On the Statistical Consistency of Risk-Sensitive Bayesian Decision-Making

Prateek Jaiswal[*], Harsha Honnappa[†], Vinayak A. Rao[‡]

[*] Department of Statistics, Texas A&M University
[†]School of Industrial Engineering, Purdue University
[*]Department of Statistics, Purdue University

37th Conference on Neural Information Processing Systems (NeurIPS 2023)

Poster Session(Poster #1222)
Where: Great Hall & Hall B1+B2
When: Wed 13 Dec,
10:45 a.m. CST - 12:45 p.m. CST

- In many applications, we must make decisions over model inference (Inventory management, Resource allocation etc.)
- (Recall) Bayes Risk

$$\text{minimize}_{a \in \mathcal{A}} \quad \mathbb{E}_{\pi(\theta|\tilde{X}_n)}[R(a,\theta)]$$

  - $\mathcal{A} \subseteq \mathbb{R}^d$: Decision space
  - $\pi(\theta|\tilde{X}_n)$: Posterior distribution over $\theta \in \Theta$
  - $R(a,\theta)$: Risk function

- We replace the expectation in the Bayes risk to a *log-exponential* or *entropic* risk measure.

$$\text{minimize}_{a \in \mathcal{A}} \overbrace{\varrho^{\gamma}_{\pi_n}(R(a,\theta))}^{\text{Entropic Risk}} := \frac{1}{\gamma} \log \mathbb{E}_{\pi_n}[\exp(\gamma \overbrace{R(a,\theta)}^{\text{Loss/Risk}})] \tag{1}$$

where $\gamma > 0$ is the risk-sensitivity parameter and $\pi_n \equiv \pi(\theta|\tilde{X}_n)$.

- In many applications, we must make decisions over model inference (Inventory management, Resource allocation etc.)
- (Recall) Bayes Risk

$$\text{minimize}_{a \in \mathcal{A}} \quad \mathbb{E}_{\pi(\theta|\tilde{X}_n)}[R(a,\theta)]$$

  - $\mathcal{A} \subseteq \mathbb{R}^d$: Decision space
  - $\pi(\theta|\tilde{X}_n)$: Posterior distribution over $\theta \in \Theta$
  - $R(a,\theta)$: Risk function

- We replace the expectation in the Bayes risk to a *log-exponential* or *entropic* risk measure.

$$\text{minimize}_{a \in \mathcal{A}} \overbrace{\varrho^{\gamma}_{\pi_n}(R(a,\theta))}^{\text{Entropic Risk}} := \frac{1}{\gamma} \log \mathbb{E}_{\pi_n}[\exp(\gamma \overbrace{R(a,\theta)}^{\text{Loss/Risk}})] \tag{1}$$

where $\gamma > 0$ is the risk-sensitivity parameter and $\pi_n \equiv \pi(\theta|\tilde{X}_n)$.

- In many applications, we must make decisions over model inference (Inventory management, Resource allocation etc.)
- (Recall) Bayes Risk

$$\text{minimize}_{a \in \mathcal{A}} \quad \mathbb{E}_{\pi(\theta|\tilde{X}_n)}[R(a, \theta)]$$

  - $\mathcal{A} \subseteq \mathbb{R}^d$: Decision space
  - $\pi(\theta|\tilde{X}_n)$: Posterior distribution over $\theta \in \Theta$
  - $R(a, \theta)$: Risk function

- We replace the expectation in the Bayes risk to a *log-exponential* or *entropic* risk measure.

$$\text{minimize}_{a \in \mathcal{A}} \overbrace{\varrho^{\gamma}_{\pi_n}(R(a, \theta))}^{\text{Entropic Risk}} := \frac{1}{\gamma} \log \mathbb{E}_{\pi_n}[\exp(\gamma \overbrace{R(a, \theta)}^{\text{Loss/Risk}})] \tag{1}$$

where $\gamma > 0$ is the risk-sensitivity parameter and $\pi_n \equiv \pi(\theta|\tilde{X}_n)$.

