TAKEHOME EXAM STAT 6869 SPRING 2000

ECONOMICS DEPARTMENT, UNIVERSITY OF UTAH

You are allowed and encouraged to cooperate while working on this exam. You may submit solutions with more than one name on them, which will count equally for all authors. But you must understand the solution you are handing in. I will perhaps ask you to demonstrate and explain the solutions in class. The exam is due back at the beginning of class on Wednesday, March 29, 2000, at the beginning of class at 3:10 pm.

Problem 16. *Also the Box-Muller transformation: two uniform variables into a bivariante normal, Hogg and Tanis example 4.8.4 p. 261. It is too complicated for a test, perhaps a takehome?*

Date of exam March 22, 2000.

Problem 17. *4 points* *Assume $t_1, \dots, t_n$ are independent random variables with the exponential distribution with the same parameter λ, i.e., each t_i has the density $f_{t_i}(t) = \lambda e^{-\lambda t}$ for $t \geq 0$, and 0 for $t < 0$.*

• **a.** *Use this density (and no other information about the expenential distribution) to show that the expected value and the variance of t are* $\mathrm{E}[t] = 1/\lambda$ *and* $\mathrm{var}[t] = 1/\lambda^2$.

Answer. $\mathrm{E}[t] = \int_0^\infty \lambda t e^{-\lambda t} dt = \int_0^\infty u v' dt = uv\Big|_0^\infty - \int_0^\infty u' v dt$, where $\begin{array}{ll} u = t & v' = \lambda e^{-\lambda t} \\ u' = 1 & v = -e^{-\lambda t} \end{array}$. Therefore $\mathrm{E}[t] = -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t} dt = (-1/\lambda) e^{-\lambda t}|_0^\infty = \frac{1}{\lambda}$.

Alternative by differentiation under the integral sign: from $\int_0^\infty \lambda e^{-\lambda t} dt = 1$ follows $\int_0^\infty e^{-\lambda t} dt = 1/\lambda$, now differentiating with respect to λ gives $\int_0^\infty -te^{-\lambda t} dt = -1/\lambda^2$, therefore $\mathrm{E}[t] = \int_0^\infty t \lambda e^{-\lambda t} dt = 1/\lambda$.

BTW, Differentiating the above one more time gives $\int_0^\infty (-t)^2 e^{-\lambda t} dt = 2/\lambda^3$, therefore $\mathrm{E}[t^2] = \int_0^\infty t^2 \lambda e^{-\lambda t} dt = 2/\lambda^2$.

For $\mathrm{E}[t^2]$ one can use that $\Gamma(r) = \int_0^\infty \lambda^r t^{r-1} e^{-\lambda t}\, dt$ for $r = 3$ to get: $\mathrm{E}[t^2] = (1/\lambda^2)\Gamma(3) = 2/\lambda^2$. Or all from scratch: $\mathrm{E}[t^2] = \int_0^\infty \lambda t^2 e^{-\lambda t} dt = \int_0^\infty u v' dt = uv\Big|_0^\infty - \int_0^\infty u' v dt$, where $\begin{array}{ll} u = t^2 & v' = \lambda e^{-\lambda t} \\ u' = 2t & v = -e^{-\lambda t} \end{array}$. Therefore $\mathrm{E}[t^2] = -t^2 e^{-\lambda t}\Big|_0^\infty + \int_0^\infty 2te^{-\lambda t}\, dt$. The first term vanishes, for the second do it again: $\int_0^\infty 2te^{-\lambda t}\, dt = \int_0^\infty u v' dt = uv\Big|_0^\infty - \int_0^\infty u' v dt$, where $\begin{array}{ll} u = t & v' = e^{-\lambda t} \\ u' = 1 & v = -(1/\lambda) e^{-\lambda t} \end{array}$. Therefore the second term becomes $2(t/\lambda) e^{-\lambda t}\Big|_0^\infty + 2\int_0^\infty (1/\lambda) e^{-\lambda t} dt = 2/\lambda^2$. $\square$

• **b.** *Show that the minimum* MSE *linear estimator of* $1/\lambda$, *i.e., the estimator which has minimum* MSE *among all estimators of the form* $s = a_1 t_1 + \cdots + a_n t_n$, *is* $s = \frac{1}{n+1}\sum t_i$. *Derive* $\text{MSE}[s; 1/\lambda]$.

