Γ-Minimax: A Paradigm for Conservative Robust Bayesians

Brani Vidakovic

ABSTRACT In this chapter a tutorial overview of Gamma minimaxity (Γ-minimaxity) is provided. One of the assumptions of the robust Bayesian analysis is that prior distributions can seldom be quantified or elicited exactly. Instead, a family of priors, Γ, reflecting prior beliefs is elicited. The Γ-minimax decision-theoretic approach to statistical inference favors an action/rule which incorporates information specified via Γ, and guards against the least favorable prior in Γ. This paradigm falls between Bayesian and minimax paradigms; it coincides with the former when prior information can be summarized in a single prior, and with the latter when no prior information is available (or equivalently, possible priors belong to the class of *all* distributions).

1 What is Γ-Minimax?

The Gamma minimax (Γ-minimax) approach was originally proposed by Robbins (1951). Under this approach the statistician (decision maker) is unable (or unwilling) to specify a single prior distribution reflecting his or her prior knowledge about model parameters of interest. Rather, the statistician is able to elicit a class Γ of plausible prior distributions. This idea is vividly expressed by Efron and Morris (1971):

> ... We have referred to the "true prior distribution" ... but in realistic situations there is seldom any one population or corresponding prior distribution that is "true" in an absolute sense. There are only more or less relevant priors, and the Bayesian statistician chooses among those as best he can, compromising between his limited knowledge of subpopulation distributions and what is usually an embarrassingly large number of identifying labels attached to the particular problem.

Optimality of a decision rule is generally judged by some form of cost for the inaccuracy of the rule. In Γ-minimax, the rule or action that minimizes the supremum of the cost functional over distributions in Γ is selected. If prior information is scarce, the class Γ of priors under consideration is large and the decision maker's actions are close to the minimax actions. In the extreme case when no information is available, the Γ-minimax setup is equivalent to the usual minimax setup.

If, on the other hand, the statistician has substantial prior information, then the class Γ can be narrow. An extreme case is a class Γ that contains a single prior. In this case, the Γ-minimax framework becomes the usual Bayes framework.

When the model is regular, the Γ-minimax decision coincides with the Bayes decision with respect to the least favorable prior. Thus, establishing the regularity of the model is fundamental in solving minimax-type problems.

Depending on the cost functional, the spirit of the Γ-minimax paradigm can be (i) classical, if the payoff for the decision maker is measured by the Bayes risk [integrated frequentist risk] and (ii) Bayesian, if the payoff is measured by the posterior expected loss. For such payoffs, Γ-minimax is often called *conditional* Γ*-minimax;* see Section 4.

Next we give a simple example to illustrate some of the notions.

Example 1. Suppose we observe $X|\theta \sim \mathcal{N}(\theta, 1)$ where $\theta \in \{-1, 1\}$. Let

$$\theta \sim \pi_p(\theta) \in \Gamma = \left\{ \left(\begin{array}{cc} -1 & 1 \\ 1-p & p \end{array} \right),\ 0 \le p \le 1 \right\},$$

and let the class of decision rules under consideration $\mathcal{D}_a$ be indexed by a,

$$\delta_a = \left\{ \begin{array}{ll} -1, & x < a \\ 1, & x \ge a \end{array} \right. .$$

What is the Γ-minimax rule in the class $\mathcal{D}_a$ when the loss is squared error?

The frequentist risk of δ_a is

$$R(\theta, \delta_a) = E^{X|\theta}(\theta - \delta_a)^2 = (\theta+1)^2\Phi(a-\theta) + (\theta-1)^2[1 - \Phi(a-\theta)].$$

The integrated frequentist risk (or payoff function) is

$$r(\pi_p, \delta_a) = 4p\Phi(a-1) + 4(1-p)[1 - \Phi(a+1)]. \tag{1.1}$$

The problem is regular. That is, the payoff function as a function of two variables p and a has a saddle point at $a = 0$ and $p = 1/2$. This can be demonstrated by simple calculus arguments (by equating partial derivatives to 0 and verifying analytic requirements for saddle point).

Since δ_0 is Bayes with respect to $\pi_{1/2}$, δ_0 is Γ-minimax in $\mathcal{D}_a$ and $\pi_{1/2}$ is the least favorable prior. One can demonstrate that δ_0 is overall Γ-minimax, i.e., Γ-minimax with respect to all (measurable) decision rules.

When either of Γ, $\mathcal{D}_a$ or the payoff function is more complicated, the above approach may not be possible. Results described in the next section help solve more general Γ-minimax problems.

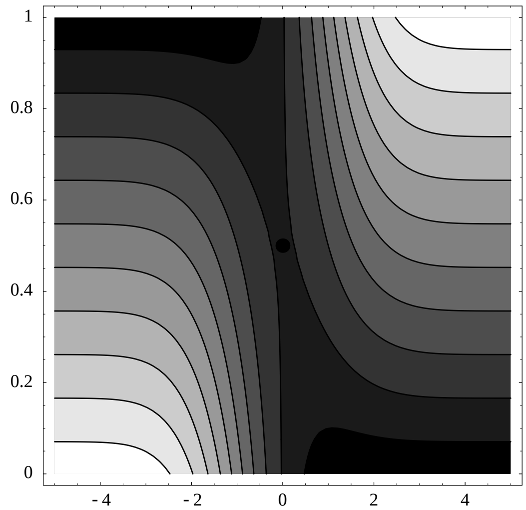

FIGURE 1. Contour plot of the integrated frequentist risk in (1.1). The horizontal axis is the value of a in $\mathcal{D}_a$ and the vertical axis is p in π_p. Note the saddle-point at $(0, 1/2)$.

The Γ-minimax paradigm has been criticized on several grounds. Some Bayesians object that belief in the Γ-minimax principle may produce "demonstrable incoherence," since there are examples in which the Γ-minimax rule is not Bayes.

Other complaints concern the use of frequentist measures as cost (payoff) measures, as well as the fact that the Γ-minimax rules often guard against priors from Γ deemed "unreasonable." While in principle agreeing with such concerns, we emphasize that the cost function may not be the decision maker's choice, especially when the inference is interpreted as a statistical game. Also, least favorable priors, against which Γ-minimax rules are guarding, can be made "more reasonable" by careful elicitation of Γ.

