Econ 4130 Eksamen 2008 H

ANSWERS

Settet består av 9 delspørsmål som alle anbefales å telle likt. Svar er gitt i << ….. >>.

Problem 1

a. Suppose that the random variable (rv) *X* has the following cumulative distribution function (cdf)

$$G(x) = \begin{cases} 1-(1+3x+\frac{9}{2}x^2)\cdot e^{-3x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Show that $X \sim \Gamma(3,3)$, i.e., that the corresponding density function (pdf) derived from *G* is the same as the pdf of the gamma distribution $\Gamma(3,3)$.

[**Hint:** Calculate the pdf based on *G* and compare with the pdf-formula for the $\Gamma(3,3)$-distribution]

<<**Answer:** The pdf is for $x > 0$

$$g(x) = G'(x) = -(3+9x)e^{-3x} - (1+3x+(9/2)x^2)e^{-3x}(-3) = 3e^{-3x}(1+3x+(9/2)x^2 - 1 - 3x) =$$
$$= \frac{27}{2}x^2e^{-3x}$$

The $\Gamma(3,3)$ pdf is $\frac{3^3}{\Gamma(3)}x^{3-1}e^{-3x} = \frac{27}{2}x^2e^{-3x}$ for $x > 0$; i.e. the same. >>

b. Let x_p denote the *p*-quantile in the distribution of *X*, i.e., $P(X \le x_p) = p$. Table 1 shows some quantiles of this distribution that may be needed later in the problem set:

Table 1 ***p*-quantiles of** $\Gamma(3,3)$

p	0.01	0.025	0.05	0.10	0.5	0.90	0.95	0.975	0.99
x_p	0.145	0.206	0.273	0.367	0.891	1.774	2.099	2.408	2.802

(i) Check that the median of *X* is equal to the value indicated in the table.

(ii) Find (from the table) a 98% variation interval for *X*, i.e., an interval $[c_1, c_2]$ such that $P(c_1 \le X \le c_2) = 0.98$

<<**Answer:** (i) According to the table, the median is $x_{0.5} = 0.891$. We get

$$G(0.891) = 1 - (1 + 3(0.891) + \frac{9}{2}(0.891)^2) \cdot e^{-3(0.891)} = 0.500$$

(ii) [0.145, 2.802] >>

c. Suppose (X, Y) has a joint distribution such that the marginal distribution of *X* is $\Gamma(3,3)$ distributed, and such that the regression function (i.e., the conditional expectation given $X = x$) of *Y* with respect to *X*, is $E(Y \mid x) = 1 + 2x$. Suppose further that the conditional variance is: $\text{var}(Y \mid x) = 4$ for all *x*. Calculate under these conditions

(i) $E(Y)$

(ii) $\text{var}(Y)$

(iii) The correlation coefficient between *X* and *Y*.

[**Hint:** You don't need to derive the expectation and variance of *X* - use the formulas for the gamma distribution. To find the covariance, calculate first $E(XY) = E[X \cdot E(Y \mid X)] = \cdots \text{etc}$.]

<<**Answer:** In $\Gamma(3,3)$ we have $E(X) = 1, \quad \text{var}(X) = 1/3$. Now

$$E(Y) = E\left[E(Y \mid X)\right] = E(1 + 2X) = 1 + 2E(X) = 3$$

$$\text{var}(Y) = E\left[\text{var}(Y \mid X)\right] + \text{var}\left[E(Y \mid X)\right] = E(4) + \text{var}(1 + 2X) = 4 + \frac{4}{3} = \frac{16}{3}$$

$$E(XY) = E[X \cdot E(Y \mid X)] = E(X + 2X^2) = 1 + 2\left(\frac{1}{3} + 1\right) = \frac{11}{3}$$

$$\operatorname{cov}(X,Y) = \frac{11}{3} - 1 \cdot 3 = \frac{2}{3}$$

$$\operatorname{corr}(X,Y) = \frac{2/3}{\sqrt{(1/3)(16/3)}} = \frac{1}{2}$$ >>

Problem 2

a. Suppose that the rv X has a distribution with pdf given by

$$f(x;\alpha) = \begin{cases} \frac{\alpha}{3^\alpha} x^{\alpha-1} & \text{for } 0 < x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha > 0$ is a parameter.

(i) Make rough sketches of the graph of $f(x;\alpha)$ in the three cases, $\alpha = 0.5,\ \alpha = 1,\ \alpha = 3$ respectively.