Answer. For a general s we have

$$\text{MSE}[s;1/\lambda] = \text{var}[s] + (\text{E}[s - \frac{1}{\lambda}])^2 = \frac{1}{\lambda^2}\Big(\sum a_i^2 + (\sum a_j - 1)^2\Big). \tag{1}$$

Now call $a_{n+1} = 1 - \sum a_j$; then minimizing this MSE is minimizing $\sum_{i=1}^{n+1} a_i^2$ subject to $\sum_{j=1}^{n+1} a_j = 1$. The solution of this is $a_i = 1/(n+1)$. Therefore the optimal estimator is $s = \frac{1}{n+1}\sum t_i$, and $\text{MSE}[s; 1/\lambda] = \frac{1}{(n+1)\lambda^2}$ (one can get this as the minimum value of the above objective function, or by adding variance plus squared bias $\frac{n}{(n+1)^2\lambda^2} + \frac{1}{(n+1)^2\lambda^2}$. Its expected value is $\text{E}[s] = \frac{n}{(n+1)\lambda}$, therefore its bias is $\text{E}[s] - \frac{1}{\lambda} = \frac{-1}{(n+1)\lambda}$ and if we call $\theta = 1/\lambda$ then $\frac{\partial}{\partial\theta}\text{E}_\theta[s] = \frac{n}{n+1}$, □

• **c.** *Write out the joint distribution of* $\boldsymbol{t}$ *and, calling* $\theta = 1/\lambda$, *compute the* score $q = \frac{\partial}{\partial\theta}\log f_{\boldsymbol{t}}(\boldsymbol{t})$ *and show that* $\text{E}[q] = 0$ *and* $\text{var}[q] = n/\lambda^2$.

Answer. The joint density is

$$f_{\boldsymbol{t}}(\boldsymbol{t}) = \lambda^n e^{-\lambda(t_1+\cdots+t_n)} = \theta^{-n} e^{-\frac{t_1+\cdots+t_n}{\theta}} \tag{2}$$

$$\log f_{\boldsymbol{t}}(\boldsymbol{t}) = -n\log\theta - \frac{t_1 + \cdots + t_n}{\theta} \tag{3}$$

$$q(ravect;\theta) = \frac{\partial}{\partial\theta}\log f_{\boldsymbol{t}}(\boldsymbol{t};\theta) = -\frac{n}{\theta} + \frac{t_1+\cdots+t_n}{\theta^2} \tag{4}$$

Clearly, $\mathrm{E}[q] = 0$, therefore $\mathrm{E}[q^2] = \mathrm{var}[q] = \frac{n}{\theta^4}\,\mathrm{var}[t_i] = n/\theta^2$. Therefore the biased CR inequality for the above s would be

$$\mathrm{MSE}[s; 1/\lambda] = \frac{1}{(n+1)\lambda^2} \geq \frac{1}{(n+1)\lambda^2} + \frac{1}{(n+1)^2\lambda^2} \tag{5}$$

There is still something wrong, look at it again! □

Problem 18. [BD77, Problem 4.2.6 on p. 144] *If a statistic* $\boldsymbol{s}$ *is* complete *and* sufficient, *then every function* $g(\boldsymbol{s})$ *is the minimum* MSE *unbiased estimator of* $\mathrm{E}[g(\boldsymbol{s})]$ *(*Lehmann-Scheffé theorem*). This gives a systematic approach to finding minimum* MSE *unbiased estimators. Here are the definitions:* $\boldsymbol{s}$ *is* sufficient *for* θ *if for any event* E *and any value* $\boldsymbol{s}$*, the conditional probability* $\Pr[E|\boldsymbol{s} \leq \boldsymbol{s}]$ *does not involve* θ*.* $\boldsymbol{s}$ *is* complete *for* θ *if the only function* $g(\boldsymbol{s})$ *of* $\boldsymbol{s}$*, which has zero expected value whatever the value of* θ*, is the function which is identically zero, i.e.,* $g(\boldsymbol{s}) = 0$ *for all* $\boldsymbol{s}$*.*

