Stat 582
Emerson, Win 12
Midterm Key
February 17, 2012

This exam is closed book, closed notes. Show sufficient detail to clearly establish your method of proof or solution.

1. (10 pt) State the elementary Skorokhod's theorem.

Ans: Suppose that $X_n \to_d X_0$. Then there exist random variables X_n^* defined on the probability space $([0,1], \mathcal{B}[0,1], \lambda)$ for which $X_n^* =_d X_n$ for all $n \geq 0$ with $X_n^* \to_{as} X_0^*$.

2. (15 pt) State and prove the delta method.

Ans: 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 *continuously* differentiable at θ. Then

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

Pf: 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.

Alt: Alternatively: 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

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

Pf 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 $a_n\left(X_n - \theta\right) \to_d X$ 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

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

3. (20 pt) State and prove the Glivenko-Cantelli theorem.

Ans: Let $I(t) = t$ and for independent identically distributed $\xi_i \sim \mathcal{U}(0,1)$, define empirical distribution function

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(\xi_i) \qquad \text{for } t \in [0,1].$$

Then as $n \to \infty$

$$||G_n - I||_\infty \equiv \sup_{0 \le t \le 1} |G_n(t) - t| \to_{as} 0.$$

Pf: Let M be an arbitrary large number. Then

$$\begin{aligned}
||G_n - I||_\infty &= \max_{1 \le j \le M}\left[\sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} |G_n(t) - t|\right] \\
&= \max_{1 \le j \le M}\left[\sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} |G_n(t) - t|\right] \\
&= \max_{1 \le j \le M}\left[\max\left(\sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} (G_n(t) - t), \sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} (t - G_n(t))\right)\right] \\
&\le \max_{1 \le j \le M}\left[\max\left(\sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} \left(G_n\left(\frac{j}{M}\right) - \frac{j-1}{M}\right), \sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} \left(\frac{j}{M} - G_n\left(\frac{j-1}{M}\right)\right)\right)\right] \\
&\le \max_{1 \le j \le M}\left[\max\left(\sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} \left(G_n\left(\frac{j}{M}\right) - \frac{j}{M}\right), \sup_{\frac{j-1}{M} \le t \le \frac{j}{M}} \left(\frac{j-1}{M} - G_n\left(\frac{j-1}{M}\right)\right)\right)\right] + \frac{1}{M} \\
&\to_{as} \frac{1}{M},
\end{aligned}$$

where the last step follows from the strong law of large numbers, which provides that

$$G_n\left(\frac{j}{M}\right) \to_{as} \frac{j}{M} \quad \text{so} \quad G_n\left(\frac{j}{M}\right) - \frac{j}{M} \to 0.$$

Now, because the above holds for arbitrarily large M, we have $||G_n - I||_\infty \to_{as} 0$.

Note that for independent identically distributed $X_i \sim F$, $P(F(X_i) \le t) = t$, and because $F^{-1}(\xi_i) \sim F$, the above results also can be used to show that for empirical distribution function $F_n(t)$,

$$||F_n - F||_\infty \le ||G_n - I||_\infty \to_{as} 0.$$

4. (20 pt) Let $X_1, X_2, \ldots$ be independent random variables distributed according to a mixture of normals in which $X_i \sim f_X(x)$ where density f_X is derived from a mixture having probability $1-p$ of a $\mathcal{N}(\mu, 1)$ distribution and probability p of a $\mathcal{N}(\mu, \sigma^2)$ distribution. Hence,

$$f_X(x) = (1-p)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2}\right) + p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Let $\overline{X}_n \sum_{i=1}^n X_i/n$ and $\tilde{\mu}_n$ be the sample median of $(X_1, \ldots, X_n)$. Find the asymptotic relative efficiency of $\overline{X}_n$ to $\tilde{\mu}_n$ as estimators of μ.

Ans: Letting Z represent a latent variable that indicates which of the two normal distributions an observation might be from, we can find the moments of the distribution of X_i as

$$\begin{aligned} E[X_i] &= E_Z\left[E[X_i|Z]\right] = (1-p)\mu + p\mu = \mu \\ Var(X_i) &= E_Z\left[Var(X_i|Z)\right] + Var\left(E[X_i|Z]\right) \\ &= (1-p)\cdot 1 + p\sigma^2 + 0 = 1 - p + p\sigma^2 \end{aligned}$$

By the Levy central limit theorem, we thus have

$$\sqrt{n}(\overline{X}_n - \mu) \to_d \mathcal{N}\left(0, 1 - p + p\sigma^2\right).$$

Now

$$Pr(X_i \le x) = (1-p)Pr(\mathcal{N}(\mu, 1) \le x) + pPr(\mathcal{N}(\mu, \sigma^2) \le x),$$

so $Pr(X_i \le \mu) = 0.5$, and μ is both the mean and the median of X_i. We can therefore use sample median $\tilde{\mu}_n$ as an estimator of μ, and from the asymptotic distribution of the sample median of a continuous random variable having positive density at the median, we know

$$\sqrt{n}(\tilde{\mu}_n - \mu) \to_d \mathcal{N}\left(0, \frac{0.5(1-0.5)}{[f(\mu)]^2}\right).$$

