Stat 582
Emerson, Win 12
Homework #3 Key
January 28, 2012

Written problems to be handed in Friday, February 3, 2012.

1. Prove: If sequence of random variables $X_n \to_d a$, where a is constant, then $X_n \to_p a$.

2. Prove: If sequence of random variables $X_n \to_d X$, then $Var(X) \leq \liminf_{n\to\infty} Var(X_n)$.

3. Prove the delta method: Suppose $a_n \to \infty$ as $n \to \infty$ and X_n is a sequence of random variables. Further suppose $g : \mathcal{R}^1 \to \mathcal{R}^1$ is a continuous function that is differentiable at θ. Then

$$a_n\,(X_n - \theta) \to_d X \qquad \Rightarrow \qquad a_n\,(g(X_n) - g(\theta)) \to_d g'(\theta)X.$$

Ans: If g is *continuously* differentiable at θ (the far more useful case in real life), by the mean value theorem

$$g(X_n) - g(\theta) = (X_n - \theta)g'(\phi_n)$$

for some ϕ_n between θ and X_n. Since $a_n(X_n - \theta) \to_d X$, then by Slutsky's $(X_n - \theta) \to_d 0$, and $X_n \to_p \theta$. By Mann-Wald, $g(X_n) \to_p g(\theta)$ and $g'(X_n) \to_p g'(\theta)$.

Because ϕ_n is between X_n and θ, we must also have $\phi_n \to_p \theta$ and by Mann-Wald that $g'(\phi_n) \to_p g'(\theta)$. Then using Slutsky's

$$a_n(g(X_n) - g(\theta)) = g'(\phi_n)a_n(X_n - \theta) \to_d g'(\theta)X$$

gives us the result.

If we only know that g is differentiable at θ, then (we are essentially proving Taylor's theorem for a first order expansion of a continuous function of a consistent estimator here) we know that for

$$h(x) = \begin{cases} \frac{g(x)-g(\theta)}{x-\theta} - g'(\theta) & x \neq \theta, \text{ and} \\ 0 & x = \theta \end{cases},$$

we have by the continuity of g that h is continuous at all $x \neq \theta$ and by the differentiability of g at θ that $\lim_{x\to\theta} h(x) = 0$, so $h(x)$ is continuous at θ. Now

$$g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + h(X_n)(X_n - \theta),$$

so

$$a_n(g(X_n) - g(\theta)) = g'(\theta)a_n(X_n - \theta) + h(X_n)a_n(X_n - \theta).$$

Now if $a_n\,(X_n - \theta) \to_d X$ and $a_n \to \infty$, then $1/a_n \to 0$ and by Slutsky's theorem

$$X_n = \frac{1}{a_n}a_n(X_n - \theta) + \theta \to_d 0 \cdot X + \theta = \theta.$$

Furthermore $X_n \to_d \theta$ a constant implies $X_n \to_p \theta$. so as $n \to \infty$, $h(X_n) \to_p 0$ by the Mann-Wald (continuous mapping) theorem. Hence, again by Slutsky's theorem we have

$$h(X_n)a_n(X_n - \theta) \to_d 0 \cdot X = 0 \quad \Rightarrow \quad h(X_n)a_n(X_n - \theta) \to_p 0,$$

which with another application of Slutsky's theorem yields

$$a_n(g(X_n) - g(\theta)) = g'(\theta)a_n(X_n - \theta) + h(X_n)a_n(X_n - \theta) \to_d g'(\theta)X + 0 = g'(\theta)X.$$

4. Consider known sequence of predictors $\{x_i\}_{i=1}^{\infty}$, independent identically distributed random variables $\{\epsilon_i\}_{i=1}^{\infty}$ with $\epsilon_i \sim (0, \sigma^2 < \infty)$, and derived random variables $Y_i = \alpha + \beta x_i + \epsilon_i$. Let $\hat{\theta} = (\hat{\alpha}, \hat{\beta})^T$ be the ordinary least squares estimates of $\theta = (\alpha, \beta)^T$.

a. Show that for suitable restrictions on the x_i's (and make clear what those restrictions are)

$$V_n = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

we have

$$V_n^{1/2}(\hat{\theta} - \theta) \to_d \mathcal{N}_2(0, \sigma^2 I_2),$$

where I_2 is the identity matrix. (Hint: consider the Cramér-Wold device and the Lindeberg-Feller central limit theorem.)

