On predictive density estimation for Gamma models with parametric constraints

AZIZ L'MOUDDEN[a], ÉRIC MARCHAND[a], OTHMANE KORTBI[b], WILLIAM E. STRAWDERMAN[c]

a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mails: aziz.lmoudden@usherbrooke.ca; eric.marchand@usherbrooke.ca)

b Department of Statistics, CBE, UAE University (e-mail: okortbi@uaeu.ac.ae)

c Rutgers University, Department of Statistics and Biostatistics, 501 Hill Center, Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

SUMMARY

This paper is concerned with prediction for Gamma models, and more specifically the estimation of a predictive density for $Y \sim \text{Ga}(\alpha_2, \beta)$ under Kullback-Leibler loss, based on $X \sim \text{Ga}(\alpha_1, \beta)$. The main focus pertains to situations where there is a parametric constraint of the form $\beta \in C = (a, b)$. We obtain representations for Bayes predictive densities and the minimum risk equivariant predictive density in the unconstrained problem. It is shown that the generalized Bayes estimator against the truncation of the non-informative prior onto C dominates the minimum risk equivariant predictive density and is minimax whenever $a = 0$ or $b = \infty$. Analytical comparisons of plug-in predictive densities $\text{Ga}(\alpha_2, \hat{\beta})$, which include the predictive mle density, are obtained with results applying as well for point estimation under dual entropy loss $\frac{\beta}{\hat{\beta}} - \log(\frac{\beta}{\hat{\beta}}) - 1$. Numerical evaluations confirm that such predictive densities are much less efficient than some Bayesian alternatives in exploiting the parametric restriction. Finally, it is shown that variance expansion improvements of the form $\text{Ga}(\frac{\alpha_2}{k}, k\hat{\beta})$ of plug-in predictive densities can always be found for a subset of $k > 1$ values and non-degenerate $\hat{\beta}$.

AMS 2010 subject classifications: 62C20, 62C86, 62F10, 62F15, 62F30

Keywords and phrases: Bayes estimator; Dominance; Frequentist risk; Gamma; Kullback-Leibler loss; Minimax; Minimum risk equivariant; Plug-in; Predictive density; Restricted parameter space; Variance expansion.

1 Introduction

Prediction of randomly generated quantities is a central theme in statistics. The richest prediction takes the form of a predictive density over the domain of possible values. Bayesian predictive densities are optimal in response to a given prior and loss function.

Nonnegative random variables arise in a wide variety of applications, such as life-lengths of components or biological organisms typical of reliability and survival analysis, magnitudes related to physical objects (e.g., plant diameters or heights, wind speeds, measures of precipitation), and measurable quantities in economics (e.g., incomes, prices).

For parametric models of positive-valued random variables, it is often the case that bounds on the parameters can be stipulated, especially when these bounds relate to the practical context in which they arise. The perhaps most straightforward way to visualize this is through restrictions on quantiles or means which translate to parametric restrictions.

In this paper, we focus on Gamma models which play an important role for nonnegative data. Situations where the distribution of a summary statistic relates to a Gamma distribution are plentiful. They include sum of squares statistics in normal linear models and benchmark estimators of variance components. Other occurrences originate with Poisson processes and queueing theory. Furthermore, Gamma distributed summary statistics also emanate when sampling from various models such as Exponential, Weibull, Rayleigh, Pareto, Lognormal, Inverse Gaussian, etc. (see for instance so-called transformed chi-square distributions as presented by Rahman and Gupta, 1993). Lower bound restrictions arise for instance in variance estimation of random effects, and in particular for a heteroscedastic random effects setting of meta-analysis recently analyzed by Rukhin (2016).

In all of the above cases, it is desirable to have methods for obtaining efficient predictive densities. We thus study the estimation of predictive densities with restrictions on the scale parameter for independently Gamma distributed

$$X|\,\beta \sim \text{Ga}(\alpha_1, \beta),\, Y|\,\beta \sim \text{Ga}(\alpha_2, \beta); \tag{1}$$

where α_1 and α_2 are known, not necessarily equal, and β is unknown but restricted to $C = (a, b)$. The set-up in (1) includes i.i.d. samples from X by virtue of sufficiency. For a future i.i.d. sample $Y_1, \ldots, Y_m$; $m > 1$; from Y, our set-up does not include the prediction of $Y_1, \ldots, Y_m$ but it does include the prediction of $\sum_i Y_i$ which is $\text{Ga}(m\alpha_2, \beta)$ distributed.
Our analysis will address: **(i)** the lower bounded case (i.e., $a > 0$ and $b = \infty$), **(ii)** the upper bounded case (i.e., $a = 0$ and $b < \infty$), and **(iii)** the doubly-bounded case with $a > 0$ and $b < \infty$. For predictive analysis purposes, researchers are interested in specifying a predictive density $\hat{q}\,(\cdot\,; x)$ as an estimate of the density $q\,(\cdot\,|\beta\,)$ of Y. In turn, such a density may play a surrogate role for generating either future or missing values of Y. To evaluate the performance of such predictive densities, we resort to the familiar Kullback-Leibler loss

$$L_{KL}\,(\beta, \hat{q}\,(\cdot\,; x)) = \int_{R_+} \log\left(\frac{q\,(y\,|\beta\,)}{\hat{q}\,(y; x)}\right) q\,(y|\beta)\; dy, \tag{2}$$

with corresponding frequentist risk given by

$$R_{KL}(\beta, \hat{q}) \;=\; E^X\left[L_{KL}\,(\beta, \hat{q}\,(\cdot\,; X))\right]. \tag{3}$$

The main findings and contributions of this paper consist of:

- **(A)** A general dominance result (Theorem 3.1) establishing that the Bayes predictive density $\hat{q}_{\pi_C}(\cdot; X)$ generated from the prior $\pi_C(\beta) = \frac{1}{\beta}\,\mathbb{I}_C(\beta)$ (in other words the truncation of the non-informative prior $\pi_0(\beta) = \frac{1}{\beta}\,\mathbb{I}_{\mathbb{R}_+}(\beta)$ onto the restricted parameter space $C = (a, b)$); dominates the minimum risk equivariant predictive density estimator $\hat{q}_{\pi_0}(\cdot; X)$, which is also Bayes with respect to π_0. The approach is unified for all $C = (a, b)$ and exploits results in **(B)** and a star-shaped property of Gamma distributions. For the doubly-bounded case, the result is new, while for the lower bounded or upper bound case, the result duplicates in a much different manner that of Kubokawa et al. (2013). Moreover, for doubly-bounded parameter spaces $C = (a, b)$, we show that the Bayesian predictive estimator $\hat{q}_{\pi_0,C}(\cdot; X)$ dominates $\hat{q}_{\pi_0,C'}(\cdot; X)$ for both $C' = (a, \infty)$ and $C' = (0, b)$.

- **(B)** Various representation and properties of Bayesian predictive densities (Section 2) which are required for **(A)**, and which are of interest on their own.

(C) An explicit form for the minimax risk for both the lower bounded ($\beta \geq a$) and the upper bounded ($\beta \leq b$) cases (Section 3.2).

(D) An analysis (Section 4) of the risk performance under the parametric restriction $\beta \in (a,b)$ of plug-in predictive density estimators; which are of $\text{Ga}(\alpha_2, \hat{\beta}(X))$ densities for an estimator $\hat{\beta}(X)$ of β; associated with truncated linear estimators, and with corresponding point estimation results for the dual entropy loss $L_0(\beta,\hat{\beta}) = \left(\frac{\beta}{\hat{\beta}} - \log(\frac{\beta}{\hat{\beta}}) - 1\right)$.

(E) A general result in Section 5 which gives a universal and unified class of variance expansion improvements on plug-in predictive density estimators. These improvements are obtained as $\text{Ga}(\frac{\alpha_2}{k}, k\,\hat{\beta}(X))$ densities with $1 < k < k_0$, where k_0 depends only on α_2 and the infimum entropy risk of the plug-in estimator $\hat{\beta}(X)$. These predictive densities are constructed in order to be equivalent in terms of the corresponding expectations (i.e., constant in k), but with variability increasing in k. Variance expansion have surfaced recently in location models as a method for obtaining improvements on plug-in predictive density estimators (e.g., Fourdrinier et al. 2011; Kubokawa, Marchand and Strawderman, 2015A, 2015B).

These above developments are further introduced throughout and constitute valuable additions to both the predictive density estimation body of work and inference problems under parametric restrictions. The former area of research has been quite active in recent years with key contributions in particular for normal and Poisson models without parametric restrictions, as witnessed by the work of Boisbunon and Maruyama (2014); Kato (2009); Komaki (2006, 2004, 2001); George, Liang and Xu (2006); Brown, George and Xu (2008); Liang and Barron (2004); among others. We also refer to Mukherjee and Johnstone (2015), as well as Mukherjee (2013), for studies of predictive densities with sparsity constraints on the parameters. Predictive density estimation with parametric restrictions has been studied by Fourdrinier et al. (2011) for normal models, as well as by Kubokawa et al. (2013) for location and scale families of distributions, while van Eeden (2006), as well as Marchand and Strawderman (2004) review estimation in restricted parameter spaces.

2 Bayes predictive densities

We begin here with a general representation for Bayesian predictive density estimators $\hat{q}_\pi(\cdot;X)$ where π is a prior for β, and where a key role is played by the predictive density estimator $\hat{q}_{\pi_0}(\cdot;X)$, with $\pi_0(\beta) = \frac{1}{\beta}\mathbb{I}_{(0,\infty)}(\beta)$ being the usual non-informative prior for β.

