Econ 241A Probability, Statistics and Econometrics

Final Exam

December 14, 2017

Please answer all questions. Show your work. The exam is open book/open note; closed any devices that can communicate. (No laptops, cell phones, Morse code keys, signal res, etc.)

1. You are in an establishment where wagers are made. (Liquid refreshments are also served.) Specifically, you are asked to guess what number will be drawn from an urn. If your guess is within ± 1 of the number drawn, you win a dollar. If your guess is not, you lose a dollar. You would like to maximize your expected earnings. The numbers are drawn from a normal distribution with distribution $x \sim N(\mu, 1)$. Unfortunately, you have no idea what the value of μ is. However, you may sit for a while and watch draws from the urn before you place your bet. But to watch a round you are required to purchase an item of liquid refreshment at cost c. You are not required to drink the item of liquid refreshment. If you do you derive no utility from such consumption, nor does drinking impair your judgement. How many drinks should you purchase? (Seeing as how the answer may be nonlinear, you can give an answer in terms of a function of the number of drinks on the left.)

Solution: This probably should be thought of as a Bayesian problem, but we know that if you have a diffuse prior the posterior of the mean of normals just mimics the usual classical distribution. So the posterior is

$$\mu | n \sim N\left(\bar{x}, \frac{1}{n}\right)$$

where n is the number of drinks.

Obviously, you are going to guess $\bar{x}$ since the problem is symmetric. Your expected winnings are

$$V = 1 \times (F(\bar{x}+1|n) - F(\bar{x}-1|n)) - 1 \times ((1 - F(\bar{x}+1|n) + F(\bar{x}-1|n)) - nc$$

Note that $F(z|n) = \Phi\left(\frac{z-\bar{x}}{\sqrt{1/n}}\right) = \Phi(\sqrt{n}(z-\bar{x}))$. And the first derivative with respect to n is $\phi(\sqrt{n}(z-\bar{x}))(z-\bar{x}) \times 0.5 \times n^{-0.5}$. Thus

$$\frac{dV}{dn} = 2(\phi(\sqrt{n}(\bar{x}+1-\bar{x}))(\bar{x}+1-\bar{x}) \times 0.5 \times n^{-0.5} - \phi(\sqrt{n}(\bar{x}-1-\bar{x}))(\bar{x}-1-\bar{x}) \times 0.5 \times n^{-0.5}) - c$$

$$\frac{dV}{dn} = 2(\phi(\sqrt{n}) \times 0.5 \times n^{-0.5} - \phi(-\sqrt{n})(-1) \times 0.5 \times n^{-0.5}) - c = 0$$

$$\frac{dV}{dn} = 2(\phi(\sqrt{n}) \times n^{-0.5} - c = 0$$

$$\phi(\sqrt{n}) \times n^{-0.5} = \frac{c}{2}$$

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Gamma distribution with parameters α and β with pdf given by

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty, \quad \alpha, \beta > 0$$

   (a) Find the method of moments estimator for α and β. (Hint: $\mathrm{E}[X] = \alpha\beta$ and $\mathrm{Var}(X) = \alpha\beta^2$)

Solution: Denote $\hat{\mu}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\hat{\mu}(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$. By the analogy principle,

$$\mathrm{E}[X] = \alpha\beta = \hat{\mu}(X)$$
$$\mathrm{Var}(X) = \alpha\beta^2 = \hat{\mu}(X^2) - \hat{\mu}(X)^2$$

So

$$\beta_{MME} = \frac{\hat{\mu}(X^2) - \hat{\mu}(X)^2}{\hat{\mu}(X)}$$
$$\alpha_{MME} = \frac{\hat{\mu}(X)^2}{\hat{\mu}(X^2) - \hat{\mu}(X)^2}$$

3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Pareto distribution with parameter β with pdf given by

$$f(x|\beta) = \frac{\beta}{x^{\beta+1}}, \quad x > 1, \quad \beta > 2$$

(a) Find the maximum likelihood estimator of β, β_{MLE}.

