Solution of assignment in STK4011/STK9011-f18

Problem 1

a) X and Y has probability density $f_X(x) = I_{[0,1]}$, where $I_A(\cdot)$ is the indicator function of the set A. Also $0 \leq X + Y \leq 1$. Let $U = X + Y$. Then

$$\begin{aligned}
f_U(u) &= \int_{-\infty}^{\infty} I_{[0,1]}(u-x)I_{[0,1]}(x)dx = \int_0^1 I_{[0,1]}(u-x)dx \\
&= \begin{cases} \int_0^u I_{[0,1]}(u-x)dx & \text{if } 0 \leq u \leq 1 \\ \int_{u-1}^1 I_{[0,1]}(u-x)dx & \text{if } 1 \leq u \leq 2 \end{cases} \\
&= \begin{cases} u & \text{if } 0 \leq u \leq 1 \\ 2-u & \text{if } 1 \leq u \leq 2 \end{cases}
\end{aligned}$$

b) The moment generating function of X is $M_X(t) = E[e^{tX}] = \int_0^1 e^{tx}dx = \frac{1}{t}(e^t - 1)$. Hence $M_U(t) = M_{X+Y}(t) = [\frac{1}{t}(e^t-1)][\frac{1}{t}(e^t-1)] = (e^{2t} - 2e^t + 1)/t^2$.

On the other hand also

$$\begin{aligned}
M_U(t) = & \quad \int_0^2 e^{tu} f_U(u)du = \int_0^1 ue^{tu}du + \int_1^2 (2-u)e^{tu}du \\
= & \quad 2\tfrac{1}{t}e^{tu}|_{u=1}^{u=2} + u\tfrac{1}{t}e^{tu}|_{u=0}^{u=1} - \int_0^1 \tfrac{1}{t}e^{tu}du - u\tfrac{1}{t}e^{tu}|_{u=1}^{u=2} + \int_1^2 \tfrac{1}{t}e^{tu}du \\
= & \quad \tfrac{2}{t}e^{2t} - \tfrac{2}{t}e^t + \tfrac{1}{t}e^t - \tfrac{1}{t^2}e^t + \tfrac{1}{t^2} - \tfrac{2}{t}e^{2t} + \tfrac{1}{t}e^t + \tfrac{1}{t^2}e^{t2} - \tfrac{1}{t^2}e^t \\
= & \quad (e^{2t} - 2e^t + 1)/t^2.
\end{aligned}$$

c) Let $V = U + Z = X + Y + Z$. Then $0 \leq V \leq 3$ and

$$\begin{aligned}
f_V(v) &= \int_{-\infty}^{\infty} f_U(v-z)f_Z(z)dz \\
&= \begin{cases} \int_0^v (v-z)I_{[0,1]}(v-z)dz & \text{if } 0 \leq v \leq 1 \\ \int_{v-1}^1 (v-z)I_{[0,1]}(v-z)dz + \int_0^{v-1}[2-(v-z)]I_{(1,2]}(v-z)dz & \text{if } 1 \leq v \leq 2 \\ \int_{v-2}^1 [2-(v-z)]I_{(1,2]}(v-z)dz & \text{if } 2 \leq v \leq 3 \end{cases} \\
&= \begin{cases} v^2 - v^2/2 = v^2/2 & \text{if } 0 \leq v \leq 1 \\ v(2-v) - \frac{1}{2} + \frac{1}{2}(v-1)^2 + (v-1)(2-v) + \frac{1}{2}(v-1)^2 & \text{if } 1 \leq v \leq 2 \\ (3-v)(2-v) + \frac{1}{2} - \frac{1}{2}(v-2) & \text{if } 2 \leq v \leq 3 \end{cases} \\
&= \begin{cases} v^2 - v^2/2 = v^2/2 & \text{if } 0 \leq v \leq 1 \\ -v^2 + 3v - 1 - \frac{1}{2} & \text{if } 1 \leq v \leq 2 \\ \frac{1}{2}v^2 - 3v + 4 + \frac{1}{2} & \text{if } 2 \leq v \leq 3 \end{cases}
\end{aligned}$$