Theorem 2.1. *Consider model (1), the noninformative prior $\pi_0(\beta) = \frac{1}{\beta}\mathbb{I}_{(0,\infty)}(\beta)$, and an arbitrary generalized prior $\pi(\beta)$ (with respect to σ-finite measure ν) for which the posterior $\pi(\beta|x)$ exists. Then, we have for the predictive densities $\hat{q}_{\pi_0}(\cdot;x)$ and $\hat{q}_\pi(\cdot;x)$*

(a) $\hat{q}_{\pi_0}(y;x) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\frac{1}{x}\left(\frac{y}{x}\right)^{\alpha_2-1}\left(1+\frac{y}{x}\right)^{-(\alpha_1+\alpha_2)}\mathbb{I}_{(0,\infty)}(y)$;

(b) $\hat{q}_\pi(y;x) = \hat{q}_{\pi_0}(y;x)\,\frac{(y+x)}{x}\,\frac{m_\pi(y+x|\,\alpha_1+\alpha_2)}{m_\pi(x|\,\alpha_1)}$;

where $m_\pi(z|\alpha) = \int_0^\infty \frac{z^{\alpha-1}e^{-z/\beta}}{\Gamma(\alpha)\,\beta^\alpha}\,\pi(\beta)\,d\nu(\beta)$ is the marginal density of $Z \sim \text{Ga}(\alpha,\beta)$ associated with prior π.

Proof. Bayes predictive density estimators under Kullback-Leibler are posterior expectations of the model density (e.g., Aitchison, 1975). For **(b)**, we thus have

$$\begin{aligned}
\hat{q}_\pi\,(y;x) &= \int_0^\infty q(y|\beta)\,\pi(\beta|x)\,d\nu(\beta) &(4)\\
&= \frac{\int_0^\infty q\,(y\,|\beta)\,p\,(x\,|\beta)\,\pi\,(\beta)\;d\nu(\beta)}{\int_0^\infty p\,(x\,|\beta)\,\pi\,(\beta)\;d\nu(\beta)}\\
&= \frac{\Gamma\,(\alpha_1+\alpha_2)}{\Gamma\,(\alpha_1)\,\Gamma\,(\alpha_2)}\,\frac{y^{\alpha_2-1}\,x^{\alpha_1}}{(y+x)^{\alpha_1+\alpha_2}}\,\left(\frac{y+x}{x}\right)\,\frac{\int_0^\infty \frac{(y+x)^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1+\alpha_2)}\,\exp\left(-\frac{y+x}{\beta}\right)\pi\,(\beta)\;d\nu(\beta)}{\int_0^\infty \frac{x^{\alpha_1-1}}{\beta^{\alpha_1}\Gamma(\alpha_1)}\,\exp\left(-\frac{x}{\beta}\right)\pi\,(\beta)\;d\nu(\beta)}\\
&= \hat{q}_{\pi_0}\,(y\,;x)\,\left(\frac{y+x}{x}\right)\,\frac{m_\pi\,(y+x\,|\,\alpha_1+\alpha_2\,)}{m_\pi\,(x\,|\,\alpha_1\,)}\,. &(5)
\end{aligned}$$

Part **(a)** follows from (5) with the evaluation $m_{\pi_0}(z|\alpha) = z^{-1}$ obtained from its definition and the Gamma function identity

$$\int_0^\infty z^{-(c+1)}\,\exp\left(-\frac{d}{z}\right)\,dz = \Gamma\,(c)\,d^{-c}, \qquad \forall c,d>0\,. \qquad \square \tag{6}$$

Part **(a)** was obtained by Aitchison (1975). Part **(b)** of Theorem 2.1, which represents the Bayes predictive density $\hat{q}_\pi$ as a weighted version of $\hat{q}_{\pi_0}$ is general with respect to the choice of prior, will be illustrated and exploited below. The representation in **(b)**, with the appearance of the marginal distributions of X and $X+Y$, is quite reminiscent of representations given by George, Liang and Xu (2006), as well as Komaki (2006), for normal and Poisson observables respectively. The density $\hat{q}_{\pi_0}(\cdot\,;x)$ is the density of a Beta2$(c=\alpha_2, d=\alpha_1, \sigma=x)$ distribution of type II, with shape parameters c and d, scale parameter σ, defined as follows (see the Appendix for further interpretation).

Definition 2.1. *We will say that $U \sim Beta2(c,d,\sigma)$, with $c,d,\sigma>0$, whenever U has density on $\mathbb{R}_+$ given by*

$$f(u) \;=\; \frac{\Gamma(c+d)}{\Gamma(c)\,\Gamma(d)}\,\frac{1}{\sigma}\,\frac{(u/\sigma)^{c-1}}{(1+u/\sigma)^{c+d}}\,. \tag{7}$$

Equivalently, such Beta2 distributions are also known as Fisher distributions. They are related to Beta (type I) distributions as follows : $\frac{U}{\sigma} \sim Beta2(c,d,1) =^d \frac{V}{1-V}$ where $V \sim Beta(c,d)$ with density $\frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} v^{c-1}(1-v)^{d-1}\,\mathbb{I}_{(0,1)}(v)$.

Here are some instructive examples of Theorem 2.1.

Example 2.1. *(Conjugate inverse Gamma priors)*
For inverse Gamma priors $\pi(\beta) \propto \beta^{-\gamma_1} e^{-\gamma_2/\beta}$ with $\gamma_1 > 1-\alpha_1$ and $\gamma_2>0$, we have $\hat{q}_\pi(y;x) \sim Beta2(\alpha_2, \gamma_1+\alpha_1-1, x+\gamma_2)$. This was obtained by Amaral and Dunsmore (1980). It can be derived from (5) via the evaluation $m_\pi(z|\alpha) = \frac{\Gamma(\alpha+\gamma_1-1)}{\Gamma(\alpha)}\,\frac{z^{\alpha-1}}{(z+\gamma_2)^{\alpha+\gamma_1-1}}$, which makes use of (6) again.

Example 2.2. *(Priors truncated to $C=(a,b)$) With the parametric restriction $\beta \in C=(a,b)$, it seems natural to consider as priors the truncation onto C of the noninformative prior $\pi_0(\beta)=\frac{1}{\beta}$ and, more generally, truncations of inverse Gamma priors given by prior densities*

$$\pi_{(a,b),\gamma_1,\gamma_2}(\beta) \propto \beta^{-\gamma_1} e^{-\gamma_2/\beta}\,\mathbb{I}_{(a,b)}(\beta)\,, \tag{8}$$

with $\gamma_1 > 1-\alpha_1$ *and* $\gamma_2 \geq 0$. *Evaluating Theorem 2.1's marginals* m_π, *we have*

$$\begin{aligned} z\, m_\pi\,(z\,|\alpha\,) \;&\propto\; \frac{z^\alpha}{\Gamma(\alpha)} \int_a^b \beta^{-(\gamma_1+\alpha)}\, e^{-(z+\gamma_2)/\beta} d\beta \\ &\propto\; \frac{z^\alpha}{\Gamma(\alpha)} \int_{(z+\gamma_2)/b}^{(z+\gamma_2)/a} \frac{u^{\alpha+\gamma_1-2}\, e^{-u}}{(z+\gamma_2)^{\alpha+\gamma_1-1}}\, du. \\ &\propto\; \frac{\Gamma\,(\alpha+\gamma_1-1)\, z^\alpha}{\Gamma(\alpha)\,(z+\gamma_2)^{\alpha+\gamma_1-1}} \left(F_{\alpha+\gamma_1-1}\left(\frac{z+\gamma_2}{a}\right) - F_{\alpha+\gamma_1-1}\left(\frac{z+\gamma_2}{b}\right)\right), \end{aligned}$$

where $F_\gamma(\cdot)$ *is the* $Ga(\gamma,1)$ *cdf. Using the above in (5), the Bayes estimator can be expressed as*

$$\hat{q}_{\pi_{(a,b)},\gamma_1,\gamma_2}(y;x) \;=\; \hat{q}_{\pi,\gamma_1,\gamma_2}(y;x)\, \frac{F_{\alpha_1+\alpha_2+\gamma_1-1}(\frac{x+y+\gamma_2}{a}) \;-\; F_{\alpha_1+\alpha_2+\gamma_1-1}(\frac{x+y+\gamma_2}{b})}{F_{\alpha_1+\gamma_1-1}(\frac{x+\gamma_2}{a}) \;-\; F_{\alpha_1+\gamma_1-1}(\frac{x+\gamma_2}{b})}, \qquad (9)$$

with $\hat{q}_{\pi,\gamma_1,\gamma_2}(\cdot;x)$ *being Example 2.1's (unrestricted) Bayes* $Beta2(\alpha_2,\gamma_1+\alpha_1-1,x+\gamma_2)$ *predictive density. In the expression above, and throughout the paper, it shall be understood that* $F_\alpha(t/a)$ *and* $F_\alpha(t/b)$ *evaluated at* $a=0$ *or* $b=+\infty$ *are taken to be equal to* 1 *and* 0 *respectively. Expression (9) encompasses all parameter spaces of the form* (a,b) *and prior parameters* γ_1,γ_2. *The expression for the case* $a=0, b<\infty$, $\gamma_1=1,\gamma_2=0$ *was given by Kubokawa et al. (2013).*

3 Dominance results and minimax predictive densities

Let the parameter space be restricted to $C=(a,b)$. We focus in this section on the Bayesian predictive estimators $\hat{q}_{\pi_0}(\cdot;X)$ and $\hat{q}_{\pi_0,C}(\cdot;X)$ associated respectively with the noninformative prior $\pi_0(\beta)=\frac{1}{\beta}$ and its truncation to the parameter space $\pi_{0,C}(\beta)=\pi_0(\beta)\,\mathbb{I}_C(\beta)$. The former predictive estimator is given in Theorem 2.1, while the latter is given by $\hat{q}_{\pi,1,0}(y;x)$ in (9) as

$$\hat{q}_{\pi_{0,C}}\,(y\,|x\,) = \hat{q}_{\pi_0}\,(y\,|x\,)\; \frac{F_{\alpha_1+\alpha_2}\left(\frac{x+y}{a}\right) - F_{\alpha_1+\alpha_2}\left(\frac{x+y}{b}\right)}{F_{\alpha_1}\left(\frac{x}{a}\right) - F_{\alpha_1}\left(\frac{x}{b}\right)}, \qquad (10)$$

(as $\pi_{0,C}$ corresponds to (8) with $\gamma_1=1,\gamma_2=0$). We show below in Theorem 3.1 that $\hat{q}_{\pi_0,C}(\cdot;X)$ dominates $\hat{q}_{\pi_0}(\cdot;X)$ for all C under Kullback-Leibler risk R_{KL} given in (3). For both the upper bounded ($a=0,b<\infty$) and lower-bounded ($a>0,b=\infty$) cases, the dominance result was established by Kubokawa et al. (2013) using an adaptation of Kubokawa's IERD method (e.g., Kubokawa, 1994), but we provide a direct, alternative and unified in C route in establishing the dominance. With Kubokawa et al. (2013) having shown that $\hat{q}_{\pi_0}(\cdot\,;X)$ is minimax for such one-sided restrictions, dominating predictive density estimators such as $\hat{q}_{\pi_0,C}(\cdot\,;X)$ are also minimax. In the doubly-bounded case with $a>0,b<\infty$, our dominance finding is new, but neither $\hat{q}_{\pi_0}(\cdot\,;X)$ nor $\hat{q}_{\pi_0,C}(\cdot\,;X)$ is minimax. We also show that $\hat{q}_{\pi_0,C}(\cdot;X)$ dominates $\hat{q}_{\pi_0,C'}(\cdot\,;X)$ for doubly-bounded parameter spaces $C=(a,b)$ and $C'=(a,\infty)$ or $C'=(0,b)$. Finally, in Subsection 3.2, we provide further insight on minimax estimators and minimax risk.