Solution:

$$L(\beta|\boldsymbol{x}) = \prod_{i=1}^{n} \frac{\beta}{x_i^{\beta+1}}$$
$$L(\beta|\boldsymbol{x}) = \frac{\beta^n}{\left(\prod_{i=1}^{n} x_i\right)^{\beta+1}}$$
$$\ln L(\beta|\boldsymbol{x}) = n\ln\beta - (\beta+1)\sum_{i=1}^{n}\ln x_i$$

The F.O.C. is

$$\frac{d\ln L(\beta|\boldsymbol{x})}{d\beta} = \frac{n}{\beta} - \sum_{i=1}^{n}\ln x_i = 0$$
$$\beta_{MLE} = \frac{n}{\sum_{i=1}^{n}\ln x_i}$$

(b) Use the transformation method to derive the distribution of $Y = \ln(X)$.

Solution:

$$X = e^Y$$
$$\frac{dX}{dY} = e^Y$$
$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|$$
$$f_Y(y) = \frac{\beta}{(e^y)^{\beta+1}}e^y$$
$$f_Y(y) = \beta e^{-\beta y}$$

(c) Is it β_{MLE} a consistent estimator?

Solution: We can rewrite $\beta_{MLE} = \left(\sum_{i=1}^{n} \frac{\ln x_i}{n}\right)^{-1}$, by the CMT and that $\ln x$ is a continuous function for $x > 0$, we have that

$$\beta_{MLE} = \left(\sum_{i=1}^{n} \frac{\ln x_i}{n}\right)^{-1} \xrightarrow{p} E[\ln X]^{-1} = \beta$$

β_{MLE} is a consistent estimator of β

(d) Use the delta method to derive the asymptotic distribution of β_{MLE}

Solution: By the CLT we know that

$$\sqrt{n}\left(\sum_{i=1}^{n} \frac{\ln x_i}{n} - \frac{1}{\beta}\right) \xrightarrow{d} N\left(0, \frac{1}{\beta^2}\right)$$

given that $\ln X_i \sim exponential(\beta)$, $E[\ln X_i] = 1/\beta$ and $\text{Var}(\ln X_i) = 1/\beta^2$ $(i = 1, 2, \ldots, n)$. By the delta method

$$\sqrt{n}\left(g\left(\sum_{i=1}^{n} \frac{\ln x_i}{n}\right) - g\left(\frac{1}{\beta}\right)\right) \xrightarrow{d} N\left(0, g'(E[\ln X_i])^2 \frac{1}{\beta^2}\right)$$

If $g(u) = u^{-1}$ (continuous for $u \neq 0$) and $g'(u) = -1/u^2$

$$\sqrt{n}\left(\left(\sum_{i=1}^{n} \frac{\ln x_i}{n}\right)^{-1} - \beta\right) \xrightarrow{d} N\left(0, \left(-\frac{1}{1/\beta^2}\right)^2 \frac{1}{\beta^2}\right)$$

$$\sqrt{n}\left(\beta_{MLE} - \beta\right) \xrightarrow{d} N\left(0, \beta^2\right)$$

(e) Find the method of moments estimator of β, β_{MME}. (Hint: $E[X] = \frac{\beta}{\beta-1}$)

Solution:

$$E[X] = \frac{\beta}{\beta - 1} = \bar{X}$$

$$\beta_{MME} = \frac{\bar{X}}{\bar{X} - 1}$$

(f) What is the distribution of β_{MME} for a large sample (Hint: use the CLT for the sample average and the delta method).

Solution: By the CLT we know that

$$\sqrt{n}(\bar{X} - E[X]) \xrightarrow[n\to\infty]{d} N(0, \text{Var}(X))$$

we know that $g(u) = \frac{u}{u-1}$ is continuous for $u \neq 1$ and $g'(u) = -\frac{1}{(u-1)^2}$. By the delta method

$$\sqrt{n}(g(\bar{X}) - g(E[X])) \xrightarrow[n\to\infty]{d} N(0, g'(E[X])^2 \text{Var}(X))$$

The asymptotic variance is

$$
\begin{aligned}
g'(E[X])^2\text{Var}(X) &= \frac{1}{(E[X]-1)^4}\frac{\beta}{(\beta-1)^2(\beta-2)} \\
g'(E[X])^2\text{Var}(X) &= (\beta-1)^4\frac{\beta}{(\beta-1)^2(\beta-2)} \\
g'(E[X])^2\text{Var}(X) &= \frac{\beta(\beta-1)^2}{(\beta-2)} \\
\sqrt{n}(\beta_{MME}-\beta) &\xrightarrow[n\to\infty]{d} N\left(0, \frac{\beta(\beta-1)^2}{(\beta-2)}\right)
\end{aligned}
$$

(g) Compute the Cramer-Rao Lower Bound for an unbiased estimator of the exponential distribution.

