Minimax estimation of a restricted mean for a one-parameter exponential family [1]

Éric Marchand[a], Fanny Rancourt[b] & William E. Strawderman[c]
a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)
b Commission scolaire de Montréal, 3737 rue Sherbrooke E, Montréal, Qc, CANADA, H1X 3B3 (e-mail: rancourt.f@csdm.qc.ca)
c Rutgers University, Department of Statistics, 501 Hill Center, Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

Summary

For one-parameter exponential families, we provide a unified minimax result for estimating the mean under weighted squared error losses in presence of a lower-bound restriction. The finding recovers cases for which the result is known, as well as others which are new such as for a negative binomial model. We also study a related *self-minimaxity* property, obtaining several non-minimax results. Finally, for discrete models such as Poisson and negative binomial, we obtain classes of minimax estimators of a lower-bound restriction on the mean, which include range preserving solutions.

AMS 2010 subject classifications: 62C15, 62C20, 62F10, 62F30

Keywords and phrases: Admissibility; Discrete distributions; Exponential family; Karlin's theorem; Minimax; Negative binomial; Poisson; Restricted parameters.

1. Introduction

Restricted parameter space inference bring many challenges, several of which are less apparent in corresponding unrestricted parameter space versions. Over the years, a better understanding of decision-theoretic properties of such problems has emerged and several techniques and methods have been developed to address such challenges (e.g., Marchand & Strawderman, 2004; van Eeden, 2006), including minimax analysis (e.g., Casella & Strawderman, 1981; Marchand & Strawderman, 2012).

Interestingly, there exist a large number of problems where the minimax risk for $X \sim p_\theta, \theta \in \Theta$, remains the same for the parametric restriction $\theta \in \Theta_0 \subset \Theta$, and this for many choices of loss functions (e.g., Marchand & Strawderman, 2012). An early example was addressed by Katz (1961) for $X \sim N(\theta, 1)$, squared error loss, $\Theta = \mathbb{R}$, and $\Theta_0 = \mathbb{R}_+$. In this case, the estimator $\delta_0(X) = X$ is clearly not adapted to the parametric restriction, but knowledge of its minimaxity nevertheless is still a useful benchmark as dominating estimators are necessarily minimax themselves. Then, in such situations, it is further of interest to specify classes of plausible estimators, Bayesian or otherwise, that dominate the given benchmark and that are necessarily minimax themselves.

[1] September 30, 2020

This paper deals with one-parameter exponential families and a corresponding minimax property for normalized squared error loss. We find it useful to record a unifying minimax finding, which not only includes existing results such as for the normal case among others, but also leads to new findings such as for a lower-bounded negative binomial mean. We also focus on determining dominating minimax estimators, in particular for discrete models as there exists a relative paucity of dominance findings in the literature.

Let

$$X \sim p_\eta(x) \;=\; e^{\eta x \,-\, A(\eta)}\,, \tag{1.1}$$

with $\eta \in (\underline{\eta}, \bar{\eta})$, be a one-parameter exponential family density with respect to a $\sigma-$finite measure ν. Set $\theta = \mathbb{E}(X) = A'(\eta)$, $\mathbb{V}(\theta) = \mathrm{Var}_\theta(X) = A''(\eta)$, assume $\theta \to \bar{\theta}$ as $\eta \to \bar{\eta}$ and $\theta \to \underline{\theta}$ as $\eta \to \underline{\eta}$. The minimax and non-minimax results of this paper depend critically on the variation of $\lambda(\theta) = \frac{\mathbb{V}(\theta)}{\theta^2}$.

We study the potential minimaxity of estimators $\delta_{\lambda,\gamma}(X) = \frac{X}{1+\lambda} + \frac{\gamma\lambda}{1+\lambda}$ of θ under weighted squared error loss

$$L_{\lambda,\gamma}(\theta,\delta) \;=\; \frac{(\delta-\theta)^2}{w(\theta)}\;,\ \text{with}\ \ w(\theta) \;=\; \mathbb{E}\left\{\delta_{\lambda,\gamma}(X)-\theta\right\}^2\ \text{and}\ \lambda \geq 0. \tag{1.2}$$

Karlin's Theorem (Karlin, 1958) concerns exponential families (1.1) and provides sufficient conditions (see Section 2.2) for the admissibility of $\delta_{\lambda,\gamma}(X)$ under squared error loss, and hence under loss (1.2). Moreover, in cases where admissibility holds, it follows immediately that $\delta_{\lambda,\gamma}(X)$ is unique minimax under loss (1.2) with constant risk equal to 1.

When $\bar{\theta} = \infty$, taking $\lambda = \lim_{\theta\to\infty} \lambda(\theta)$, we show with Theorem 2.2 that the estimator $\delta_{\lambda,0}(X) = \frac{X}{1+\lambda}$ is also minimax under a lower bound restriction $\theta \geq \theta_0$, under loss $L_{\lambda,0}$ provided Karlin's conditions for admissibility in the full parameter space hold. The finding is unified and we provide examples, some of which being known minimax results, and some which being new to the best of our knowledge. Also, the result is significant as dominating estimators of $\delta_0(X)$ will necessarily be minimax. Extensions to a much large class of loss functions (Corollary 2.1 are then obtained. An analogue minimax result (Theorem 2.3) is also established for an upper-bound restriction and implications for the potential minimaxity of estimators $\delta_{\lambda,\gamma}(X)$ under loss $L_{\lambda,\gamma}$ is also addressed with Corollary 2.2.

The angle examined here can be described as one of *self-minimaxity* as loss (1.2) is linked to the risk function of the estimator being assessed for minimaxity. This interesting question is first studied in Section 2.1 with examples of non-minimaxity provided for estimators of a mean θ of the type $\delta(X) = aX$. Although such findings are not necessarily linked to exponential families, they do apply for families in (1.1) and we identify many cases of non *self-minimaxity* when the choice $a = \frac{1}{1+\lambda}$ is not made. Section 2.2 contains the principal minimax results described in the paragraph above, and examples are presented in Section 2.3. Section 3 deals with improved estimators under squared error loss of a lower-bounded expectation for discrete models using an adaptation of Kubokawa's integral expression of risk difference technique. For Poisson and negative binomial models, the classes of improved estimators are minimax under normalized squared error loss by virtue of the minimax findings of Section 2.2.

2. Main results

2.1. Non-minimaxity results

The following non-minimaxity result applies to one-parameter exponential families, but as well more generally to estimating a mean θ with finite variance.

Theorem 2.1. *Let X be a random variable on the real line with $\theta = \mathbb{E}(X)$, $Var_\theta(X) = \mathbb{V}(\theta)\,, \theta \in \Theta$, and $\lambda(\theta) = \frac{\mathbb{V}(\theta)}{\theta^2}$. Assume that $\lambda(\theta)$ is continuous and that Θ is connected. Let $\underline{\lambda} = \inf_{\theta\in\Theta} \lambda(\theta)$ and $\bar{\lambda} = \sup_{\theta\in\Theta} \lambda(\theta)$, and assume that $\underline{\lambda} > 0$ or $\bar{\lambda} < \infty$ (or both). Consider estimating θ under loss $L_{\lambda,0}$ in (1.2). Then, for any $\lambda < \underline{\lambda}$ or $\lambda > \bar{\lambda}$, the estimator $X/(1+\lambda)$ is inadmissible and dominated in both risk and maximum risk, by $X/(1+\underline{\lambda})$ in the former case, and by $X/(1+\bar{\lambda})$ in the latter case.*

Proof. With

$$\begin{aligned}
\mathbb{E}_\theta(aX-\theta)^2 &= a^2\,\mathbb{V}(\theta) \,+\, (a-1)^2\,\theta^2 \\
&= \theta^2\,\{a^2\lambda(\theta) \,+\, (a-1)^2\,\} \\
&= \theta^2\,(1+\lambda(\theta))\,\left\{(a-\frac{1}{1+\lambda(\theta)})^2 \,+\, \frac{\lambda(\theta)}{(1+\lambda(\theta))^2}\right\}\,, \theta \neq 0\,,
\end{aligned}$$

we have

$$R(\theta, aX) \,=\, \frac{(a-\frac{1}{1+\lambda(\theta)})^2 \,+\, \frac{\lambda(\theta)}{(1+\lambda(\theta))^2}}{(\frac{1}{1+\lambda}-\frac{1}{1+\lambda(\theta)})^2 \,+\, \frac{\lambda(\theta)}{(1+\lambda(\theta))^2}}\,, \theta \neq 0. \tag{2.3}$$

Since $R(0, aX) \,=\, a^2(1+\lambda)^2$, and with the above risk depending on a only through the distance $|a - \frac{1}{1+\lambda(\theta)}|$, we infer that

$$1 = R\,(\theta, X/(1+\lambda)) > R\,(\theta, X/(1+\underline{\lambda})) \quad \text{for all } \theta \in \Theta, \lambda < \underline{\lambda}\,,$$

and

$$1 = R\,(\theta, X/(1+\lambda)) > R\,\big(\theta, X/(1+\bar{\lambda})\big) \quad \text{for all } \theta \in \Theta, \lambda > \bar{\lambda}\,,$$

which establishes the dominance parts. Since the risk $R(\theta, aX)$ depends on θ through $\lambda(\theta)$ only, it suffices to show for the dominance in maximum risk part that:

$$\begin{aligned}
(i)\; L_1 \,&=\, \lim_{\lambda(\theta)\to\bar{\lambda}} R\,(\theta, X/(1+\underline{\lambda})) \,<\, 1 \quad \text{for} \quad \lambda < \underline{\lambda}\,, \\
(ii)\; L_2 \,&=\, \lim_{\lambda(\theta)\to\underline{\lambda}} R\,\big(\theta, X/(1+\bar{\lambda})\big) \,<\, 1 \quad \text{for} \quad \lambda > \bar{\lambda}\,.
\end{aligned}$$

Clearly, (i) holds if $\bar{\lambda} < \infty$, and (ii) holds if $\underline{\lambda} > 0$. And finally, we obtain readily from (2.3) that $L_1 \,=\, \left(\frac{1+\lambda}{1+\underline{\lambda}}\right)^2 < 1$ in (i) if $\bar{\lambda} = \infty$, and $L_2 \,=\, \left\{\big(\bar{\lambda}\,(1+\lambda)/\lambda(1+\bar{\lambda}\big)\right\}^2 < 1$ in (ii) if $\underline{\lambda} = 0$. The proof is thus complete. □

It is immediate that the dominance part of the above result holds as well for weighted squared error losses $q(\theta)\,(\delta-\theta)^2$ with $q(\cdot) > 0$. As well, the above result is applicable both under a parametric restriction on θ, or not. Here are some applications of Theorem 2.1.