3.1 Dominance results

We will require the next two lemmas. The first of these involves inequalities given and exploited by Misra and Arshad (2014), although we are stating these as strict inequalities. It is an immediate consequence of the star-shaped property of the family of Ga$(\alpha, 1)$ distributions, i.e., $\frac{F_\alpha^{-1}(F_{\alpha'}(x))}{x}$ is a strictly increasing function of $x \in (0,\infty)$ whenever $\alpha < \alpha'$, which in turn follows from the convexity of $F_\alpha^{-1} \circ F_{\alpha'}$ (van Zwet, 1964), for $\alpha < \alpha'$, and $F_\alpha^{-1}(F_{\alpha'}(0)) = 0$.

Lemma 3.1. *Let F_α be the cdf of a Ga$(\alpha,1)$ distribution. Then, for $0 < \alpha < \alpha'$, $c_1, c_2 \in (0,1)$, and for all $x > 0$, we have*

(i) $c_1 F_\alpha^{-1}(F_{\alpha'}(x)) > F_\alpha^{-1}(F_{\alpha'}(c_1 x))$;

(ii) $\frac{1}{c_2} F_\alpha^{-1}(F_{\alpha'}(x)) < F_\alpha^{-1}\left(F_{\alpha'}(\frac{x}{c_2})\right)$.

Here is our main dominance result.

Theorem 3.1. *Let $C = (a,b)$ be the restricted parameter space, either lower bounded with $b = +\infty$, upper bounded with $a = 0$, or doubly-bounded with $a > 0$ and $b < \infty$. Consider estimating the density of $Y \sim Ga(\alpha_2, \beta)$ under Kullback-Leibler loss (2) with $\beta \in C$, and based on $X \sim Ga(\alpha_1, \beta)$.*

(a) *Then, the predictive density estimator $\hat{q}_{\pi_0,C}(\cdot\,; X)$ dominates $\hat{q}_{\pi_0,C'}(\cdot\,; X)$ where $C \subset C' = (a', b')$ with either **(i)** $a' = a, b' = +\infty$; **(ii)** $a' = 0, b' = b$; or **(iii)** $C' = \mathbb{R}_+$,*

(b) *with the Kullback-Leibler risks being equal if and only if $C' = \mathbb{R}_+$ and $\beta = a(> 0), b = +\infty$ or $a = 0, \beta = b(< +\infty)$.*

Proof. Representation (10) allows us to express the difference of risks as

$$\begin{aligned}
\Delta_{KL}(\beta) &= R_{KL}\left(\beta, \hat{q}_{\pi_{0,C'}}\right) - R_{KL}\left(\beta, \hat{q}_{\pi_{0,C}}\right) \\
&= E_\beta^{X,Y}\left[\log \frac{\hat{q}_{\pi_{0,C}}(Y\,; X)}{\hat{q}_{\pi_{0,C'}}(Y\,; X)}\right] \\
&= E_\beta^{X,Y}\left[\log \frac{F_{\alpha_1+\alpha_2}\left(\frac{X+Y}{a}\right) - F_{\alpha_1+\alpha_2}\left(\frac{X+Y}{b}\right)}{F_{\alpha_1}\left(\frac{X}{a}\right) - F_{\alpha_1}\left(\frac{X}{b}\right)} - \log \frac{F_{\alpha_1+\alpha_2}\left(\frac{X+Y}{a'}\right) - F_{\alpha_1+\alpha_2}\left(\frac{X+Y}{b'}\right)}{F_{\alpha_1}\left(\frac{X}{a'}\right) - F_{\alpha_1}\left(\frac{X}{b'}\right)}\right] \\
&= E_1^{X,Y}\left[\log \frac{F_{\alpha_1+\alpha_2}\left(\beta\frac{X+Y}{a}\right) - F_{\alpha_1+\alpha_2}\left(\beta\frac{X+Y}{b}\right)}{F_{\alpha_1+\alpha_2}\left(\beta\frac{X+Y}{a'}\right) - F_{\alpha_1+\alpha_2}\left(\beta\frac{X+Y}{b'}\right)} - \log \frac{F_{\alpha_1}\left(\beta\frac{X}{a}\right) - F_{\alpha_1}\left(\beta\frac{X}{b}\right)}{F_{\alpha_1}\left(\beta\frac{X}{a'}\right) - F_{\alpha_1}\left(\beta\frac{X}{b'}\right)}\right] \\
&= \phi(\alpha_1 + \alpha_2, \beta) - \phi(\alpha_1, \beta),
\end{aligned}$$

with

$$\phi(\alpha, \beta) = E_\alpha^T\left[\log\left(\frac{F_\alpha\left(\frac{\beta}{a}T\right) - F_\alpha\left(\frac{\beta}{b}T\right)}{F_\alpha\left(\frac{\beta}{a'}T\right) - F_\alpha\left(\frac{\beta}{b'}T\right)}\right)\right],$$

the expectation E_α^T taken with respect to $T \sim Ga(\alpha, 1)$. For establishing the result in **(a)**, it will suffice to show, for cases **(i)**, **(ii)**, **(iii)**, that $\phi(\alpha', \beta) \geq \phi(\alpha, \beta)$ for all $\beta \in C = (a,b)$ and $\alpha' > \alpha$. We have by definition of $\phi(\alpha,\beta)$ and with the change of variable $x = F_{\alpha'}^{-1}(F_\alpha(t))$

$$\begin{aligned}
\phi(\alpha,\beta) &= \int_0^\infty \log\left(\frac{F_\alpha\left(\frac{\beta}{a}t\right) - F_\alpha\left(\frac{\beta}{b}t\right)}{F_\alpha\left(\frac{\beta}{a'}t\right) - F_\alpha\left(\frac{\beta}{b'}t\right)}\right) dF_\alpha(t) \qquad (11) \\
&= \int_0^\infty \log\left(\frac{F_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}(F_{\alpha'}(x))\right) - F_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}(F_{\alpha'}(x))\right)}{F_\alpha\left(\frac{\beta}{a'}F_\alpha^{-1}(F_{\alpha'}(x))\right) - F_\alpha\left(\frac{\beta}{b'}F_\alpha^{-1}(F_{\alpha'}(x))\right)}\right) dF_{\alpha'}(x)\,. \qquad (12)
\end{aligned}$$

We now separate the remainder of the proof into cases (**i**), (**ii**), and (**iii**).

(i) For $a' = a, b' = +\infty$, $\alpha' > \alpha$, we obtain from (12)

$$\begin{aligned}\phi(\alpha,\beta) &= \int_0^{\infty} \log\left(1 - \frac{F_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}\left(F_{\alpha'}(x)\right)\right)}{F_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}\left(F_{\alpha'}(x)\right)\right)}\right) dF_{\alpha'}(x) \\ &< \int_0^{\infty} \log(1 - \frac{F_{\alpha'}(\frac{\beta}{b}x)}{F_{\alpha'}(\frac{\beta}{a}x)})\, dF_{\alpha'}(x) \;=\; \phi(\alpha',\beta)\,,\end{aligned}$$

by making use of Lemma 3.1 with $c_2 = \frac{a}{\beta}$ and $c_1 = \frac{\beta}{b}$. This now completes the proof required in **(i)**.

(ii) Similarly, we have for $a' = 0$, $b = b'$, $\alpha' > \alpha$ via (12), with $\bar{F} \equiv 1 - F$,

$$\begin{aligned}\phi(\alpha,\beta) &= \int_0^{\infty} \log\left(\frac{\bar{F}_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}\left(F_{\alpha'}(x)\right)\right)}{\bar{F}_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}\left(F_{\alpha'}(x)\right)\right)} - 1\right) dF_{\alpha'}(x) \\ &< \int_0^{\infty} \log(\frac{\bar{F}_{\alpha'}(\frac{\beta}{a}x)}{\bar{F}_{\alpha'}(\frac{\beta}{b}x)} - 1)\, dF_{\alpha'}(x) \;=\; \phi(\alpha',\beta)\,,\end{aligned}$$

completing the proof for **(ii)**.

(iii) As above for $C' = \mathbb{R}_+$ and $\alpha' > \alpha$, the result follows from (12) as

$$\begin{aligned}\phi(\alpha,\beta) &= \int_0^{\infty} \log\left(F_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}\left(F_{\alpha'}(x)\right)\right) - F_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}\left(F_{\alpha'}(x)\right)\right)\right) dF_{\alpha'}(x) \\ &\le \int_0^{\infty} \log\left(F_{\alpha'}(\frac{\beta}{a}x) - F_{\alpha'}(\frac{\beta}{b}x)\right) dF_{\alpha'}(x) = \phi(\alpha',\beta)\,,\end{aligned}$$

with equality if and only if $\beta = a, b = +\infty$ or $a = 0, \beta = b$.