Ans: The basic idea:

- Anticipate the restrictions that will need to be placed on the distribution of the x_i's by considering known exact results and the Lindeberg condition.
- Simplify the problem under those restrictions to make the notation more tractable.
- Specify the appropriate normalization that will lead to a bivariate normal asymptotic distribution.
- Construct an arbitrary linear combination of the bivariate normalized statistic.
- Use the Lindeberg-Feller central limit theorem to show asymptotic normality of the linear combination.
- Appeal to the Cramér-Wold device to establish the bivariate normal asymptotic distribution.

First, we note that because the x_i's are known we can easily reparameterize our problem to $Y_i = \beta_0 + \beta_1(x_i - \overline{x}) + \epsilon_i$, with $\alpha = \beta_0 - \beta_1\overline{x}$ and $\beta = \beta_1$. Defining $\vec{\beta} = (\beta_0 \;\; \beta_1)^T$, we also note that we have very extensive theory about the conditional (on the x_i's) distribution of least squares estimates when the ϵ_i of our problem are independent and identically distributed according to $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. In that setting, we define $\mathbf{X} = (1_n^T \;\; \vec{x}^T - \overline{x})$, and we note that for the ordinary least squares estimator

$$\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$$

has distribution

$$\widehat{\vec{\beta}} \sim \mathcal{N}_2\left(\vec{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right),$$

providing that the x_i's are not all equal to a single constant. Knowing that we would eventually like to have an asymptotic distribution independent of n, we consider

$$(\mathbf{X}^T\mathbf{X})^{1/2}\left(\widehat{\vec{\beta}} - \vec{\beta}\right) \sim \mathcal{N}_2\left(0, \sigma^2\mathbf{I_2}\right),$$

and any limiting distribution will have to agree with that. Under our reparameterization

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & 0 \\ 0 & ss_{xx} \equiv \sum_{i=1}^n (x_i - \overline{x})^2 \end{bmatrix} \quad \Rightarrow \quad (\mathbf{X}^T\mathbf{X})^{1/2} = \begin{bmatrix} \sqrt{n} & 0 \\ 0 & \sqrt{ss_{xx}} \end{bmatrix}.$$

Now onward...Let

$$\begin{aligned} \vec{Z}_n &= (\mathbf{X}^T\mathbf{X})^{1/2}(\widehat{\vec{\beta}} - \vec{\beta}) \\ &= \begin{pmatrix} \sqrt{n}(\hat{\beta}_0 - \beta_0) \\ \sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}(\hat{\beta}_1 - \beta_1) \end{pmatrix} \end{aligned}$$

with $\widehat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\vec{Y}$. Then by doing the matrix arithmetic,

$$\hat{\beta}_0 = \overline{Y} \qquad (\text{and } \hat{\alpha} = \overline{Y} - \hat{\beta}_1\overline{x})$$

$$\hat{\beta}_1 = \frac{ss_{xy}}{ss_{xx}} = \frac{\sum_{i=1}^n (x_i - \overline{x})Y_i}{\sum_{j=1}^n (x_j - \overline{x})^2}.$$

Further,

$$\begin{aligned}\hat{\beta}_0 - \beta_0 &= \frac{1}{n}\sum_{i=1}^n (Y_i - \beta_0)\\ &= \frac{1}{n}\sum_{i=1}^n [(x_i - \overline{x})\beta_1 + \epsilon_i]\\ &= \overline{\epsilon}\end{aligned}$$

and

$$\begin{aligned}\hat{\beta}_1 - \beta_1 &= \frac{1}{\sum_{j=1}^n (x_j - \overline{x})^2}\sum_{i=1}^n (x_i - \overline{x})(Y_i - (x_i - \overline{x})\beta_1)\\ &= \frac{1}{\sum_{j=1}^n (x_j - \overline{x})^2}\sum_{i=1}^n (x_i - \overline{x})(\beta_0 + \epsilon_i)\\ &= \frac{\sum_{i=1}^n (x_i - \overline{x})\epsilon_i}{\sum_{j=1}^n (x_j - \overline{x})^2}.\end{aligned}$$

