UPPSALA UNIVERSITY
Department of Mathematics
Rolf Larsson

Exam in Mathematical Statistics
Inference Theory II, 1MS037
2021–01–14
Solutions

1. Consider the random variable

$$X = \begin{cases} 0, & \text{with probability } 4\theta_1\theta_2, \\ 1, & \text{with probability } \theta_1^2, \\ 2, & \text{with probability } 4\theta_2^2, \end{cases}$$

where $\theta_1 + 2\theta_2 = 1$. Suppose that we have an independent sample $\mathbf{X} = (X_1, ..., X_n)$ where all X_i are distributed as X.

Does the distribution belong to a strictly k-parametric family? In that case, determine k, the natural parameters(s) and the sufficient statistic(s). (5p)

Solution: Let $I\{A\}$ be 1 if A occurs and 0 otherwise. The probability function for X may be written as

$$\begin{aligned} p(x;\theta_1,\theta_2) &= (4\theta_1\theta_2)^{I\{x=0\}}(\theta_1^2)^{I\{x=1\}}(4\theta_2^2)^{I\{x=2\}} \\ &= 4^{I\{x=0\}+I\{x=2\}}\theta_1^{I\{x=0\}+2I\{x=1\}}\theta_2^{I\{x=0\}+2I\{x=2\}}. \end{aligned}$$

Letting n_0 be the number of zeros in the sample, etcetera, and observing that $n_0 = n - n_1 - n_2$, this gives the probability function of the sample, i.e. the likelihood, as

$$L(\theta_1,\theta_2;\mathbf{x}) = 4^{n_0+n_2}\theta_1^{n_0+2n_1}\theta_2^{n_0+2n_2} = 4^{n-n_1-n_2}\theta_1^{n+n_1-n_2}\theta_2^{n-n_1+n_2}.$$

Moreover, utilizing $\theta_1 = 1 - 2\theta_2$, we find that the likelihood may be expressed solely as a function of θ_2 as

$$\begin{aligned} L(\theta_2;\mathbf{x}) &= 4^{n-n_1}(1-2\theta_2)^{n+n_1-n_2}\theta_2^{n-n_1+n_2} \\ &= 4^{n-n_1}(1-2\theta_2)^n\theta_2^n \exp\left\{(n_2-n_1)\log\left(\frac{\theta_2}{1-2\theta_2}\right)\right\}, \end{aligned}$$

from which we see that we have a strictly 1-dimensional exponential family with natural parameter $\log\left(\frac{\theta_2}{1-2\theta_2}\right)$ and sufficient statistic $n_2 - n_1$.

2. A continuous random variable X is said to be Weibull distributed with parameters $\gamma > 0$ and $\beta > 0$ if it has density function

$$f(x) = \frac{\gamma}{\beta} x^{\gamma-1} \exp\left(-\frac{x^\gamma}{\beta}\right),$$

for $x \geq 0$, and 0 otherwise.

Suppose that we have an independent sample $\mathbf{X} = (X_1, ..., X_n)$ where all X_i are distributed as X.

You may without proof use that for $\gamma = 2$, $E(X) = 3\sqrt{\pi\beta}/2$, $E(X^2) = \beta$, $E(X^4) = 2\beta^2$.

(a) Assume that γ is fixed. Give a sufficient statistic for β. (2p)

Solution: With observations $\mathbf{x} = (x_1, ..., x_n)$, the likelihood is

$$\begin{aligned} L(\gamma, \beta) &= \prod_{i=1}^{n} \frac{\gamma}{\beta} x_i^{\gamma-1} \exp\left(-\frac{x_i^\gamma}{\beta}\right) \\ &= \frac{\gamma^n}{\beta^n} \left(\prod_{i=1}^{n} x_i^{\gamma-1}\right) \exp\left(-\frac{1}{\beta}\sum_{i=1}^{n} x_i^\gamma\right). \end{aligned}$$

Hence, by the factorization theorem, $\sum_{i=1}^{n} x_i^\gamma$ is a sufficient statistic for β.

