Probability and Stochastic Processes II
Lecture 2

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC632024.html

2024

II. Convergence

- various modes of convergence, and their relationships, were discussed in PSPI and we review these here but for random vectors and establish some useful results

Suppose $\mathbf{X}_n$ is a sequence of random vectors and $\mathbf{X}$ is a random vector all defined wrt probability space $(\Omega, \mathcal{A}, P)$ mapping into $\mathbb{R}^k$.

1. Convergence in Distribution (weak convergence): $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ if $\lim_{n\to\infty} P_{\mathbf{X}_n}(B) = P_{\mathbf{X}}(B)$ for every continuity set $B \in \mathcal{B}^k$ of $P_{\mathbf{X}}$ (if $P_{\mathbf{X}}(\partial B) = 0$ then B is a continuity set)

- for $k = 1$ then $\partial(a, b] = \{a, b\}$ and $(a, b]$ is a continuity set of P_X when $P_X(\{a\}) = P_X(\{b\}) = 0$ so a and b are continuity points of F_X

- also $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ iff $F_{\mathbf{X}_n}(\mathbf{x}) \to F_{\mathbf{X}}(\mathbf{x})$ at all continuity points for $F_{\mathbf{X}}$ (see van der Vaart (1998) Asymptotic Statistics)

Proposition II.1 (*Continuous mapping theorem for convergence in distribution*) If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ is continuous, then $g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$.

Proof: Suppose $B \in \mathcal{B}^l$ is a continuity set for the distribution of $g(\mathbf{X})$. Then

$$0 = P(g(\mathbf{X}) \in \partial B) = P(\mathbf{X} \in g^{-1}(\partial B)) \geq P(\mathbf{X} \in \partial g^{-1}B)$$

since $\partial g^{-1}B \subset g^{-1}(\partial B)$ for a continuous g (see below). Therefore $g^{-1}B$ is a continuity set for $\mathbf{X}$ which implies

$$\lim_{n\to\infty} P(g(\mathbf{X}_n) \in B) = \lim_{n\to\infty} P(\mathbf{X}_n \in g^{-1}B) = P(\mathbf{X} \in g^{-1}B) = P(g(\mathbf{X}) \in B)$$

which establishes the result. ■

- if $x \in \partial g^{-1}B$, by definition for any open set O with $x \in O$, then $O \cap g^{-1}B \neq \phi$ and $O \cap (g^{-1}B)^c = O \cap g^{-1}B^c \neq \phi$

- so for any open ball $B_\epsilon(g(x))$ then $B_\epsilon(g(x)) \cap B \neq \phi$ since $g^{-1}(B_\epsilon(g(x)) \cap B) = g^{-1}B_\epsilon(g(x)) \cap g^{-1}B \neq \phi$ as $x \in g^{-1}B_\epsilon(g(x))$ and $g^{-1}B_\epsilon(g(x))$ is open, similarly $B_\epsilon(g(x)) \cap B^c \neq \phi$ and so $g(x) \in \partial B$ which implies $x \in g^{-1}(\partial B)$

- CLT holds for random vectors, namely, $\mathbf{X}_1, \mathbf{X}_2, \ldots \overset{i.i.d.}{\sim} P$ with mean $\boldsymbol{\mu}$ and variance Σ, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i - \boldsymbol{\mu}\right) \xrightarrow{d} \mathbf{X} \sim N_k\left(0, \Sigma\right)$$

Example II.1

- if $\mathbf{X}_n \xrightarrow{d} \mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$, then we know $g(\mathbf{X}_n)$ is approximately distributed like $g(\mathbf{X})$ but this doesn't tell us the approximate distribution (but we can simulate from the distribution of $g(\mathbf{X})$ if we can simulate from the distribution of $\mathbf{X}$) ■

- recall from PSPI that $\mathbf{Y}_n \xrightarrow{d} \mathbf{c}$ iff $\mathbf{Y}_n \xrightarrow{P} \mathbf{c}$

Proposition II.2 If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $\mathbf{W}_n \xrightarrow{P} \mathbf{0}$, then $\mathbf{W}_n + \mathbf{X}_n \xrightarrow{d} \mathbf{X}$.