Example 2.1. *(Poisson) For $X \sim Poisson(\theta)$, we have $\mathbb{E}(X) = \theta, \mathbb{V}(\theta) = \theta$, and $\lambda(\theta) = \frac{\mathbb{V}(\theta)}{\theta^2} \,=\, \frac{1}{\theta}$.*

(A) *Under an upper bound restriction* $\theta \leq \theta_0$*, we have* $\underline{\lambda} = \inf_{\theta \leq \theta_0} \lambda(\theta) = \frac{1}{\theta_0}$*. Theorem 2.1 thus applies and tells us that estimators* $X/(1+\lambda)$ *are, for* $0 \leq \lambda < \frac{1}{\theta_0}$ *and loss* $L_{\lambda,0}$*, not minimax for* $\theta \leq \theta_0$ *and dominated by* $X/(1+\underline{\lambda}) = \frac{\theta_0}{\theta_0+1}X$ *in both risk and maximum risk. The dominated estimators include* X *which is, for loss* $(\delta-\theta)^2/\theta$ *and in contrast, unique minimax for the unconstrained parameter space* $\theta \in \mathbb{R}_+$*, and minimax under a lower bound restriction* $\theta \geq \theta_0$ *as addressed below (e.g., Theorem 2.2, Section 2.3). As a complement, interesting minimax analyses for* $\theta \leq \theta_0$ *were given by Johnstone & MacGibbon (1992).*

(B) *Under a lower bound restriction* $\theta \geq \theta_0$*, we have* $\bar{\lambda} = \sup_{\theta \geq \theta_0} \lambda(\theta) = \frac{1}{\theta_0}$*. It thus follows from Theorem 2.1 that estimators* $X/(1+\lambda)$ *are, for* $\lambda > \frac{1}{\theta_0}$ *and loss* $L_{\lambda,0}$*, not minimax for* $\theta \geq \theta_0$ *and dominated by* $X/(1+\bar{\lambda}) = \frac{\theta_0}{\theta_0+1}X$ *in both risk and maximum risk. Finally, observe as well that we can fix* $\lambda \in (0,\infty)$ *and infer the same dominance result for* $\theta \geq \theta_0$ *with* $\theta_0 > 1/\lambda$*. This tells us that the non self-minimaxity of* $X/(1+\lambda)$ *is inevitable for sufficiently large* θ_0 *and* $\lambda > 0$*. As mentioned above in (A), it thus follows that only the case* $\lambda = 0$ *is immune to this property.*

Example 2.2. *(Negative binomial) Consider a negative binomial model* $NB(r,\theta)$ *for* X *with*

$$p_\theta(x) = \frac{(r)_x}{x!}\left(\frac{r}{r+\theta}\right)^r\left(\frac{\theta}{r+\theta}\right)^x, \text{ for } x \in \mathbb{N}. \tag{2.4}$$

Here, we have $\mathbb{E}(X) = \theta$*,* $\mathbb{V}(\theta) = \theta(\theta+r)/r$*, and* $\lambda(\theta) = \frac{1}{\theta} + \frac{1}{r}$*.*

(A) *Under an upper bound restriction* $\theta \leq \theta_0$*, we have* $\underline{\lambda} = \inf_{\theta \leq \theta_0} \lambda(\theta) = \frac{1}{\theta_0} + \frac{1}{r}$*. Theorem 2.1 thus applies and tells us that estimators* $X/(1+\lambda)$ *are for* $0 \leq \lambda < \frac{1}{\theta_0} + \frac{1}{r}$ *and loss (1.2) not minimax for* $\theta \leq \theta_0$ *and dominated by* $X/(1+\underline{\lambda}) = \frac{X}{1+1/\theta_0+1/r}$ *in both risk and maximum risk. The dominated estimators in both risk and maximum risk include* $rX/(r+1)$ *for loss (1.2) with* $w(\theta) = \mathbb{E}\left\{(\frac{r}{r+1}X-\theta)^2\right\} = \frac{r\theta(\theta+r)+\theta^2}{(r+1)^2}$*. The dominance persists of course for other weighted squared error losses* $q(\theta)(\delta-\theta)^2$*, namely for* $q(\theta) = 1/\mathbb{E}\left\{(X-\theta)^2\right\} = \frac{r}{\theta(\theta+r)}$*. With respect to this latter loss, in contrast to the non-minimaxity example here for* $\theta \leq \theta_0$*, the estimator* $rX/(r+1)$ *is unique minimax for the unconstrained parameter space* $\theta \in \mathbb{R}_+$*, and minimax under a lower bound restriction* $\theta \geq \theta_0$ *as addressed below (see Example 2.6, iv).*

(B) *Under a lower bound restriction* $\theta \geq \theta_0$*, we have* $\underline{\lambda} = \frac{1}{r}$ *and* $\bar{\lambda} = \frac{1}{\theta_0} + \frac{1}{r}$*. It follows from Theorem 2.1 that, for* $\lambda < 1/r$*,* $X/(1+\lambda)$ *is not minimax for loss* $L_{\lambda,0}$ *and dominated in both risk and maximum risk by* $X/(1+\underline{\lambda}) = \frac{rX}{r+1}$*. Actually this is true for* $\theta_0 = 0$ *as well. For* $\lambda > \frac{1}{\theta_0} + \frac{1}{r}$*, Theorem 2.1 tells us that* $X/(1+\lambda)$ *is not minimax for loss* $L_{\lambda,0}$ *and dominated in both risk and maximum risk by* $\frac{X}{1+1/\theta_0+1/r}$*. Alternatively, for any fixed* $\lambda > 1/r$*, the same dominance result applies as long as* $\theta_0 > r/(\lambda r - 1)$*.*

Example 2.3. *(Location families) Let* $X \sim N(\theta,\sigma^2)$ *with known* σ^2*, or more generally let* $X \sim f(x-\theta)$ *with* $\mathbb{E}(X) = \theta$*,* $\mathbb{V}(\theta) = \sigma^2$*, and* $\lambda(\theta) = \sigma^2/\theta^2$*. Then, for the parameter restriction* $|\theta| \leq \theta_0$*, we have* $\underline{\lambda} = \sigma^2/\theta_0^2$ *and Theorem 2.1 tells us that estimators* $X/(1+\lambda)$ *are, for* $\lambda \in [0,\underline{\lambda})$ *and loss (1.2), not minimax for* $|\theta| \leq \theta_0$ *and dominated by* $X/(1+\underline{\lambda}) = \frac{\theta_0^2 X}{\theta_0^2+\sigma^2}$ *in both risk and maximum risk. Such dominated estimators include* X *(e.g., Lehmann & Casella, 1998, page 327).*

Example 2.4. *(Inverse gaussian) Consider an inverse Gaussian distribution for* X *with density*

$$p(x) = \sqrt{\frac{\xi}{2\pi x^3}}\, e^{-\frac{\xi}{2x}}\, e^{-\frac{x\xi}{2\theta^2}+\frac{\xi}{\theta}}\, \mathbb{I}_{(0,\infty)}(x).$$

Here, for known $\xi > 0$, we have an example of model (1.1) with $\eta = -\frac{\xi}{2\theta^2} \in (-\infty, 0)$, $A(\eta) = (-2\xi\eta)^{1/2}$, $\mathbb{E}_\theta(X) = A'(\eta) = \theta$, $\mathbb{V}(\theta) = A''(\eta) = \theta^3/\xi$, and $\lambda(\theta) = \theta/\xi$. Under an upper bound restriction $\theta \le \theta_0$, we have $\bar{\lambda} = \theta_0/\xi$ and it follows from Theorem 2.1 that estimators $X/(1+\lambda)$ are, for $\lambda > \theta_0/\xi$ and loss (1.2), not minimax and dominated by $X/(1+\bar{\lambda}) = \frac{\xi X}{\xi+\theta_0}$. As well, under the restriction $\theta \ge \theta_0$, we have $\underline{\lambda} = \theta_0/\xi$ and we infer from Theorem 2.1 that estimators $X/(1+\lambda)$ are, for $\lambda \in [0, \theta_0/\xi)$ and loss (1.2), dominated in risk and maximum risk by $\frac{\xi X}{\xi+\theta_0}$.

