ECON 546 - MidTerm Test
Spring 2009

(Solution)

Q. 1. (a) Suppose that a researcher has some loss function, $L[\hat{\theta}, \theta]$, in mind when using $\hat{\theta}$ as an estimator of θ. This loss must satisfy:

$$\begin{cases} L(\hat{\theta}, \theta) \geq 0 & ; \forall \theta \\ L(\hat{\theta}, \theta) = 0 & ; \text{if } \theta = \hat{\theta}. \end{cases}$$

To eliminate the random nature of $L(\hat{\theta}, \theta)$, arising from the fact that $\hat{\theta} = \hat{\theta}(y)$ is random, we might average the loss across all possible random values, the result of which we call the risk of $\hat{\theta}$:

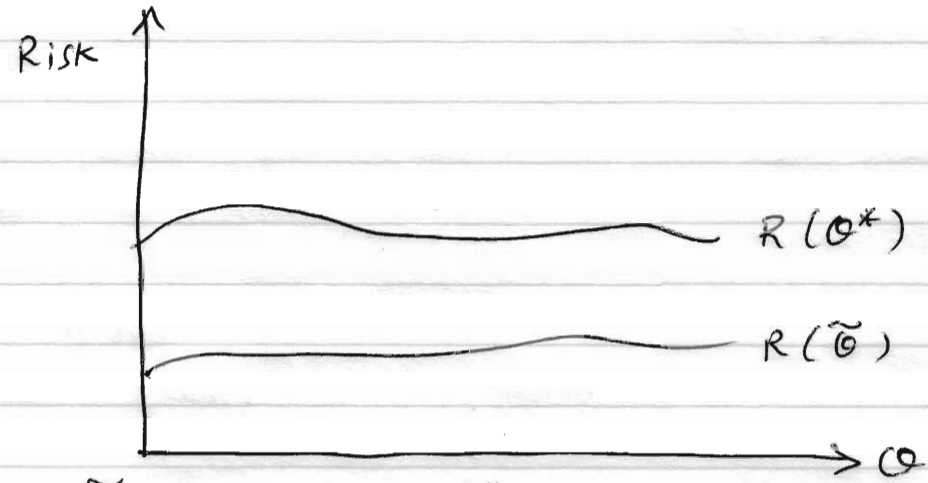
$$R(\hat{\theta}, \theta) = E_y[L(\hat{\theta}; \theta)] = \int_y L(\hat{\theta}, \theta) p(y|\theta) dy.$$

An estimator should, ideally have small risk, for all possible values of θ. Now suppose I intend to use a specific estimator, θ^*. If there is at least one alternative estimator (say $\tilde{\theta}$) such that:

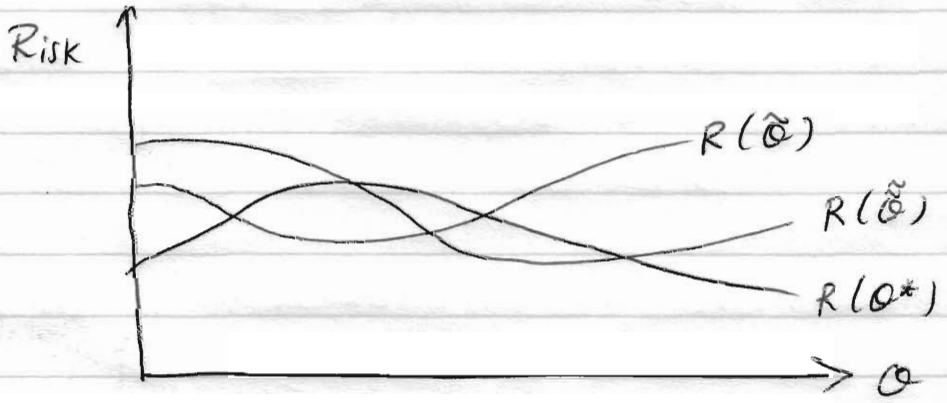
$$R(\tilde{\theta}) \leq R(\theta^*) \quad ; \text{ for all } \theta$$
$$\& \; R(\tilde{\theta}) < R(\theta^*) \quad ; \text{ for some } \theta$$

then we say that my θ^* is "inadmissible", as it is risk-dominated by another estimator.

If it is impossible to find any such $\tilde{\theta}$, then any θ^* is "admissible".



(A) $\tilde{\theta}$ dominates θ^*, so θ^* is "inadmissible".



