

# Minimax Estimation of a Bounded Discrete Parameter

ÉRIC MARCHAND<sup>a,1</sup>, AHMAD PARSIAN<sup>b, 2</sup>,

*a University of New Brunswick and Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)*

*b School of Mathematical Sciences, Isfahan University of Technology, Isfahan, 84156, IRAN (e-mail: ahmad\_p@cc.iut.ac.ir)*

## ABSTRACT

For a vast class of discrete model families with cdf's  $F_\theta$ , and for estimating  $\theta$  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  $m \leq m(F)$ , 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.

**Key words and phrases:** Minimax estimation, restricted parameter space, discrete distributions, squared error loss

---

<sup>1</sup>research supported by NSERC of Canada and by the University of New Brunswick Academic Fund

<sup>2</sup>research supported by a grant of the Research Council of the Isfahan University of Technology

# 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  $\theta$  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 (1986) 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 small enough  $m$ , say  $m \leq m^*$  with  $m^* > 0$ . Since the phenomenon also extends in various univariate and multivariate settings (DasGupta, 1986), 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  $X = (X_1, \dots, X_n); n \geq 1$ , 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  $\theta$ , under squared error loss, in situations where  $\theta$  is constrained to a known, and small enough interval  $[a, b]$ , and where the distribution of  $X$  under  $\theta = a$  is degenerate at  $(\underline{s}, \dots, \underline{s})$ . Without loss of generality, given that we can reparametrize and translate, we assume hereafter that  $[a, b] = [0, m]$  and  $\underline{s} = 0$ . 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  $A = \{x \in \mathfrak{R}^n : \sum_{i=1}^n x_i = 0\}$ , 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  $\theta \in [0, m]$ . We now describe the class  $C$  of families of distributions for which the main result of this paper, Theorem ??, applies.

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

$$(-1)^k \frac{\partial^k}{\partial \theta^k} G(n, \theta) \geq 0, \text{ on } [0, m], \text{ for } k = 1, 2, 3.$$

Here are some immediate examples of families  $p_\theta$  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, Kotz and Kemp (1993).

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

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

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

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

(e)  $X_i$ 's are independently distributed  $\text{Waring}(k, \rho)$ , with  $P_\theta(X_i = x) = \rho \frac{\binom{k}{x}}{(k+\rho)_{x+1}} I_{\{0,1,\dots\}}(x)$ ,  $k > 0, \rho > 0$  and  $\theta = k/\rho$ . Whenever (i)  $k$  is known and  $\rho$  is unknown, (which includes the Yule distribution for  $k = 1$ ), or (ii)  $k$  is unknown and  $\rho$  is known, the families of  $p_\theta$ 's belong to  $C$  with  $G(n, \theta) = (\frac{1}{1+\theta})^n$ . Observe that this expression matches the  $\text{Negative Binomial } G(n, \theta)$  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  $X_1, \dots, X_n$  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  $E_1[\lambda^k] = \int_0^\infty \lambda^k f_1(\lambda) d\lambda < \infty$ ;  $k = 1, 2, \dots$ ; <sup>3</sup> then*

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

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

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

*with  $E_1[\lambda^k] = \int_0^\infty \lambda^k f_1(\lambda) d\lambda < \infty$ ;  $k = 1, 2, \dots$ ; 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, page 417, Feller 1966). Observe for part (b) that the given representation implies:

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

from which the result follows by the assumed complete 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 ???. Indeed, denoting  $T = \sum_{i=1}^n X_i$  and referring to Example 1(c), we have as illustrations of Lemma ???'s part (b) and part (c) respectively:*

(i)  $T \sim NBI(\epsilon, p = (1 + n\theta)^{-1})$ , (with  $G(n, \theta) = (1 + n\theta)^{-\epsilon}$ ), whenever  $X_1, \dots, X_n|\lambda$  are independent  $Poisson(\lambda)$ , with  $\lambda|\theta \sim Gamma(\epsilon, \theta)$ ;

(ii)  $T \sim NBI(n\epsilon, p = (1 + \theta)^{-1})$ , (with  $G(n, \theta) = (\frac{\epsilon}{\epsilon + \theta})^{n\epsilon}$ ), whenever  $X_1, \dots, X_n$  are independent with  $X_i|\lambda \sim Poisson(\lambda)$ , and  $\lambda|\theta \sim Gamma(\epsilon, \theta)$ .