• **a.** *3 points Given an unknown parameter* θ*, and a complete sufficient statistic* $\boldsymbol{s}$*, how can one find that function of* $\boldsymbol{s}$ *whose expected value is* θ*? There is an easy trick: start with any statistic* p *with* $\mathrm{E}[p] = \theta$*, and use the conditional expectation* $\mathrm{E}[p|\boldsymbol{s}]$*. Argue why this conditional expectation does not depend on the unknown parameter* θ*, is an unbiased estimator of* θ*, and why this leads to the same estimate regardless which* p *one starts with.*

Answer. You need sufficiency for the first part of the problem, the law of iterated expectations for the second, and completeness for the third.

Set $E = \{p \le p\}$ in the definition of sufficiency given at the beginning of the Problem to see that the cdf of p conditionally on s being in any interval does not involve θ, therefore also $\mathrm{E}[p|s]$ does not involve θ.

Unbiasedness follows from the theorem of iterated expectations $\mathrm{E}\big[\mathrm{E}[p|s]\big] = \mathrm{E}[p] = \theta$.

The independence on the choice of p can be shown as follows: Since the conditional expectation conditionally on s is a function of s, we can use the notation $\mathrm{E}[p|s] = g_1(s)$ and $\mathrm{E}[q|s] = g_2(s)$. From $\mathrm{E}[p] = \mathrm{E}[q]$ follows by the law of iterated expectations $\mathrm{E}[g_1(s) - g_2(s)] = 0$, therefore by completeness $g_1(s) - g_2(s) \equiv 0$. □

• **b.** *2 points Assume $y_i \sim \mathrm{NID}(\mu, 1)$ $(i = 1, \dots, n)$, i.e., they are independent and normally distributed with mean μ and variance 1. Without proof you are allowed to use the fact that in this case, the sample mean $\bar{y}$ is a complete sufficient statistic for μ. What is the minimum* MSE *unbiased estimate of μ, and what is that of μ^2?*

Answer. We have to find functions of $\bar{y}$ with the desired parameters as expected values. Clearly, $\bar{y}$ is that of μ, and $\bar{y}^2 - 1/n$ is that of μ^2. □

• **c.** *1 point For a given j, let π be the probability that the j^{th} observation is nonnegative, i.e., $\pi = \Pr[y_j \ge 0]$. Show that $\pi = \Phi(\mu)$ where Φ is the cumulative distribution function of the standard normal. The purpose of the remainder of this Problem is to find a minimum* MSE *unbiased estimator of π.*

Answer.

$$\pi = \Pr[y_i \geq 0] = \Pr[y_i - \mu \geq -\mu] = \Pr[y_i - \mu \leq \mu] = \Phi(\mu) \tag{6}$$

because $y_i - \mu \sim N(0,1)$. We needed symmetry of the distribution to flip the sign. □

• **d.** *1 point As a first step we have to find an unbiased estimator of π. It does not have to be a good one, any ubiased estimator will do. And such an estimator is indeed implicit in the definition of π. Let q be the "indicator function" for nonnegative values, satisfying $q(y) = 1$ if $y \geq 0$ and 0 otherwise. We will be working with the random variable which one obtains by inserting the j^{th} observation y_j into q, i.e., with $q = q(y_j)$. Show that q is an unbiased estimator of π.*

Answer. $q(y_j)$ has a discrete distribution and $\Pr[q(y_j) = 1] = \Pr[y_j \geq 0] = \pi$ by (6) and therefore $\Pr[q(y_j) = 0] = 1 - \pi$

The expected value is $\mathrm{E}[q(y_j)] = (1-\pi) \cdot 0 + \pi \cdot 1 = \pi$. □

