

Math Review Notes—Asymptotics and Convergence

Gregory Faletto

Contents

1	Asymptotics and Convergence	3
1.1	Preliminaries (5.9 and 7.1, Grimmett and Stirzaker)	3
1.2	Inequalities (8.6 of Pesaran)	4
1.3	Modes of Convergence (7.2 of Grimmett and Stirzaker, 8.2 and 8.4 of Pesaran)	8
1.4	More on convergence (7.2 of Grimmett and Stirzaker)	10
1.4.1	Slutsky's Convergence Theorems (8.4.1 of Pesaran, 7.3 of Grimmett and Stirzaker)	14
1.5	Stochastic orders $\mathcal{O}_p(\cdot)$ and $o_p(\cdot)$ (Pesaran 8.5)	15
1.6	Laws of Large Numbers and Central Limit Theorems (Pesaran 8.6; Grimmett and Stirzaker 7.4, 7.5)	15
1.7	The case of dependent and heterogeneously distributed observations (Pesaran 8.8)	17
1.8	Worked Examples from Math 505A Midterm 2	17

Last updated January 19, 2019

1 Asymptotics and Convergence

These notes are based on my notes from chapter 8 of *Time Series and Panel Data Econometrics* (1st edition) by M. Hashem Pesaran and coursework for Economics 613: Economic and Financial Time Series I at USC, as well as Math 505A and Math 541A at USC and chapter 7 from *Probability and Random Processes* (Grimmett and Stirzaker) 3rd edition.

1.1 Preliminaries (5.9 and 7.1, Grimmett and Stirzaker)

Definition 1.1. Definition 7.1.4, Grimmett and Stirzaker. If for all $x \in [0, 1]$ the sequence $\{f_n(x)\}$ of real numbers satisfies $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ then we say $f_n \rightarrow f$ **pointwise**.

Remark. In practice pointwise convergence is often not useful for functions because a sequence of functions may be continuous while its limit is not. For instance, consider $\{f_n : f_n = x^n \forall x \in [0, 1]\}$. Then f_n is continuous for all n but

$$\lim_{n \rightarrow \infty} f_n = \begin{cases} 0 & x < 1 \\ 1 & x = 1 \end{cases}$$

Instead, the following definition is often more useful.

Definition 1.2. (from class notes.) We say that f_n **uniformly converges to f on $[a, b]$** if for every $\epsilon > 0$ there exists N such that for every $n > N$,

$$\forall x \in [a, b] \quad |f_n(x) - f(x)| < \epsilon$$

Definition 1.3. (Definition 7.1.5, Grimmett and Stirzaker.) Let V be a collection of functions mapping $[0, 1]$ into \mathbb{R} and assume V is endowed with a function $\|\cdot\| : V \rightarrow \mathbb{R}$ satisfying

- (a) $\|f\| \geq 0$ for all $f \in V$
- (b) $\|f\| = 0$ if and only if f is the zero function (or equivalent to it)
- (c) $\|af\| = |a| \cdot \|f\|$ for all $a \in \mathbb{R}$, $f \in V$
- (d) $\|f + g\| \leq \|f\| + \|g\|$ (triangle inequality)

The function $\|\cdot\|$ is called a **norm**. If $\{f_n\}$ is a sequence of members of V then we say that $f_n \rightarrow f$ **with respect to the norm $\|\cdot\|$** if $\|f_n - f\| \rightarrow 0$ as $n \rightarrow \infty$.

Definition 1.4. (Definition 7.16, Grimmett and Stirzaker.) Let $\epsilon > 0$ be prescribed, and define the distance between two functions $g, h : [0, 1] \rightarrow \mathbb{R}$ by

$$d_\epsilon(g, h) = \int_E dx$$

where $E = \{u \in [0, 1] : |g(u) - h(u)| > \epsilon\}$. We say that $f_n \rightarrow f$ **in measure** if

$$d_\epsilon(f_n, f) \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for all } \epsilon > 0$$

Theorem 1. Inversion Theorem (Theorem 5.9.2, Grimmett and Stirzaker). Let X have distribution function F and characteristic function ϕ . Define $\bar{F} : \mathbb{R} \rightarrow [0, 1]$ by

$$\bar{F}(x) = \frac{1}{2} [F(x) + \lim_{y \rightarrow x^-} F(y)]$$

Then

$$\bar{F}(b) - \bar{F}(a) = \lim_{N \rightarrow \infty} \int_{-N}^N \frac{\exp(-iat) - \exp(-ibt)}{2\pi it} \cdot \phi(t) dt$$

Proof. See Kingman and Taylor (1966). □

Corollary 1.1. Corollary 5.9.3. Random variables X and Y have the same characteristic function if and only if they have the same distribution function.

