Department of Mathematics	Partial exam II and full exam – Statistics
VU University Amsterdam	20 December 2016

Write your name, your student number and that of your teaching assistant at the top of the page! **The use of a (non-graphical) calculator is allowed.** After correction exams can be inspected through the educational office.

Full or partial exam:
→ *Full exam* : **exercises 1,2,3,4**. Grade= $(total + 4)/4$.
→ *Partial exam II* : **exercises 2,3,4 (not 2c)**. Grade= $(total + 3)/3$.
Any exam handed in later than two hours after the start will be marked as a full exam and the result of partial exam I will be dropped automatically.

Good luck!

Exercise 1 [7 points (full exam only)]

Let $X_1, \dots, X_n$ be a sample from the lognormal distribution (see Appendix 1) parameters $\mu = 0$ and unknown $\sigma > 0$.

a. [3 points (full exam only)] Determine the distribution function F_{σ^2} and the quantile function. *Hint:* Express both in terms of the (inverse of the) cumulative distribution of the standard normal denoted by $\Phi_{0,1}(x)$.
b. [1 points (full exam only)] Give the moment estimator of σ^2.
c. [3 points (full exam only)] Assume that the prior on σ^2 is an inverse gamma distribution with shape parameter α and scale parameter β. Compute the corresponding Bayes estimator of σ^2.

Exercise 2 [7 points (partial exam); 9 points (full exam)]

Two groups of athletes (amateur and professional) compete on the 400 meter sprint. To investigate which group runs faster nine representatives of each group are randomly selected and their time on the 400 meter sprint measured. This yields measured times $X_1, \dots, X_9$ and $Y_1, \dots, Y_9$. The following has been observed: $\overline{x} = 45.22$ seconds, $\overline{y} = 41.56$ seconds, $\sum_{i=1}^{9}(x_i - \overline{x})^2 = 195.56$ and $\sum_{i=1}^{9}(y_i - \overline{y})^2 = 160.22$.

a. [2 points] Formulate an appropriate statistical model and suitable null (H_0) and alternative (H_a) hypothesis for this situation.
b. [5 points] Test the null hypothesis from part a.) at (significance) level $\alpha_0 = 0.05$ (see Appendix 2). Report *i)* the test statistic, *ii)* the distribution of the test statistic under the null hypothesis, *iii)* the exceedance probability or the critical region, and *iv)* the conclusion of the test. Clearly state the assumptions made.
c. [2 points (full exam only)] Give the definition of the power function of a test and describe how, in the current situation, the power of the test (under H_a) may be increased.

Exercise 3 [10 points]

Let $X_1, \dots, X_n$ be a sample from the (unconventionally parametrized) normal distribution with mean zero, unknown variance $\theta > 0$ and density $p_\theta(x) = (2\pi\theta)^{-1/2}\exp[-x^2/(2\theta)]$.

a. [5 points] Report the likelihood-ratio statistic $\lambda_n(X_1, \dots, X_n)$ for testing the null hypothesis $H_0 : \theta = 1$ vs. the alternative hypothesis $H_a : \theta \neq 1$.
b. [3 points] Report the limiting distribution of $2\log[\lambda_n(X_1, \dots, X_n)]$ and give the $(1 - \alpha_0)$-confidence interval for θ based on the likelihood-ratio test of $H_0 : \theta = 1$ vs. the alternative hypothesis $H_a : \theta \neq 1$.
c. [2 points] Explain the difference between a pivot and an approximate pivot. Is the confidence interval reported in part (b) based on a pivot or approximate pivot?

Exercise 4 [10 points]
Let $X_1, \ldots, X_n$ be a sample from the distribution with probability density $p_\theta(x) = \frac{1}{\theta}\exp(-x/\theta)$ for $x > 0$ (and zero otherwise) and with unknown parameter $\theta > 0$. The maximum likelihood estimator of θ is equal to $\overline{X}$.

a. [3 points] Give a sufficient and complete statistic.
b. [2 points] Give a UMVU estimator of θ. Clarify why it is a UMVU estimator.
c. [2 points] Report the Fisher-information i_θ. (*Hint: you may wish to make use of the lemma:* $i_\theta = \mathbb{V}_\theta\big[\dot{\ell}_\theta(X_1)\big] = -\mathbb{E}_\theta\big[\ddot{\ell}_\theta(X_1)\big]$.)
d. [2 points] Assume the UMVU estimator of θ sought for in part (b) is $\overline{X}$. Report its variance. (*Hint: you may wish to make use of the lemma on* $\mathbb{V}_\theta[\dot{\ell}(X_1)]$ *again*).
e. [1 points] Is the Cramér-Rao underbound for the variance of an unbiased estimator of θ sharp here?

Appendix 1: Some relevant distributions

Lognormal distribution
A random variable X follows a lognormal distribution with support over $\mathbb{R}_{\geq 0}$ with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$ if it has density:

$$f_{\mu,\sigma^2}(x) = (2\pi\sigma^2 x^2)^{-1/2} \exp\{-[\log(x) - \mu]^2/(2\sigma^2)\}.$$

Its expectation and variance are $\mathbb{E}(X) = \exp(\mu^2 + \frac{1}{2}\sigma^2)$ and $\mathbb{V}(X) = [\exp(\sigma^2) - 1]\exp(2\mu^2 + \sigma^2)$.

