Assignment sheet 5 — Spring 2025

Assignment 1. Let $X_1, X_2, \dots, X_n$ be an i.i.d. sample from the $N(\mu, 1)$ distribution. Let $\widehat{\mu}$ be the MLE of μ.
(a) Find $\widehat{\mu}$.
(b) Find the asymptotic distribution of $\widehat{\mu}$.
(c) Using part (b) and without direct calculations, find the Cramer-Rao lower bound for the variance of an unbiased estimator of μ.
(d) Is there an estimator that satisfies this lower bound for each fixed n ?
(e) Suppose that we are interested in estimating $g(\mu) = \mathbb{P}[X_1 \leq 2]$. Find an explicit expression for $g(\mu)$.
(f) Find the MLE of $g(\mu)$. Denote it by T.
(g) Using the delta method, find the asymptotic distribution of T.

Assignment 2. Let $X_1, X_2, \dots, X_n$ be an i.i.d. sample from the distribution with density function

$$f_X(x) = \begin{cases} \frac{\alpha\pi^\alpha}{x^{\alpha+1}}, & x \geq \pi \\ 0 & x < \pi. \end{cases}$$

(This is a Pareto distribution and α is called the tail index or the Pareto index.)
(a) Find $\mathbb{E}[\log X_1]$ and $\mathbb{E}[(\log X_1)^2]$.
Hint : Instead of calculating painful integrals, notice that this is an exponential family with sufficient statistic related to $\log X$*, and use the theorem from Week 3 slides : pg 19 (Sampling from an Exponential Family).*
(b) Find the MLE $\widehat{\alpha}$ of α.
(c) Use MLE theory to find the asymptotic distribution of $\widehat{\alpha}$. Are the assumptions satisfied ?
(d) Let $Y = \log(X/\pi)$. Find the distribution of Y directly, i.e., without using transformation of variables.
(e) Find the asymptotic distribution of $T(Y_1, Y_2, \dots, Y_n) := \sum_{i=1}^n Y_i$.
(f) Express $\widehat{\alpha}$ in terms of $T(Y_1, Y_2, \dots, Y_n)$, and use this along with part (d) to find the asymptotic distribution of $\widehat{\alpha}$.
(*Hint : Use the delta method.*)
(g) Find the method of moments estimator $\widetilde{\alpha}$ of α and compare with the maximum likelihood estimator $\widehat{\alpha}$.
(h) Assuming $\alpha > 2$, compare the *asymptotic* variance of the two estimators for α.
Hint : for $\widetilde{\alpha}$ *use the central limit theorem and the delta method.*

Assignment 3. (optional)
In this assignment we shall see empirically that Stein's estimator has a lower mean squared error than the maximum likelihood estimator.
Let y_1, y_2 and y_3 be independent normal random variables with unit variance and unknown means μ_1, μ_2 and μ_3.
(a) Use R to simulate one realisation of the random vector $y = (y_1, y_2, y_3)$ for the parameter value $\mu = (\mu_1, \mu_2, \mu_3) = (-1, 0, 1)$. *Hint : the command* `rnorm` *can take vector values.*
(b) What is the optimal value of a in terms of the mean squared error of the James–Stein estimator $\widetilde{\mu}_a$? Write an R command that calculates it, for a sample stored in a vector $Y \in \mathbb{R}^3$.
Hint : you can use `sum(Y^2)`.
(c) Repeat the simulation 1000 times. For each repetition, calculate the errors $\|\mu - \widehat{\mu}\|^2$ and $\|\mu - \widetilde{\mu}_a\|^2$. Store these in two vectors of length 1000, `MSE.mle` and `MSE.stein`. Use these

vectors to approximate the mean squared error of the two estimators. Which one is smaller? Try changing the values of μ (and perhaps n and a).

Assignment 4. In this assignment we give an alternative approach to shrinkage by means of adding a penalty term to the optimisation problem.
(a) Let $X \sim Gamma(k, \lambda)$ with $k > 1$. Using the property that $\Gamma(x) = (x-1)\Gamma(x-1)$ for $x > 1$, show that $\mathbb{E}\,\frac{1}{X} = \lambda/(k-1)$.
(b) Using part (a), show that if $X \sim \chi^2_n$ with $n > 2$, then $\mathbb{E}\,\frac{1}{X} = 1/(n-2)$.
(c) Now that we know that shrinking is a good idea, we approach the estimation from a different point of view, that of *penalisation.*
Recall Stein's setup (Week 5 slides) : let $Y_i \sim N(\mu_i, 1)$ be independent, $i = 1, \ldots, n$. Explain why the maximum likelihood estimator $\widehat{\mu}$ can be obtained as the minimiser

$$\min_{\mu_1,\ldots,\mu_n} \sum_{i=1}^{n} (y_i - \mu_i)^2.$$

(d) We can shrink $\widehat{\mu}$ by adding a penality term that renders large values disadvantageous : for $\lambda \geq 0$ define $\tilde{\mu}_\lambda$ as the solution of

$$\min_{\mu_1,\ldots,\mu_n} \sum_{i=1}^{n} (y_i - \mu_i)^2 + \lambda \sum_{i=1}^{n} \mu_i^2.$$

By solving this minimisation problem, show that $\tilde{\mu}_\lambda = y/(1+\lambda)$.
(e) Find the mean squared error of $\tilde{\mu}_\lambda$ as a function of λ. *Hint :* $\mathbb{E}\, y_i - \mu_i = 0$.
(f) Show that for some values of λ, the mean squared error of $\tilde{\mu}_\lambda$ is smaller than that of $\widehat{\mu} = \tilde{\mu}_0$.
(g) Find the optimal value of λ in terms of the mean squared error. Can one use this value in practice?
Remark. This is a particular case of ridge regression that will be seen later in the course.