**Norges teknisk–**
**naturvitenskapelige universitet**
**Institutt for matematiske fag**



**LØSNINGSFORSLAG**
EXAM IN TMA4295 STATISTICAL INFERENCE
Friday 6 June 2008
Time: 09:00–13:00

## Oppgave 1

Let $X_1, ..., X_n$ be iid from a beta distribution with parameters $(\theta, 4\theta)$, i.e. from a distribution with pdf

$$\frac{\Gamma(5\theta)}{\Gamma(\theta)\Gamma(4\theta)} x^{\theta-1}(1-x)^{4\theta-1},\ 0 < x < 1,\ \theta > 0.$$

**a)** List at least three different one-dimensional sufficient statistics.

**Solution.** The likelihood function is

$$f(\mathbf{x}|\theta) = \left(\frac{\Gamma(5\theta)}{\Gamma(\theta)\Gamma(4\theta)}\right)^n \left(\prod_{i=1}^n x_i(1-x_i)^4\right)^\theta \left(\prod_{i=1}^n x_i(1-x_i)\right)^{-1} = g(T(\mathbf{x});\theta)h(\mathbf{x}),$$

where

$$T(\mathbf{x}) = \prod_{i=1}^n x_i(1-x_i)^4,$$

$$g(t;\theta) = \left(\frac{\Gamma(5\theta)}{\Gamma(\theta)\Gamma(4\theta)}\right)^n t^\theta,$$

$$h(\mathbf{x}) = \left(\prod_{i=1}^n x_i(1-x_i)\right)^{-1}.$$

Therefore (Factorization theorem)

$$T(\mathbf{X}) = \prod_{i=1}^n X_i(1-X_i)^4$$

is a (one-dimensional) sufficient statistic. Any one-to-one transformation of a sufficient statistic is also a sufficient statistic, therefore, for example, each

$$T_k(\mathbf{X}) = [T(\mathbf{X})]^k,\ k = 2, 3, ...,$$

is a sufficient statistic.

**b)** Can the method of moments estimator (MME) of $\theta$ be found using the first moment (expectation)? Find MME using the second moment.

**Solution.** The first moment is $1/5$ independently of $\theta$, therefore it cannot be used. Let $\mu_2$ be the second moment. Then

$$\mu_2 = \frac{4}{25(5\theta + 1)} + \frac{1}{25}$$

or

$$\theta = \frac{4}{5(25\mu_2 - 1)} - \frac{1}{5}.$$

Thus MME is

$$\hat{\theta} = \frac{4}{5(25m_2 - 1)} - \frac{1}{5},$$

where

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

## Oppgave 2

Let $X_1, ..., X_n$ be iid from a geometric distribution with parameter $\theta$, i.e. from a distribution with pmf

$$\theta(1-\theta)^{x-1},\ x = 1, 2, ...;\ 0 < \theta < 1.$$

**a)** Find the maximum likelihood estimator $\hat{\tau}_1$ of $\tau(\theta) = 1/\theta$.

**Solution.** MLE of $\theta$ is solution of the equation

$$\frac{\partial \ln f(\mathbf{X}|\theta)}{\partial \theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} X_i - n}{1-\theta} = 0$$

that is

$$\hat{\theta}_{\text{MLE}} = \frac{1}{\bar{X}}.$$

Therefore (the invariance property of MLE) $\hat{\tau}_1 = \bar{X}$.

**b)** Find the asymptotic variance of the estimator $\hat{\tau}_1$.

**Solution.** The asymptotic variance is (see theory)

$$v(\theta) = \frac{[\tau'(\theta)]^2}{I_0(\theta)},$$

where $I_0(\theta)$ is the Fisher information of one observation.

$$I_0(\theta) = -E\frac{\partial^2 \ln f(X_i|\theta)}{\partial\theta^2} = \frac{1}{\theta^2(1-\theta)}.$$

Therefore, since $\tau'(\theta) = -1/\theta^2$, the asymptotic variance is

$$v(\theta) = \frac{1-\theta}{\theta^2}.$$

**c)** Find asymptotic $(1-\alpha)$ maximum likelihood confidence interval for $\theta$.

**Solution.** Since

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{I_0(\theta)}\right),$$

the variance is estimated by

$$\widehat{Var}\hat{\theta} = \frac{1}{nI_0(\hat{\theta})} = \frac{\hat{\theta}^2(1-\hat{\theta})}{n} = \frac{\bar{X}-1}{n\bar{X}^3}.$$

Then

$$1-\alpha \approx P\left(-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta}{\sqrt{\widehat{Var}\hat{\theta}}} \leq z_{\alpha/2}\right) = P\left(\frac{1}{\bar{X}} - z_{\alpha/2}\sqrt{\frac{\bar{X}-1}{n\bar{X}^3}} \leq \theta \leq \frac{1}{\bar{X}} + z_{\alpha/2}\sqrt{\frac{\bar{X}-1}{n\bar{X}^3}}\right).$$

Thus

$$\left[\frac{1}{\bar{X}} - z_{\alpha/2}\sqrt{\frac{\bar{X}-1}{n\bar{X}^3}}, \frac{1}{\bar{X}} + z_{\alpha/2}\sqrt{\frac{\bar{X}-1}{n\bar{X}^3}}\right]$$

is an asymptotic $(1-\alpha)$ maximum likelihood confidence interval for $\theta$.

