JAMES K. PRINGLE
550.621 Probability
Dr. Frangakis
Personal Assignment
May 29, 2013

Personal Assignment
2011, 2012 Comprehensive Exam questions

1   2011

1. You wish to obtain an estimator $\hat{\mu}$ of μ that is consistent for μ, in the sense that, as n grows, $\hat{\mu} \to_P \mu$. Decide which of the two estimators,

$$\frac{\sum_h P_h \bar{E}_h}{\sum_h P_h}, \text{ or } \frac{1}{n} \sum_h \bar{E}_h \tag{1}$$

   you will choose as your estimator $\hat{\mu}$ and justify your choice. From now on call $\hat{\mu}$ your chosen estimator.

   Proof. Since μ is defined as "the average per-person health-related expenditure that was spent out of pocket by the P persons last year," to me that means

$$\mu = \frac{\sum_{i=1}^{P} X_i}{P} \tag{2}$$

   where X_i is the expenditure of person i. The second estimate is clearly a weighted average, weighted towards households with smaller counts of people. The first estimate is better because in the numerator we sum $P_h \bar{E}_h$, the total expenditures in house h, over all houses canvassed. Then we divide by the total number of people we got information for. Conceptually, this is exactly the definition of μ, just taken the average of a smaller sample than the population. □

2. From now on, assume that the vectors $(P_h, \bar{E}_h), h = 1, \cdots, n$, are i.i.d. from the list $\{(P_h, \bar{E}_h), h = 1, \cdots, H\}$. Show that, under reasonable conditions, $\hat{\mu}$ is consistent for μ.

   Proof. If H is finite, then clearly as n grows, eventually it will be equal to H. In plain terms, we will poll all houses in Baltimore. The statistic, the first of (1), will equal (2), when h varies from 1 to H, assuming that everyone lives in a household. Thus you would account for everyone. □

3. Find the large sample distribution of $\sqrt{n}(\hat{\mu} - \mu)$ as n grows. Hence, find an approximate 95% confidence interval for μ.

4. Now suppose we model $\mathrm{pr}(\log(\bar{E}_h \mid P_h)$ as a normal distributions with mean $\beta_0 P_h + \beta_1 P_h^2$ and variance σ^2; and we model $\mathrm{pr}(P_h)$ as a discrete distribution that will place point mass at each observed value of P_h. Describe precisely how you would obtain the maximum likelihood estimator $\tilde{\mu}$ of μ based on these models.

5. Discuss when, if ever, you would empirically choose to use $\tilde{\mu}$ versus $\hat{\mu}$.

1.1 QUESTIONS

- How do you show consistency? I know the definition: as the sample size grows to infinity, then the statistic converges (in probability) to the true value. Perhaps in problem 2 we are to treat H as a VERY large number (infinite). In that case how could we show the statistic is consistent?

- I am having a hard time with problem 3. How to find the large sample distribution? Is $\hat{\mu}$ an MLE? How do we show that? Then what are the parameter of the population? The score for each observation would then follow how?

- For problem 4, is this Bayesian statistics, with a likelihood $f(x \mid \theta) = \mathrm{pr}(\log(\bar{E}_h) \mid P_h) \sim N(\beta_0 P_h + \beta_1 P_h^2, \sigma^2)$ and a prior given, $\pi(\theta) = \mathrm{pr}(P_h)$, is a (discrete) PMF. Then we look at the posterior,

$$\pi(\theta \mid x) = f(x \mid \theta)\pi(\theta) / \int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta \tag{3}$$

From here do we just evaluate the first of (1) using our posterior distribution? This is a little confusing for me.

- I am struggling to understand the word "empirically" in question 5.