There is also widespread justification for Γ-minimax and for a detailed discussion we direct the reader to Berger (1984) and Rios (1990). Experiments and observations showed that in the presence of uncertainty, the decision-maker often takes conservative actions even though they might be suboptimal. We give an example which is based on *Ellsberg's Paradox* (Ellsberg, 1954) representing a behavioral defense of Γ-minimax.

> Suppose box I contains 50 white and 50 black balls while box II also contains 100 white and black balls, although their proportion is unknown (in absence of information all proportions might be considered equally likely).
>
> A subject in the experiment first chooses a box and then draws

> a ball at random from the selected box. If the selected ball is white, the subject receives a prize of, say, \$1,000. If a black ball is selected, the subject receives no prize. Which box should the subject choose?
>
> Although preferences for selecting either of the boxes should be equal, most subjects in the experiment prefer the box I. An explanation for such decisions is that subjects act in Γ-minimax fashion, protecting themselves against unfavorable proportions, or *unreliable probabilities*, as Ellsberg calls them.

There is substantial research published on applications of Γ-minimax to statistical estimation, testing, and ranking and selection. For the foundations and the philosophy of Γ-minimaxity the reader is referred to, among others, Robbins (1951), George (1969), Good (1952), Berger (1984, 1985), and Rios (1990).

A group of researchers affiliated with the Technische Hochschule at Darmstadt (Bischoff, W., Chen, L., Eichenauer-Herrmann, J., Ickstadt, K., Lehn, J., Weiß, E., and others) have a substantial body of research on Γ-minimax estimation, foundations, and certain testing problems.

Γ-minimax research in problems of ranking and selection has been studied by the Purdue decision theory group (Gupta, S. S., Huang, D-Y., Berger, R., Kim, W.-C., Miescke, K. J., and Hsiao, P.).

Relevant references on Γ-minimaxity can be found in Berger (1994).

2 The Mathematical Formulation and Some Standard Results

In this section we formulate the Γ-minimax problem in the estimation setup and provide the basic results and definitions.

Let X be a random variable whose distribution is in $\{P_\theta,\ \theta \in \Theta\}$, a family which is indexed by a parameter (random variable) θ. The goal is to make an inference about the parameter θ, given an observation X. A solution is a *decision procedure (decision rule)* $\delta(x)$, which identifies a particular inference for each value of x that can be observed. Let $\mathcal{A}$ be the class of all *actions*; that is, all possible realizations of $\delta(x)$. The *loss function* $L(\theta, a)$ maps $\Theta \times \mathcal{A}$ into the set of real numbers and defines the cost to the statistician when action a is taken and the true value of the parameter is θ. A *risk function* $R(\theta, \delta)$ characterizes the performance of the rule δ for each value of parameter $\theta \in \Theta$. The risk is usually defined in terms of the underlying loss function $L(\theta, a)$ as

$$R(\theta, \delta) = E^{X|\theta} L(\theta, \delta(X)) \tag{1.2}$$

where $E^{X|\theta}$ is the expectation with respect to P_θ. Since the risk function is defined as an average loss with respect to a sample space, it is called the *frequentist risk*. Let $\mathcal{D}$ be the collection of all measurable decision rules. There are several principles for determining preferences among the rules in $\mathcal{D}$. The three most relevant here are the Bayes principle, the minimax principle, and the Γ-minimax principle.

Under the Bayes principle, the prior distribution π is specified on the parameter space Θ. Any rule δ is characterized by its *Bayes risk*

$$r(\pi, \delta) = \int R(\theta, \delta)\pi(d\theta) = E^{\pi} R(\theta, \delta). \tag{1.3}$$

The rule δ_π that minimizes Bayes risk is called *Bayes rule* and is given by

$$\delta_\pi = \arg \inf_{\delta \in \mathcal{D}} r(\pi, \delta). \tag{1.4}$$

The *Bayes risk of the prior distribution* π (*Bayes envelope function*) is

$$r(\pi) = r(\pi, \delta_\pi). \tag{1.5}$$

The distribution π^* for which

$$r(\pi^*) \geq r(\pi), \text{ for any } \pi, \tag{1.6}$$

is called the *least favorable prior*.

Under the minimax principle, the optimal rule δ^* which minimizes the maximum of the frequentist risk $R(\theta, \delta)$ and is given by

$$\delta^* = \arg \inf_{\delta \in \mathcal{D}} (\sup_{\theta \in \Theta} R(\theta, \delta)) \tag{1.7}$$

is called the *minimax rule*.

Now suppose that instead of a single prior on θ, the statistician elicits a family of priors, Γ. Under the Γ-minimax principle, the rule δ_0 is optimal if it minimizes $\sup_{\pi \in \Gamma} r(\pi, \delta)$; specifically

$$\delta_0 = \arg \inf_{\delta \in \mathcal{D}} (\sup_{\pi \in \Gamma} r(\pi, \delta)). \tag{1.8}$$

Such a rule δ_0 is called Γ-*minimax rule*. The corresponding Γ-*minimax risk* is given by

$$\bar{r}_\Gamma = \inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta). \tag{1.9}$$

Let $\underline{r}_\Gamma$ be the supremum of $r(\pi)$ over the class Γ,

$$\underline{r}_\Gamma = \sup_{\pi \in \Gamma} \inf_{\delta \in \mathcal{D}} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi) = r(\pi_0). \tag{1.10}$$

Then

$$\underline{r}_\Gamma \leq \bar{r}_\Gamma.$$

This fact has an interpretation as a lower-upper value inequality in the theory of statistical games. For instance, an intelligent player chooses a prior distribution on θ, and the statistician responds by choosing the rule $\delta \in \mathcal{D}$. The payoff function for the statistician is $r(\pi, \delta)$. The rule δ_0 is the minimax strategy for the statistician and $\bar{r}_\Gamma$ is an upper value of the game. It is of interest that a statistical game has a value since in this case finding the Γ-minimax rules is straightforward. The following theorem gives conditions under which the statistical game has a value.