(b) Suppose that $\gamma = 2$. Consider the estimator

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

Show that this is an unbiased estimator of β. (1p)

Solution: Because all $E(X_i^2) = \beta$, we have

$$E(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) = \frac{1}{n} n\beta = \beta,$$

which proves that $\hat{\beta}$ is unbiased for β

(c) Is the estimator in (b) efficient for β? Motivate your answer. (3p)

Solution: We need to check if the variance of $\hat{\beta}$ attains the lower bound. As above, we find that since

$$\mathrm{Var}(X^2) = E(X^4) - \{E(X^2)\}^2 = 2\beta^2 - \beta^2 = \beta^2,$$

the variance is

$$\mathrm{Var}(\hat{\beta}) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}(X_i^2) = \frac{1}{n^2} n\beta^2 = \frac{\beta^2}{n}.$$

To find the Cramér-Rao lower bound, we calculate the log likelihood

$$l(\gamma, \beta) = n \log \gamma - n \log \beta + (\gamma - 1) \sum_{i=1}^{n} \log x_i - \frac{1}{\beta} \sum_{i=1}^{n} x_i^{\gamma},$$

and its first two derivatives w.r.t. β as

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i^{\gamma}, \\ \frac{\partial^2 l}{\partial \beta^2} &= \frac{n}{\beta^2} - \frac{2}{\beta^3} \sum_{i=1}^{n} x_i^{\gamma}. \end{aligned}$$

We find that the Fisher information is given as minus the expectation of the second derivative (as a stochastic variable) as, inserting $\gamma = 2$,

$$I(\beta) = -\frac{n}{\beta^2} + \frac{2}{\beta^3} \sum_{i=1}^{n} E(X_i^2) = -\frac{n}{\beta^2} + \frac{2}{\beta^3} n\beta = \frac{n}{\beta^2}.$$

Hence,

$$\mathrm{Var}(\hat{\beta}) = \frac{\beta^2}{n} = \frac{1}{I(\beta)},$$

which means that the variance attains the Cramér-Rao lower bound. Consequently, the estimator is efficient.

3. Let X be a continuous random variable which is uniform on $(\theta, \theta+1)$, where $-\infty < \theta < \infty$. Suppose that we have an independent sample $\mathbf{X} = (X_1, ..., X_n)$ where all X_i are distributed as X.

   (a) Show that $(\min_i X_i, \max_i X_i)$ is a sufficient statistic for θ. (2p)

   Solution: With $I\{A\}$ as the indicator function of the event A, and the observations $\mathbf{x} = (x_1, ..., x_n)$, we have the likelihood

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} I\{\theta \leq x_i \leq \theta + 1\} = I\{\theta \leq \min_i x_i \leq \max_i x_i \leq \theta + 1\},$$

   which is a function of the parameter and the sample only through $\min_i x_i$ and $\max_i x_i$. Hence, by the factorization theorem, $(\min_i x_i, \max_i x_i)$ is sufficient for θ.

   (b) Show that $(\min_i X_i, \max_i X_i)$ is a *minimal* sufficient statistic for θ. (3p)

   Solution: With non equal samples $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$, we have, for $L(\theta; \mathbf{y}) > 0$,

$$\frac{L(\theta; \mathbf{x})}{L(\theta; \mathbf{y})} = \frac{I\{\theta \leq \min_i x_i \leq \max_i x_i \leq \theta + 1\}}{I\{\theta \leq \min_i y_i \leq \max_i y_i \leq \theta + 1\}}$$

   and this ratio is no function of θ if.f. $(\min_i x_i, \max_i x_i) = (\min_i y_i, \max_i y_i)$. This argument may be extended to the case $L(\theta; \mathbf{y}) = 0$ if we interpret $0/0$ as one.

   Alternatively, we could see that $L(\theta; \mathbf{x})$ and $L(\theta; \mathbf{y})$ are proportional (with proportionality constant as no function of θ) if.f. they are one for the same set of θ and zero for the same set of θ. For this to be the case, we must have that $(\min_i x_i, \max_i x_i) = (\min_i y_i, \max_i y_i)$.

4. Suppose that X is Bernoulli distributed with parameter p, i.e. that $P(X=1)=p=1-P(X=0)$. Suppose that we have an independent sample $\mathbf{X}=(X_1,...,X_n)$ where all X_i are distributed as X.

