x)\) are positive and satisfy condition A, and the pairs
\(\{g_1(x), [g_1(x)]^{\frac{1}{r}}\}\) and \(\{f_1(x), [g_1(x)]^{\frac{1}{r}}\}\) satisfy condition B, then
\[
\phi(n) \sim \frac{1}{n^r} \left[ \frac{s}{r} + \frac{1}{1} \right] f_1\left( \frac{1}{n^r} \right) \left[ g_1\left( \frac{1}{n^r} \right) \right] - \frac{s}{r} n - \frac{s}{r},
\]
as \(n \to \infty\).

**Proof.** Let \(f_1(x)x^{-s-1}\) and \(g_1(x)x^{-r}\) play the roles of \(f^*(x)\) and \(g^*(x)\) in lemma 3, so that
\[
S^*(t) = \{x: g_1(x)x^{-r} \leq t\}.
\]
If for each \(t\) we let \(x_t\) be the smallest \(x\) in \(S^*(t)\) we clearly have
\[
(31) \quad g_1(x_t)x_t^{-r} \sim t, \quad \text{as } t \to 0.
\]
Using the fact that \(\{g_1(x), [g_1(x)]^{\frac{1}{r}}\}\) satisfies condition B we see that (31) is equivalent to
\[
(32) \quad x_t \sim \left( g_1\left( t^{-\frac{1}{r}} \right) \right)^{-\frac{1}{r}}, \quad \text{as } t \to 0.
\]
Now as \( t \to 0 \), \( x_t \to \infty \) and

\[
v^*(t) = \sum_{x > x_t} f_1(x)x^{-(s+1)}
\]

\[
\sim \int_{x_t}^{\infty} f_1(x)x^{-(s+1)} \, dx = f_1(x_t)x_t^{-s} \int_1^{\infty} \frac{f_1(\frac{yx_t}{x_t})}{f_1(x_t)} y^{-(s+1)} \, dy.
\]

By Karamata [5] (page 45), we may write

\[(33) \quad f_1(x) = c(x) \exp \left\{ \int_0^x \frac{\epsilon(u)}{u} \, du \right\},\]

where \( c(x) + c > 0 \) as \( x \to \infty \) and \( \epsilon(u) \to 0 \) as \( u \to \infty \). For arbitrary \( \epsilon > 0 \), we may choose \( t_\epsilon \) small enough so that \( 0 < t < t_\epsilon \) implies \n
\[1 - \epsilon < c(yx_t)/c(x_t) < 1 + \epsilon \]

for all \( y > 1 \), and \( |\epsilon(x)| < \epsilon \) for \( x > x_t \).

Then for \( t < t_\epsilon \),

\[
\int_1^{\infty} \frac{f_1(\frac{yx_t}{x_t})}{f_1(x_t)} y^{-(s+1)} \, dy \leq (1+\epsilon) \int_1^{\infty} y^{-(s+1)} \exp \left\{ \int_{x_t}^{\infty} \frac{\epsilon(y)}{y} \, dy \right\} dy
\]

\[= (1+\epsilon) \int_1^{\infty} y^{-(s+1)+\epsilon} \, dy = \frac{1+\epsilon}{s-\epsilon}.
\]

Similarly, the integral may be bounded below for \( 0 < t < t_\epsilon \) by \( (1-\epsilon)/(s+\epsilon) \) and since \( \epsilon \) is arbitrary we have

\[v^*(t) \sim \frac{1}{s} f_1(x_t)x_t^{-s}, \text{ as } t \to 0.\]
This together with (32) and the assumption that \( \left( f_1(x), \left[ g_1(x) \right]^\frac{1}{r} \right) \) satisfies condition B yields

\[
\nu^*(t) \sim \frac{1}{e} f_1(t^{-\frac{1}{r}}) \left( \frac{t}{g_1(t^{-\frac{1}{r}})} \right)^\frac{s}{r}, \quad \text{as } t \to 0.
\]

The desired result follows by lemmas 2 and 3 upon noting that since \( \nu^*(t) \) behaves like a power of \( t \) as \( t \to 0 \), we have \( \nu^*(t+o(t)) \sim \nu^*(t) \).

