Spring 2012

1. a) Let Z_i be independent $N(0,1)$, $i = 1, 2, \ldots, n$. Are $\bar{Z} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^n (Z_i - \bar{Z})^2$ independent? Prove your claim.

Solution. Define $\boldsymbol{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}^T$. Then, $\bar{Z} = \frac{1}{n}\mathbf{1}^T\boldsymbol{Z}$, and

$$\boldsymbol{Z} - \mathbf{1}\bar{Z} = \begin{bmatrix} Z_1 - \bar{Z} \\ Z_2 - \bar{Z} \\ \vdots \\ Z_n - \bar{Z} \end{bmatrix} = \boldsymbol{Z} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\boldsymbol{Z} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\boldsymbol{Z}.$$

It is easily verified that $P := I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is an orthogonal projection; i.e. $P^2 = P$ and $P^T = P$. Now, we claim that $\bar{Z}$ and $P\boldsymbol{Z}$ are independent. To see this, we consider the covariance:

$$\begin{aligned} \operatorname{Cov}\left(\frac{1}{n}\mathbf{1}^T\boldsymbol{Z}, P\boldsymbol{Z}\right) &= \frac{1}{n}\mathbf{1}^T\operatorname{Var}(\boldsymbol{Z})P^T = \frac{1}{n}\mathbf{1}^T IP = \frac{1}{n}\mathbf{1}^T\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) = \frac{1}{n}\left(\mathbf{1}^T - \frac{1}{n}\mathbf{1}^T\mathbf{1}\mathbf{1}^T\right) \\ &= \frac{1}{n}\left(\mathbf{1}^T - \frac{1}{n}n\mathbf{1}^T\right) = 0, \end{aligned}$$

and so, $\bar{Z} = \frac{1}{n}\mathbf{1}^T\boldsymbol{Z}$ and $\boldsymbol{Z} - \mathbf{1}\bar{Z} = P\boldsymbol{Z}$ are independent. Now, for any measurable f, we know that $\bar{Z}$ and $f(\boldsymbol{Z} - \mathbf{1}\bar{Z})$ will be independent. In particular, take $f(\boldsymbol{x}) = \frac{1}{n-1}\boldsymbol{x}^T\boldsymbol{x}$. Then,

$$f\left(\boldsymbol{Z} - \mathbf{1}\bar{Z}\right) = \frac{1}{n-1}\begin{bmatrix} Z_1 - \bar{Z} & \cdots & Z_n - \bar{Z} \end{bmatrix}\begin{bmatrix} Z_1 - \bar{Z} \\ \vdots \\ Z_n - \bar{Z} \end{bmatrix} = S_Z^2,$$

and so, $\bar{Z}$ and S_Z^2 are independent. □

b) Let $X_1, X_2, \ldots, X_n$ be independent identically distributed normal with mean θ and variance θ^2, where $\theta > 0$ is unknown. Let

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Are $\bar{X}$ and S^2 independent? Prove your claim. [HINT: you can directly use the result from part (a).]

Solution. Note that $X_i = \theta + \theta Z_i$, and so, $\bar{X} = \theta + \theta\bar{Z}$. Also,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^n \left(\theta + \theta Z_i - \theta - \theta\bar{Z}\right)^2 = \theta^2 S_Z^2.$$

Hence,

$$\begin{aligned} P(\bar{X} \le a, S^2 \le b) &= P\left(\theta + \theta\bar{Z} \le a, \theta^2 S_Z^2 \le b\right) = P\left(\bar{Z} \le \frac{a-\theta}{\theta}, S_Z^2 \le \frac{b}{\theta^2}\right) \\ &\overset{(a)}{=} P\left(\bar{Z} \le \frac{a-\theta}{\theta}\right)P\left(S_Z^2 \le \frac{b}{\theta^2}\right) = P(\bar{X} \le a)P(S^2 \le b), \end{aligned}$$

and so, $\bar{X}$ and S^2 are independent. □

c) Show that $(\bar{X}, S^2)$ is a sufficient statistic for θ, but it is not complete.

