Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen:
Nikolai Ushakov 918 04 616

EXAM IN TMA4295 STATISTICAL INFERENCE

Tuesday 1 June 2004
Time: 09:00–14:00

Tillatte hjelpemidler:
K.Rottman: Mathematische Formelsamlung
Statistiske tabeller og formler, TAPIR
Godkjent lommekalkulator med tomt minne
Selvskrevet gult titteark på A4-ark utdelt av faglærer

Sensur: 22. juni 2004

Oppgave 1

A random variable X has distribution with probability mass function (pmf)

$$f(x) = Cf_1(x)f_2^2(x), \quad x = 0, 1, ..., m,$$

where $f_1(x)$ and $f_2(x)$ are respectively a binomial pmf with parameters (m, θ) (m is known and fixed) and a geometric pmf with parameter θ, i.e.

$$f_1(x) = \binom{m}{x}\theta^x(1-\theta)^{m-x}, \quad x = 0, 1, ..., m$$

and

$$f_2(x) = \theta(1-\theta)^x, \quad x = 0, 1, 2, ...$$

a) Show that $C = \theta^{-2}(1-\theta)^{-m}[1+\theta(1-\theta)]^{-m}$.

b) Show that this distribution belongs to the exponential family and find the natural parameter.

c) Find EX.

d) Let $X_1, ..., X_n$ be a sample from the distribution $f(x)$. Find a one-dimensional sufficient statistic.

Solution. a) We have

$$1 = \sum_{x=0}^{m} f(x) = C\theta^2(1-\theta)^m \sum_{x=0}^{m} \binom{m}{x} \theta^x (1-\theta)^x =$$

$$= C\theta^2(1-\theta)^m \sum_{x=0}^{m} \binom{m}{x} [\theta(1-\theta)]^x 1^{m-x} = C\theta^2(1-\theta)^m \Big(\theta(1-\theta)+1\Big)^m$$

thererfore

$$C = \theta^{-2}(1-\theta)^{-m}[1+\theta(1-\theta)]^{-m}.$$

b) We have

$$f(x) = [1+\theta(1-\theta)]^{-m} \binom{m}{x} \theta^x (1-\theta)^x = e^{x\ln[\theta(1-\theta)] - m\ln[1+\theta(1-\theta)] + \ln\binom{m}{x}}$$

i.e. $f(x)$ has form

$$e^{a(x)b(\theta)+c(\theta)+d(x)}$$

and, therefore, $f(x)$ belongs (by definition) to the exponential family with

$$a(x) = x,\ b(\theta) = \ln[\theta(1-\theta)],\ c(\theta) = -m\ln[1+\theta(1-\theta)],\ d(x) = \ln\binom{m}{x}.$$

Since $a(x) = x$, the natural parameter is (by definition)

$$b(\theta) = \ln[\theta(1-\theta)].$$

c) The shortest way: using formula $EX = -c'(\theta)/b'(\theta)$, we immediately get

$$EX = \frac{m\theta(1-\theta)}{1+\theta(1-\theta)}.$$

The direct way:

$$EX = \sum_{x=0}^{m} xf(x) = [1+\theta(1-\theta)]^{-m} \sum_{x=0}^{m} x\binom{m}{x} \theta^x (1-\theta)^x =$$

$$= [1+\theta(1-\theta)]^{-m}\sum_{x=0}^{m} x\frac{m!}{x!(m-x)!}\theta^x(1-\theta)^x =$$

$$= [1+\theta(1-\theta)]^{-m}\sum_{x=1}^{m} x\frac{m!}{x!(m-x)!}\theta^x(1-\theta)^x =$$

$$= [1+\theta(1-\theta)]^{-m}\sum_{x=1}^{m} \frac{m!}{(x-1)!(m-x)!}\theta^x(1-\theta)^x =$$

$$= [1+\theta(1-\theta)]^{-m}\sum_{y=0}^{m-1} \frac{(m-1)!}{y!(m-y-1)!}m\theta^{y+1}(1-\theta)^{y+1} =$$

$$= m\theta(1-\theta)[1+\theta(1-\theta)]^{-m}\sum_{y=0}^{m-1} \binom{m-1}{y}\theta^y(1-\theta)^y \cdot 1^{(m-1)-y} =$$

$$= m\theta(1-\theta)[1+\theta(1-\theta)]^{-m}[1+\theta(1-\theta)]^{m-1} = \frac{m\theta(1-\theta)}{1+\theta(1-\theta)}.$$

d) Using representation of part (b) for $f(x)$, we get

$$L(\theta; X) = \prod_{i=1}^{n} e^{X_i \ln[\theta(1-\theta)] - m\ln[1+\theta(1-\theta)] + \ln\binom{m}{X_i}} =$$

$$= e^{\ln[\theta(1-\theta)]\sum X_i - nm\ln[1+\theta(1-\theta)]}\prod\binom{m}{X_i},$$

therefore, due to the factorization theorem, $T(X) = \sum X_i$ is a univariate sufficient statistic.

