# IT for Statistics and Learning
## 2024

## Assignment 8
Assigned: Fr, Jan 12, 2024
Due: before the lecture on Fr, Jan 19, 2024 T. Oechtering

**Problem 8.1:** *Maximum Likelihood Estimator.* Consider the linear regression model with observations $y_i = x_i\theta + n_i$, $i = 1, \ldots, n \geq p$ in independent Gaussian noise $n_i|x_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ given known vector $x_i \in \mathbb{R}^{1\times p}$ and parameter $\theta \in \mathbb{R}^{p\times 1}$ to be estimated.

(i) Show that the pdf of $y_i|x_i$ is Gaussian with mean $x_i\theta$ and variance $\sigma^2$.

(ii) Show that the maximum likelihood estimator of $\theta$ is $\hat{T}_\theta(y, X) = (X^T X)^{-1} X^T y$ with $y = (y_1, \ldots, y_n)^T \in \mathbb{R}^{n\times 1}$ and $X^T = (x_1, \ldots, x_n) \in \mathbb{R}^{p\times n}$.

**Problem 8.2:** *Complete the proofs.* Show that

(i) $f^*(y) = y + \frac{1}{4}y^2$ is the convex conjugate of $f(x) = (x-1)^2$ which is used in the variational representation of the $\chi^2$-divergence;

(ii) Find coefficients $a, b$ in linear function $h(x) = ax + b$ such that we have

$$\sup_{a,b\in\mathbb{R}} 2(aE_P[X] + b) - E_Q\left[(aX+b)^2\right] - 1 = \frac{(E_P[X] - E_Q[X])^2}{\mathrm{Var}_Q[X]}.$$

**Problem 8.3:** *Complete the proofs.* Show that for Fisher information we have

(i) Regularity condition: $I(\theta) = -E_\theta\left[\frac{\mathrm{d}^2 \log P_\theta}{\mathrm{d}\theta^2}\right]$ if $P_\theta$ is twice differentiable and we have

$$\int \frac{\mathrm{d}^2 P_\theta(x)}{\mathrm{d}\theta^2}\mathrm{d}x = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\int P_\theta(x)\mathrm{d}x = 0.$$

(ii) Multiple samples: Let $X_1, ..., X_n \sim P_\theta$ iid, then

$$I_n(\theta) = nI(\theta)$$

**Problem 8.4:** *Coin flips [W].* Consider the experiment flipping a coin with a bias. Let $X_1, \ldots X_n \overset{iid}{\sim} \mathrm{Bern}(\theta)$ denote the random experiments with bias $\theta \in [0,1]$. We consider the quadratic loss $l(\theta, \hat{\theta}] = (\theta - \hat{\theta})^2$ and denote the minimax risk by $R^*$.

(i) Use the empirical frequency as estimator $\hat{\theta}_{\mathrm{emp}}$ to estimate $\theta$. Compute and plot the risk $R_\theta(\hat{\theta}_{\mathrm{emp}})$ and show that $R^* \leq \frac{1}{4n}$.

(ii) Compute the Fisher information of $P_\theta = \mathrm{Bern}(\theta)^{\otimes n}$ and $Q_\theta = \mathrm{Binom}(n, \theta)$. Explain why they are equal.

(iii) Invoke the Bayesian Cramér-Rao lower bound to show that $R^* = \frac{1+o(1)}{4n}$.

(iv) The risk of $\hat{\theta}_{\mathrm{emp}}$ is maximized at $\frac{1}{2}$, which suggests that it might be possible to hedge against the situation by using the following randomized estimator

$$\hat{\theta}_{\mathrm{rand}} = \begin{cases} \hat{\theta}_{\mathrm{emp}}, & \text{with probability } \delta, \\ 1/2, & \text{with probability } 1-\delta \end{cases}$$

Find the worst-case risk of $\hat{\theta}_{\mathrm{rand}}$ as a function of $\delta$. Choose the best $\delta$ ans show that this leads to a better upper bound $R^* \leq \frac{1}{4(n+1)}$.

(v) Randomized estimators are always improvable when the loss is convex by averaging out the randomness by considering the estimator $\hat{\theta}^* = E[\hat{\theta}_{\mathrm{rand}}] = \hat{\theta}_{\mathrm{emp}}\delta + 1/2(1-\delta)$. Find the optimal $\delta$ that minimizes the worst-case risk and show that it is independent of $\delta$ and $R^* \leq \frac{1}{4(1+\sqrt{n})^2}$.

(vi) (optional to trade one of the previous subquestion) Show that $\hat{\theta}^*$ is exactly minimax optimal and hence $R^* = \frac{1}{4(1+\sqrt{n})^2}$. Consider the following prior Beta$(a, b)$ with density

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}, \qquad \theta \in [0, 1]$$

where $\Gamma(a) = \int_0^\infty x^{a-1}\mathrm{e}^{-x}$. Show that if $a = b = \sqrt{n}/2$, then $\hat{\theta}^*$ coincides with Bayes estimator $E[\theta|X_1, \ldots X_n]$ for this prior, which is therefore least favourable. (Hint: Work with the sufficient statistic $S = X_1 + \cdots + x_n$)