Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Statistics & Probability Letters 76 (2006) 547–554
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Minimax estimation of a bounded parameter of a discrete distribution

Éric Marchand[a,*,1], Ahmad Parsian[b,2]

[a]*University of New Brunswick and Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, Canada, J1K 2R1*
[b]*School of Mathematical Sciences, Isfahan University of Technology, Isfahan, 84156, Iran*

Received 26 October 2004; received in revised form 6 July 2005
Available online 29 September 2005

Abstract

For a vast class of discrete model families with cdf's F_θ, and for estimating θ under squared error loss under a constraint of the type $\theta \in [0, m]$, we present a general and unified development concerning the minimaxity of a boundary supported prior Bayes estimator. While the sufficient conditions obtained are of the expected form $m \leqslant m(F)$, the approach presented leads, in many instances, to both necessary and sufficient conditions, and/or explicit values for $m(F)$. Finally, the scope of the results is illustrated with various examples that, not only include several common distributions (e.g., Poisson, Binomial, Negative Binomial), but many others as well.

MSC: primary 62C10; 62F10; 62F30; 60E99

Keywords: Minimax estimation; Restricted parameter space; Discrete distributions; Squared error loss

1. Introduction

In many statistical problems, there exists bounds on the values that unknown parameters can take. Correspondingly, a large body of work concerned with estimation problems in restricted parameter spaces has emerged, as reviewed for instance by van Eeden (1996) or Marchand and Strawderman (2004). One influential, much studied, and useful criterion to select or to evaluate a procedure is minimaxity (e.g., Brown, 1994; Strawderman, 2000).

We are concerned here with minimax estimation, under squared error loss, of an unknown parameter θ for certain kinds of discrete families of distributions under a constraint of the type $\theta \in [0, m]$. More specifically, we focus on explicit conditions for the minimax estimator to be Bayes with respect to a boundary supported prior. It follows here from the work of DasGupta (1985) that the least favourable prior is quite generally supported on the boundary $\{0, m\}$ of the parameter space; and that the corresponding Bayes estimator is minimax; for

*Corresponding author. Tel.: +1 819 821 8000; fax: +1 819 821 7189.
E-mail addresses: eric.marchand@usherbrooke.ca (É. Marchand), ahmad_p@cc.iut.ac.ir (A. Parsian).
[1]Research supported by NSERC of Canada and by the University of New Brunswick Academic Fund.
[2]Research supported by a grant of the Research Council of the Isfahan University of Technology.

0167-7152/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2005.09.007

small enough m, say $m \leqslant m^*$ with $m^* > 0$. Since the phenomenon also extends in various univariate and multivariate settings (DasGupta, 1985), and to other strictly convex loss functions (e.g., Bader and Bischoff, 2003), it has lead researchers to analytical investigations of m^* in many specific settings. In particular for discrete models, analytical investigations of m^* appeared in: (i) Marchand and MacGibbon (2000) for the Binomial model under both scaled and unscaled squared error losses, in (ii) Wan et al. (2000) for a Poisson model and Linex loss, and in (iii) Johnstone and MacGibbon (1992) for a Poisson model and scaled squared error loss.

But, such analytical investigations often focus on either a single model and/or sufficient conditions for the least favourable prior to be boundary supported, in other words lower bounds for m^*. Our work provides a general and unified development applicable to a large class of discrete models. Members of this class include Poisson and Consul's (1989) Generalized Poisson, Binomial, Negative Binomial, Waring, many members of the "Power Series" family of distributions, and various types of mixtures and "Stopped Sum" distributions. Moreover, we are able to specify not only sufficient conditions, but also in many cases necessary and sufficient conditions for least favourable priors to be supported on the boundary of the parameter space. Finally, in Section 3, we give explicitly these conditions for various cases.

