6 November 2017, 14:00 – 17:00

Rijksuniversiteit Groningen
Statistiek

Tentamen

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 normal table and a chi-squared table.
- Your exam mark : 10 + your score.

1. **Point estimation** $\boxed{\textbf{40 Marks}}$**.** Let $X_1, \ldots, X_n$ be a random sample of independent, identically distributed Exponential(θ) random variables, with density

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find a sufficient statistic $\hat{\theta}(X_1, \ldots, X_n)$ for θ. [5 Marks]

ANSWER: Factorization theorem:

$$\begin{aligned} f(X) &= \prod_{i=1}^{n} \frac{1}{\theta} e^{-X_i/\theta} \\ &= \frac{1}{\theta^n} e^{-\sum_{i=1}^{n} X_i/\theta} \end{aligned}$$

So, therefore $\hat{\theta}(X) = \sum_{i=1}^{n} X_i$ is a sufficient statistic.

(b) Determine the Cramer-Rao lower bound for an unbiased estimator of θ. [10 Marks]

ANSWER:

$$\begin{aligned} l_{X_1}(\theta) &= -\log\theta - X_1/\theta \\ l'_{X_1}(\theta) &= -\frac{1}{\theta} + X_1/\theta^2 \\ l''_{X_1}(\theta) &= \frac{1}{\theta^2} - 2X_1/\theta^3 \end{aligned}$$

Then

$$El''_{X_1} = \frac{1}{\theta^2} - 2EX_1/\theta^3 = \frac{1}{\theta^2} - 2\theta/\theta^3 = \frac{-1}{\theta^2}.$$

So,

$$CRLB(\theta) = \frac{-1}{n \times (-1/\theta^2)} = \frac{\theta^2}{n}.$$

(c) Determine the maximum likelihood estimator (MLE) of θ. [5 Marks]

ANSWER: Set $l_X(\theta) = 0$, i.e., solve

$$-\frac{n}{\theta} + \sum_{i=1}^{n} X_i/\theta^2 = 0,$$

i.e.,

$$\hat{\theta} = \bar{X}.$$

We need to check that this is a maximum, so substitute into second derivative:

$$l''_X(\hat{\theta}) = \frac{n}{\bar{X}^2} - 2\sum_{i=1}^{n} X_i/\bar{X}^3 = -\frac{n}{\bar{X}^2} < 0,$$

So it is a maximum.

(d) Let $\hat{\theta}_n$ be the MLE of θ,

i. Determine whether $\hat{\theta}_n$ is unbiased. [5 Marks]
ANSWER:

$$E\hat{\theta}_n = E\bar{X} = \theta,$$

so the estimator is unbiased.

ii. Prove that $\hat{\theta}$ is efficient. [5 Marks]
ANSWER:

$$V(\hat{theta}_n) = \frac{\theta^2}{n},$$

which achieves the CRLB and therefore it is efficient.

(e) Assume that $\bar{x} = 4$ and $n = 16$, then determine an approximate 95% confidence interval for θ, using

i. a Wald approach – assuming asymptotic normality. [5 Marks]
ANSWER: We want to find

$$I_W(X) = \{\theta_0 \mid \text{given } W(X) \text{ do not reject } H_0 : \theta = theta_0\},$$

The Wald statistic is given by

$$W(X) = \hat{\theta}_n = \bar{X}.$$

We know that asymptotically

$$\sqrt{n}\frac{\bar{X} - \theta}{\theta} \xrightarrow{D} N(0, 1).$$

So, we do not reject θ_0, if

$$\sqrt{n}\frac{\bar{X} - \theta_0}{\theta_0} \in (-1.96, 1.96),$$

i.e.,

$$I(X) = \left\{\theta_0 \mid \frac{4}{1 + 1.96/4} < \theta_0 < \frac{4}{1 - 1.96/4}\right\} = (2.7, 7.8).$$

ii. a Likelihood Ratio approach – assuming asymptotic chi-squared [Hint: graphical approximations of the interval are allowed, but show the graph]. [5 Marks]
ANSWER: We want to find

$$I_{LR}(X) = \{\theta_0 \mid \text{given } LR(X) \text{ do not reject } H_0 : \theta = theta_0\},$$