Example 2.5. *Consider a random variable X with $\theta = \mathbb{E}(X) \in (\underline{\theta}, \infty)$, $\mathbb{V}(\theta) = c\,\theta^\alpha$, $\lambda(\theta) = c\theta^{\alpha-2}$ with $\alpha \ge 0$; the location families in Example 2.3 covered here by cases $\alpha = 0$, the Poisson case covered by $\alpha = 1$, and the inverse gaussian arising with $\alpha = 3$. For $\alpha = 2$, such as for Gamma models, the risk of aX is, under any loss $L_{\lambda,0}$ constant in θ and an optimal choice among multiples of X is given by $X/(1+c)$. We do not need Theorem 2.1, but it applies anyway with $\underline{\lambda} = \bar{\lambda} = c$. For $0 \le \alpha < 2$ and $\theta \ge \theta_0 > 0$, it follows from Theorem 2.1 that $X/(1+\lambda)$ is, for $\lambda > \bar{\lambda}$ and loss $L_{\lambda,0}$, dominated in risk and maximum risk by $X/(1+\bar{\lambda})$ with $\bar{\lambda} = \lambda(\theta_0)$. For $0 \le \alpha < 2$ and $0 < \theta \le \theta_0$, we obtain that $X/(1+\lambda)$ is, for $0 \le \lambda < \underline{\lambda}$ and loss $L_{\lambda,0}$, dominated in risk and maximum risk by $X/(1+\underline{\lambda})$ with $\underline{\lambda} = \lambda(\theta_0)$. Similar inferences, such as those in the inverse gaussian example, follow for cases $\alpha > 2$.*

2.2. Minimaxity results

As implied by Theorem 2.1 for exponential family model (1.1), when $\lambda(\theta)$ is monotone increasing or decreasing with $\lim_{\theta\to\bar{\theta}} \lambda(\theta) = \lambda_0$, there exists some (and many) values θ_0 such that $X/(1+\lambda)$ for $\lambda \ne \lambda_0$ is not minimax for $\theta \ge \theta_0$ under weighted squared error loss (1.2). In contrast, as expanded upon in this section for one-parameter exponential families, for $\lambda = \lambda_0$ and whenever $X/(1+\lambda)$ is admissible as an estimator of $\theta \in (\underline{\theta}, \infty)$ by Karlin's theorem, the estimator $X/(1+\lambda)$ is minimax for $\theta \ge \theta_0$ under loss $L_{\lambda,0}$ in (1.2).

For X exponentially family distributed as in model (1.1), Karlin's sufficient condition for $\frac{X}{1+\lambda} + \frac{\gamma\lambda}{1+\lambda}$ to be an admissible estimator of θ, $\theta \in (\underline{\theta}, \infty)$ is that the integrals

$$\int_{\eta_0}^{\bar{\eta}} e^{\lambda\{A(\eta)\,-\,\gamma\eta\}}\, d\eta \text{ and } \int_{\underline{\eta}}^{\eta_0} e^{\lambda\{A(\eta)\,-\,\gamma\eta\}}\, d\eta \tag{2.5}$$

diverge for some η_0. Under the given model, the quadratic risk of $X/(1+\lambda)$, or weight in (1.2), is given by:

$$w(\theta) \;=\; \mathbb{E}\,(\frac{X}{1+\lambda} - \theta)^2 \;=\; \frac{\mathbb{V}(\theta)}{(1+\lambda)^2} \;+\; (\frac{\lambda}{1+\lambda})^2\,\theta^2\,. \tag{2.6}$$

Here is the main result of this section.

Theorem 2.2. *Assume model (1.1), $\bar{\theta} = \infty$, $\lim_{\theta\to\infty} \lambda(\theta) = \lambda \in [0, \infty)$, and that $\delta_0(X) = \delta_{\lambda,0}(X) = X/(1+\lambda)$ is admissible by Karlin's theorem with the integrals in (2.5) divergent. Then, $\delta_0(X)$ is minimax under the parametric restriction $\theta \ge \theta_0$ for weighted squared error loss $L_{\lambda,0}$ in (1.2) with $w(\theta)$ as in (2.6).*

Proof. Assume that $\delta_0(X)$ is not minimax. Then, there would exist an estimator $\delta(X)$ and $\epsilon > 0$ such that for all $\theta \ge \theta_0$:

$$\frac{\mathbb{E}\,\{(\delta(X) - \theta)^2\}}{\mathbb{E}\,\{(\delta_0(X) - \theta)^2\}} \;<\; 1 - \epsilon\,,$$

or, equivalently via (2.6)

$$\mathbb{E}\left\{(\delta(X)-\theta)^2\right\} \;\leq\; (1-\epsilon)\,\{\frac{\mathbb{V}(\theta)}{(1+\lambda)^2} \;+\; (\frac{\lambda}{1+\lambda})^2\,\theta^2\}\,. \tag{2.7}$$

Now, with $I(\theta) \;=\; 1/\mathbb{V}(\theta)$ for this given model, where $I(\theta)$ is the Fisher information, and setting

$$b(\theta) \;=\; \mathbb{E}_\theta(\delta(X)) \;-\; \mathbb{E}_\theta(\delta_0(X)) \;=\; \mathrm{Bias}_\theta(\delta(X)) \;+\; \frac{\lambda}{1+\lambda}\theta\,,$$

the Cramér-Rao inequality

$$\begin{aligned}\mathbb{E}\left((\delta(X)-\theta)^2\right) \;&\geq\; \{\mathrm{Bias}_\theta(\delta(X))\}^2 \;+\; \frac{\left\{1+\frac{d}{d\theta}\mathrm{Bias}_\theta(\delta(X))\right\}^2}{I(\theta)} \\ &=\; \left(b(\theta)-\frac{\lambda}{1+\lambda}\theta\right)^2 \;+\; \left(b'(\theta)+\frac{1}{1+\lambda}\right)^2\,\mathbb{V}(\theta)\,,\end{aligned}$$

along with inequality (2.7) would imply

$$\left\{b'(\theta)^2+\frac{2b'(\theta)}{1+\lambda}\right\}\mathbb{V}(\theta) \;+\; b^2(\theta) \;-\; 2\theta b(\theta)\frac{\lambda}{1+\lambda} \;\leq\; -\frac{\epsilon}{(1+\lambda)^2}\left\{\mathbb{V}(\theta) \;+\; \lambda^2\,\theta^2\right\}\,,$$

and, as well,

$$\frac{2b'(\theta)}{1+\lambda}\,\mathbb{V}(\theta) \;+\; b^2(\theta) \;-\; 2\theta\,b(\theta)\frac{\lambda}{1+\lambda} \;\leq\; -\frac{\epsilon}{(1+\lambda)^2}\left\{\mathbb{V}(\theta) \;+\; \lambda^2\,\theta^2\right\}\,. \tag{2.8}$$

We pursue by showing that there cannot exist a $b(\theta)$ (equivalently an estimator $\delta(X)$) such that (2.8) is satisfied for all $\theta \geq \theta_0$. We first make use of the fact that if $b(\theta') < 0$ for some $\theta' \geq \theta_0$, there exists no solution to (2.8). This is due to the proof of Karlin's theorem (e.g., Lehmann & Casella, 1998; Brown, 1986). We cannot have $b(\theta') = 0$ for some θ' neither since (2.8) would imply $b'(\theta') < 0$ and then $b(\theta'+h) < 0$ for some $h > 0$. We therefore assume hereafter that $b(\theta) > 0$ for all $\theta \geq \theta_0$ and we proceed by treating the cases : (I) $\lambda = 0$ and (II) $\lambda > 0$ separately.

(I) For $\lambda = 0$, expression (2.8) implies that $b'(\theta) < -\epsilon/2$ for all $\theta \geq \theta_0$ which in turn implies that $b(\theta') < 0$ for some $\theta' \geq \theta_0$ which leads to a contradiction as outlined in the above paragraph.

(II) For $\lambda > 0$, we establish and make use of three intermediate results which are as follows, the proofs of which are relegated to the Appendix.

Lemma 2.1. *For some $\tilde{\theta} \geq \theta_0$ and $K > 0$, we have $b(\theta) \leq K\theta$ for all $\theta \geq \tilde{\theta}$.*

Lemma 2.2. *We have $\lim_{\theta\to\infty} \frac{b(\theta)}{\theta} \;=\; 0$.*

Lemma 2.3. *There exists a sequence $\{\theta_n : n \in \mathbb{N}_+\}$ such that $\theta_n \to \infty$ and $b'(\theta_n) \to 0$ as $n \to \infty$.*

Now, Lemma 2.3, together with inequality (2.8), imply that for the sequence $\{\theta_n : n \geq 1\}$, there exists $0 < \epsilon' < \frac{\lambda}{1+\lambda}$ such that

$$\frac{b^2(\theta_n)}{\theta_n^2} \;-\; \frac{2\lambda}{1+\lambda}\,\frac{b(\theta_n)}{\theta_n} \;\leq\; -\epsilon'\,\frac{\lambda}{1+\lambda}\,,$$

for sufficiently large n. But, this quadratic inequality implies that

$$\frac{b(\theta_n)}{\theta_n} \geq \frac{\lambda}{1+\lambda} - \sqrt{\frac{\lambda}{1+\lambda}\left(\frac{\lambda}{1+\lambda} - \epsilon'\right)} > 0\,,$$

contradicting Lemma 2.2, showing finally that (2.8) cannot hold and establishing the result that $X/(1+\lambda)$ is a minimax estimator under the given conditions. □

The above finding relates to the minimaxity of $\delta_0(X) = X/(1+\lambda)$ for weighted squared error loss (1.2) with weight associated to mean square error of $\delta_0(X)$. The following intermediate result will permit us to infer the minimaxity of $\delta_0(X)$ for a larger class of weighted squared error losses.