Hence $Z_{n_1} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i$ and $Z_{n_2} = \frac{\sum_{i=1}^n (x_i - \overline{x})\epsilon_i}{\sqrt{\sum_{j=1}^n (x_j - \overline{x})^2}}$. Now consider $W_n = aZ_{n_1} + bZ_{n_2}$.

$$\Rightarrow W_n = \sum_{i=1}^n \left(\frac{a}{\sqrt{n}} + \frac{b(x_i - \bar{x})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2}}\right)\epsilon_i \equiv \sum_{i=1}^n w_{n,i}$$

so that

$$E[w_{n,i}] = \sum_{i=1}^n \left(\frac{a}{\sqrt{n}} + \frac{b(x_i - \bar{x})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2}}\right) E[\epsilon_i] = 0$$

$$Var[w_{n,i}] = \left(\frac{a}{\sqrt{n}} + \frac{b(x_i - \bar{x})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2}}\right)^2 \sigma^2$$

$$V_n \equiv \sum_{i=1}^n Var[w_{n,i}] = (a^2 + b^2)\sigma^2$$

We now want to use the L-F CLT to show that $W_n/V_n \to_d \mathcal{N}(0,1)$, so we must show that

$$\frac{1}{V_n}\sum_{i=1}^n E[w_{n,i}^2 \times 1_{[|w_{n,i}| > M\sqrt{V_n}]}] \to 0 \qquad \forall M > 0.$$

To this end, let $c_{n,i} = \frac{a}{\sqrt{n}} + \frac{b(x_i - \overline{x})}{\sqrt{\sum_{j=1}^n (x_j - \overline{x})^2}}$ and $c_n^* = \max\{c_{n,i}\}$ and suppose that

$$\max\left\{\frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}\right\} \to 0 \text{ as } n \to \infty.$$

$$\Rightarrow c_n^* \to 0 \text{ as } n \to \infty.$$

So,

$$\frac{1}{V_n}\sum_{i=1}^n E[c_{n,i}^2\epsilon_i^2 \times 1_{[|c_{n,i}\epsilon_i|>M\sqrt{V_n}]}] \leq \frac{1}{V_n}\sum_{i=1}^n E[c_{n,i}^2\epsilon_i^2 \times 1_{[|\epsilon_{n,i}|>M\sqrt{V_n}/c_n^*]}]$$

$$= \frac{1}{V_n}E[\epsilon_i^2 \times 1_{[|\epsilon_{n,i}|>M\sqrt{V_n}/c_n^*]}]\sum_{i=1}^n c_{n,i}^2 \qquad \text{(since } \epsilon_i \text{ are iid)}$$

$$= \frac{1}{(a^2+b^2)\sigma^2}E[\epsilon_i^2 \times 1_{[|\epsilon_{n,i}|>M\sqrt{V_n}/c_n^*]}] \times (a^2+b^2)$$

$$= \frac{1}{\sigma^2}E[\epsilon_i^2 \times 1_{[|\epsilon_{n,i}|>M\sqrt{V_n}/c_n^*]}] \to 0$$

$$\text{(since } M\sqrt{V_n}/c_n^* \to \infty \text{ since } c_n^* \to 0 \text{ as } n \to \infty\text{)}$$

So the Lindeberg condition holds and $aZ_{n_1} + bZ_{n_2} \to_d \mathcal{N}(0,1)$. Finally, by appealing Cramer-Wold we have the desired result.