Proof. Available in Grimmett and Stirzaker section 5.9, pp. 189 - 190. □

Definition 1.5. (Definition 5.9.4, Grimmett and Stirzaker.) We say that the sequence F_1, F_2, \dots of distribution functions **converges** to the distribution function F (written $F_n \rightarrow F$) if $F(x) = \lim_{n \rightarrow \infty} F_n(x)$ at each point x where F is continuous.

Theorem 2. Continuity theorem (Theorem 5.9.5; in notes from Friday 10/26, Lecture 28). Suppose that F_1, F_2, \dots is a sequence of distribution functions with corresponding characteristic functions ϕ_1, ϕ_2, \dots

- (a) If $F_n(x) \rightarrow F(x)$ for some distribution function F with characteristic function ϕ (at x where F is continuous), then $\phi_n(t) \rightarrow \phi(t)$ for all t .
- (b) Conversely, if $\phi(t) = \lim_{n \rightarrow \infty} \phi_n(t)$ exists and $\phi(t)$ is continuous at $t = 0$, then ϕ is the characteristic function of some distribution function F , and $F_n \rightarrow F$.

Proof. See Kingman and Taylor (1966). □

1.2 Inequalities (8.6 of Pesaran)

Inequalities

- Probabilities

—

Lemma 3. Markov's Inequality (Grimmett and Stirzaker p. 311, 319) : Let $X : \Omega \rightarrow [-\infty, \infty]$ be a random variable. Then for all $a > 0$,

$$\Pr(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

Proof. Note $a \cdot \mathbf{1}_{\{|X| \geq a\}} \leq |X|$, where $\mathbf{1}$ is the indicator function. Dividing both sides by a and taking expectations yields the result. \square

Corollary 3.1. If n is a positive integer, then

$$\Pr(|X| \geq t) \leq \frac{\mathbb{E}(|X|^n)}{t^n} \quad \forall t > 0$$

Proof. By Markov's Inequality (Theorem 3),

$$\Pr(|X| \geq t) = \Pr(|X|^n \geq t^n) \leq \frac{\mathbb{E}(|X|^n)}{t^n}$$

\square

Theorem 4. Chebyshev's Inequality: (probability p. 319) Let $X : \Omega \rightarrow [-\infty, \infty]$ be an (integrable) random variable with $\mathbb{E}(X^2) < \infty$. Then for any real number $k > 0$

$$\Pr(|X - \mathbb{E}(X)| \geq k\sqrt{\text{Var}(X)}) \leq \frac{1}{k^2}$$

This can also be written as

$$\Pr(|X - \mathbb{E}(X)| \geq k) \leq \frac{\text{Var}(X)}{k^2}$$

(Can be used to demonstrate consistency of estimators: if we can show that as $T \rightarrow \infty$ $\text{Var}(X) = \sigma^2 \rightarrow 0$, then this implies $\Pr(|X - \mu| \geq k\sigma) \rightarrow 0$ as $T \rightarrow \infty$, showing consistency.)

Theorem 5. Chernoff For $x \geq 0$, $a > 0$, $\forall t > 0$,

$$\Pr(X \geq a) = \Pr(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}}$$

• Moments

Theorem 6. Cauchy-Schwarz (and **Bunyakovsky**). If X and Y are random variables with finite variance then

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

Note that this can be a corollary of Theorem 10 with $p = q = 2$. We can also prove this theorem on its own in a different one. We first prove a useful result.

Lemma 7. If $\text{Var}(X) = 0$ then X is almost surely constant; that is, $\Pr(X = a) = 1$ for some $a \in \mathbb{R}$.

Proof. Note that because $\text{Var}(X) = 0 < \infty$, we know that $\mathbb{E}(X)$ and $\mathbb{E}(X^2)$ exist. We have

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = 0$$

Let $Y = (X - \mathbb{E}(X))^2$. Note that $Y = (X - \mathbb{E}(X))^2 \geq 0$ and that $\mathbb{E}(Y) = \text{Var}(X) = 0$. Therefore $\Pr(Y = 0) = 1$, so $\Pr(Y \neq 0) = 0$. To see why, in the case that X is discrete,

$$\mathbb{E}(Y) = \sum_{k=0}^{\infty} k \cdot \Pr(Y = k) = \text{Var}(X) = 0$$

which is true if and only if $\Pr(Y = k) = 0$ for all $k > 0$. Since we already showed that $\Pr(Y < 0) = 0$, it follows that $\Pr(Y = 0) = 1$. In the continuous case,

$$\mathbb{E}(Y) = \int_0^{\infty} y \cdot f_Y(y) dy = \text{Var}(X) = 0$$

which implies that $f_Y(x) = 0$ for all $x > 0$. Again, since $\Pr(Y < 0) = 0$, we have $\Pr(Y \neq 0) = 0$. But $Y = 0 \iff X = \mathbb{E}(X)$ so we have $\Pr(X = \mathbb{E}(X)) = 1$. \square

Remark. Note that Lemma 7 along with Proposition ?? imply that X has variance 0 if and only if it is (almost surely) constant.