---

<sup>3</sup>Weaker conditions can be given. Moreover, for the specification of members of  $C$ , and for applying the results below, we only need to assume  $E_1[\lambda^3] < \infty$ .

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 ??, can be found among the family of power series distributions, where  $X_1, \dots, X_n$  are independent with a common probability function given by:

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

with  $a_0 = 1$  (without loss of generality),  $\gamma(\theta)$  being a positive and nondecreasing function of  $\theta$ , and  $c(\gamma) = \sum_{x \geq 0} a_x \gamma^x$ . Here  $G(n, \theta) = (\frac{1}{c(\gamma(\theta))})^n$ . For such families to belong to  $C$ , it will suffice, as just observed with the previous lemma, that  $\frac{1}{c(\gamma(\theta))}$  be completely monotone in  $\theta$ . We note that, although a sufficiency reduction is available here with the statistic  $\sum_{i=1}^n X_i$ , and its family of distributions being power series (as in (??) with the coefficients  $a_x$  being the coefficients in the power series representation of  $(c(\gamma(\theta)))^n$ , we will not (and it is not necessary for our purposes) make use of it, which will be make it somewhat easier to highlight the role of the sample size  $n$  in some of our results below. We conclude this section with two more interesting examples of classes of families of distributions which belong to  $C$ , and which greatly increase the number of applications of Theorem ?? below.

**Example 2** (*Generalized Hypergeometric Factorial Moment Distributions*)

Described by Johnson, Kotz, and Kemp (1993, section 4.2), these distributions (which include, although with a possibly different parametrization, the distributions in Example 1, parts (a), (b), and (c) have probability generating functions of the form  $E_\theta[t^{X_1}] = {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; \theta(t-1))$ , where  ${}_pF_q$  is a Generalized Hypergeometric function. Therefore, for  $X_1, \dots, X_n$  independent and identically distributed arising from one such family, we have:

$$G(n, \theta) = \{P_\theta(X_1 = 0)\}^n = \{{}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; -\theta)\}^n$$

(note that, for our general i.i.d. case,  $G(n, \theta) = (E_\theta[t^{X_1}]|_{t=0})^n$ ). Now, for cases where both the  $a_i$ 's and  $b_i$ 's are positive, it is easy to see from the above representation and part (a) of Lemma ?? that  $G(n, \theta)$  is completely monotone and, consequently, the corresponding family of  $p_\theta$ 's belong to  $C$ .

**Example 3** (*Stopped-Sum Distributions*) Closely related to mixture models, such as those in Lemma ??, are Stopped-Sum distributions which admit the representation:  $S = \sum_{i=1}^N X_i$ , with  $N, X_1, X_2, \dots$  being independent random variables such that the  $X_i$ 's are identically distributed with  $P_\theta(X_i = 0) = G_0(1, \theta)$  (for some  $G_0$ ); the distribution of  $N$  is known with  $P(N \in \{1, 2, \dots\}) = 1$ ,  $E[N^k] < \infty$  for  $k = 1, 2, \dots$ ; and  $P(S < \infty) = 1$ . Now, suppose  $G_0(1, \theta)$  is completely monotone in  $\theta$ . Then, it is easy to see that  $G_S(1, \theta) = P_\theta(S = 0)$  is also completely monotone in  $\theta$  given the representation  $P_\theta(S = 0) = E[(G_0(1, \theta))^N]$ , and part (a) of Lemma ?? . Hence, such a family of probability functions  $P_\theta(S = s); \theta \in [0, m]$ ; belongs to the class  $C$ . The function  $G_0$  is general, with Poisson and Bernoulli  $G_0$ 's being perhaps of particular importance.

## 2.2 Minimavity

Our results rely on a following well-known criteria for minimavity applied to boundary two-point priors.

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

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

Now, for our discrete models, Bayes estimators  $\delta_\pi(X)$  corresponding to two-point boundary supported priors  $\pi$  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  $\pi$  exactly one of them leads to an equalizer Bayes rule under squared error loss, and it is given by*

$$\delta^*(x) = y^*[x \in A] + m[x \notin A], \quad (4)$$

with

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

**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  $\pi(0) = 1 - \pi(m)$ :

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

and

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

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

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

with  $y$  minimizing (??) 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  $y$ .