Lemma 2.4. *Consider a model $X \sim f_\theta$, a non-negative loss function $L_0(\theta,\delta)$ for estimating $\theta \in \Theta_0 = (\theta_0, \infty)$, and positive functions $w_1(\theta)$ and $w_2(\theta)$ such that $\frac{w_1(\theta)}{w_2(\theta)}$ is non-decreasing on Θ_0 with limit $c = \lim_{\theta\to\infty} \frac{w_1(\theta)}{w_2(\theta)}$. Further consider weighted loss functions $L_i(\theta,\delta) = \frac{L_0(\theta,\delta)}{w_i(\theta)}$ for i=1,2, and suppose that $\delta^*(X)$ is, under $L_1(\theta,\delta)$ and for $\theta \in \Theta_0$, a minimax estimator with constant risk equal to 1. Now, assume*

<u>Condition C0</u>: For any estimator $\delta(X)$, there exists a subsequence $\{\theta_n : n \in \mathbb{N}_+\}$ tending to infinity such that for any $\epsilon > 0$: there is a n_ϵ with $\mathbb{E}\, L_1(\theta_n, \delta(X)) > \mathbb{E}\, L_1(\theta_n, \delta^*(X)) - \epsilon = 1 - \epsilon$ for all $n > n_\epsilon$.

Then, $\delta^*(X)$ is a minimax estimator of θ under loss $L_2(\theta,\delta)$ and for $\theta \in \Theta_0$ with minimax risk equal to c.

Proof. Denote $R_i(\theta,\delta)$ the risk associated with loss $L_i(\theta,\delta)$. For any δ, we have

$$\begin{aligned}
\sup_{\theta\in\Theta_0} R_2(\theta,\delta) &\geq \sup_{n>n_\epsilon} R_2(\theta_n,\delta) \\
&= \sup_{n>n_\epsilon} \left\{ R_1(\theta_n,\delta)\, \frac{w_1(\theta_n)}{w_2(\theta_n)} \right\} \\
&\geq (1-\epsilon) \sup_{n>n_\epsilon} \frac{w_1(\theta_n)}{w_2(\theta_n)} \\
&= (1-\epsilon) \lim_{\theta_n\to\infty} \frac{w_1(\theta_n)}{w_2(\theta_n)} \\
&= (1-\epsilon)\, c\,.
\end{aligned}$$

Since ϵ is arbitrary, we infer that $\sup_{\theta\in\Theta_0} R_2(\theta,\delta) \geq c = \sup_{\theta\in\Theta_0} R_2(\theta,\delta^*)$, which establishes the result. □

Corollary 2.1. *Consider the context of Theorem 2.2, but with loss $\frac{(\delta-\theta)^2}{w_2(\theta)}$ instead for positive and continuous $w_2(\theta)$ such that $\frac{\mathbb{E}(\delta_0(X)-\theta)^2}{w_2(\theta)}$ is non-decreasing in $\theta \in \Theta_0$ with $c = \lim_{\theta\to\infty} \frac{\mathbb{E}(\delta_0(X)-\theta)^2}{w_2(\theta)}$. Then, the estimator $\delta_0(X) = X/(1+\lambda)$ remains minimax with minimax risk equal to c.*

Proof. Apply Lemma 2.4 with $\delta^* = \delta_0$, $L_0(\theta,\delta) = (\delta-\theta)^2$, and $w_1(\theta) = \mathbb{E}(\delta_0(X)-\theta)^2$. It follows from Theorem 2.2 that $\delta^* = \delta_0$ is a minimax estimator for $L_1(\theta,\delta) = \frac{(\delta-\theta)^2}{w_1(\theta)}$. As a consequence, Condition C0 is met and the result follows. □

For an upper bound restriction, we have the following analogous result, the proof of which is essentially the same as that of Theorem 2.2.

Theorem 2.3. *Assume model (1.1),* $\underline{\theta} = -\infty$, $\lim_{\theta\to-\infty} \lambda(\theta) = \lambda \in [0,\infty)$, *and that* $\delta_0(X) = X/(1+\lambda)$ *is admissible by Karlin's theorem with the integrals in (2.5) divergent. Then,* $\delta_0(X)$ *is minimax under the parametric restriction* $\theta \le \theta_0$ *and for weighted squared error loss* $L_{\lambda,0}$ *in (1.2).*

We conclude this section with an extension of Theorem 2.2 to estimators $\delta_{\lambda,\gamma}(X)$.

Corollary 2.2. *Assume model (1.1),* $\bar{\theta} = \infty$, $\lim_{\theta\to\infty} \lambda(\theta) = \lambda \in [0,\infty)$, *and that* $\delta_{\lambda,\gamma}(X) = \frac{X}{1+\lambda} + \frac{\gamma\lambda}{1+\lambda}$ *is admissible by Karlin's theorem with the integrals in (2.5) divergent. Then,* $\delta_{\lambda,\gamma}(X)$ *is minimax under the parametric restriction* $\theta \ge \theta_0$ *and for weighted squared error loss* $L_{\lambda,\gamma}(\theta,\delta) = \frac{(\delta-\theta)^2}{w(\theta)}$ *with* $w(\theta) = \mathbb{E}(\delta_{\lambda,\gamma}(X)-\theta)^2 = \frac{\mathbb{V}(\theta)}{(1+\lambda)^2} + \frac{\lambda^2}{(1+\lambda)^2}(\theta-\gamma)^2$.

Proof. We make use of an argument similar to Brown (1986). Let $Y = X - \gamma$. With model (1.1) representative of the distribution of Y with $E_\theta(Y) = \theta - \gamma$ and with $A_Y(\eta) = A(\eta) - \eta\gamma$, we have

$$\lambda_Y(\theta) = \frac{\mathrm{Var}_\theta(Y)}{(\mathbb{E}_\theta(Y))^2} = \frac{\lambda(\theta)\theta^2}{(\theta-\gamma)^2} \to \lambda \text{ as } \theta \to \infty\,.$$

Hence, $\frac{Y}{1+\lambda}$ is by Theorem 2.2 minimax for estimating $\mathbb{E}_\theta(Y)$ under the restriction $\mathbb{E}(Y) \ge \theta_0 - \gamma$ under weighted squared error loss $\frac{\{\delta-(\theta-\gamma)\}^2}{w_Y(\theta-\gamma)}$ with $w_Y(\theta-\gamma) = \mathbb{E}(\frac{Y}{1+\lambda} - (\theta-\gamma))^2$. Finally, the result follows as $w_Y(\theta-\gamma) = \mathbb{E}(\frac{X+\gamma\lambda}{1+\lambda} - \theta)^2$. □

2.3. Examples

This section is devoted to illustration and further observations.

Example 2.6. **(i)** *(Poisson) Let* $X \sim$ *Poisson*(θ) *with* $\theta \ge \theta_0$. *Here* $\lambda = \lim_{\theta\to\infty} \frac{\mathbb{V}(\theta)}{\theta^2} = \lim_{\theta\to\infty} \frac{1}{\theta} = 0$. *With* $\eta = \log\theta \in (-\infty,\infty)$, *it is easily verified that the integrals in (2.5) are divergent with* $\lambda = \gamma = 0$, *and it thus follows from Theorem 2.2 that* $\delta_0(X) = X$ *is minimax under loss* $\frac{(\delta-\theta)^2}{w(\theta)}$ *with* $w(\theta) = \theta$ *obtained from (2.6). This is a known result from van Eeden (2006, page 57).*

(ii) *(Normal) Theorem 2.2 applies for* $X \sim N(\theta,\sigma_0^2)$ *for the restriction* $\theta \ge \theta_0$ *and with known* σ_0^2. *With* $\lambda(\theta) = \sigma_0^2/\theta^2$, $\lambda = 0$, $\eta = \theta \in (-\infty,\infty)$, *it is easily verified that the integrals in (2.5) are divergent with* $\lambda = \gamma = 0$. *It thus follows from Theorem 2.2 that* $\delta_0(X) = X$ *is minimax under loss* $(\delta-\theta)^2/\sigma_0^2$, *and squared error loss as well. The result is due to Katz (1961) (also see van Eeden, 2006). Analogously, Theorem 2.3 applies yielding the minimaxity of* $\delta_0(X)$ *for the same loss and for* $\theta \le \theta_0$.

(iii) *(Gamma) Consider* $X \sim G(\alpha,\beta)$ *with known* α, *unknown* β, *and mean* $\theta = \alpha\beta$. *Density (1.1) is representative of the density of* X *for* $\eta = -1/\beta \in (-\infty,0)$ *and* $A(\eta) = -\alpha\log(-\eta)$. *With* $\mathbb{V}(\theta) = \theta^2/\alpha$, *we have* $\lambda(\theta) = 1/\alpha = \lim_{\theta\to\infty}\lambda(\theta) = \lambda$ *and it is thus pertinent to study the applicability of Theorems 2.2, Corollary 2.2 and Theorem 2.3 to estimators* $\delta_{\frac{1}{\alpha},\gamma}(X) =$

$\frac{\alpha}{\alpha+1} X + \frac{\gamma}{\alpha+1}$. *As well known (e.g., Lehmann & Casella, 1998), the integrals in (2.5), which reduce to*

$$\int_{\eta_0}^{0} -\frac{1}{\eta}\, e^{-\gamma\eta/\alpha}\, d\eta \text{ and } \int_{-\infty}^{\eta_0} -\frac{1}{\eta}\, e^{-\gamma\eta/\alpha}\, d\eta\,, \tag{2.9}$$

diverge for some η_0 if and only if $\gamma \geq 0$. Theorem 2.2 hence applies and tells us that $\delta_{\frac{1}{\alpha},0}(X)$ is minimax for estimating $\theta \in [\theta_0, \infty)$ under loss (1.2) with $w(\theta) = \theta^2/(\alpha+1)$, or equivalently for loss $(\delta-\theta)^2)/\theta^2$. Translating the above to the problem of estimating β, we infer that $\hat{\beta}_0(X) = X/(\alpha+1)$ is minimax for estimating β with $\beta \geq \beta_0$ and loss $(\hat{\beta}-\beta)^2/\beta^2$. This is a known result from van Eeden (1995). Corollary 2.2 also applies and tells us that $\delta_{\frac{1}{\alpha},\gamma}(X)$ is minimax for estimating $\theta \in [\theta_0, \infty)$ under weighted squared error loss $\frac{(\delta-\theta)^2}{w(\theta)}$ with $w(\theta) = \frac{\alpha\theta^2}{(\alpha+1)^2} + \frac{(\theta-\gamma)^2}{(\alpha+1)^2}$.