So we obtained the desired result by placing the restriction that

$$\max\left\{\frac{(x_i-\bar{x})^2}{\sum_{j=1}^n(x_j-\bar{x})^2}\right\} \to 0 \text{ as } n \to \infty.$$

In general, this will be satisfied so long as the sum of the squared x residuals becomes infinite. Among other possibilities, this would be violated if

- There were some x_0 such that $x_i = x_0$ almost always. In this case, $\overline{x}_n \to x_0$ and $(x_i - \overline{x}_n)^2$ is 0 almost always. Hence, the sum of squared residuals will not approach infinity, and nonzero residuals will cause the maximum to be nonzero. will not approach 0.
- $x_1 = 1$ and $x_i = -2^{i-1}$ for $i \geq 2$. In this case, $\overline{x}_n \to 0$ and the sum of the squared residuals will approach 1.5, with the maximum squared residual equalling 1.

b. Show that $\hat{\sigma}_n^2 = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}x_i)^2/(n-2)$ is a consistent estimator for σ^2.

Ans: Notationally, let $\mu_i = \alpha + \beta x_i$ and $\hat{\mu}_i = \hat{\alpha} + \hat{\beta}x_i$. Note further that

$$\frac{1}{n}\sum_{i=1}^n(Y_i-\hat{\mu}_i)^2 = \frac{1}{n}\sum_{i=1}^n(Y_i-\mu_i)^2 - \frac{2}{n}\sum_{i=1}^n(Y_i-\mu_i)(\hat{\mu}_i-\mu_i) + \frac{1}{n}\sum_{i=1}^n(\hat{\mu}_i-\mu_i)^2$$

We will deal with the three terms on the right hand side in sequence.

The first term is handled easily because by the weak law of large numbers

$$\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_i)^2 = \frac{1}{n}\sum_{i=1}^n(\epsilon_i)^2 \to_p \sigma^2$$

The second and third terms are handled a bit easier in matrix notation. The second term thus becomes

$$\frac{2}{n}(\vec{Y} - \mathbf{X}\vec{\beta})^T(\mathbf{X}\widehat{\vec{\beta}} - \mathbf{X}\vec{\beta}) = \frac{2}{n}\vec{Y}^T\mathbf{X}(\widehat{\vec{\beta}} - \vec{\beta}) - \vec{\beta}^T\mathbf{X}^T\mathbf{X}(\widehat{\vec{\beta}} - \vec{\beta}).$$

Now from part a we immediately obtain that so long as $ss_{xx} \to \infty$, $\widehat{\vec{\beta}} \to_p \beta$. We thus want to consider $\vec{Y}^T\mathbf{X}/n$ and $\mathbf{X}^T\mathbf{X}/n$. Now

$$\vec{Y}^T\mathbf{X}/n = (\mathbf{X}\beta + \vec{\epsilon})^T\mathbf{X} = \mathbf{X}^T\mathbf{X}/n + (\bar{\epsilon} \quad \vec{\epsilon}^T\vec{x}/n)^T.$$

By the WLLN, we know $\overline{\epsilon}_n \rightarrow \infty$. Furthermore, if we assume that $\vec{\epsilon}^T\vec{x}/n \rightarrow 0$ (i.e., the errors are not correlated with predictors), we only need to consider whether $\mathbf{X}^T\mathbf{X}/n$ has finite elements in order to have both the second and third terms converge in probability to 0. But this obtains so long as $ss_{xx}/n \rightarrow v_x < \infty$, so we will add this to our conditions on the predictor distribution. We thus have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu}_i)^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_i)^2 - \frac{2}{n}\sum_{i=1}^{n}(Y_i - \mu_i)(\hat{\mu}_i - \mu_i) + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_i - \mu_i)^2 \\ &\rightarrow_p \sigma^2 + 0 + 0\end{aligned}$$

and because $n/(n-2) \rightarrow_p 1$, we obtain the desired consistency. Recapping our restrictions on the sampling distribution

- $ss_{xx} \rightarrow \infty$
- $x_i - \overline{x} < \infty$
- $\vec{\epsilon}^T\vec{x} \rightarrow 0$