• **e.** *2 points Given q we can apply the Lehmann-Scheffé theorem: $\mathrm{E}[q(y_j)|\bar{y}]$ is the best unbiased estimator of π. We will compute $\mathrm{E}[q(y_j)|\bar{y}]$ in four steps which build on each other. First step: since for every indicator function follows $\mathrm{E}[q(y_j)|\bar{y}] = \Pr[y_j \geq 0|\bar{y}]$, we need for every given value $\bar{y}$, the conditional distribution of y_j conditionally on $\bar{y} = \bar{y}$. (Not just the conditional mean but the whole conditional distribution.) In order to construct this, we first have to specify exactly the joint distribution of y_j and $\bar{y}$:*

Answer. They are jointly normal:

$$\begin{bmatrix} y_j \\ \bar{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & 1/n \\ 1/n & 1/n \end{bmatrix} \right) \tag{7}$$

□

• **f.** *2 points Second step: From this joint distribution derive the conditional distribution of y_j conditionally on $\bar{y} = \bar{y}$. (Not just the conditional mean but the whole conditional distribution.) For this you will need formula* (**??**) *and* (**??**).

Answer. Here are these two formulas: if u and v are jointly normal, then the conditional distribution of v conditionally on $u = u$ is Normal with mean

$$\mathrm{E}[v|u = u] = \mathrm{E}[v] + \frac{\operatorname{cov}[u, v]}{\operatorname{var}[u]}(u - \mathrm{E}[u]) \tag{8}$$

and variance

$$\operatorname{var}[v|u = u] = \operatorname{var}[v] - \frac{(\operatorname{cov}[u, v])^2}{\operatorname{var}[u]}. \tag{9}$$

Plugging $u = \bar{y}$ and $v = y_j$ into (**??**) and (**??**) gives: the conditional distribution of y_j conditionally on $\bar{y} = \bar{y}$ has mean

$$\mathrm{E}[y_j|\bar{y} = \bar{y}] = \mathrm{E}[y_j] + \frac{\operatorname{cov}[\bar{y}, y_j]}{\operatorname{var}[\bar{y}]}(\bar{y} - \mathrm{E}[\bar{y}]) \tag{10}$$

$$= \mu + \frac{1/n}{1/n}(\bar{y} - \mu) = \bar{y} \tag{11}$$

and variance

$$\operatorname{var}[y_j|\bar{y} = \bar{y}] = \operatorname{var}[y_j] - \frac{(\operatorname{cov}[\bar{y}, y_j])^2}{\operatorname{var}[\bar{y}]} \tag{12}$$

$$= 1 - \frac{(1/n)^2}{1/n} = 1 - \frac{1}{n}. \tag{13}$$

Therefore the conditional distribution of y_j conditional on $\bar{y}$ is $N(\bar{y}, (n-1)/n)$. How can this be motivated? if we know the actual arithmetic mean of the variables, then our best estimate is that each variable is equal to this arithmetic mean. And this additional knowledge cuts down the variance by $1/n$. □

• **g.** *2 points The variance decomposition* (**??**) *gives a decomposition of* $\operatorname{var}[y_j]$*: give it here:*

Answer.

$$\operatorname{var}[y_j] = \operatorname{var}\Big[\operatorname{E}[y_j|\bar{y}]\Big] + \operatorname{E}\Big[\operatorname{var}[y_j|\bar{y}]\Big] \tag{14}$$

$$= \operatorname{var}[\bar{y}] + \operatorname{E}\left[\frac{n-1}{n}\right] = \frac{1}{n} + \frac{n-1}{n} \tag{15}$$

□

• **h.** *Compare the conditional with the unconditional distribution.*

Answer. Conditional distribution does not depend on unknown parameters, and it has smaller variance! □

• **i.** *2 points Third step: Compute the probability, conditionally on* $\bar{y} = \bar{y}$*, that* $y_j \geq 0$.

