UPPSALA UNIVERSITY
Department of Mathematics
Rolf Larsson

Exam in Mathematical Statistics
Inference Theory II, 1MS037
2019–01–04
Solutions

1. Suppose $X_1, ..., X_n$ are independent negative binomial with parameters (r, p), where r is a positive integer and $0 < p < 1$, i.e. the probability mass function of X_i is

$$p(x) = \binom{r+x-1}{x} p^r (1-p)^x$$

for $x = 0, 1, 2, ...$ Suppose that r is known and that p is unknown.

(a) Find a sufficient statistic for p. (3p)

Solution: Let the observed sample be $\mathbf{x} = (x_1, ..., x_n)$. The joint probability mass function (likelihood) is

$$\begin{aligned} p(\mathbf{x}; p) &= \prod_{i=1}^{n} \binom{r+x_i-1}{x_i} p^r (1-p)^x \\ &= \left\{ \prod_{i=1}^{n} \binom{r+x_i-1}{x_i} \right\} p^{nr} (1-p)^{\sum_{i=1}^{n} x_i} \\ &= \left\{ \prod_{i=1}^{n} \binom{r+x_i-1}{x_i} \right\} p^{nr} \exp\left\{ \log(1-p) \sum_{i=1}^{n} x_i \right\}. \end{aligned}$$

The observations enter together with the parameter only through $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$. Hence, by the factorization theorem, $T(\mathbf{x})$ is sufficient.

(b) Find a minimal sufficient statistic for p. (It could be the same one as in (a).) (2p)

Solution: A statistic $T(\mathbf{x})$ is minimal sufficient if for two distinct samples $\mathbf{x}$ and $\mathbf{y}$, the ratio of the likelihoods $L(\theta; \mathbf{x})/L(\theta; \mathbf{y})$ does not depend on θ only if $T(\mathbf{x}) = T(\mathbf{y})$. In our case, we have

$$\begin{aligned} \frac{L(p; \mathbf{x})}{L(p; \mathbf{y})} &= \frac{\left\{ \prod_{i=1}^{n} \binom{r+x_i-1}{x_i} \right\} p^{nr} (1-p)^{\sum_{i=1}^{n} x_i}}{\left\{ \prod_{i=1}^{n} \binom{r+y_i-1}{y_i} \right\} p^{nr} (1-p)^{\sum_{i=1}^{n} y_i}} \\ &= \frac{\prod_{i=1}^{n} \binom{r+x_i-1}{x_i}}{\prod_{i=1}^{n} \binom{r+y_i-1}{y_i}} (1-p)^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}, \end{aligned}$$

which does not depend on p only if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.
Hence, $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ is also minimal sufficient.
Minimal sufficiency also follows because we have a strictly one-parameter exponential family.

2. Suppose that the discrete random variable X can take the values 1, 2, 3 according to

$$P(X=1)=\theta_1^2,\ P(X=2)=\theta_2^2, P(X=3)=2\theta_1\theta_2,$$

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

(a) Does the distribution belong to a strictly k-parametric family? In that case, determine k, the natural parameter(s) and the sufficient statistic(s). (2p)

Solution: Let the number of observed ones and twos be n_1 and n_2, respectively. Then, the number of observed threes is $n-n_1-n_2$. The likelihood function is (with $\theta_2=1-\theta_1$)

$$\begin{aligned} L(\theta_1) &= \left(\theta_1^2\right)^{n_1}\left(\theta_2^2\right)^{n_2}(2\theta_1\theta_2)^{n-n_1-n_2} = 2^{n-n_1-n_2}\theta_1^{n+(n_1-n_2)}\theta_2^{n-(n_1-n_2)} \\ &= 2^{n-n_1-n_2}\theta_1^n\theta_2^n\left(\frac{\theta_1}{\theta_2}\right)^{n_1-n_2} = 2^{n-n_1-n_2}\theta_1^n(1-\theta_1)^n\left(\frac{\theta_1}{1-\theta_1}\right)^{n_1-n_2} \\ &= 2^{n-n_1-n_2}\theta_1^n(1-\theta_1)^n\exp\left\{(n_1-n_2)\log\left(\frac{\theta_1}{1-\theta_1}\right)\right\}. \end{aligned}$$

By inspection, we find that this is a strictly one-parameter exponential family with natural parameter $\log\left(\frac{\theta_1}{1-\theta_1}\right)$ and sufficient statistic n_1-n_2.