(B) Only alternative estimators are $\hat{\theta}$ and $\tilde{\theta}$, so θ^* is "admissible". (In fact all 3 estimators are admissible here.)

(b) Let $\hat{\theta}$ be any unbiased estimator of θ, where θ is a scalar parameter. The Cramér-Rao lower-bound tells us the smallest value that the variance of this estimator can be, whatever the (unbiased) estimator is. It gives a lower limit — this limit may not actually be achievable by any estimator.

In the scalar case,

$$\text{var.}(\hat{\theta}) \geq -1 / E\left[\partial^2 \log L / \partial \theta^2\right]$$

(The expression on the RHS is the CRLB.)
So, if we find a $\hat{\theta}$ whose variance is equal to this lower bound, we have found the "most efficient" unbiased estimator of θ.

If $\underset{\sim}{\theta}$ is a vector of parameters, the result becomes
$[V(\underset{\sim}{\hat{\theta}}) - I^{-1}(\theta)]$ is positive semi-definite, where
$I(\theta) = -E\left[\partial^2 \log L / \partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'\right]$.

(c) A statistic is some function of the random sample data: $S_n = f(y_1, \ldots, y_n)$. S_n is "sufficient" if knowledge of this statistic is all that we need to draw inferences about the parameter, θ. That is, if the individual data-points tell us nothing extra. S_n will be sufficient for θ if $p(y_1, \ldots, y_n | S_n)$ does not depend on θ. We have a sufficient statistic if we can factor the joint data density as follows -

$$p(y_1, \ldots, y_n | \theta) = g(S_n | \theta) \cdot h(y_1, \ldots, y_n)$$

where $h(\cdot)$ does <u>not</u> involve θ, and $g(\cdot)$ depends on the data <u>only</u> through S_n. We call this the "Factorization Theorem".

Q.2 (a) $p(y_1, \dots, y_n) = \prod_i p(y_i|\lambda)$ (indep.)

$$= \frac{\lambda^{nk} \prod_i y_i^{k-1} \exp(-\lambda \sum_i y_i)}{[(k-1)!]^n}$$

$$= [\lambda^{nk} \exp(-\lambda \sum_i y_i)] \cdot [\prod_i y_i^{k-1} / (n(k-1)!)]$$

$$= g(\lambda, s_n) \cdot h(y_1, \dots, y_n)$$

where $s_n = \sum_i y_i$. k is known, so $h(\cdot)$ does not depend on λ. So, $\sum_i y_i$ is a sufficient statistic.

(b) $L = p(y_1, \dots, y_n)$

So, $\log L = nk \log\lambda + (k-1) \sum_i \log y_i - \lambda \sum_i y_i - n \log[(k-1)!]$

$(\partial \log L / \partial \lambda) = (nk/\lambda) - \sum_i y_i = 0$; for max

$\Rightarrow \tilde{\lambda} = \left(\frac{kn}{\sum y_i}\right) = (k/\bar{y})$; $\bar{y} = \frac{1}{n} \sum_i y_i$.

$(\partial^2 \log L / \partial \lambda^2) = -nk/\lambda^2 < 0$ (everywhere $\Rightarrow$ max.)

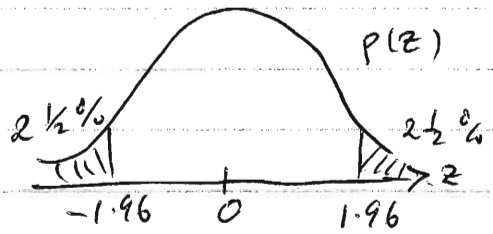
(c) $I = -E(\partial^2 \log L / \partial \lambda^2) = \frac{nk}{\lambda^2}$

$$IA = \lim_{n \to \infty} \left[\frac{1}{n} I\right] = k/\lambda^2.$$

A consistent estimate of IA would be $(k/\bar{y}^2)$, or $k/(k/\bar{y})^2 = (\bar{y}^2/k)$.

(d) $\sqrt{n}(\tilde{\lambda} - \lambda) \xrightarrow{d} N[0, IA^{-1}] = N[0, \lambda^2/k]$.

(e)



The 95% asymptotic confidence interval would be:

$$\tilde{\lambda} \pm 1.96 \text{ ase}(\tilde{\lambda})$$

where "ase" $\Rightarrow$ asymptotic standard error.

Now, asy. var. $\sqrt{n}(\tilde{\lambda} - \lambda) = \lambda^2/k$.

