2 February 2018, 9:00–12:00

# Rijksuniversiteit Groningen
# Statistics

*Resit*

RULES FOR THE EXAM:

- The use of a normal, non-graphical calculator is permitted.
- This is a CLOSED-BOOK exam.
- At the end of the exam you can find a chi-squared table.
- Your exam mark : $10 + 90\times$ your score/75 .

1. **Point estimation [20 Marks].** Let $X_1, \dots, X_n$ be a sample of independent, identically distributed $N(1,\theta)$ random variables, with density for $\theta \in \mathbb{R}^+$,

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-1)^2}.$$

   (a) [5 Marks] Determine the Method of Moments estimator $\hat{\theta}$ of $\theta$.

   (b) [5 Marks] Determine whether $\hat{\theta}$ is unbiased.

   (c) [10 Marks] Does $\hat{\theta}$ achieve the Cramer-Rao lower bound? [Hint: You can use the fact that $EX_1^4 = 3\theta^2 + 6\theta + 1$]

2. **Asymptotic distribution Likelihood Ratio Test statistic** **20 Marks**.
Let $X = (X_1, \dots, X_n)$ be the observed data, such that

$$X_i \overset{\text{i.i.d.}}{\sim} f_{\theta_0}.$$

Let $f_\theta$ be twice continuously differentiable with support not depending on $\theta$.
Let $\hat{\theta} = \hat{\theta}(X)$ be the maximum likelihood estimate of $\theta$.

   (a) Derive the second order Taylor approximation of the log-likelihood $\ell(\theta) = \log f_\theta(X_1, \dots, X_n)$ at the true value $\ell(\theta_0)$ around the MLE $\hat{\theta}$. [5 Marks]

   (b) Show that $\frac{1}{n}\frac{d^2\ell}{d\theta^2}(\theta_0) \to -I(\theta_0) = E\frac{d^2}{d\theta^2}\log f_\theta(X_1)|_{\theta=\theta_0}$ as $n \to \infty$. [5 Marks]

   (c) Use the second order Taylor approximation to show that

$$-2\log LRT \approx -(\hat{\theta}-\theta_0)^2 \frac{d^2\ell}{d\theta^2}(\hat{\theta}).$$

   where LRT is the likelihood ratio test statistic.[5 Marks]

   (d) Taking the approximation in (c) as an equality, use (b) and (c) together with the asymptotic efficiency of the MLE $\hat{\theta}$ to show that

$$-2\log LRT \to \chi_1^2$$

   in distribution as $n \to \infty$. [5 Marks]

3. **Confidence interval** **30 Marks**.
Let $(Y_1, x_1), ..., (Y_n, x_n)$ be the data, where $\{Y_i\}_{i=1}^n$ are independently and exponentially distributed random variables in the following way:

$$Y_i \sim Ex(\lambda_0 x_i), \quad i = 1, 2, \ldots, n$$

i.e.

$$f_{Y_i}(y) = \lambda_0 x_i e^{-\lambda_0 x_i y} 1_{y \geq 0}.$$

The known constants $\{x_i\}$ are strictly positive.

(a) Derive the maximum likelihood estimator $\hat{\lambda}_n$ for $\lambda_0$. [Hint: Don't forget to show that this is really a maximum.] [10 Marks]

(b) Determine whether the MLE is a sufficient statistic. [5 Marks]

(c) Show that for large $n$ approximately, $\hat{\lambda}_n \sim N(\lambda_0, \frac{\lambda_0^2}{n})$. [5 Marks]

(d) Find an expression for a 95% asymptotic confidence interval for $\lambda$ based on inverting the test statistic statistic $\hat{\lambda}_n$ [Hint: $z_{0.975} = 1.96$]. [10 Marks]

4. **Optimal testing** **15 Marks**. Consider 10 i.i.d. observations $X_1, \ldots, X_{10} \sim N(\mu, 81)$, i.e., with density

$$f_X(x) = \frac{1}{\sqrt{2\pi 9}} e^{-\frac{1}{162}(x-\mu)^2}.$$

We want to test the following hypotheses:

$$H_0 : \quad \mu = 1$$
$$H_1 : \quad \mu = 2$$

(a) We want to perform an optimal test with a significance level of at most 5% of the null hypothesis against the alternative. Determine the critical region. [10 Marks]

(b) What is the power of this test? [5 Marks]

**Below statistical tables which may be used in the calculations.**

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.0 | 0.341 | 0.344 | 0.346 | 0.348 | 0.351 | 0.353 | 0.355 | 0.358 | 0.360 | 0.362 |
| 1.1 | 0.364 | 0.367 | 0.369 | 0.371 | 0.373 | 0.375 | 0.377 | 0.379 | 0.381 | 0.383 |
| 1.2 | 0.385 | 0.387 | 0.389 | 0.391 | 0.393 | 0.394 | 0.396 | 0.398 | 0.400 | 0.401 |
| 1.3 | 0.403 | 0.405 | 0.407 | 0.408 | 0.410 | 0.411 | 0.413 | 0.415 | 0.416 | 0.418 |
| 1.4 | 0.419 | 0.421 | 0.422 | 0.424 | 0.425 | 0.426 | 0.428 | 0.429 | 0.431 | 0.432 |
| 1.5 | 0.433 | 0.434 | 0.436 | 0.437 | 0.438 | 0.439 | 0.441 | 0.442 | 0.443 | 0.444 |
| 1.6 | 0.445 | 0.446 | 0.447 | 0.448 | 0.449 | 0.451 | 0.452 | 0.453 | 0.454 | 0.454 |
| 1.7 | 0.455 | 0.456 | 0.457 | 0.458 | 0.459 | 0.460 | 0.461 | 0.462 | 0.462 | 0.463 |
| 1.8 | 0.464 | 0.465 | 0.466 | 0.466 | 0.467 | 0.468 | 0.469 | 0.469 | 0.470 | 0.471 |
| 1.9 | 0.471 | 0.472 | 0.473 | 0.473 | 0.474 | 0.474 | 0.475 | 0.476 | 0.476 | 0.477 |
| 2.0 | 0.477 | 0.478 | 0.478 | 0.479 | 0.479 | 0.480 | 0.480 | 0.481 | 0.481 | 0.482 |

Table 1: Standard Normal Distribution as found in the book. This means that values in the table correspond to probabilities $P(0 < Z \leq z)$, where $Z$ is a standard normal distributed variable.