CSI771 / Stat 751 Midterm
October 24, 2005
Comments/Solutions

The exam was open book and open notes. In attempting to answer a question, you should write down formulas, but of course in this kind of exam, you usually cannot expect to get any credit for merely writing a formula from the book.

1. *(35) Let $Y_1, \ldots, Y_n$ be a random sample from a distribution with CDF P that has two unknown parameters θ and λ. We can write θ as $\Theta(P)$ and λ as $\Lambda(P)$.*

   Our interest is in using the sample to make inferences about θ. There are various ways of doing this.

   In each of the following questions, do not forget also to tell what to do about λ.

   It is often a good idea to think of a concrete example. An example that is very similar to this is a normal distribution with unknown parameters μ and σ^2, in which we want to make inferences about μ without worrying about σ^2. In this case, σ^2 is a nuisance parameter as is λ in the questions below.

   (a) *How would you estimate θ using the plug-in principle?*

   Determine the ECDF P_n, and then compute $\Theta(P_n)$. If the functional is free of λ, we do not need to consider λ. If the functional $\Theta(P_n)$ is not free of λ, we also form the expression $\Lambda(P_n)$, and evaluate the two expressions simultaneously.

   (b) *How would you estimate θ using maximum likelihood?*

   This requires the likelihood function, which is formally the same as the probability density or the probability function, which we can write in a generalized form as $\mathrm{d}P$, with the role of the value of the random variable interchanged with that of the parameters; i.e., the parameters are the variables, for given realizations of the random variables:

   $$L_n(t, l; y_1, \ldots, y_n).$$

   We maximize $L_n(t, l; y_1, \ldots, y_n)$ with respect to both t and l. The value of t at the maximum is the MLE of θ.

   An alternate approach would be to compute an estimate of λ, say $\hat{\lambda}$ (perhaps as a plug-in estimator) and then maximize the conditional likelihood, $L_n(t; \hat{\lambda}, y_1, \ldots, y_n)$.

   (Notice I use t and l in place of θ and λ to emphasize that these are variables instead of fixed parameters.)

   I counted off if instead of maximization, differentiation was stated as the step to perform. I counted off if both maximization and differentiation were mentioned, but no qualifications were put on the differentiation step.

   (c) *How would you estimate θ using least squares?*

   To use least squares, we must form a predictive model that involves θ, and possibly λ, for the observable variables or some function of the observable variables. The predicted values are a function of θ and λ, for which we substitute the variables t and l. Then we form the sum of the squares of the differences between what the model predicts (as a function of t and l) and what is observed, and we minimize this with respect to t and l. The value of t at the minimum is the LS estimate of θ.

   Exactly how this these residuals are formed is very much dependent on the problem. The simplest case is when θ is the mean (in which case, λ is not a part of the problem). In that simple case, we minimize $\sum(y_i - t)^2$ with respect to t and that value of t is the LS estimate of θ.

   An alternate approach if λ is in the residuals would be to compute an estimate of λ, say $\hat{\lambda}$ (perhaps as a plug-in estimator) and then minimize the conditional sum of squared residuals, $s(t; \hat{\lambda}, y_1, \ldots, y_n)$.

(d) *How would you form a 95% confidence set for θ? (Specify all of the things needed to do this.)*

In general, a 95% confidence set for θ is random set, C such that $\Pr(C \ni \theta) = .95$. In this case, C must not involve λ; otherwise the confidence set is conditional.

Although the question did not describe θ, for simplicity let us assume it is a real scalar. The question also did not state what kind of confidence set to form. Let us assume we want a two-sided interval. (*Any other reasonable assumptions you made would be OK.*) We will also assume we want a confidence interval with equal probabilities on either side of the interval.