2. Preliminaries

Our results apply to observable random vectors $X = (X_1, \ldots, X_n); n \geqslant 1$; where: (i) the X_i's are identically, but not necessarily independently distributed, discrete random variables with joint probability function $p_\theta(x) = P_\theta(X = x)$, and (ii) to situations where the support of $P_\theta(X_1 = x_1)$ is lower bounded (say by $\underline{s}$). We will be concerned with minimax estimation of θ, under squared error loss, in situations where θ is constrained to a known, and small enough interval $[a, b]$, and where the distribution of X under $\theta = a$ is degenerate at $(\underline{s}, \ldots, \underline{s})$. Without loss of generality, given that we can reparametrize and translate, we assume hereafter that $[a, b] = [0, m]$ and $\underline{s} = 0$. In contrast to this constrained parameter space, we will refer to the unconstrained parameter space as $\Theta = \{\theta : p_\theta(\cdot) \text{ is a valid probability function}\}$. The degenerate property of X at $\theta = 0$ renders possible the minimaxity results below, but it is hardly an unusual or exceptional property as it is shared by many common discrete distributions.

2.1. Distributional assumptions

It will be convenient to set $A = \{x \in \Re^n : \sum_{i=1}^n x_i = 0\}$, and denote $G(n, \theta)$ as follows.

Definition 1. We define $G(n, \theta) = P_\theta(X \in A)$; and in cases where the X_i's are independent with common cdf F, we will sometimes denote $G(n, \theta)$ by $G_F(n, \theta)$; (recall that our above conditions imply that $G(n, 0) = 1$).

We will assume throughout this paper that $G(n, \theta)$ is three times differentiable for $\theta \in \Theta$. We now describe the class C of families of distributions for which the main result of this paper, Theorem 1, applies.

Definition 2. We define C as the class of families p_θ for X such that $G(n, 0) = 1$, and

$$(-1)^k \frac{\partial^k}{\partial \theta^k} G(n, \theta) > 0; \;\; \theta \in \Theta \quad \text{for } k = 1, 2, 3.$$

Here are some immediate examples of families p_θ that are members of C (other examples follow later in this section). Various properties of these families, as well as additional examples, can be found in the excellent reference book on discrete distributions by Johnson et al. (1993).

Example 1. (a) X_i's are independently distributed Poisson(θ), with $G(n, \theta) = e^{-n\theta}$.

(b) X_i's are independently distributed Bernoulli(θ) with $G(n, \theta) = (1 - \theta)^n$.

(c) X_i's are independently distributed as Negative Binomial (NBI(α, p)) with parameters (α, p); $\alpha > 0$ (known), $0 < p \leqslant 1$; reparametrized such that $E_\theta(X_i) = \theta = \alpha(1/p - 1)$, and $P_\theta(X_i = x) = (\Gamma(\alpha + x)/x!\Gamma(\alpha))(\alpha/(\theta + \alpha))^\alpha(\theta/(\theta + \alpha))^x I_{\{0,1,\ldots\}}(x)$. In such a case, we have $G(n, \theta) = (\alpha/(\alpha + \theta))^{n\alpha}$. Observe here that the constraint $\theta \leqslant m$ is equivalent to a lower bound constraint for p (i.e., $p \geqslant \alpha/(\alpha + m)$).

(d) X_i's are independently distributed "Generalized Poisson" (as denoted by Consul, 1989; or Lagrangian Poisson as denoted by others including Johnson et al., 1993) with parameters (θ, λ); $\lambda \geqslant 0$, and $\theta \geqslant 0$. Here $P_\theta(X_i = x) = (\theta(\theta + x\lambda)^{x-1} \mathrm{e}^{-\theta - x\lambda})/x! I_{\{0,1,\ldots\}}(x)$, and $G(n, \theta) = \mathrm{e}^{-n\theta}$ as in the above Poisson case.

(e) X_i's are independently distributed Waring (k, ρ), with $P_\theta(X_i = x) = \rho(k)_x/((k+\rho)_{x+1}) I_{\{0,1,\ldots\}}(x)$, $k > 0, \rho > 0$ and $\theta = k/\rho$. Whenever: (i) k is known and ρ is unknown, (which includes the Yule distribution for $k = 1$), or (ii) k is unknown and ρ is known, the families of p_θ's belong to C with $G(n, \theta) = (1/(1+\theta))^n$. Observe that this expression matches the Negative Binomial $G(n, \theta)$ above for $\alpha = 1$.