Answer. If $x \sim N(\bar{y}, (n-1)/n)$ (I call it x here instead of y_j since we use it not with its familiar unconditional distribution $N(\mu, 1)$ but with a conditional distribution), then $\Pr[x \geq 0] = \Pr[x - \bar{y} \geq -\bar{y}] = \Pr[x - \bar{y} \leq \bar{y}] = \Pr\left[(x - \bar{y})\sqrt{n/(n-1)} \leq \bar{y}\sqrt{n/(n-1)}\right] = \Phi(\bar{y}\sqrt{n/(n-1)})$ because $(x - \bar{y})\sqrt{n/(n-1)} \sim N(0, 1)$ conditionally on $\bar{y}$. Again we needed symmetry of the distribution to flip the sign. □

• **j.** *1 point Finally, put all the pieces together and write down* $\mathrm{E}[q(y_j)|\bar{y}]$*, the conditional expectation of* $q(y_j)$ *conditionally on* $\bar{y}$*, which by the Lehmann-Scheffé theorem is the minimum* MSE *unbiased estimator of* π*. The formula you should come up with is*

$$\hat{\pi} = \Phi(\bar{y}\sqrt{n/(n-1)}), \tag{16}$$

where Φ *is the standard normal cumulative distribution function.*

Answer. The conditional expectation of $q(y_j)$ conditionally on $\bar{y} = \bar{y}$ is, by part d, simply the probability that $y_j \geq 0$ under this conditional distribution. In part i this was computed as $\Phi(\bar{y}\sqrt{n/(n-1)})$. Therefore all we have to do is replace $\bar{y}$ by $\bar{y}$ to get the minimum MSE unbiased estimator of π as $\Phi(\bar{y}\sqrt{n/(n-1)})$. □

Remark: this particular example did not give any brand new estimators, but it can rather be considered a proof that certain obvious estimators are unbiased and efficient.

But often this same procedure gives new estimators which one would not have been able to guess. Already when the variance is unknown, the above example becomes quite a bit more complicated, see [Rao73, p. 322, example 2]. *When the variables have an exponential distribution then this example (probability of early failure) is discussed in* [BD77, example 4.2.4 on pp. 124/5].

Problem 19. *Show that the median minimizes the expected absolute deviation.* [Zel71, p. 25].

Problem 20. *Zellner's example about income distribution with the Pareto distribution.* [Zel71, p. 34].

Problem 21. *You perform a Bernoulli experiment with success probability p, but you don't know p; your prior about p is uniform between 0 and 1. Then the probability that the $n+1$st trial is successful, given that the first n trials were already successful,*

- *rises steadily with n*
- *first rises and then declines*
- *declines steadily with n*
- *first declines and then rises*
- *is constant over n.*

Extra credit if you can give a formula for this conditional probability

Answer. The first choice is correct. The formula is $\frac{n+1}{n+2}$. This is Laplace's rule of succession, see [Ros, example 5d. on p. 73]. □

Problem 22. *How to get the original parameter estimates from the reverse regression? Show that it is not what one should expect arithmetically. The product of slope parameters is not 1 but R^2.*

Problem 23. *What is the formula for Mallow's C_p for nonspherical covariance matrix, or if the `weights` variable has a nondefault value?*

Problem 24. *3 points Assume $\hat{\hat{\boldsymbol\beta}}$ is the constrained least squares estimate, and $\boldsymbol\beta_0$ is any vector satisfying $\boldsymbol R\boldsymbol\beta_0 = \boldsymbol u$. Show that in the decomposition*

$$\boldsymbol y - \boldsymbol X\boldsymbol\beta_0 = \boldsymbol X(\hat{\hat{\boldsymbol\beta}} - \boldsymbol\beta_0) + \hat{\hat{\boldsymbol\varepsilon}} \tag{17}$$

the two vectors on the righthand side are orthogonal.