The Likelihood Ratio statistic is given by

$$LR(X) = L_X(\theta_0)/L_X(\hat{\theta}_n).$$

We use the fact that approximately under H_0

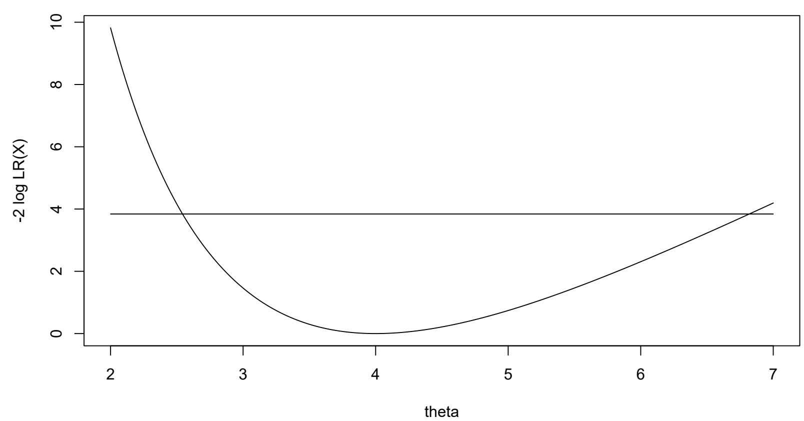
$$-2\log LR(X) \sim \chi^2_1,$$

whereby

$$\begin{aligned}
-2\log LR(X) &= -2\times\{(-\log\theta_0 - \sum_{i=1}^n X_i/\theta_0) - (-\log\bar{X} - \sum_{i=1}^n X_i/\bar{X})\} \\
&= -2\{\log\frac{\bar{X}}{\theta_0} - n\times(\frac{\bar{X}}{\theta_0} - 1)\}
\end{aligned}$$

The figure below shows the -2 times the log likelihood ratio together with the $\chi^2_{1;0.95} = 3.841$, which results into an approximate

$$I_{LR}(X) = (2.5, 6.8).$$



2. **Cramer-Rao: best unbiased estimators** $\boxed{\textbf{25 Marks}}$**.**
   Let $X = (X_1, \ldots, X_n)$ be the observed data, such that

   $$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta.$$

   Let $\hat{\theta} = \hat{\theta}(X)$ be an unbiased estimator of θ. Let $Y = \frac{d}{d\theta}\log f_{\theta,\text{joint}}(X)$.

   (a) Show that $EY = 0$. [5 Marks]

   (b) Show that $\text{Cov}(\hat{\theta}, Y) = 1$. [10 Marks]

   (c) Show that $V(\hat{\theta}) \geq 1/E(Y^2)$. [5 Marks]

(d) Use the above to show that

$$V(\hat{\theta}) \geq \frac{1}{nE(\frac{d}{d\theta}\log f_\theta(X_1))^2}.$$

[5 Marks]

3. **Optimal testing** $\boxed{\textbf{25 Marks}}$**.** An Atomic Energy Agency is worried that a particular nuclear plant has leaked radio-active material. They do 5 independent Geiger counter measurements in the direct neighbourhood of the reactor. They find the following measurements (per unit time):

observation	1	2	3	4	5
count	1	2	6	2	7

The natural background radiation has an average of $\lambda = 2$ (per unit time). The agency would only be worried if the radiation rate would be in the order of $\lambda = 5$. They therefore decide to test,

$$\begin{aligned} H_0: &\quad \lambda = 2 \\ H_1: &\quad \lambda = 5 \end{aligned}$$

On the basis of these 5 measurements, they want to device the optimal test to see if there is any reason for alarm, i.e. whether the rate of radioactivity is 5. They make the following assumptions. Let $X_1, ..., X_5$ be independently Poisson distributed with parameter λ, i.e.

$$p_{X_i}(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

We use significance *level* $\alpha = 0.05$.

(a) [10 Marks]Device the **most powerful test** for deciding between the two hypotheses on the basis of these five measurements. Determine the Critical Region. [Hint: sum of independent Poissons is Poisson].

a) Given the simple hypotheses, the most powerful test is given by

$$CR = \left\{ x \in \mathbb{N}^5 \;\middle|\; \frac{L_0(x)}{L_1(x)} \leq k_\alpha \right\}$$