Problem 2

a)

$$F_X(x) = \beta\alpha^\beta \int_\alpha^x \frac{1}{y^{\beta+1}} dy = \beta\alpha^\beta \frac{1}{-\beta} y^{-\beta}|_{y=\alpha}^x = 1 - (\frac{\alpha}{x})^\beta \text{ for } x > \alpha.$$

b)

$$E[X^k] = \beta\alpha^\beta \int_\alpha^\infty \frac{x^k}{x^{\beta+1}} = \beta\alpha^\beta \frac{1}{k-\beta} x^{k-\beta}|_{x=\alpha}^\infty = \begin{cases} \frac{\beta\alpha^k}{\beta-k} & \text{if } k < \beta. \\ \infty & \text{else} \end{cases}$$

c)

$$\begin{aligned} f(x|\alpha,\beta) &= \begin{cases} \frac{\beta\alpha^\beta}{x^{\beta+1}} & \text{if } \alpha \le x < \infty \quad \alpha, \beta > 0 \\ 0 & \text{else} \end{cases} \\ &= \exp(-(\beta+1)\log x)\beta\alpha^\beta I_{[\alpha,\infty)}(x). \end{aligned}$$

By defining $h(x) = I_{[\alpha,\infty)}(x)$ which is not involving unknown parameters since α is known, $c(\beta) = \beta\alpha^\beta$, $t(x) = \log x$ and $w(\beta) = -(\beta+1)$ $f(x|\beta)$ can be written on the form $h(x)c(\beta)\exp[w(x)t(x)]$ which is the definition of an exponential family.

If $U = \log X/\alpha$, the inverse transformation is $X = \alpha\exp(U)$. Hence the density of U is $g_U(u|\beta) = \beta\alpha^\beta \exp(-(\beta+1)u)\alpha^{-\beta-1}\alpha\exp(u) = \beta\exp(-\beta u)$ for $u > 0$ since $x > \alpha$.

d) The likelihood equals

$$L(\alpha,\beta|x_1,\dots,x_n) = \prod_{i=1}^n \beta\alpha^\beta \frac{1}{x_i^{\beta+1}} I_{[\alpha,\infty)}(x_i) = \alpha^{n\beta}\exp(n\log\beta - (\beta+1)\Sigma_{i=1}^n \log x_i) I_{[\alpha,\infty)}(x_{(1)})$$

where $x_{(1)} = \min_{i=1,\dots,n} x_i$. Because $\beta\alpha^\beta I_{[\alpha,\infty)}(x_{(1)})$ is increasing for $\alpha \le x_{(1)}$ and equal to 0 for $\alpha > x_{(1)}$, $\hat\alpha = x_{(1)}$ and

$$\begin{aligned} L(\hat\alpha(\beta),\beta|x_1,\dots,x_n) &= \exp(n\log\beta - (\beta+1)\Sigma_{i=1}^n \log x_i - [(n(\beta+1) - n]\log x_{(1)}) \\ &= \exp(n\log\beta - (\beta+1)\Sigma_{i=1}^n \log x_i/x_{(1)} - n\log x_{(1)}) \end{aligned}$$

Hence

$$\log L(\hat\alpha(\beta),\beta|x_1,\dots,x_n) = \exp(n\log\beta - (\beta+1)\Sigma_{i=1}^n \log x_i/x_{(1)} - n\log x_{(1)})$$

so

$$\frac{\partial}{\partial\beta}\log L(\hat\alpha(\beta),\beta|x_1,\dots,x_n) = \frac{n}{\beta} - \Sigma_{i=1}^n \log x_i/x_{(1)}$$

and $\hat\beta = n/\Sigma_{i=1}^n \log x_i/x_{(1)}$. Also $\frac{\partial^2}{\partial\beta^2}\log L(\hat\alpha(\beta),\beta|x_1,\dots,x_n) = -n/\beta^2 < 0$ so the stationary point is a maximum since there is only one stationary point.