- *Entropic* risk measure captures higher moments, and thus tail effects.
- γ encodes the risk sensitivity of the decision maker.

$$\underbrace{\lim_{\gamma\to 0^+} \varrho^{\gamma}_{\pi_n}(R(a,\theta)) = \mathbb{E}_{\pi_n}[R(a,\theta)]}_{\text{Risk Neutral}} \quad \underbrace{\lim_{\gamma\to\infty} \varrho^{\gamma}_{\pi_n}(R(a,\theta)) = \text{ess-sup}_{\pi_n} R(a,\theta)}_{\text{Completely Risk Averse}}$$

Recall that the posterior distribution is intractable in general

- *Entropic* risk measure captures higher moments, and thus tail effects.
- γ encodes the risk sensitivity of the decision maker.

$$\overbrace{\lim_{\gamma \to 0^+} \varrho^{\gamma}_{\pi_n}(R(a,\theta)) = \mathbb{E}_{\pi_n}[R(a,\theta)]}^{\text{Risk Neutral}} \quad \overbrace{\lim_{\gamma \to \infty} \varrho^{\gamma}_{\pi_n}(R(a,\theta)) = \textsf{ess-sup}_{\pi_n} R(a,\theta)}^{\text{Completely Risk Averse}}$$

Recall that the posterior distribution is intractable in general

Risk-Sensitive Variational Bayes (RSVB)

For any $\gamma > 0$ and $a \in \mathcal{A}$ [Using Donsker-Varadhan variational lemma]

$$\frac{1}{\gamma}\log \mathbb{E}_{\pi_n}\left[\exp(\gamma R(a,\theta))\right] = \max_{q\in\mathcal{M}} \underbrace{\left[\mathbb{E}_q[R(a,\theta)] - \frac{1}{\gamma}\mathrm{KL}(q\|\pi_n)\right]}_{=:\mathcal{F}(a;q,\tilde{X}_n,\gamma)},$$

$$\Rightarrow \frac{1}{\gamma}\log \mathbb{E}_{\pi_n}\left[\exp(\gamma R(a,\theta))\right] \geq \max_{q\in\mathcal{Q}} \mathcal{F}(a;q,\tilde{X}_n,\gamma)$$

RSVB decision rule

$$\mathsf{a}^*_{\mathrm{RS}} \equiv \mathsf{a}^*_{\mathrm{RS}}(\gamma,\tilde{X}_n) := \mathsf{argmin}_{a\in\mathcal{A}} \max_{q\in\mathcal{Q}} \mathcal{F}(a;q,\tilde{X}_n,\gamma)$$

RSVB posterior (for any $a' \in \mathcal{A}$)

$$q^*_{a',\gamma}(\theta|\tilde{X}_n) \in \mathsf{argmax}_{q\in\mathcal{Q}}\mathcal{F}(a';q,\tilde{X}_n,\gamma),$$

1. For $\gamma = 1$, it recovers **Loss-calibrated VB (LCVB)**[1]

$$\min_{a\in\mathcal{A}}\max_{q\in\mathcal{Q}}\left[\mathbb{E}_q[R(a,\theta)] - \mathrm{KL}(q\|\pi_n)\right]$$

[1] Lacoste–Julien et. al., Approximate inference for the loss-calibrated Bayesian. AISTAT(2011)

Risk-Sensitive Variational Bayes (RSVB)

For any $\gamma > 0$ and $a \in \mathcal{A}$ [Using Donsker-Varadhan variational lemma]

$$\frac{1}{\gamma} \log \mathbb{E}_{\pi_n} \left[\exp(\gamma R(a,\theta))\right] = \max_{q \in \mathcal{M}} \underbrace{\left[\mathbb{E}_q[R(a,\theta)] - \frac{1}{\gamma}\mathrm{KL}(q\|\pi_n)\right]}_{=:\mathcal{F}(a;q,\tilde{X}_n,\gamma)},$$

$$\Rightarrow \frac{1}{\gamma} \log \mathbb{E}_{\pi_n} \left[\exp(\gamma R(a,\theta))\right] \geq \max_{q \in \mathcal{Q}} \mathcal{F}(a;q,\tilde{X}_n,\gamma)$$