The problem of specifying dominating estimators of $\delta_{\frac{1}{\alpha},0}(X)$ for the parametric restriction $\theta \geq \theta_0$ is of interest and has been studied by van Eeden (1995), as well as Marchand & Strawderman (2005), among others. Such dominating estimators include the Bayes estimator of θ with respect to the truncation onto $[\theta_0, \infty)$ of the non-informative prior $1/\theta$, for a large class of losses which include weighted squared error loss $\frac{(\delta-\theta)^2}{\theta^2}$.

(iv) *(Negative binomial) Consider a negative binomial model for X with $p(x) = \frac{(r)_x}{x!}\, p^r (1-p)^x\, \mathbb{I}_{\mathbb{N}}(x)$, and the parametric restriction $p \in (0, p_0]$. We have $\theta = \mathbb{E}(X) = r(1-p)/p$, $\mathbb{V}(\theta) = \theta(\theta+r)/r$, and the restriction on p corresponds to $\theta \geq \theta_0$ with $\theta_0 = r(1-p_0)/p_0$. With $\lambda = \lim_{\theta\to\infty} \frac{(r+\theta)}{r\theta} = \frac{1}{r}$, and the integrals in (2.5) shown to be divergent, Theorem 2.2 applies confirming the minimaxity of $\delta_{\frac{1}{r},0}(X) = rX/(r+1)$ for weighted squared error loss (1.2) with $w(\theta) = \mathbb{E}\left\{(\frac{r}{r+1}X-\theta)^2\right\} = \frac{r\theta(\theta+r)+\theta^2}{(r+1)^2}$ obtained from (2.6). Furthermore, an application of Corollary 2.1 with $w_1(\theta) = w(\theta)$ and $w_2(\theta) = \theta(\theta+r)/r$ tells us that $\delta_{\frac{1}{r},0}(X)$ is also minimax for loss $\frac{(\delta-\theta)^2}{w_2(\theta)}$ with minimax risk $c = \lim_{\theta\to\infty} \frac{w_1(\theta)}{w_2(\theta)} = \frac{r}{r+1}$. In the latter case, the minimaxity property of $\delta_{\frac{1}{r},0}(X)$, applicable here for $\theta \geq \theta_0$ with $\theta_0 > 0$, extends the known minimaxity result for $\delta_{\frac{1}{r},0}(X)$ and $\theta_0 = 0$. The findings of this example have not been previously established to the best of our knowledge.*

Remark 2.1. *The minimax properties in parts* **(ii)** *and* **(iii)** *above extend much more generally to location families in* **(ii)** *and scale families in* **(iii)** *with either a lower-bound or upper-bound parametric constraint on θ, with the minimum risk equivariant estimator, of the form $X + c^*$ in* **(ii)** *and a^*X in* **(iii)**, *being minimax for invariant loss of the form $\rho(\delta-\theta)$ in* **(ii)** *and of the form $\rho(\delta/\theta)$ in* **(iii)**. *This follows from findings of Marchand and Strawderman (2012), which are applicable to a general framework with an invariance structure. Their results however do not cover discrete exponential families such as the Poisson and Negative Binomial families studied above.*

In the context of Theorem 2.2, dominating estimators of $\delta_0(X) = X/(1+\lambda)$ are necessarily minimax. These include for instance their truncations onto the restricted parameter space given by $\max\{\theta_0, \delta_0(X)\}$. For the Poisson and negative binomial models, further dominating estimators are given in Section 3. It is instructive to immediately identify estimators $\delta_{\lambda,\gamma}(X)$ that dominate $\delta_{\lambda,0}(X)$ whenever $\lambda > 0$.

Lemma 2.5. *Let X be a non-negative random variables with $X \sim p_\theta$ for $\theta \geq 0$ such that $\mathbb{E}(X) = \theta$ and $Var_\theta(X) = \mathbb{V}(\theta) < \infty$. Consider estimators $\delta_{\lambda,\gamma}(X)$ for $\lambda > 0$ and the parametric restriction*

$\theta \geq \theta_0 > 0$. *Then, estimators* $\delta_{\lambda,\gamma}(X) = \frac{X}{1+\lambda} + \frac{\gamma\lambda}{1+\lambda}$ *dominate* $\delta_{\lambda,0}(X)$ *under squared error loss* $(\delta-\theta)^2$ *if and only if* $0 < \gamma \leq 2\theta_0$.

Proof. The result follows immediately with the risk evaluation $R(\theta, \delta_{\lambda,\gamma}) = \frac{1}{(1+\lambda)^2}\,\mathbb{V}(\theta) + (\frac{\lambda}{1+\lambda})^2\,\theta^2 + (\frac{\gamma\lambda}{1+\lambda})^2 - 2(\frac{\lambda}{1+\lambda})^2\,\gamma\theta \;=\; R(\theta, \delta_{\lambda,0}) + (\frac{\lambda}{1+\lambda})^2\,(\gamma^2 - 2\,\gamma\,\theta)$. □

The following is an immediate consequence of the above and Theorem 2.2.

Corollary 2.3. *In the context of Theorem 2.2 with* $\lambda > 0$, *estimators* $\delta_{\lambda,\gamma}(X)$ *are minimax, and dominate* $\delta_{\lambda,0}(X)$ *whenever* $0 < \gamma \leq 2\theta_0$, *under loss (1.2).*

The above dominating estimators $\delta_{\lambda,\gamma}(X)$ can actually take values solely on the restricted parameter space. With $\mathbb{P}(X \leq t) > 0$ for all $t > 0$, such a condition requires $\gamma \geq \frac{1+\lambda}{\lambda}\,\theta_0$ as well as $\gamma \leq 2\theta_0$, in other words $\lambda \geq 1$ in Corollary 2.3. The above corollary is applicable in both the Gamma and Negative Binomial cases.

Example 2.7.

- *For* $X \sim G(\alpha, \beta)$ *as above in* **(iii)**, *estimators* $\delta_{\frac{1}{\alpha},\gamma}(X)$ *of* $\theta = \mathbb{E}_\theta(X)$ *with* $0 < \gamma \leq 2\theta_0$ *are minimax and dominate* $\delta_{\frac{1}{\alpha},0}(X)$ *for weighted squared error loss* $(\delta-\theta)^2/\theta^2$ *and the parametric restriction* $\theta \geq \theta_0$. *For instance, in the exponential case with* $\alpha = 1$, *the estimator* $(X/2) + \theta_0$ *is minimax and takes values solely on the parameter space.*

- *For the Negative Binomial model as above in part* **(iv)** *of Example 2.6, estimators* $\delta_{\frac{1}{r},\gamma}(X)$ *of* $\theta = \mathbb{E}_\theta(X)$ *with* $0 < \gamma \leq 2\theta_0$ *are minimax and dominate* $\delta_{\frac{1}{r},0}(X)$ *for weighted squared error losses* $\frac{(\delta-\theta)^2}{w_i(\theta)}, i = 1, 2$, *and the parametric restriction* $\theta \geq \theta_0$, *with the given* w_i*'s. For instance, in the geometric case with* $r = 1$, *the estimator* $(X/2) + \theta_0$ *is minimax and takes values solely on the parameter space.*

3. Minimax estimation of a lower-bounded expectation for discrete models

For $X \sim \text{Poisson}(\theta)$ as well as $X \sim NB(r, \theta)$ with $\theta \geq \theta_0 > 0$, we have seen in Section 2 (e.g., Examples 2.6 **(i), (iv)**) that $\delta_0(X) = X$ and $\delta_0(X) = rX/(r+1)$ are respectively for these models minimax under weighted squared error loss (1.2). As well, following Corollary 2.1 and as illustrated for the negative binomial model with loss $\frac{r(\delta-\theta)^2}{\theta(\theta+r)}$, the minimax property carries over to a large class of weighted squared error loss functions. In this section for a class of discrete distributions, we provide estimators that dominate a target estimator $\delta_0(X)$ under squared error loss. Target estimators such as the above $\delta_0(X)$ are of interest because of their minimaxity and because we wish to have at our disposal alternative estimators that pass the minimum test of dominating $\delta_0(X)$. As particular cases, we obtain for both the Poisson and negative binomial models above with $\theta \geq \theta_0 > 0$ minimax estimators under weighted squared error (1.2) and Corollary 2.1's variants. To achieve this, we make use of Kubokawa's Integral Expression for Risk Difference (IERD; Kubokawa, 1994) adapted to count variables taking values on $\mathbb{N}$ and conditional moment properties of these distributions. Analogous results have been obtained for continuous models (see for instance Kubokawa, 1999; Marchand & Strawderman, 2004; for reviews and additional references), but the discrete model analysis presented here is not available in the literature to the best of our knowledge.

Hereafter, we consider target estimators $\delta_0(X)$, such that $\delta_0(x)$ is non-decreasing in $x \in \mathbb{N}$, and (possibly) shrinkers with respect to X, i.e., $\delta_0(x) \leq x$ for all $x \in \mathbb{N}$. These plausible conditions are satisfied by the target minimax estimators in both the Poisson and negative binomial given at the outset of this section, as well as for other situations described next.