In this case,

$$f(\mu) = \frac{1-p}{\sqrt{2\pi}} + \frac{p}{\sqrt{2\pi}\sigma},$$

so

$$\sqrt{n}(\tilde{\mu}_n - \mu) \to_d \mathcal{N}\left(0, \frac{\pi\sigma^2}{2\left(\sigma(1-p)+p\right)^2}\right).$$

The asymptotic relative efficiency of $\overline{X}_n$ to $\tilde{\mu}_n$ is thus the ratio of the variances of the asymptotic distributions:

$$e_{\overline{X}_n, \tilde{\mu}_n} = \frac{\pi\sigma^2}{2\left(\sigma(1-p)+p\right)^2\left(1-p+p\sigma^2\right)}.$$

(Note that for $p = 0$ or $p = 1$, we have a normal distribution and the ARE of the sample mean relative to the sample median is $\pi/2$. For fixed $p \in (0,1)$, the ARE approaches 0 as σ^2 becomes large. For fixed $\sigma^2 \ne 1$, the ARE achieves a local minimum somewhere between 0 and 1, with that minimum being above 1 for $\sigma^2 < 4.94$ or so. The larger the value of σ^2, the wider the range of p for which the sample median might be more efficient than the sample mean.)

5. (35 pt) Consider the "regression through the origin" model in which for known sequence of predictors $\{x_i\}_{i=1}^n$ and independent identically distributed "errors" $\{\epsilon_i\}_{i=1}^n$ with $\epsilon_i \sim (0, \sigma^2 < \infty)$, we observe derived random variables

$$Y_i = \beta x_i + \epsilon_i,$$

for some unknown parameter β. Let $\hat{\beta}$ be the ordinary least squares estimator of β, and let

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - x_i\hat{\beta})^2.$$

Show that for some suitable scaling factor a_n and under suitable restrictions on the x_i's

$$a_n \frac{\hat{\beta} - \beta}{\hat{\sigma}_n} \to_d \mathcal{N}(0,1).$$

(Be sure to specify the form of a_n and the sufficient restrictions on the x_i's, noting that in order to get maximal credit on this problem, the restrictions on the x_i's should be as least restrictive as possible.)

Ans: First, we find a_n by considering the general theory regarding ordinary least squares. We know that for $\vec{Y}|\vec{x} \sim (\mathbf{X}\beta, \sigma^2\mathbf{I})$, the OLS estimator $\widehat{\vec{\beta}} \sim (\beta, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1})$. Furthermore, if $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, then $\widehat{\vec{\beta}}$ is normally distributed: $\widehat{\vec{\beta}} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1})$. Because the general asymptotic results must include this case, we consider that

$$(\mathbf{X}^t\mathbf{X})^{\frac{1}{2}} \frac{\hat{\beta} - \beta}{\sigma} \quad \dot{\sim} \quad \mathcal{N}(0, 1)$$

as suggesting that we should use

$$a_n = (\mathbf{X}^t\mathbf{X})^{\frac{1}{2}}.$$

In this problem, $\mathbf{X} = \vec{x}$, so $\mathbf{X}^t\mathbf{X} = ss_{xx} = \sum_{i=1}^n x_i^2$, so we consider $a_n = \sqrt{ss_{xx}}$. Further,

$$\hat{\beta}_n = \frac{ss_{xy}}{ss_{xx}} = \frac{\sum_{i=1}^n x_i Y_i}{ss_{xx}},$$

hence

$$\begin{aligned}
W_n = a_n(\hat{\beta}_n - \beta) &= \sqrt{ss_{xx}} \left(\frac{\sum_{i=1}^n x_i Y_i}{ss_{xx}} - \beta \right) \\
&= \sqrt{ss_{xx}} \left(\frac{\sum_{i=1}^n x_i (x_i\beta + \epsilon_i)}{ss_{xx}} - \beta \right) \\
&= \sqrt{ss_{xx}} \left(\frac{ss_{xx}\beta}{ss_{xx}} + \frac{\sum_{i=1}^n x_i\epsilon_i}{ss_{xx}} - \beta \right) \\
&= \frac{\sum_{i=1}^n x_i\epsilon_i}{\sqrt{ss_{xx}}} \\
&\equiv \sum_{i=1}^n c_{n:i}\epsilon_i \\
&\equiv \sum_{i=1}^n W_{n:i}
\end{aligned}$$

where independently distributed $W_{n:i}$ have mean 0 and variance $v_{n:i} = c_{n:i}^2\sigma^2$, and

$$V_n = \sum_{i=1}^n v_{n:i} = \sum_{i=1}^n \frac{x_i^2}{ss_{xx}}\sigma^2 = \sigma^2.$$