Oppgave 2

Let $X_1, ..., X_n$ be a sample taken from the distribution with (probability density function) pdf

$$f(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x},\ x > 0,\ \theta > 0.$$

a) Which of the following three functions of θ

$$\tau_1(\theta) = \theta,\quad \tau_2(\theta) = 1/\theta,\quad \tau_3 = \ln\theta$$

admits an efficient estimator (we call an unbiased estimator efficient if its variance coincides with the lower bound of the Cramer-Rao inequality)? Why? Find this estimator.

b) Find maximum likelihood estimators of $\tau_1(\theta)$, $\tau_2(\theta)$ and $\tau_3(\theta)$.

c) Let $\tau(\theta)$ be the function from part (a) which admits an efficient estimator, and $T(X)$ be this (efficient) estimator. Does there exist a consistent estimator of $\tau(\theta)$ whose variance is strongly smaller than variance of $T(X)$ for each n?

Solution. a) A function $\psi(\theta)$ admits an efficient estimator iff the score function $\partial \ln L(\theta; X)/\partial\theta$ is represented in the form

$$\frac{\partial \ln L(\theta; X)}{\partial \theta} = c(\theta)(T(X) - \psi(\theta)).$$

Then $T(X)$ is the efficient estimator of $\psi(\theta)$. Find the score function in our case. The likelihood function is

$$L(\theta; X) = \frac{1}{2^n}\theta^{3n}\left(\prod_{i=1}^{n} X_i\right)^2 e^{-\theta\sum X_i},$$

and the score function is

$$\frac{\partial \ln L(\theta; X)}{\partial \theta} = \frac{3n}{\theta} - \sum_{i=1}^{n} X_i = -3n\left(\frac{1}{3n}\sum_{i=1}^{n} - \frac{1}{\theta}\right)$$

therefore $\tau_2(\theta) = 1/\theta$ (and only this function of the three) admits an efficient estimator. The estimator is

$$T(X) = \frac{1}{3}\bar{X} = \frac{1}{3n}\sum_{i=1}^{n} X_i.$$

b) Solving equation

$$\frac{\partial \ln L(\theta; X)}{\partial \theta} = 0$$

with respect to θ (the left hand side of the equation was found in part (a)), we get: MLE of $\tau_1(\theta) = \theta$ is

$$T_1(X) = \frac{3}{\bar{X}}.$$

Using the invariance property of MLE, we immediately obtain

MLE of $\tau_2(\theta) = 1/\theta$ is $T_2(X) = \bar{X}/3$,

MLE of $\tau_3(\theta) = \ln\theta$ is $T_3(X) = \ln(3/\bar{X})$.

c) The estimator $T(X)$ is consistent (this can be easily proved). Therefore, any estimator of the form $c_n T(X)$, where $c_n \to 1$, as $n \to \infty$, is consistent. Now it suffices to take c_n satisfying two conditions: 1) $c_n \to 1$, as $n \to \infty$ and 2) $|c_n| < 1$ (for example $c_n = n/(n+1)$). Then $S(X) = c_n T(X)$ is consistent, and

$$VarS(X) = c_n^2 VarT(X) < VarT(X).$$

Oppgave 3

Let $X_1, ..., X_n$ be a sample drawn from a Poisson distribution with parameter θ.

a) Suppose that n is large enough so that the Central Limit Theorem can be used. For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ find the acceptance region of the significance level α a UMP (uniformly most powerful) test.

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

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

Solution. a) 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 UMP test has form

$$\sum_{i=1}^{n} X_i > c \Longrightarrow H_1$$

where c is determined from condition

$$P_{\theta_0}(\sum X_i > c) = \alpha.$$

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 acceptance region has form

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

or

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

b)

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

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

Simple analysis shows that

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

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

and

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

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

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

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

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