e) Because $U_i = \log X_i - \alpha,\ i = 1, \ldots, n$ are independent and exponentially distributed random variables, $(\log X_{(1)} - \alpha, \log X_{(2)}/X_{(1)}, \ldots, \log X_{(n)}/X_{(1)})$ has the same distribution as $(U_{(1)}, U_{(2)} - U_{(1)}, \ldots, U_{(n)} - U_{(1)})$ where $(X_{(1)}, \ldots, X_{(n)})$ and $(U_{(1)}, \ldots, U_{(n)})$ are the order statistics. Hence independence of $X_{(1)}$ and $\Sigma_{i=1}^n \log X_i/X_{(1)} = \Sigma_{i=2}^n \log X_{(i)}/X_{(1)}$ will follow fram independence of $U_{(1)}$ and $\Sigma_{i=2}^n (U_{(i)} - U_{(1)})$.

To show that consider the transformation $V_1 = U_{(1)}, V_i = (n - i + 1)(U_{(i)} - U_{(i-1)}),,\ i = 2, \ldots, n$ with inverse transformation

$$\begin{aligned} U_{(1)} &= V_1 \\ U_{(2)} &= V_2/(n-1) + V_1 \\ U_{(3)} &= V_3/(n-2) + U_{(2)} = V_3/(n-2) + V_2/(n-1) + V_1 \\ &\vdots \\ U_{(n)} &= \Sigma_{i=2}^n V_i/(n-i+1) + V_1 \end{aligned}$$

Because $\Sigma_{i=1}^n U_{(i)} = \Sigma_{i=2}^n V_i + nV_1$ and the Jacobian is equal to $1/(n-1)!$, the simultanous density of $(V_1, \ldots, V_n)$ is $n! \exp(-\beta(v_2 + \cdots + v_n) - n\beta v_1)/(n-1)!$. Thus the simultanous density factorizes, so $V_1 = U_{(1)}$ and $(V_2, \ldots, V_n) = (n-2)(U_{(2)} - U_{(1)}), \ldots, U_{(n)} - U_{(1)})$ are independent and therefore also $U_{(1)}$ and $\Sigma_{i=2}^n (U_{(i)} - U_{(1)})$.

The cumulative distribution function of a Pareto distributed random variable is $F_X(x) = \int_\alpha^x \frac{\beta\alpha^\beta}{y^{\beta+1}} dy = -\alpha^\beta y^{-\beta}|_\alpha^x = 1 - (\frac{\alpha}{x})^\beta$.

The cumulative distribution function of $\hat{\alpha}$ is $F_{\hat{\alpha}} = P(\hat{\alpha} \leq x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, \cdots, X_n > x) = 1 - [1 - F_X(x)]^n = 1 - (\frac{\alpha}{x})^{n\beta}$, so $\hat{\alpha}$ is Pareto distributed with parameters α and $n\beta$.

From the results above $\Sigma_{i=2}^n \log X_{(i)}/X_{(1)}$ has the same distribution as $\Sigma_{i=2}^n (U_{(i)} - U_{(1)}) = \Sigma_{i=2}^n V_i$. The random variables $V_2, \cdots, V_n$ are independent gamma$(1, 1/\beta)$ distributed variables, so $2\beta V_1, \cdots, 2\beta V_n$ are independently distributed χ_2^2 random variables. Then $2\beta\Sigma_{i=2}^n V_i$ is $\chi^2_{2(n-1)}$ distributed and so is $2n\beta/\hat{\beta} = 2\beta\Sigma_{i=2}^n \log X_{(i)}/X_{(1)}$.