Answer. We have to show $(\hat{\hat{\boldsymbol\beta}} - \boldsymbol\beta_0)^\top \boldsymbol X^\top \hat{\hat{\boldsymbol\varepsilon}} = 0$. Since $\hat{\hat{\boldsymbol\varepsilon}} = \boldsymbol y - \boldsymbol X\hat{\hat{\boldsymbol\beta}} = \boldsymbol y - \boldsymbol X\hat{\boldsymbol\beta} + \boldsymbol X(\hat{\boldsymbol\beta} - \hat{\hat{\boldsymbol\beta}}) = \hat{\boldsymbol\varepsilon} + \boldsymbol X(\hat{\boldsymbol\beta} - \hat{\hat{\boldsymbol\beta}})$, and we already know that $\boldsymbol X^\top\hat{\boldsymbol\varepsilon} = \boldsymbol o$, it is necessary and sufficient to show that $(\hat{\hat{\boldsymbol\beta}} - \boldsymbol\beta_0)^\top \boldsymbol X^\top \boldsymbol X(\hat{\boldsymbol\beta} - \hat{\hat{\boldsymbol\beta}}) = 0$. By (**??**),

$$\begin{aligned}(\hat{\hat{\boldsymbol\beta}} - \boldsymbol\beta_0)^\top \boldsymbol X^\top \boldsymbol X(\hat{\boldsymbol\beta} - \hat{\hat{\boldsymbol\beta}}) &= (\hat{\hat{\boldsymbol\beta}} - \boldsymbol\beta_0)^\top \boldsymbol X^\top \boldsymbol X(\boldsymbol X^\top \boldsymbol X)^{-1}\boldsymbol R^\top\Big(\boldsymbol R(\boldsymbol X^\top\boldsymbol X)^{-1}\boldsymbol R^\top\Big)^{-1}(\boldsymbol R\hat{\boldsymbol\beta} - \boldsymbol u)\\ &= (\boldsymbol u - \boldsymbol u)^\top\Big(\boldsymbol R(\boldsymbol X^\top\boldsymbol X)^{-1}\boldsymbol R^\top\Big)^{-1}(\boldsymbol R\hat{\boldsymbol\beta} - \boldsymbol u) = 0.\end{aligned}$$

□

Problem 25. [Seb77, 6.1.3 on p. 144]. *Assume the true model has the form*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{18}$$

where the $n \times k$ matrix $\boldsymbol{X}$ is nonrandom and contains a constant term, while the $n \times h$ matrix $\boldsymbol{Z}$ is random with the following structure: the rows of $\boldsymbol{Z}$ are independent observations of a given random h-vector $\boldsymbol{z}$. $\boldsymbol{z}$ is independent of $\boldsymbol{\varepsilon}$ and the first and second moments of $\boldsymbol{z}$ exist.

• **a.** *By regressing $\boldsymbol{y}$ on $\boldsymbol{X}$ and $\boldsymbol{Z}$ you obtain the estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, and s^2. Prove that these estimators are unbiased and consistent, and compute their* MSE-*matrix.*

• **b.** *Now you regress $\boldsymbol{y}$ on $\boldsymbol{X}$ only, omitting $\boldsymbol{Z}$. This gives you the estimators $\hat{\hat{\boldsymbol{\beta}}}$ and $\hat{\hat{\sigma}}^2$. Show that in $\hat{\hat{\boldsymbol{\beta}}}$, only the estimator of the constant term is biased, while the slope parameters are unbiased.*

Unfortunately, the random processes generating the omitted explanatory variables are usually more complex than in this simple case.

Problem 26. *I should make a plot of surfaces and ask which ones are additive.*

-

Maximum number of points: 23.

References

[BD77] Peter J. Bickel and Kjell A. Doksum, *Mathematical statistics: Basic ideas and selected topics*, Holden-Day, San Francisco, 1977. 4, 10

[Rao73] C. Radhakrishna Rao, *Linear statistical inference and its applications*, second ed., Wiley, New York, 1973. 10

[Ros] Sheldon Ross, *A first course in probability*. 11

[Seb77] G. A. F. Seber, *Linear regression analysis*, Wiley, New York, 1977. 12

[Zel71] Arnold Zellner, *An introduction to Bayesian inference in econometrics*, Wiley, New York, 1971. 10

Economics Department, University of Utah, 1645 Campus Center Drive, Salt Lake City, UT 84112-9300, U.S.A

E-mail address: `ehrbar@econ.utah.edu`

URL: `http://www.econ.utah.edu/ehrbar`