We are now ready to prove the Cauchy-Schwarz Inequality.

Proof. if $\mathbb{E}(X^2) = 0$ or $\mathbb{E}(Y^2) = 0$, the Cauchy-Schwarz Inequality follows immediately. To see why, suppose without loss of generality that $\mathbb{E}(X^2) = 0$. Then the right side is 0. Also, $0 \leq \text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = -\mathbb{E}(X)^2$. Since $\mathbb{E}(X)^2 \geq 0$, we must have $\mathbb{E}(X)^2 = 0$ and therefore $\text{Var}(X) = 0$. Therefore by Lemma 7, X is almost surely constant, which means that $\text{Cov}(X, Y) = 0$.

In the case that $\mathbb{E}(X^2) > 0$ and $\mathbb{E}(Y^2) > 0$, for $a, b \in \mathbb{R}$, let $Z = aX - bY$. Then

$$0 \leq \mathbb{E}(Z^2) = a^2\mathbb{E}(X^2) - 2ab\mathbb{E}(XY) + b^2\mathbb{E}(Y^2) \quad (1)$$

The right side of (1) is quadratic in a . Because it is greater than or equal to zero, it has at most one real root, which means its discriminant must be non-positive. That is, if $b \neq 0$,

$$(-2b\mathbb{E}(XY))^2 - 4b^2\mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0 \iff \mathbb{E}(XY)^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$$

which yields the result. Note that equality holds if and only if $\Pr(aX = bY) = 1$ because the discriminant is zero if and only if the quadratic has a real root, which occurs if and only if

$$\mathbb{E}[(aX - bY)^2] = 0$$

which is true if and only if $\Pr(aX = bY) = 1$ by Lemma 7 and Proposition ??. \square

– **Krylov**

–

Definition 1.6. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$. We say that ϕ is **convex** if for any $x, y \in \mathbb{R}$ and for any $t \in [0, 1]$, we have

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y)$$

Proposition 8. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$. Then ϕ is convex if and only if for any $y \in \mathbb{R}$ there exists a constant a and there exists a function $L : \mathbb{R} \rightarrow \mathbb{R}$ defined by $L(x) = a(x - y) + \phi(y)$, $x \in \mathbb{R}$ such that $L(y) = \phi(y)$ and such that $L(x) \leq \phi(x)$ for all $x \in \mathbb{R}$.

Proof. Hint: Suppose ϕ is convex. If x is fixed and y varies, show that $\frac{\phi(y)-\phi(x)}{y-x}$ increases as y increases. Draw a picture. What slope a should L have at x ?

□

Theorem 9. Jensen's Inequality (Grimmett and Stirzaker p.181, 349) Let $X : \Omega \rightarrow [-\infty, \infty]$ be a random variable. Let $u : \mathbb{R} \rightarrow \mathbb{R}$ be convex. If $\mathbb{E}|X| < \infty$ and $\mathbb{E}|u(X)| < \infty$, then

$$u(\mathbb{E}(X)) \leq \mathbb{E}(u(X))$$

Proof. Hint: use Proposition 8 with $y := \mathbb{E}(X)$.

□

Corollary 9.1. Triangle Inequality. Let $X : \Omega \rightarrow [-\infty, \infty]$ be a random variable. If $\mathbb{E}|X| < \infty$, then $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Proof. Follows immediately from Jensen's Inequality using $u(X) = |X|$.

□

Corollary 9.2. If $\mathbb{E}(X^2) < \infty$, then $\mathbb{E}(|X|) < \infty$, so $\mathbb{E}(X) \in \mathbb{R}$.

Proof. Follows immediately from Jensen's Inequality using $u(X) = X^2$.