The standard way to form a confidence interval for a scalar parameter is to find a pivotal function that

- involves only the parameter of interest and a statistic
- allows us to solve for parameter of interest
- has a known distribution

Let $f(T(Y), \theta)$ be such a pivotal function. $f(T(Y), \theta)$ is a random variable, and we assume we know its distribution. Let $f_{0.025}$ be such that $\Pr(f(T(Y), \theta) < f_{0.025}) = 0.025$, and $f_{0.975}$ be such that $\Pr(f(T(Y), \theta) > f_{0.975}) = 0.975$. (In a specific situation, this may require a randomization procedure.) A 95% confidence interval for θ is form by first writing

$$\Pr(f_{0.025} \leq f(T(Y), \theta) \leq f_{0.975}) = 0.95$$

and then solving for θ to get

$$\Pr(g(T(Y), f_{0.025}) \leq \theta \leq g(T(Y), f_{0.975})) = 0.95$$

The confidence interval is $(g(T(Y), f_{0.025}),\ g(T(Y), f_{0.975}))$. (Think of a concrete example, say a two-sided confidence for μ in a normal distribution with unknown parameters μ and σ^2. Recall how to form a pivotal function that has a Student's t distribution.)

(e) *In the preceding question, you may not have everything that is needed. Describe one way you could use the bootstrap to form a 95% confidence set for θ.*

The main thing we may not have that we need is the distribution of $f(T(Y), \theta)$. There are various bootstrap confidence intervals we could form, and this question and the next one are to describe a couple of ways. The actual specifics may depend on the form of θ and $f(T(Y), \theta)$. In general, we will use bootstrap samples to estimate the necessary quantiles of $f(T(Y), \theta)$.

The simplest is the percentile confidence interval given at the bottom of page 90.

For the next question, the simplest answer would be the BC_α method described on page 92. Alternatively, you could use other intervals described on pages 89 through 93. For some you may have to make more assumptions about the form of θ and $f(T(Y), \theta)$.

Note that confidence intervals formed in this way are "estimated" confidence intervals, not "approximate" confidence intervals. (We often use approximate confidence intervals, based on some approximate distribution, perhaps an asymptotic distribution.)

(f) *Describe another way (different from your answer in the preceding question) to use the bootstrap to form a 95% confidence set for θ.*

See above.

(Continued. Use all the definitions on the previous page.)

(g) *How would you do a Monte Carlo test, at the 0.05 level, of the null hypothesis that $\theta = 100$ against the alternative that $\theta \neq 100$? (Don't forget about λ.)*

We first need a test statistic, say T, whose distribution depends on θ in some predictable way.

This requires either that we know something about the role that λ plays in the distribution of the test statistic. If the distribution does not depend on λ (for example, if λ is a scale parameter and our test statistics is a ratio), λ can be given an arbitrary value. Otherwise, our test is conditional on a fixed value of λ.

In either case we simulate m samples of size n from the distribution with $\theta = 100$ (and λ at a fixed value) and compute the values of T for each. Call these values t_j^* and rank them, $t_{(1)}^* \leq \cdots \leq t_{(m)}^*$. Now compute the value of T based on the given sample; call this t_0. If and only if t_0 is less than the 0.025 empirical quantile of the t^*'s or greater than the 0.975 empirical quantile of the t^*'s, we reject the null hypothesis.

(h) *Suppose you have a statistic $T(Y)$ to use as an estimator of θ. You need the variance of $T(Y)$. Describe the steps to use the jackknife to estimate the variance of $T(Y)$.*

I will describe the leave-out-one jackknife because that is the most commonly used one.

We form the n statistics $T_{(-j)}$ and their mean

$$\overline{T}_{(\bullet)} = \frac{1}{n}\sum_{j=1}^{n} T_{(-j)}.$$

We also form the n pseudovalues

$$T_j^* = nT - (n-1)T_{(-j)}.$$

The bootstrap estimator of θ is

$$\mathrm{J}(T) = nT - (n-1)\overline{T}_{(\bullet)},$$

and the sample variance of the mean of the pseudovalues is the jackknife estimator of $\mathrm{V}(T)$:

$$\widehat{\mathrm{V}(T)}_{\mathrm{J}} = \frac{\sum_{j=1}^{n}(T_j^* - \mathrm{J}(T))^2}{n(n-1)}.$$

(i) *Describe the steps to use the bootstrap to estimate the variance of $T(Y)$.*