   (a) Show that X_1 is an unbiased estimator of p. (1p)

   Solution: This follows because

   $$E(X_1)=P(X_1=1)=p.$$

   (b) Show that $T=\sum_{i=1}^n X_i$ is sufficient for p. (1p)

   Solution: Let $I\{A\}$ be the indicator function of the event A. With observations $\mathbf{x}=(x_1,...,x_n)$, we have the likelihood

   $$L(p;\mathbf{x})=\prod_{i=1}^n p^{I\{x_i=1\}}(1-p)^{I\{x_i=0\}}=p^{n_1}(1-p)^{n_0},$$

   where n_1 is the number of ones in the sample, and n_0 is the number of zeros. Because $n_1=\sum_i x_i=t$ (say) and $n_0=n-n_1$, we have

   $$L(p;\mathbf{x})=p^t(1-p)^{n-t},$$

   and so, by the factorization theorem, T is sufficient for p.

   (c) Use the Rao-Blackwell theorem to find an unbiased estimator of p with smaller variance that X_1. (2p)

   Hint: $\binom{n}{t}=\frac{n}{t}\binom{n-1}{t-1}$.

   Solution: Because T is $\text{Bin}(n,p)$ and

   $$T-X_1=\sum_{i=2}^n X_i$$

   is $\text{Bin}(n-1,p)$, independent of X_1, we have

   $$\begin{aligned}
   &E(X_1|T=t)\\
   &=P(X_1=1|T=t)=\frac{P(X_1=1,T=t)}{P(T=t)}\\
   &=\frac{P(X_1=1,T-X_1=t-1)}{P(T=t)}=\frac{P(X_1=1)P(T-X_1=t-1)}{P(T=t)}\\
   &=\frac{p\binom{n-1}{t-1}p^{t-1}(1-p)^{(n-1)-(t-1)}}{\binom{n}{t}p^t(1-p)^{n-t}}=\frac{\binom{n-1}{t-1}}{\binom{n}{t}}=\frac{t}{n},
   \end{aligned}$$

   by the hint. Hence, by the Rao-Blackwell theorem, $T/n=\bar{X}$ is an unbiased estimator that has smaller variance than X_1.

(d) Is your estimator in (c) the best unbiased estimator (BUE) of p? Motivate your answer! (2p)

Solution: By the Lehmann-Sheffé theorem, this holds because T is complete since it is a sufficient statistic in an exponential family, $\bar{X} = T/n$, and $\bar{X}$ is derived by Rao-Bleckwellization through

$$E(\bar{X}|T) = E\left(\frac{T}{n} \,\Big|\, T\right) = \frac{T}{n} = \bar{X}.$$

It is also possible to solve this problem by showing that the variance of the estimator attains the Cramér-Rao lower bound.

5. Consider testing that the observation x comes from a discrete distribution with probability function $p_0(x)$ vs the alternative that it comes from a discrete distribution with probability function $p_1(x)$, where these two probability functions are given in the following table:

x	1	2	3	4	5
$p_0(x)$	0.10	0.10	0.04	0.30	0.46
$p_1(x)$	0.10	0.30	0.20	0.15	0.25

(a) Which is the most powerful (MP) test at level $\alpha = 0.05$? (2p)

Solution: By the Neyman-Pearson lemma, the MP test is based on the smallest possible values of $p_0(x)/p_1(x)$. We complement the table with such a row below:

x	1	2	3	4	5
$p_0(x)$	0.10	0.10	0.04	0.30	0.46
$p_1(x)$	0.10	0.30	0.20	0.15	0.25
$\frac{p_0(x)}{p_1(x)}$	1.00	0.33	0.20	2.00	1.84

The smallest possible value is $p_0(3)/p_1(3) = 0.2$. This has probability $p_0(3) = 0.04 < 0.05$ under H_0. Hence, we reject for $x = 3$. The next to smallest value is $p_0(2)/p_1(2) \approx 0.33$, which has probability $p_0(2) = 0.10$. Because $0.04 + 0.10 = 0.14 > 0.05$, we cannot always reject for $x = 2$ to achieve a test level of 0.05. Instead, we reject with probability γ such that $0.04 + \gamma * 0.10 = 0.05$. This yields $\gamma = 0.1$.
Hence, the MP test has test function

$$\varphi(x) = \begin{cases} 1 & \text{if } x = 3, \\ 0.1 & \text{if } x = 2, \\ 0 & \text{otherwise.} \end{cases}$$