**Lemma 5.** If \( \phi(n) \) is given by (26) and if \( f(x) \sim f_1(e^{x^\beta}) e^{-sx^\beta} \) and \( g(x) \sim g_1(e^{x^\beta}) e^{-rx^\beta} \) as \( x \to \infty \), where \( 0 < \beta \leq 1 \) and \( r, s > 0 \), and where \( f_1(x) \) and \( g_1(x) \) are positive and satisfy condition A,

\[
\left[ g_1(x), \left[ g_1(x) \right]^\delta \right]
\]

satisfies condition B and \( \left( f_1(x), \left[ g_1(x) \right]^\delta \right) \) satisfies condition B uniformly for \( \delta \) in an open interval containing \( \frac{1}{r} \), then

\[
\phi(n) \sim C_\beta \left( \frac{s}{r} + 1 \right) f_1(n^\frac{1}{r}) \left( \frac{1}{g_1(n^\frac{1}{r})} \right)^{-\frac{s}{r}} \left( \log n \right)^{\frac{1}{r} - 1} \frac{s}{n},
\]

as \( n \to \infty \), where

\[
C_\beta = \begin{cases} 
\frac{1}{\beta sr^\beta - 1}, & \beta < 1 \\
1/ \left( 1 - e^{-s} \right), & \beta = 1.
\end{cases}
\]

**Proof.** Proceeding as in lemma 4 and setting \( y = e^{x^\beta} \) we have

\[
\nu^*(t) = \{ x : g_1(y)y^{-r} \leq t \}.
\]
If $x_t$ is the smallest $x$ in $S^*(t)$ it corresponds to the smallest $y$ which is given asymptotically by (32). Thus

$$x_t \approx \left[ \log \left( \frac{t}{g_1(t - \frac{1}{r})} \right) \right]^{\frac{1}{r}} \frac{1}{\beta r},$$

that is

$$x_t = r^{\frac{1}{\beta}} \left[ \log \frac{1}{t} + (1+o(1)) \log g_1(t) \right]^{\frac{1}{\beta}}.$$

Now using (33) and choosing $t_\varepsilon$ for arbitrary $\varepsilon$, $0 < \varepsilon < s$, such that $0 < t < t_\varepsilon$ implies $1 - \varepsilon < c(e^{x_\beta})/c(e^{x_t}) < 1 + \varepsilon$ and $|\varepsilon(x)| < \varepsilon$ for $x \geq x_t$, we have for $t > t_\varepsilon$,

$$v^*(t) = \sum_{x > x_t} f_1(e^{x_\beta}) e^{-sx_\beta} < (1+\varepsilon)f_1(e^{x_t}) \sum_{x > x_t} \exp \left\{ -sx_\beta + \int_{x_t}^{e^{x_\beta}} \frac{\varepsilon}{u} du \right\} \left( e^{x_\beta} \int_{x_t}^{e^{x_\beta}} \frac{\varepsilon}{u} du \right) \left( e^{x_\beta} \int_{x_t}^{e^{x_\beta}} \frac{\varepsilon}{u} du \right)$$

$$= (1+\varepsilon)f_1(e^{x_t}) e^{-x_t^\beta} \sum_{x > x_t} e^{-(s-\varepsilon)x_\beta}.$$

For the case $\beta < 1,$
\[
\sum_{x \geq x_t} e^{-(s-\varepsilon)x^\beta} \leq \int_{x_t-1}^{\infty} e^{-(s-\varepsilon)x^\beta} \, dx,
\]
and making the transformation \( v = x^\beta - (x_t-1)^\beta \), we have
\[
\sum_{x \geq x_t} e^{-(s-\varepsilon)x^\beta} \leq \frac{1}{\beta} e^{-(s-\varepsilon)(x_t-1)^\beta} (x_t-1)^{1-\beta} \int_0^\infty e^{-(s-\varepsilon)v} \left( \frac{v}{(x_t-1)^\beta} + 1 \right)^{\frac{1}{\beta} - 1} \, dv.
\]