With this last observation and the strict inequalities in **(i)** and **(ii)**, part **(b)** is established and the proof is complete. □

Remark 3.1. *(A) The dominance findings of Theorem 3.1 are illustrated below in Section 6. The general theme concerns the plausible frequentist risk improvement when comparing the predictive density estimators $\hat{q}_{\pi_0,C}(\cdot;X)$ and $\hat{q}_{\pi_0,C'}(\cdot;X)$ for $\beta \in C$, and C a strict subset of C'. Others have come across and established similar results in the point estimation literature when C' corresponds to the unrestricted parameter space. For instance, Hartigan's result (Hartigan, 2003) applies for $p-$variate normal models, squared error loss and $C' = \mathbb{R}^p$. In this sense, Theorem 3.1's dominance result applied to $C = (a,b)$, and $C' = (a,\infty)$ or $C' = (0,b)$, is a departure on known restricted parameter space results.*

(B) Along the same theme, and as an example of non-dominance when one truncates the prior, we point out that $\hat{q}_{\pi_0,C}(\cdot;X)$ **does not** *dominate $\hat{q}_{\pi_0,C'}(\cdot;X)$ when $C = (a,\infty)$ and $C' = (a',\infty)$ with $a' < a$. Indeed, Theorem 3.1 tells us that $R_{KL}(a,\hat{q}_{\pi_0,C}) = R_{KL}(a,\hat{q}_{\pi_0}) > R_{KL}(a,\hat{q}_{\pi_0,C'})$. The same is true for $C = (0,b)$ and $C' = (0,b')$ with $b < b'$.*

3.2 Minimax estimators and minimax risk

The prior $\pi_0(\beta) = \frac{1}{\beta}\,\mathbb{I}_{(0,\infty)}(\beta)$ is not adapted to the constraint $\beta \in C = (a,b)$. And, we have just seen that its truncation $\pi_{0,C}$ is efficient in the sense that the predictive density estimator $\hat{q}_{\pi_0,C}(\cdot;X)$ dominates, under KL loss, $\hat{q}_{\pi_0}(\cdot;X)$ as an estimator of $q \sim \text{Ga}(\alpha_2,\beta)$ based on $X \sim \text{Ga}(\alpha_1,\beta)$. However, as shown by Kubokawa et al. (2013), $\hat{q}_{\pi_0}(\cdot;X)$ remains minimax whenever $a=0$ or $b=+\infty$. For such parametric restrictions, it thus provides a useful benchmark with dominating estimators, such as $\hat{q}_{\pi_0,C}$, being necessarily minimax, and with its constant risk yielding the minimax risk.
We evaluate this constant risk below and we first recall and establish some facts about the digamma function given by $\psi(t) = \frac{\Gamma'(t)}{\Gamma(t)} = \frac{d}{dt}\log\Gamma(t)$, where $\Gamma(t)$ is the Gamma function given by $\Gamma(t) = \int_0^\infty y^{t-1}\,e^{-y}\,dy$ and log is hereafter used in base e.

Lemma 3.2. **(a)** *We have* $\psi(\alpha) = E_\alpha[\log Z]$*, where* $Z \sim Ga(\alpha,1)$*.*

(b) *Also* $\psi(1) = -\gamma$*, the recursive property* $\psi(t+1) = \psi(t) + \frac{1}{t}$ *holds for* $t > 0$ *which is equivalent to* $\psi(t+1) = -\gamma + \sum_{k=1}^{t}\frac{1}{k}$ *for positive integer* t*, where* γ *is Euler's constant given by* $\gamma = \lim_{m\to\infty}(\sum_{k=1}^{m}\frac{1}{k} - \log m) \approx 0.577216$*.*

(c) *The function* $\psi(t) - \log t$ *increases in* t*, for* $t > 0$*.*

Proof. Part **(a)** is verified by differentiating $\Gamma(\alpha) = \int_{\mathbb{R}_+} t^{\alpha-1}\,e^{-t}\,dt$ under the integral sign, while the properties in **(b)** are well documented (see for instance Andrews, Askey and Roy, 1999). For **(c)**, the stronger result $\psi'(t) - \frac{1}{t} - \frac{1}{2t^2} > 0$, for all $t \in \mathbb{R}_+$ is given in Muldoon (1978). □

Corollary 3.1. *The Kullback-Leibler risk of* $\hat{q}_{\pi_0}(\cdot;X)$ *for estimating the density of* $Y \sim Ga(\alpha_2,\beta)$ *based on* $X \sim Ga(\alpha_1,\beta)$*, is constant and given by*

$$R_{KL}(\beta,\hat{q}_{\pi_0}) \;=\; (\alpha_1+\alpha_2)\psi(\alpha_1+\alpha_2) + \log\Gamma(\alpha_1) - \log\Gamma(\alpha_1+\alpha_2) - \alpha_2 - \alpha_1\psi(\alpha_1). \tag{13}$$

For the unrestricted parameter space $(0,\infty)$ *and for restricted parameter spaces of the form* $\beta \in (a,b)$ *with* $a=0$ *or* $b=+\infty$*, the above is equal to the minimax risk, which is moreover independent of the parameter space endpoints.*

Proof. The minimaxity follows from Kubokawa et al. (2013). For the evaluation of the risk, we have from (3) and Theorem 2.1

$$\begin{aligned}
R_{KL}(\beta,\hat{q}_{\pi_0}) &= E^{X,Y}\left[\log\left(\frac{q\,(Y\,|\beta)}{\hat{q}_{\pi_0}(Y;X)}\right)\right] \\
&= E^{X,Y}\left[\log\left(\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1+\alpha_2)}\right) + \alpha_2\log(\frac{X}{\beta}) - \frac{Y}{\beta} + (\alpha_1+\alpha_2)\,\{\log(X+Y) - \log X\}\right] \\
&= \log(\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1+\alpha_2)}) - \alpha_1\psi(\alpha_1) - \alpha_2 + (\alpha_1+\alpha_2)\psi(\alpha_1+\alpha_2)\,,
\end{aligned}$$

with the property $X+Y \sim \text{Ga}\,(\alpha_1+\alpha_2,\beta)$ and Lemma 3.2. □
We will make use of the above constant risk for comparisons later on (see Example 4.1). With evaluations of the digamma and the gamma functions, we obtain explicit values for integer α_1 and α_2. As an example, for $\alpha_1 = 2, \alpha_2 = 1$, we obtain $R_{KL}(\beta,\hat{q}_{\pi_0}) = \frac{3}{2} - \log 2 - \gamma \approx 0.22964$ by making use of Lemma 3.2.

4 Plug-in predictive density estimators

Plug-in density estimators are ubiquitous in statistical theory and practice. It may well be appealing to reduce the action space to plug-in density estimators. For our model, they are of the form $q_{\hat{\beta}}(\cdot; X) \sim \text{Ga}(\alpha_2, \hat{\beta}(X))$, where $\hat{\beta}(X)$ is a point estimator of β. Such estimators include the maximum likelihood predictive density estimator $\hat{q}_{mle} = q_{\hat{\beta}_{mle}} \sim \text{Ga}(\alpha_2, \hat{\beta}_{mle}(X))$, where $\hat{\beta}_{mle}(X)$ is the point estimator mle of β. As the next result describes, the Kullback-Leibler loss and frequentist risk incurred by plug-in density estimators are connected to a dual entropy loss and risk.

Lemma 4.1. *The Kullback-Leibler loss associated with the plug-in density estimate $q_{\hat{\beta}}(\cdot\,; x) \sim Ga(\alpha_2, \hat{\beta}(x))$ is given by the entropy loss*

$$L_0(\beta, \hat{\beta}) \,=\, \alpha_2 \left(\frac{\beta}{\hat{\beta}} \,-\, \log(\frac{\beta}{\hat{\beta}}) \,-\, 1 \right) \tag{14}$$

in the corresponding point estimation problem. Consequently, the frequentist Kullback-Leibler risk of the plug-in density estimator $q_{\hat{\beta}}(\cdot; X) \sim Ga(\alpha_2, \hat{\beta}(X))$ is equal to the frequentist risk of the point estimator $\hat{\beta}(X)$ of β under the above loss L_0.

Proof. We have by a direct evaluation

$$\begin{aligned} L_{KL}(\beta, q_{\hat{\beta}}(\cdot\,; x)) &= \int_{\mathbb{R}_+} \log(\frac{q(y|\beta)}{q_{\hat{\beta}}(y; x)})\, q(y|\beta)\, dy \\ &= \int_{\mathbb{R}_+} \left(y\,(\frac{1}{\hat{\beta}(x)} - \frac{1}{\beta}) + \alpha_2\,(\log(\frac{\hat{\beta}(x)}{\beta}) \right) q(y|\beta)\, dy \\ &= \alpha_2 \left(\frac{\beta}{\hat{\beta}(x)} \,-\, \log(\frac{\beta}{\hat{\beta}(x)}) - 1 \right), \end{aligned}$$

since $E_\beta(Y) = \alpha_2\, \beta$. □

Remark 4.1. *For the problem of estimating β under entropy loss L_0, equivariant estimators (under changes of scale) are multiples of X, i.e., of the form $\hat{\beta}_c(X) = cX$. For $\alpha_1 > 1$, such estimators have constant entropy risk, and the optimal or best equivariant estimator is easily shown to be given by $\hat{\beta}_{mre}(X) = \frac{X}{\alpha_1 - 1}$. This result is well known and it is also the case that the minimum risk equivariant (mre) estimator is minimax for the unrestricted parameter space, and Generalized Bayes with respect to the prior $\frac{1}{\beta}\mathbb{I}_{(0,\infty)}(\beta)$ (e.g., Berger, 1985). In view of Lemma 4.1's duality and these known results for entropy loss L_0, it is immediate that the subclass of plug-in $Ga(\alpha_2, cX)$ predictive density estimators have constant Kullback-Leibler risk, and that the optimal choice is given by $c = \frac{1}{\alpha_1 - 1}$.*
More generally, comparisons at the entropy risk level carry-over to the predictive Kullback-Leibler risk for assessing the performance of plug-in predictive density estimators. As a further observation, Marchand and Strawderman (2005) showed for the restricted parameter space $\beta \in C = (a, b)$ that the generalized Bayes estimator $\hat{\beta}_{\pi_C}(X)$ with respect to the prior $\pi_C(\beta) = \frac{1}{\beta}\, \mathbb{I}_C(\beta)$, dominates $\hat{\beta}_{mre}(X)$. It thus follows that the plug-in predictive density estimator $q_{\hat{\beta}_{\pi_C}}(\cdot\,; X)$ dominates the plug-in $q_{\hat{\beta}_{mre}}(\cdot\,; X)$ in terms of Kullback-Leibler risk for the restricted parameter space $\beta \in C = (a, b)$.