Theorem 2.1 *If δ_0 is the Bayes rule with respect to the prior $\pi_0 \in \Gamma$, and for all $\pi \in \Gamma$*

$$r(\pi, \delta_0) \leq r(\pi_0, \delta_0), \tag{1.11}$$

then $\underline{r}_\Gamma = \bar{r}_\Gamma$, δ_0 is Γ-minimax and π_0 is the least favorable prior.

Example 2. Let $X \sim \mathcal{N}(\theta, 1)$ and let $\Gamma = \{\pi \mid E^\pi\theta = \mu, E^\pi(\theta - \mu)^2 = \tau^2,\ \mu, \tau \text{ fixed }\}$. Let $\delta_0(x) = \frac{\tau^2}{1+\tau^2}x + \frac{1}{1+\tau^2}\mu$. Then, for any $\pi \in \Gamma$, $r(\pi, \delta_0) = \frac{\tau^2}{1+\tau^2}$. In addition, δ_0 is Bayes with respect to $\pi_0 = \mathcal{N}(\mu, \tau^2)$ from Γ and $r(\pi, \delta_0) = r(\pi_0, \delta_0) = \frac{\tau^2}{1+\tau^2}$. By Theorem 2.1, δ_0 is the Γ-minimax rule. For more details about extensions of this example consult Berger (1985) and Jackson *et al.* (1970).

The following example discusses a case in which the form of the Γ-minimax rule is known in principle. However, effective calculations are possible only for limited parameter spaces.

Example 3. Assume that $X|\theta \sim \mathcal{N}(\theta, 1)$ and the prior for θ belongs to the class of all symmetric and unimodal distributions supported on $[-m, m]$. The form of the Γ-minimax rule in estimating θ is given in Vidakovic (1992) and Vidakovic and DasGupta (1996). The least favorable prior π_0 depends on m. In general, π_0 is a finite linear combination of uniform distributions and the point mass at zero,

$$\pi_0(\theta) = \alpha_0 \mathbf{1}(\theta = 0) + \sum_{i=1}^{n} \frac{\alpha_i}{2m_i} \mathbf{1}(-m_i \leq \theta \leq m_i),$$
$$0 < m_1 < \ldots < m_n = m, \alpha_i \geq 0, \sum_{i=0}^{n} \alpha_i = 1.$$

The marginal density of X is

$$m(x) = \alpha_0 \phi(x) + \sum_{i=1}^{n} \frac{\alpha_i}{2m_i} (\Phi(x + m_i) - \Phi(x - m_i)),$$

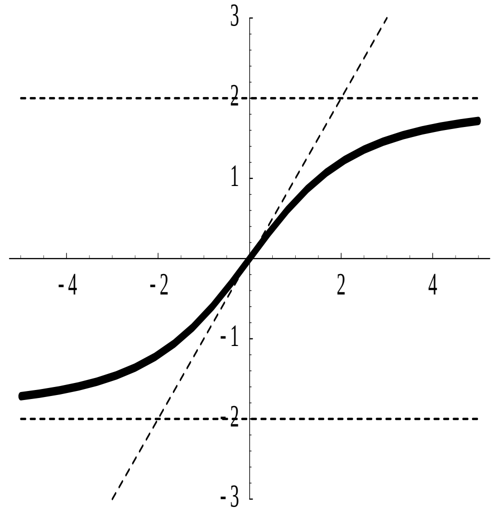

FIGURE 2. Γ-minimax rule for $m = 2$.

and the Bayes rule $\delta_0(x) = x + \frac{m'(x)}{m(x)}$ has the form

$$\delta_0(x) = x - \frac{\alpha_0 x\phi(x) - \sum_{i=1}^{n} \frac{\alpha_i}{2m_i}(\phi(x+m_i) - \phi(x-m_i))}{\alpha_0\phi(x) + \sum_{i=1}^{n} \frac{\alpha_i}{2m_i}(\Phi(x+m_i) - \Phi(x-m_i))}$$
$$(= \delta(x; 0, \alpha_0, m_1, \alpha_1, \dots, m_n, \alpha_n)).$$

For instance, for $m \le m_0 = 2.532258$,[1] the least favorable prior is uniform $\mathcal{U}[-m, m]$, the Γ-minimax rule is

$$\delta_0(x) = x + \frac{\phi(x+m) - \phi(x-m)}{\Phi(x+m) - \Phi(x-m)}, \tag{1.12}$$

and Γ-minimax risk is given by

$$r_\Gamma = 1 - \frac{1}{m}\int_0^\infty \frac{(\phi(x+m) - \phi(x-m))^2}{\Phi(x+m) - \Phi(x-m)} dx.$$

Figure 2 depicts the Γ-minimax rule for $m = 2$, for which the uniform prior $\mathcal{U}[-2, 2]$ is the least favorable.

[1]This constant has been obtained independently by Eichenauer-Herrmann and Ickstadt (1992).

3 Constrained Rules in Gamma Minimax Estimation

It is often beneficial to the statistician to restrict the class of all decision rules to a particular subset of all decision rules. Why would the statistician restrict his choices in the inference process? There are two reasons. Although unrestricted Γ-minimax rules may be difficult or even impossible to find, finding restricted Γ-minimax rules is usually straightforward. Furthermore, restricted rules are usually simple which means that their implementation is easy.

Let $\mathcal{D}_L \subset \mathcal{D}$ be a subclass of all decision rules, let

$$\delta_L^* = \arg \inf_{\delta \in \mathcal{D}_L} (\sup_{\pi \in \Gamma} r(\pi, \delta)) \tag{1.13}$$

be the *restricted* Γ*-minimax rule*, and let

$$\bar{r}_L = \inf_{\delta \in \mathcal{D}_L} \sup_{\pi \in \Gamma} r(\pi, \delta) \tag{1.14}$$

be the Γ*-minimax risk of the restricted rule.*

The most common subclass $\mathcal{D}_L$ is the class of all linear decision rules; that is, rules of the form $\delta_L(x) = ax + b$. In the multivariate case the analogy is the class of all affine rules.