Proof: Let $\epsilon > 0$ and $\mathbf{x}$ be a cty point for $F_{\mathbf{X}}$. Then

$$\begin{aligned}
& P(\mathbf{W}_n + \mathbf{X}_n \leq \mathbf{x}) \\
= \; & P(\mathbf{X}_n \leq \mathbf{x} - \mathbf{W}_n, ||\mathbf{W}_n|| \leq \epsilon) + P(\mathbf{X}_n \leq \mathbf{x} - \mathbf{W}_n, ||\mathbf{W}_n|| > \epsilon) \\
\leq \; & P(\mathbf{X}_n \leq \mathbf{x} + \epsilon \mathbf{1}) + P(||\mathbf{W}_n|| > \epsilon) \text{ and} \\
& P(\mathbf{X}_n \leq \mathbf{x} - \epsilon \mathbf{1}, ||\mathbf{W}_n|| \leq \epsilon) \\
= \; & P(\mathbf{X}_n \leq \mathbf{x} - \epsilon \mathbf{1}) - P(\mathbf{X}_n \leq \mathbf{x} - \epsilon \mathbf{1}, ||\mathbf{W}_n|| > \epsilon) \\
\leq \; & P(\mathbf{W}_n + \mathbf{X}_n \leq \mathbf{x}).
\end{aligned}$$

We can choose ϵ s.t. $\mathbf{x} \pm \epsilon \mathbf{1}$ are cty points of $F_{\mathbf{X}}$. Now $P(\mathbf{X}_n \leq \mathbf{x} - \epsilon \mathbf{1}, ||\mathbf{W}_n|| > \epsilon) \leq P(||\mathbf{W}_n|| > \epsilon) \to 0$ and so

$$\begin{aligned}
F_{\mathbf{X}}(\mathbf{x} - \epsilon \mathbf{1}) & \leq \liminf P(\mathbf{W}_n + \mathbf{X}_n \leq \mathbf{x}) \\
& \leq \limsup P(\mathbf{W}_n + \mathbf{X}_n \leq \mathbf{x}) \leq F_{\mathbf{X}}(\mathbf{x} + \epsilon \mathbf{1})
\end{aligned}$$

and since ϵ can be chosen arbitrarily small and $\mathbf{x}$ is a cty point of $F_{\mathbf{X}}$, the result is proved. ■

Proposition II.3 If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{d} \mathbf{Y}$ (i) it is not generally true that $(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{d} (\mathbf{X}, \mathbf{Y})$ (ii) but if $\mathbf{Y}$ is degenerate at $\mathbf{c}$ then this is true.

Proof: (i) If $(X, Y)^t \sim N_2(\mathbf{0}, I)$, then $X \sim N(0, 1)$, $Y \sim N(0, 1)$. Now putting $X_n = X$, $Y_n = X$ for all n, then $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} Y$ but for all n

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim N_2\left(\mathbf{0}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right) \neq N_2(\mathbf{0}, I).$$

(ii) We have

$$P(||(\mathbf{X}_n, \mathbf{Y}_n) - (\mathbf{X}_n, \mathbf{c})|| > \epsilon) = P(||\mathbf{Y}_n - \mathbf{c}|| > \epsilon) \to \mathbf{0}$$

and so $\mathbf{W}_n = (\mathbf{X}_n, \mathbf{Y}_n) - (\mathbf{X}_n, \mathbf{c}) \xrightarrow{P} \mathbf{0}$ and clearly $(\mathbf{X}_n, \mathbf{c}) \xrightarrow{d} (\mathbf{X}, \mathbf{c})$. Therefore, by Prop. II.2 $(\mathbf{X}_n, \mathbf{Y}_n) = \mathbf{W}_n + (\mathbf{X}_n, \mathbf{c}) \xrightarrow{d} (\mathbf{X}, \mathbf{c})$. ■

Corollary II.4 (*Slutsky's Theorem*) If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{d} \mathbf{c}$, then (i) $\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{d} \mathbf{X} + \mathbf{c}$ (ii) $\mathbf{Y}_n^t \mathbf{X}_n \xrightarrow{d} \mathbf{c}^t \mathbf{X}$. (iii) If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $Y_n \xrightarrow{d} c$, then provided $c \neq 0$, $\mathbf{X}_n / Y_n \xrightarrow{d} \mathbf{X}/c$.

Proof: The functions $g(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \mathbf{y}, g(\mathbf{x}, \mathbf{y}) = \mathbf{y}^t \mathbf{x}$ are continuous and $g(x, y) = \mathbf{x}/y$ is continuous provided $y \neq 0$. The continuous mapping theorem then gives the result. ■

Definition A sequence $\{\mathbf{X}_n : n \in \mathbb{N}\}$ of random vectors is *bounded in probability* if for every $\epsilon > 0$ there is a constant M such that

$$\sup_n P(||\mathbf{X}_n|| > M) < \epsilon.$$

Proposition II.5 If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$, then $\{\mathbf{X}_n : n \in \mathbb{N}\}$ is bounded in probability.