Observe that the above $G(n, \theta)$ are also examples of completely monotone (CM) functions, which possess alternating derivative signs of all positive order. Now, completely monotone functions (see Feller, 1966, Section XIII.4) are perhaps best known for their usefulness in characterizing functions which are Laplace transforms, but we draw the connection here mostly because they possess rather interesting properties; for instance that the product of completely monotone functions is completely monotone. The next lemma provides a springboard, or a method of identifying or generating easily new members in C from others.

Lemma 1. (a) *If $X_1, \ldots, X_n$ are independent (with common cdf F), then*

$$G_F(1, \theta) \text{ is CM} \Rightarrow G_F(n, \theta) \text{ is CM}.$$

(b) *For the hierarchical (or mixture) model such that,*

$$P(X \in A|\lambda) = G_0(n, \lambda), \quad \lambda|\theta \sim f_\theta(\lambda) = \frac{1}{\theta} f_1\left(\frac{\lambda}{\theta}\right)$$

with $E_1[\lambda^k] = \int_0^\infty \lambda^k f_1(\lambda)\,\mathrm{d}\lambda < \infty$; $k = 1, 2, \ldots$; then

$$G_0(n, \lambda) \text{ is CM in } \lambda \Rightarrow G(n, \theta) \text{ is CM in } \theta.$$

(c) *If $X_1, \ldots, X_n$ are independent with*

$$P(X_i = 0|\lambda) = G_0(1, \lambda), \quad \text{and} \quad \lambda|\theta \sim f_\theta(\lambda) = \frac{1}{\theta} f_1\left(\frac{\lambda}{\theta}\right)$$

with $E_1[\lambda^k] = \int_0^\infty \lambda^k f_1(\lambda)\,\mathrm{d}\lambda < \infty$; $k = 1, 2, \ldots$; then

$$G_0(1, \lambda) \text{ is CM in } \lambda \Rightarrow G(n, \theta) \text{ is CM in } \theta.$$

Proof. Part (a) follows given that the product of completely monotone functions is completely monotone (see Criterion 1, p. 417, Feller, 1966). Observe for part (b) that the given representation implies

$$G(n, \theta) = \int_0^\infty G_0(n, \theta\lambda) f_1(\lambda)\,\mathrm{d}\lambda \tag{1}$$

from which the result follows by the assumed completely monotonicity of G_0 and the finiteness conditions on $E_1[\lambda^k]$. Finally, part (c) is a direct consequence of parts (a) and (b). □

Remark 1. The Negative Binomial representation as a Poisson mixture is well known, and its $G(n, \theta)$ actually arises as particular instances of both models (b) and (c) of Lemma 1. Indeed, denoting $T = \sum_{i=1}^n X_i$ and referring to Example 1(c), we have as illustrations of Lemma 1's part (b) and part (c) respectively:

(i) $T \sim \mathrm{NBI}(\varepsilon, p = (1 + n\theta)^{-1})$, (with $G(n, \theta) = (1 + n\theta)^{-\varepsilon}$), whenever $X_1, \ldots, X_n|\lambda$ are independent Poisson(λ), with $\lambda|\theta \sim$ Gamma(ε, θ).
(ii) $T \sim \mathrm{NBI}(n\varepsilon, p = (1 + \theta)^{-1})$, (with $G(n, \theta) = (\varepsilon/(\varepsilon + \theta))^{n\varepsilon}$), whenever $X_1, \ldots, X_n$ are independent with $X_i|\lambda \sim$ Poisson(λ), and $\lambda|\theta \sim$ Gamma(ε, θ).

Observe that the X_i's are not independent in (i), while scenario (ii) reduces directly to the situation in part (c) of Example 1 given the independence of the X_i's and their Poisson mixture representation.