There are several measures for comparing restricted and unrestricted rules. The measure defined as a ratio of Γ-minimax risks of restricted and unrestricted rules,

$$\rho = \frac{r_L}{r_\Gamma}, \tag{1.15}$$

was investigated by Donoho, Liu and MacGibbon (1990) in the context of estimating a normal mean. The restricted rule δ_L is "good" and can be used instead of an unrestricted one if ρ is "close" to one.

For example, Vidakovic and DasGupta (1996) show that, under the model discussed in Example 3, the loss of efficiency due to the use of linear rules is at most 7.4% uniformly over m ($\rho \leq 1.074$). This efficiency improves even more if one considers n-degree polynomial rules. In the example that follows, we give the form of a Γ-minimax polynomial rule and work out details for the cubic case.

Example 4. Let $\mathcal{D}_n$ denote the class of all polynomial rules of the form

$$\delta_n(x) = \sum_{i=0}^{n} a_i x^i, n \in \mathbf{N}.$$

Exact Γ-minimax rules can be approximated arbitrarily well by Γ-minimax polynomial rules. It is straightforward to show that the polynomial Γ-minimax rules are skewed symmetric; that is, $a_0 = a_2 = \cdots = a_{2k} = 0$ for

$n = 2k+1$. Define $\mathbf{a} = (a_1, a_3, \ldots, a_{2k+1})'$ and $\mathbf{y} = (x, x^3, x^5, \ldots, x^{2k+1})'$. The frequentist risk of $\delta_n(x) = \mathbf{a}'\mathbf{y}$ is

$$R(\theta, \delta_n(x)) = (\theta - \mathbf{a}'E\mathbf{y})^2 + \mathbf{a}'\Sigma\mathbf{a} \tag{1.16}$$

where $\Sigma = Cov(\mathbf{y}, \mathbf{y})$. The quantities $E\mathbf{y}$ and Σ can be expressed through Chebyshev-Hermite-like polynomials of θ. Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $D = \frac{d}{dx}$. The polynomials defined as

$$H_n(x) = \frac{(-D)^n \phi(x)}{\phi(x)}$$

are the standard Chebyshev-Hermite polynomials. Let $Q_n(x) = \frac{1}{i^n} H_n(ix)$ and $t_k = Q_k(\theta)$. Then,

$$E\mathbf{y} = (t_1, t_3, \ldots, t_{2k+1})'$$

and

$$\Sigma = \begin{pmatrix} t_2 - t_1^2 & t_4 - t_1 t_3 & \ldots & t_{2k+2} - t_1 t_{2k+1} \\ t_4 - t_3 t_1 & t_6 - t_3^2 & \ldots & t_{2k+4} - t_3 t_{2k+1} \\ \vdots & \vdots & \ddots & \vdots \\ t_{2k+2} - t_{2k+1} t_1 & t_{2k+4} - t_{2k+1} t_3 & \ldots & t_{4k+2} - t_{2k+1}^2 \end{pmatrix}.$$

The frequentist risk needed for finding the Γ-minimax solution is obtained by simplifying (1.16):

$$R(\theta, \mathbf{a}'\mathbf{y}) = \sum_{i \in \mathbf{O}_k} a_i^2 t_{2i} + 2 \sum\sum_{i,j \in \mathbf{O}_k, i<j} a_i a_j t_{i+j} - 2\theta \sum_{i \in \mathbf{O}_k} a_i t_i + \theta^2 \tag{1.17}$$

where $\mathbf{O}_k = \{1, 3, 5, \ldots, 2k+1\}$.

We elaborate on the case $n = 3$. Indeed, larger values of odd n differ from the case $n = 3$ only by the complexity of calculation.

In finding the minimax solution we can interchange the *sup* and *inf*, because the corresponding statistical game can be formulated as a finite S-game and therefore has a value, see Berger (1985). Thus, to find the polynomial Γ-minimax rule we first minimize $E^\pi R(\theta, \mathbf{a}'\mathbf{y})$ with respect to $\mathbf{a}$, for fixed π, and then maximize this minimum with respect to $\pi \in \Gamma$. By standard moment theory, for a fixed k, the least favorable distributions are linear combinations of at most $k+2$ uniform distributions, if we consider a point mass at zero as a degenerate uniform distribution.

Theorem 3.1

$$\inf_{\delta \in \mathcal{D}_n} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{0 \le p_1, p_2, p_3 \le 1} 2BDE - CD^2 - AE^2 + F, \tag{1.18}$$

where

$$\begin{aligned} A &= \frac{1}{3}m^2\nu_1 + 1, \\ B &= \frac{1}{5}m^4\nu_2 + 2m^2\nu_1 + 3, \\ C &= \frac{1}{7}m^6\nu_3 + 3m^4\nu_2 + 15m^2\nu_1 + 15, \\ D &= \frac{1}{3}m^2\nu_1, \\ E &= \frac{1}{5}m^4\nu_2 + m^2\nu_1, \quad \textit{and} \\ F &= \frac{1}{3}m^2\nu_1, \end{aligned}$$

and $\nu_i = \nu_i(p_1, p_2, p_3)$ *are first three canonical moments:* $\nu_1 = p_1$, $\nu_2 = p_1(p_1+q_1p_2)$, *and* $\nu_2 = p_1(p_1(p_1+q_1p_2)+q_1p_2(p_1+q_1p_2+q_2p_3))$, $q_i = 1-p_i$.

Proof: Let

$$\delta_3(x) = (a_1, a_3)\begin{pmatrix} x \\ x^3 \end{pmatrix} = a_1x + a_3x^3.$$

Then

$$\begin{aligned} R(\theta, \delta_3(x)) = \left(\theta - \mathbf{a}'\begin{pmatrix} \theta \\ \theta^3 + 3\theta \end{pmatrix}\right)^2 + \mathbf{a}'\begin{pmatrix} 1 & 3\theta^2+3 \\ 3\theta^2+3 & 9\theta^4+36\theta^2+15 \end{pmatrix}\mathbf{a} \\ = a_1^2(\theta^2+1) + 2a_1a_3(\theta^4+6\theta^2+3) + a_3^2(\theta^6+15\theta^4+45\theta^2+15) \\ -2a_1\theta^2 - 2a_3(\theta^4+3\theta^2) + \theta^2. \end{aligned}$$

If we take the expectation of $R(\theta, \delta_3(x))$ with respect to θ, and use the representation [2] $\theta = U \cdot Z$ to replace $E\theta^n$ with $\frac{1}{n+1}EZ^n$, we get (1.18).