Proof: Let $\epsilon > 0$. Choose M such that $P(||\mathbf{X}|| \geq M) < \epsilon/2$ and such that $\{||\mathbf{X}|| \geq M\}$ is a continuity set for $P_{\mathbf{X}}$. Then, since

$$\lim_{n\to\infty} P(||\mathbf{X}_n|| \geq M) = P(||\mathbf{X}|| \geq M),$$

there exists n_ϵ such that for all $n > n_\epsilon$ then $P(||\mathbf{X}_n|| \geq M) < \epsilon$. Now choose $M' > M$ and such that $P(||\mathbf{X}_n|| \geq M') < \epsilon$ for $n = 1, \ldots, n_\epsilon$. This implies that $\sup_n P(||\mathbf{X}_n|| > M') < \epsilon$ and the result is proven. ■

Proposition II.6 (*The delta theorem*) Suppose $r_n(\mathbf{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{X}$ for some real sequence $r_n \to \infty$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ is continuously differentiable at $\boldsymbol{\mu}$ with derivative

$$G(\boldsymbol{\mu}) = \left(\left. \frac{\partial g_i(\mathbf{x})}{\partial x_j} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right) \in \mathbb{R}^{l \times k},$$

then $r_n(g(\mathbf{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{d} G(\boldsymbol{\mu})\mathbf{X}$.

Proof: By Lemma II.2 the sequence $r_n(\mathbf{X}_n - \boldsymbol{\mu})$ is bounded in probability and so for given $\delta > 0$ there exists M such that $P(||\mathbf{X}_n - \boldsymbol{\mu}|| > M/r_n) < \delta$ for every n. This implies that $\lim_{n\to\infty} P(||\mathbf{X}_n - \boldsymbol{\mu}|| > \epsilon) = 0$ so $\mathbf{X}_n \xrightarrow{P} \boldsymbol{\mu}$. Since g is continuously differentiable at $\boldsymbol{\mu}$ this implies $G(\mathbf{X}_n) \xrightarrow{P} G(\boldsymbol{\mu})$. Taking the first term of a Taylor expansion with remainder of g about $\boldsymbol{\mu}$ gives

$$g(\mathbf{X}_n) = g(\boldsymbol{\mu}) + G(\boldsymbol{\mu}^*(\mathbf{X_n}))(\mathbf{X}_n - \boldsymbol{\mu})$$

for some point $\boldsymbol{\mu}^*(\mathbf{X_n})$ on the line segment joining $\mathbf{X}_n$ to $\boldsymbol{\mu}$. Then $||\boldsymbol{\mu}^*(\mathbf{X_n}) - \boldsymbol{\mu}|| \leq ||\mathbf{X}_n - \boldsymbol{\mu}||$ so

$$P(||\boldsymbol{\mu}^*(\mathbf{X_n}) - \boldsymbol{\mu}|| > \epsilon) \leq P(||\mathbf{X_n} - \boldsymbol{\mu}|| > \epsilon) \to 0$$

as $n \to \infty$, then $\boldsymbol{\mu}^*(\mathbf{X_n}) \xrightarrow{P} \boldsymbol{\mu}$ which implies $G(\boldsymbol{\mu}^*(\mathbf{X_n})) \xrightarrow{P} G(\boldsymbol{\mu})$. Therefore, by Slutsky's Theorem (generalized)

$$r_n(g(\mathbf{X}_n) - g(\boldsymbol{\mu})) = G(\boldsymbol{\mu}(\mathbf{X_n}))r_n(\mathbf{X}_n - \boldsymbol{\mu}) \xrightarrow{d} G(\boldsymbol{\mu})\mathbf{X}.$$

■

Corollary II.7 (*Asymptotic normality)* Suppose $r_n(\mathbf{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{X} \sim N_k(\mathbf{0}, \Sigma)$, then for g satisfying the delta theorem

$$r_n(g(\mathbf{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{d} N_l(\mathbf{0}, G(\boldsymbol{\mu})\Sigma G^t(\boldsymbol{\mu})).$$

Example II.2

- suppose $X_1, \ldots, X_n$ is an $i.i.d.$ sequence from a distribution with mean μ and variance σ^2