We generate m bootstrap samples from the given data, and compute T^* for each. Call these T^{*j}. The bootstrap estimate of the variance is the sample variance of T^* based on the m samples of size n taken from P_n:

$$\begin{aligned} \widehat{\mathrm{V}}(T) &= \widehat{\mathrm{V}}(T^*) \\ &= \frac{1}{m-1}\sum(T^{*j} - \overline{T}^*)^2. \end{aligned}$$

(j) *Suppose $\Theta(P) = \theta$ is such that*

$$\int_{-\infty}^{\theta} \mathrm{d}\,P(t) = 0.5.$$

What is the plug-ing estimate of θ?

It is the sample median. You might use some slight variation, such as described on page 14, but the median is OK.

2. *(15) Consider the problem of estimating θ in a uniform distribution over $[0, \theta]$. From a sample of size n, an estimator is $T = x_{(n)}$, the maximum order statistic.*

We assume $\theta \neq 0$.

(a) *Show that T is biased.*

The PDF of T is

$$\frac{n}{\theta^n} t^{n-1} \quad \text{for } 0 \leq t \leq \theta.$$

(This is a beta distribution.)

$$\begin{aligned} \mathrm{E}(T) &= \int_0^\theta \frac{n}{n\theta^n} t^n \\ &= \frac{n}{n+1}\theta \\ &\neq \theta. \end{aligned}$$

(b) *Derive (describe) the nonparametric bootstrap bias-corrected estimator for θ.*

What we need is $\mathrm{E}\Big(T(P_n^{(1)}) \;\Big|\; P_n\Big)$. This can be worked out analytically based on the distribution of the max order statistic in a discrete uniform distribution over the mass points of the given sample. Another way is to generate m bootstrap samples of size n from P_n and compute the mean of the maximum order statistics. This is an estimate of the quantity $\mathrm{E}\Big(T(P_n^{(1)}) \;\Big|\; P_n\Big)$.

Whether we use the analytic value or the estimated value, the bootstrap bias-corrected estimate of θ is

$$2x_{(n)} - \mathrm{E}\Big(T(P_n^{(1)}) \;\Big|\; P_n\Big).$$

3. *(15) Assume you have a source of uniform* U$(0,1)$ *random numbers and you want to generate random numbers from the distribution with density*

$$\begin{aligned} p_X(x) &= \frac{3}{2}x^2, \quad \text{for } -1 \le x \le 1, \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

(This is a type of beta distribution.)

(a) *Write out the exact steps you would use to do this by the inverse CDF method.*

The CDF is

$$\begin{aligned} P_X(x) &= 0, \quad \text{for } x \le -1, \\ &= \frac{1}{2} + \frac{1}{2}x^3, \quad \text{for } -1 \le x \le 1, \\ &= 1, \quad \text{otherwise.} \end{aligned}$$

(Note that the inverse only exists over the support of the distribution, and of course that is the only place we are interested in.)

1. generate u from U$(0,1)$.
2. deliver $x = (2u-1)^{1/3}$.

(b) *Write out the exact steps you would use to do this by the acceptance/rejection method. Clearly identify any function or constant that you use.*

First, choose a simple distribution from which to form a majorizing function. For simplicity, I will use the uniform over $(-1,1)$. So $g_Y(y) = 1/2$ over $(-1,1)$, and $cg_Y(y) = 3/2$.

1. generate u_1 from $U(0,1)$, and form $y = 2u_1 - 1$. (This means y is from g_Y.)
2. generate u_2 from $U(0,1)$.
3. If $u_2 \le y^2$, take y as the realization; otherwise reject it, and go back to step 1.

4. *(15) Describe how you would evaluate these integrals using Monte Carlo. (Give the formulas, and tell what kinds of variables you use.)*

(a) $\int_{-1}^{1} \cos(x)x^2 \mathrm{d}x$

A simple solution is

$$2\sum_{i=1}^{m} \cos(u_i)u_i^2/m,$$

where the u_i are iid U$(-1,1)$.