If we minimize $r(\pi, \delta_3)$ with respect to a_1 and a_3 first, then by standard calculus arguments the minimum

$$r(\pi, \delta_3^\star) = 2BDE - CD^2 - AE^2 + F,$$

is achieved for the rule $\delta_3^\star = a_1^\star x + a_3^\star x^3$, where

$$a_1^\star = \frac{DC - BE}{AC - B^2} \quad \text{and} \quad a_3^\star = \frac{AE - BD}{AC - B^2}.$$

To maximize $r(\pi, \delta_3^\star)$ with respect to the moments ν_1, ν_2, and ν_3, the *canonical moments* (see Skibinsky, 1968) are employed. After expressing

[2] Any symmetric and unimodal random variable θ supported on $[-m, m]$ is equal in distribution to $U \cdot Z$, where U is uniform on $[-1, 1]$, and Z is the corresponding random variable supported on $[0, m]$.

the ν_i's through the canonical moments p_1, p_2 and p_3, the original extremal moment problem with complex boundary conditions is transformed to an equivalent problem where the boundary conditions are independent and simple. In fact, the maximization over the unit cube $[0,1] \times [0,1] \times [0,1]$ is performed.

The numerical maximization (Fortran IMSL routine for constrained maximization DBCONF) used suggests that only two types of distributions can be least favorable in the cubic Γ-minimax problem. For $m < 2.7599$, the maximizing p_1 is equal to 1, which corresponds (regardless of values for p_2 and p_3) to the uniform $\mathcal{U}[-m,m]$ distribution on θ. Some selected values of $m < 2.7599$, and the corresponding values of $a_1^\star, a_3^\star, r_C, \frac{r_C}{r_\Gamma}$ (where r_C is the Γ-minimax risk of the cubic rule $\delta_3^\star$), are given in Table 1.1.

TABLE 1.1. Values of coefficients for the cubic rule and corresponding ρ.

m	$a_1^\star$	$a_3^\star$	r_C	r_C/r_Γ
0.3	0.02962	-0.00016	0.029126	1
0.5	0.08023	-0.00102	0.076915	1
1	0.28032	-0.00777	0.24922	1.0001
1.5	0.50624	-0.01597	0.42241	1.0014
2	0.69311	-0.02	0.55315	1.0035
2.5	0.82672	-0.02	0.64193	1.0047
2.7	0.86704	0.01928	0.66862	1.0038

If $m \geq 2.7599$, then the maximizing p_1 is strictly less than 1, p_2 is equal to 1, and p_3 is arbitrary. This corresponds to the least favorable distribution on θ which is a linear combination of uniforms $\mathcal{U}[-m,m]$ and a point mass at zero, namely $\pi_0(\theta) = \alpha\delta(\{0\}) + (1-\alpha)\frac{1}{2m}\mathbf{1}(-m \leq \theta \leq m)$. Table 1.2 gives the risks and the efficiency in this case.

4 Conditional Gamma Minimax and Gamma Minimax Regret

Conditional Γ-minimax had been at hinted by Watson (1974), but was first explored in a particular context by DasGupta and Studden (1989). As the name suggest, conditional on observations, the statistician is interested in devising an action that minimizes the payoff expressed in terms of posterior expected loss. For an action a, the posterior expected loss is $\rho(\pi, a) = E^{\pi^*} L(\theta, a)$. For any action a, let π_a be a density such that $\rho(\pi_a, a) = \sup_\pi \rho(\pi, a)$. Such π_a (not necessarily unique) is called a *least favorable prior.* An action a^* that minimizes the supremum of posterior expected loss is a *conditional* Γ*-minimax action.* Existence of conditional Γ-minimax

TABLE 1.2. Continuation of the previous table when $m \geq 2.7599$.

m	$a_1^\star$	$a_3^\star$	$\alpha = 1 - p_1$	r_C	r_C/r_Γ
2.8	0.87906	-0.01845	0.01011	0.68054	1.0027
3	0.88499	-0.01581	0.05447	0.70300	1.0015
3.5	0.89928	-0.01080	0.13326	0.75217	1.0043
4	0.91228	-0.00748	0.18379	0.79215	1.0087
m	$a_1^\star$	$a_3^\star$	$\alpha = 1 - p_1$	r_C	$r_C/r_\Gamma <$
5	0.93347	-0.00379	0.24253	0.85037	1.0151
6	0.94885	-0.00207	0.27412	0.88861	1.0192
8	0.96793	-0.00075	0.30535	0.93256	1.0228
10	0.97835	-0.00033	0.31976	0.95526	1.0177
12	0.98451	-0.00016	0.32758	0.96831	1.0147
15	0.98984	-0.00007	0.33398	0.97938	1.0116
20	0.99417	-0.00002	0.33895	0.98825	1.0094
50	0.99905	-5.82 10^{-7}	0.34428	0.99809	1.0020
100	0.99976	-3.65 10^{-8}	0.34513	0.99952	~ 1

actions for general models/losses was explored in Betrò and Ruggeri (1992).

Although this approach is more in the Bayesian spirit, there is an abundance of examples in which the conditional Γ-minimax actions fail to be Bayes actions with respect to any single prior in the class Γ. We give an example by Watson (1974). For more interesting theory and examples about conditional Γ-minimax, we direct the reader to Betrò and Ruggeri (1992) and DasGupta and Studden (1989). The stability of conditional Γ-minimax actions is explored in the work of Boratyńska (1997), Męczarski (1993, 1998), and Męczarski and Zieliński (1991).