2 2012

1. In practice (i.e. non-ideally), we only have access to the data S_i, P_i, Y_i for an i.i.d. sample from the population of patients $\{i\}$, where S_i is the *number* of comorbidities $\sum_{k=1}^{20} M_{i,k}$ of a patient i. Consider the following models for these data:

$$\mathrm{logit}\ \mathrm{pr}(Y_i = 1 \mid P_i, S_i) = \alpha' + \beta' P_i + \gamma' S_i \tag{4}$$

$$\mathrm{logit}\ \mathrm{pr}(Y_i = 1 \mid P_i, S_i = 0) = \alpha'' + \beta'' P_i \tag{5}$$

Which of the parameters, β' or β'', is better suited to answer the question of interest and why?

Proof. β' represents the change in log-odds that the patient dies within one month given the same *number* of comorbidities associated with having private health insurance. This averages patients with all different comorbidity profiles together, given they have the same number of comorbidities. This is not very close to what we want to study.

On the other hand, β'' represents the change in log-odds that a patient dies within one month given no comorbidities (this is one specific profile) associated with having private health insurance. This would give us information that we want for the profile of no comorbidities. Thus, β'' is better suited to answer the question of interest. □

2. We have approximately:

$$\hat{\beta}_{h1} \mid \beta_{h1} \sim N(\beta_{h1}, V_{h1}) \text{ and, independently, } \hat{\beta}_{h2} \mid \beta_{h2} \sim N(\beta_{h2}, V_{h2}) \tag{6}$$

where we assume V_{h1}, V_{h2} are known. By treating $\hat{\beta}_{h1}, \hat{\beta}_{h2}$ now as the only available data and (6) as their likelihoods, and assuming that both β_{h1} and β_{h2} equal to a common value, say β, find the MLE of β.

Proof. We have

$$f(\hat{\beta}_{h1}, \hat{\beta}_{h2}) = \varphi_{\beta, V_{h1}}(\hat{\beta}_{h1})\varphi_{\beta, V_{h2}}(\hat{\beta}_{h2}) = \frac{1}{\sqrt{2\pi V_{h1}}} \exp\left\{\frac{-(\hat{\beta}_{h1} - \beta)^2}{2V_{h1}}\right\} \frac{1}{\sqrt{2\pi V_{h2}}} \exp\left\{\frac{-(\hat{\beta}_{h2} - \beta)^2}{2V_{h2}}\right\} \tag{7}$$

Taking the logarithm, we get

$$\ell = \log(f) = \frac{-1}{2}\log(2\pi V_{h1}) - \frac{(\hat{\beta}_{h1} - \beta)^2}{2V_{h1}} - \frac{1}{2}\log(2\pi V_{h2}) - \frac{(\hat{\beta}_{h2} - \beta)^2}{2V_{h2}} \tag{8}$$

Taking the derivative with respect to β, we get

$$\frac{d\ell}{d\beta} = \frac{\hat{\beta}_{h1} - \beta}{V_{h1}} + \frac{\hat{\beta}_{h2} - \beta}{V_{h2}} \tag{9}$$

Setting that to 0 and solving for β, we get

$$\beta = \frac{V_{h2}\hat{\beta}_{h1} + V_{h1}\hat{\beta}_{h2}}{V_{h1} + V_{h2}} \tag{10}$$

□

3. Suppose that based on the large sample distribution of the MLE estimators, we have approximately:

$$\begin{pmatrix}\hat{\beta}_{1mo} \\ \hat{\beta}_{2mo}\end{pmatrix} \middle| \begin{pmatrix}\beta_{1mo} \\ \beta_{2mo}\end{pmatrix} \sim N\left(\begin{pmatrix}\beta_{1mo} \\ \beta_{2mo}\end{pmatrix}, \Sigma\right) \tag{11}$$

where we assume Σ known. Explain why Σ will generally not be diagonal.