□

—

Theorem 10. Holder (Grimmett and Stirzaker p. p. 143, 319) Generalization of Cauchy-Schwarz. For $p, q > 1$ satisfying $1/p + 1/q = 1$ we have

$$\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|^p))^{1/p} (\mathbb{E}(|Y|^q))^{1/q}$$

Proof. Assume without loss of generality that $\|X\|_p = \|Y\|_q = 1$. Also, the case $p = 1, q = \infty$ follows from the triangle inequality, so we assume $1 < p < \infty$. From concavity of the log function, we have

$$\log((x^p)^{1/p}(y^q)^{1/q}) = (1/p)\log(x^p) + (1/q)\log(y^q)$$

$$\leq \log\left(\frac{1}{p}x^p + \frac{1}{q}y^q\right)$$

$$\implies (x^p)^{1/p}(y^q)^{1/q} \leq \frac{1}{p}x^p + \frac{1}{q}y^q$$

Fixing an $\omega \in \Omega$, we have

$$|X(\omega)Y(\omega)| = (|X(\omega)|^p)^{1/p}(|Y(\omega)|^q)^{1/q} \leq \frac{1}{p}|X(\omega)|^p + \frac{1}{q}|Y(\omega)|^q$$

Integrating we have...

□

—

Theorem 11. Minkowski (Grimmett and Stirzaker p. p. 143) For $p \geq 1$,

$$[\mathbb{E}(|X + Y|^p)]^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}$$

— Useful for showing lower order moments are finite (e.g. finite variance implies finite mean).

Lemma 12. Lyapunov's Inequality (Grimmett and Stirzaker p. 143). For $0 < r \leq s < \infty$,

$$\mathbb{E}(|X|^r)^{1/r} \leq \mathbb{E}(|X|^s)^{1/s}$$

—

Theorem 13. Triangle Inequality: Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Let $1 \leq p \leq \infty$. Then

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p, 1 \leq p \leq \infty$$

Proof. The case $p = \infty$ follows from the scalar triangle inequality, so assume $1 \leq p < \infty$. By scaling, we may assume $\|X\|_p = 1 - t, \|Y\|_p = t$, for some $t \in (0, 1)$ (zeroes and infinities being trivial). Define $V := X/(1 - t), W := Y/t$. Then by convexity of $x \rightarrow |x|^p$ on \mathbb{R} ,

$$|(1 - t)V(\omega) + t(W(\omega))|^p \leq (1 - t)|V(\omega)|^p + t|W(\omega)|^p$$

Take expectation of both sides:

$$\mathbb{E}|X + Y|^p \leq (1 - t)^{1-p} \mathbb{E}(|X|^p) + t^{1-p} \mathbb{E}(|Y|^p)$$

Since $\|X\|_p = t, \|Y\|_p = 1 - t$, we have that the right side is $1 - t + t = 1$. (Note: $\|Y\|_p = t, \mathbb{E}|Y|^p = t^p, \|X\|_p = 1 - t$ Therefore

$$(\mathbb{E}|X + Y|^p)^{1/p} = \|X + Y\|_p \leq 1$$

□

Monotone convergence theorem.

Dominated convergence theorem.

1.3 Modes of Convergence (7.2 of Grimmett and Stirzaker, 8.2 and 8.4 of Pesaran)

Let $\{X_n\} = \{X_1, X_2, \dots\}$ and X be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.7. Convergence in probability. $\{X_n\}$ is said to **converge in probability** to X if

- Grimmett and Strizaker definition:

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \epsilon) = 0, \text{ for every } \epsilon > 0$$

- Pesaran definition:

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| < \epsilon) = 1, \text{ for every } \epsilon > 0$$

This mode of convergence is also often denoted by $X_n \xrightarrow{p} X$ and when X is a fixed constant it is referred to as the **probability limit** of X_n , written as $Plim(X_n) = x$, as $n \rightarrow \infty$.

The above concept is readily extended to multivariate cases where $\{\mathbf{X}_n, n = 1, 2, \dots\}$ denote m -dimensional vectors of random variables. Then the condition is

$$\lim_{n \rightarrow \infty} \Pr(\|\mathbf{X}_n - \mathbf{X}\| < \epsilon) = 1, \text{ for every } \epsilon > 0$$

where $\|\cdot\|$ denotes an appropriate norm (say ℓ_2). Convergence in probability is often referred to as "weak convergence" (in contrast to convergence with probability 1, below).

Definition 1.8. Convergence with probability 1 or almost surely. The sequence of random variables $\{X_n\}$ is said to **converge with probability 1** (or **almost surely**) to X if

- (505A class notes definition)

$$\Pr\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

(Note: pointwise convergence can hardly ever be shown here and is not useful.)