As \( x_t \to \infty \) the integral on the right approaches \((s-\varepsilon)^{-1}\) by dominated convergence so that by (35)
\[
v^*(t) < \frac{1+\varepsilon}{\beta(s-\varepsilon)} f_1(e^{x_t^\beta/(x_t-1)^\beta}) e^{x_t^\beta} - (s-\varepsilon)(x_t-1)^\beta,
\]
for all sufficiently small \( t \). Similarly, reversing the inequality in (36) by changing the lower limit to \( x_t \), we have for all sufficiently small \( t \),
\[
v^*(t) > \frac{1-\varepsilon}{\beta(s+\varepsilon)} f_1(e^{x_t^\beta}) x_t^{1-\beta} e^{-x_t^\beta} - s x_t^\beta.
\]

Hence, since \( \varepsilon \) is arbitrary and \( x_t^\beta - (x_t-1)^\beta = o(1) \) for \( \beta < 1 \), we have
\[
v^*(t) \sim \frac{1}{\beta s} f_1(e^{x_t^\beta}) x_t^{1-\beta} e^{-x_t^\beta},
\]
as \( t \to 0 \) for the case \( \beta < 1 \).
For the case \( \beta = 1 \), the integral approximation (36) is inappropriate and instead we observe that

\[
\sum_{x \geq x_t} e^{-(s^t \varepsilon) x} = \frac{e^{-(s^t \varepsilon) x_t}}{1 - e^{-(s^t \varepsilon)}},
\]

and obtain upper and lower bounds for \( \upsilon^*(t) \) using (35) and its lower analogue. Thus, for the case \( \beta = 1 \),

\[
(38) \quad \upsilon^*(t) \sim \frac{1}{1 - e^{-s}} f_1(e^{x_t}) e^{-sx_t},
\]

and we note that the only difference between (38) and (37) with \( \beta = 1 \) is the constant factor. Hence, for \( \beta < 1 \), by (34), (37), (38) and the fact that \( \log g_1(t^{-\frac{1}{\beta}}) = o\left(\log \frac{1}{t}\right) \), we have

\[
\upsilon^*(t) \sim C_{\beta} f_1\left( t^{-\frac{1}{\beta}} \left[ g_1\left( t^{-\frac{1}{\beta}} \right) \right]^{\frac{1}{\beta}} \right) \left[ g_1\left( t^{-\frac{1}{\beta}} \right) \right]^{-\frac{s}{\beta}} \left[ \log \frac{1}{t} \right]^{\frac{1}{\beta}} - 1 \frac{s}{t^\alpha}.
\]

Recalling that \( (f_1, [g_1]^\varepsilon) \) satisfies condition B uniformly for \( \varepsilon \) in an open interval containing \( \frac{1}{\beta} \) this becomes

\[
\upsilon^*(t) \sim C_{\beta} f_1\left( t^{-\frac{1}{\beta}} \left[ g_1\left( t^{-\frac{1}{\beta}} \right) \right]^{\frac{1}{\beta}} \right) \left[ g_1\left( t^{-\frac{1}{\beta}} \right) \right]^{-\frac{s}{\beta}} \left[ \log \frac{1}{t} \right]^{\frac{1}{\beta}} - 1 \frac{s}{t^\alpha}.
\]

The desired result follows by lemmas 2 and 3.

The next three lemmas elucidate the asymptotic behavior of the unconditional probabilities \( p(x) \) for various families \( p_\lambda(x) \) given
by (1) and certain classes of prior distributions \( G(\lambda) \). The first two lemmas are concerned with the case where the function \( h(x) \) appearing in (1) behaves asymptotically like a power of \( x \). For this case the natural parameter range is clearly \( 0 \leq \lambda < 1 \).