**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  $\rho$ . This was observed for a Linex loss  $\rho$  by Wan, Zou, and Lee (2000, Theorem 2.1).*

### 3 Main Results

The general development that follows with Lemma ?? and Theorem ?? capitalizes on:

(i) the necessary condition:

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

for (??) to hold with  $\delta(X) = \delta^*(X)$ ; and

(ii) the sufficient condition:

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

for (??) to hold with  $\delta(X) = \delta^*(X)$  in cases where (??) is satisfied.

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

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

**Proof.** Given representation (??) of  $\delta^*(X)$ , we obtain directly

$$\begin{aligned} R(\theta, \delta^*) &= (m - \theta)^2 + G(n, \theta)[(y^* - \theta)^2 - (m - \theta)^2] \\ \frac{\partial}{\partial \theta} R(\theta, \delta^*) &= 2(\theta - m) + 2G(n, \theta)(m - y^*) + \frac{\partial}{\partial \theta} G(n, \theta)\{(y^* - \theta)^2 - (m - \theta)^2\} \end{aligned} \quad (10)$$

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

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

$$\begin{aligned} m &\leq \left( \frac{-2}{\frac{\partial}{\partial \theta} G(n, \theta)|_{\theta=0}} \right) \frac{\frac{y^*}{m}}{1 - \left(\frac{y^*}{m}\right)^2} \\ \iff m &\leq \left( \frac{-2}{\frac{\partial}{\partial \theta} G(n, \theta)|_{\theta=0}} \right) \frac{(\sqrt{G(n, m)})(1 + \sqrt{G(n, m)})}{1 + 2\sqrt{G(n, m)}}. \end{aligned} \quad (12)$$

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

Before pursuing, we define  $z(\theta) = \theta \frac{\partial^2}{\partial \theta^2} G(n, \theta)$  and  $m_1 = \inf\{\theta : z'(\theta) < 0\}$ . Here, the infimum is taken over the set of all  $\theta$  such that  $p_\theta$  is a valid probability measure. Also, 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)  $\frac{\partial^2}{\partial \theta^2} R(\theta, \delta^*)$  is nondecreasing for  $\theta \in [0, m_1]$ ;

(b)  $\delta^*(X)$  is minimax whenever  $m \leq m_0 \wedge m_1$ ;

(c) Whenever  $m_0 \leq m_1$ ; or equivalently  $T(m_1) \geq 0$ ;  $\delta^*(X)$  is minimax iff  $m \leq m_0$ .

**Proof.** Part (b) follows part (a), criterion (??), and Lemma ?? while part (c) follows from part (b) and Lemma ?. Hence, there remains only to prove (a). From (??), 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  $p_\theta \in C$ ,  $y^* \leq m$ , and  $z(\theta)$  is nondecreasing by definition for  $\theta \leq m_1$ .

We now turn to various illustrations and implications of Theorem ?. Given the generality of Theorem ?, 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  $p_\theta$  given in Example ?. Also, observe that whenever Theorem ?'s  $\delta^*(X)$  is minimax, the minimax risk is given by  $R(0, \delta^*(X)) = y^{*2}$ , with  $y^*$  given in (??). 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 4** (*Poisson*( $\theta$ ) or *Generalized Poisson*( $\theta, \lambda$ )) Here  $G(n, \theta) = e^{-n\theta}$ ,  $z(\theta) = n^2\theta e^{-n\theta}$ , and  $z'(\theta) = n^2 e^{-n\theta}(1 - n\theta)$ , which gives  $m_1(n) = \frac{1}{n}$ . From (??), we obtain that  $m_0(n) = \frac{c_0}{n}$  with  $c_0 = 2 \frac{e^{-c_0/2}(1+e^{-c_0/2})}{1+2e^{-c_0/2}}$ . With the numerical evaluation  $c_0 \approx 0.912955$ , part (c) of Theorem ?? tells us that  $\delta^*(X)$ , given in (??), is minimax iff  $m \leq m_0(n) \approx \frac{0.912955}{n}$ . Interestingly, the above minimaxity result still holds when  $\lambda$  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  $\lambda$ , we arrive at a contradiction.