Many members of the class of families of distributions C, including those in parts (a), (b), and (c) of Example 1, can be found among the family of power series distributions, where $X_1, \ldots, X_n$ are independent with a common probability function given by

$$P_\theta(X_i = x) = \frac{a_x[\gamma(\theta)]^x}{c(\gamma(\theta))} I_{\{0,1,\dots\}}(x) \tag{2}$$

with $a_0 = 1$ (without loss of generality), $\gamma(\theta)$ being a positive and nondecreasing function of θ, and $c(\gamma) = \sum_{x\geqslant 0} a_x\gamma^x$. Here $G(n,\theta) = (1/(c(\gamma(\theta))))^n$. For such families to belong to C, it will suffice, as remarked upon in part (a) of Lemma 1, that $1/(c(\gamma(\theta)))$ be completely monotone in θ. Finally, further interesting examples of classes of families of distributions which belong to C are provided in Marchand and Parsian (2004).

2.2. Minimaxity

Our results rely on a following well-known criteria for minimaxity applied to boundary two-point priors. In the spirit of Kempthorne (1987, Theorem 2.2), the condition is presented as necessary and sufficient.

Lemma 2. *A two-point boundary prior π on $\{0, m\}$ is least favourable, and the corresponding Bayes estimator $\delta_\pi(X)$ is minimax, iff*

$$R(0,\delta_\pi) = R(m,\delta_\pi) = \sup\{R(\theta,\delta_\pi); 0\leqslant\theta\leqslant m\}. \tag{3}$$

Now, for our discrete models, Bayes estimators $\delta_\pi(X)$ corresponding to two-point boundary supported priors π are of a simple form; and, as we now show, exactly one of these is an "equalizer" rule (i.e., $R(0,\delta_\pi) = R(m,\delta_\pi)$).

Lemma 3. *Among two-point boundary priors π exactly one of them leads to an equalizer Bayes rule under squared error loss, and it is given by*

$$\delta^*(x) = y^*[x\in A] + m[x\notin A] \tag{4}$$

with

$$y^* = m\,\frac{\sqrt{G(n,m)}}{1+\sqrt{G(n,m)}}. \tag{5}$$

Proof. Given that $P_0(X\in A) = 1$ and $P_m(X\in A) = G(n,m)$, the expected posterior loss $E[(d-\theta)^2|x]$ becomes, for the prior with $\pi(0) = 1-\pi(m)$:

$$(d-m)^2 \quad \text{if } x\notin A$$

and

$$\pi(0|x)d^2 + (1-\pi(0|x))(d-m)^2 \quad \text{if } x\in A, \tag{6}$$

where $\pi(0|x) = (1+[\pi(m)/\pi(0)]G(n,m))^{-1}$. From this, it follows that Bayes estimators $\delta_\pi(X)$ corresponding to two-point boundary priors π are of the form

$$\delta_\pi(x) = y[x\in A] + m[x\notin A]$$

with y minimizing (6) in d. Furthermore, observe that y is a continuous and strictly decreasing function of $\pi(0|x)$, and hence of $\pi(0)$, taking values on $[0,m]$ as $\pi(0|x)$ (or $\pi(0)$) varies on $[0,1]$. Now, evaluating the difference in risks of $\delta_\pi(X)$ at $\theta = 0$ and $\theta = m$, we have $R(0,\delta_\pi) - R(m,\delta_\pi) = y^2 - (y-m)^2G(n,m)$, which clearly admits the unique root y^* in y. $\square$

Remark 2. It is not too difficult to see that the above development giving the uniqueness of an "equalizer" two-point boundary Bayes rule also remains valid for general losses $\rho(d-\theta)$ with strictly convex ρ. This was observed for a Linex loss ρ by Wan et al. (2000, Theorem 2.1).

3. Main results

The general development that follows with Lemma 4, Lemma 5, and Theorem 1 capitalizes on:

(i) The necessary condition

$$\frac{\partial}{\partial\theta} R(\theta, \delta^*)|_{\theta=0} \leqslant 0 \tag{7}$$

for (3) to hold with $\delta_\pi(X) = \delta^*(X)$; and

(ii) the condition

$$\frac{\partial^2}{\partial^2\theta} R(\theta, \delta^*) \text{ nondecreasing in } \theta; \theta \in [0, m] \tag{8}$$

which is shown in Lemma 5 below to be sufficient for (3) to hold, with $\delta_\pi(X) = \delta^*(X)$, in cases where (7) is satisfied.

