STAT 630 Fall 2014
Homework 8 Solution

6.1.7

The likelihood function of this model is

$$\begin{aligned} L(\theta|x_1,\cdots,x_n) &= \prod_{i=1}^{n} \frac{\theta^{x_i}e^{-\theta}}{x_i!} \\ &= \left\{\prod_{i=1}^{n} x_i!\right\}^{-1} \theta^{n\bar{x}}e^{-n\theta}. \end{aligned}$$

From the factorization theorem, we know $\bar{x}$ is a sufficient statistic.

6.1.19

The likelihood function is

$$\begin{aligned} L(\theta|x_1,\cdots,x_n) &= \prod_{i=1}^{n} \frac{\theta^{\alpha_0}}{\Gamma(\alpha_0)} x_i^{\alpha_0}\exp\{-\theta x_i\} \\ &= \left\{\frac{\theta^{\alpha_0}}{\Gamma(\alpha_0)}\right\}^{n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha_0} \exp\{-\theta n\bar{x}\}. \end{aligned}$$

Hence, $\bar{x}$ is a sufficient statistic.

6.2.4

(a) Since $f(x_i;\theta) = \frac{e^{-\theta}\theta^{x_i}}{x_i!}$, then we can write down the log-likelihood function: $l(\theta|x_1,\cdots,x_n) = \sum_{i=1}^{n}(-\theta + x_i * \log(\theta) - \log(x_i!) = -n\theta + \log(\theta)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\log(x_i!)$. Then we take the derivative with respect to θ and let it be zero: $\frac{\partial l}{\partial \theta} = -n + \frac{1}{\theta}\sum_{i=1}^{n} x_i = 0$. So the MLE estimate is $\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$. (Note that since $\bar{x} \geq 0$, we have $\frac{\partial S(\theta|x_1,\cdots,x_n)}{\partial\theta}|_{\theta=\bar{x}} = -\frac{n}{\bar{x}} < 0$)

(b) Since $x_1, x_2, \cdots, x_n$ are i.i.d samples and $E[x_i] = \theta$, thus $E[\hat{\theta}] = E[x_i] = \theta$ which means the MLE is unbiased. For the variance, $\text{Var}(\hat{\theta}) = \text{Var}(\bar{x}) = \frac{\text{Var}(x_i)}{n} = \frac{\theta}{n}$. Therefore, $\text{MSE}(\hat{\theta}) = \text{Bias}^2 + \text{Variance} = \frac{\theta}{n}$.

6.2.5

(a) Since $f(x_i;\theta) = \frac{\theta^{\alpha_0} x_i^{\alpha_0 - 1}}{\Gamma(\alpha_0)} e^{-\theta x_i}$, the log-likelihood function is $l(\theta|X) = \sum_{i=1}^n [\alpha_0 \log\theta + (\alpha_0 - 1)\log x_i - \theta x_i - \Gamma(\alpha_0)] = n\alpha_0 \log\theta - \theta\sum_{i=1}^n x_i - (\alpha_0 - 1)\sum_{i=1}^n \log x_i - n\Gamma(\alpha_0)$. By calculating $\frac{\partial l}{\partial\theta}(\theta|X) = \frac{n\alpha_0}{\theta} - \sum_{i=1}^n x_i = 0$, we get the MLE is $\hat{\theta} = \frac{n\alpha_0}{\sum_{i=1}^n x_i} = \frac{\alpha_0}{\bar{x}}$ (It's easy to check that $\frac{\partial S}{\partial\theta}|_{\theta=\hat{\theta}} < 0$).