Inverse-gamma distribution
A random variable X follows an inverse gamma distribution with support over $\mathbb{R}_{>0}$ with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ if it has density:

$$f_{\alpha,\beta}(x) = \beta^\alpha[\Gamma(\alpha)]^{-1}x^{-\alpha-1}\exp(-\beta/x).$$

Its expectation and variance are $\mathbb{E}(X) = \frac{\beta}{\alpha-1}$ (for $\alpha > 1$) and $\mathbb{V}(X) = \frac{\beta}{(\alpha-1)^2(\alpha-2)}$ (for $\alpha > 2$), respectively.

Appendix 2: Table t-distribution

df	0.6	0.7	0.75	0.8	0.85	0.9	0.925	0.95	0.975	0.98	0.99	0.999
1	0.32	0.73	1	1.38	1.96	3.08	4.17	6.31	12.71	15.89	31.82	318.31
2	0.29	0.62	0.82	1.06	1.39	1.89	2.28	2.92	4.3	4.85	6.96	22.33
3	0.28	0.58	0.76	0.98	1.25	1.64	1.92	2.35	3.18	3.48	4.54	10.21
4	0.27	0.57	0.74	0.94	1.19	1.53	1.78	2.13	2.78	3	3.75	7.17
5	0.27	0.56	0.73	0.92	1.16	1.48	1.7	2.02	2.57	2.76	3.36	5.89
6	0.26	0.55	0.72	0.91	1.13	1.44	1.65	1.94	2.45	2.61	3.14	5.21
7	0.26	0.55	0.71	0.9	1.12	1.41	1.62	1.89	2.36	2.52	3	4.79
8	0.26	0.55	0.71	0.89	1.11	1.4	1.59	1.86	2.31	2.45	2.9	4.5
9	0.26	0.54	0.7	0.88	1.1	1.38	1.57	1.83	2.26	2.4	2.82	4.3
10	0.26	0.54	0.7	0.88	1.09	1.37	1.56	1.81	2.23	2.36	2.76	4.14
11	0.26	0.54	0.7	0.88	1.09	1.36	1.55	1.8	2.2	2.33	2.72	4.02
12	0.26	0.54	0.7	0.87	1.08	1.36	1.54	1.78	2.18	2.3	2.68	3.93
13	0.26	0.54	0.69	0.87	1.08	1.35	1.53	1.77	2.16	2.28	2.65	3.85
14	0.26	0.54	0.69	0.87	1.08	1.35	1.52	1.76	2.14	2.26	2.62	3.79
15	0.26	0.54	0.69	0.87	1.07	1.34	1.52	1.75	2.13	2.25	2.6	3.73
16	0.26	0.54	0.69	0.86	1.07	1.34	1.51	1.75	2.12	2.24	2.58	3.69
17	0.26	0.53	0.69	0.86	1.07	1.33	1.51	1.74	2.11	2.22	2.57	3.65
18	0.26	0.53	0.69	0.86	1.07	1.33	1.5	1.73	2.1	2.21	2.55	3.61
19	0.26	0.53	0.69	0.86	1.07	1.33	1.5	1.73	2.09	2.2	2.54	3.58
20	0.26	0.53	0.69	0.86	1.06	1.33	1.5	1.72	2.09	2.2	2.53	3.55
21	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.72	2.08	2.19	2.52	3.53
22	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.72	2.07	2.18	2.51	3.5
23	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.71	2.07	2.18	2.5	3.48
24	0.26	0.53	0.68	0.86	1.06	1.32	1.49	1.71	2.06	2.17	2.49	3.47
25	0.26	0.53	0.68	0.86	1.06	1.32	1.49	1.71	2.06	2.17	2.49	3.45
26	0.26	0.53	0.68	0.86	1.06	1.31	1.48	1.71	2.06	2.16	2.48	3.43
27	0.26	0.53	0.68	0.86	1.06	1.31	1.48	1.7	2.05	2.16	2.47	3.42
28	0.26	0.53	0.68	0.85	1.06	1.31	1.48	1.7	2.05	2.15	2.47	3.41
29	0.26	0.53	0.68	0.85	1.06	1.31	1.48	1.7	2.05	2.15	2.46	3.4
30	0.26	0.53	0.68	0.85	1.05	1.31	1.48	1.7	2.04	2.15	2.46	3.39
31	0.26	0.53	0.68	0.85	1.05	1.31	1.48	1.7	2.04	2.14	2.45	3.37
32	0.26	0.53	0.68	0.85	1.05	1.31	1.47	1.69	2.04	2.14	2.45	3.37
33	0.26	0.53	0.68	0.85	1.05	1.31	1.47	1.69	2.03	2.14	2.44	3.36
34	0.26	0.53	0.68	0.85	1.05	1.31	1.47	1.69	2.03	2.14	2.44	3.35
35	0.26	0.53	0.68	0.85	1.05	1.31	1.47	1.69	2.03	2.13	2.44	3.34
36	0.26	0.53	0.68	0.85	1.05	1.31	1.47	1.69	2.03	2.13	2.43	3.33
37	0.26	0.53	0.68	0.85	1.05	1.3	1.47	1.69	2.03	2.13	2.43	3.33
38	0.26	0.53	0.68	0.85	1.05	1.3	1.47	1.69	2.02	2.13	2.43	3.32
39	0.26	0.53	0.68	0.85	1.05	1.3	1.47	1.68	2.02	2.12	2.43	3.31
40	0.26	0.53	0.68	0.85	1.05	1.3	1.47	1.68	2.02	2.12	2.42	3.31

Table 1: Quantiles (columns) of the t-distribution with 1 to 40 degrees of freedom (rows).