Homework 2

Solutions: 09.04.2025

1. Let $X_1, \dots, X_n$ be independent and identically distributed random variables whose density is given by

$$p_\theta(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1, \\ 0, & \text{else}, \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) Show that this is an exponential family in canonical form.
We have for $0 < x < 1$

$$p_\theta(x) = \exp[\log\theta + (\theta-1)\log x] = \exp\left[\log\theta + \log x^{\theta-1}\right] = \exp[c(\theta)T(x) - d(\theta)]\, h(x),$$

where $c(\theta) = \theta$, $T(x) = \log x$, $d(\theta) = -\log\theta$, and $h(x) = \frac{1}{x}$. The exponential family is in canonical form since $c(\theta) = \theta$.

(b) Let $T = T(X_1, \dots, X_n) := -\frac{1}{n}\sum_{i=1}^n \log X_i$. Show that T is a sufficient and complete statistic.
For the sample $X_1, \dots, X_n$, we have

$$\prod_{i=1}^n p_\theta(x_i) = \exp[nc(\theta)] \sum_{i=1}^n T(x_i) - nd(\theta) \prod_{i=1}^n h(x_i),$$

again with $c(\theta) = \theta$, $T(x) = \log x$, $d(\theta) = -\log\theta$, and $h(x) = \frac{1}{x}$. Hence, $T = T(X_1, \dots, X_n) := \frac{1}{n}\sum_{i=1}^n \log X_i$ is a sufficient statistic. The set $C := \{\theta : \theta > 0\}$ surely contains an open interval in $\mathbb{R}$, so that T is also complete.

(c) Prove that

$$\mathbb{E}_\theta[T] = \frac{1}{\theta}, \quad \mathrm{var}_\theta(T) = \frac{1}{n\theta^2}.$$

We have $d'(\theta) = -\frac{1}{\theta}$, and $d''(\theta) = \frac{1}{\theta^2}$. Thus, it follows from Lemma 4.5.1 that

$$E_\theta[T(X_1)] = d'(\theta) = -\frac{1}{\theta},$$

Hence,

$$E_\theta[T] = \frac{1}{n}\sum_{i=1}^n E_\theta[T(X_i)] = -\frac{1}{\theta},$$

and

$$\mathrm{var}_\theta(T) = \mathrm{var}_\theta\left(\frac{1}{n}\sum_{i=1}^n \log X_i\right) = \frac{1}{n^2}\mathrm{var}_\theta(T(X_i)) = \frac{1}{n\theta^2}.$$

2. If T is a survival time in $(0, \infty)$ with density f, then one object worth studying is the so-called *hazard function*:

$$\lambda(t) = \lim_{h\to 0} \frac{\mathbb{P}(t \le T \le t+h \mid T > t)}{h} = \frac{f(t)}{1-F(t)} = -\frac{d}{dt}\log(1-F(t)).$$

Consider the one-parameter family of distributions for survival times which is defined via the hazard function

$$\lambda_\theta(t) = \theta\lambda_0(t), \quad \theta \in \mathbb{R}^+,$$

where

$$\lambda_\theta(t) = -\frac{d}{dt}\log(1 - F_\theta(t))$$

as above and with $\lambda_0(t)$ the (known) hazard function under the standard treatment.

(a) Show that this is an exponential family.
By integration, we get

$$-\log(1 - F_\theta(t)) = \int_0^t \lambda_0(s)ds$$

and hence

$$1 - F_\theta(t) = \exp\left\{-\int_0^t \lambda_0(s)ds\right\}.$$

Plugging in $\lambda_\theta(t) = \theta\lambda_0(t)$ gives

$$F_\theta(t) = 1 - \exp\left\{-\int_0^t \theta\lambda_0(s)ds\right\}.$$

The density can now be obtained by taking the derivative with respect to t:

$$f_\theta(t) = \theta\exp\{-\theta A_0(t)\}\lambda_0(t) = \exp\{-\theta A_0(t) + \log(\theta)\}\lambda_0(t),$$

where $A_0(t) = \int_0^t \lambda_0(s)ds$. Hence, the densities form an exponential family with $c(\theta) = -\theta$, $T(t) = A_0(t)$, $d(\theta) = -\log(\theta)$ and $h(t) = \lambda_0(t)$.

(b) Compute a sufficient statistic for the parameter θ.
From exercise a), it immediately follows that $A_0(t)$ is sufficient for the parameter θ.

3. (Optional) Let $X_1, \dots, X_n$ be i.i.d. Poisson(θ)-distributed. We want to estimate $g(\theta) := \theta\exp(-\theta)$.

(a) Show that $T(X) := \mathbb{1}(\{X_1 = 1\})$ is an unbiased estimator of $g(\theta)$, where $\mathbb{1}(A)$ denotes the indicator RV on the event A.
We have that

$$E_\theta[T(X)] = E_\theta[I_{\{X_1=x_1\}}] = P_\theta(X_1 = 1) = \theta\exp(-\theta) = g(\theta).$$

Hence, $T(X)$ is unbiased.