(b) It's easy to verify that $\bar{x} \sim \Gamma(\alpha,\lambda)$ using moment generating function, where $\alpha = n\alpha_0, \lambda = n\theta$. Hence, $E(\frac{1}{\bar{x}}) = \int_0^\infty \frac{1}{x}\frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\lambda\Gamma(\alpha-1)}{\Gamma(\alpha)}\int_0^\infty \frac{\lambda^{\alpha-1}x^{\alpha-2}}{\Gamma(\alpha-1)} e^{-\lambda x} dx = \frac{\lambda\Gamma(\alpha-1)}{\Gamma(\alpha)} = \frac{\lambda}{\alpha-1} = \frac{n\theta}{n\alpha_0-1}$. Similarly, we get $Var(\frac{1}{\bar{x}}) = E(\frac{1}{\bar{x}^2}) - E(\frac{1}{\bar{x}})^2 = \frac{\lambda^2}{(\alpha-1)(\alpha-2)} - (\frac{\lambda}{\alpha-1})^2 = \frac{\lambda^2}{(\alpha-1)^2(\alpha-2)} = \frac{n^2\theta^2}{(n\alpha_0-1)^2(n\alpha_0-2)}$. Hence, Bias($\hat{\theta}$)$= E(\hat{\theta}) - \theta = \alpha_0 E(\frac{1}{\bar{x}}) - \theta = \frac{\theta}{n\alpha_0-1}$, Var($\hat{\theta}$)$= \alpha_0^2 Var(\frac{1}{\bar{x}}) = \frac{(n\alpha_0\theta)^2}{(n\alpha_0-1)^2(n\alpha_0-2)}$ and MSE($\hat{\theta}$)=Bias^2 + Variance $= \frac{(n\alpha_0+2)\theta^2}{(n\alpha_0-1)(n\alpha_0-2)}$.

(c) The first moment of Gamma(α_0,θ) is $\frac{\alpha_0}{\theta}$. Let $\frac{\alpha_0}{\theta} = \bar{x}$, we get the method-of-moments estimator $\hat{\theta} = \frac{\alpha_0}{\bar{x}}$, which is the same as the MLE.

6.2.6

(a) Since $x_i \sim$ Geometric(θ), then the likelihood function is $L(\theta|x_1,\ldots,x_n) = \theta^n(1-\theta)^{n\bar{x}}$, the log-likelihood function is $l(\theta|x_1,\ldots,x_n) = nln(\theta) + n\bar{x}ln(1-\theta)$. The score function is $S(\theta|x_1,\ldots,x_n) = \frac{n}{\theta} - \frac{n\bar{x}}{1-\theta}$. Sovling the score equation gives $\hat{\theta} = \frac{1}{1+\bar{x}}$. Since $0 \le \bar{x} \le 1$, we have $\frac{\partial S(\theta|x_1,\cdots,x_n)}{\partial\theta}|_{\theta=\frac{1}{1+\bar{x}}} = -n\left((1+\bar{x})^2 + \frac{(1+\bar{x})^2}{\bar{x}}\right) < 0$. Thus, $\hat{\theta} = \frac{1}{1+\bar{x}}$ is the MLE.

(b) since $E(X) = \frac{1-\theta}{\theta}$, then $\theta = \frac{1}{1+E(X)}$. By replacing $E(X)$ with $\bar{x}$, we get the method-of-moments estimator $\hat{\theta} = \frac{1}{1+\bar{x}}$.

6.2.8

The density function for weibull distribution is $f(x) = \beta x^{\beta-1}e^{-x^\beta}$. Then the log-likelihood function is $l(\beta|x_1,\cdots,x_n) = \sum_{i=1}^n(\log\beta + (\beta-1)\log x_i - x_i^\beta)$. So take derivative of the log-likelihood function with respect to β and let it be zero, we can obtain the score function is:

$$\sum_{i=1}^n (1/\beta + \log x_i - x_i^\beta \log x_i) = 0$$

6.2.12

(a) For this normal model, the log-likelihood function is $l(\mu_0,\sigma^2|x_1,\cdots,x_n) = \sum_{i=1}^n(-\log(2\pi)/2 - \log(\sigma) - \frac{(x_i-\mu_0)^2}{2\sigma^2})$. Then we take derivative of it with respect to σ^2, we can obtain: $\frac{\partial l}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu_0)^2}{\sigma^4} = 0$. Thus the MLE of σ^2 is $\hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i-\mu_0)^2}{n}$. (It's easy to check the second derivative of the loglikelihood function at $\hat{\sigma}^2$ is negative.) For the location-scale normal model, the μ_0 is unknown. The MLE of σ^2 for this model is $\tilde{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}$. So