- then by the CLT $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

- let $g(x) = \cos x$ so $G(x) = \sin x$ and by the delta theorem $\sqrt{n}(\cos(\bar{X}) - \cos(\mu)) \xrightarrow{d} N(0, (\sin \mu)^2 \sigma^2)$

- or $g(x) = x^2$ with $G(x) = 2x$ so by the delta theorem $\sqrt{n}(\bar{X}^2 - \mu^2) \xrightarrow{d} N(0, 4\mu^2\sigma^2)$ ■

- **note** - if $G(\boldsymbol{\mu}) = \mathbf{0}$ the delta theorem is still valid but not very useful as the approximating distribution is degenerate

- in such a case an approximation can be worked out based on a higher order Taylor expansion, for example, in Example II.2 with $\mu = 0$, then $\sqrt{n}(\bar{X}^2 - \mu^2) = \sqrt{n}\bar{X}^2 \xrightarrow{d} 0$ but

$$n\bar{X}^2 = (\sqrt{n}\bar{X})^2 = \sigma^2(\sqrt{n}\bar{X}/\sigma)^2 \xrightarrow{d} \sigma^2 W$$

where $W \sim$ chi-squared(1)

2. Convergence in Probability: $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(||\mathbf{X}_n - \mathbf{X}|| > \epsilon) = 0$$

where $||\mathbf{x}|| = \left\{\sum_{i=1}^{k} x_i^2\right\}^{1/2}$ is the Euclidean norm on $\mathbb{R}^k$

Proposition II.8 $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ iff $X_{in} \xrightarrow{P} X_i$ for each $i = 1, \ldots, k$.

Proof: Suppose $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$, then $P(|X_{in} - X_i| > \epsilon) \le P(||\mathbf{X}_n - \mathbf{X}|| > \epsilon)$ which implies $X_{in} \xrightarrow{P} X_i$. Now suppose $X_{in} \xrightarrow{P} X_i$ for each $i = 1, \ldots, k$. Then

$$\begin{aligned} & P(||\mathbf{X}_n - \mathbf{X}|| > \epsilon) \le P(\max\{|X_{1n} - X_1|, \ldots, |X_{kn} - X_k|\} > \epsilon/\sqrt{k}) \\ = \; & P(\cup_{i=1}^{k}\{|X_{in} - X_i| > \epsilon/\sqrt{k}\}) \le \sum_{i=1}^{k} P(|X_{in} - X_i| > \epsilon/\sqrt{k}) \to 0 \end{aligned}$$

as $n \to \infty$. ■

Proposition II.9 (*Continuous mapping theorem for convergence in probability*) If $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ is continuous, then $g(\mathbf{X}_n) \xrightarrow{P} g(\mathbf{X})$.

Proof: Let $\epsilon, \epsilon' > 0$. Since $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ we have $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and so $\{\mathbf{X}_n\}$ is bounded in probability by Proposition II.2. So there exists M such that $P(||\mathbf{X}_n|| > M) < \epsilon'/2$ for every n and $P(||\mathbf{X}|| > M) < \epsilon'/2$. Now $C = \{\mathbf{x} : ||\mathbf{x}|| \leq M\}$ is compact and so there exists $\delta > 0$ s.t. $||\mathbf{x} - \mathbf{y}|| < \delta$ implies $||g(\mathbf{x}) - g(\mathbf{y})|| < \epsilon$ for all $\mathbf{x}, \mathbf{y} \in C$.Therefore,

$$\begin{aligned}
& P(||g(\mathbf{X}_n) - g(\mathbf{X})|| > \epsilon) \\
= \;& P(||g(\mathbf{X}_n) - g(\mathbf{X})|| > \epsilon \text{ and } \mathbf{X}_n, \mathbf{X} \in C) + \\
& P(||g(\mathbf{X}_n) - g(\mathbf{X})|| > \epsilon \text{ and } \mathbf{X}_n \text{ or } \mathbf{X} \notin C) \\
\leq \;& P(||\mathbf{X}_n) - \mathbf{X}|| \geq \delta) + P(||\mathbf{X}_n|| > M) + P(||\mathbf{X}|| > M) \\
\leq \;& P(||\mathbf{X}_n) - \mathbf{X}|| \geq \delta) + \epsilon' \\
\to \;& \epsilon' \text{ as } n \to \infty
\end{aligned}$$

and this implies the result. ■

3. Convergence with Probability 1 *(convergence almost surely)*:
$\mathbf{X}_n \stackrel{wp1}{\to} \mathbf{X}$ if $P(\lim_{n\to\infty} \mathbf{X}_n = \mathbf{X}) = 1$

Proposition II.10 (*Continuous mapping theorem for convergence wp1)* If $\mathbf{X}_n \stackrel{wp1}{\to} \mathbf{X}$ and $g : \mathbb{R}^k \to \mathbb{R}^l$ is continuous, then $g(\mathbf{X}_n) \stackrel{wp1}{\to} g(\mathbf{X})$.