**Lemma 6.** If \( p_\lambda(x) \) is given by (1) with

\[
h(x) \sim h_1(x)x^\gamma, \quad \gamma > -1, \quad \text{as} \quad x \to \infty,
\]

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

\[
G'(\lambda) \sim G_1(1-\lambda)(1-\lambda)^\sigma, \quad \sigma > -1, \quad \text{as} \quad \lambda \to 1,
\]

where \( h_1(x) \) and \( G_1(1-\lambda) \) satisfy condition A, then

\[
p(x) \sim \frac{\Gamma(\gamma+\sigma+2)}{\Gamma(\gamma+1)} G_1 \left( \frac{1}{2} \right) x^{-\sigma-2}.
\]

**Proof.** Letting \( \lambda = e^{-\theta} \), we have

\[
(39) \quad p(x) = h(x) \int_{0}^{1} \lambda^x h(\lambda) dG(\lambda) = -h(x) \int_{0}^{\infty} e^{-\theta x} \beta(e^{-\theta}) dG(e^{-\theta}).
\]

Now

\[
\beta(e^{-\theta})^{-1} = \sum_{x} e^{-\theta x} h(x) = \int_{0}^{\infty} e^{-\theta x} h(x) d\mu(x),
\]

where \( \mu(x) \) is counting measure on the integers. Thus, by the Abelian theorem cited in lemma 2,
\[ (40) \quad \beta (e^{-\theta})^{-1} \sim \Gamma(\gamma+1) h_1 \left( \frac{1}{\beta} \right) e^{-(\gamma+1)}, \quad \text{as} \quad \theta \to 0. \]

Also, letting \( G_0(\theta) = -G(e^{-\theta}) \), we have

\[ G_0'(\theta) \sim G_1(\theta) \sigma, \quad \text{as} \quad \theta \to 0. \]

The Abelian theorem then yields

\[ -\log(1-\varepsilon) \int_0 e^{-\theta x} \beta (e^{-\theta}) G_0'(\theta) d\theta \sim \frac{\Gamma(\gamma+\sigma+2)}{\Gamma(\gamma+1)} \frac{G_1 \left( \frac{1}{x} \right)}{h_1(x)} x^{-(\gamma+\sigma+2)}, \]

as \( x \to \infty \). The desired result then follows from (39) upon dividing the range of the right-hand integral into \((0, -\log(1-\varepsilon))\) and \([-\log(1-\varepsilon), \infty)\), and observing that the contribution of the second interval is exponentially small in \( x \).

**Lemma 7.** If \( p_\lambda(x) \) is given by (1) with

\[ h(x) \sim h_1(x) x^\gamma, \quad \gamma > -1, \quad \text{as} \quad x \to \infty, \]

where \( h_1(x) \) satisfies condition A, and if \( G'(\lambda) \) exists for \( 1 - \varepsilon < \lambda < 1 \) for some \( \varepsilon > 0 \), and

\[ G'(\lambda) \sim C \delta \sigma, \quad \delta, \sigma > 0, \quad \text{as} \quad \lambda \to 1, \]

then
\[ p(x) \sim C_1 \frac{h_1(x)}{h_1(x^{\sigma+1})} \frac{\sigma(x^{\sigma+1})^{-2}}{x} \exp\left\{-C_2 x \frac{\sigma}{1+\sigma}\right\}, \]

as \( x \to \infty \), where

\[ C_1 = C_0 \frac{\Gamma(\frac{\gamma+\frac{1}{2}}{\sigma+1})}{\Gamma(\gamma+1)(\sigma+1)^{1/2}}, \quad C_2 = \left(1 + \frac{1}{\sigma}\right)(\delta \sigma)^{1/2}. \]