$$\begin{aligned}
\hat{\sigma}^2 - \tilde{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu_0)^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - \mu_0)^2 + \frac{2}{n}\sum_{i=1}^{n}(x_i - \bar{x})(\bar{x} - \mu_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - \mu_0)^2 + \frac{2}{n}(n\bar{x} - n\bar{x})(\bar{x} - \mu_0) \\
&= \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - \mu_0)^2
\end{aligned}$$

Since $n \to \infty$, $\bar{x} \to \mu_0$, thus the difference will tend to be zero when n goes to infinity.

(b) First $E[\hat{\sigma}^2] = E[\sigma^2 \cdot \frac{\hat{\sigma}^2}{\sigma^2}]$. Since $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_n$, thus $E[\frac{\hat{\sigma}^2}{\sigma^2}] = 1$. So $E[\hat{\sigma}^2] = \sigma^2$ which means it is unbiased. Then $\text{Var}(\hat{\sigma}^2) = \text{Var}(\frac{\sigma^2}{n} \cdot \frac{n\hat{\sigma}^2}{\sigma^2}) = \frac{\sigma^4}{n^2} * \text{Var}(\chi_n) = \frac{2\sigma^4}{n}$. The mean square error is $0 + \text{Variance} = \frac{2\sigma^4}{n}$.

The mean square error of $\hat{\sigma}^2 = \frac{2\sigma^4}{n}$, which is smaller than mean square error of S^2 but greater than that of $\hat{\sigma}^2$ in example 45.

6.2.19

(a) It is multinomial $(\theta^2, 2\theta(1-\theta), (1-\theta)^2)$ distributed .

(b) The likelihood function is $L(\theta|y_1, \cdots, y_n) = (\theta^2)^{x_1}(2\theta(1-\theta))^{x_2}(1-\theta)^{x_3} = 2^{x_2}\theta^{2x_1+x_2}(1-\theta)^{x_2+2x_3}$; The log-likelihood function is $l(\theta|y_1, \cdots, y_n) = x_2 \log 2 + (2x_1 + x_2)\log(\theta) + (x_2 + 2x_3)\log(1-\theta)$; Taking derivative of log-likelihood function with respect to θ we can obtain the score function: $\frac{2x_1+x_2}{\theta} - \frac{x_2+2x_3}{1-\theta}$.

(c) By solving the score function, we can obtain the MLE of θ which is $\hat{\theta} = \frac{2x_1+x_2}{2x_1+2x_2+2x_3}$, since $\frac{\partial S(\theta|s_1,\cdots,s_n)}{\partial\theta} = -\frac{2x_1+x_2}{\theta^2} - \frac{x_2+2x_3}{(1-\theta)^2} < 0$ for every $\theta \in [0,1]$

6.3.15

(a) Since $E[x_1] = 1 * \theta + 0 * (1-\theta) = \theta$, thus x_1 is an unbiased estimator of θ.

(b) Since $x_1 = 1 \; or \; 0$, then $x_1^2 = x_1$. Thus $E[x_1^2] = \theta$. Hence, x_1^2 is not an unbiased estimator for θ^2. An unbiased estimator is not transformation invariant.

6.3.24

(a) $E[\alpha T_1 + (1-\alpha)T_2] = \alpha E[T_1] + (1-\alpha)E[T_2] = \alpha\psi(\theta) + (1-\alpha)\psi(\theta) = \psi(\theta)$. So $\alpha T_1 + (1-\alpha)T_2$ is an unbiased estimator for $\psi(\theta)$.

