Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

LØSNINGSFORSLAG
EXAM IN ST2201 MATHEMATICAL STATISTICS
Wednsday 18 May 2007
Time: 09:00–14:00

Oppgave 1

Let $X_1, .., X_n$ be iid from a distribution with the density

$$\frac{1}{2\theta^3} x^2 e^{-x/\theta}, \ x \geq 0, \ \theta > 0.$$

a) Find the Method of moments estimator.

Solution. This is a gamma distribution with parameters $(3, \theta)$. The first moment is $\mu_1 = EX_1 = 3\theta$, therefore $\hat{\theta}_{MME}$ is solution of the equation

$$3\theta = m_1, \quad m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

i.e.

$$\hat{\theta}_{MME} = \frac{1}{3n} \sum_{i=1}^{n} X_i = \frac{1}{3} \bar{X}.$$

b) Find the MLE.

Solution. The likelihood function is

$$f(X|\theta) = \prod_{i=1}^{n} \frac{1}{2\theta^3} X_i^2 e^{-X_i/\theta} = \frac{1}{2^n \theta^{3n}} e^{-(1/\theta) \sum_{i=1}^{n} X_i} \left(\prod_{i=1}^{n} X_i^2 \right),$$

therefore

$$\frac{\partial \ln f(\mathbf{X}|\theta)}{\partial \theta} = -\frac{3n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} X_i,$$

and

$$\hat{\theta}_{MLE} = \frac{1}{3n}\sum_{i=1}^{n} X_i = \frac{1}{3}\bar{X}.$$

c) Find a one-dimensional sufficient statistic.

Solution. From part (b) we have

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x}), \theta))h(\mathbf{x}),$$

where

$$T(\mathbf{x}) = \sum_{i=1}^{n} x_i, \quad g(t, \theta) = \frac{1}{2^n\theta^{3n}} e^{-(t/\theta)}, \quad h(\mathbf{x}) = \prod_{i=1}^{n} x_i^2,$$

therefore, due to the factorization theorem,

$$T(\mathbf{X}) = \sum_{i=1}^{n} X_i$$

is a (one-dimensional) sufficient statistic.

Oppgave 2

Let $X_1, ..., X_n$ be iid from a normal distribution with expectation θ and variance 1. Two tests are used for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$. The rejection regions of these tests are $R_1 = \{x : \bar{x} > a\}$ and $R_2 = \{x : \max_i x_i > b\}$, where a and b are such that

$$P_{\theta_0}(\bar{X} > a) = P_{\theta_0}(\max_i X_i > b) = \alpha, \quad 0 < \alpha < 1.$$

a) Prove that

$$a = \theta_0 + \frac{z_\alpha}{\sqrt{n}}$$

(z_α is $(1-\alpha)$ quantile of the standard normal distribution, i.e. $\Phi(z_\alpha) = 1 - \alpha$).

Solution. Under H_0 $\bar{X} \sim n(\theta_0, 1/n)$, therefore

$$P_{\theta_0}(\bar{X} > a) = P_{\theta_0}(\sqrt{n}(\bar{X} - \theta_0) > \sqrt{n}(a - \theta_0)) = 1 - \Phi(\sqrt{n}(a - \theta_0)).$$

Solving the equation $1 - \Phi(\sqrt{n}(a - \theta_0)) = \alpha$, obtain

$$a = \theta_0 + \frac{z_\alpha}{\sqrt{n}}.$$

b) Prove that the first test is unbiased, i.e.

$$P_{\theta_0}(\mathbf{X} \in R_1) \leq P_{\theta_1}(\mathbf{X} \in R_1)$$

(the second test is also unbiased but prove this only for the first one).

Solution. Since $\theta_0 < \theta_1$, and $\Phi(x)$ increases,

$$P_{\theta_0}(\bar{X} > a) = P_{\theta_0}(\sqrt{n}(\bar{X} - \theta_0) > \sqrt{n}(a - \theta_0)) = 1 - \Phi(\sqrt{n}(a - \theta_0)) <$$

$$< 1 - \Phi(\sqrt{n}(a - \theta_1)) = P_{\theta_1}(\sqrt{n}(\bar{X} - \theta_1) > \sqrt{n}(a - \theta_1)) = P_{\theta_1}(\bar{X} > a).$$

c) Prove that the second test is consistent, i.e.

$$P_{\theta_1}(\mathbf{X} \in R_2) \to 1 \quad \text{as} \quad n \to \infty$$

(the first test is also consistent but prove this only for the second one).

Solution.

From condition $P_{\theta_0}(\max_i X_i > b) = \alpha$ we find

$$b = \theta + u_{(1-\alpha)^{1/n}}$$

where $u_\gamma = z_{1-\gamma}$ is γ-quantile of $N(0,1)$ i.e. $\Phi(u_\gamma) = \gamma$. For short notations, denote $u = u_{(1-\alpha)^{1/n}}$.