We further focus in this section on both **(i)** the lower bound constraint $\beta \geq a$, and **(ii)** the upper bound constraint $\beta \leq b$, in the study of the performance of plug-in predictive density estimators $q_{\hat{\beta}_c}$, $c > 0$, which are taken as $\text{Ga}(\alpha_2, \hat{\beta}_c)$ densities, with $\hat{\beta}_c$ a truncated linear estimator of β given by

$$\hat{\beta}_c(x) = \begin{cases} a & \text{if } cx \leq a \\ cx & \text{if } a < cx < b \\ b & \text{if } cx \geq b \end{cases} . \tag{15}$$

To further simplify the presentation, we will denote plug-in predictive density estimators $q_{\hat{\beta}_c}$ as $\hat{q}_c$. These predictive density or point estimators are interesting since they include the mle for $c = \frac{1}{\alpha_1}$, as well as the truncation of $\hat{\beta}_{mre}$ onto the parameter space for $c = \frac{1}{\alpha_1 - 1}$. Furthermore, for estimating β under entropy loss L_0, the truncated MRE estimator is minimax for both the restrictions $\beta \geq a$ and $\beta \leq b$, since it dominates the MRE estimator, with the latter being minimax despite the fact that it is not adapted to the restricted parameter space (e.g., Marchand and Strawderman, 2005). However, for the predictive density estimation problem, the situation is quite different and we will show that even the minimax choice among the $\hat{q}_c$'s is "far" from being minimax.

Lemma 4.2. *Let F_{α_1} and $\bar{F}_{\alpha_1} \equiv 1 - F_{\alpha_1}$ be the $Ga(\alpha_1, 1)$ cumulative and survivor functions respectively. For estimating the density of $Y \sim Ga(\alpha_2, \beta)$ with $\beta \in C$ based on $X \sim Ga(\alpha_1, \beta)$, the Kullback-Leibler risk of the plug-in predictive density estimator $\hat{q}_c$ is equal to*

$$(i) \qquad \alpha_2 \left(\left(\frac{\beta}{a} - \log(\frac{\beta}{a}) \right) F_{\alpha_1}(\frac{a}{c\beta}) - 1 + \int_{\frac{a}{c\beta}}^{\infty} (\frac{1}{cx} + \log c + \log x) \, p(x|1) \, dx \right), \ \textit{for } \beta \geq a \,, \tag{16}$$

$$(ii) \qquad \alpha_2 \left(\left(\frac{\beta}{b} - \log(\frac{\beta}{b}) \right) \bar{F}_{\alpha_1}(\frac{b}{c\beta}) - 1 + \int_{0}^{\frac{b}{c\beta}} (\frac{1}{cx} + \log c + \log x) \, p(x|1) \, dx \right), \ \textit{for } \beta \leq b \,. \tag{17}$$

Proof. By virtue of Lemma 4.1, we have in **(i)**

$$\begin{aligned} R_{KL}(\beta, \hat{q}_c) &= \alpha_2 E \left(\frac{\beta}{\hat{\beta}_c(X)} - \log(\frac{\beta}{\hat{\beta}_c(X)}) - 1 \right) \\ &= -\alpha_2 + \alpha_2 \left(\int_0^{a/c} (\frac{\beta}{a} - \log(\frac{\beta}{a})) \, p(x|\beta) \, dx + \int_{a/c}^{\infty} (\frac{\beta}{cx} - \log(\frac{\beta}{cx})) \, p(x|\beta) \, dx \right) , \end{aligned}$$

which leads to (16) with a change of variables $x \to x/\beta$. Finally, the Kullback-Leibler risk in (17) for situation **(ii)** $\beta \leq b$ is obtained in a similar fashion. □

The following is an analysis of the performance of the predictive density estimators $\hat{q}_c$ in terms of a maximum risk or minimax criterion.

Theorem 4.1. *Consider plug-in predictive density estimators $\hat{q}_c$ for estimating the density of $Y \sim Ga(\alpha_2, \beta)$, with $\beta \in C$, based on $X \sim Ga(\alpha_1, \beta)$ with $\alpha_1 > 1$. Then,*

- **(a)** *The Kullback-Leibler risk of $\hat{q}_c$ is increasing in $\beta \in C$ for situation:* **(i)** *$\beta \geq a$, and decreasing in $\beta \in C$ for situation:* **(ii)** *$\beta \leq b$;*

(b) *The minimax procedure and risk among the $\hat{q}_c$'s are given by the choice $c = \frac{1}{\alpha_1 - 1}$, and the value*

$$\inf_{c>0} \{\sup_{\beta \in C} R_{KL}(\beta, \hat{q}_c)\} \;=\; \alpha_2 \left(\psi(\alpha_1) - \log(\alpha_1 - 1)\right) \,, \tag{18}$$

independently of $C = [a, \infty)$ or $C = (0, b]$;

(c) *The difference in risks $\Delta_{c_1,c_2}(\beta) \;=\; R_{KL}(\beta, \hat{q}_{c_1}) - R_{KL}(\beta, \hat{q}_{c_2})$, where $0 < c_1 < c_2$ increases in β for both situations* **(i)** *$\beta \geq a$, and situation* **(ii)** *$\beta \leq b$. Consequently, $\hat{q}_{(\alpha_1-1)^{-1}}$ dominates all $\hat{q}_c$ with $c > \frac{1}{\alpha_1-1}$ in situation* **(i)** *$\beta \geq a$, and $\hat{q}_{(\alpha_1-1)^{-1}}$ dominates all $\hat{q}_c$ with $c < \frac{1}{\alpha_1-1}$, including $\hat{q}_{mle}$, in situation* **(ii)** *$\beta \leq b$.*

Proof. **(a)** A calculation using (16) yields the expression $\frac{d}{d\beta} R_{KL}(\beta, \hat{q}_c) \;=\; \frac{\alpha_2}{\beta}\,(\frac{\beta}{a} - 1)\, F_{\alpha_1}(\frac{a}{c\beta}) > 0$ for $\beta > a$, which implies the result for **(i)**. Similarly, we obtain $\frac{d}{d\beta} R_{KL}(\beta, \hat{q}_c) \;=\; \frac{\alpha_2}{\beta}\,(\frac{\beta}{b} - 1)\, \bar{F}_{\alpha_1}(\frac{b}{c\beta}) < 0$ for $\beta < b$.
(b) From part **(a)** and expression (16), we obtain for case **(i)** $\beta \geq a$

$$\begin{aligned}
\sup_{\beta \geq a} R_{KL}(\beta, \hat{q}_c) \;&=\; \lim_{\beta \to \infty} R_{KL}(\beta, \hat{q}_c) \\
&=\; -\alpha_2 \,+\, \alpha_2 \left(\int_0^{\infty} (\frac{1}{cx} + \log c + \log x)\, p(x|1)\, dx \right) \\
&=\; \alpha_2 \left(\frac{1}{c(\alpha_1 - 1)} + \log c + \psi(\alpha_1) - 1 \right) \,,
\end{aligned} \tag{19}$$

making use of $E_1(\frac{1}{X}) \;=\; \frac{1}{\alpha_1-1}$, as well as part (a) of Lemma 3.2, and observing that $\lim_{z\to\infty} z F_{\alpha_1}(1/z) \;=\; \lim_{z\to\infty} p(\frac{1}{z}\,|\,\alpha_1) = 0$ for $\alpha_1 > 1$. Similarly, for case **(ii)** $\beta \leq b$, the supremum risk is attained as $\beta \to 0$ yielding (19) as well. In both cases, minimization of (19) leads to the stated result.
(c) Again here, both the lower bounded and upper bounded cases **(i)** and **(ii)** are analogous, so we only tackle the upper bounded case $\beta \leq b$. It follows from the derivative in **(a)** that $\frac{\partial}{\partial\beta} \Delta_{c_1,c_2}(\beta) \;=\; \frac{\alpha_2}{\beta}\,(\frac{\beta}{b} - 1)\, \left(\bar{F}_{\alpha_1}(\frac{b}{c_1\beta}) - \bar{F}_{\alpha_1}(\frac{b}{c_2\beta}) \right) > 0$ for $0 < \beta < b$ and $c_1 < c_2$. Furthermore, for $c_1 \;<\; \frac{1}{\alpha_1-1}$ and $0 \;<\; \beta \;\leq\; b$, we have $\Delta_{c_1,\frac{1}{\alpha_1-1}}(\beta) \;>\; \Delta_{c_1,\frac{1}{\alpha_1-1}}(0^+) \;>\; 0$ given that $\hat{q}_{\frac{1}{\alpha_1-1}}$ minimizes the risk at $\beta \to 0^+$ among estimators $\hat{q}_c$. $\square$

Both plug-in predictive density estimators $\hat{q}_c$ and the Bayes predictive density estimator $\hat{q}_{\pi_0,C}$ exploit the restricted parameter information $\beta \in C$. As expanded on with the next example, and in opposition to a large collection of various point estimation problems (e.g., Marchand and Strawderman, 2012), they differ in efficiency and the difference can be significant. Actually, the maximum risk of all plug-in's $\hat{q}_c$ even exceeds the constant risk of the predictive MRE density $\hat{q}_{\pi_0}$ which ignores the lower bound restriction.