Lemma 4. *For a family $p_\theta \in C$ and squared error loss, a necessary condition for* (3) *to be satisfied with $\delta_\pi(X) = \delta^*(X)$ is $m \leqslant m_0$, where m_0 is the unique* (*positive*) *solution in m of the equation $T(m) = 0$ with*

$$T(m) = m + \left(\frac{2}{\frac{\partial}{\partial\theta} G(n, \theta)|_{\theta=0}}\right) \frac{\left(\sqrt{G(n, m)}\right)\left(1 + \sqrt{G(n, m)}\right)}{1 + 2\sqrt{G(n, m)}}. \tag{9}$$

Proof. We use (7). Given representation (4) of $\delta^*(X)$, we obtain directly

$$R(\theta, \delta^*) = (m-\theta)^2 + G(n, \theta)[(y^*-\theta)^2 - (m-\theta)^2],$$

$$\frac{\partial}{\partial\theta} R(\theta, \delta^*) = 2(\theta - m) + 2G(n, \theta)(m - y^*) + \frac{\partial}{\partial\theta} G(n, \theta)\{(y^*-\theta)^2 - (m-\theta)^2\} \tag{10}$$

and

$$\frac{\partial}{\partial\theta} R(\theta, \delta^*)|_{\theta=0} = \frac{\partial}{\partial\theta} G(n, \theta)|_{\theta=0}(y^{*2} - m^2) - 2y^*, \tag{11}$$

since $G(n, 0) = 1$. Now, substituting y^* as in (5), and solving (7) with the help of (11) yields the necessary condition

$$m \leqslant \left(\frac{-2}{\frac{\partial}{\partial\theta} G(n, \theta)|_{\theta=0}}\right) \frac{\frac{y^*}{m}}{1 - \left(\frac{y^*}{m}\right)^2} \Longleftrightarrow m \leqslant \left(\frac{-2}{\frac{\partial}{\partial\theta} G(n, \theta)|_{\theta=0}}\right) \frac{\left(\sqrt{G(n, m)}\right)\left(1 + \sqrt{G(n, m)}\right)}{1 + 2\sqrt{G(n, m)}}. \tag{12}$$

Finally, the result follows since, for $p_\theta \in C$, $G(n, m)$ is decreasing in m implying that the rhs of (12) is decreasing in m. □

Lemma 5. *For a family $p_\theta \in C$, with $m \leqslant m_0$, condition* (8) *is sufficient* (*and necessary by virtue of Lemma* 4) *for* (3) *to hold, under squared error loss.*

Proof. Condition (8) tells us that $R(\theta, \delta^*)$ is, as θ varies on $(0, m)$, either: (i) convex, (ii) concave then convex, or (iii) concave. Now, given that $m \leqslant m_0$ by assumption, it must be the case that (7) holds. This renders (iii) impossible given that $R(0, \delta^*) = R(m, \delta^*)$. Otherwise, both (i) and (ii) do not allow for the risk function $R(\theta, \delta^*)$ to be maximized at an interior point $\theta_0 \in (0, m)$. Therefore, we infer that (3) must hold, yielding the result. □

Before pursuing, we define $z(\theta)=\theta[\partial^2/\partial\theta^2]G(n,\theta)$ and $m_1=\inf\{\theta\in\Theta: z'(\theta)<0\}$. Observe that for $p_\theta\in C$, we must have $m_1>0$.

Theorem 1. *For squared error loss and* $p_\theta\in C$,

(a) $[\partial^2/\partial\theta^2]R(\theta,\delta^*)$ *is nondecreasing for* $\theta\in[0,m_1]$.
(b) $\delta^*(X)$ *is minimax whenever* $m\leqslant m_0\wedge m_1$.
(c) *Whenever* $m_0\leqslant m_1$; *or equivalently* $T(m_1)\geqslant 0$; $\delta^*(X)$ *is minimax iff* $m\leqslant m_0$.

