NMSA332 EXERCISE SESSIONS: QUESTIONS AND ANSWERS

STANISLAV NAGY AND MAREK OMELKA

Week 5: the use of sufficient statistics

Example 40: Geometric distribution.

Question. *I don't understand part (iii). How did we get the exponential function? Why is it a complete sufficient statistic? How did we get $a(\theta)$?*

To show that the estimator is UMVUE*, we want to use Theorem 6. We know that $\sum_i X_i$, or equivalently $\bar{X}$ is a sufficient statistic for p, and we have an estimator that is a function of $\bar{X}$, that is at the same time unbiased. It remains to show that the statistic $\bar{X}$ is also complete†. For this we typically use Theorem 3 — most of the used distributions are from exponential systems. We write the density of the geometric distribution in the form of an exponential system

$$f(x) = p(1-p)^x = p\exp(x\log(1-p)),$$

for x a non-negative integer, and see that it is an exponential system, but for parameter $\theta = \log(1-p)$. Parameter θ is called the canonical parameter of an exponential system. We also see (Theorem 3 again) that $\sum X_i$ (or $\bar{X}$) is complete sufficient for θ. We express p as a parametric function of θ

$$p = a(\theta) = 1 - \exp(\theta),$$

and apply Theorem 6 with this parametric function of the canonical parameter θ. We get that our estimator $u(\bar{X})$ is indeed UMVUE for p.

Example 41: Special multinomial distribution.

Question. *Part (ii) is not clear. How does it follow that the statistic is sufficient? In part (iv) I don't see why the expression is rewritten as an exponential.*

This is a multinomial distribution. Its density can be expressed as

$$f(x) = p^{\mathbb{I}[x=-1]}(1-2p)^{\mathbb{I}[x=0]}p^{\mathbb{I}[x=1]} = p^{\mathbb{I}[x\neq 0]}(1-2p)^{1-\mathbb{I}[x\neq 0]} \tag{1}$$

for $x \in \{-1,0,1\}$, zero elsewhere. The second equality is true because $X_i \in \{-1,0,1\}$ almost surely. We wrote the density only in terms of $\mathbb{I}[x \neq 0]$; if we now take the product of densities of all X_i, we get

$$\prod_{i=1}^{n} f(X_i) = p^{\sum \mathbb{I}[X_i\neq 0]}(1-2p)^{n-\sum \mathbb{I}[X_i \neq 0]}.$$

By Neyman's factorization criterion (Theorem 1) the statistic $S(\boldsymbol{X})$ is sufficient.

E-mail addresses: `nagy@karlin.mff.cuni.cz`.

Date: April 11, 2020.

*uniformly minimum-variance unbiased estimator — najlepší nestranný odhad

†úplná

In part (iv) we again need, for the second Lehmann-Scheffé's theorem, that statistic $S(\boldsymbol{X})$ is complete. For that we use Theorem 3 again. We need to write the multinomial distribution as a member of the exponential system of distributions. For that we use a common trick that helps to work with the general multinomial distribution, and instead of the indicators as in (1) I simply write in the solutions x_1, x_2, and x_3 (it's not correct to write it like that, I just simplified the notation not to write the indicators everywhere). For example, by x_2 in the solutions I mean $\mathbb{I}[x_i = 0]$ etc. What we get in the end can be written as[‡]

$$f(x) = \exp(\mathbb{I}[x_i \neq 0] \log(p/(1-2p)))(1-2p),$$

meaning that this is an exponential system with canonical parameter $\theta = \log(p/(1-2p))$, p is a parametric function of θ, and $S(\boldsymbol{X})$ is a complete sufficient statistic for θ (Theorem 3). By Theorem 5 we now know that our estimator is UMVUE.

Example 44: Normal distribution.

Question. *How do I know that* $\mathrm{E}\,\chi_{n-1} = 1/a_n$*? Or is this something else? I don't see this from Example 13. I don't understand part (iii) — how is this computed, and how does this follow from the L-S theorem (and which L-S theorem)? How do I find the estimator in part (v)?*

The expectation of χ_{n-1}-distribution is a known constant[§]. It can be computed as an integral, but this is not required from you. In part (iv) of Example 13, in the special case when $\mu = 0$ (but the computation is the same also for general μ) we showed that for $\widetilde{\sigma}_n$ to be unbiased we need to scale it by a_n.

In part (iii) we can use any Lehmann-Scheffé's theorem (Theorem 5 or 6). We know that the conditions of these theorems are satisfied, and from them we know that there is a *unique* UMVUE estimator of μ that must be a function of the complete sufficient statistic. But, the median is clearly not a function of this statistic, therefore it cannot be UMVUE.