- Grimmett and Stirzaker textbook definition:

$$\Pr\left(\left\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega) \text{ as } n \rightarrow \infty\right\}\right) = 1$$

- Pesaran textbook definition:

$$\Pr\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$$

This is often written as $X_n \xrightarrow{w.p.1} X$ or $X_n \xrightarrow{a.s.} X$. An equivalent condition for convergence with probability 1 is given by

$$\lim_{n \rightarrow \infty} \Pr(|X_m - X| < \epsilon, \text{ for all } m \geq n) = 1, \text{ for every } \epsilon > 0$$

which shows that convergence in probability is a special case of convergence with probability 1 (obtained by setting $m = n$). Convergence with probability 1 is stronger than convergence in probability and is often referred to as "strong convergence."

Definition 1.9. Convergence in r -th mean. $X_n \rightarrow X$ in r th mean where $r \geq 1$ if $\mathbb{E}|X_n^r| < \infty$ for all n and

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^r) = 0$$

Convergence in r th mean is often written $X_n \xrightarrow{r} X$.

Definition 1.10. Convergence in Distribution. Let X_1, X_2, \dots have distribution functions $F_1(\cdot), F_2(\cdot), \dots$ respectively. Then X_n is said to **converge in distribution to X** if

$$\lim_{n \rightarrow \infty} \Pr(X_n \leq u) = \Pr(X \leq u)$$

for all u at which $F_X(x) = \Pr(X \leq x)$ is continuous. This can also be written

$$\lim_{n \rightarrow \infty} F_n(u) = F(u)$$

for all u at which F is continuous. Convergence in distribution is usually denoted by $X_n \xrightarrow{d} X$, $X_n \xrightarrow{L} X$, or $F_n \Rightarrow F$. By the Continuity Theorem (section 1.1), this is equivalent to

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \phi_X(t), \quad t \in \mathbb{R}$$

Theorem 14. (Theorem 7.2.3, Grimmett and Stirzaker.) The following implications hold:

- $(X_n \xrightarrow{a.s.} X) \implies (X_n \xrightarrow{p} X)$
- $(X_n \xrightarrow{r} X) \implies (X_n \xrightarrow{p} X)$ for any $r \geq 1$
- $(X_n \xrightarrow{p} X) \implies (X_n \xrightarrow{d} X)$

Also, if $r > s \geq 1$, then $(X_n \xrightarrow{r} X) \implies (X_n \xrightarrow{s} X)$. No other implications hold in general.

Theorem 15. Some exceptions (Theorem 7.2.4).

- If $X_n \xrightarrow{d} c$ where c is constant, then $X_n \xrightarrow{p} c$.
- If $X_n \xrightarrow{p} X$ and $\Pr(|X_n| \leq k) = 0$ for all n and some k , then $X_n \xrightarrow{r} X$ for all $r \geq 1$.
- If $P_n(\epsilon) = \Pr(|X_n - X| > \epsilon)$ satisfies $\sum_n P_n(\epsilon) < \infty$ for all $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$.

Proof. (Part (c).) Let $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ (so that $P_n(\epsilon) = \Pr[A_n(\epsilon)]$), and let $B_m(\epsilon) = \bigcup_{n \geq m} A_n(\epsilon)$. Then

$$\Pr(B_m(\epsilon)) \leq \sum_{n=m}^{\infty} \Pr(A_n(\epsilon))$$

so $\lim_{m \rightarrow \infty} \Pr(B_m(\epsilon)) = 0$ whenever $\sum_n \Pr(A_n(\epsilon)) < \infty$. See also Lemma 17 part (b). \square

1.4 More on convergence (7.2 of Grimmett and Stirzaker)

Other theorems to include: Fatou's Lemma, Fubini's Theorem, Kolmogorov's Maximal Inequality, Kolmogorov Three-Series Test, Lindeberg Feller Central Limit Theorem, **this and more at beginning of Mike's 505A qual solutions.**

Definition 1.11. Cauchy Convergence. We say that the sequence $\{X_n : n \geq 1\}$ of random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is **almost surely Cauchy convergent** if

$$\Pr(\{\omega \in \Omega : X_m(\omega) - X_n(\omega) \rightarrow 0 \text{ as } m, n \rightarrow \infty\}) = 1$$

That is, the set of points ω of the sample space for which the real sequence $\{X_n(\omega) : n \geq 1\}$ is Cauchy convergent is an event having probability 1.

Lemma 16. (Lemma 7.2.6 from Grimmett and Stirzaker)

- (a) If $r > s \geq 1$ and $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{s} X$.
- (b) If $X_n \xrightarrow{1} X$ then $X_n \xrightarrow{p} X$.