A better solution is

$$\frac{2}{3}\sum_{i=1}^{m} \cos(x_i)/m,$$

where the x_i are iid from the distribution in the previous question.

(b) $\int_0^2 \int_0^\infty y\cos(x)e^{-x}\mathrm{d}x\,\mathrm{d}y$

The important part of this is the improper inner integral. To evaluate such an integral, you can generate $x_1,\ldots,x_m$ from the standard exponential distribution (do this by taking $x_i = -\log(u_i)$, where u_i are iid $\mathrm{U}(0,1)$).

The integral as written is should be evaluated by Monte Carlo only after some analytic reduction:

$$\begin{aligned}\int_0^2 \int_0^\infty y\cos(x)e^{-x}\mathrm{d}x\,\mathrm{d}y &= \int_0^2 y\left(\int_0^\infty \cos(x)e^{-x}\mathrm{d}x\right)\mathrm{d}y \\ &= 2\int_0^\infty \cos(x)e^{-x}\mathrm{d}x\end{aligned}$$

The Monte Carlo estimate is

$$2\sum_{i=1}^m \cos(x_i)/m,$$

where the x_i are iid from a standard exponential distribution.

The integral that I had meant for this question is

$$\int_0^2 \int_0^\infty y\cos(xy)e^{-x}\mathrm{d}x\,\mathrm{d}y.$$

For this you would generate $x_1,\ldots,x_m$ from the standard exponential distribution and generate $y_1,\ldots,y_m$ from $\mathrm{U}(0,1)$.

Then the MC estimate is

$$2\sum_{i=1}^m y_i\cos(x_iy_i)/m.$$

5. *(20) Given the vector* $x = (2, 3, 4)$.

(a) *What is the matrix of the transformation that will rotate* x *into the vector* $(y_1, y_2, 0)$*?*

We can rotate the vector within the e_1, e_3 plane or within the e_2, e_3 plane.

Within the e_1, e_3 plane, the rotation matrix has the form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & s \\ 0 & -s & c \end{bmatrix},$$

where $c^2 + s^2 = 1$. Then writing

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & s \\ 0 & -s & c \end{bmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ 0 \end{pmatrix},$$

we arrive at the additional equation $3s = 4c$. Solving these two equations in two unknowns, we get $c = 3/5$ and $s = 4/5$.

(b) *What are* y_1 *and* y_2*? (Give their numeric values.)*

Substituting and multiplying we get

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3/5 & 4/5 \\ 0 & -4/5 & 3/5 \end{bmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 0 \end{pmatrix}.$$

(c) *Given the vector* $z = (1, 1, 1)$*. Determine a transformation of* z *that makes it orthogonal to the vector* $(y_1, y_2, 0)$ *above.*

There are many transformations of z that will make it orthogonal to $y = (y_1, y_2, 0)$. A vector $\tilde{z}$ that is orthogonal to y must be such that $\tilde{z}_1 y_1 = -\tilde{z}_2 y_2$; hence, the simplest transformation is

$$\tilde{z} = \begin{bmatrix} -y_2 & 0 & 0 \\ 0 & y_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} z.$$

We could also use the Gram-Schmidt transformation, either following a normalization of y or not. (If a vector is orthogonal to a normalized vector, it is orthogonal to the unnormalized vector.)

An unnormalized Gram-Schmidt transformation is

$$\begin{aligned} \tilde{z} &= z - y^{\mathrm{T}} z y / y^{\mathrm{T}} y \\ &= z - \frac{y_1 + y_2}{y_1^2 + y_2^2} y. \end{aligned}$$

(d) *Is your transformation in Part (c) linear? orthogonal? isometric? affine? projective? (The answers may depend on the specific transformation you use.)* The transformations given above are linear and affine. They are not orthogonal, isometric, or projective.

We could make them both orthogonal and isometric by normalizing the matrices of the transformations.

We could make a projective transformation by forming $\tilde{z}$ as $(-y_2, y_1, 0)$ (using a rotation).

(e) *What is the Euclidean distance between* x *and* z *above?*

$\sum_{i=1}^{3}(x_i - z_i)^2 = 14$.