Example 5. Consider estimating the parameter θ of a Poisson process, and suppose that prior is known to belong to the family Γ of gamma distributions,

$$\pi(\theta) \in \left\{ \frac{m^r \theta^{r-1}}{(r-1)!} e^{-\theta m}, r \text{ known}, m \in [m_L, m^U] \right\}.$$

Assume that the loss is weighted squared error, $L(\theta, a) = (\theta - a)^2/\theta$. If the process is observed up to time T and n events have been recorded, the Γ minimax action in estimating θ is

$$a_{GM} = \sqrt{\frac{(n+r-1)(n+r)}{(m_L+T)(m_U+T)}}.$$

However, for a particular prior in Γ (indexed by m) the Bayes action is

$$a_B = \frac{n+r-1}{m+T},$$

and clearly no such value of m exist for which $a_{GM} = a_B$ for all values of n.

DasGupta and Studden (1989) discuss conditional Γ minimax actions for the squared error loss, multivariate normal model, and multivariate normal prior. We discuss one of their interesting results. Let $\mathbf{y} \sim \mathcal{MVN}_p(\theta, \Sigma_0)$ be a multivariate normal distribution where θ is unknown and Σ_0 is a known positive definite matrix.

Let the distribution for θ belong to the class

$$\Gamma = \{\pi|\ \pi \sim \mathcal{N}(\mu, \Sigma_0), \mu \text{ fixed}, \Sigma_1 \le \Sigma \le \Sigma_2\},$$

where the matrix inequality $A \le B$ means that $B - A$ is a non-negative definite matrix. For a given c, suppose that $c'\theta$ is the parameter of interest. Let $S(c)$ be the set of two-dimensional vectors consisting of the posterior mean and posterior variance,

$$S(c) = \{(E(c'\theta|\mathbf{y}), Var(c'\theta|\mathbf{y})|\pi \in \Gamma\}.$$

It is curious that the set $S(c)$ forms an ellipse

$$\{u|\ (u-u_0)'D^{-1}(u-u_0) \le 1\}$$

where $u_0 = (c'\mu + c'\bar{\Lambda}v, c'\bar{\Lambda}c)'$ and

$$D = A^2 \begin{pmatrix} v'(\Lambda_2 - \Lambda_1)v & c'(\Lambda_2 - \Lambda_1)v \\ & c'(\Lambda_2 - \Lambda_1)c \end{pmatrix},$$

with $\Lambda_i = (\Sigma_0^{-1} + \Sigma_i^{-1}),\ i = 1, 2;\ \bar{\Lambda} = (\Lambda_1 + \Lambda_2)/2,\ v = \Sigma_0^{-1}(\mathbf{y} - \mu)$, and $A^2 = \frac{1}{4}c'(\Lambda_2 - \Lambda_1)c$.

The conditional Γ-minimax action in estimating $c'\theta$ is

$$a^* = \begin{cases} c'(\Lambda_2 v + \mu) & \text{if } v'(\Lambda_2 - \Lambda_1)v \le 1 \\ c'(\Lambda^* v + \mu) & \text{if } v'(\Lambda_2 - \Lambda_1)v > 1 \end{cases}, \tag{1.19}$$

where $\Lambda^* = \bar{\Lambda} + \frac{\Lambda_2 - \Lambda_1}{2v'(\Lambda_2 - \Lambda_1)v}$.

It is interesting that

$$\theta^* = \begin{cases} (\Lambda_2 v + \mu) & \text{if } v'(\Lambda_2 - \Lambda_1)v \le 1 \\ (\Lambda^* v + \mu) & \text{if } v'(\Lambda_2 - \Lambda_1)v > 1 \end{cases}$$

is *not* a conditional Γ-minimax action for θ, although for each c, (1.19) is a conditional Γ-minimax estimator of $c'\theta$.

However, DasGupta and Studden show that θ^* is a conditional Γ-minimax action for θ if $\Lambda_2 - \Lambda_1$ is a multiple of the identity matrix, I. For instance, if $\mu = 0$, $\Sigma_0 = I$, $\Sigma_1 = \alpha I$, and $\Sigma_2 = \infty I$ $(\Sigma \ge \alpha I)$, the conditional Γ-minimax action is

$$\theta^* = \begin{cases} \mathbf{y} & \text{if } \mathbf{y}'\mathbf{y} \le \alpha + 1 \\ y - \frac{1}{2(\alpha+1)}\left(1 - \frac{\alpha+1}{\mathbf{y}'\mathbf{y}}\right)\mathbf{y} & \text{if } \mathbf{y}'\mathbf{y} > \alpha + 1. \end{cases}$$

This estimator resembles a shrinkage James-Stein-type estimator as well as some restricted-risk Bayesian estimates; see Berger (1982) and DasGupta and Rubin (1987).

There is another interesting example from Betrò and Ruggeri (1992).

Example 6. Assume $X = 1$ was observed from the model $f(x|\theta) = \theta^x(1-\theta)^{1-x}$, and Γ has two distributions only, $\pi_1(\theta) = \mathbf{1}(0 \le \theta \le 1)$ and $\pi_2(\theta) = \frac{3}{2}\mathbf{1}(0 \le \theta \le \frac{1}{2}) + \frac{1}{2}\mathbf{1}(\frac{1}{2} < \theta \le 1)$. The posterior expected losses for an action a are $\rho(\pi_1, a) = a^2 - \frac{4}{3}a + \frac{1}{2}$ and $\rho(\pi_2, a) = a^2 - \frac{10}{9}a + \frac{3}{8}$. Thus $a^\star = \frac{9}{16}$ is the conditional Γ-minimax action.

It is curious that the Γ-minimax rule $\delta^*(x) = \frac{x+1}{3}$ does not produce $\frac{9}{16}$ for $X = 1$.

4.1 Conditional Gamma Minimax Regret

Regret-type rules in decision theory have been time honored; see, for example, the discussion in Berger (1984, 1985). In the conditional context, Γ-minimax regret rules had been introduced by Zen and DasGupta (1993).