Example 4.1. *We illustrate here some of the above findings as applied to restricted parameter spaces $C = (a, b)$ with either $a = 0$ or $b = +\infty$. Denote $\overline{R}_{KL}(\hat{q})$ the supremum risk of a predictive density estimator $\hat{q}$. Consider the plug-in predictive density estimators $\hat{q}_{mle}$ and $\hat{q}_{1/(\alpha_1-1)}$, as well as the minimax estimator $\hat{q}_{\pi_0,C}$. Our findings above tell us that $\overline{R}_{KL}(\hat{q}_{mle}) > \overline{R}_{KL}(\hat{q}_{1/(\alpha_1-1)}) > \overline{R}_{KL}(\hat{q}_{\pi_0,C})$. From (19), we have $\overline{R}_{KL}(\hat{q}_{mle}) = \alpha_2 \left(\psi(\alpha_1) + \frac{\alpha_1}{\alpha_1-1} - \log \alpha_1 - 1 \right)$, while $\overline{R}_{KL}(\hat{q}_{1/(\alpha_1-1)})$ and $\overline{R}_{KL}(\hat{q}_{\pi_0,C})$ are given by (18) and (13) respectively. As a numerical illustration, consider $\alpha_1 = 2$ and $\alpha_2 = 1$. Evaluations yield $\overline{R}_{KL}(\hat{q}_{mle}) = 2 - \log 2 - \gamma \approx 0.72964$, $\overline{R}_{KL}(\hat{q}_{1/2}) = 1 - \gamma \approx$*

0.423784, and $\overline{R}_{KL}(\hat{q}_{\pi_0,C}) = 3/2 - \log 2 - \gamma \approx 0.22964$*. The differences are quite important here in relative terms, but are attenuated for larger* α_1*. This is further illustrated in Figure 1, where we have exhibited the ratio of minimax risks*

$$\frac{\inf_{c>0}\{\sup_{\beta\geq a} R_{KL}(\beta,\hat{q}_c)\}}{\inf_{\hat{q}}\{\sup_{\beta\geq a} R_{KL}(\beta,\hat{q})\}} = \frac{\alpha_2\,(\psi(\alpha_1) - \log(\alpha_1 - 1))}{(\alpha_1+\alpha_2)\psi(\alpha_1+\alpha_2) + \log\Gamma(\alpha_1) - \log\Gamma(\alpha_1+\alpha_2) - \alpha_2 - \alpha_1\psi(\alpha_1)},$$

as a function of α_1*, for a selection of* α_2 *values. Finally, we point out the elegant difference*

$$\bar{R}_{KL}(\hat{q}_{mle}) - \bar{R}_{KL}(\hat{q}_{1/(\alpha_1-1)}) = \alpha_2\left(\frac{\alpha_1}{\alpha_1-1} - \log(\frac{\alpha_1}{\alpha_1-1}) - 1\right),$$

and reiterate that this expression, as well as the findings of this section relative to the performance of the predictive density estimators $\hat{q}_c$*, also represent point estimation results under entropy loss* L_0 *for truncated linear estimators of the form* $\hat{\beta}_c$*.*

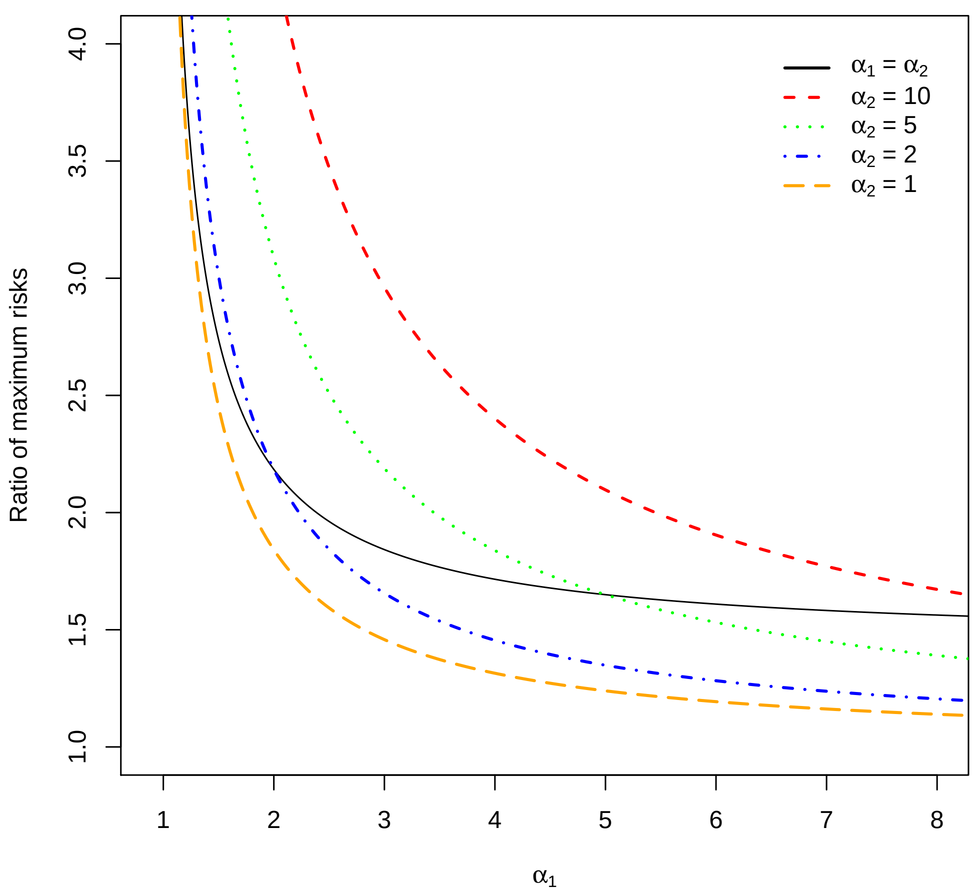

Figure 1: Ratios $\frac{\inf_c \overline{R}(\hat{q}_c)}{\inf_{\hat{q}} \overline{R}(\hat{q})}$ of minimax risks for $\beta \geq a$.

5 Variance expansion improvements on plug-in predictive density estimators

Previous predictive density estimation findings for location models have shown that improvements on plug-in predictive density estimators can be found by expanding the variance up to a certain level (Fourdrinier et al. 2011; Kubokawa, Marchand and Strawderman, 2015A, 2015B). In the spirit of searching for plausible improvements, we investigate below predictive density estimators of the form $\mathrm{Ga}(\frac{\alpha_2}{k}, k\hat{\beta}(X))$ for $k \geq 1$, a given plug-in estimator $\hat{\beta}(X)$ of β. For fixed x, the related expectation is constant as a function of k, but the variance $k\alpha_2\hat{\beta}(x)$ increases in k. We now show that improvements on the plug-in predictive density estimator (i.e., $k = 1$) can always be found, regardless of the parameter space and of $\hat{\beta}(X)$.

Theorem 5.1. *Let $q_{\hat{\beta}}(\cdot; X) \sim Ga(\alpha_2, \hat{\beta}(X))$ be a plug-in estimator, with non-degenerate $\hat{\beta}(X)$, for estimating the density of $Y \sim Ga(\alpha_2, \beta)$ under Kullback-Leibler loss with $\beta \in C = (a, b)$, and based on $X \sim Ga(\alpha_1, \beta)$. Denote $R(\beta, \hat{\beta}) = E(\frac{\beta}{\hat{\beta}(X)} - \log(\frac{\beta}{\hat{\beta}(X)}) - 1)$ the entropy risk of $\hat{\beta}(X)$ and denote $\underline{R}(\hat{\beta}) = \inf_{\beta \in C} R(\beta, \hat{\beta})$. Then, $q_{\hat{\beta}}(\cdot; X)$ is inadmissible and dominated by the subclass of predictive density estimators $q_{\hat{\beta},k}(\cdot; X) \sim Ga(\frac{\alpha_2}{k}, k\hat{\beta}(X))$ with $1 < k \leq k_0(\underline{R}(\hat{\beta}))$, $k_0(\underline{R}(\hat{\beta}))$ being the unique solution in $k \in (1, \infty)$ of the equation $G_{\underline{R}(\hat{\beta})}(k) = 0$, with*

$$G_s(k) = \alpha_2(\frac{1}{k} - 1)\,(s + 1 - \psi(\alpha_2)) \,+\, \frac{\alpha_2}{k}\,\log k \,+\, \log\frac{\Gamma(\frac{\alpha_2}{k})}{\Gamma(\alpha_2)}\,.$$

Proof. First fix $\beta \in (a, b)$ and set

$$h(k, \beta) \;=\; R_{KL}(\beta, q_{\hat{\beta},k})\,.$$

From (3), we obtain

$$\begin{aligned} h(k,\beta) \;&=\; E_{\alpha_2}\left\{\alpha_2(1 - \frac{1}{k})\log(\frac{Y}{\beta}) + \frac{Y}{\beta}\left(\frac{\beta}{k\hat{\beta}(X)} - 1\right) + \frac{\alpha_2}{k}\,\log(\frac{k\hat{\beta}(X)}{\beta}) + \log\Gamma(\frac{\alpha_2}{k}) - \log\Gamma(\alpha_2)\right\} \\ &=\; \alpha_2\,(\psi(\alpha_2) - 1) - \log\Gamma(\alpha_2) + \frac{\alpha_2}{k}\left(\log k - \psi(\alpha_2) + 1 + R(\beta, \hat{\beta})\right) + \log\Gamma(\frac{\alpha_2}{k})\,, \qquad (20) \end{aligned}$$

by the independence of X and Y, and by using the identities $E_{\alpha_2}(\frac{Y}{\beta}) = \alpha_2$ and $E_{\alpha_2}(\log(\frac{Y}{\beta})) = \psi(\alpha_2)$. By differentiation, we obtain $\frac{d}{dk}h(k,\beta) = \frac{\alpha_2}{k^2}T(k)$, with

$$T(k) = \psi(\alpha_2) - R(\beta, \hat{\beta}) - \psi(\frac{\alpha_2}{k}) - \log k\,. \qquad (21)$$

Now, observe that $T(k)$ increases in $k \geq 1$ as an application of part **(c)** of Lemma 3.2. Observe as well that $\lim_{k\to\infty} h(k,\beta) = +\infty$. Since $\frac{d}{dk}h(k,\beta)|_{k=1^+} = -\alpha_2 R(\beta, \hat{\beta}) < 0$, it follows that $\frac{\partial}{\partial k}h(k,\beta)$ changes signs from $-$ to $+$ as k increases from 1 to ∞, which implies that $h(k,\beta) - h(1,\beta) < 0$ iff $1 < k < k_0(\beta)$, where $G_{R(\beta,\hat{\beta})}(k_0(\beta)) = 0$. With $k_0(\beta)$ increasing in $R(\beta, \hat{\beta})$, the result follows. □