(b) $\text{Var}_\theta(\alpha T_1+(1-\alpha)T_2) = \alpha^2\text{Var}_\theta(T_1)+(1-\alpha)^2\text{Var}_\theta(T_2)+2\alpha(1-\alpha)\text{cov}(T_1,T_2)$. Since T_1, T_2 are independent thus $\text{cov}(T_1,T_2) = 0$. Therefore $\text{Var}_\theta(\alpha T_1 + (1-\alpha)T_2) = \alpha^2\text{Var}_\theta(T_1) + (1-\alpha)^2\text{Var}_\theta(T_2)$.

(c) To obtain the optimal value of α, we take derivative of $\text{Var}_\theta(\alpha T_1 + (1-\alpha)T_2)$ with respect to α and let it be zero, then we can obtain: $2\alpha\text{Var}_\theta(T_1) = 2(1-\alpha)\text{Var}_\theta(T_2)$. Thus $\hat{\alpha}_{opt} = \frac{\text{Var}_\theta(T_2)}{\text{Var}_\theta(T_1)+\text{Var}_\theta(T_2)}$. Then if $\text{Var}_\theta(T_1)$ is much larger than $\text{Var}_\theta(T_2)$, the $\hat{\alpha}_{opt}$ will be very small and the estimator tends to be T_2.

(d) If T_1, T_2 are not independent, then $\text{Var}_\theta(\alpha T_1+(1-\alpha)T_2) = \alpha^2\text{Var}_\theta(T_1)+(1-\alpha)^2\text{Var}_\theta(T_2)+2\alpha(1-\alpha)\text{Cov}_\theta(T_1,T_2)$. Then we take derivative of it with respect to α and let it be zero, we can solve the optimal value of α which is

$$\hat{\alpha}_{opt} = \frac{\text{Var}_\theta(T_2) - \text{Cov}_\theta(T_1,T_2)}{\text{Var}_\theta(T_1) - 2\text{Cov}_\theta(T_1,T_2) + \text{Var}_\theta(T_2)} = \frac{\text{Var}_\theta(T_2) - \text{Cov}_\theta(T_1,T_2)}{\text{Var}_\theta(T_1 - T_2)}$$

when $T_1 \neq T_2$. If $T_1 = T_2$, then the variance of $\alpha T_1 + (1-\alpha)T_2$ is free of α, so we assume $P(T_1 = T_2) < 1$. In this case, if $\text{Var}_\theta(T_1)$ is much larger than $\text{Var}_\theta(T_2)$, the $\hat{\alpha}_{opt}$ is still very small and the estimator will highly depend on T_2.

Additional problem

$T_C = \frac{C}{\sum_{i=1}^{n} x_i} = \frac{C}{n} \cdot \frac{1}{\bar{x}} = \frac{C}{n}\hat{\lambda}$, where $\hat{\lambda}$ is the maximum likelihood estimator of λ. To obtain the mean square error of $\frac{C}{\sum_{i=1}^{n} x_i}$, we will use the results in the lecture notes of chapter6: $E[\hat{\lambda}] = \frac{n}{n-1}\lambda$ and $\text{Var}(\hat{\lambda}) = \frac{n^2\lambda^2}{(n-2)(n-1)^2}$ (detailed proof available in lecture chapter six). So the bias of $\frac{C}{n}\hat{\lambda}$ is $(\frac{C}{n} * \frac{n}{n-1}\lambda - \lambda) = \frac{C-n+1}{n-1}\lambda$ and the variance of it is $\frac{C^2}{n^2} * \frac{n^2\lambda^2}{(n-2)(n-1)^2} = \frac{C^2\lambda^2}{(n-2)(n-1)^2}$. Therefore the mean square error is $\frac{(C-n+1)^2}{(n-1)^2}\lambda^2 + \frac{C^2\lambda^2}{(n-2)(n-1)^2}$ which is a function of C. Then we take the derivative of the mean square error with respect to C and let it be zero, we can easily obtain the optimal value $\hat{C} = n-2$ which minimizes the mean square error.