The converse assertions fail in general.

Proof. (a) Using Lyapunov's Inequality (Lemma 12), if $r > s \geq 1$

$$[\mathbb{E}(|X_n - X|^s)]^{1/s} \leq [\mathbb{E}(|X_n - X|^r)]^{1/r}$$

Therefore if $X_n \xrightarrow{r} X$ (meaning $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^r) = 0$), (then $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^s) = 0$, so $X_n \xrightarrow{s} X$. We show the converse fails by counterexample:

$$X_n = \begin{cases} n & \text{with probability } n^{(-1/2)(r+s)} \\ 0 & \text{with probability } 1 - n^{(-1/2)(r+s)} \end{cases}$$

Then $\mathbb{E}|X_n^s| = n^{(1/2)(s-r)} \rightarrow 0$ and $\mathbb{E}|X_n^r| = n^{(1/2)(r-s)} \rightarrow \infty$.

- (b) By Markov's Inequality (Lemma 3),

$$\Pr(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}|X_n - X|}{\epsilon} \quad \text{for all } \epsilon > 0$$

Therefore if $X_n \xrightarrow{1} X$; that is, $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|) = 0$, then $\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \epsilon) = 0$ for every $\epsilon > 0$, so $X_n \xrightarrow{p} X$.

To see the converse fails, define an independent sequence $\{X_n\}$ by

$$X_n = \begin{cases} n^3 & \text{with probability } n^{-2} \\ 0 & \text{with probability } 1 - n^{-2} \end{cases}$$

Then $\Pr(|X| > \epsilon) = n^{-2}$ for all large n , and so $X_n \xrightarrow{p} 0$. However, $\mathbb{E}|X_n| = n \rightarrow \infty$.

□

Lemma 17. (Lemma 7.2.10, Grimmett and Stirzaker.) Let $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ and $B_m(\epsilon) = \cup_{n \geq m} A_n(\epsilon)$. Then:

- (a) $X_n \xrightarrow{a.s.} X$ if and only if $\Pr(B_m(\epsilon)) \rightarrow 0$ as $m \rightarrow \infty$ for all $\epsilon > 0$.

- (b) $X_n \xrightarrow{a.s.} X$ if $\sum_n \Pr(A_n(\epsilon)) < \infty$ for all $\epsilon > 0$.
(c) If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{p} X$, but the converse fails in general.

Proof. (a)

(b) As for Theorem 15 part (c).

(c) To see the converse fails, define an independent sequence $\{X_n\}$ by

$$X_n = \begin{cases} 1 & \text{with probability } n^{-1} \\ 0 & \text{with probability } 1 - n^{-1} \end{cases}$$

Clearly $X_n \xrightarrow{p} 0$. However, if $0 < \epsilon < 1$,

$$\Pr(B_m(\epsilon)) = 1 - \lim_{r \rightarrow \infty} \Pr(X_n = 0 \text{ for all } n \text{ such that } m \leq n \leq r) \text{ (by Lemma 1.3.5)}$$

$$= 1 - \left(1 - \frac{1}{m}\right) \left(1 - \frac{1}{m+1}\right) \cdots \text{ (by independence)}$$

$$= 1 - \lim_{M \rightarrow \infty} \left(\frac{m-1}{m} \cdot \frac{m}{m+1} \cdot \frac{m+1}{m+2} \cdots \frac{M}{M+1} \right)$$

$$= 1 - \lim_{M \rightarrow \infty} \frac{m-1}{M+1} = 1$$

and so $\{X_n\}$ does not converge almost surely. □

Lemma 18. (Lemma 7.2.12, Grimmett and Stirzaker.) There exist sequences which

- (a) converge almost surely but not in mean,
(b) converge in mean but not almost surely.

Proof. (a) As for Lemma 16 part (b). □

Theorem 19. (Theorem 7.2.13, Grimmett and Stirzaker.) If $X_n \xrightarrow{p} X$, there exists a non-random increasing sequence of integers n_1, n_2, \dots such that $X_{n_i} \xrightarrow{a.s.} X$ as $i \rightarrow \infty$.

Theorem 20. Skorokhod's representation theorem (Theorem 7.2.14, Grimmett and Stirzaker). If $\{X_n\}$ and X with distribution functions $\{F_n\}$ and F are such that $X_n \xrightarrow{d} X$ (or equivalently, $F_n \rightarrow F$) as $n \rightarrow \infty$, then there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random variables $\{Y_n\}$ and Y mapping Ω' into \mathbb{R} such that

- (a) $\{Y_n\}$ and Y have distribution functions $\{F_n\}$ and F
(b) $Y_n \xrightarrow{a.s.} Y$ as $n \rightarrow \infty$

Therefore, although X_n may fail to converge to X in any mode other than in distribution, there exists a sequence $\{Y_n\}$ such that Y_n is distributed identically to X_n for every n , which converges almost surely to a copy of X .