So, asy. var. $(\tilde{\lambda}) = \lambda^2/kn$

asy. s.d. $(\tilde{\lambda}) = \lambda/\sqrt{nk}$

asy. s.e. $(\tilde{\lambda}) = \tilde{\lambda}/\sqrt{nk} = \dfrac{k}{y\sqrt{nk}}$

$= \dfrac{\sqrt{k}}{\sqrt{n}\, y}$

The interval would be

$$(k/y) \pm 1.96 \left(\frac{\sqrt{k}}{\sqrt{n}\, y}\right).$$

(f) $\phi_y(t) = (1 - it/\lambda)^{-k}$

$\phi_y'(t) = (-k)(-i/\lambda)(1 - it/\lambda)^{-(k+1)}$

$+\left(\phi_y'(0)/i\right) = k/\lambda \quad = E(y)$

$\phi_y''(t) = \left(\frac{ik}{\lambda}\right)(-1)(k+1)(-i/\lambda)(1 - it/\lambda)^{-(k+2)}$

& $\left(\phi_y''(0)/i^2\right) = \dfrac{k(k+1)}{\lambda^2} = E(y^2)$

So, $\text{var}(y) = E(y^2) - [E(y)]^2$

$= \dfrac{k(k+1)}{\lambda^2} - \left(\dfrac{k}{\lambda}\right)^2$

$= (k/\lambda^2)$.

(g) MLE of mean is $(k/\tilde{\lambda}) = k/(k/\bar{y}) = \bar{y}$.
MLE of variance is $(k/\tilde{\lambda}^2) = k/(k/\bar{y})^2$
$= (\bar{y}^2/k)$
(using the invariance principle.)

(h) The unrestricted log-likelihood function, evaluated at $\tilde{\lambda}$, is:

$$\log L_u = nk \log \tilde{\lambda} + (k-1)\sum_i \log y_i - \tilde{\lambda}\sum_i y_i - \log[n(k-1)!]$$
$$= nk[\log k - \log \bar{y}] + (k-1)\sum_i \log y_i - nk - \log[n(k-1)!]$$

When we impose the restriction that $\lambda = 1$, we get:

$$\log L_R = (k-1)\sum_i \log y_i - n\bar{y} - \log[n(k-1)!]$$

The LRT statistic is $-2\log\left[L_R/L_u\right] = 2[\log L_u - \log L_R]$.
So,

$$LRT = 2[nk[\log k - \log \bar{y}] - n(k-\bar{y})] \quad \#$$

(Note that as $\tilde{\lambda} \to 1$, $k \to \bar{y}$ & $LRT \to 0$. This seems sensible.)

(i) $LRT \xrightarrow{d} \chi^2_{(1)}$ if H_0 is true.

Here, $k = 2$ & $\bar{y} = 1.8$. So, with $n = 1{,}000$:

$$\begin{aligned} LRT &= 2\left[2000\{\log 2 - \log 1.8\} - 1000(2 - 1.8)\right] \\ &= 2\left[2000(0.10536) - 200\right] \\ &= 2(10.7210) \\ &= 21.442. \end{aligned}$$

The 5% critical value is 3.84. We strongly reject H_0 at this (or any other reasonable) significance level. The result of the test is not sensitive to the choice of α.

Q.3.

(a) We can see that the Newton-Raphson algorithm has been used to maximize or minimize some objective function. We see that the "method" is "ML" — presumably this implies "maximum likelihood". We see that instead of "t-statistics" we have "z-statistics", reflecting the asymptotic normality of the MLE's for the coefficients.

(b) The "Huber-White" estimator of the covariance matrix has been used. As you might guess, this is done to ensure that the standard errors are estimated consistently, even in the face of heteroskedastic errors.

(c) $-8.4651 \pm 1.645\,(4.4694)$

or $[-15.8173 \;;\; -1.1129]$.

(We are not told any units, so we can't report those here.) The interval will valid if the sample size is infinite. We have 753 observations, so this should be reasonable.

(d) Using (EXPER + AGE) as one variable implies that EXPER & AGE are constrained to have the same coefficient (rather than 2 separate coefficients, as in OUTPUT 1). There is one restriction here.

$\log \tilde{L}_U = -3825.319$; $\log \tilde{L}_R = -3899.164$

$$\begin{aligned} LRT &= 2\left[\log \tilde{L}_U - \log \tilde{L}_R\right] \\ &= 2\left[-3825.319 + 3899.164\right] \\ &= 147.69. \end{aligned}$$

With 1 d.o.f. the 5% critical value = 3.84
Even the 0.5% critical value = 7.88.

So, we very strongly reject H_0.