4. Convergence in mean of order r : $\mathbf{X}_n \xrightarrow{r} \mathbf{X}$ (also denoted as $\mathbf{X}_n \xrightarrow{L^r} \mathbf{X}$) if $\lim_{n\to\infty} E\left(\sum_{i=1}^{k} |X_{in} - X_i|^r\right) = 0$

- $r = 1$ (*convergence in mean*) and $r = 2$ (*convergence in mean square*) are the most important cases

Proposition II.11 For both convergence *wp*1 and convergence in mean of order r convergence of $\mathbf{X}_n$ to $\mathbf{X}$ occurs iff X_{in} converges to X_i for all $i = 1, \ldots, k$.

5. *Relationships*

- as for r.v.'s we have

$$\mathbf{X}_n \stackrel{wp1}{\to} \mathbf{X} \implies \mathbf{X}_n \stackrel{P}{\to} \mathbf{X} \implies \mathbf{X}_n \stackrel{d}{\to} \mathbf{X}$$

$$r > s \text{ then } \mathbf{X}_n \stackrel{r}{\to} \mathbf{X} \implies \mathbf{X}_n \stackrel{s}{\to} \mathbf{X}$$

$$\mathbf{X}_n \stackrel{1}{\to} \mathbf{X} \implies \mathbf{X}_n \stackrel{P}{\to} \mathbf{X}$$

$$\mathbf{X}_n \stackrel{d}{\to} \mathbf{X} \not\Rightarrow \mathbf{X}_n \stackrel{P}{\to} \mathbf{X} \not\Rightarrow \mathbf{X}_n \stackrel{wp1}{\to} \mathbf{X}$$

$$\mathbf{X}_n \stackrel{1}{\to} \mathbf{X} \not\Rightarrow \mathbf{X}_n \stackrel{wp1}{\to} \mathbf{X}$$

Example II.3 $\mathbf{X}_n \xrightarrow{P} \mathbf{X} \nRightarrow \mathbf{X}_n \xrightarrow{wp1} \mathbf{X}$ **and** $\mathbf{X}_n \xrightarrow{1} \mathbf{X} \nRightarrow \mathbf{X}_n \xrightarrow{wp1} \mathbf{X}$

- let $\omega \sim \text{Uniform}(0,1)$

- define
$X_1 = I_{[0,1/2]}(\omega), X_2 = I_{(1/2,1]}(\omega)$
$X_3 = I_{[0,1/4]}(\omega), X_4 = I_{(1/4,1/2]}(\omega), X_5 = I_{(1/2,3/4]}(\omega), X_6 = I_{(3/4,1]}(\omega)$
$\vdots$

- then $P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) \to 0$ as $n \to \infty$ so $X_n \xrightarrow{P} 0$

- but $\lim_{n\to\infty} X_n(\omega)$ doesn't exist as for any $\omega \in [0,1]$ and N there is $n_0, n_1 > N$ s.t. $X_{n_0}(\omega) = 0, X_{n_1}(\omega) = 1$ so convergence $wp1$ doesn't hold

- also $E(|X_n - 0|) = E(X_n) = P(X_n = 1) \to 0$ as $n \to \infty$ ■

Exercises

II.1 Prove that if $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ then $F_{\mathbf{X}_n}(\mathbf{x}) \to F_{\mathbf{X}}(\mathbf{x})$ at all continuity points $\mathbf{x}$ for $F_{\mathbf{X}}$.

II.2 Prove that if $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ then $X_{in} \xrightarrow{d} X_i$.

II.3 Suppose $X_1, \ldots, X_n$ is an $i.i.d.$ sequence from a distribution with finite first 4 moments $\mu_1, \mu_2, \mu_3, \mu_4$. Prove that

$$\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i - \mu_1 \\ \frac{1}{n}\sum_{i=1}^n X_i^2 - \mu_2 \end{pmatrix} \xrightarrow{d} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix}\right).$$

Using the function $g(x, y) = y - x^2$ determine the asymptotic disribution of $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$ and express the asymptotic variance in terms of the 4th central moment.

II.4 Prove Proposition II.10.

II.5 Prove Proposition II.11.