Proof. Parts (b) and (c) follows from part (a), as well as Lemmas 2, 4 and 5. Hence, there remains only to prove (a). From (10), we obtain

$$\frac{\partial^2}{\partial\theta^2}R(\theta,\delta^*)=2+4(m-y^*)\frac{\partial}{\partial\theta}G(n,\theta)+(m^2-(y^*)^2)\left(-\frac{\partial^2}{\partial\theta^2}G(n,\theta)\right)+2(m-y^*)z(\theta).$$

The result follows since $p_\theta\in C$, $y^*\leqslant m$, and $z(\theta)$ is nondecreasing by definition for $\theta\leqslant m_1$. □

We now turn to various illustrations and implications of Theorem 1. Given the generality of Theorem 1, remaining interest lies in the specification (and properties) of m_0 and m_1 for various members $p_\theta\in C$. We do this below for the families p_θ given in Example 1. Also, observe that whenever Theorem 1's $\delta^*(X)$ is minimax, the minimax risk is given by $R(0,\delta^*(X))=y^{*2}$, with y^* given in (5). In what follows, we will sometimes denote m_0 and m_1 as $m_0(n)$ and $m_1(n)$ to emphasize the dependence on the sample size n.

Example 2 (*Poisson*(θ) *or Generalized Poisson*(θ,λ)). Here $G(n,\theta)=e^{-n\theta}, z(\theta)=n^2\theta e^{-n\theta}$, and $z'(\theta)=n^2e^{-n\theta}(1-n\theta)$, which gives $m_1(n)=1/n$. From (9), we obtain that $m_0(n)=c_0/n$ with $c_0=2(e^{-c_0/2}(1+e^{-c_0/2}))/(1+2e^{-c_0/2})$. With the numerical evaluation $c_0\approx 0.912955$, part (c) of Theorem 1 tells us that $\delta^*(X)$, given in (4), is minimax iff $m\leqslant m_0(n)\approx(0.912955)/n$. Interestingly, the above minimaxity result still holds when λ is unknown. This is so, because otherwise with both $P_{(\theta,\lambda)}(X\in A)$ and the risk $R((\theta,\lambda),\delta^*)$ being independent of λ, a contradiction is arrived at.

Example 3 (*Binomial; Marchand and MacGibbon, 2000*[1]). Here $G(n,\theta)=(1-\theta)^n$, $z'(\theta)=n(n-1)(1-\theta)^{n-3}(1-\theta(n-1))$, $m_1(1)=1$, and $m_1(n)=1/(n-1)$ for $n\geqslant 2$. From (9), we may show that $T(m_1(n))>0$. Therefore, part (c) of Theorem 1 applies, and implies that the estimator $\delta^*(X)$, as given in (4) is minimax iff $m\leqslant m_0(n)$, with $m_0(n)=T^{-1}(0)$. Marchand and MacGibbon (2000) give a graph for $m_0(n)$, and show that $\lim_{n\to\infty} nm_0(n)\approx 0.912955\ldots$.

Remark 3. The above large sample approximation for a Binomial $m_0(n)$ (which follows from Corollary 1 below with $\gamma(\theta)=\theta/(1-\theta)$ and $a_1=1$) also can be deduced from Example 2 given the limiting Poisson distribution for a sequence of $\{Bi(n,c/n);n\geqslant 1\}$ random variables. With the help of the following Poisson approximation result, the above may be generalized as follows for power series families given in (2) which are members of C.

Lemma 6 (*Pérez-Abreu, 1991*). *For* $X_1,\ldots,X_n$ *distributed as in* (2) *with parameter* θ_n, *such that* $n\gamma(\theta_n)\to\lambda$, *as* $n\to\infty$, *it follows that* $S_n=\sum_{i=1}^n X_i$ *converges to a Poisson*(λa_1) *distribution. In particular*, $G(n,\theta_n)\to e^{-\lambda a_1}$ *as* $n\to\infty$.

Corollary 1. *For power series families as in* (2) *which are members of* C, *we have*

$$nm_1(n)\to a_1,\quad nm_0(n)\to c_0a_1\approx 0.912955a_1,\quad \text{as } n\to\infty.$$

Moreover, for sufficiently large n, the estimator $\delta^*(X)$, *given in* (4) *is minimax iff* $m\leqslant m_0(n)\approx(0.912955a_1)/n$.