(b) Show that the Fisher information for θ_1 is $\frac{2n}{\theta_1(1-\theta_1)}$. (2p)

Solution: The log likelihood is

$$l(\theta_1)=C+n\log\{\theta_1(1-\theta_1)\}+(n_1-n_2)\log\left(\frac{\theta_1}{1-\theta_1}\right),$$

where C is a constant not depending on θ_1. By simplification, we find the first two derivatives as

$$\begin{aligned} l'(\theta_1) &= \frac{n(1-2\theta_1)+n_1-n_2}{\theta_1(1-\theta_1)}, \\ l''(\theta_1) &= \frac{-2n\theta_1(1-\theta_1)-\{n(1-2\theta_1)+n_1-n_2\}(1-2\theta_1)}{\theta_1^2(1-\theta_1)^2}. \end{aligned}$$

Letting N_1 and N_2 be the random variables corresponding to n_1 and n_2, and using $E(N_1)=n\theta_1^2$, $E(N_2)=n(1-\theta_1)^2$, implying

$$E(N_1-N_2)=n\left\{\theta_1^2-(1-\theta_1)^2\right\}=n(2\theta_1-1),$$

we obtain the Fisher information

$$I(\theta_1)=-E\left\{l''(\theta_1;N_1,N_2)\right\}=\frac{2n}{\theta_1(1-\theta_1)}.$$

(c) Is there any unbiased estimator of θ_1 with variance strictly less than $\frac{\theta_1(1-\theta_1)}{2n}$? Motivate your answer. (2p)

Solution: The Cramér-Rao lower bound for the variance is

$$\frac{1}{I(\theta)} = \frac{\theta_1(1-\theta_1)}{2n}.$$

This means that an unbiased estimator of θ_1 with strictly smaller variance than this cannot exist. The answer is no.

3. Suppose $X_1, ..., X_n$ are independent, distributed according to a continuous uniform distribution on $[-\theta, \theta]$. We have observations $x_1, ..., x_n$.

(a) Show that the maximum likelihood estimate (MLE) is $\hat{\theta}_{\text{MLE}} = \max_i |x_i|$, where $|a|$ is the absolute value of a. (3p)

Solution: We may write the probability mass function as

$$p(x) = \frac{1}{2\theta} I\{-\theta \leq x \leq \theta\},$$

where $I\{A\} = 1$ if A holds and 0 otherwise. This gives the likelihood

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2\theta} I\{-\theta \leq x_i \leq \theta\} = (2\theta)^{-n} I\{-\theta \leq \text{all } x_i \leq \theta\}$$

Observing that all $x_i \in [-\theta, \theta]$ is equivalent to $\max_i |x_i| \leq \theta$, and that $(2\theta)^{-n}$ is maximized at the smallest possible θ, we deduce that the MLE must be $\hat{\theta}_{\text{MLE}} = \max_i |x_i|$.

(b) Assume that the observations are $2.5, -3.2, -0.5, 2.0$. Consider testing H_0: $\theta = 4$ vs H_1: $\theta > 4$, using the MLE as test statistic.
Calculate the p value. (3p)

Hint: If X is uniform on $[-\theta, \theta]$, then $|X|$ is uniform on $[0, \theta]$. This fact may be used without proof.

Solution: Let $T = \max_i |X_i|$. The alternative points at higher values of T. We observe $t_{obs} = 3.2$. Hence, the p value is the probability under H_0 of getting $T \geq 3.2$. Using the hint, this gives the p value

$$\begin{aligned} P_0(T \geq 3.2) &= 1 - P_0(T < 3.2) = 1 - P_0(\text{all } |X_i| < 3.2) \\ &= 1 - P_0(|X_1| < 3.2)^4 = 1 - \left(\frac{3.2}{4}\right)^4 = 1 - 0.8^4 = 0.59. \end{aligned}$$

4. Suppose $X_1, ..., X_n$ are independent, distributed as $X \sim N(\mu, 1)$. We have observations $x_1, ..., x_n$.

   (a) Show that $\bar{X}$ is a sufficent statistic for μ. (1p)

   Solution: The likelihood is

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

   Here, the observations enter together with μ only via $\bar{x}$, from which it follows by the factorization theorem that $\bar{x}$ is sufficient.

   (b) Let $\theta = P(X \leq a)$ for some real number a. In the rest of this problem, the aim is to estimate θ. Show that the estimator

$$U = \begin{cases} 1, & \text{if} \quad X_1 \leq a, \\ 0, & \text{if} \quad X_1 > a \end{cases}$$

   is unbiased for θ. (2p)

   Solution: We have

$$E(U) = P(X_1 \leq a) = P(X \leq a) = \theta,$$

   showing that U is unbiased for θ.