Remark 5.1. *Theorem 5.1 is appealing since it is universal in our context, applying for all parameter spaces $C = (a, b)$ and all plug-in predictive density estimators $q_{\hat{\beta}}$ with non-degenerate $\hat{\beta}(X)$. We also point out that, while the Gamma distributed assumption for Y is necessary,* *the same is not true for the distribution of X and the theorem also holds for other scale family models $X \sim \frac{1}{\beta} f_1(\frac{x}{\beta})$, which only impacts the risk $R(\beta, \hat{\beta})$, the corresponding minimum risk $\underline{R}(\hat{\beta})$, and the upper cut-off point $k_0(\underline{R}(\hat{\beta}))$. Furthermore, Theorem 5.1 does provide explicit predictive density estimators that dominate plug-in predictive density estimators, such as the predictive mle $\hat{q}_{1/\alpha_1}$, the minimax $\hat{q}_c$ procedure (e.g., Theorem 4.1) given by the choice $c = \frac{1}{\alpha_1 - 1}$ for $\alpha_1 > 1$, and $q_{\hat{\beta}_{\pi_C}}$ with $\hat{\beta}_{\pi_C}(X)$ the Bayes estimator of β with respect to the prior $\pi_C(\beta) = \frac{1}{\beta} \mathbb{I}_C(\beta)$ (see Remark 4.1).*

Remark 5.2. *Theorem 5.1 applies for the case of the unrestricted parameter space $C = (0, \infty)$. In particular, it applies for the plug-in MRE estimator $\hat{\beta}(X) = \frac{X}{\alpha_1 - 1}$ when $\alpha_1 > 1$. However, a stronger inference is available as there exists an optimal predictive density estimator in the class of $Ga(\frac{\alpha_2}{k}, \frac{kX}{\alpha_1 - 1})$ predictive densities with $k > 1$. This follows as the risk of such estimators are, as seen by (20), constant as a function of β with $R(\beta, \hat{\beta}_{mre}) = \psi(\alpha_1) - \log(\alpha_1 - 1)$ (see 18), and uniquely minimized at $k_0 = T^{-1}(0)$ where, from (21), we have*

$$T(k) = \psi(\alpha_2) + \log(\alpha_1 - 1) - \psi(\alpha_1) - \psi(\frac{\alpha_2}{k}) - \log k\,.$$

Notwithstanding the above illustrative discussion, all these predictive density estimators remain inadmissible and, being invariant with respect to changes of scale, they are dominated by the minimum risk equivariant predictive density estimator $\hat{q}_{\pi_0}$.

6 Illustrations and final comments

Numerical risk evaluations are presented in Figures 2, 3 and 4 and commented on in this section. These serve to illustrate most of the findings of this paper, are useful in assessing the strengths and weaknesses of different predictive density estimators, and are particularly representative of the degree of improvement obtained when a theoretical dominance result applies.

- Figure 2 relates to $\alpha_1 = 3, \alpha_2 = 5, C = [1, \infty)$. First, the risk function of the Bayes predictive density estimator $\hat{q}_{\pi_0,C}$ is represented. It performs well overall and its minimaxity is seen to be quite beneficial. In accordance to Theorem 3.1, it dominates the Bayes predictive density $\hat{q}_{\pi_0}$, which is also minimax (Corollary 3.1), with matching risks at $\beta = 1$ and at $\beta \to \infty$ (Theorem 3.1). The exact minimax risk is given by (13) and is equal to $8\psi(8) + \log\Gamma(3) - \log\Gamma(8) - 5 - 3\psi(3) \;=\; -5\gamma - \log 2520 + \frac{787}{70} \approx 0.524763$.

- Looking again at Figure 2, the risk functions of the plug-in predictive mle $\hat{q}_{1/3} \sim Ga(5, \hat{\beta}_{1/3}(X))$ and $\hat{q}_{1/2} \sim Ga(5, \hat{\beta}_{1/2}(X))$, as well as variance expansion improvements derived from Theorem 5.1, are also represented. The improvements are $\hat{q}_{c,k_1} \sim Ga(\frac{5}{k_1}, k_1 \hat{\beta}_c(X))$, $c = 1/3, 1/2$, predictive densities, with $k_1 = k_0(R(1, \hat{\beta}_c)) \approx 1.88, 5.14$ (respectively) taken as the upper cutoff point of Theorem 5.1, and with $\underline{R}(\hat{\beta}_c) = \inf_{\beta \geq 1} R(\beta, \hat{\beta}_c) = R(1, \hat{\beta}_c)$ by virtue of part (a) of Theorem 4.1 and Lemma 4.1. The degree of improvement is striking with $\hat{q}_{1/2,k_1}$ even dominating $\hat{q}_{1/3}$. Compared to the Bayes predictive density estimator $\hat{q}_{\pi_0,C}$, the predictive density mle $\hat{q}_{1/2,k_1}$ is efficient for small β with smaller risk for $\beta \leq \beta_0$, but with a less favorable performance otherwise and supremum risk well above the minimax risk. The variance inflated predictive density $\hat{q}_{1/3,k_1}$ is more efficient in terms of supremum risk, but performs poorly being dominated by both $\hat{q}_{\pi_0}$ and $\hat{q}_{\pi_0,C}$, with the former not even exploiting the parametric restriction.

- Other cases α_1, α_2 and choices of k, many of which we looked at, yield similar comparisons with the Bayes predictive density estimator performing well in comparison to the plug-in predictive density estimators $\hat{q}_c$ and their variance inflated improvements.

- Figure 3 is such an illustration with $\alpha_1 = 2$, $\alpha_2 = 1$, $\beta \leq 1.0$. The plug-in MRE estimator is an Exponential(X) density, while the chosen variance expansion improvement is a $Ga(\frac{1}{k_1}, k_1 X)$ density with $k_1 \approx 1.745$. The inefficiency of $\hat{q}_{mle}(\cdot; x) \sim Exponential(\frac{x}{2})$ is more striking here, with $\hat{q}_{1/(\alpha_1-1)}$ offering much lower risk in accordance with Theorem 4.1, and with

$$\begin{aligned} R_{KL}(\beta, \hat{q}_{mle}) - R_{KL}(\beta, \hat{q}_{1/(\alpha_1-1)}) &\geq \lim_{\beta \to 0+} (R_{KL}(\beta, \hat{q}_{mle}) - R_{KL}(\beta, \hat{q}_{1/(\alpha_1-1)})) \\ &= 1 - \log 2\,, \end{aligned}$$

  for all $\beta \leq 1.0$, as a consequence of Theorem 4.1 and (19).

- Finally, Figure 4 illustrates a doubly-bounded case with $\alpha_1 = 3, \alpha_2 = 3$ and $C = [1, 8]$. In accordance with Theorem 3.1, the Bayes predictive density estimator $\hat{q}_{\pi_0,C}$ dominates both $\hat{q}_{\pi_0,C'}$ and the MRE predictive density estimator $\hat{q}_{\pi_0}$ with $C' = [1, \infty)$. It also performs quite well in comparison to the plug-in predictive density estimators $\hat{q}_{\text{mle}}$ and $\hat{q}_{\frac{1}{\alpha_1 - 1}}$ with risk improvement on a very large part of the parameter space.

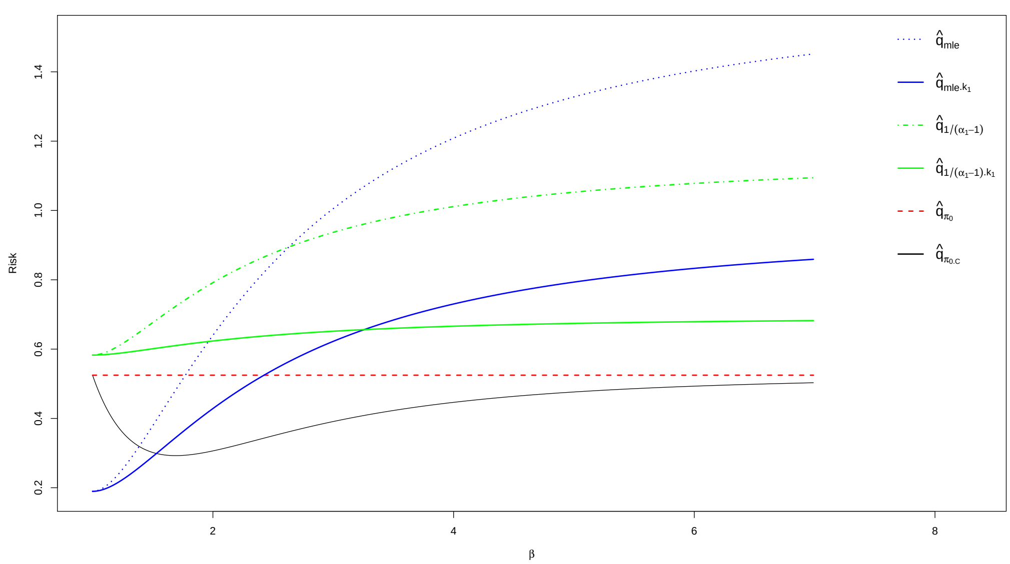

Figure 2: Risks of predictive density estimators : $\alpha_1 = 3$, $\alpha_2 = 5$, $\beta \geq 1$.