Solution. The joint density is

$$\begin{aligned} f(\boldsymbol{x};\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta^2}} \exp\left(-\frac{(x_i-\theta)^2}{2\theta^2}\right) = (2\pi)^{-\frac{n}{2}}\theta^{-n}\exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2\right) \\ &= (2\pi)^{-\frac{n}{2}}\theta^{-n}\exp\left(-\frac{1}{2\theta^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(\bar{x}-\theta)^2\right)\right) \\ &= (2\pi)^{-\frac{n}{2}}\theta^{-n}\exp\left(-\frac{n-1}{2\theta^2}\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\exp\left(-\frac{n}{2\theta^2}\left(\bar{x}^2-2\theta\bar{x}+\theta^2\right)\right), \end{aligned}$$

and so, by Neymar-Fisher factorization theorem, $(\bar{X}, S^2)$ is a sufficient statistic for θ. Now, consider the function $g(x,y) = (x-\theta)^2 - \frac{1}{n}y$. Then,

$$\mathbb{E}g\left(\bar{X}, S^2\right) = \mathbb{E}\left[\left(\bar{X}-\theta\right)^2 - \frac{1}{n}S^2\right] = \mathrm{Var}(\bar{X}) - \frac{1}{n}\mathbb{E}S^2 = \frac{\theta^2}{n} - \frac{\theta^2}{n} = 0.$$

However, $P_\theta(g=0) \neq 1$, and so, $(\bar{X}, S^2)$ is not complete. □

2. a) Let $X_1, X_2, \ldots, X_n$ be exponentially distributed with density

$$f(x) = \lambda \exp(-\lambda x), \qquad x > 0.$$

Let $c > 0$ be a constant and if $X_i < c$, we observe X_i, otherwise we observe c.

$$S_n = \sum_{i=1}^{n} X_i I(X_i < c), \qquad T_n = \sum_{i=1}^{n} I(X_i > c),$$

where $I(A) = 1$ if event A occurs and $I(A) = 0$ otherwise. Write down the likelihood function of the observed values in terms of T_n and S_n.

Solution. If $T_n = t_n$ c's are observed, without loss of generality, we can assume it is the first t_n X_i's that are greater than c. The likelihood function is

$$\begin{aligned}
\mathcal{L}(\lambda; \boldsymbol{x}) &= \prod_{i=1}^{t_n} P(X_i > c) \prod_{i=t_n+1}^{n} f_{X_i}(x_i) = (\exp(-c\lambda))^{t_n} \prod_{i=t_n+1}^{n} \lambda \exp(-\lambda x_i) \\
&= \lambda^{n-t_n} \exp\left(-c\lambda t_n - \lambda \sum_{i=t_n+1}^{n} x_i\right) \\
&= \boxed{\lambda^{n-t_n} \exp(-c\lambda t_n - \lambda s_n)}.
\end{aligned}$$

b) Show the maximum likelihood estimator of λ is

$$\hat{\lambda}_n = \frac{n - T_n}{S_n + cT_n}.$$

Solution. The log-likelihood function is

$$\log \mathcal{L}(\lambda; \boldsymbol{x}) = (n - t_n) \log \lambda - c\lambda t_n - \lambda s_n,$$

and so, we have

$$\frac{d}{d\lambda} \log \mathcal{L}(\lambda; \boldsymbol{x}) = \frac{n - t_n}{\lambda} - ct_n - s_n,$$

and

$$\frac{d^2}{d\lambda^2} \log \mathcal{L}(\lambda; \boldsymbol{x}) = -\frac{n - t_n}{\lambda} < 0.$$

Thus,

$$\hat{\lambda} = \frac{n - T_n}{cT_n + S_n}$$

is the MLE of λ. □