<u>Consider</u>

$$\frac{L_0(x)}{L_1(x)} = \frac{\prod_{i=1}^{5} e^{-2} \frac{2^{x_i}}{x_i!}}{\prod_{i=1}^{5} e^{-5} \frac{5^{x_i}}{x_i!}}$$

$$= \prod_{i=1}^{5} e^{3} \left(\frac{2}{5}\right)^{x_i}$$

$$= e^{15} \left(\frac{2}{5}\right)^{\sum_{i=1}^{5} x_i}$$

<u>Note</u>: $\sum_{i=1}^{5} X_i \mid H_0 \sim \text{Poisson}(10)$

Critical region of level α

$$CR_\alpha = \left\{ x \in \mathbb{N}^5 \;\middle|\; \frac{L_0(x)}{L_1(x)} \leq k_\alpha \right\}$$

$$= \left\{ x \in \mathbb{N}^5 \;\middle|\; e^{15} \left(\frac{2}{5}\right)^{\sum_{i=1}^{5} x_i} \leq k_\alpha \right\}$$

$$= \left\{ x \in \mathbb{N}^5 \;\middle|\; \sum_{i=1}^{5} x_i \geq k^*_\alpha \right\}$$

To find k^*_α, we need

$$\alpha = P_{H_0}\left(\sum_{i=1}^{5} X_i \geq k^*_\alpha\right)$$

So k^*_α = upper α quantile of Poisson 10.

$$CR_\alpha = \left\{ x \in \mathbb{N}^5 \;\middle|\; \sum_{i=1}^{5} x_i \geq \underbrace{\text{upper Poi}(10)\ \alpha \text{ quantile}}_{\substack{\text{For } \alpha = 0.0487 \\ \cong 15}} \right\}$$

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

b) Power of the test

$$1-\beta = P(\underbrace{\sum X_i}_{Poi\,(25)} \geqslant 16 \mid H_1)$$

$$= 0.9777$$

(c) Clearly, in practice, it is difficult to set up the hypothesis test as two simple hypotheses. In fact, we would like to test,

$$\begin{aligned} H_0: &\quad \lambda \leq 2 \\ H_1: &\quad \lambda > 2 \end{aligned}$$

i. [5 Marks] Determine a sufficient statistic T w.r.t. λ of $X_1, ..., X_5$ be independently Poisson distributed with parameter λ and show that it has a *monotone likelihood ratio*.
ANSWER: From the factorization theorem, it is clear that

$$f_X(x) = \prod_{i=1}^{5} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-5\lambda} \lambda^{\sum_{i=1}^{5} x_i} / \prod_{i=1}^{5} x_i!$$

and therefore

$$T = \sum_{i=1}^{5} X_i \sim \text{Poisson}(5\lambda) \text{ is a sufficient statistic.}$$

The likelihood ratio for $\lambda_2 > \lambda_1$ as a function of $t = \sum_{i=1}^{5} x_i$ is given as

$$LR(t) = e^{5(\lambda_1 - \lambda_2)} \left(\frac{\lambda_2}{\lambda_1} \right)^t.$$

LR is monotone increasing in t as $\lambda_2/\lambda_1 > 1$.

ii. [5 Marks] Derive a uniform most powerful test of level $\alpha = 0.05$.

ANSWER: In (a) we have shown that the sufficient statistic $T = \sum_{i=1}^{5} X_i$ has a monotone likelihood ratio. The conditions of the Karlin-Rubin theorem are thereby fullfilled. From the Karlin-Rubin theorem, it follows that if we reject H_0 if and only if $T > t_0$, where

$$0.0487 = P_{5\lambda=10}(T > t_0)$$

the test is UMP of level $\alpha = 0.0487$. From this it is clear that $t_0 = 15$.

Below statistical tables which may be used in the calculations.

$\nu \setminus \alpha$	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188

Table 1: Values of $\chi^2_{\alpha,\nu}$: the entries in the table correspond to values of x, such that $P(\chi^2_\nu > x) = \alpha$, where χ^2_ν correspond to a chi-squared distributed variable with ν degrees of freedom.

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

	x											
λ	5	6	7	8	9	10	11	12	13	14	15	16
2	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.56	0.38	0.24	0.13	0.07	0.03	0.01	0.00	0.00	0.00	0.00	0.00
10	0.97	0.93	0.87	0.78	0.67	0.54	0.42	0.30	0.21	0.14	0.08	0.05
25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98

Table 3: Exceedance probabilities for Poison(λ) distribution, i.e., $P(X \geq x)$ where $X \sim$ Poison(λ), where $\lambda \in \{2, 5, 10, 25\}$ and $x \in \{5, 6, \ldots, 16\}$.