**Proof.** Letting \( \lambda = e^{-\theta} \), we have \( \beta(e^{-\theta})^{-1} \) satisfying (40) so that for sufficiently small \( \epsilon > 0 \),

\[ p(x) \sim C_0 \frac{h_1(x) x^\gamma}{\Gamma(\gamma+1)} \int_{0}^{\epsilon} e^{-\theta(x+\delta \sigma) - \sigma \delta \sigma^{1/2}} \frac{1}{h_1(\theta)} \theta^{1/2} \theta^{1+1} d\theta, \]

as \( x \to \infty \), provided the right hand integral is of larger order than \( e^{-\epsilon x} \) which bounds the order of the neglected portion of the integral. Now letting \( t = \theta(x/\delta \sigma)^{1/2} - 1 \), noting that \( h_1(x^{\sigma+1})/h_1((t+1)^{-1} x^{\sigma+1}) \sim 1 \) uniformly for \( -\epsilon < t < \epsilon \) as \( x \to \infty \), and neglecting exponentially small contributions, the integral in (41) is seen to be asymptotic to

\[ h_1(x^{\sigma+1})^{-1} \left(\frac{\delta \sigma}{x}\right)^{\gamma+2} \exp\left\{-x \frac{\sigma}{\sigma+1} \left[\frac{1}{\sigma+1} + \delta(\delta \sigma)^{-\frac{\sigma}{\sigma+1}}\right]\right\} \int_{-\epsilon}^{\epsilon} \exp\left\{-x \frac{\sigma}{\sigma+1}(\delta \sigma)^{1/2} \left[t + \frac{1}{\sigma}(t+1)^{1/2} - \frac{1}{\sigma}\right]\right\} (t+1)^{1+1} dt, \]

for large \( x \). The expression in [ ] in the exponential term of the integrand is of the form \( \frac{1}{2}(\sigma+1)t^2(l+o(t)) \) as \( t \to 0 \), so that by the well-
known asymptotic theorem for Laplace integrals (see e.g., De Bruijn [ ] pp. 63-65) the integral in (42) is asymptotic to \( \frac{1}{2} \left( \sigma + 1 \right)^{\frac{1}{2}} \left( \sigma \right)^{\frac{1}{2}} \left( x^{\sigma + 1} \right)^{\frac{1}{2}} \) as \( x \to \infty \), and the desired result follows.

**Remark 2.** In lemmas 6 and 7 we required \( \gamma > -1 \) in the expressions 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 6 this leads to \( \lambda^{-(\sigma - \gamma + 1)} \) for the power of \( x \) in the asymptotic expression for \( p(x) \), and \( h_1(x) \) appears as a factor. In lemma 7 the power of \( x \) is reduced by \( (\gamma + 1)/(\sigma + 1) \) and the factor \( h_1 \left( \frac{1}{\sigma + 1} \right)^{1} \) disappears. The constant factors also change in both lemmas.

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

**Lemma 8.** If \( h(x) = \Gamma(x+1)^{-1} \) and \( G'(\lambda) \) exists for all sufficiently large \( \lambda \) with

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

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

\[
p(x) = \frac{1}{\Gamma(x+1)} \int_0^\infty \lambda^x e^{-\lambda} d\lambda \sim \frac{C_0}{\Gamma(x+1)} \int_0^\infty \lambda^{x+\delta} e^{-(\sigma + 1)\lambda} d\lambda,
\]
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 $\Gamma(x+\delta+1)/(x+1)^{x+\delta+1}$ and the desired result follows since

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

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 9.** 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

$$G'(\lambda) \sim G_1(\lambda_0 - \lambda)(\lambda_0 - \lambda)^\sigma, \sigma > -1, \text{ as } \lambda \to \lambda_0,$$

where $G_1$ satisfies condition A, then

$$p(x) \sim \beta(\lambda_0)\Gamma(\sigma+1)h(x)G_1(\frac{1}{2})x^{-\sigma-1}\lambda_0^{x+\sigma+1},$$

as $x \to \infty$.