   (c) Find a unbiased estimator of θ with smaller variance than U. (3p)

   Hint: It may be used without proof that $(X_1|\bar{X} = t)$ is normal with expectation t and variance $1 - n^{-1}$.

   Solution: Use the Rao-Blackwell theorem. Since U is unbiased and $\bar{X}$ is sufficient, an unbiased estimator with variance less than or equal to the variance of U corresponds to the estimate (from the hint)

$$E(U|\bar{X} = t) = P(X_1 \leq a|\bar{X} = t) = \Phi\left(\frac{a - t}{\sqrt{1 - \frac{1}{n}}}\right),$$

   where Φ is the standard normal distribution function.

   Since this estimate is a function of t, which U is not, it cannot equal U with probability one. Hence, since by the Rao-Blackwell theorem that would be the only possibility to get an estimator with the same variance using the above procedure, we conclude that we have arrived at an estimator with strictly smaller variance than U.

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	6
$p_0(x)$	0.10	0.02	0.05	0.20	0.30	0.33
$p_1(x)$	0.05	0.10	0.20	0.25	0.20	0.20

(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	6
$p_0(x)$	0.10	0.02	0.05	0.20	0.30	0.33
$p_1(x)$	0.05	0.10	0.20	0.25	0.20	0.20
$\frac{p_0(x)}{p_1(x)}$	2.00	0.20	0.25	0.80	1.50	1.65

The smallest possible value is $p_0(2)/p_1(2) = 0.2$. This has probability $p_0(2) = 0.02$ under H_0. Hence, we reject for $x = 2$. The next to smallest value is $p_0(3)/p_1(3) = 0.25$, which has probability $p_0(3) = 0.05$.

Because $0.02 + 0.05 = 0.07 > 0.05$, we cannot always reject for $x = 3$ to achieve a test level of 0.05. Instead, we reject with probability γ such that $0.02 + \gamma * 0.05 = 0.05$. This yields $\gamma = 0.6$.

Hence, the MP test has test function

$$\varphi(x) = \begin{cases} 1 & \text{if } x = 2, \\ 0.6 & \text{if } x = 3, \\ 0 & \text{otherwise.} \end{cases}$$

We may confirm by calculating the test level as

$$E_0\{\varphi(X)\} = p_0(2) + 0.6p_0(3) = 0.02 + 0.6 * 0.05 = 0.05.$$

(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.

$$E_1\{\varphi(X)\} = p_1(2) + 0.6p_1(3) = 0.10 + 0.6 * 0.20 = 0.22.$$

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

$$\beta = 1 - E_1\{\varphi(X)\} = 1 - 0.22 = 0.78.$$

(c) Calculate sizes of the errors of type I and II as well as the power for the test with critical region $\{x = 3\}$. Compare to the power for the MP test. (2p)

Solution: Analogous to (b), the test size is $\alpha = p_0(3) = 0.05$, i.e. equal to the test size of the MP test. The power is $p_1(3) = 0.20$ and the probability of a type II error is $\beta = 1 - 0.20 = 0.80$.

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

6. Suppose $X_1, ..., X_n$ are independent, distributed as X which is exponential with intensity $\beta > 0$, i.e. with density function

$$f(x) = \beta \exp(-\beta x), \quad x > 0,$$

and 0 otherwise. Let $x_1, ..., x_n$ be the observations.

(a) Show that this distribution belongs to a one-parameter exponential family. (1p)

Solution: The likelihood is

$$L(\beta) = \prod_{i=1}^{n} \beta \exp(-\beta x_i) = \beta^n \exp\left(-\beta \sum_{i=1}^{n} x_i\right),$$

which is of the one parameter exponential family form
$A(\theta) \exp\{\zeta(\beta) T(\mathbf{x})\} h(\mathbf{x})$.

(b) Give the natural parameter and the sufficient statistic. (1p)

Solution: From the expression above, we find the natural parameter $\zeta(\beta) = -\beta$ and the sufficient statistic $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$.