As before, let X be an observation from a distribution P_θ with density $p_\theta(x)$, indexed by the parameter $\theta \in \Theta$. Suppose that θ has a prior distribution with density $\pi(\theta)$. Let $\mathcal{A}$ be the action space, $L(\theta, a)$ the loss if the action $a \in \mathcal{A}$ is adopted, and θ the state of nature. Let $\pi_x(\theta)$ be the posterior density when x is observed, and $\rho(\pi_x, a)$ the posterior expected loss of a.

The *posterior regret* of an action a is

$$d(\pi_x, a) = \rho(\pi_x, a) - \rho(\pi_x, a_{\pi_x}),$$

where a_{π_x} is an action minimizing $\rho(\pi_x, a)$. Informally, d measures the loss of optimality due to choosing a instead of the optimal action a_{π_x}.

Definition 4.1 *$a_M \in \mathcal{A}$ is the posterior regret Γ-minimax (PRGM) action if*

$$\inf_{a\in\mathcal{A}} \sup_{\pi\in\Gamma} d(\pi_x, a) = \sup_{\pi\in\Gamma} d(\pi_x, a_M). \tag{1.20}$$

Suppose that $\Theta = \mathcal{A} \subset \mathbf{R}$ is an interval, and the loss function is $L(\theta, a) = (\theta - a)^2$. It is easy to see that

$$r(\pi_x, a) = (a - a_{\pi_x})^2,$$

where a_{π_x} is the posterior expected value of π_x. The following proposition provides a guide for finding the posterior regret action.

Proposition 1 *Let* $\underline{a} = \inf_{\pi\in\Gamma} a_{\pi_x}$ *and* $\bar{a} = \sup_{\pi\in\Gamma} a_{\pi_x}$ *be finite. Then*

$$a_M = \frac{1}{2}(\underline{a} + \bar{a}). \tag{1.21}$$

Thus, besides its heuristic appeal, computing a_M is simple provided that we have procedures to compute the range of posterior expectations; see Berliner and Goel (1990), DasGupta and Bose (1988), Rios *et al.,* (1995), Sivaganesan and Berger (1989), and Zen and DasGupta (1993), for details.

In the following example of Zen and DasGupta (1993), it is demonstrated how the PRGM action changes with enrichment of the class Γ.

Example 7. Let $X|\theta \sim \mathcal{B}in(n,p)$ and suppose the prior on θ belongs to
(i) $\Gamma_1 = \{\mathcal{B}eta(\alpha,\alpha),\ \alpha_1 \le \alpha \le \alpha_2\}$,
(ii) $\Gamma_2 = \Gamma_{SU}[0,1]$ (symmetric, unimodal, supported on [0,1]), or
(iii) $\Gamma_3 = \Gamma_S[0,1]$ (symmetric, supported on [0,1]).
In case (i), the PRGM action is

$$\delta_1(X) = \frac{X + \frac{n(\alpha_1+\alpha_2)+4\alpha_1\alpha_2}{2(n+\alpha_1+\alpha_2)}}{n + \frac{n(\alpha_1+\alpha_2)+4\alpha_1\alpha_2}{n+\alpha_1+\alpha_2}}.$$

Hence, δ_1 is Bayes with respect to $\pi^*(\theta) = \text{Beta}(\alpha^*,\alpha^*)$, where $\alpha^* = \frac{n(\alpha_1+\alpha_2)+4\alpha_1\alpha_2}{n+\alpha_1+\alpha_2}$. Note that $\pi^*(\theta) \in \Gamma_1$ and $\alpha^* \to \frac{\alpha_1+\alpha_2}{2}$, when $n \to \infty$.

If $\theta \sim \pi \in \Gamma_2$, the PRGM action is

$$\delta_2(X) = \frac{X + \frac{n}{2} + 2}{2n+4},$$

which is Bayes with respect to $\pi^*(\theta) = \text{Beta}(\frac{n}{2}+2, \frac{n}{2}+2) \in \Gamma$.

When $\theta \sim \pi \in \Gamma_3$, the PRGM rule is rather unattractive,

$$\delta_3(X) = \begin{cases} \frac{1}{4}, & 0 \le X < [\frac{n}{2}] \\ \frac{1}{2}, & X = [\frac{n}{2}] \\ \frac{3}{4}, & [\frac{n}{2}] < X \le n \end{cases}.$$

Example 8. Let $X_1,\ldots,X_n \sim \mathcal{N}(\theta,1)$. We seek the PRGM action for θ. Let Γ_S and Γ_{SU} be the families of all symmetric distributions and all symmetric, unimodal distributions on $[-m,m]$, respectively. In deriving the PRGM action, we use the well-known fact that any distribution $\pi \in \Gamma_S$ can be represented as a mixture of symmetric two-point priors, and

$$\sup_{\pi\in\Gamma_S} a_{\pi_x} = \sup_{\pi\in\Gamma'_S} a_{\pi_x},$$
$$\inf_{\pi\in\Gamma_S} a_{\pi_x} = \inf_{\pi\in\Gamma'_S} a_{\pi_x},$$

where

$$\Gamma'_S = \left\{\pi \in \Gamma_S \mid \pi(\{-z\}) = \pi(\{z\}) = \frac{1}{2},\ z \in [0,m]\right\}.$$

Then, for $\theta \sim \pi \in \Gamma_S$, the PRGM action is

$$a_M = \frac{m}{2} \tanh mn\bar{X}.$$

Consider now $\pi \in \Gamma_{SU}[-m, m]$. Here,

$$\sup_{\pi \in \Gamma_{SU}} a_{\pi_x} = \sup_{0 \leq z \leq m} f(z, x),$$
$$\inf_{\pi \in \Gamma_{SU}} a_{\pi_x} = \inf_{0 \leq z \leq m} f(z, x),$$

where

$$f(z, x) = \frac{\int_{-z}^{z} \theta \phi_{1/n}(x - \theta) d\theta}{\int_{-z}^{z} \phi_{1/n}(x - \theta) d\theta},$$

and $\Phi_{1/n}(x)$ and $\phi_{1/n}(x)$ denote the cdf and pdf of the normal $\mathcal{N}(0, \frac{1}{n})$ law.