Example 4 (*NBI*(α,p) *and Waring*(k,ρ)). Here $G(n,\theta)=(\alpha/(\alpha+\theta))^{n\alpha}$ (as in Example 1 c,e), $z(\theta)\propto\theta/((\theta+\alpha)^{n\alpha+2})$, $z'(\theta)\propto(1/(\theta+\alpha)^{n\alpha+3})(\alpha-\theta(n\alpha+1))$, which gives $m_1(n)=\alpha/(1+n\alpha)$. Now,

[1] Note that their Theorem 2.1 is correct but that some of the derivatives leading up to it are in error.

we obtain from (9),

$$T(m) = m - \frac{2}{n} \frac{\left(\frac{\alpha}{\alpha+m}\right)^{n\alpha/2}\left(1+\left(\frac{\alpha}{\alpha+m}\right)^{n\alpha/2}\right)}{1+2\left(\frac{\alpha}{\alpha+m}\right)^{n\alpha/2}} \tag{13}$$

and

$$T(m_1(n)) = \frac{1}{n}\left\{\frac{n\alpha}{1+n\alpha} - 2v\left[\left(\frac{1+n\alpha}{2+n\alpha}\right)^{n\alpha/2}\right]\right\}$$

with $v(y) = (y(1+y))/(1+2y)$. Now, observe that $((1+s)/(2+s))^{s/2}$ is decreasing in s; $s\geqslant 0$; so that

$$m_0(n)\geqslant m_1(n) \iff T(m_1(n))\geqslant 0 \iff n\alpha\geqslant s_0,$$

where $s_0/(1+s_0) = 2v(((1+s_0)/(2+s_0))^{s_0/2})$. With the numerical evaluation $s_0 \approx 11.876904$ we have the following summaries for both the Negative Binomial and Waring cases.

Corollary 2. *For* $X_1,\ldots,X_n$ *independent* $NBI(\alpha,p)$ (*as in Example* 1c) *with* $\theta = E(X_i)$; $\theta\leqslant m$; *the estimator* $\delta^*(X)$ *given by* (4) *is minimax*:

(a) *Whenever* $n\alpha < s_0 \approx 11.876904$, *and* $m\leqslant \alpha/(1+n\alpha)$.
(b) *iff* $m\leqslant m_0(n) = T^{-1}(0)$ *whenever* $n\alpha\geqslant s_0 \approx 11.876904$.

Corollary 3. *For* $X_1,\ldots,X_n$ *independent Waring*(k,ρ) *with* $\theta = k/\rho$ (*as in Example* 1e); $\theta\leqslant m$; *the estimator* $\delta^*(X)$ *given by* (4) *is minimax*:

(a) *For* $m\leqslant 1/(1+n)$ *whenever* $n\leqslant 11$ (*i.e.*, $n\leqslant s_0$).
(b) *iff* $m\leqslant m_0(n) = T^{-1}(0)$ *whenever* $n\geqslant 12$, *with* $T(m)$ *given in* (13) *with* $\alpha = 1$.

We consider next a Poisson hierarchical model given in part (b) of Lemma 1 with $G_0(n,\lambda) = \mathrm{e}^{-n\lambda}$. As witnessed in the Poisson case above where both $nm_0(n)$ and $nm_1(n)$ are constants independent of n, we show in Lemma 7 that this property holds in general for such Poisson mixtures. Consequently, the comparison of $m_0(n)$ and $m_1(n)$, as needed for applications of Theorem 1, does not depend on n and only requires a handling of the case $n = 1$.

Lemma 7. *For Poisson mixture cases as in* (1) *with* $G_0(n,\lambda) = \mathrm{e}^{-n\lambda}$, *and* $E_1[\lambda^k] < \infty$ *for* $k\geqslant 1$, *we have* $m_0(n) = c_0/n$ *and* $m_1(n) = c_1/n$, *where* c_0 *and* c_1 *are independent of* n.

