SAPIENZA
UNIVERSITÀ DI ROMA

Optimal credible intervals under alternative loss functions

Fulvio De Santis & Stefania Gubbiotti

Dipartimento di Scienze Statistiche :: Sapienza Università di Roma

Outline

- ▷ Bayesian interval estimation from a decision-theoretic perspective
- ▷ Focus on the class of monotone loss functions
- ▷ Comparisons between linear, rational, exponential loss functions

Methodology

▷ Assumptions

- $\{f_n(\cdot|\theta),\ \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^1$
- $\pi(\theta)$ prior density for $\theta \Rightarrow \pi(\theta|\mathbf{x}_n)$
- $\mathbb{L}(\theta, C)$ loss function for $C \in \mathcal{C}$

▷ Decision-theoretic approach for set estimation [1, 4]

Select $C^\star$ that minimizes $\rho(C, \mathbf{x}_n) = \int_\Theta \mathbb{L}(\theta, C)\pi(\theta|\mathbf{x}_n)d\theta$.

▷ Class of monotone loss functions

$$\mathbb{L}(\theta, C) = \mathbb{S}[\mathcal{L}(C)] + \mathbb{I}_{\bar{C}}(\theta),$$

- $\mathbb{S}(\cdot)$ size, i.e. increasing function of $\mathcal{L}(C)$
- $\mathcal{L}(C)$ Lebesgue measure of C
- $\mathbb{I}_{\bar{C}}(\cdot)$ indicator function of $\bar{C} = \Theta \setminus C$
- $\mathbb{L}(\theta, C)$ loss function for $C \in \mathcal{C}$

The posterior expected loss $\rho(C, \mathbf{x}_n) = \mathbb{S}[\mathcal{L}(C)] + 1 - \mathbb{P}(C|\mathbf{x}_n)$ is a compromise between size and posterior probability.

If θ is an absolutely continuous r.v., optimal actions are HPD sets $C^\star = \{\theta \in \Theta\ :\ \pi(\theta|\mathbf{x}_n) \geq k\}\ ,\ k \geq 0\}$. (see [3])
Also, assume HPD sets are intervals $C = [L, U] \Rightarrow \mathcal{L}(C) = U - L$

By specifying the size function, different losses are obtained:

- $\mathbb{S}_\ell[\mathcal{L}(C)] = a\mathcal{L}(C), \quad a > 0 \Rightarrow$ **Linear loss**

Paradox: in the case of unbounded parameter space, optimal sets under the linear loss function may be dominated by unreasonable sets

Solution by [2]: pick a nonlinear $\mathbb{S}(\cdot)$ that ranges monotonically in $[0, 1]$

- $\mathbb{S}_e[\mathcal{L}(C)] = 1 - e^{-\frac{a\mathcal{L}(C)^2}{2}} \quad \Rightarrow$ **Exponential loss**
- $\mathbb{S}_r[\mathcal{L}(C)] = \frac{a\mathcal{L}(C)}{a\mathcal{L}(C)+1} \quad \Rightarrow$ **Rational loss**

Note: the role of a is not equivalent in the different size functions $\mathbb{S}_j$

Posterior comparison of loss functions

Posterior expected losses

$$\rho_j(C, \mathbf{x}_n) = \mathbb{S}_j[\mathcal{L}(C)] + 1 - \mathbb{P}(C|\mathbf{x}_n), \qquad j = \ell, e, r,$$

as functions of $\mathbb{P}(C|\mathbf{x}_n)$ are compared for different values of a, under the following assumptions:

- $X_i|\theta \sim$ Pois(θ), $i = 1, \ldots, n$ (i.i.d.), $\theta > 0$
- $\theta \sim$ Ga(α, β), $\alpha, \beta > 0$.
- $\theta|\mathbf{x}_n \sim$ Ga$(\bar{\alpha}, \bar{\beta})$, $\bar{\alpha} = \alpha + s_n$, $\bar{\beta} = \beta + n$, $s_n = \sum_{i=1}^n x_i$.

Figure 1: $\rho_j(C, \mathbf{x}_n)$ for Gamma posteriors $(\bar{\alpha}, \bar{\beta}) = (6, 6)$ (left column) and $(\bar{\alpha}, \bar{\beta}) = (14, 6)$ (right column). For each ρ_j, $j = \ell, r, e$, circles denote $\mathbb{P}(C^\star|\mathbf{x}_n)$, i.e. the posterior probabilities of optimal sets.