(b) Show that $S := \sum_{i=1}^n X_i$ is a sufficient and complete statistic.
Let $X = (X_1, \dots, X_n)$, $x = (x_1, \dots, x_n)$ and $s = \sum_{i=1}^n x_i$. By Lemma 3.3.1, the distribution of S is Poisson($n\theta$). Thus, we have that

$$\begin{aligned} P(X = x|S = s) &= \frac{P(X = x, S = s)}{P(S = s)} \\ &= \frac{\exp(-n\theta)\frac{(n\theta)^s}{s!}}{\exp(-n\theta)\frac{(n\theta)^s}{s!}} \\ &= \frac{s!}{n^s\prod_{i=1}^n x_i!}, \end{aligned}$$

which does not depend on θ. For completeness, we note that if

$$E_\theta[h(S)] = \sum_{k=0}^\infty \exp(-n\theta)\frac{(n\theta)^k}{k!}h(k) = 0\ \forall\theta,$$

then $h(k) = 0\ \forall k$, P_θ-a.s.

(c) Construct an UMVU estimator of $g(\theta)$.
From exercise a), we know that T is unbiased, and it obviously has finite variance. From exercise b), S is sufficient and complete. Hence, by the Lehmann-Scheffé Lemma, the estimator $T^* = E[T|S]$ is UMVUE. We compute

$$\begin{aligned} E[T|S=s] &= E[I_{\{x_1=1\}}|S=s] \\ &= P(X_1=1|S=s) \\ &= \frac{P(X_1=1, S=s)}{P(S=s)} \\ &= \frac{P(X_1=1)P(\sum_{i=2}^n X_i = s-1)}{P(S=s)} \\ &= \frac{\exp(-\theta)\theta\exp(-\theta(n-1))\frac{((n-1)\theta)^{s-1}}{(s-1)!}}{\exp(-n\theta)\frac{(n\theta)^s}{s!}} \\ &= \frac{s}{n}\left(\frac{n-1}{n}\right)^{s-1}. \end{aligned}$$

Hence, we have

$$T^* = E[T|S] = \frac{s}{n}\left(\frac{n-1}{n}\right)^{s-1}.$$

4. Consider the Beta distribution, a family of continuous probability distributions defined on the interval [0, 1] parameterized by two positive shape parameters, α and β, which appear as exponents of the random variable and control the shape of the distribution. The probability density function is given by:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},$$

where $B(\alpha,\beta)$ is the Beta function which serves as a normalization constant to ensure that the total probability is 1.

(a) Show that the Beta distribution is an exponential family.

$$f(x;\alpha,\beta) = \exp\left((\alpha-1)\sum \log X_j + (\beta-1)\sum \log(1-X_j) - \log B(\alpha,\beta)\right)\prod_j \mathbb{1}(x_j \in (0,1))$$

(b) Bring this distribution into the canonical form of the exponential family.

(c) Consider the statistic $T(X) = (\sum_{i=1}^n \log(X_i), \sum_{i=1}^n \log(1-X_i))$ for a random sample $X_1, X_2, \ldots, X_n$ from the Beta distribution. Prove whether $T(X)$ is complete.
The set $\{(\alpha,\beta) \in [2,3]^2\}$ is full-dimensional, so the exponential family is complete by the lemma seen in class.

(d) Now assume that $\Theta = \{(\alpha,\beta) \in \mathbb{R}^2_{>0} \ : \ \beta = \alpha + 2\}$. Is the statistic given above still complete? If not, can you find a sufficient and complete statistic for this new model?
We can write

$$f(x;\alpha,\beta) = \exp\left((\alpha-1)\sum \log\frac{x_j}{1-x_j} - \log B(\alpha,\beta)\right)\prod_j (1-x_j)^2 \mathbb{1}(x_j \in (0,1))$$

so the statistic $\tilde{T}(x) = \sum_j \log\frac{x_j}{1-x_j}$ is sufficient and complete.

5. Let $X_1, X_2, \ldots, X_n, n > 2$, be a random sample from the binomial distribution $b(1,\theta)$.

(a) Show that $T_n(X) = \sum_{j=1}^n X_j$ is a complete sufficient statistic for θ.
Since the Binomial distribution is a member of the exponential family and

$$f(x;\theta) = \theta^x(1-\theta)^{1-x} = e^{x\log(\theta)+(1-x)\log(1-\theta)}, \quad x = 0, 1,$$

$T_n = \sum_{i=1}^n X_i$ is a complete sufficient statistic for θ.

(b) Find the function $\psi(T_n)$ that is the UMVU estimator of θ.
Since $T_n \sim b(n,\theta)$, $E(T_n) = n\theta$. Thus, $\psi(T_1) = T_n/n$ is the MVUE of θ by part (a).

(c) Let $T_2 = (X_1 + X_2)/2$ and compute $E(T_2)$.
$2T_2 = X_1 + X_2 \sim b(2,\theta)$ gives $E(2T_2) = 2\theta$ hence $E(T_2) = \theta$.

(d) Determine $E(T_2|T_n = s)$.
By the iterative expectation and part (c), $E(E(T_2|T_n)) = E(T_2) = \theta$. Thus, $E(T_2|T_n = T_n)$ is MVUE of θ by the Rao-Blackwell and Lehmann-Scheffé theorems. By part (b), we found that $T_n/n = \bar{X}$ is MVUE of θ, which shows $E(T_2|T_n = T_n) = T_n/n$.