Remark 3.2. *A large class of interesting models for which the dominance results below will apply, other than Poisson and negative binomial, are Poisson mixtures such that*

$$X|\lambda \sim Poisson(\lambda)\,,\ \lambda \sim g_\theta\,,$$

with $\mathbb{E}(\lambda) = \theta$ *and* $Var(\lambda) = k^2\,\theta^2$, $k > 0$ *known and representing the coefficient of variation of* λ. *The* $NB(r, \theta)$ *model is one such model with* $k = 1/\sqrt{r}$. *For such Poisson mixtures, we have*

$$\mathbb{E}(X) \,=\, \theta\,,\ \textit{and}\ \mathbb{V}(\theta) \,=\, \theta \,+\, k^2\,\theta^2\,.$$

From this, for weighted loss $\frac{(\delta-\theta)^2}{\mathbb{V}(\theta)}$, *the frequentist risk of a multiple* cX *of* X *with* $c \geq 0$ *is given by*

$$R(\theta, cX) \,=\, c^2 \,+\, \frac{(c-1)^2\theta}{1+k^2\theta}\,,$$

It is easy to see then that cX *has lower risk than* X *for all* $\theta \geq 0$ *whenever* $c \in [0 \vee \frac{1-k^2}{1+k^2}, 1)$. *As well, the choice* $\frac{X}{1+k^2}$ *is unique minimax among such multiples for* $\theta \geq \theta_0$ *and any* $\theta_0 > 0$. *We thus see that these estimators are shrinkers with respect to the unbiased estimator* X.

The next result represents the principal dominance finding of this section. Motivated by the Poisson and negative binomial models, it is cast more generally not only in continuation of the previous Remark, but also to handle cases where X is supported on the finite set $\{0, 1, \ldots, M\}$ with a parametric restriction of the form $\theta = \mathbb{E}(X) \in [\theta_0, M]$ with $\theta_0 > 0$. A prominent example, which is further detailed in Example 3.8 below, is given by a Binomial model $X \sim B(n, p)$ with $\theta = np$, $M = n$, and $p \geq p_0 = \theta_0/n$.

Theorem 3.4. *Consider* $X \sim p_\theta$, $\theta \geq 0$, *with* $p_\theta(t) = \mathbb{P}_\theta(X = t)$, $\mathbb{E}_\theta(X) = \theta$, $\mathbb{E}_\theta(X^2) < \infty$, *and support* S *independent of* θ *either of the form* $S = \mathbb{N}$ *or* $S = \{0, 1, \ldots, M\}$. *Consider estimators* $\delta_\psi(X) = \delta_0(X) + \psi(X)$, *where* $\delta_0(X)$ *is a target estimator such that* $\delta_0(x)$ *is non-decreasing in* x *and* $\delta_0(x) \leq x$ *for* $x \in S$, *and* ψ *is a non-increasing function on* S *such that* $\lim_{t\to\bar{s}} \psi(t) = 0$ *where* $\bar{s} = \sup S$. *Let*

$$\psi_1(t) \,=\, \inf_{\theta \geq \theta_0}\, \mathbb{E}_\theta\left(\theta - \delta_0(X)\,|X \leq t\,\right)\,, t \in S\,,\ \theta_0 > 0\,.$$

Then, with the further specification that $\psi(M+1) = 0$ *when* $S = \{0, 1, \ldots, M\}$, *under squared error loss and the lower-bound constraint* $\theta \geq \theta_0$,

(a) $\delta_\psi(X)$ *dominates* $\delta_0(X)$ *whenever*

$$\frac{\psi(t+1) + \psi(t)}{2} \,\leq\, \psi_1(t)\,,\ \textit{for all}\ t \in S\,; \tag{3.10}$$

(b) $\delta_{\psi^*}(X)$ *dominates* $\delta_0(X)$ *where* $\psi^*(t) = \psi_1(t) - \ell$ *with* $\ell = \lim_{t\to\bar{s}} \psi_1(t)$ *;*

(c) $\delta_{\psi^*}(X)$ *is range-preserving whenever* $\delta_0(X) = X$, *i.e.,* $\mathbb{P}_\theta(\delta_{\psi^*}(X) \geq \theta_0) = 1$ *for all* $\theta \geq \theta_0$, *while* $\delta^*(X) = \max\{\theta_0, \delta_{\psi^*}(X)\}$ *is range-preserving and dominates* $\delta_0(X)$.

Proof. **(a)** First, condition (3.10) is not vacuous since

$$\mathbb{E}_\theta\,(\theta - \delta_0(X)\,|X \le t\,) \;\ge \mathbb{E}_\theta\,(\theta - X\,|X \le t\,) \;\ge \theta - \min\{\theta, t\} \;=\; \max\{0, \theta - t\}\,, \tag{3.11}$$

which implies that

$$\psi_1(t) \;\ge\; \inf_{\theta \ge \theta_0} (\max\{0, \theta - t\}) \;\ge \max\{0, \theta_0 - t\}\,, \text{ for all } t \in S. \tag{3.12}$$

Now, we have for the difference in losses:

$$\begin{aligned}
(\delta_0(x) \,-\, \theta)^2 - (\delta_\psi(x) \,-\, \theta)^2 \;&=\; (\delta_0(x) \,-\, \theta \,+ \psi(t)\,)^2 \,|_{t=x}^{\bar{s}} \\
&=\; \sum_{t \ge x} \left[\{\delta_0(x) \,-\, \theta \,+ \psi(t+1)\,\}^2 \,-\, \{\delta_0(x) \,-\, \theta \,+ \psi(t)\,\}^2\right] \\
&=\; 2 \sum_{t \ge x} \{\psi(t+1) - \psi(t)\}\,\{\delta_0(x) - \theta + \frac{\psi(t+1) + \psi(t)}{2}\}\,.
\end{aligned}$$

As a difference in risks, we hence obtain

$$\begin{aligned}
\Delta(\theta) \;&=\; \mathbb{E}_\theta\,(\delta_0(X) - \theta)^2 \,-\, \mathbb{E}_\theta\,(\delta_\psi(X) - \theta)^2 \\
&=\; 2 \sum_{x \in S} p_\theta(x) \sum_{t \ge x} \{\psi(t+1) - \psi(t)\}\,\{\delta_0(x) - \theta + \frac{\psi(t+1) + \psi(t)}{2}\} \\
&=\; 2 \sum_{t \in S} \{\psi(t+1) - \psi(t)\} \sum_{x \le t} p_\theta(x)\,\{\delta_0(x) - \theta + \frac{\psi(t+1) + \psi(t)}{2}\}\,.
\end{aligned}$$

Since by assumption $\psi(t+1) - \psi(t) \le 0$ for all $t \in S$, a sufficient condition to have $\Delta(\theta) \ge 0$ is:

$$\frac{\sum_{x \le t} p_\theta(x)\,(\theta \,-\, \delta_0(x))}{\mathbb{P}_\theta(X \le t)} \;\ge\; \frac{\psi(t+1) + \psi(t)}{2}\,, \text{ for all } t \in S\,.$$

Finally, the result follows by minimizing the l.h.s. of the above expression for $\theta \ge \theta_0$.

(b) Since $\delta_0(x)$ is non-decreasing in x, it follows that $\mathbb{E}_\theta\,(\delta_0(X)|X \le t)$ is non-decreasing in $t \in S$. Since this holds for all θ, it follows that both $\psi_1(t)$ and $\psi^*(t)$ are non-increasing in $t \in S$. With $\lim_{t \to \bar{s}} \psi^*(t) \;=\; 0$ and $\frac{\psi^*(t+1) + \psi^*(t)}{2} \;\le\; \psi^*(t) \;\le\; \psi_1(t)$, condition (3.10) is satisfied and the result follows from part **(a)**.

(c) Clearly, $\delta^*(X)$ is range preserving and dominates $\delta_{\psi^*}(X)$. Part (**b**) therefore implies that $\delta^*(X)$ dominates $\delta_0(X)$. Finally for $\delta_0(X) = X$, since we have $\lim_{t \to \bar{s}} \mathbb{E}_\theta(\theta - X|X \le t\,) = 0$ for all $\theta \ge \theta_0$, it follows that $\ell = 0$ and

$$\delta_{\psi^*}(t) \;=\; \delta_{\psi_1}(t) \;=\; \inf_{\theta \ge \theta_0} \mathbb{E}_\theta\,(\theta + t - X\,|X \le t\,) \;\ge \theta_0\,,$$

for all $t \in S$. □

Remark 3.3. *The dominance results of Theorem 3.4 can be extended directly for an upper-bounded restriction $\theta \le \theta_1$ in cases where S is finite. This is achieved with the reflection $X' = M - X$ and the correspondence $\theta' = M - \theta$, because the risk of $\delta(X)$ at θ for estimating $\theta \in [0, \theta_1]$ is equivalent to the risk of $M - \delta(M - X')$ at θ' for estimating $\theta' \in [\theta_0, M]$.*

Example 3.8. *As previously mentioned, Theorem 3.4 applies for Binomial models $X \sim B(n,p)$ with $M = n$ and for a lower bound restriction $\theta = np \geq \theta_0$ with $\theta_0 \in (0,n)$. Such situations correspond to cases where the success parameter p is lower-bounded and applications also cover upper-bounded restrictions as pointed out in Remark 3.3. There has been some earlier work on estimating an upper-bounded or lower-bounded Binomial proportion. See for instance Marchand & MacGibbon (2000), as well as van Eeden (2006), for findings and useful references. A specific application of Theorem 3.4 is given for the target estimator $\delta_0(X) = X$ with $\ell = 0$ and $\psi_1(t) = \inf_{\theta \geq \theta_0} \mathbb{E}(\theta - X | X \leq t)$ for $t = 0, 1, \ldots, n$.*

The dominance finding of Theorem 3.4 leaves untouched the question of where the minimum value of $\mathbb{E}_\theta(\theta - \delta_0(X)|X \leq t)$ is attained on the set $\theta \in [\theta_0, \infty)$. We show below, for both the Poisson and negative binomial models and when $\delta_0(X)$ is the target minimax estimator given in examples **(i)** and **(iv)** of Section 2.3, that the infimum value is attained at the boundary $\theta = \theta_0$. We first address the Poisson case.

