WISB263 Mathematical Statistics
Resit
20th July 2021, 15.00-18.00
Total amount of points: 100
Lecturer: Wioletta Ruszel

Exercise 1 (25 p.)
Suppose that $\mathbf{X} = (X_1, ..., X_n)$ is a random sample such that each X_i can take values in $\mathbb{N} \cup \{0\}$ and has the probability mass function

$$p_\theta(x_i) = ce^{-\theta x_i}$$

with $\theta = \mathbb{R}_+$ and $x_i \in \mathbb{N} \cup \{0\}$.

(i) (3 p.) Show that $c = 1 - e^{-\theta}$.

(ii) (5 p.) Find the MoM estimator $\hat{\theta}_n$ of θ. (Hint: You can use that $\sum_{k=0}^{\infty} kz^k = \frac{z}{(1-z)^2}$ for $|z| < 1$.)

(iii) (6 p.) Show that $Var_\theta(X_1) = \frac{e^{-\theta}}{(1-e^{-\theta})^2}$ and that $\hat{\theta}_n$ is asymptotically normal. Determine the variance of the asymptotic distribution. (Hint: You can use that $\sum_{k=0}^{\infty} k^2 z^k = \frac{z(z+1)}{(1-z)^3}$ for $|z| < 1$.)

(iv) (6 p.) Show that the MLE θ_n^* is equal to the MoM estimator.

(v) (5 p.) Assuming that θ_n^* is consistent and asymptotically normal, verify that the asymptotic variance is the same as the one found in (iii).

Exercise 2 (20 p.)
Two surveys were independently conducted to estimate a population mean μ in a population of size N using simple random sampling. Denote their unbiased estimators by $\overline{X}_{n,1}$ and $\overline{X}_{n,2}$ respectively and standard errors by $\sigma_{\overline{X}_{n,1}}$ and $\sigma_{\overline{X}_{n,2}}$. We want to understand if combining both estimators will give a better result. Let $a, b \in \mathbb{R}$ and define:

$$X_n' := a\overline{X}_{n,1} + b\overline{X}_{n,2}.$$

(i) (2 p.) Under which condition on a, b is the new estimator X_n' unbiased?

(ii) (8 p.) Determine which a, b minimizes the variance, subjected to unbiasedness.

(iii) (5 p.) Assume now that we have a standard error for the first survey of $\sqrt{0.1}$ and for the second of $\sqrt{0.3}$. What is the standard error of the combined estimator? What do you observe?

(iv) (5 p.) Assume that $\overline{X}_{n,1}$ and $\overline{X}_{n,2}$ are sample mean estimators. Compute the standard error $\sigma_{X_n'}$ when $\sigma^2 = 0.2$, $N = 200$ and $n = 50$ and compare with the standard error for the estimator X_n' from random sampling. Which one is smaller?

Exercise 3 (25 p.)
Consider a random sample $\mathbf{X} = (X_1, ..., X_n)$ such that $X_i \sim N(\theta, \theta)$ were $\theta \in \mathbb{R}_+$.

(i) (5 p.) Prove that the parametric family is of exponential type.

(ii) (5 p.) Design a u.m.p. test for testing $(H_0) : \theta \geq \theta_0$ against $(H_1) : \theta < \theta_0$ at significance α. Describe the rejection region implicitely.

(iii) (1 p.) Does the rejection region change when we change the null hypothesis to $(H_0) : \theta = \theta_0$?

(iv) (5 p.) Show that the MLE $\hat{\theta}_n$ for θ is equal to

$$\hat{\theta}_n = \frac{1}{2}\left(\sqrt{\frac{4\sum_{i=1}^n X_i^2}{n} + 1} - 1\right).$$

(v) (5 p.) Compute the Fisher information $I(\theta)$ and determine the Wald random confidence interval for θ at significance α.

(vi) (4 p.) For $\alpha = 0.05$, $z(0.025) = 1.96$ and $\sum_{i=1}^{100} x_i^2 = 80$ determine the confidence interval.

Exercise 4 (30 p.) Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d random variables with common density f. Define for $s > 0$

$$\tau(s) = \inf\{n \geq 1; X_n > s\}$$

the index of the first X_n-random variable which is exceeding a certain level s.

(i) (3 p.) Let $k \in \mathbb{N}$, express the event $\{\tau(s) = k\}$ in terms of $X_1, ..., X_k$.

(ii) (5 p.) Compute $\mathbb{P}(\tau(s) = k)$ for $k \in \mathbb{N}$ in terms of $p_s := \mathbb{P}(X_1 > s)$. Show that it is geometric. Identify the parameter.

(iii) (5 p.) Show that the MGF $M_s(t)$ of $p_s\tau(s)$ is equal to

$$M_s(t) = \frac{p_s e^{tp_s}}{1 - e^{tp_s}(1 - p_s)}.$$

(iv) (5 p.) Show that $\lim_{s\to\infty} p_s = 0$ and compute $\lim_{s\to\infty} M_s(t)$. Show the limiting distribution is $Exp(1)$.

(v) (7 p.) Application: every day the water level of the Maas river is measured. X_n measures the water level on day n in standard units. If the level exceeds 8 then there is a high chance of floods to occur and the Rijkswaterstaat needs to be called. Assume now that the common density of $(X_n)_{n\geq 1}$ is given by

$$f(x) = \frac{3}{(1 + x)^4} \mathbb{1}_{x\geq 0}.$$

What is the probability that there is danger of flooding for the first time after exactly 5 days?

(vi) (5 p.) Find the maximal $k \geq 1$ such that $\mathbb{P}(\tau(8) > k) \geq 0.9$. What does the result mean in combination with (v)?