(ii) Is there any value of α for which the distribution is symmetric?

(iii) Find $E(X)$.

<<**Answer:** (i)….(ii)… (iii) $E(X) = \frac{\alpha}{3^\alpha}\int_0^3 x^\alpha dx = \frac{\alpha}{3^\alpha} \Big|_0^3 \frac{x^{\alpha+1}}{\alpha+1} = \frac{3\alpha}{\alpha+1}$ >>

b. Show that $Y = \ln(3/X) \sim \exp(\alpha)$, i.e., Y is exponentially distributed with parameter α.

<<**Answer:** Clearly Y varies in $[0,\infty)$ when X varies in $(0, 3]$. Let $y > 0$. Then

$$P(Y \le y) = P\left(\frac{3}{X} \le e^y\right) = P(X \ge 3e^{-y}) = 1 - F_X(3e^{-y}).$$

Hence $f_Y(y) = \frac{\partial}{\partial y}(1 - F_X(3e^{-y})) = -f(3e^{-y};\alpha)(-3e^{-y}) = \frac{\alpha}{3^\alpha}\left(3e^{-y}\right)^{\alpha-1} 3e^{-y} = \alpha e^{-\alpha y}$, (or, directly from substituting in $F_X(x)$, which is easily seen to be $\left(\frac{x}{3}\right)^{\alpha}$:

$F_Y(y) = 1 - F_X(3e^{-y}) = 1 - e^{-\alpha y}$) >>

c. Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (*iid*) with $X_i \sim f(x;\alpha)$ as in question **a**.

(i) Show that the maximum likelihood estimator (mle) for α is $\hat{\alpha} = \frac{1}{\overline{Y}}$ where $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $Y_i = \ln\left(\frac{3}{X_i}\right)$.

(ii) Show that $\frac{\alpha}{\hat{\alpha}}$ is distributed exactly as $\Gamma(n,n)$.
[**Hint:** You may first exploit a result saying that if $Y_1, Y_2, \ldots, Y_n \sim iid$ with $Y_i \sim e\mathrm{xp}(\lambda)$, then $\overline{Y} \sim \Gamma(n, n\lambda)$. Then, for example, next utilize the moment generating function (mgf) of $\overline{Y}$.]

<< **Answer:** (i) Log likelihood: $l(\alpha) = (\alpha - 1)\sum \ln(x_i) + n\ln(\alpha) - n\alpha\ln(3)$ with derivative $l' = \sum \ln(x_i) + \frac{n}{\alpha} - n\ln(3) = \frac{n}{\alpha} - \sum \ln\left(3/x_i\right) = 0$, which gives the suggested solution. The second derivative being negative, we have a maximum.

(ii) We have $V = \frac{\alpha}{\hat{\alpha}} = \alpha\overline{Y}$, where $\overline{Y} \sim \Gamma(n, \alpha n)$ with mgf $M_{\overline{Y}}(t) = \left(\frac{n\alpha}{n\alpha - t}\right)^n$.

Then *V* has mgf $M_V(t) = M_{\overline{Y}}(\alpha t) = \left(\frac{n\alpha}{n\alpha - \alpha t}\right)^n = \left(\frac{n}{n - t}\right)^n$, which identifies $\Gamma(n,n)$. >>

d. We want to investigate if there is strong evidence that the distribution of *X* is right skewed, i.e., with tail to the right, based on observing $X_1, X_2, \ldots, X_n$ in **c**. In other words, we want to test $H_0 : \alpha \geq 1$ versus $H_1 : \alpha < 1$. The preferred level of significance is 5%. Explain how the following test,

(1) Reject H_0 if $\sqrt{n}\dfrac{\hat{\alpha}-1}{\hat{\alpha}} \leq -1.645$

can be justified by asymptotic theory for mle's and Slutsky's lemma with an approximate level of significance 5%.

[**Hint:** Find first the Fisher information for one observation.]

