Quadratic risk

Exercise 1 (Gaussian model – mean)**:** 1. $\widehat{\theta}(x) = \frac{1}{n}\sum_{i=1}^n x_i$.

2. $\mathrm{E}(\widehat{\theta}(X)) = \theta$ so no bias.
3. $\mathrm{var}(\widehat{\theta}(X)) = \frac{1}{n}\sigma^2$ so quadratic risk:
$$R(\theta, \widehat{\theta}) = \frac{\sigma^2}{n}.$$
4. Log-likelihood for $n = 1$:
$$\log p_\theta(x) = c - \frac{(x-\theta)^2}{2\sigma^2}$$
Fisher information for $n = 1$:
$$I(\theta) = \mathrm{var}(\frac{\partial \log p_\theta(X)}{\partial \theta}) = \frac{1}{\sigma^2}$$
Fisher information for $n \geq 1$:
$$I(\theta) = \frac{n}{\sigma^2}$$
Cramer-Rao bound:
$$R(\theta, \widehat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\sigma^2}{n}.$$
The estimator is efficient.

Exercise 2 (Poisson model)**:** 1. $\widehat{\theta}(x) = \frac{1}{n}\sum_{i=1}^n x_i$.

2. Unbiased: $\mathrm{E}(\widehat{\theta}(X)) = \theta$
3. $\mathrm{var}(\widehat{\theta}(X)) = \frac{1}{n}\theta$ so quadratic risk:
$$R(\theta, \widehat{\theta}) = \frac{\theta}{n}.$$
4. Log-likelihood for $n = 1$:
$$\log p_\theta(x) = c - \theta + x \log \theta.$$
Fisher information for $n = 1$:
$$I(\theta) = \mathrm{var}(\frac{\partial \log p_\theta(X)}{\partial \theta}) = \frac{\mathrm{var}(X)}{\theta^2} = \frac{1}{\theta}.$$
Fisher information for $n \geq 1$:
$$I(\theta) = \frac{n}{\theta}$$
Cramer-Rao bound:
$$R(\theta, \widehat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\theta}{n}.$$
The estimator is efficient.

Exercise 3 (Poisson model bis)**:** 1. $\widehat{g}_{\text{MLE}}(x) = e^{-\frac{1}{n}\sum_{i=1}^n x_i}$ and $\widehat{g}_{\text{MME}}(x) = \frac{1}{n}\sum_{i=1}^n 1_{x_i=0}$.

2. The MLE is biased (by Jensen's inequality), not the MME.

3. The MLE cannot be efficient since biased. For the MME, we get:

$$\text{var}(\widehat{g}_{\text{MME}}(X)) = \frac{1}{n}e^{-\theta}(1-e^{-\theta}).$$

   So risk:

$$R(\theta, \widehat{g}_{\text{MME}}) = \frac{1}{n}e^{-\theta}(1-e^{-\theta}).$$

   Fisher information:

$$I(\theta) = \frac{n}{\theta}$$

   Cramer-Rao bound:

$$R(\theta, \widehat{g}) \geq \frac{g'(\theta)^2}{I(\theta)} = \frac{\theta e^{-2\theta}}{n}.$$

   The estimator is not efficient.

Exercise 4 (Translated exponential)**:**

1. $\widehat{\theta}_{\text{MLE}}(x) = \min_{i=1,\dots,n} x_i$, $\widehat{\theta}_{\text{MME}}(x) = \frac{1}{n}\sum_{i=1}^n x_i - 1$.

2. $\text{E}(\widehat{\theta}_{\text{MLE}}(x)) = \theta + \frac{1}{n}$, biased. $\text{E}(\widehat{\theta}_{\text{MME}}(x)) = \theta$, unbiased.

3. $\text{var}(\widehat{\theta}_{\text{MLE}}) = \frac{1}{n^2}$, $R(\theta, \widehat{\theta}_{\text{MLE}}) = \frac{2}{n^2}$.
   $\text{var}(\widehat{\theta}_{\text{MME}}) = \frac{1}{n}$, $R(\theta, \widehat{\theta}_{\text{MME}}) = \frac{1}{n}$
   The MLE is better than the MM estimator in terms of quadratic risk (except for $n = 1$), although not efficient because biased.

Exercise 5 (Mixture model)**:** 1. $\widehat{\theta}(x) = \frac{\frac{1}{n}\sum_{i=1}^n x_i - \frac{1}{\lambda}}{\frac{1}{\mu} - \frac{1}{\lambda}}$

2. Unbiased. Risk:

$$R(\theta, \widehat{\theta}) = \text{var}(\widehat{\theta}) = \frac{2}{n}\frac{(\frac{\theta}{\mu^2} + \frac{1-\theta}{\lambda^2})}{(\frac{1}{\mu} - \frac{1}{\lambda})^2}$$

3. With $n = 1$,

$$\log p_\theta(x) = \log(\theta\mu e^{-\mu x} + (1-\theta)\lambda e^{-\lambda x})$$

   so that

$$I(\theta) = n\text{E}\left(\left(\frac{\mu e^{-\mu X} - \lambda e^{-\lambda X}}{\theta\mu e^{-\mu X} + (1-\theta)\lambda e^{-\lambda X})}\right)^2\right)$$

   For $\mu \to +\infty$,

$$R(\theta, \widehat{\theta}) \to \frac{2(1-\theta)}{n}, \quad I(\theta) \to \frac{(1-\theta)^2}{n}.$$

   Inefficient.

Exercise 6 (Gaussian vector)**:**
Consider the estimator:

$$\widehat{g}(x) = w^T\widehat{\theta} \quad \text{with} \quad \widehat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i.$$

1. No bias:

$$\mathrm{E}(\widehat{g}(X)) = w^T\mathrm{E}(\frac{1}{n}\sum_{i=1}^n X_i) = w^T\theta.$$

   Variance:

$$\begin{aligned}\mathrm{var}(\widehat{g}(X)) &= \mathrm{var}(w^T\frac{1}{n}\sum_{i=1}^n X_i)\\ &= \mathrm{var}(\frac{1}{n}\sum_{i=1}^n w^T X_i)\\ &= \frac{1}{n}\mathrm{var}(w^T X_1)\\ &= \frac{1}{n}w^T\Gamma w.\end{aligned}$$

   Quadratic risk:

$$R(\theta, \widehat{g}) = \frac{1}{n}w^T\Gamma w.$$

2. Log-likelihood for $n = 1$:

$$\log p_\theta(x) = c - \frac{1}{2}(x-\theta)^T\Gamma^{-1}(x-\theta).$$

   Gradient:

$$\nabla \log p_\theta(x) = -\Gamma^{-1}(x-\theta).$$

   Fisher information for $n = 1$:

$$I(\theta) = \mathrm{cov}\left(\nabla \log p_\theta(X)\right) = \Gamma^{-1}.$$

   Fisher information for $n \geq 1$:

$$I(\theta) = n\Gamma^{-1}.$$

   Cramer-Rao:

$$R(\theta, \widehat{g}) \geq \nabla g(\theta)^T I(\theta)^{-1}\nabla g(\theta) = \frac{1}{n}w^T\Gamma w.$$

   The estimator is efficient.