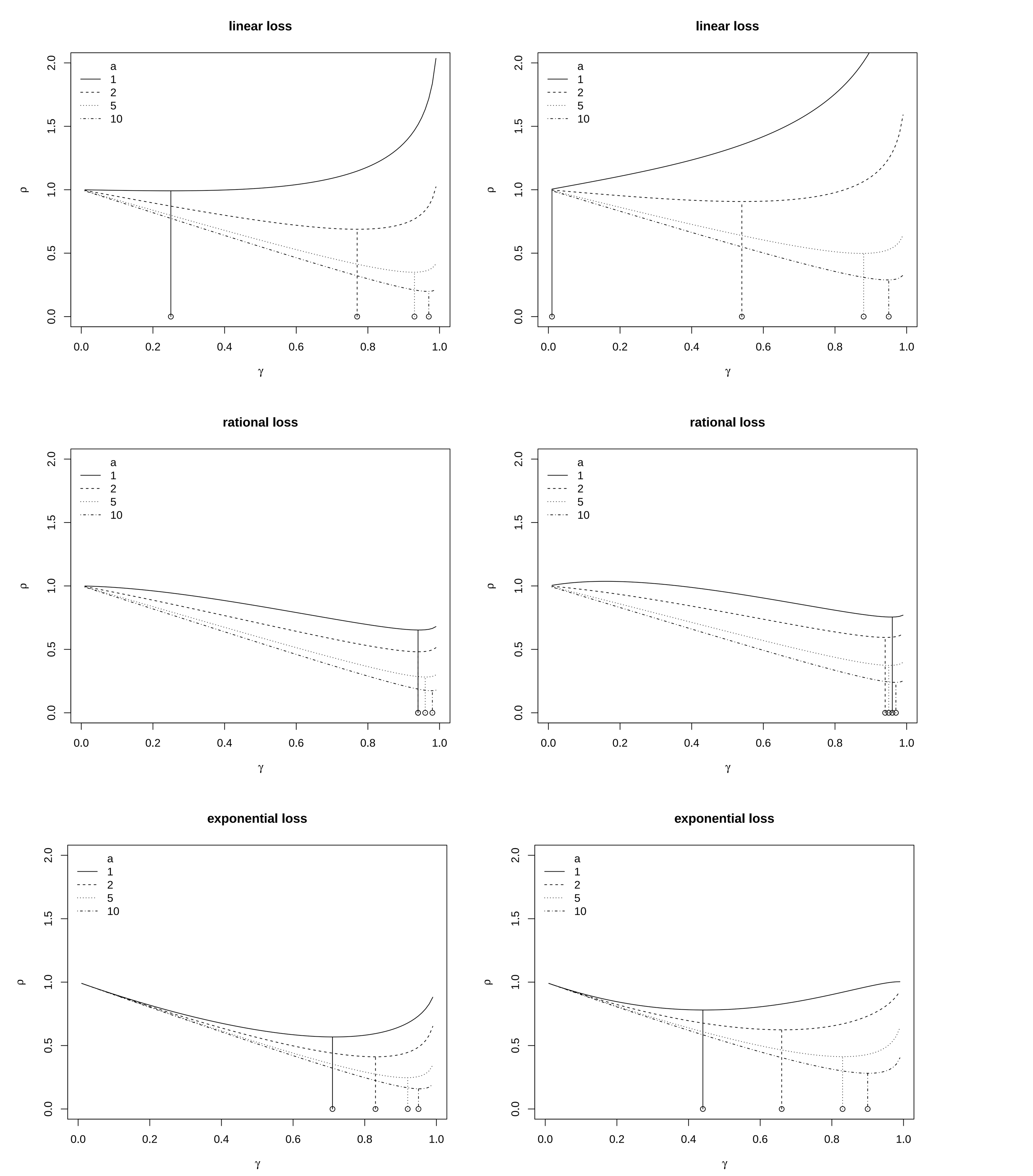

Table 1: Bounds, length, posterior probability and posterior expected loss for $C^\star$ under the three loss functions.

Loss	a	$\mathcal{L}(C^\star)$	$\mathbb{P}(C^\star\|\mathbf{x}_n)$	$\rho_j(C^\star, \mathbf{x}_n)$	$\mathcal{L}(C^\star)$	$\mathbb{P}(C^\star\|\mathbf{x}_n)$	$\rho_j(C^\star, \mathbf{x}_n)$
			$(\bar{\alpha}, \bar{\beta}) = (6, 6)$			$(\bar{\alpha}, \bar{\beta}) = (14, 6)$	
linear	1.0	0.242	0.250	0.992	0.015	0.010	1.005
	0.5	0.917	0.770	0.689	0.895	0.540	0.907
	0.2	1.399	0.930	0.350	1.890	0.880	0.498
	0.1	1.689	0.970	0.199	2.390	0.950	0.289
rational	1.0	1.454	0.940	0.652	2.507	0.960	0.755
	0.5	1.454	0.940	0.481	2.292	0.940	0.594
	0.2	1.594	0.960	0.282	2.390	0.950	0.373
	0.1	1.817	0.980	0.174	2.652	0.970	0.240
exponential	1.0	0.807	0.710	0.568	0.706	0.440	0.780
	0.5	1.051	0.830	0.411	1.156	0.660	0.624
	0.2	1.350	0.920	0.247	1.667	0.830	0.412
	0.1	1.518	0.950	0.159	2.001	0.900	0.281

▷ Remarks

- the larger a the smaller $\gamma^\star$ for linear and exponential loss
- $\rho_\ell(C_\gamma, \mathbf{x}_n)$ is highly sensitive to the values of a
- $\rho_r(C_\gamma, \mathbf{x}_n)$ is excessively robust with respect to a
- the exponential loss represents a sensible trade-off
- the larger a, the smaller $\mathcal{L}(C^\star)$ and $\mathbb{P}(C^\star|\mathbf{x}_n)$, the larger $\rho_j(C^\star, \mathbf{x}_n)$
- this effect is mostly remarkable in the linear loss case
- variations in the values of $\rho_r(C^\star, \mathbf{x}_n)$ depend almost entirely on a
- $\rho_\ell(C^\star, \mathbf{x}_n)$ and $\rho_e(C^\star, \mathbf{x}_n)$ change according to $\mathcal{L}(C^\star)$ and $\mathbb{P}(C^\star|\mathbf{x}_n)$.

References

[1] Berger J. O. Statistical decision theory and Bayesian analysis. New York: Springer-Verlag. Chicago (1985)

[2] Casella G., Hwang J.T.G., Robert C. A paradox in decision-theoretic interval estimation. Statistica Sinica, 3(1), 141–155 (1993a)

[3] Casella G., Hwang J.T.G., Robert C. Loss function for set estimation. In Statistical Decision Theory and related topics V, J.O. Berger and S.S. Gupta (Eds.), 237–252. Springer-Verlag, New York (1993b)

[4] Robert C. The Bayesian choice: from decision-theoretic foundations to computational implementation Springer Science & Business Media (2007)