<<**Answer:** The Fisher information:

$\ln(f(X;\alpha) = (\alpha-1)\ln(X) + \ln(\alpha) - \alpha\ln(3)$ gives

$$\frac{\partial}{\partial\alpha}\ln f = \ln(X) + \frac{1}{\alpha} - \ln(3), \quad \text{and} \quad \frac{\partial^2}{\partial\alpha^2}\ln f = -\frac{1}{\alpha^2}$$

Hence the Fisher information: $I(\alpha) = -E\left(\dfrac{\partial^2}{\partial\alpha^2}\ln f\right) = \dfrac{1}{\alpha^2}$

Large sample mle theory gives

$$\sqrt{nI(\alpha)}(\hat{\alpha}-\alpha) = \sqrt{n}\frac{\hat{\alpha}-\alpha}{\alpha} \underset{n\to\infty}{\overset{D}{\to}} N(0,1)$$

Then Slutsky gives

$$\sqrt{n}\frac{\hat{\alpha}-\alpha}{\hat{\alpha}} = \frac{\alpha}{\hat{\alpha}} \cdot \sqrt{n}\frac{\hat{\alpha}-\alpha}{\alpha} \underset{n\to\infty}{\overset{D}{\to}} N(0,1) \text{ , since } \frac{\alpha}{\hat{\alpha}} \underset{n\to\infty}{\overset{P}{\to}} 1.$$

From this we take the test statistic $V = \sqrt{n}\dfrac{\hat{\alpha}-1}{\hat{\alpha}}$, which is approximately $\sim N(0,1)$ if $\alpha = 1$. Rejecting H_0 when *V* is smaller than the 0.05 quantile in $N(0,1)$ (i.e., -1.645), then gives a test with approximate level of significance, 5%.

\>>

e. Suppose that X represents the processing time for a new, standardized but expensive industrial process, and that we have only three independent observations of *X*. The observations are $x_1 = 0.2, x_2 = 0.6, x_3 = 0.6$, of X_1, X_2, X_3

respectively. Recognizing that $n = 3$ is somewhat small to appeal to asymptotic theory, we prefer exact small-sample inference, which, in fact, is possible in this case.

(i) Using **c**(ii) for $n = 3$ and **table 1** in **Problem 1**, derive and calculate from the data, an exact 95% confidence interval for α, i.e., find observable rv's, L and U such that $P(L \le \alpha \le U) = 0.95$.

[**Hint:** Start with a 95% variation interval for $\alpha/\hat{\alpha}$ similar to that in **Problem 1b**(ii).]

(ii) Using **c**(ii) again, derive and perform an exact 5%-level test for $H_0 : \alpha \ge 1$ versus $H_1 : \alpha < 1$. In other words, find a critical value, *c*, such that $P(\hat{\alpha} < c) = 0.05$ if $\alpha = 1$ is true.

<<**Answer:** (i): Table 1 gives

$$0.95 = P(0.206 \le \alpha/\hat{\alpha} \le 2.408) = P(0.206 \cdot \hat{\alpha} \le \alpha \le 2.408 \cdot \hat{\alpha})$$

Data

X_i	Y_i
0.2	2.708
0.6	1.609
0.6	1.609
$\bar{Y}$	1.976
$\hat{\alpha}$	0.506

95% CI: $[0.206 \cdot \hat{\alpha},\ 2.408 \cdot \hat{\alpha}] = [0.104,\ 1.219]$

(ii) The test statistic here is simply $\hat{\alpha}$. If $\alpha = 1$, $\frac{1}{\hat{\alpha}} \sim \Gamma(3,3)$,

$$0.05 = P(\hat{\alpha} \le c) = P(1/\hat{\alpha} \ge 1/c) \quad \Rightarrow \quad 1/c = 2.099 \quad \Rightarrow \quad c = 0.477$$

Hence, the exact 5%-level test is "Reject H_0 if $\hat{\alpha} \le 0.477$"
Conclusion: Don't reject. >>

f. Suppose you insist to use the asymptotic 5%-level test in (1) in spite of *n* being only 3. Calculate the exact level of significance for that test and compare with the approximate (also called *nominal*) level of significance used, i.e., 5%.

<<**Answer:** The approximate test is

Reject H_0 if

$$\sqrt{3}\,\frac{\hat{\alpha}-1}{\hat{\alpha}} \le -1.645 \quad \Leftrightarrow \quad 1-\frac{1}{\hat{\alpha}} \le -\frac{1.645}{\sqrt{3}} \quad \Leftrightarrow \quad \frac{1}{\hat{\alpha}} \ge 1+\frac{1.645}{\sqrt{3}} = 1.950$$

(Since $1/1.95 = 0.513$, the approximate test would give rejection)

Hence the exact level is, using G from Problem 1,

$P_{\alpha=1}(1/\hat{\alpha} \ge 1.95) = 1 - G(1.95) = 0.069$ >>