**Example 5** (*Binomial*; Marchand and MacGibbon, 2000<sup>4</sup>) 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  $m_1(n) = \frac{1}{n-1}$  for  $n \geq 2$ . From (??), we obtain

$$T(m) = m - \frac{2(1 - m)^{n/2}(1 + (1 - m)^{n/2})}{n(1 + 2(1 - m)^{n/2})},$$

and, for  $n \geq 2$ ,

$$T(m_1(n)) = T\left(\frac{1}{n-1}\right) \geq T\left(\frac{1}{n}\right) = \frac{1 - 2(1 - \frac{1}{n})^n}{n(1 + 2(1 - \frac{1}{n})^{n/2})} > 0.$$

Here again, part (c) of Theorem ?? applies, and implies that the estimator  $\delta^*(X)$ , as given in (??) is minimax iff  $m \leq m_0(n)$ , with  $m_0(n) = T^{-1}(0)$ . Marchand and MacGibbon (2000) give a graph for  $m_0(n)$ , and show that  $\lim_{n \rightarrow \infty} nm_0(n) = 0.912955\dots$  This last result can equally be deduced by the Poisson example above given the limiting Poisson distribution for a sequence of  $\{Bi(n, c/n); n \geq 1\}$  random variables.

**Remark 3** The above limiting result and Binomial large sample approximation for  $m_0(n)$  (which follows from Corollary 1 below with  $\gamma(\theta) = \frac{\theta}{1-\theta}$  and  $a_1 = 1$ ) can be generalized as follows for power series families given in (??) which are members of  $C$ . This is done in Corollary 1 with the help of Example ?? the following Poisson approximation result.

<sup>4</sup>Note that their Theorem 2.1 is correct but that some of the derivatives leading up to it are in error.

**Lemma 5** (Pérez-Abreu, 1991)

For  $X_1, \dots, X_n$  distributed as in (??) with parameter  $\theta_n$ , such that  $n\gamma(\theta_n) \rightarrow \lambda$ , as  $n \rightarrow \infty$ , it follows that  $S_n = \sum_{i=1}^n X_i$  converges to a Poisson( $\lambda a_1$ ) distribution. In particular,  $G(n, \theta_n) \rightarrow e^{-\lambda a_1}$  as  $n \rightarrow \infty$ .

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

$$nm_1(n) \rightarrow a_1, nm_0(n) \rightarrow c_0 a_1 \approx 0.912955 a_1, \text{ as } n \rightarrow \infty.$$

Moreover, for sufficiently large  $n$ , the estimator  $\delta^*(X)$ , given in (??) is minimax iff  $m \leq m_0(n) \approx \frac{0.912955 a_1}{n}$ .

**Example 6** (Negative Binomial) Here  $G(n, \theta) = \left(\frac{\alpha}{\alpha+\theta}\right)^{n\alpha}$  (as in Example 1 c),  $z(\theta) \propto \frac{\theta}{(\theta+\alpha)^{n\alpha+2}}$ ,  $z'(\theta) \propto \frac{1}{(\theta+\alpha)^{n\alpha+3}}(\alpha - \theta(n\alpha + 1))$ , which gives  $m_1(n) = \frac{\alpha}{1+n\alpha}$ . Now, we obtain from (??),

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

and

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

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

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

where  $\frac{s_0}{1+s_0} = 2v\left(\left(\frac{1+s_0}{2+s_0}\right)^{s_0/2}\right)$ . With the numerical evaluation  $s_0 \approx 11.876904$  we have the following summary for the Negative Binomial case.

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

(a) whenever  $n\alpha < s_0 \approx 11.876904$ , and  $m \leq \frac{\alpha}{1+n\alpha}$ ;

(b) iff  $m \leq m_0(n) = T^{-1}(0)$  whenever  $n\alpha \geq s_0 \approx 11.876904$ .

**Example 7** (Waring( $k, \rho$ )) In terms of the pair  $(k, \rho)$ , Theorem ?? and Corollary ?? below apply for estimating  $k$  with  $\rho$  known, or estimating  $\frac{1}{\rho}$  with  $k$  known, with  $G(n, \theta) = \left(\frac{1}{1+\theta}\right)^n$  as in Example 1(e). Matching Example ??'s  $G(n, \theta)$  for  $\alpha = 1$ , we immediately have the following summary.

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

(a) for  $m \leq \frac{1}{1+n}$  whenever  $n \leq 11$  (i.e.,  $n \leq s_0$ );

(b) iff  $m \leq m_0(n) = T^{-1}(0)$  whenever  $n \geq 12$ , with  $T(m)$  given in (??) with  $\alpha = 1$ .