Proof. Here $G(n,\theta) = \int_0^\infty \mathrm{e}^{-n\theta\lambda} f_1(\lambda)\,\mathrm{d}\lambda$, and an easy calculation shows that, for a constant $c>0$,

$$z'\left(\frac{c}{n}\right) = n^2 \int_0^\infty \lambda^2 \mathrm{e}^{-c\lambda}(1-c\lambda) f_1(\lambda)\,\mathrm{d}\lambda,$$

which implies that $m_1(n)$ (i.e., the solution of $z'(\theta) = 0$) is of the form c_1/n with c_1 independent of n. Similarly, it is easy to verify by expanding (9) that, for $c>0$, $nT(c/n)$ is free of n, implying that Lemma 4's $m_0(n)$ is of the form c_0/n with c_0 independent of n. $\square$

Example 5 (*The Poisson mixture in Remark* 1, *part* (i)). Referring to Remark 1(i) with $n = 1$, we have $G(1,\theta) = (1/(1+\theta))^{\varepsilon}$. Since this matches Example 4's $G(n,\theta)$ with $\alpha = 1$ and $n = \varepsilon$, we can borrow from that analysis. Along with Lemma 7, we obtain the following.

Corollary 4. *For the model*: $X_1,\ldots,X_n|\lambda$ *independent Poisson*(λ), *with* $\lambda|\theta\sim Gamma(\varepsilon,\theta)$ (*as in Remark* 1, *part* (i)); *with* $\theta\leqslant m$; *the estimator* $\delta^*(X)$ *given by* (4) *is minimax*:

(a) *Whenever* $\varepsilon < s_0 \approx 11.876904$, *and* $m\leqslant \frac{1}{n}\,\alpha/(1+\alpha)$.

(b) *iff* $m \leqslant 1/nm_0(1) = 1/nT^{-1}(0)$ *whenever* $\varepsilon \geqslant s_0 \approx 11.876904$, *where* T^{-1} *is the inverse of the* $T(m)$ *function given in* (13) *with* $\alpha = 1$ *and* $n = \varepsilon$.

Acknowledgements

The authors benefited from discussions with Yiping Dou and Rokhaya Gueye. The authors are thankful to a referee for constructive comments which improved the presentation of our results.

References

Bader, G., Bischoff, W., 2003. Old and new aspects of minimax estimation of a bounded parameter. Mathematical Statistics and Applications, Festschrift for Constance van Eeden. IMS Lecture Notes-Monograph Series, vol. 42, pp. 15–30.

Brown, L.D., 1994. Minimaxity, more or less. In: Gupta, S.S., Berger, J.O. (Eds.), Statistical Decision Theory and Related Topics, vol. 5. Springer, New York, pp. 1–18.

Consul, P.C., 1989. Generalized Poisson Distributions: Properties and Applications. Marcel Dekker, New York.

DasGupta, A., 1985. Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. Sankhyā: Ser. A 47, 326–332.

van Eeden, C., 1996. Estimation in restricted parameter spaces—some history and recent developments. CWI Quart. 9, 69–76.

Feller, W., 1966. An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York.

Johnson, N.L., Kotz, S., Kemp, A.W., 1993. Univariate Discrete Distributions. Wiley, New York.

Johnstone, I., MacGibbon, B., 1992. Minimax estimation of a constrained Poisson vector. Ann. Statist. 20, 807–831.

Kempthorne, P.J., 1987. Numerical specification of discrete least favorable distributions. Siam J. Sci. Comput. 8, 171–184.

Marchand, É., MacGibbon, B., 2000. Minimax estimation of a constrained binomial proportion. Statist. Decisions 18, 129–167.

Marchand, É., Parsian, A., 2004. Minimax estimation of a bounded discrete parameter, Rapport de recherche 2004–2007. Département de mathématiques, Université de Sherbrooke, www.usherbrooke.ca/mathematiques/telechargement/

Marchand, É., Strawderman, W.E., 2004. Estimation in restricted parameter spaces: a review. Festschrift for Herman Rubin, IMS Lecture Notes-Monograph Series, vol. 45, pp. 1–24.

Pérez-Abreu, V., 1991. Poisson approximation to power series distributions. Am. Statist. 45, 42–45.

Strawderman, W.E., 2000. Minimaxity. J. Am. Statist. Assoc. 95, 1364–1368.

Wan, T., Zou, G., Lee, A., 2000. Minimax and Γ-minimax estimation for a Poisson distribution under linex loss when the parameter space is restricted. Statist. Probab. Lett. 50, 23–32.