f) The joint probability density is

$$\begin{aligned} & f(x_1, \ldots, x_n|\alpha, \beta) \\ = & \prod_{i=1}^n \beta\alpha^\beta \frac{1}{x_i^{\beta+1}} I_{[\alpha,\infty)}(x_i) \\ = & \exp(n\beta \log\alpha + n\log\beta - (\beta+1)\Sigma_{i=1}^n \log x_i) I_{[\alpha,\infty)}(x_{(1)}) \\ = & \exp(-(\beta+1)\Sigma_{i=1}^n \log x_i/x_{(1)} - n(\beta+1)\log x_{(1)}/\alpha + n\log\beta\alpha + n\log x_{(1)}) I_{[\alpha,\infty)}(x_{(1)}). \end{aligned}$$

The density can therefore be expressed as a function of the parameters α and β and the functions $\Sigma_{i=1}^n \log x_i/x_{(1)}$ and $x_{(1)}$. From the factorization theorem, it then follows that the statistic $(X_{(1)}, \Sigma_{i=1}^n \log X_i/X_{(1)})$ is sufficient for the parameters α and β.

g) From part d) $2n\beta/\hat{\beta}$ is $\chi^2_{2(n-1)}$ distributed so $E[\frac{1}{2n\beta/\hat{\beta}}] = 1/2(n-2)$ since $E[X] = 1/(k-2)$ when $X \sim \chi^2_k$, $k > 2$, see Casella and Berger p.130. Thus $E[\hat{\beta}] = 2n\beta/2(n-2) = n\beta/(n-2)$. Therefore $(n-2)\hat{\beta}/n$ is an unbiased estimator of β. Since it is a function of the complete and sufficient statistic $(X_{(1)}, \Sigma_{i=1}^n \log X_i/X_{(1)})$ it is UMVUE or the best unbiased estimator.

h) For $\hat{\alpha}$, write

$$X_{(1)}[1 - \frac{1}{(n-1)}\frac{1}{\hat{\beta}}] = X_{(1)}[1 - \frac{1}{(n-1)}\frac{2n\beta}{\hat{\beta}}\frac{1}{2n\beta}].$$

Using that $2n\beta/\hat{\beta}$ is $\chi^2_{2(n-1)}$ distributed and that $X_{(1)}$ and $\hat{\beta}$ are independent it follows that

$$E\{X_{(1)}[1 - \frac{1}{(n-1)}\frac{1}{\hat{\beta}}]\} = \frac{n\beta\alpha}{n\beta-1}[1 - \frac{1}{(n-1)}\frac{2(n-1)}{2n\beta}] = \frac{n\beta\alpha}{n\beta-1}\frac{n\beta-1}{n\beta} = \alpha.$$

Hence, $X_{(1)}[1 - \frac{1}{(n-1)}\frac{1}{\hat{\beta}}]$ is an unbiased estimator of α. It is also a function of the complete and sufficient statistic $(X_{(1)}, \Sigma_{i=1}^n \log X_i/X_{(1)})$ and is therefore UMVUE or the best unbiased estimator for α.

Problem 3

a) If $X \sim f_X(x|p)$, $\sum_{x=1}^{\infty} p(1-p)^{x-1} = 1$ means that $\sum_{x=1}^{\infty}(1-p)^{x-1} = 1/p$. Multiplying with $1-p$ yields $\sum_{x=1}^{\infty}(1-p)^x = (1-p)/p$, so after differentiating with respect to p and multiplying with p $E(X) = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = 1/p$. Similarly, multiplying the last identity with $(1-p)$ and differentiating with respect to p, yields $E(X^2) = \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} = (2-p)/p$, so $Var(X) = (1-p)/p^2$.
Thus $E(\bar{X}) = 1/p$ and $Var(\bar{X}) = (1-p)/np^2$.

b)

$$\log\prod_{i=1}^n f_X(x_i|p) = \log\prod_{i=1}^n p(1-p)^{x_i-1} = n\log p + (\sum x_i - n)\log(1-p)$$

so

$$\frac{\partial}{\partial p}\log\prod_{i=1}^n f_X(x_i|p) = n/p - (\sum x_i - n)/(1-p) = \frac{n - np - np\bar{x} + np}{p(1-p)} = \frac{n - np\bar{x}}{p(1-p)}.$$