We thus use the Lindeberg-Feller central limit theorem by requiring that the Lindeberg condition is satisfied:

$$\forall M > 0 \quad \lim_{n\to\infty} \frac{1}{V_n} \sum_{i=1}^n E\left[|W_{n:i}|^2 \, 1_{[\,|W_{n:i}| > M\sqrt{V_n}\,]} \right] \to 0.$$

We thus define $c_n^* = \max\{c_{n:i}\}$, and place a restriction on the $\{x_i\}$ such that $c_n^* \to 0$ as $n \to \infty$. Then,

$$\begin{aligned}
\frac{1}{V_n}\sum_{i=1}^n E\left[\,|W_{n:i}|^2\, 1_{[\,|W_{n:i}|>M\sqrt{V_n}\,]}\right] &= \frac{1}{V_n}\sum_{i=1}^n E\left[\,c_{n:i}^2|\epsilon_i^2|\, 1_{[\,|\epsilon_i|>M\sqrt{V_n}/c_{n:i}\,]}\right] \\
&\le \frac{1}{V_n}\sum_{i=1}^n E\left[\,c_{n:i}^2|\epsilon_i^2|\, 1_{[\,|\epsilon_i|>M\sqrt{V_n}/c_n^*\,]}\right] \\
&= \frac{1}{V_n}\sum_{i=1}^n c_{n:i}^2 E\left[\,|\epsilon_i^2|\, 1_{[\,|\epsilon_i|>M\sqrt{V_n}/c_n^*\,]}\right] \\
&= \frac{1}{V_n} E\left[\,|\epsilon_1^2|\, 1_{[\,|\epsilon_1|>M\sqrt{V_n}/c_n^*\,]}\right]\sum_{i=1}^n c_{n:i}^2 \\
&= \frac{1}{V_n} E\left[\,|\epsilon_1^2|\, 1_{[\,|\epsilon_1|>M\sqrt{V_n}/c_n^*\,]}\right] \\
&\to 0
\end{aligned}$$

where the last step follows from the finite variance for the ϵ_i's and the fact that $c_n^* \to 0$ (so $M\sqrt{V_n}/c_n^* \to \infty$).

We have thus shown

$$a_n\frac{\hat\beta-\beta}{\sigma} \to_d \mathcal{N}(0,1).$$

We now need to show that $\hat\sigma_n \to_p \sigma$, in which case $\sigma/\hat\sigma_n \to_p 1$ and Slutsky's will provide that

$$a_n\frac{\hat\beta-\beta}{\hat\sigma_n} = \frac{\sigma}{\hat\sigma_n}a_n\frac{\hat\beta-\beta}{\sigma} \to_d 1\cdot\mathcal{N}(0,1) \;\sim\; \mathcal{N}(0,1).$$

Now, by the weak law of large numbers we know

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

But we also have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n (Y_i - x_i\beta)^2 &= \frac{1}{n}\sum_{i=1}^n (Y_i - x_i\hat\beta + x_i\hat\beta - x_i\beta)^2 \\
&= \frac{1}{n}\sum_{i=1}^n (Y_i - x_i\hat\beta)^2 + \frac{2}{n}\sum_{i=1}^n (Y_i - x_i\hat\beta)(x_i\hat\beta - x_i\beta) + \frac{1}{n}\sum_{i=1}^n (x_i\hat\beta - x_i\beta)^2 \\
&= \frac{1}{n}\sum_{i=1}^n (Y_i - x_i\hat\beta)^2 + \frac{2(\hat\beta-\beta)}{n}\sum_{i=1}^n (x_iY_i - x_i^2\hat\beta) + \frac{(\hat\beta-\beta)^2}{n}\sum_{i=1}^n x_i^2 \\
&= \frac{1}{n}\sum_{i=1}^n (Y_i - x_i\hat\beta)^2 + \frac{2(\hat\beta-\beta)}{n}\left[\sum_{i=1}^n x_iY_i - \hat\beta\sum_{i=1}^n x_i^2\right] + \frac{(\hat\beta-\beta)^2}{n}\sum_{i=1}^n x_i^2 \\
&= \frac{1}{n}\sum_{i=1}^n (Y_i - x_i\hat\beta)^2 + \frac{a_n^2(\hat\beta-\beta)^2}{n}
\end{aligned}$$

where the middle term drops out because $\hat\beta = ss_{xy}/ss_{xx}$.

Now, by the Mann-Wald theorem we have $a_n^2(\hat\beta-\beta)^2 \to_d [\mathcal{N}(0,1)]^2 \sim \chi^2$, and because $1/n \to 0$, Slutsky's theorem provides that

$$\frac{a_n^2(\hat\beta-\beta)^2}{n} \to_d 0\cdot\chi^2.$$

Hence using Slutsky's

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - x_i\hat{\beta})^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - x_i\hat{\beta})^2\right] - \frac{n}{n-1}\left[\frac{a_n^2(\hat{\beta}-\beta)^2}{n}\right]$$
$$\to_d 1\cdot\sigma^2 - 1\cdot 0 = \sigma^2,$$

and Mann-Wald then gives us $\hat{\sigma}_n \to_p \sigma$.