Theorem 21. (Theorem 7.2.19, Grimmett and Stirzaker; same as Portmanteau Theorem?)

The following three statements are equivalent:

- (a) $X_n \xrightarrow{d} X$
- (b) $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$ for all bounded continuous functions g .
- (c) $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$ for all functions g of the form $g(x) = f(x)\mathbf{1}_{[a,b]}(x)$ where f is continuous on $[a, b]$ and a and b are points of continuity of the distribution function of the random variable X .

Theorem 22. (Grimmett and Stirzaker Theorem 7.3.9.)

- (a) If $X_n \xrightarrow{a.s.} X$ and $Y_n \xrightarrow{a.s.} Y$ then $X_n + Y_n \xrightarrow{a.s.} X + Y$.
- (b) If $X_n \xrightarrow{r} X$ and $Y_n \xrightarrow{r} Y$ then $X_n + Y_n \xrightarrow{r} X + Y$.
- (c) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ then $X_n + Y_n \xrightarrow{p} X + Y$.
- (d) It is not in general true that $X_n + Y_n \xrightarrow{d} X + Y$ whenever $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$.

Theorem 23. Borel-Cantelli lemmas (Grimmett and Stirzaker Theorem 7.3.10.) Let $\{A_n\}$ be an infinite sequence of events from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $A = \bigcap_n \bigcup_{m=n}^{\infty} A_m = \limsup_{n \rightarrow \infty} A_n = \{A_n \text{ i.o.}\}$ be the event that infinitely many of the A_n occur. Then:

- (a) $\Pr(A) = 0$ if $\sum_n \Pr(A_n) < \infty$
- (b) $\Pr(A) = 1$ if $\sum_n \Pr(A_n) = \infty$ and A_1, A_2, \dots are independent events.

Proof. (a) We have that $A \subseteq \bigcup_{m=n}^{\infty} A_m$ for all n , so

$$\Pr(A) \leq \sum_{m=n}^{\infty} \Pr(A_m) \rightarrow 0 \text{ as } n \rightarrow \infty$$

whenever $\sum_n \Pr(A_n) < \infty$.

- (b) One can confirm that

$$A^c = \bigcup_n \bigcap_{m=n}^{\infty} A_m^c$$

But

$$\begin{aligned} \Pr\left(\bigcap_{m=n}^{\infty} A_m^c\right) &= \lim_{r \rightarrow \infty} \Pr\left(\bigcap_{m=n}^r A_m^c\right) = \prod_{m=n}^{\infty} [1 - \Pr(A_m)] \text{ (by independence)} \leq \prod_{m=n}^{\infty} \exp(-\Pr(A_m)) \\ &= \exp\left(-\sum_{m=n}^{\infty} \Pr(A_m)\right) = 0 \end{aligned}$$

whenever $\sum_n \Pr(A_n) = \infty$, where the fourth step follows since $1 - x \leq e^{-x}$ if $x \geq 0$. Thus

$$\Pr(A^c) = \lim_{n \rightarrow \infty} \Pr\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0$$

so $\Pr(A) = 1$. □

Theorem 24. Kolmogorov's Two-Series Theorem. Let X_1, X_2, \dots be independent random variables with $\mathbb{E}(X_n) = \mu_n$ and $\text{Var}(X_n) = \sigma_n^2$ such that $\sum_{n=1}^{\infty} \mu_n < \infty$ and $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. Then $\sum_{n=1}^{\infty} X_n$ converges in \mathbb{R} almost surely.

Proof. Available on wikipedia, https://en.wikipedia.org/wiki/Kolmogorov%27s_two-series_theorem. □

1.4.1 Slutsky's Convergence Theorems (8.4.1 of Pesaran, 7.3 of Grimmett and Stirzaker)

Theorem 25. Theorem 6 of Pesaran, Section 8.4.1, p. 173. Let $\{x_t, y_t\}, t = 1, 2, \dots$ be a sequence of pairs of random variables with $y_t \xrightarrow{d} y$ and $|y_t - x_t| \xrightarrow{p} 0$. Then $x_t \xrightarrow{d} y$.