In part (v) you need to find an unbiased estimator of μ^2 that is a function of the complete sufficient statistic. Then Theorem 6 gives you that it must be the unique UMVUE estimator. Think about how to find such an estimator.

Example 46: Shift in exponential distribution.

Question. *In the beginning, how do I see that* $\min X_i$ *is sufficient?*

This is directly Theorem 1. We know that λ is a known constant, so we just wrote the density as a function where δ couples with X only in the indicator via $\min X_i$.

Example 47: λ in exponential distribution.

Question. *How do I know immediately that* $\sum X_i$ *is complete sufficient? In part (i) I found an unbiased estimator. How do I know that it is UMVUE? How did we get the estimator in part (iii)?*

Completeness and sufficiency follow from Theorem 3. We know that our unbiased estimator is a function of a complete sufficient statistic, so we just use Theorem 6 to get that it must be UMVUE.

For part (iii) again as in part (v) in Example 44. You need to find an unbiased estimator of λ^k that is a function of $\bar{X}$, and use Theorem 6.

[‡]The multinomial coefficient in the formula in my solution should not be there, that's an error.

[§]https://en.wikipedia.org/wiki/Chi_distribution

Example 48: θ in uniform distribution.

Question. *How did we get $x\, n\, x^{n-1}$ in the formula for the expectation?*

Find the distribution function of $\max_{i=1,\dots,n} X_i$ and take a derivative of that. You get its density. An integral of the density is the expectation.

Example 49: General multinomial distribution.

Question. *How did we find the estimator in (i)?*

It is not an estimator, it is a complete sufficient statistic. We rewrite the density of $\boldsymbol{X} = (X_1, \dots, X_K)^\mathsf{T}$ (that is, a single K-dimensional random vector with multinomial distribution $\mathsf{M}\,(1; p_1, \dots, p_K)$) as a density from an exponential system

$$f(\boldsymbol{x}) = \prod_{k=1}^{K} {p_k}^{x_k} = \exp\left(\sum_{k=1}^{K} x_k \log(p_k)\right) \quad \text{for } \boldsymbol{x} = (x_1, \dots, x_K)^\mathsf{T} \text{ with } x_k \in \{0, 1\}, \sum_{k=1}^{K} x_k = 1,$$

zero elsewhere. Now we want to use Theorem 3, but with parameter $(\log(p_1), \dots, \log(p_K))^\mathsf{T}$ the condition about a non-degenerate interval in the parametric space is not fulfilled[¶] because $p_K = 1 - \sum_{k=1}^{K-1} p_k$. We rewrite the density further to

$$\begin{aligned} f(\boldsymbol{x}) &= \exp\left(\sum_{k=1}^{K-1} x_k \log(p_k) + \left(n - \sum_{k=1}^{K-1} x_k\right) \log\left(1 - \sum_{k=1}^{K-1} p_k\right)\right) \\ &= \exp\left(\sum_{k=1}^{K-1} x_k \log\left(\frac{p_k}{1 - \sum_{j=1}^{K-1} p_j}\right)\right) \left(1 - \sum_{k=1}^{K-1} p_k\right)^n. \end{aligned}$$

Now we are allowed to use Theorem 3, and we get that a canonical parameter for this distribution is $\boldsymbol{\theta} = \left(\log\left(\frac{p_1}{p_K}\right), \dots, \log\left(\frac{p_{K-1}}{p_K}\right)\right)^\mathsf{T}$, and the complete sufficient statistic for $\boldsymbol{\theta}$ is a $(K-1)$-dimensional vector whose k-th element, $k = 1, \dots, K-1$, is the sum of the k-th elements of the vectors of the random sample.

Week 6: introduction to maximum likelihood

Example 53: Geometric distribution.

Question. *In our solution the asymptotic distribution of the MLE of p is $\mathrm{N}(0, p^2(1-p))$. There appears to be a problem with the sign of the second term in the computation of $J_n(p)$. This problem affects also the result in part (ii).*

That's true, this is a mistake in my solution. Thank you for this.

Example 55: Uniform discrete distribution.

Question. *We are able to derive the likelihood function and we see that it is decreasing. But we do not know how to formally justify that $\max X_i$ is the MLE of parameter M.*

[¶] This is also incorrect in my solutions. Think about why the complete vector $\left(\sum_{i=1}^{n} X_{i,1}, \dots, \sum_{i=1}^{n} X_{i,K}\right)^\mathsf{T}$ is not a complete statistic.