**Proof.** Letting $\lambda_0 = e^{-\theta_0}$ and $\lambda = e^{\theta-\theta_0}$ and noting that $\beta(\lambda) = \beta(\lambda_0) > 0$ as $\lambda \to \lambda_0$ for this case, we have

30
\[ p(x) = h(x) \int_0^{\lambda_0} \lambda^x \beta(\lambda) dG(\lambda) \]

\[ \sim - \beta(\lambda_0) e^{-\theta_0} x h(x) \int_0^{\infty} e^{-\theta x} dG(e^{-\theta-\theta_0}). \]

Letting \( G_0(\theta) = -G(e^{-\theta-\theta_0}) \), we have for sufficiently small \( \theta \),

\[ G'_0(\theta) = e^{-\theta-\theta_0} G'(e^{-\theta-\theta_0}) \]

\[ \sim e^{-\theta-\theta_0} G_1(e^{-\theta_0}(1-e^{-\theta})) \left[ e^{-\theta_0}(1-e^{-\theta}) \right]^\sigma \]

\[ \sim e^{-(\sigma+1) \theta} G_1(\theta) \theta^\sigma, \]

as \( \theta \to 0 \). The desired result follows by the argument used in lemma 6.
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 \( h_1(x) \) satisfies condition A, 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

\[
(i) \quad G'(\lambda) \sim G_1(1-\lambda)G, \quad \sigma > -1, \quad \text{as} \quad \lambda \to 1,
\]

where \( G_1 \) satisfies condition A and \( \left( G_{1i}, \frac{1}{\sigma+2} \right) \) satisfies condition B for \( i = 1, 2 \), then as \( n \to \infty \),

\[
r_n - r^*(G) = e^{0(1)} \left[ G_1 \left( n^{-\frac{1}{\sigma+2}} \right) \right]^{\frac{1}{\sigma+2}} n^{-\frac{\sigma+1}{\sigma+2}}.
\]

Alternatively, if

\[
(ii) \quad G'(\lambda) \sim C_0 e^{-\delta(\log \frac{1}{\lambda})^{-\sigma}} \text{, } \delta, \sigma > 0, \text{ as } \lambda \to 1,
\]
and if \[ \left[ h_1(x)/h_1(x^{\frac{1}{\alpha+1}}) \right]^{1}, \left[ h_1(x)/h_1(x^{\frac{1}{\alpha+1}}) \right]^{\xi} \]
satisfies condition B uniformly for \( \xi \) in some open interval containing \( (1+\frac{1}{\alpha})^{-1} (\delta \sigma) \)
for \( i = 1,2 \), then as \( n \to \infty \),

\[ r_n - r^*(G) = e^{o(1)} \left( \log n \right)^{\frac{1}{\sigma-1}}. \]

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

\[ \text{(iii) } G'(\lambda) \sim G_{1}(\lambda_0-\lambda)(\lambda_0-\lambda)^{\sigma}, \sigma > -1, \text{ as } \lambda \to \lambda_0, \]

where \( G_1 \) satisfies condition A and \( \left[ G_{1}(\frac{1}{x})h_1(x) \right]^{1}, \left[ G_{1}(\frac{1}{x})h_1(x) \right]^{\xi} \)
satisfies condition B uniformly for \( \xi \) in an open interval containing \( (-\log \lambda_0)^{-1} \) for \( i = 1,2 \), 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 (i) and (ii) we have \( p(x+1)/p(x) \to 1 \) as \( x \to \infty \) by lemmas 6 and 7, and for case (iii), \( p(x+1)/p(x) \to \lambda_0 > c, \) as \( x \to \infty \), by lemma 9. Thus, by (7), (11) and (12) we see that for each case, for some \( c_1, c_2 > 0 \)

\[ \alpha(x) \sim c_1 p(x) \text{, and } \sigma^2(x) \sim c_2 p(x), \text{ as } x \to \infty. \]