Proof. The variance-covariance matrix Σ will not be diagonal if and only if the covariance of β_{1mo} and β_{2mo} is not zero. In general the variance of β_{1mo} and β_{2mo} will be non-zero. Clearly, β_{2mo} depends on β_{1mo}: if you are dead before the first month, you will be dead before the second month. The same reasons that you have a good chance to survive one month may keep you alive for the second month. $\square$

4. By treating β_{1mo}, β_{2mo} now as the only available data and (11) as their likelihood for β_{com}, suppose we find the MLE of β_{com}, say $\widehat{\beta_{com}}$. Can $\widehat{\beta_{com}}$ be strictly outside the interval defined by the individual MLEs $\hat{\beta}_{1mo}, \hat{\beta}_{2mo}$? Justify your answer by either a figure or by considering simple specific values for $\hat{\beta}_{1mo}, \hat{\beta}_{2mo}$ and Σ or otherwise.

Proof. We know from Cauchy-Schwarz

$$\mathrm{cov}(X,Y)^2 \leq \mathrm{var}(X)\mathrm{var}(Y) \tag{12}$$

This will be helpful when finding a counterexample. Denote Σ as

$$\Sigma = \begin{bmatrix} x & p \\ p & y \end{bmatrix} \tag{13}$$

Then it follows that Σ^{-1} as

$$\Sigma^{-1} = |\Sigma|^{-1} \begin{bmatrix} y & -p \\ -p & x \end{bmatrix} \tag{14}$$

Given (11),

$$f(\hat{\beta}_{1mo}, \hat{\beta}_{2mo}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix}' |\Sigma|^{-1} \begin{bmatrix} y & -p \\ -p & x \end{bmatrix} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix} \right\} \tag{15}$$

Next we take the logarithm then the derivative with respect to β. Notice that only the expression in the exponent of e depend on β, so we can jump to

$$\frac{d\log(f)}{d\beta} = \frac{d\ell}{d\beta} = \frac{d}{d\beta}\left(-\frac{1}{2} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix}' |\Sigma|^{-1} \begin{bmatrix} y & -p \\ -p & x \end{bmatrix} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix} \right) \tag{16}$$

Setting $\frac{d\ell}{d\beta}$ to 0 and solving for β, we get

$$0 = \frac{d}{d\beta}\left(-\frac{1}{2} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix}' |\Sigma|^{-1} \begin{bmatrix} y & -p \\ -p & x \end{bmatrix} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix} \right) \tag{17}$$

$$0 = \frac{d}{d\beta}\left(\begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix}' \begin{bmatrix} y & -p \\ -p & x \end{bmatrix} \begin{pmatrix} \hat{\beta}_{1mo} - \beta \\ \hat{\beta}_{2mo} - \beta \end{pmatrix} \right) \tag{18}$$

$$\beta = \frac{-\frac{(\hat{\beta}_{1mo}+\hat{\beta}_{2mo})}{2}2p + x\hat{\beta}_{2mo} + y\hat{\beta}_{1mo}}{-2p + x + y} \tag{19}$$

Hence β is the MLE $\widehat{\beta_{com}}$. (NOTE: Calculations for (19) were completed on paper and with SAGE. They are not reproduced here due to the tedious nature of the algebra). We now seek a counterexample. Let $\hat{\beta}_{1mo} \leq \hat{\beta}_{2mo}$. We choose our variables such that

$$-\frac{\left(\hat{\beta}_{1mo} + \hat{\beta}_{2mo}\right)}{2}2p + x\hat{\beta}_{2mo} \leq -\hat{\beta}_{1mo}2p + x\hat{\beta}_{1mo} \tag{20}$$

Say, $y = 10$, $p = 1.7$, and $x = 0.4$, and therefore,

$$\Sigma = \begin{bmatrix} 0.4 & 1.7 \\ 1.7 & 10 \end{bmatrix} \tag{21}$$

This satisifies Cauchy-Schwarz. Let $\hat{\beta}_{1mo} = 0$ and $\hat{\beta}_{2mo} = 1$, then by (19), $\widehat{\beta_{com}} = -0.1857$. Hence $\widehat{\beta_{com}}$ is outside of the range defined by $[\hat{\beta}_{1mo}, \hat{\beta}_{2mo}]$ $\square$