(c) Consider testing H_0: $\beta \geq \beta_0$ vs H_1: $\beta < \beta_0$. Show that the uniformly most powerful (UMP) test has critical region $\bar{x} > C$ where $\bar{x}$ is the mean of the observations and C is some constant. (3p)

Solution: In terms of the natural parameter $\zeta(\beta) = -\beta$, the likelihood is monotonely increasing given T. Hence, the UMP test of H_0: $-\beta \leq -\beta_0$ vs H_1: $-\beta > -\beta_0$ (equivalent to our hypotheses above) has the critical region $T(\mathbf{x}) = \sum_{i=1}^{n} x_i > C'$, i.e. $\bar{x} > C$ where $C = C'/n$ is a constant.

7. Suppose we have one observation of X_1, which is Poisson with parameter θ, and one of X_2, which is Poisson with parameter $\theta + \delta$, where X_1 and X_2 are independent. The parameter space consists of all $\theta > 0$ and all $\delta \geq 0$.

   (a) If $\delta = 0$, is $X_1 + X_2$ complete and sufficient for θ? Why or why not?(2p)

   Solution: If $\delta = 0$, the likelihood is

$$L(\theta) = \frac{\theta^{x_1}}{x_1!} e^{-\theta} \frac{\theta^{x_2}}{x_2!} e^{-\theta} = e^{-2\theta} \exp\{(x_1 + x_2)\log(\theta)\}(x_1!x_2!)^{-1},$$

   and we see by inspection that we have a strictly one-parameter exponential family with sufficient statistic $t(x_1, x_2) = x_1 + x_2$. Hence, $X_1 + X_2$ is sufficient and complete.
   Completeness may also be proved from the fact that
   $T = X_1 + X_2 \sim \text{Po}(2\theta)$, and that

$$0 = E\{h(T)\} = \sum_{k=0}^{\infty} h(k) \frac{(2\theta)^k}{k!} e^{-2\theta}$$

   implies that all $h(k) = 0$. (If a polynomial in $\theta > 0$ is zero, then all its coefficients must be zero.)

   (b) Consider testing H_0: $\delta = 0$ vs H_1: $\delta > 0$.
   Derive the UMP α-similar test. Do we reject at level 0.05 if $x_1 = 2$ and $x_2 = 5$? (4p)

   Hint: It may be used without proof that if $\delta = 0$, then $(X_2 | X_1 + X_2 = t)$ is $\text{Bin}(t, 1/2)$.

   Solution: Without restricting δ, the likelihood is

$$\begin{aligned} L(\delta, \theta) &= \frac{\theta^{x_1}}{x_1!} e^{-\theta} \frac{(\theta + \delta)^{x_2}}{x_2!} e^{-(\theta+\delta)} \\ &= e^{-(2\theta+\delta)} \theta^{x_1+x_2} \left(1 + \frac{\delta}{\theta}\right)^{x_2} (x_1!x_2!)^{-1} \\ &= e^{-(2\theta+\delta)} \exp\left\{(x_1 + x_2)\log(\theta) + x_2 \log\left(1 + \frac{\delta}{\theta}\right)\right\} (x_1!x_2!)^{-1}. \end{aligned}$$

   This is a two-parameter exponential family with parameter of interest $\lambda = 1 + \delta/\theta$ and nuisance parameter θ. The corresponding sufficient statistics are $u = x_2$ and $t = x_1 + x_2$. Because λ is a monotonely increasing function of δ, the UMP α-similar test rejects with probability one if $u > c_0(t)$, where $c_0(t)$ is the smallest possible $c(t)$ such that $P_0\{U > c(t)|T = t\} \leq \alpha$, with $U = X_2$ and $T = X_1 + X_2$. Moreover, we reject with probability γ if $u = c_0(t)$.
   To calculate $c_0(t)$, we have from the hint that

$$P_0\{U > c(t)|T = t\} = \sum_{k=c(t)+1}^{t} \binom{t}{k} \left(\frac{1}{2}\right)^t.$$

For $x_1 = 2$ and $x_2 = 5$, we have $t = 7$ and observe $u = 5$, so the p value of the test is

$$\binom{7}{5}\left(\frac{1}{2}\right)^7 + \binom{7}{6}\left(\frac{1}{2}\right)^7 + \binom{7}{7}\left(\frac{1}{2}\right)^7 = \frac{29}{2^7} = 0.23 > 0.05.$$

Hence, we do not reject H_0 at the level 0.05.

In fact, since $P_0\{U > 5|T = 7\} = 8/2^7 = 0.0625$ and $P_0\{U > 6|T = 7\} = 1/2^7 = 0.0078$, for $t = 7$ we reject with probability 1 of we observe $u = 7$, and with probability $\gamma = 0.675$ if we observe $u = 6$.

$(1/2^7 + 0.675 * 8/2^7 = 0.05)$