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

33
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
\[
f(x) \sim c_1 p^2(x), \quad g(x) \sim 2p(x), \quad \text{as} \quad x \to \infty.
\]
If (i) holds then the desired conclusion follows for the lower bound by lemmas 4 and 6 upon setting \( g_1(x) = 2\Gamma(\gamma+\sigma+2)\Gamma(\gamma+1)^{-1}G_{\gamma} \left( \frac{1}{x} \right) \) and
\[
f_1(x) = \frac{1}{4} c_1 g^2_1(x),
\]
and observing that \( \sigma = \sigma + 2 \) and \( s = 2(\sigma+2)-1 \) in lemma 4 for this case. Similarly, the desired conclusions for the lower bound are obtained for the cases where (ii) or (iii) hold by applying lemma 5 to the results of lemma 7 for case (ii) and lemma 9 for case (iii).

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)
\[
(45) \quad m_x(t) = p(x+1)e^{w(x)t} + p(x)e^{-ct},
\]
so that condition 2 is also satisfied. Furthermore, differentiating (45) twice yields
\[
(46) \quad \ddot{m}_x(t) = w^2(x)p(x+1)e^{w(x)t} + c_1^2 p(x)e^{-ct},
\]
and by (43), \( \tau |a(x)|/\sigma^2(x) \sim \tau c_1/c_2 \) 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 (1) holds the desired conclusion for the upper bound follows by lemmas 4 and 6 upon setting \( f_1(x) = c_0^{-1} c_1^{-1} c_2 g_1(x) = c_1 \Gamma(\gamma+2) \Gamma(\gamma+1)^{-1} G_1 \left( \frac{1}{x} \right) \) and \( r = s = \sigma + 2 \) in lemma 4. The upper bounds for cases (ii) and (iii) are verified similarly using lemmas 5, 7 and 9 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 (iii) of the theorem is best possible.

Although algebraically slower rates occur under hypothesis (i) of the theorem, the prior distributions satisfying this condition put excessive weight in the right tail (near one) to the extent that the un-
conditional distribution \( p(x) \) lacks higher moments (by lemma 6). 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

\[
G'(\lambda) = (\gamma+1)(1-\lambda)^\gamma, \quad 0 \leq \lambda < 1, \quad \gamma > -1,
\]

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 [4].

**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 \( a_n(x) \) given by

36
(10), then if $G'(\lambda)$ exists for all sufficiently large $\lambda$ with

\[(i) \quad G'(\lambda) \sim C_0 \lambda^\delta, \quad \delta < -2,\]

as $\lambda \to \infty$, then

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

Alternatively, if

\[(ii) \quad G'(\lambda) \sim C_0 \lambda^\delta e^{-\sigma \lambda}, \quad \sigma > 0,\]

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 8,

\[(47) \quad \alpha(x) \sim c_1 x^{\delta + 1} (\sigma + 1)^{-x}, \quad \sigma^2(x) \sim c_1 x^{\delta + 2} (\sigma + 1)^{-x}\]

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

\[f(x) \sim c_2 x^{2\delta + 1} (\sigma + 1)^{-2x}, \quad g(x) \sim c_3 x^{\delta} (\sigma + 1)^{-x}\]

as $x \to \infty$, for suitable $c_2, c_3 > 0$, by lemma 8. The required conclusion for the lower bound then follows for case (i)($\sigma = 0$) by lemma 4 with
\[ f_1(x) = c_2 \log^2 x, \quad g_1(x) = c_3 \log x. \]

The conclusion for case (ii) follows by lemma 5 with \( \beta = 1, \quad 2r = s = 2 \log(\sigma+1), \)
\[ f_1(x) = c_2 \log^2 x \quad \text{and} \quad g_1(x) = c_3 \log x. \]