Note that we could also have obtained the desired consistency by noting that

$$\sum_{i=1}^{n}(Y_i - \hat{\mu}_i)^2 = \vec{Y}^T(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})\mathbf{X}^T)\vec{Y} = \vec{\epsilon}^T(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})\mathbf{X}^T)\vec{\epsilon} = \vec{\epsilon}^t\mathbf{Q}\vec{\epsilon} =$$

and using general results for the variance of a quadratic form which for independent, identically distributed $\epsilon_i \sim (0, \sigma^2, \mu_3, \mu_4)$

$$var(\vec{\epsilon}^T\mathbf{Q}\vec{\epsilon}) = (\mu_4 - 3\sigma^4)\vec{a}^T\vec{a} + 2\sigma^4 tr(\mathbf{Q})$$

where $\vec{a} = diag(\mathbf{Q}$. Then, noting that $\hat{\sigma}^2$ is unbiased (hence, asymptotically unbiased), we obtain the desired consistency so long as

$$var(\vec{\epsilon}^T\mathbf{Q}\vec{\epsilon})/n \rightarrow 0.$$

c. Show that the normal theory based inference for ordinary least squares regression estimates (i.e., statistical inference based on the t and F distributions) in this homoscedastic setting is thus asymptotically valid for any error distribution with a finite variance.

Ans: The normal theory based inference uses the presumption that $\widehat{\vec{\beta}} \sim \mathcal{N}_2(\vec{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$ and that $(n-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-2}$. Tests of $H_0 : \mathbf{A}\vec{\beta} = \vec{c}$ are based on the quadratic form

$$\frac{(\mathbf{A}\widehat{\vec{\beta}} - \vec{c})^T\mathbf{X}^T\mathbf{X}(\mathbf{A}\widehat{\vec{\beta}} - \vec{c})}{\hat{\sigma}} \sim F_{r,n-2}$$

where $r = rank(\mathbf{A})$. Owing to the asymptotic normality of $\widehat{\vec{\beta}}$, and the consistency of $\hat{\sigma}^2$ for σ^2, the quadratic form has an asymptotic χ^2_r distribution. Furthermore, based by the definition of the F distribution (the ratio of two independent chi square random variables, each divided by their degrees of freedom) and our results about the asymptotic distribution of a chi square random variable divided by its degrees of freedom from homework 1, we easily obtain that the limiting distribution for a $F_{r,n-2}$ random variable is χ^2_r as $n \rightarrow \infty$.

5. Consider identically distributed random variables

$$X_i \;\sim\; (\mu, \sigma^2 \in (0, \infty)) \quad \text{having correlation } corr(X_i, X_j) = \rho \text{ with } 0 < \rho < 1 \text{ for } i \neq j.$$

Prove or disprove the following statement about the sample mean: $\overline{X}_n \rightarrow_p \mu$ as $n \rightarrow \infty$. (If you want, you can extend your answer to the three additional special cases when $\rho < 1$, $\rho = 0$, and $\rho = 1$.

Ans: Note first that the expectation and variance of $\overline{X}_n$ can be derived as

$$\begin{aligned} E(\overline{X}_n) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \\ Var(\overline{X}_n) &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{j\neq i} Cov\left(X_i X_j\right)\right] \\ &= \frac{1}{n^2}\left[n\sigma^2 + n(n-1)\rho\sigma^2\right] \\ &= \frac{\sigma^2}{n}\left[1 + (n-1)\rho\right] \end{aligned}$$

From this, we note that for $Var(\overline{X}_n) \geq 0$, we must have $\rho \geq -1/(n-1)$, and if this is to be true for all $n \geq 1$, we need $\rho \geq 0$ in this "exchangeable" setting.

We further note that if $\rho = 0$, we obtain the consistency of $\overline{X}_n$ using problem 3 of homework # 1 (it is unbiased, hence, asymptotically unbiased, and $Var(\overline{X}_n) = \sigma^2/n \to 0$ or more directly by problem 7 of homework # 1. (Note that we cannot use the usual weak law of large numbers, because we do not know that our data are independent.)

Now for $\rho > 1$, we have $Var(\overline{X}_n) \to \rho\sigma^2 \neq 0$. We consider the special case in which $(X_1, \ldots, X_n)^T$ is multivariate normal for all $n \geq 1$. In that setting, we then know that $\overline{X}_n \to_d \mathcal{N}(\mu, \rho\sigma^2)$, and

$$P(|\overline{X}_n - \mu| \geq \epsilon) \to 1 - 2\Phi\left(-\frac{\epsilon}{\sigma\sqrt{\rho}}\right) > 0 \text{ for all } \epsilon > 0.$$