We have

$$P_{\theta_1}(\max_i X_i > b) = 1 - [P_{\theta_1}(X_1 - \theta_1 \leq u - (\theta_1 - \theta_0))]^n.$$

Under H_1 $X_i - \theta_1$ have the standard normal distribution, therefore, to prove consistency, it is sufficient to prove that if $Z \sim N(0,1)$, then

$$[P(Z \leq u - \Delta)]^n \to 0 \quad \text{as} \quad n \to \infty \tag{1}$$

for $\Delta > 0$. We have

$$[P(Z \leq u - \Delta)]^n = [P(Z \leq u) - P(u - \Delta < Z \leq u)]^n =$$

$$= \left[1 - (1 - (1-\alpha)^{1/n} + P(u - \Delta < Z \leq u))\right]^n =$$

$$= \left[1 - (1 - (1-\alpha)^{1/n} + P(u - \Delta < Z \leq u))\right]^{\frac{n(1-(1-\alpha)^{1/n} + P(u-\Delta<Z\leq u))}{1-(1-\alpha)^{1/n} + P(u-\Delta<Z\leq u)}}.$$

Note that

$$1 - (1-\alpha)^{1/n} + P(u - \Delta < Z \leq u) \to 0,$$

therefore

$$\left[1 - (1 - (1-\alpha)^{1/n} + P(u - \Delta < Z \leq u))\right]^{\frac{1}{1-(1-\alpha)^{1/n} + P(u-\Delta<Z\leq u)}} \to e^{-1},$$

and to prove (1), it is sufficient to show that

$$n(1-(1-\alpha)^{1/n}+P(u-\Delta<Z\le u))\to\infty.$$

It is easy to see that

$$n(1-(1-\alpha)^{1/n})\to -\ln(1-\alpha)$$

therefore we need to prove that

$$nP(u-\Delta<Z\le u)\to\infty. \tag{2}$$

Evidently

$$P(u-\Delta<Z\le u)\ge\frac{\Delta}{\sqrt{2\pi}}e^{-u^2/2}.$$

Let us use the inequality

$$e^{-u^2/2}\ge\sqrt{2\pi}u(1-\Phi(u))$$

(inequality (3.6.1) from the textbook), then

$$nP(u-\Delta<Z\le u)\ge\Delta un(1-\Phi(u))=\Delta un(1-(1-\alpha)^{1/n}).$$

Taking into account that $n(1-(1-\alpha)^{1/n})\to-\ln(1-\alpha)$ and $u\to\infty$, we see that the right hand side converges to ∞, that implies (2).

d) Which test is more powerful? *Hint.* One of these two tests is the size α Neyman-Pearson test.

Solution. Let R be the rejection region of the size α Neyman-Pearson test. Then

$$R=\left\{x:\frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)}<c\right\}$$

where c is such that $P_{\theta_0}(\mathbf{X}\in R)=\alpha$. But

$$\lambda(x)=\frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)}=\exp\left(\frac{n}{2}[(\theta_1^2-\theta_0^2)-(\theta_1-\theta_0)\bar{x}]\right)$$

is a decreasing function of $\bar{x}$, therefore $\lambda(x)<c$ iff $\bar{x}>a$ i.e. $R=R_1$. The Neyman-Pearson test is the most powerful, hence R_1 is more powerful than R_2.

Oppgave 3

Let $X_1,...,X_n$ be a random sample from a uniform on $[0,\theta]$ distribution, $\theta>0$. The prior for θ is a Pareto distribution $Pa(\alpha,\beta)$, $\alpha>0$, $\beta>0$, with the density

$$\pi(\theta)=\begin{cases}\alpha\beta^\alpha/\theta^{\alpha+1} & \text{for } \theta\ge\beta,\\ 0 & \text{for } \theta<\beta.\end{cases}$$

a) Show that the posterior is a Pareto distribution and find its parameters (for convenience, use the following short notations: $\mu = \max\{X_1, ..., X_n\}$, $\gamma = \max\{\mu, \beta\} = \max\{X_1, ..., X_n, \beta\}$).

Solution. The likelihood function is

$$f(\mathbf{X}|\theta) \propto \frac{1}{\theta^n}\left(\prod_{i=1}^{n} I_{[0,\theta]}(X_i)\right) = \frac{1}{\theta^n} I_{[0,\theta]}(\mu) = \frac{1}{\theta^n} I_{[\mu,\infty)}(\theta),$$

the prior

$$\pi(\theta) \propto \frac{1}{\theta^{\alpha+1}} I_{[\beta,\infty)}(\theta),$$

therefore

$$\pi(\theta|\mathbf{X}) \propto \frac{1}{\theta^{n+\alpha+1}} I_{[\gamma,\infty)}(\theta),$$

that is $Pa(n+\alpha, \gamma)$.

b) Find the GMLE of θ.

Solution. $\pi(\theta|\mathbf{X})$ takes its maximal value at $\theta = \gamma$, therefore

$$\hat{\theta}_{GMLE} = \gamma = \max\{X_1, ..., X_n, \beta\}.$$

c) Find the $(1-\delta)$ HPD credible interval for θ $(0 < \delta < 1)$.

Solution. We have (see solution of (a))

$$\pi(\theta|X) = \frac{(n+\alpha)\gamma^{n+\alpha}}{\theta^{n+\alpha+1}} I_{[\gamma,\infty)}(\theta)$$

i.e. $\pi(\theta|X) = 0$ for $\theta < \gamma$ and $\pi(\theta|X)$ decreases for $\theta \geq \gamma$, therefore the HPD interval has form $[\gamma, \gamma + \Delta]$. Let us find Δ.

$$1 - \delta = \int_{\gamma}^{\gamma+\Delta} \pi(\theta|X)d\theta = (n+\alpha)\gamma^{n+\alpha} \int_{\gamma}^{\gamma+\Delta} \frac{d\theta}{\theta^{n+\alpha+1}} = 1 - \left(\frac{\gamma}{\gamma+\Delta}\right)^{n+\alpha}.$$

Solving the equation

$$1 - \left(\frac{\gamma}{\gamma+\Delta}\right)^{n+\alpha} = 1 - \delta$$

we obtain

$$\Delta = \gamma \frac{1 - \delta^{\frac{1}{n+\alpha}}}{\delta^{\frac{1}{n+\alpha}}}.$$

Thus the HPD interval is

$$\left[\gamma, \gamma/\delta^{\frac{1}{n+\alpha}}\right].$$