HG
Nov. 10

ECON 4130 H10

Extra exercises for no-seminar week 47

(Solutions will be put on the net at the end of the week)

Exercise A

(*A slightly altered version of* ***Exam 2005H -"utsatt prove".***)

Problem 1

a. Let X and Y be two continuous random variables, both varying in the interval [0, 1]. The joint cumulative probability function (cdf) of X and Y is given by

$$F(x, y) = xy[1 + (0.8)(1 - x)(1 - y)] \quad \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1$$

Show that the joint density function for X and Y is given by

$$f(x, y) = 1 + (0.8)(1 - 2x)(1 - 2y) \quad \text{for } 0 \le x \le 1 \text{ og } 0 \le y \le 1 .$$

b. Show that both X and Y marginally are uniformly distributed over [0, 1]. Find the expectation and variance for X and Y. Are X and Y stochastically independent? Calculate $P(Y \le 0.5)$ and $P(Y \le 0.5 \mid X \le 0.5)$.

c. Sketch the conditional density for Y, $f(y \mid x)$, in a graph when $x = 1/2$, and $x = 1$ respectively (i.e. altogether two graphs).

Calculate the regression function, $\mathrm{E}(Y \mid x)$, and sketch a graph of this.

d. Find $\mathrm{E}(XY)$ by the "double expectation" theorem, i.e.,

$$\mathrm{E}(XY) = \mathrm{E}[\mathrm{E}(XY \mid X)] = \mathrm{E}[X\, \mathrm{E}(Y \mid X)] \text{ etc.}$$

Calculate the correlation coefficient between X and Y.

Problem 2

a. Let Z be chi-square distributed with r degrees of freedom (written in short $Z \sim \chi^2(r)$). According to the textbook, this is the same as the gamma distribution with parameters $\alpha = r/2$ and $\lambda = 1/2$ (in short $Z \sim \Gamma\left(\frac{r}{2}, \frac{1}{2}\right)$). Use this and known properties of the gamma distribution to show that

(i) $\mathrm{E}(Z) = r, \qquad \mathrm{Var}(Z) = 2r$

(ii) For $r = 2$, find the median of Z. Is the median smaller or larger than $\mathrm{E}(Z)$?

b. According to the textbook, if X is standard normally distributed ($X \sim N(0,1)$), then $X^2 \sim \chi^2(1)$.
(i) Use this and known properties of the gamma distribution to show that, if $X_1, X_2, \ldots, X_n$ are *iid* with $X_i \sim N(0,1)$, then $Z = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$. (*Note that iid means "independent and identically distributed".*)
(ii) Use for example the central limit theorem to justify that $Z = \sum_{i=1}^{n} X_i^2$ is approximately $N(n, 2n)$ distributed for large n.

c. Let q_n denote the 95% quantile (i.e., such that $P(Z \le q_n) = 0.95$) in the $\chi^2(n)$ distribution. Exact values for q_n can be found in the χ^2 - table in Rice. Let q_n^* denote the approximate 95% quantile determined by the approximate distribution for Z derived in section **b.** (i.e., $N(n, 2n)$). Show that

$$q_n^* = n + \sqrt{2n} \cdot 1{,}64$$

Calculate the approximation error $q_n - q_n^*$ for $n = 30, 60, 120$, and comment on the result.

d. The approximate quantile in section **c.** can be improved somewhat. Let $X_1, X_2, \ldots, X_n \sim N(0, 1)$ as in section **b**., and put

$$\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Then, by the central limit theorem combined with some other theory, it can be proven that $\sqrt{2n}\left(\sqrt{\overline{Z}} - 1\right) \xrightarrow{D} N(0,1)$ as $n \to \infty$ (you do not need to justify this here). Show that, based on the approximate normal distribution for $\sqrt{\overline{Z}}$, we can derive an alternative approximation, q_n^{**}, to q_n (i.e., the 95% quantile of $Z = n\overline{Z}$), given by

$$q_n^{**} = \frac{1}{2}\left(1.64 + \sqrt{2n}\right)^2.$$

Compare q_n^{**} with q_n^*.for $n = 30, 60, 120$.

[*Note that this approximation is similar but slightly different from the suggestion given in the* χ^2*-table in Rice, which represents a third approximation*.]

Problem 3

Suppose $X_1, X_2, \ldots, X_n$ are *iid* with $X_i \sim \Gamma(9, \lambda)$, where $\lambda > 0$ is unknown. We assume here that the shape parameter is known, $\alpha = 9$.

(i) Derive the maximum likelihood estimator for λ based on $X_1, X_2, \ldots, X_n$.

(ii) Develop an approximate 95% confidence interval for λ, and calculate the interval for the data

$$n = 100 \quad \text{and} \quad \sum_{i=1}^{n} X_i = 630$$

Exercise B

Problem 1 from postponed exam 2006H

a. Let the random variable (rv.) X be exponentially distributed with density function (pdf)

$$g(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

(i) Find (i.e., calculate) $P(X < E(X))$ and $P(X > E(X))$.

(ii) Show that the median of X is approximately 0,693[1].

b. Let $Y = \dfrac{1}{1 + X/\theta}$ where $\theta > 0$ is an unknown parameter. Show that the pdf of Y is given by

$$f(y;\theta) = \begin{cases} \theta e^{\theta} e^{-\frac{\theta}{y}} \cdot \dfrac{1}{y^2} & \text{for } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

c. Let $Y_1, Y_2, \ldots, Y_n$ be *iid* with $Y_i \sim f(y_i;\theta)$ as in **b.**

(i) Derive the maximum likelihood estimator (mle), $\hat{\theta}$, for θ, based on $Y_1, Y_2, \ldots, Y_n$.

(ii) Show that the Fisher information for one observation is $I(\theta) = \dfrac{1}{\theta^2}$

(iii) Calculate an approximate 95% confidence interval (CI) for θ when the estimate is $\hat{\theta} = 1{,}85$ and $n = 100$.

[1] Printing error in original version.

d. Let θ_0 denote the true value of θ.

(i) How many observations, *n*, is needed (approximately) to ensure that the estimation error, $|\hat{\theta}-\theta_0|$, is less than 0,1 with probability at least 0,95, assuming that the true θ is $\theta_0 = 2$? [**Hint:** solve $P(|\hat{\theta}-\theta_0| \leq 0,1) \geq 0,95$ with respect to *n*.]

(ii) For *any* θ_0, how many observations, *n*, is needed to ensure that the relative estimation error, $|\hat{\theta}-\theta_0| / \theta_0$ is less than 0,1 with probability at least 0,95?