Solution: The CRLB for unbiased estimator is

$$
\begin{aligned}
CRLB &= \frac{1}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]} \\
CRLB &= \frac{1}{nE\left[\frac{d^2}{d\beta^2}\log \beta e^{-\beta X}\right]} \\
CRLB &= \frac{1}{nE\left[\frac{1}{\beta^2}\right]} \\
CRLB &= \frac{\beta^2}{n}
\end{aligned}
$$

(h) Compare the asymptotic variance of β_{MLE} and β_{MME} (Hint 1: $\frac{x(x-1)^2}{(x-2)} = x^2 + \frac{2}{(x-2)} + 1$. Hint 2: use your answer to part g).

Solution:

$$
n\text{AVar}(\beta_{MLE}) = \beta^2 \leq \frac{\beta(\beta-1)^2}{(\beta-2)} = \beta^2 + \frac{2}{(\beta-2)} + 1 = n\text{AVar}(\beta_{MME})
$$

where $\text{AVar}(X)$ is the asymptotic variance of X and $\beta > 2$ by assumption.
The MLE estimator achieves asymptotically the CRLB.

4. Income is distributed (very) roughly log-normal, once people with zero income are dropped from the sample. The pdf of a lognormal distribution is

$$
f(y_i|\mu,\sigma^2) = \frac{1}{y_i\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\ln y_i-\mu)^2}{2\sigma^2}\right\}, \quad 0 \leq y_i < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0
$$

Assume that $Y_1, Y_2, \ldots, Y_n$ is a random sample from this distribution. Let FE_i be an indicator variable such that

$$
FE_i = \begin{cases} 1 & \text{if } i \text{ is a woman} \\ 0 & \text{if } i \text{ is a man} \end{cases}
$$

We suspect that the distribution for income of men and women differ so $\mu_M \neq \mu_F$ and $\sigma_M \neq \sigma_F$ in general.

(a) Suppose that $\mu_M = \mu_F = \mu$ and $\sigma_M = \sigma_F = \sigma$, write the likelihood function $L(\mu, \sigma^2 | y_1, y_2, \ldots, y_n) := f(y_1, y_2, \ldots, y_n | \mu, \sigma^2)$

Solution:

$$L(\mu, \sigma^2 | y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} \frac{1}{y_i \sigma \sqrt{2\pi}} \exp\left\{ -\frac{(\ln y_i - \mu)^2}{2\sigma^2} \right\}$$

(b) Find the maximum likelihood estimators for μ and σ^2.

Solution:

$$L(\mu, \sigma^2 | y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} \frac{1}{y_i \sigma \sqrt{2\pi}} \exp\left\{ -\frac{(\ln y_i - \mu)^2}{2\sigma^2} \right\}$$

$$L(\mu, \sigma^2 | y_1, y_2, \ldots, y_n) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\sum_{i=1}^{n} \frac{(\ln y_i - \mu)^2}{2\sigma^2} \right\} \prod_{i=1}^{n} y_i^{-1}$$

$$\ln L(\mu, \sigma^2 | y_1, y_2, \ldots, y_n) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \sum_{i=1}^{n} \frac{(\ln y_i - \mu)^2}{2\sigma^2} - \sum_{i=1}^{n} \ln y_i$$

This problem is very similar to find the MLE of a normal distribution. We know that

$$\mu_{MLE} = \sum_{i=1}^{n} \frac{\ln y_i}{n}$$

$$\sigma^2_{MLE} = \frac{\sum_{i=1}^{n} \left(\ln y_i - \sum_{i=1}^{n} \frac{\ln y_i}{n} \right)^2}{n}$$

(c) Write the likelihood function $L(\mu_F, \mu_M, \sigma^2_M, \sigma^2_F | y_1, y_2, \ldots, y_n)$ where $\mu_F \neq \mu_M$ and $\sigma_F \neq \sigma_M$.