Corollary 3.4. *For $X \sim$ Poisson(θ) with $\theta \geq \theta_0 > 0$, and for estimating θ under loss $\frac{(\delta-\theta)^2}{\theta}$, dominating estimators $\delta_\psi(X)$ of $\delta_0(X) = X$ are given by Theorem 3.4 with $\ell = 0$ and*

$$\psi_1(t) \;=\; \frac{\theta_0^{t+1}/t!}{\sum_{0 \leq j \leq t} \theta_0^j / j!}\,.$$

Moreover, these dominating estimators are minimax.

Proof. The dominating estimators $\delta_\psi(X)$ here are minimax since $\delta_0(X)$ is minimax as a consequence of Theorem 2.2. The dominance follows from Theorem 3.4, and by the calculation (e.g., Gurmu & Trivedi, 1992; Remark 3.6)

$$\mathbb{E}_\theta(\theta - X | X \leq t) \;=\; \frac{\theta^{t+1}/t!}{\sum_{0 \leq j \leq t} \theta^j / j!}\,, \tag{3.13}$$

which is clearly increasing in $\theta \in [\theta_0, \infty)$ for all $t \in \mathbb{N}$. This implies that the infimum value for the determination of $\psi_1(t)$, $t \in \mathbb{N}$, is attained at $\theta = \theta_0$, and establishes the result. □

Now, for the negative binomial case, we will make use of the following one-parameter exponential family connection between the determination of $\psi_1(t)$ and the comparison of the conditional variance $\mathrm{Var}(X|X \leq t)$ and the unconditional variance $\mathrm{Var}(X)$.

Lemma 3.6. *For exponential family model (1.1), an estimator $\delta_0(X)$ of $\theta = \mathbb{E}(X)$ such that $t - \delta_0(t)$ is nondecreasing in $t \in \mathbb{N}$, we have for t belonging to the support of X:*

$$\psi_1(t) \;=\; \inf_{\theta \geq \theta_0} \mathbb{E}_\theta\left(\theta - \delta_0(X)\,|X \leq t\,\right) \;=\; \mathbb{E}_{\theta_0}\left(\theta_0 - \delta_0(X)\,|X \leq t\,\right)\,, \tag{3.14}$$

as long as $Var_\theta(X|X \leq t) \;\leq\; Var_\theta(X)$ for all $\theta \geq \theta_0$.

Proof. The truncated distribution $X|X \leq t$ has exponential family density as in (1.1) with

$$p_\eta(x|X \leq t) \;=\; e^{\eta x - A_t(\eta)}\,, \;\text{ and } A_t(\eta) \;=\; A(\eta) \;+\; \log \mathbb{P}_\eta(X \leq t)\,.$$

From this, we have

$$\mathbb{E}_\eta(X|X \le t) \;=\; A'(\eta) \;+\; \frac{d}{d\eta}\,\log \mathbb{P}_\eta(X \le t) \;\text{ and }\; \mathrm{Var}_\eta(X|X \le t) \;=\; \mathrm{Var}(X) \;+\; \frac{d^2}{d^2\eta}\,\log \mathbb{P}_\eta(X \le t)\,. \tag{3.15}$$

From the above, recalling that $\theta = A'(\eta)$, we obtain

$$\begin{aligned}
\frac{d}{d\eta}\,\mathbb{E}_\eta(\theta - \delta_0(X)|X \le t) \;&=\; \frac{d}{d\eta}\,\mathbb{E}_\eta(\theta - X|X \le t) \;+\; \frac{d}{d\eta}\mathbb{E}_\eta(X - \delta_0(X)|X \le t) \\
&=\; \mathrm{Var}_\eta(X) \;-\; \mathrm{Var}_\eta(X|X \le t) \;+\; \frac{d}{d\eta}\mathbb{E}_\eta(X - \delta_0(X)|X \le t) \\
&\ge\; 0\,,
\end{aligned}$$

by the assumption on the conditional and unconditional variances, and since the monotonicity assumption for $t - \delta_0(t)$, paired with an increasing monotone likelihood ordering in X for the distributions $X|X \le t$ and parameter η, imply that $\frac{d}{d\eta}\mathbb{E}_\eta(X - \delta_0(X)|X \le t) \ge 0$. Finally, the result follows since θ is a monotone increasing function of η. □

Remark 3.4. *It is of interest to point out, as can be inferred from the above development, that the probability $\mathbb{P}_\eta(X \le t)$ is logconcave in $\eta \in (\underline{\eta}, \bar{\eta})$ if and only if $Var_\eta(X) \;\ge\; Var_\eta(X|X \le t)$ for $\eta \in (\underline{\eta}, \bar{\eta})$.*

Remark 3.5. *The ordering required in Lemma 3.6 between the truncated and untruncated variances is quite plausible as a property, but it is not guaranteed without some conditions on the underlying distribution, and such issues have been studied mainly for continuous models by Mailhot (1987), Burdett (1996), Chen et al. (2010), and Chen (2013), among others. As can be inferred from the above, such an ordering does hold for both the Poisson and negative binomial distributions.*

Remark 3.6. *Expression (3.13), which relates to the expectation of a truncated Poisson distribution, can be derived from (3.15), with $\theta = e^\eta$ as follows:*

$$\begin{aligned}
\mathbb{E}_\theta(\theta - X|X \le t) \;&=\; -\frac{d}{d\eta}\,\log \mathbb{P}_\eta(X \le t) \\
&=\; -\frac{d}{d\eta}\,\log \sum_{0 \le j \le t} \frac{e^{-e^\eta}\,e^{\eta j}}{j!} \\
&=\; \frac{1/t!}{\sum_{0 \le j \le t} e^{\eta(j-t-1)}/j!}\,,
\end{aligned}$$

which is (3.13) indeed.

To conclude, here is an application of our findings (Corollary 2.1, Theorem 3.4, and Lemma 3.6) to negative binomial models. The minimax implication is presented for loss $\frac{r(\delta-\theta)^2}{\theta(\theta+r)}$, but applies as well to loss (1.2) and Corollary 2.1's variants.

Corollary 3.5. *For $X \sim NB(r, \theta)$ with probability mass function as in (2.4) and with $\theta \ge \theta_0 > 0$, and for estimating θ under loss $\frac{r(\delta-\theta)^2}{\theta(\theta+r)}$, estimators $\delta_\psi(X)$ given by Theorem 3.4 with $\ell = \theta_0/(r+1)$ and*

$$\psi_1(t) \;=\; \frac{\theta_0}{r+1} \;+\; \frac{\frac{r+\theta_0}{r+1}\,\frac{(r)_{t+1}}{t!}\,\left(\frac{\theta_0}{\theta_0+r}\right)^{t+1}}{\sum_{0 \le j \le t} \frac{(r)_j}{j!}\,\left(\frac{\theta_0}{\theta_0+r}\right)^j}\,, \tag{3.16}$$

dominate $\delta_0(X) = rX/(r+1)$ and are minimax. Namely $\delta_{\psi^*}(X) = \delta_0(X) + \psi^*(X)$ dominates $\delta_0(X)$ with $\psi^*(t) \;=\; \psi_1(t) - \frac{\theta_0}{r+1}$.

Proof. As in Corollary 3.4, the dominating estimators $\delta_\psi(X)$ here are minimax since $\delta_0(X)$ is minimax as a consequence of Corollary 2.1 (see Example 2.6, iv). The dominance of the above $\delta'_\psi s$ follows from Theorem 3.4, while the determination of ℓ and representation of $\delta_{\psi^*}(X)$ follow as $\lim_{t\to\infty}\psi_1(t) = \frac{\theta_0}{r+1}$ from (3.16).

There remains to show that $\psi_1(t)$ is indeed given by (3.16). Shonkmiler (2016) gives expressions for the truncated mean and variance of a negative binomial distribution. Namely, denoting $\sigma^2 = \mathrm{Var}(X) = \theta(1+\frac{\theta}{r})$, he obtains the following expressions for $\theta_t = \mathbb{E}(X|X \le t)$ and $\sigma_t^2 = \mathrm{Var}(X|X \le t)$, $t \in \mathbb{N}_+$:

$$\begin{aligned} \theta_t &= \theta - \frac{(r+\theta)\,(t+1)\,p_\theta(t+1)}{r\mathbb{P}_\theta(X \le t)} \\ \text{and } \sigma_t^2 &= \theta_t + t(\theta_t - \theta) + \theta_t\,\theta\,(1+\frac{1}{r}) - \theta_t^2\,. \end{aligned} \tag{3.17}$$

From this, it follows that

$$\sigma^2 - \sigma_t^2 = (\theta - \theta_t)\,(1 + \frac{\theta}{r} + t - \theta_t) > 0\,,$$

as $\theta_t < \min\{\theta, t\}$ for $t \in \mathbb{N}_+$. Hence, Lemma 3.6 applies with $t - \delta_0(t) = \frac{t}{r+1}$ nondecreasing in $t \in \mathbb{N}$ and implies that

$$\psi_1(t) = \mathbb{E}_{\theta_0}(\theta_0 - \frac{rX}{r+1}|X \le t) = \theta_0 - \frac{r}{r+1}\,\theta_t\,,$$

which from (3.17) yields (3.16) and completes the proof. □

4. Concluding remarks

We have provided a unified treatment of minimaxity for one-parameter exponential family subject to a lower-bound restriction on the expectation with several instances where the minimax property persists from the unrestricted case to the restricted parameter case. Our findings recovers previous known cases, but also leads to new results, namely for negative binomial models and for a large variety of weighted squared error loss functions. In contrast, we have elaborated on a plentiful number of non-minimax results and provided illustrations. In the latter part of the work, we have focussed on determining improved estimators in discrete models under squared error loss in presence of a lower-bound constrained expectation, which lead to classes of minimax estimators for estimating lower-bounded Poisson or negative binomial expectations under normalized squared error loss.