If the likelihood is not differentiable, as we see here, the MLE is typically found by simply observing the course of the (log-)likelihood. Here we see that the likelihood $L(M)$ is non-zero only for integers M that satisfy $1 \leq \min_i X_i \leq \max_i X_i \leq M$, and for these values, it is decreasing in M. We now consider both $\min_i X_i$ and $\max_i X_i$ as fixed and given. Therefore, the likelihood is non-zero only for integers $M \geq \max_i X_i$, and decreasing in M. Thus, it must be maximized in $\max_i X_i$, and this is the MLE of M.

Example 58: Weibull distribution.

Question. *In the solutions you write that the derivative of the log-likelihood approaches $-\infty$ as $\theta \to \infty$. Is this really true? What if all X_i are smaller than one?*

Yes, we need to distinguish two cases. If there is at least one $X_i > 1$, then the derivative of the log-likelihood goes to $-\infty$ as $\theta \to \infty$. If $X_i \leq 1$ for all i, then $X_i^\theta \xrightarrow[\theta\to\infty]{} 0$, but also $\sum_{i=1}^n \log(X_i) < 0$, which means that the derivative of the log-likelihood converges to a negative constant $\sum_{i=1}^n \log(X_i)$ with $\theta \to \infty$. This is still enough to see that there must be a unique root of this function, that is a unique MLE.

Week 8: Neyman-Pearson's lemma and the likelihood ration test

Example 77: Normal distribution.

Question. *How do we know in part (i) that σ^2 cannot be taken as a fixed number, and why do we search for two maximum likelihood estimators of σ^2?*

Our distribution $X_1 \sim \mathsf{N}(\mu, \sigma^2)$ has a two-dimensional unknown vector parameter $\boldsymbol{\theta} = (\mu, \sigma^2)^\mathsf{T} \in \boldsymbol{\Theta} = \left\{(a,b)^\mathsf{T} : a \in \mathbb{R} \text{ and } b > 0\right\}$. We test the hypothesis

$$H_0 \colon \mu_X = \mu_0 \text{ against } H_1 \colon \mu_X \neq \mu_0,$$

for μ_0 given. In the notation for the general likelihood ratio test on p. 30, we have that $\boldsymbol{\Theta}_0 = \left\{(a,b)^\mathsf{T} : a = \mu_0 \text{ and } b > 0\right\} \subset \boldsymbol{\Theta}$, and $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_0$.

To compute the test statistic of the likelihood ratio test in this situation, we have to first find (i) the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$ when searched over the whole parametric space $\boldsymbol{\Theta}$; and (ii) the maximum likelihood estimator $\widetilde{\boldsymbol{\theta}}_n$ when we maximize only over the set $\boldsymbol{\Theta}_0$ of parameters that satisfy H_0. The unconstrained maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$ is clearly the vector of the average $\bar{X}_n$ and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$. In the second step, we need to compute $\widetilde{\boldsymbol{\theta}}_n$, that is we need to maximize the (log-)likelihood of our data only over $\boldsymbol{\Theta}_0$. This is what I compute in the solutions, where I obtain that

$$\widetilde{\boldsymbol{\theta}}_n = \left(\mu_0, \widetilde{\sigma}^2\right)^\mathsf{T} = \left(\mu_0, \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2\right)^\mathsf{T} \in \boldsymbol{\Theta}_0.$$

Therefore, the test statistic takes the form

$$LR_n = 2\left(\ell_n\left(\widehat{\boldsymbol{\theta}}_n\right) - \ell_n\left(\widetilde{\boldsymbol{\theta}}_n\right)\right) = n \log\left(\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}\right).$$

From the theory again we know that under H_0, LR_n has asymptotically χ^2-distribution with $\dim(\boldsymbol{\Theta}) - \dim(\boldsymbol{\Theta}_0)$ degrees of freedom. Here $\dim(\boldsymbol{\Theta}) = 2$ and $\dim(\boldsymbol{\Theta}_0) = 1$, so that we reject H_0 if and only if LR_n exceeds the $(1-\alpha)$-quantile of χ^2_1.

In general, if your parameter is a vector $\boldsymbol{\theta}$, but you are interested only in a test about a sub-vector of $\boldsymbol{\theta}$, you cannot "ignore" the nuisance parameter[‖], or take it a fixed value. Because that part of the parameter is unknown, and has to be estimated. The uncertainty in the estimation of that part of the parameter has to be considered in the testing procedure. We will see this later in Section 10, where these tests will be considered in detail.

[‖] rušivý parameter