Solution:

$$L(\mu_F, \mu_M, \sigma^2_M, \sigma^2_F | y_1, y_2, \ldots, y_n) =$$

$$\prod_{i=1}^{n} \left[\frac{1}{y_i \sigma_F \sqrt{2\pi}} \exp\left\{ -\frac{(\ln y_i - \mu_F)^2}{2\sigma^2_F} \right\} \right]^{FE_i} \left[\frac{1}{y_i \sigma_M \sqrt{2\pi}} \exp\left\{ -\frac{(\ln y_i - \mu_M)^2}{2\sigma^2_M} \right\} \right]^{1-FE_i}$$

(d) Find the maximum likelihood estimators for μ_M, μ_F, σ^2_M and σ^2_F.

Solution: By the multiplicative structure of the likelihood, the estimation of (μ_F, σ^2_F) is independent of the estimation of (μ_M, σ^2_M). This implies that

$$\mu_{j,MLE} = \sum_{i=1, FE_i=j}^{n} \frac{\ln y_i}{n_j}$$

$$\sigma^2_{j,MLE} = \frac{\sum_{i=1, FE_i=j}^{n} \left(\ln y_i - \sum_{i=1, FE_i=j}^{n} \frac{\ln y_i}{n_j} \right)^2}{n_j}$$

where $j = 1$ for women and $j = 0$ for men, and n_j is the number of people in group j.
In this case we simply compute the average and the sample variance of the log-income for both groups separately.

(e) We want to test the hypothesis that the distribution of income between men and women differ. Write down the null and alternative hypotheses (H_0 and H_1).

Solution: $H_0 : \mu_F = \mu_M \wedge \sigma_F^2 = \sigma_M^2$ and $H_1 : \mu_F \neq \mu_M \vee \sigma_F^2 \neq \sigma_M^2$

(f) Write down the likelihood ratio test statistic (LRT) for the hypothesis in part (e). (Hint: use your answers to parts (a) and (c))

Solution:

$$LRT(\boldsymbol{y}) = \frac{L(\hat{\mu}, \hat{\sigma}^2 | y_1, y_2, \ldots, y_n)}{L(\hat{\mu}_F, \hat{\mu}_M, \hat{\sigma}_M^2, \hat{\sigma}_F^2 | y_1, y_2, \ldots, y_n)}$$

(g) Assume n is large, what is the (asymptotic) distribution of $-2\log(LRT)$, where LRT is the likelihood ratio test statistic? (You may write it down from the notes).

Solution: $-2\log(LRT) \sim_A \chi_2^2$

(h) Use your answer to part (g) to write down a rejection rule for H_0 at α% of significance.

Solution: Reject H_0 if $-2\log(LRT) > \chi^2_{2,1-\alpha}$. Do not reject otherwise.

(i) Write down the Wald test statistic for the hypothesis in part (e).
Hint:
Let $\hat{\theta}$ be the MLE estimator of θ. Under some regularity conditions and for large samples $\hat{\theta} \sim_A N(\theta, I^{-1}(\theta))$ where $I(\theta) = -E[\frac{\partial^2}{\partial\theta\partial\theta'} \ln f(Y|\theta)]$ is the Fisher's information matrix (Remember that $f(Y|\theta) = L(\theta|Y)$). Assume these conditions hold.

Solution:

$$W = (R\hat{\theta} - \theta_0)'(RI^{-1}(\theta)R')^{-1}(R\hat{\theta} - \theta_0)$$

where

$$R = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$
$$\theta = (\mu_F, \sigma_F^2, \mu_M, \sigma_M^2)'$$
$$\theta_0 = (0, 0, 0, 0)'$$
$$I(\theta) = \begin{bmatrix} I(\mu_F, \sigma_F) & 0 \\ 0 & I(\mu_M, \sigma_M) \end{bmatrix}$$
$$I(\mu_j, \sigma_j) = \begin{bmatrix} \frac{n}{\sigma_j^2} & 0 \\ 0 & \frac{n}{2\sigma_j^4} \end{bmatrix} \qquad j = F, M$$