We consider next the hierarchical model given in part (b) of Lemma ?? for the particular Poisson case where  $G_0(n, \lambda) = e^{-n\lambda}$ . As witnessed in the Poisson case above where both  $nm_0(n)$  and  $nm_1(n)$  were constants independent of  $n$ , we show in Lemma ?? that this property holds in general for such Poisson mixtures. Consequently, the comparison of  $m_0(n)$  and  $m_1(n)$ , as required for applications of Theorem ??, does not depend on  $n$ . It will then suffice in applications such as the one given in Example ?? to handle the case  $n = 1$ .

**Lemma 6** *For Poisson mixture cases as in (??) with  $G_0(n, \lambda) = e^{-n\lambda}$  and  $E_1[\lambda^3] < \infty$ , we have  $m_0(n) = \frac{c_0}{n}$  and  $m_1(n) = \frac{c_1}{n}$ , where  $c_0$  and  $c_1$  are independent of  $n$ .*

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

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

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

**Example 8** *(The Poisson mixture in Remark ??, part (i)) Referring to Remark ?? (i) with  $n = 1$ , we have  $G(1, \theta) = (\frac{1}{1+\theta})^\epsilon$ . Since this matches Example ??'s  $G(n, \theta)$  with  $\alpha = 1$  and  $n = \epsilon$ , we can borrow from that analysis (or even Corollary 2). Along with Lemma ??, we obtain the following.*

**Corollary 4** *For the model:  $X_1, \dots, X_n | \lambda$  i.i.d.  $\text{Poisson}(\lambda)$ , with  $\lambda | \theta \sim \text{Gamma}(\epsilon, \theta)$  (as in Remark ??, part (i)); with  $\theta \leq m$ ; the estimator  $\delta^*(X)$  given by (??) is minimax:*

- (a) whenever  $\epsilon < s_0 \approx 11.876904$ , and  $m \leq \frac{1}{n} \frac{\alpha}{1+\alpha}$ ;
- (b) iff  $m \leq \frac{1}{n} m_0(1) = \frac{1}{n} T^{-1}(0)$  whenever  $\epsilon \geq s_0 \approx 11.876904$ , where  $T^{-1}$  is the inverse of the  $T(m)$  function given in (??) with  $\alpha = 1$  and  $n = \epsilon$ .

## ACKNOWLEDGEMENTS

The authors benefited from discussions with Yiping Dou and Rokhaya Gueye.

## REFERENCES

- Bader, G. and 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, **42**, pp. 15-30.
- Brown, L.D. (1994). Minimacity, more or less, in *Statistical Decision Theory and Related Topics, V*, editors S.S. Gupta and J.O Berger. Springer, New York, pp. 1-18.

- Consul, P.C. (1989). *Generalized Poisson distributions: Properties and Applications*, Marcel Dekker.
- DasGupta, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set, *Sankhyā: Series A*, **47**, 326-332.
- Feller, W. (1966). *An Introduction to Probability Theory and Its Applications: Volume II*, John Wiley & Sons.
- Johnson, N.L., Kotz, S. and Kemp, A.W. (1993). *Univariate Discrete Distributions*, Wiley, New York.
- Johnstone, I. and MacGibbon, B. (1992). Minimax estimation of a constrained Poisson vector, *Annals of Statistics*, **20**, 807-831.
- Marchand, É. and MacGibbon, B. (2000). Minimax estimation of a constrained binomial proportion, *Statistics and Decisions*, **18**, 129-167.
- Marchand, É. and Strawderman, W. E. (2004). Estimation in restricted parameter spaces: A review, *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, pp. 1-24.
- Pérez-Abreu, V. (1991). Poisson approximation to power series distributions, *The American Statistician*, **45**, 42-45.
- Strawderman, W. E. (2000). Minimaxity, *Journal of the American Statistical Association*, **95**, 1364-1368.
- van Eeden, C. (1996). Estimation in restricted parameter spaces- some history and recent developments, *CWI Quarterly*, **9**, 69-76.
- Wan, T., Zou, G., and Lee, A. (2000). Minimax and  $\Gamma$ -minimax estimation for a Poisson distribution under linex loss when the parameter space is restricted, *Statistics & Probability Letters*, **50**, 23-32.