The upper bound (14) is of the form of \( \phi(n) \) given by (26) with
\[ f(x) \sim c_1 x^{\delta+1} (\sigma+1)^{-x}, \quad g(x) \sim c_1 x^{\delta} (\sigma+1)^{-x}, \]
as \( x \to \infty \) by (47). Also, by (47) and (45) conditions 1 and 2 are satisfied, and by (47) \( \tau |a(x)| / \sigma^2(x) \sim x^{-1} \) as \( x \to \infty \), so that by (46) and lemma 8 condition 3 is satisfied. The desired result follows for case (i) by lemma 4 with \( f_1(x) = g_1(x) = c_1, \quad r = -\delta \quad \text{and} \quad s = -\delta-2. \)

The conclusion for case (ii) follows by lemma 5 with \( \beta = 1, \quad r = s = \log(\sigma+1), \)
\[ f_1(x) = [\log x]^{\delta+1} \quad \text{and} \quad g_1(x) = [\log x]^{\delta}. \]

This completes the proof of the theorem.

**Remark 7.** The conclusions of theorem 4 are completely analogous to those for the corresponding cases (i) and (ii) 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 (ii) of theorem 4 is related to the fact that we must have \( EA < \infty \) 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 \( n \sum_{x} x p(x) [1-2p(x)]^{2-1} \)
and if we choose \( x_n \) such that \( p(x_n) \sim \frac{1}{n} \) as \( n \to \infty \), then the expression for the bound is itself bounded below by \( e^{-2x_n/n} \), 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 (ii) we see that the exact rate is \( (\log n)n^{-1} \) for this case. Thus again, as 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 [3], to which the results of sections 2 and 3 apply. Such an example is provided by the reparameterized negative binomial distribution,

\[
p_\lambda(x) = h(x) \left( \frac{k}{k+x} \right)^k \left( \frac{\lambda}{k+x} \right)^x, \quad \lambda > 0, \quad x = 0,1, \ldots
\]

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

\[
h(x) = \frac{k(k+1) \ldots (k+x-1)}{x!}, \quad x = 1,2, \ldots.
\]
The appropriate choice for $Z_j(x)$ in this case is

$$Z_j(x) = \begin{cases} \frac{k h(x)}{h(x+y)}, & X_j = x + y, \ y = 1,2, \ldots \\ -c, & X_j = x \\ 0, & \text{otherwise} \end{cases}$$

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


Convergence Rates for Empirical Bayes Two-Action Problems I. Discrete Case

A sequence of decision problems is considered where for each problem the observation has discrete probability function of the form \( p(x) = h(x) \beta(\lambda) \lambda^x, \) \( x = 0,1,2,\ldots, \) and where \( \lambda \) is selected independently for each problem according to an unknown prior distribution \( G(\lambda) \). It is supposed that for each problem one of two possible actions (e.g., "accept" or "reject") must be selected. Under various assumptions about \( h(x) \) and \( G(\lambda) \) the rate at which the risk of the \( n \)th problem approaches the smallest possible risk is determined for standard empirical Bayes procedures. It is shown that for most practical situations, the rate of convergence to "optimality" will be at least as fast as \( L(n)n^{-1} \) where \( L(n) \) is a slowly varying function (e.g., \( \log n \)). The rate cannot be faster than \( n^{-1} \) and this exact rate is achieved in some cases. Arbitrarily slow rates will occur in certain pathological situations.
### UNCLASSIFIED

#### Security Classification

<table>
<thead>
<tr>
<th>KEY WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision theory</td>
</tr>
<tr>
<td>Repeated problems</td>
</tr>
<tr>
<td>Empirical Bayes</td>
</tr>
</tbody>
</table>

### INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

3. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

4. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

8. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

9. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

10. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

11. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

12. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

13. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   (1) "Qualified requesters may obtain copies of this report from DDC."

   (2) "Foreign announcement and dissemination of this report by DDC is not authorized."

   (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through"

   (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through"

   (5) "All distribution of this report is controlled. Qualified DDC users shall request through"

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

14. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

15. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (F), (O), (C), or (U). There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

16. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.