2 November 2015, 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** **10 Marks**.
   Let $X_1, \ldots, X_n$ be a sample of independent, identically distributed random variables, with density

$$f_\theta(x) = \begin{cases} \frac{2}{3\theta}(1 - \frac{x}{3\theta}) & 0 < x < 3\theta \\ 0 & \text{elsewhere} \end{cases}$$

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

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

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

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

   Let $\hat{\theta} = \hat{\theta}(X)$ be an unbiased estimate 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) Use Cauchy-Schwarz to 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. **Survival regression** $\boxed{\textbf{35 Marks}}$.

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:

$$Y_i \sim Ex(\lambda x_i), \quad i = 1, 2, \dots, n$$

i.e.

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

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

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

(b) Determine whether the MLE is a sufficient statistic. $\boxed{\text{5 Marks}}$

(c) Exponential distributions are often used to model survival times. A researcher wants to find the relationship between the price of a light bulb and its life span. She samples 40 different light bulbs and hypothesizes the above model, whereby for $i = 1, \dots, 40$

- $x_i = \frac{1}{\text{(euro) price of light bulb } i}$
- Y_i = life span of light bulb i (in years).

The plot of the relationship between price and lifespan is given in the plot below. We are given that $\sum_{i=1}^{40} x_i y_i = 80$.

i. A consumer organization wants to know whether, based on these data, it can reject $\lambda = 1$. Based on the asymptotic distribution of the MLE as test-statistic, set up a hypothesis test with null and alternative hypothesis to see if there is sufficient evidence for a relationship at a significance level of $\alpha = 0.05$. $\boxed{\text{10 Marks}}$

ii. Calculate the (numeric!) 95% confidence interval for λ based on the asymptotic distribution of the likelihood ratio statistic. Draw this confidence interval in a plot of the likelihood ratio statistic versus λ. [Hint: you are allowed to "read off" the numeric CI from this plot, but provide the inequality that you would like to solve.] $\boxed{\text{10 Marks}}$

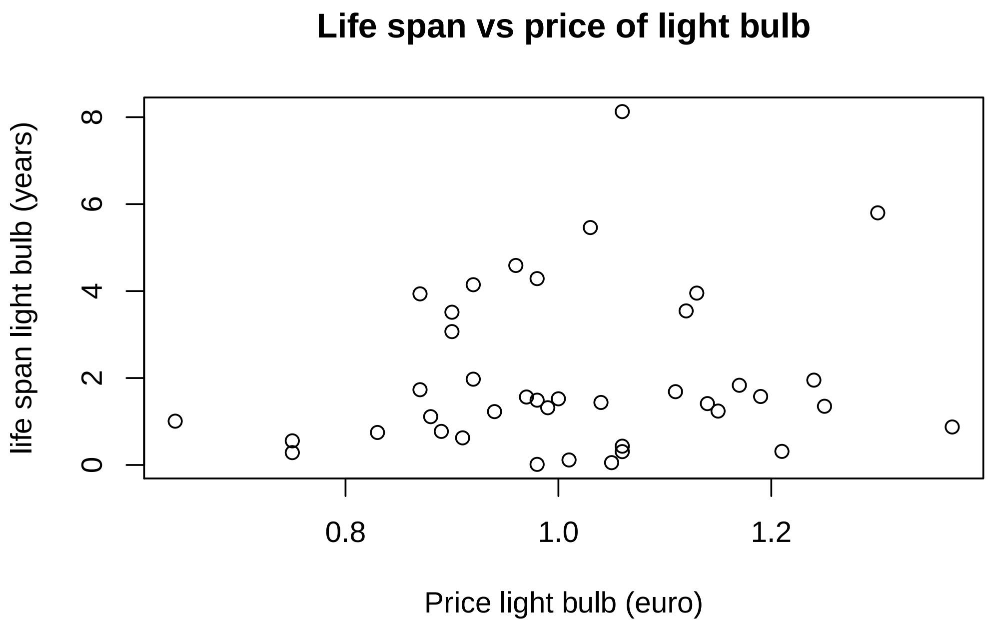

Figure 1: 40 measurements of light bulb life spans versus their price.

4. **Optimal testing** **20 Marks**. A demographer wants to know whether the birth probability for boys versus girls is 0.5 vs 0.5 OR 0.51 vs 0.49. He observes a sequence of 1600 independent (single) births. Let p be the probability of a boy. He therefore decides to test,

$$\begin{aligned} H_0 : &\quad p = 0.5 \\ H_1 : &\quad p = 0.51 \end{aligned}$$

On the basis of these 1600 measurements, he wants to device an optimal test, using a significance level $\alpha = 0.05$.

(a) Device the **most powerful test** for deciding between the two hypotheses on the basis of these 1600 measurements. In other words, determine the Critical Region. Hint: You are allowed to make a normal approximation for calculating the critical region. [15 Marks]

(b) What is the power of this test? Hint: you can again make use of a normal approximation. [5 Marks]

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.