RSVB decision rule

$$\mathsf{a}^*_{\mathrm{RS}} \equiv \mathsf{a}^*_{\mathrm{RS}}(\gamma, \tilde{X}_n) := \mathsf{argmin}_{a \in \mathcal{A}} \max_{q \in \mathcal{Q}} \mathcal{F}(a;q,\tilde{X}_n,\gamma)$$

RSVB posterior (for any $a' \in \mathcal{A}$)

$$q^*_{a',\gamma}(\theta|\tilde{X}_n) \in \mathsf{argmax}_{q \in \mathcal{Q}} \mathcal{F}(a';q,\tilde{X}_n,\gamma),$$

1. For $\gamma = 1$, it recovers **Loss-calibrated VB (LCVB)**[1]

$$\min_{a \in \mathcal{A}} \max_{q \in \mathcal{Q}} \left[\mathbb{E}_q[R(a,\theta)] - \mathrm{KL}(q\|\pi_n)\right]$$

[1] Lacoste–Julien et. al., Approximate inference for the loss-calibrated Bayesian. AISTAT(2011)

Our Contributions

Statistical Guarantees: Under verifiable regularity conditions on the prior, likelihood model, and the risk function

1. RSVB posterior converges to δ_{θ_0} at the same convergence rate (wrt sample size n) as the true posterior,
2. Quantify the rate of convergence of the RSVB decision rule (when $\mathcal{A}$ is compact.)
3. Our theoretical results also imply the asymptotic properties of the LCVB posterior and the associated decision rule.

Empirical Results

Performance Measures

1. Variance of $\theta \sim q^*_{a,\gamma}(\theta|\tilde{X}_n)$, at $a = \mathsf{a}^*_{\mathrm{RS}}$.
2. Optimality Gap (OG) in values: $R(\mathsf{a}^*_{\mathrm{RS}}, \theta_0) - R(a^*, \theta_0)$, where $a^* = \mathrm{argmin}_{a\in\mathcal{A}} R(a, \theta_0)$.

Our Contributions

Statistical Guarantees: Under verifiable regularity conditions on the prior, likelihood model, and the risk function

1. RSVB posterior converges to δ_{θ_0} at the same convergence rate (wrt sample size n) as the true posterior,
2. Quantify the rate of convergence of the RSVB decision rule (when $\mathcal{A}$ is compact.)
3. Our theoretical results also imply the asymptotic properties of the LCVB posterior and the associated decision rule.

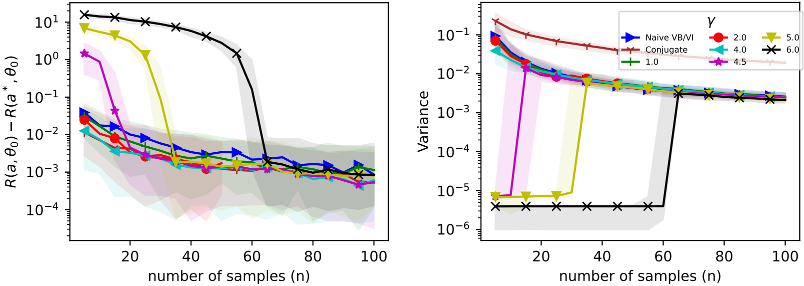
Empirical Results

Performance Measures

1. Variance of $\theta \sim q^*_{a,\gamma}(\theta|\tilde{X}_n)$, at $a = \mathtt{a}^*_{\mathtt{RS}}$.
2. Optimality Gap (OG) in values: $R(\mathtt{a}^*_{\mathtt{RS}}, \theta_0) - R(a^*, \theta_0)$, where $a^* = \text{argmin}_{a \in \mathcal{A}} R(a, \theta_0)$.

Demonstrating the effect of γ and n

$$\max_{q \in \mathcal{Q}} \underbrace{\mathbb{E}_q[R(a,\theta)]}_{\text{Risk-Aversion}} - \frac{1}{\gamma} \underbrace{\mathrm{KL}(q||\pi_n)}_{\text{Closesness to } \pi_n}$$



(a) OG in Values v/s n (b) Variance v/s n

Questions?

Poster Session(Poster #1222)
Where: Great Hall & Hall B1+B2
When: Wed 13 Dec,
10:45 a.m. CST - 12:45 p.m. CST

Paper link