For $\theta \sim \pi \in \Gamma_{SU}$, the PRGM action is

$$a_M = \frac{\bar{X}}{2} + \frac{1}{2n} \frac{\phi_{1/n}(\bar{X} + m) - \phi_{1/n}(\bar{X} - m)}{\Phi_{1/n}(\bar{X} + m) - \Phi_{1/n}(\bar{X} - m)}.$$

5 Conclusions

In this chapter we gave an overview of the Γ-minimax inference standpoint as a way to select a robust action/rule from a multitude of possible rules. When the problem is "regular", i.e., when the *sup* and *inf* can interchange places in the game theoretic formulation of the problem, the selected action becomes Bayes with respect to a *least favorable prior.* We also provided a list of examples in connection with the different flavors of the Γ-minimax formulation, namely standard, conditional, regret, and restricted.

Acknowledgment. The author thanks Editors for their kind invitation and for providing some recent references. Constructive comments of Anirban DasGupta, Michael Kozdron, and an anonymous referee improved the presentation.

References

Berger, J. (1982). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. In *Statistical Decision Theory and Related Topics III,* **1**, 109–141, Academic Press, New York.

BERGER, J. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analyses*, Ed. J. Kadane, Elsevier Science Publishers, 63–124.

BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis, Second Edition*. Springer-Verlag, New York.

BERGER, J. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Jour, Statist. Planning and Inf.,* **25**, *303–328.*

BERGER, J. (1994). An overview of robust Bayesian analysis (Disc. p 59–124). *Test*, **1**, 5–59.

BERGER, R. (1979). Gamma minimax robustness of Bayes rules. *Communications in Statistics, Part A – Theory and Methods*, **8**, 543–560.

BERLINER, M. and GOEL, P. (1990). Incorporating partial prior information: Ranges of posterior probabilities. In *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, 397–406.

BETRÒ, B. and RUGGERI, F. (1992). Conditional Γ-minimax actions under convex losses. *Communications in Statistics, Part A – Theory and Methods*, **21**, 1051–1066.

BORATYŃSKA, A. (1997). Stability of Bayesian inference in exponential families. *Statistics & Probability Letters*, **36**, 173–178.

DASGUPTA, A. and BOSE, A. (1988). Γ-minimax and restricted-risk Bayes estimation of multiple Poisson means under ϵ-contaminations of the subjective prior. *Statistics & Decisions*, **6**, 311–341.

DASGUPTA, A. and RUBIN, H. (1987). Bayesian estimation subject to minimaxity of the mean of a multivariate normal distribution in the case of a common unknown variance. In *Statistical Decision Theory and Related Topics IV,* S. S. Gupta and J. O. Berger eds. Volume 1, 325–345.

DASGUPTA, A. and STUDDEN, W. (1989). Frequentist behavior of robust Bayes estimates of normal means. *Statistics and Decisions,* **7**, 333–361.

DONOHO, D., LIU, R., and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.,* **18**, *1416–1437.*

EICHENAUER-HERRMANN, J. and ICKSTADT, K. (1992). Minimax estimators for a bounded location parameter. *Metrika*, **39**, 227–237.

EFRON, B. and MORRIS, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators - Part I: The Bayes case. *J. Amer. Statist. Assoc.,* **66,** *807–815.*

ELLSBERG, D. (1954). Classic and current notions of "measurable utility." *Economic Journal,* LXIV, 528–556.

GEORGE, S. (1969). Partial prior information: Some empirical Bayes and G-minimax decision functions. Ph. D. thesis, Southern Methodist University.

GOOD, I. J. (1952). Rational decisions. *J. Royal Statist. Soc.* (Ser. B), **14,** 107–114.

JACKSON, D., O'DONOVAN, T., ZIMMER, W., and DEELY, J (1970). $\mathcal{G}_2$-minimax estimators in the exponential family. *Biometrika,* **57**, 439–443.

MĘCZARSKI, M. (1993). Stability and conditional gamma-minimaxity in Bayesian inference. *Applicationes Mathematicae*, **22**, 117–122.

MĘCZARSKI, M. (1998). *Robustness Problems in Bayesian Statistical Analysis* (in Polish). Monograph Series No. 446, Publishing House of Warsaw School of Economics, Warszawa.

MĘCZARSKI, M. AND ZIELIŃSKI, R. (1991). Stability of the Bayesian estimator of the Poisson mean under the inexactly specified Gamma prior. *Statistics & Probability Letters,* **12**, 329–333.

RIOS, D. (1990). *Sensitivity Analysis in Multiobjective Decision Making.* Lecture Notes in Economics and Mathematical Systems, 347, Springer-Verlag.

RIOS, D., RUGGERI, F. AND VIDAKOVIC, B. (1995). Some results on posterior regret Γ-minimax estimation. *Statistics & Decisions,* **13**, 315-331.

ROBBINS, H. (1951). Asymptotically sub-minimax solutions to compound statistical decision problems. In *Proc. Second Berkeley Symp. Math. Stat. Probab.*, **1**, University of California Press, Berkeley.

SKIBINSKY, M. (1968). Extreme nth moments for distributions on $[0, 1]$ and the inverse of a moment space map. *J. Appl. Prob.,* **5,** *693–701.*

SIVAGANESAN, S. and BERGER, J. (1989). Range of posterior measures for priors with unimodal contaminations. *The Annals of Statistics,* **17**, 868–889.

VIDAKOVIC, B. (1992). A study of the properties of computationally simple rules in estimation problems. Ph. D. Thesis, Department of Statistics, Purdue University.

VIDAKOVIC, B. and DASGUPTA, A. (1996). Efficiency of linear rules for estimating a bounded normal mean. *Sankhyā, Series A, Indian Journal of Statistics*, **58**, 81–100.

WATSON, S. R. (1974). On Bayesian inference with incompletely specified prior distributions. *Biometrika*, **61**, 193–196.

ZEN, M. and DASGUPTA, A. (1993). Estimating a binomial parameter: Is robust Bayes real Bayes? *Statistics & Decisions*, **11**, 37–60.

ZIELIŃSKI, R. (1994). Comment on "Robust Bayesian methods in simple ANOVA models." *Journal of Statistical Planning and Inference*, **40**, 308–310.