We may confirm by calculating the test level as

$$E_0\{\varphi(X)\} = p_0(3) + 0.1p_0(2) = 0.04 + 0.1 * 0.10 = 0.05.$$

(b) Calculate the size of the type II error and the power for the MP test.(2p)

Solution: The power is the expected value of the test function (the probability to reject) under the alternative, i.e.

$$E_1\{\varphi(X)\} = p_1(3) + 0.1p_1(2) = 0.20 + 0.1 * 0.30 = 0.23.$$

The type II error is not to reject when the alternative is true. Thus, the probability for committing a type II error is

$$\beta = 1 - E_1\{\varphi(X)\} = 1 - 0.23 = 0.77.$$

(c) Calculate sizes of the errors of type I and II as well as the power for the test that rejects with probability 0.5 if $x = 2$, and otherwise does not reject. Compare to the power for the MP test. (2p)

Solution: Analogous to (b), the test size is

$$\alpha = 0.5p_0(2) = 0.5 * 0.10 = 0.05,$$

i.e. equal to the test size of the MP test. The power is

$$0.5p_1(2) = 0.5 * 0.30 = 0.15$$

and the probability of a type II error is

$$\beta = 1 - 0.15 = 0.85.$$

Hence, as we should from theory, we get a smaller power than for the MP test, as well as a larger β.

6. Let X be normally distributed with expectation 0 and variance σ^2, and suppose that we have an independent sample $\mathbf{X} = (X_1, ..., X_n)$ where all X_i are distributed as X. The corresponding observations are $\mathbf{x} = (x_1, ..., x_n)$.

(a) Consider testing H_0: $\sigma^2 \leq 1$ vs H_1: $\sigma^2 > 1$. Let $\chi^2_\alpha(n)$ be such that $P\{Y > \chi^2_\alpha(n)\} = \alpha$ for $Y \sim \chi^2(n)$.
Let

$$T(\mathbf{x}) = \sum_{i=1}^{n} x_i^2.$$

Show that the test that rejects H_0 if.f. $T(\mathbf{x}) > \chi^2_\alpha(n)$ is a UMP (uniformly most powerful) size α test for this situation. (3p)

Solution: We want to use the Blackwell theorem. In order to do so, we must att first prove that the model has an MLR in $T(\mathbf{x})$. To this end, we calculate the likelihood

$$L(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right),$$

so we have an exponential family model with natural parameter $-1/(2\sigma^2)$, which is non decreasing in σ^2, and sufficient statistic $T(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$. Hence, by theorem 5.3, we have an MLR in $T(\mathbf{x})$.
Consequently, it follows from the Blackwell theorem (5.4) that the test that rejects H_0 if.f. $T(\mathbf{x}) > \chi^2_\alpha(n)$ is a UMP (uniformly most powerful) size α test.

(b) Show that as $\sigma^2 \to 0$, the probability of rejecting H_0 in (a) tends to 0. (1p)

Hint: You may use without proof that $\sum_i^n X_i^2/\sigma^2$ is $\chi^2(n)$.

Solution: Using the hint, for a general σ^2 the probability of rejecting H_0 in (a) is

$$P\{T(\mathbf{X}) > \chi^2_\alpha(n)\} = P\left\{\frac{\sum_{i=1}^{n} X_i^2}{\sigma^2} > \frac{\chi^2_\alpha(n)}{\sigma^2}\right\} = P\left\{Y > \frac{\chi^2_\alpha(n)}{\sigma^2}\right\},$$

where Y is $\chi^2(n)$. As σ^2 tends to 0, $\chi^2_\alpha(n)/\sigma^2$ tends to infinity, which gives that the probability of rejecting H_0 tends to zero.