**d)** Suppose that the first ten observations and each even observation are lost, and $\tau(\theta) = 1/\theta$ is estimated by

$$\hat{\tau}_2 = \frac{2}{n-10}\sum_{i=6}^{n/2} X_{2i-1}$$

(assume for simplicity that the sample size $n$ is always even). Find the asymptotic efficiency of $\hat{\tau}_2$ (that is asymptotic relative efficiency of $\hat{\tau}_2$ with respect to asymptotically efficient estimator $\hat{\tau}_1$, for which all observations are used).

**Solution.** Since

$$\sqrt{\frac{n-10}{2}}(\hat{\tau}_2 - \tau(\theta)) \xrightarrow{D} \mathcal{N}\left(0, \frac{[\tau'(\theta)]^2}{I_0(\theta)}\right),$$

and since

$$\frac{\sqrt{n}}{\sqrt{\frac{n-10}{2}}} \to \sqrt{2},$$

we have

$$\sqrt{n}(\hat{\tau}_2 - \tau(\theta)) \xrightarrow{D} \mathcal{N}\left(0, \frac{2[\tau'(\theta)]^2}{I_0(\theta)}\right).$$

Therefore the asymptotic efficiency of $\hat{\tau}_2$ is $1/2$.

## Oppgave 3

Let $X_1, ..., X_n$ be a random sample drawn from a Poisson distribution with parameter $\theta$.

**a)** Show that for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, the rejection region of a uniformly most powerful test has form

$$R = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i > c\right\}.$$

Let $\alpha$ be the significance level. Find (approximately) $c$ if $n$ is large enough so that the Central Limit Theorem can be used.

**Solution.** The likelihood function is

$$L(\theta; X) = e^{-n\theta}\theta^{\sum X_i}\left(\prod X_i!\right)^{-1},$$

therefore, if $\theta' < \theta''$, then the ratio

$$\frac{L(\theta'; X)}{L(\theta''; X)} = e^{n(\theta''-\theta')}\left(\frac{\theta'}{\theta''}\right)^{\sum X_i}$$

is a monotone (decreasing) function of $T(X) = \sum X_i$. Therefore the rejection region of UMP test has form

$$R = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i > c\right\}.$$

To find $c$ let us use CLT. We have $EX_i = \theta$, $Var(X_i) = \theta$ therefore

$$\alpha = P_{\theta_0}(\sum X_i > c) = P_{\theta_0}\left(\frac{\sum X_i - n\theta_0}{\sqrt{n\theta_0}} > \frac{c - n\theta_0}{\sqrt{n\theta_0}}\right) \approx 1 - \Phi\left(\frac{c - n\theta_0}{\sqrt{n\theta_0}}\right)$$

and

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

Thus the hypothesis is accepted if

$$\sum_{i=1}^{n} X_i \leq n\theta_0 + \sqrt{n\theta_0} z_\alpha$$

or

$$\bar{X} \leq \theta_0 + \sqrt{\frac{\theta_0}{n}} z_\alpha.$$

**b)** Prove that the test of part (a) is unbiased.

**Solution.** Prove the following. Let $Y_1 \sim \text{Poisson}(\lambda_1)$, $Y_2 \sim \text{Poisson}(\lambda_2)$, and $\lambda_1 < \lambda_2$. Then

$$P(Y_2 > c) > P(Y_1 > c)$$

for any $c > 0$. Indeed, consider $Y_3$ independent on $Y_1$ and such that $Y_3 \sim \text{Poisson}(\lambda_2 - \lambda_1)$. Then $Y_1 + Y_3 \sim \text{Poisson}(\lambda_2)$, and

$$P(Y_2 > c) = P(Y_1 + Y_3 > c) > P(Y_1 > c).$$

Now part (b) follows from the fact that

$$\sum_{i=1}^{n} X_i \sim \text{Poisson}(n\theta).$$

Less strong but also valid solution is based on the normal approximation (see solution of part (c)).

**c)** Find (approximately) and plot the power function $\pi(\theta)$ of the test of part (a). Find, in particular, $\lim_{\theta \downarrow 0} \pi(\theta)$, $\pi(\theta_0)$ and $\lim_{\theta \uparrow \infty} \pi(\theta)$

**Solution.**

$$\pi(\theta) = P_\theta\left(\sum_{i=1}^{n} X_i > c\right) = P_\theta\left(\frac{\sum X_i - n\theta}{\sqrt{n\theta}} > \frac{n(\theta_0 - \theta) + \sqrt{n\theta_0} z_\alpha}{\sqrt{n\theta}}\right) \approx$$

$$\approx 1 - \Phi\left(\frac{n(\theta_0 - \theta) + \sqrt{n\theta_0} z_\alpha}{\sqrt{n\theta}}\right).$$

Simple analysis shows that

$$\lim_{\theta \downarrow 0} \pi(\theta) = 0,$$

$$\pi(\theta_0) = \alpha$$

and

$$\lim_{\theta \uparrow \infty} \pi(\theta) = 1.$$

**d)** Find the $(1-\alpha)$ one-sided confidence interval that results from inverting the test of part (a).

**Solution.** Inverting the test of part (a), i.e. solving the inequality

$$\bar{X} \leq \theta + \sqrt{\frac{\theta}{n}} z_\alpha$$

with respect to $\theta$, we obtain the following $(1-\alpha)$ one-sided confidence interval:

$$\left[ \frac{1}{4} \left( \sqrt{\frac{z_\alpha^2}{n} + 4\bar{X}} - \frac{z_\alpha}{\sqrt{n}} \right)^2, \infty \right).$$