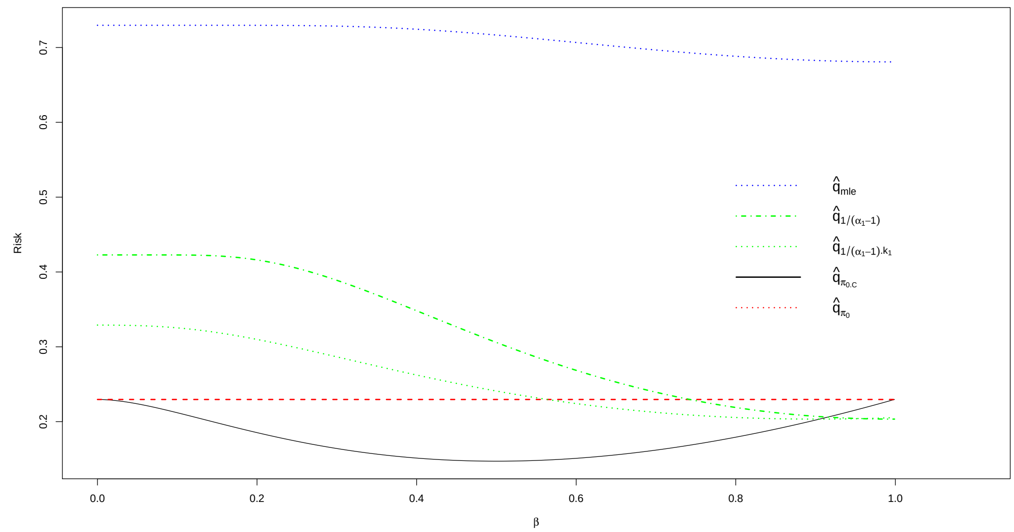

Figure 3: Risks of predictive density estimators : $\alpha_1 = 2$, $\alpha_2 = 1$, $\beta \le 1.0$.

Acknowledgements

We are thankful to two anonymous reviewers for quite thoughtful and constructive comments. Eric Marchand's research is supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by grants from the Simons Foundation (#209035 and #418098).

Appendix

We briefly expand here on a general representation for Bayes predictive density estimators in scale family problems associated with the usual noninformative prior (e.g., Berger, 1985; chapter 6). Such predictive densities are also Minimum Risk Equivariant (MRE) with respect to changes in scale $X \to gX$, $g > 0$, (e.g., Liang and Barron, 2004). In our set-up, as well as the more general scale family set-up of Theorem 6.1 below, both the loss and the family of distributions leave the problem invariant, equivariant predictive densities are characterized by the property $\hat{q}(y; gx) = \frac{1}{g}\hat{q}(\frac{y}{g}; x)$, for all $g > 0$, and such equivariant predictive densities have constant Kullback-Leibler risk.

Now, observe that the predictive density $\hat{q}_{\pi_0}(\cdot; x)$, given in Theorem 2.1, is the density of $x\frac{T_2}{T_1}$, where $T_1 \sim \text{Ga}(\alpha_1, \beta)$, $T_2 \sim \text{Ga}(\alpha_2, \beta)$ are independently distributed (as in model (1)), as such ratios are Beta2(α_2, α_1, x) distributed. Here is a generalization to other scale parameter families followed by a pair of examples.

Theorem 6.1. *Let $X|\sigma \sim \frac{1}{\sigma}\, p(\frac{x}{\sigma})$, $Y|\sigma \sim \frac{1}{\sigma}\, q(\frac{y}{\sigma})$ be independently distributed nonnegative random variables. Let h be the density of the ratio $\frac{Y}{X}$ (which does not depend on σ) given by $h(v) = \int_0^\infty u\, q(vu)\, p(u)\, du$. Then, the Bayes predictive density estimator under Kullback-Leibler loss in (2) of the density of Y associated with the prior $\pi_0(\sigma) = \frac{1}{\sigma}\, \mathbb{I}_{(0,\infty)}(\sigma)$, is given by $\hat{q}_{\pi_0}(y; x) = \frac{1}{x}\, h(\frac{y}{x})$.*

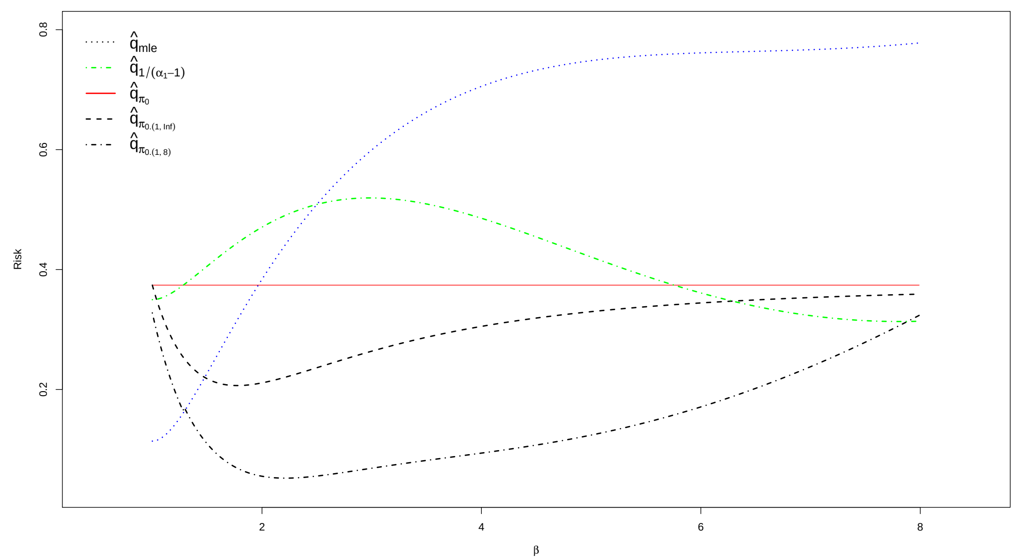

Figure 4: Risks of predictive density estimators : $\alpha_1 = 3$, $\alpha_2 = 3$, $\beta \in [1, 8]$.

Proof. We have

$$\begin{aligned} \hat{q}_{\pi_0}(y;x) &= \frac{\int_0^\infty \frac{1}{\sigma}q(\frac{y}{\sigma})\frac{1}{\sigma}p(\frac{x}{\sigma})\,\frac{d\sigma}{\sigma}}{\int_0^\infty \frac{1}{\sigma}q(\frac{x}{\sigma})\,\frac{d\sigma}{\sigma}} \\ &= \frac{1}{x}\frac{\int_0^\infty uq(\frac{y}{x}u)p(u)du}{\int_0^\infty p(u)du} \\ &= \frac{1}{x}\,h(\frac{y}{x})\,. \qquad \square \end{aligned}$$

Example 6.1. *(Uniform model)*
For $X|\sigma \sim U(0,\sigma)$ $Y|\sigma \sim U(0,\sigma)$, we have $h(v) = \int_0^\infty u\,\mathbb{I}_{(0,1)}(uv)\,\mathbb{I}_{(0,1)}(u)\,du = \frac{1}{2}\,\mathbb{I}_{(0,1]}(v) + \frac{1}{2v^2}\,\mathbb{I}_{(1,\infty)}(v)$; i.e., the density of the ratio $\frac{Y}{X}$, and thus $\hat{q}_{\pi_0}(y;x) = \frac{1}{x}\,h(\frac{y}{x}) = \frac{1}{2x}\,\mathbb{I}_{(0,x]}(y) + \frac{x}{2y^2}\,\mathbb{I}_{(x,\infty)}(y)$.

Example 6.2. *(Beta type 2 model)*
For $X|\sigma \sim Beta2(c_1,d_1,\sigma)$ and $Y|\sigma \sim Beta2(c_2,d_2,\sigma)$ as in Definition 2.1, we obtain $\hat{q}_{\pi_0}(y;x) = \frac{1}{x}\,h(\frac{y}{x})$, with the distribution of the ratio $\frac{Y}{X}$ and its density h studied in Pham-Gia and Turkkan (2002) for instance.

References

Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547-554.

Amaral, M.A. and Dunsmore, I.R. (1980). Optimal estimates of predictive distributions. *Biometrika*, **67**, 685-689.

Andrews, G.E., Askey, R. and Roy, R. (1999). *Special functions.* Encyclopedia of Mathematics and its Applications, **71**, Cambridge University Press.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 2nd edition.

Boisbunon, A. and Maruyama, Y. (2014). Inadmissibility of the best equivariant predictive density in the unknown variance case. *Biometrika*, **101**, 733-740.

Brown, L.D., George, E.I., and Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170.

Fourdrinier, D., Marchand, É., Righi, A. and Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191.

George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91.

Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.

Kato K. (2009). Improved prediction for a multivariate normal distribution with unknown mean and variance. *Annals of the Institute of Statistical Mathematics*, **61**, 531-542.

Komaki, T. (2006). A class of proper priors for Bayesian simultaneous prediction of independent Poisson observables. *Journal of Multivariate Analysis*, **97**, 1815-1828.

Komaki, T. (2004). Simultaneous prediction of Poisson observables. *Annals of Statistics*, **32**, 1744-1769.

Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88**, 859-864.

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

Kubokawa, T., Marchand, É & Strawderman, W.E. (2015A). On predictive density estimation for location families under integrated absolute error loss. To appear in *Bernoulli*.

Kubokawa, T., Marchand, É & Strawderman, W.E. (2015B). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, **142**, 57-74.

Kubokawa, T., Marchand, É., Strawderman, W.E. and Turcotte, J-P. (2013). Minimaxity in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, **116**, 382-397.

Liang, F. and Barron, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inform. Theory*, **50**, 2708-2726.

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

Marchand, É., and Strawderman, W. E. (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.

Marchand, É. and Strawderman, W.E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, 21-44.

Misra, N. and Arshad, M. (2014). Monotonicity of certain integrals involving gamma distributions and their applications in multiple comparisons. *Statistics & Probability Letters*, **85**, 144-152.

Muldoon, M.E. (1978). Some monotonicity properties and characterizations of the gamma function. *Aequationes Math.*, **18**, 54-63.

Mukherjee, G. and Johnstone, I.M. (2015). Exact minimax estimation of the predictive density in sparse Gaussian models. *Annals of Statistics*, **43**, 937-961.

Mukherjee, G. (2013). *Sparsity and shrinkage in predictive density estimation.* Doctoral dissertation, Stanford University.

Pham-Gia, T. and Turkkan, N. (2002). Operations on Generalized-F variables and applications. *Statistics*, **36**, 195-209.

Rahman, M.S. and Gupta, R.P. (1993). Family of transformed chi-square distributions. *Communications in Statistics: Theory and Methods*, **22**, 135-146.

Rukhin, A. (2016). Decision-theoretic issues in heterogeneity variance estimation. *Annals of the Institute of Statistical Mathematics*, **68**, 571-588.

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

van Zwet, W.R. (1964). Convex transformations of random variables. *Mathematical Centre Tracts, 7 Mathematisch Centrum, Amsterdam vi+116 pp*