(c) Now, consider testing H_0: $\sigma^2 = 1$ vs H_1: $\sigma^2 \neq 1$. Is the test in (a) an unbiased size α test for this situation? Why or why not? (2p)

Solution: For an unbiased size α test, the minimum power under H_1 is greater than or equal to α. This would not be true for the test in (a) under the alternative that $\sigma^2 \neq 1$, since from (b) the power (i.e. the probability to reject) can be made arbitrarily small by choosing σ^2 small enough.

7. Suppose that we have a sample $\mathbf{X} = (X_1, X_2)$, where for $i = 1, 2$, X_i is Poisson with parameter μ_i. Moreover, suppose that $\mu_1 = \psi\mu_2$. Hence, the parameters of our model are ψ and μ_2.

   (a) Show that X_1 is sufficient for ψ and that $T = X_1 + X_2$ is sufficient for μ_2. (2p)

   Solution: Say that we have observations (x_1, x_2). The likelihood is

$$L(\psi, \mu_2) = \frac{(\psi\mu_2)^{x_1}}{x_1!} e^{-\psi\mu_2} \frac{\mu_2^{x_2}}{x_2!} e^{-\mu_2} = \frac{\psi^{x_1}\mu_2^{x_1+x_2}}{x_1!x_2!} e^{-\mu_2(1+\psi)}.$$

   Hence, from the factorization theorem, we find that X_1 is sufficient for ψ and that $T = X_1 + X_2$ is sufficient for μ_2.

   (b) Show that $X_1|T = t$ is Binomial with parameters t and $\psi/(1+\psi)$. (1p)

   Solution: From the definition of conditional probability and independence of X_1 and X_2, we find

$$\begin{aligned}
&P(X_1 = x_1|T = t) \\
&= \frac{P(X_1 = x_1, X_2 = t - x_1)}{P(T = t)} = \frac{P(X_1 = x_1)P(X_2 = t - x_1)}{P(T = t)} \\
&= \frac{\frac{\mu_1^{x_1}}{x_1!} e^{-\mu_1} \frac{\mu_2^{t-x_1}}{(t-x_1)!} e^{-\mu_2}}{\frac{(\mu_1+\mu_2)^t}{t!} e^{-(\mu_1+\mu_2)}} = \frac{t!}{x_1!(t-x_1)!} \frac{\mu_1^{x_1}\mu_2^{t-x_1}}{(\mu_1+\mu_2)^t} \\
&= \binom{t}{x_1} \left(\frac{\mu_1}{\mu_1+\mu_2}\right)^{x_1} \left(\frac{\mu_2}{\mu_1+\mu_2}\right)^{t-x_1},
\end{aligned}$$

   which shows that $X_1|T = t$ is Binomial with parameters t and

$$\frac{\mu_1}{\mu_1+\mu_2} = \frac{\psi\mu_2}{\psi\mu_2+\mu_2} = \frac{\psi}{\psi+1}.$$

(c) Suppose that we want to test H_0: $\psi \geq 1$ vs H_1: $\psi < 1$, and that we have the observations $x_1 = 0$ and $x_2 = 4$.

Does the UMP α similar test reject H_0 at the 10% level? (3p)

Solution: Since T is sufficient for the nuisance parameter μ_2, a similar test on hypotheses on ψ should be based on the distribution of X_1 conditional on $T = t$. Theorem 5.7 (the generalization of Blackwell's theorem to similar tests) gives that the UMP α similar test rejects H_0 if $X_1 < c_0(t)$, and with probability γ it rejects if $X_1 = c_0(t)$, where $c_0(t)$ and γ are derived so that the test has the correct size α.

We observe $x_1 = 0$ and $x_2 = 4$, i.e. $t = x_1 + x_2 = 4$. The UMP α similar test rejects H_0 at the 10% level if $P(X_1 = 0|T = 4) < 0.1$. (This means that the observed P value is smaller that 10%.) Now, we find from (b) that, for $\psi = 1$ (i.e. $\psi/(1+\psi) = 1/2$),

$$P(X = 0|T = 4) = \left(\frac{1}{2}\right)^4 = \frac{1}{16} < 0.1,$$

which means that we reject H_0 at the 10% level with this test.