An open question in relationship to the class of minimax estimators for a lower-bounded Poisson mean given in Corollary 3.4 is the determination of a Bayesian solution. In fact, to the best of our knowledge, no such solution has been found. A plausible candidate is the Bayes estimator $\delta_\pi(X)$ with respect to the *uniform* prior $\pi_0(\theta) = \mathbb{I}_{[\theta_0,\infty)}(\theta)$; which is the truncation of $\pi(\theta) = \mathbb{I}_{(0,\infty)}(\theta)$ for which $\delta_\pi(X) = X$ is minimax. However, it it is known (i.e., van Eeden, 2006, pp. 58-59, reporting on a personal communication of Bill Strawderman) that $\delta_{\pi_0}(X)$ is never minimax for any $\theta_0 > 0$. In a related problem with interesting techniques and findings that may be helpful in addressing the open question described above as well as related ones, Hamura & Kubokawa (2019) consider $X \sim \mathrm{Poisson}(\theta)$, $\theta \le \theta_0$, and loss $L(\theta,\delta) = \theta\{\delta/\theta - \log(\delta/\theta) - 1\}$. They further consider the class of prior densities $\pi(\theta) = \theta^{\beta-1}$ and $\pi_{\theta_0}(\theta) = \theta^{\beta-1}\,\mathbb{I}_{(0,\theta_0)}(\theta)$, with $\beta > 0$ and including the Jeffreys

prior for $\beta = 1/2$ in the former case, and the associated Bayes estimators $\hat{\theta}_\beta(X)$ and $\hat{\theta}_{\beta,\theta_0}(X)$, showing that $\hat{\theta}_{\beta,\theta_0}(X)$ dominates $\hat{\theta}_\beta(X)$ for sufficiently small θ_0.

Appendix

PROOF OF LEMMA 2.1. Since $\lim_{\theta\to\infty} \lambda(\theta) = \lambda$ by assumption, for any $\delta \in (0,1)$ there exists $\tilde{\theta} > \max\{0, \theta_0\}$ such that

$$\lambda\,(1-\delta) \;\le \lambda(\theta) \;\le \lambda\,(1+\delta)\,, \text{ for all } \theta \ge \tilde{\theta}\,. \tag{4.18}$$

From (2.8), it follows for $\theta \ge \tilde{\theta}$ that:

$$\begin{aligned}
b'(\theta) \;-\; \frac{b(\theta)}{\theta}\,\frac{\lambda}{\lambda(\theta)} \;+\; \frac{1+\lambda}{2}\,\frac{b^2(\theta)}{\theta^2\lambda(\theta)} \quad &\le \quad -\frac{\epsilon}{2\,(1+\lambda)}\left(\frac{\lambda(\theta)+\lambda^2}{\lambda(\theta)}\right) \\
\Longleftrightarrow b'(\theta) \;+\; \frac{1+\lambda}{2\lambda(\theta)}\left(\frac{b(\theta)}{\theta} - \frac{\lambda}{1+\lambda}\right)^2 \quad &\le \quad -\frac{\epsilon}{2\,(1+\lambda)}\left(\frac{\lambda(\theta)+\lambda^2}{\lambda(\theta)}\right) + \frac{\lambda^2}{2\,\lambda(\theta)(1+\lambda)} \\
\Longrightarrow b'(\theta) \quad &\le \quad \frac{1}{2\,\lambda(\theta)(1+\lambda)}\left\{\lambda^2(1-\epsilon) - \epsilon\lambda(\theta)\right\} \\
\Longrightarrow b'(\theta) \quad &\le \quad \frac{1}{2(1+\lambda)}\left\{\frac{\lambda(1-\epsilon)}{1-\delta} \;-\; \epsilon\right\} = K' > 0\,,
\end{aligned} \tag{4.19}$$

for $\epsilon < \frac{\lambda}{\lambda+1-\delta}$. Recalling that $\tilde{\theta} > 0$ and that $b(\theta) > 0$ for all $\theta > \tilde{\theta}$, an integration of both sides of inequality $b'(t) < K'$ on $t \in (\tilde{\theta}, \theta)$ establishes that $b(\theta) \le K'(\theta - \tilde{\theta}) + b(\tilde{\theta}) \le K'\theta + b(\tilde{\theta}) \le K\theta$ with $K = K' + b(\tilde{\theta})/\tilde{\theta}$, establishing the result. □

PROOF OF LEMMA 2.2. Since $b(\theta) > 0$ for all $\theta \ge \theta_0$, expression (2.8) along with (4.18) and Lemma 2.1 imply for sufficiently large θ, say $\theta \ge \theta_2$

$$\begin{aligned}
\frac{b'(\theta)}{b(\theta)} \quad &\le \quad \frac{\lambda}{\theta\lambda(\theta)} \;-\; \frac{\epsilon}{2b(\theta)(1+\lambda)}\left\{1 + \frac{\lambda^2}{\lambda(\theta)}\right\} \\
&\le \quad \frac{1}{\theta}\left\{1 + \delta - \frac{\epsilon}{2(1+\lambda)K}\left\{1 + \lambda/(1+\delta)\right\}\right\} \\
&\le \quad \frac{1}{\theta}(1\,-\epsilon')\,,
\end{aligned} \tag{4.20}$$

for some $\epsilon' > 0$ since δ may be chosen arbitrarily small and ϵ is fixed here and positive. Hence, integrating both sides of the inequality $\frac{b'(t)}{b(t)} \le \frac{1-\epsilon'}{t}$ for $t \in (\theta_2, \theta)$, we infer that $\log b(\theta) \;\le\; \log(c\,\theta^{1-\epsilon'})$, for some $c > 0$ and θ sufficiently large. Finally, this implies that $\frac{b(\theta)}{\theta} \le \frac{c}{\theta^{\epsilon'}}$ which goes to 0 as θ goes to $+\infty$. □

PROOF OF LEMMA 2.3. It follows from (4.20) and the last line of the proof of Lemma 2.2, that

$$b'(\theta) \le \frac{(1-\epsilon')\,b(\theta)}{\theta} \;\le \frac{(1-\epsilon')c}{\theta^{\epsilon'}} \to 0 \text{ as } \theta \to \infty. \tag{4.21}$$

Hence, for any $h > 0$, $b'(\theta) < h$ for all large θ. In addition, if $b'(\theta) < -h$ for all large θ, $b(\theta') < 0$ for many θ' and the contradiction in the proof of Karlin's theorem would arise. Therefore, for every $h_n > 0$, there exists a θ_n such that $-h_n < b'(\theta_n) < h_n$. By choosing $h_n \to 0$, it follows that $b'(\theta_n) \to 0$ as $\theta_n \to \infty$. □

Acknowledgements

We are quite thankful for the useful comments and suggestions of two reviewers. Éric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by a grant from the Simons Foundation (#418098).

References

[1] Burdett, K. (1999). Truncated means and variances. *Economics Letters*, **52**, 263-267.

[2] Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Lecture Notes-Monograph Series, Vol. 9, Institute of Mathematical Statistics.

[3] Casella, G. & Strawderman, W.E. (1981). Estimating a bounded normal mean. *Annals of Statistics*, **9**, 870–878.

[4] Chen, J., van Eeden, C. & Zidek, J. V. (2010). Uncertainty and the conditional variance. Statistics & Probability Letters, **80**, 1764-1770.

[5] Chen, J. (2013). A partial order on uncertainty and information. *Journal of Theoretical Probability*, **26**, 349-359.

[6] Gurmu, S. & Trivedi, P.K. (1992). Overdispersion tests for truncated Poisson regression models. *Journal of Econometrics*, **54**, 347-370.

[7] Hamura, Y. & Kubokawa, T. (2019). Bayesian predictive distributions for a Poisson model with a parametric restriction. *Communications in Statistics: Theory and Methods*, **49**, 3257–3266.

[8] Johnstone, I.M. & MacGibbon, K. B. (1992). Minimax estimation of a constrained Poisson vector. *Annals of Statistics*, **20**, 807–831.

[9] Karlin, S. (1958). Admissibility for estimation with quadratic loss. *Annals of Mathematical Statistics*, **29**, 406–436.

[10] Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces, *Annals of Mathematical Statistics*, **32**, 136-142.

[11] Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: a review. *Communications in Statistics: Theory and Methods*, **28**, 613-650.

[12] Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics*, **22**, 290-299.

[13] Lehmann, E.L. & Casella, G. (1998). *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York.

[14] Mailhot, L. (1987). Ordre de dispersion et lois tronquées. *Cahiers de recherche de l'Académie des sciences de Paris, Sér. I Math.* **304**, 499-501.

[15] Marchand, É. & MacGibbon, K.B. (2000). Minimax estimation of a constrained Binomial proportion. *Statistics & Decisions*, **18**, 129-167.

[16] Marchand, É. & Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.

[17] Marchand, É. & William E. Strawderman (2005). On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint. *Journal of Statistical Planning and Inference*, **134**, 90-101.

[18] Marchand, É. & Strawderman, W.E. (2004). Estimation in Restricted Parameter Spaces: A Review, *Festschrift for Herman Rubin*. Institute of Mathematical Statistics Lecture Notes-Monograph Series, pp. 21-44.

[19] Shonkwiler, J.S. (2016). Variance of the truncated negative binomial distribution. *Journal of Econometrics*, **195**, 209-210.

[20] van Eeden, C. (2006). *Restricted parameter space problems: Admissibility and minimaxity properties*. Lecture Notes in Statistics, **188**, Springer.

[21] van Eeden, C. (1995). Minimax estimation of a lower-bounded scale parameter of a gamma distribution for scale-invariant squared-error loss. *Canadian Journal of Statistics*, **23**, 245–256.