<u>WISB263 Mathematical Statistics</u>
Resit - 14th July 2020
Total amount of points: 100
Grade exam: number of points/10 rounded to 0.5 or integers

Lecturer: Dr. Wioletta Ruszel
Teaching assistants: Leandro Chiarini, Ben Balkenende, Esther Steenkamer

Exercise 1 (25 p.) We want to investigate the percentage of products on a production line which are defective. Components are inspected independently until the first defective is encountered. We observe that 5 distinct components are found defective after the first inspection, 11 distinct components are found defective at their second inspection (the same distinct 11 components were checked twice and at the second inspection they were broken), 12 at their third inspection, 8 at their forth inspection, 6 at their fifth inspection and 8 at their sixth inspection.

We have checked in total $n = 50$ components (the components which were checked and not defective are not considered here) and performed $\sum_{i=1}^{50} x_i = 173$ inspections. x_i denotes the number of inspections of component i until the first time it is defective.

1. (10 p.) Perform a hypothesis test to decide whether the number of inspections until the first defective one is encountered is geometrically distributed at $1 - \alpha = 0.99$ significance. What is the p-value?

2. (10 p.) Under the assumption that the data comes from a geometric distribution with parameter p, perform a hypothesis test for $(H_0) : p = 0.3$ against $(H_1) : p \neq 0.3$ using a generalized likelihood ratio statistic $R(\mathbf{X})$ at $\alpha = 0.01$. You can use the appropriate asymptotic distribution of $R(\mathbf{X})$. Compute the p-value. (Hint: You can use that the MLE is equal to $\hat{p}_n = \frac{1}{\bar{X}_n}$.)

3. (5 p.) Compare both tests, which one is more appropriate? Comment also on the assumption that the underlying distribution is geometric in the second test. Was it a reasonable assumption?

Exercise 2 (25 p.) Let $X_1, ..., X_N$ be independent Bernoulli random variables $B(p)$ modelling the outcome whether a person i has a sickness or not. We consider the estimator $\hat{p}_n$ for the proportion of people with the sickness in a sample of size n taken out of a population of size N using simple random sample scheme.

1. (6 p.) Determine for which $p \in (0, 1)$ the standard error of the estimator $\hat{p}_n$ is maximal.

2. (4 p.) Show that the maximal standard error of the estimator of the variance of $\hat{p}_n$ is equal to $\frac{1}{2}\sqrt{\frac{N-n}{N(n-1)}}$.

3. (4 p.) Choose now $N = 1000$. How many samples n do you need to ensure that the standard error is at most 0.05?

4. (8 p.) Use the result from (2) to construct an asymptotic confidence interval for p at confidence of at least 0.99.

5. (3 p.) Let $N = 1000$. If in the simple random sample 40 out of 100 tested people have the sickness, determine the 99% (asymptotic) confidence interval for p.

Exercise 3 (25 p.) We want to study different estimators for the parameter λ in an exponential distribution with parameter $\frac{1}{\lambda}$ for $\lambda > 0$.

1. (5 p.) Determine the MoM estimator $\hat{\lambda}_1$ of λ using an i.i.d. sample $X_1, ..., X_n$ of exponential random variables with parameter $\frac{1}{\lambda}$. Compute the mean and variance of $\hat{\lambda}_1$.

2. (5 p.) Prove that the rescaled estimator $\sqrt{n}(\hat{\lambda}_1 - \lambda)$ is asymptotically normal as $n \to \infty$ and determine the limiting variance.

3. (5 p.) Determine the MoM estimator $\hat{\lambda}_2$ of λ using an i.i.d. sample $X_1, ..., X_n$ of exponential random variables with parameter $\frac{1}{\sqrt{\lambda}}$. Compute the mean and variance of $\hat{\lambda}_2$. (Hint: You can use that $\int_0^\infty x^4 e^{-ax} dx = \frac{24}{a^5}$.)

4. (5 p.) Prove that the rescaled estimator $\sqrt{n}(\hat{\lambda}_2 - \lambda)$ is asymptotically normal as $n \to \infty$ and determine the limiting variance.

5. (5 p.) Is one of the estimators $\hat{\lambda}_1$ or $\hat{\lambda}_2$ most efficient in their respective class K_b? Which one is asymptotically more efficient?

Exercise 4 (25 p.) Consider independent random variables $X_1, ..., X_n$ such that $X_i \sim N(\mu, \sigma^2 x_i)$ and $x_i \neq 0$ for all $i = 1, ..., n$.

1. (5 p.) Write the random variables $X_1, ..., X_n$ as a linear regression model $Y_1, ..., Y_n$ with respect to errors ϵ_i which have mean 0 and common variance σ^2 for all $i = 1, ..., n$.

2. (7 p.) Compute directly the unbiased MLE's $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ for μ resp. σ^2.

3. (5 p.) Determine the distribution of $\frac{n-1}{\sigma^2}\hat{\sigma}_n^2$.

4. (5 p.) Discuss the hypothesis test $(H_0) : \sigma = 1$ against $(H_1) : \sigma > 1$. Which test statistic could you choose? Determine the rejection region for $\alpha = 0.05$ and $n = 10$. For which $\hat{\sigma}_{10}^2$ do we accept H_0?

5. (3 p.) Interpret now the vector $(Y_1, ..., Y_n)$ in (1) as a linear regression model. Which design matrix $\mathbf{X}$ and vector of parameters B correspond to $(Y_1, ..., Y_n)$? Discuss the connection of the MLE $\hat{\mu}_n$ from (2) with the least-square estimator in a simple linear regression model.