Hence

$$E[\frac{\partial}{\partial p}\log\prod_{i=1}^n f_X(X_i|p)]^2 = n^2 E[\frac{(1-p\bar{X})^2}{p^2(1-p)^2}] = n^2 E[\frac{(1 - 2p\bar{X} + p^2\bar{X}^2)}{p^2(1-p)^2}]$$

But $E(\bar{X}^2) = Var(\bar{X}) + (E(\bar{X})^2 = \frac{1-p}{np^2} + \frac{1}{p^2} = \frac{1+(1-p)/n}{p^2}$ so

$$E[\frac{\partial}{\partial p} \log \prod_{i=1}^{n} f_X(X_i|p)]^2 = \frac{n^2(1-2+1+(1-p)/n)}{p^2(1-p)^2} = \frac{n}{p^2(1-p)}.$$

By the Cramer Rao inequality

$$Var(\bar{X}) \geq \frac{(\frac{\partial}{\partial p} E(\bar{X}))^2}{E[\frac{\partial}{\partial p} \log \prod_{i=1}^{n} f_X(X_i|p)]^2} = \frac{\frac{1}{p^4}}{\frac{n}{p^2(1-p)}} = \frac{1}{np^2(1-p)}$$

The lower bound is equal to $Var(\bar{X})$ so $\bar{X}$ is a best unbiased estimator for $1/p$.

c) The likelihood function is

$$L(p|x_1, \ldots, x_n) = \prod_{i=1}^{n} p(1-p)^{x_i - 1} = p^n (1-p)^{\sum x_i - n}$$

so

$$\log L(p|x_1, \ldots, x_n) = n \log p + (\sum x_i - n) \log(1-p).$$

Differentiating with respect to p yields first order conditions

$$\frac{n}{p} - \frac{(\sum x_i - n)}{1-p} = \frac{n - p\bar{x}}{p(1-p)} = 0$$

with solution $\hat{p} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$ which is the maximum likelihood estimator for p. Also remark that the derivative changes from positive to negative, which means that the stationary point is a maximum

d) $E[I_{[x=1]}(X_1)] = P(X_1 = 1) = p$ so the estimator is unbiased. Using the factorization theorem we see that $\sum X_i$ is a sufficient statistic. From the description of the negative binominal distribution in Casella and Berger, p.95, l. 10-14 one can see that the sum of n geometric variables are negative binominally distributed with parameters n and p.

$$P(\sum X_i = x) = \binom{x-1}{n-1} p^n (1-p)^{x-n}, \ x = n, n+1, \ldots$$

Then

$$\begin{aligned} \tilde{p} &= P(X_1 = 1| \sum_{i=1}^{n} X_i = x) = \frac{P(X_1 = 1 \text{ and } \sum_{i=2}^{n} X_i = x - 1)}{P(\sum X_i = x)} \\ &= \frac{p\binom{x-2}{n-2} p^{n-2} (1-p)^{x-2-n+2}}{\binom{x-1}{n-1} p^{n-1} (1-p)^{x-1-n+1}} = \frac{\frac{(x-2)!}{(n-2)!(x-n)!}}{\frac{(x-1)!}{(n-1)!(x-n)!}} = \frac{n-1}{x-1}. \end{aligned}$$

e) Notice that

$$\tilde{p} = \frac{n-1}{n\bar{X} - 1} = \frac{n-1}{n\frac{1}{\hat{p}} - 1} = \hat{p}\frac{n-1}{n - \hat{p}}.$$

Thus $\hat{p} < \hat{p}\frac{n-1}{n-\hat{p}} = \tilde{p}$. Since $\tilde{p}$ is unbiased, $E(\hat{p}) < p$, so $\hat{p}$ is negatively biased.