Theorem 26. Theorem 7 in Pesaran, on p.318 (section 7.3) of Grimmett and Stirzaker. (Section 8.4.1, p. 174) If $x_t \xrightarrow{d} x$ and $y_t \xrightarrow{p} c$ where c is a finite constant, then

(i) $x_t + y_t \xrightarrow{d} x + c$

(ii) $y_t x_t \xrightarrow{d} cx$

(iii) $x_t/y_t \xrightarrow{d} x/c$, if $c \neq 0$.

Theorem 27. on p.318 (section 7.3) of Grimmett and Stirzaker. Suppose that $X_n \xrightarrow{d} 0$ and $Y_n \xrightarrow{p} Y$, and let $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ be such that $g(x, y)$ is a continuous function of y for all x , and $g(x, y)$ is continuous at $x = 0$ for all y . Then $g(X_n, Y_n) \xrightarrow{p} g(0, Y)$.

Theorem 28. Continuous Mapping Theorem (Theorem 9 of Pesaran, Section 8.4.1, p. 176: convergence properties of transformed sequences.) Suppose $\{x_t\}, \{y_t\}, \mathbf{x}$, and \mathbf{y} are $m \times 1$ vectors of random variables on a probability space, and let $\mathbf{g}(\cdot)$ be a continuous vector-valued function. (Alternatively, suppose \mathbf{g} has the set of discontinuity points D_g such that $\Pr(X \in D_g) = 0$.) Then

(i) $x_t \xrightarrow{a.s.} x \implies \mathbf{g}(x_t) \xrightarrow{a.s.} \mathbf{g}(x)$

(ii) $x_t \xrightarrow{p} x \implies \mathbf{g}(x_t) \xrightarrow{p} \mathbf{g}(x)$

(iii) $x_t \xrightarrow{d} x \implies \mathbf{g}(x_t) \xrightarrow{d} \mathbf{g}(x)$

(iv) $x_t - y_t \xrightarrow{p} \mathbf{0}$ and $y_t \xrightarrow{d} \mathbf{y} \implies \mathbf{g}(x_t) - \mathbf{g}(y_t) \xrightarrow{d} \mathbf{0}(x)$

Proof. See Serfling (1980) or Rao (1973). □

See also:

Theorem 29. (Theorem 7.2.18, Grimmett and Stirzaker.) If $X_n \xrightarrow{d} X$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ is continuous, then $g(X_n) \xrightarrow{d} g(X)$.

1.5 Stochastic orders $\mathcal{O}_p(\cdot)$ and $o_p(\cdot)$ (Pesaran 8.5)

Definition 1.12. (Pesaran 8.5 Definition 6.) Let $\{a_t\}$ be a sequence of positive numbers and $\{x_t\}$ be a sequence of random variables. Then

- (i) $x_t = \mathcal{O}_p(a_t)$, or x_t/a_t is bounded in probability, if for every $\epsilon > 0$ there exist real numbers M_ϵ and N_ϵ such that

$$\Pr \left(\frac{|x_t|}{a_t} > M_\epsilon \right) < \epsilon, \quad \text{for } t > N_\epsilon$$

- (ii) $x_t = o_p(a_t)$ if

$$\frac{x_t}{a_t} \xrightarrow{p} 0$$

1.6 Laws of Large Numbers and Central Limit Theorems (Pesaran 8.6; Grimmett and Stirzaker 7.4, 7.5)

Theorem 30. Weak Law of Large Numbers (Khinchine) (Pesaran 8.6 Theorem 10, Grimmett and Stirzaker Theorem 7.4.7). Suppose that $\{X_k\}$ is a sequence of (i) IID random variables with (ii) constant means, i.e., $\mathbb{E}(X_k) = \mu < \infty$. Then

$$\bar{X}_k = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{p} \mu$$

Theorem 31. Weak Law of Large Numbers (Chebyshev) (Pesaran Section 8.6, p. 178, Theorem 11.) Let $\{X_k\}$ be a sequence of random variables. If (i) $\mathbb{E}(X_k) = \mu_k$, (ii) $\text{Var}(X_k) = \sigma_k^2$, and (iii) $\text{Cov}(X_k, X_j) = 0$, $k \neq j$, and (iv)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sigma_k^2 < \infty$$

then we have $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$, where $\bar{\mu}_n = n^{-1} \sum_{k=1}^n \mu_k$.

Theorem 32. Strong Law of Large Numbers (Grimmett and Stirzaker Theorem 7.4.3). Let $\{X_k\}$ be a sequence of (i) independent (ii) identically distributed random variables with (iii) $\mathbb{E}(X_k) = \mu$ and (iv) $\mathbb{E}(X_k^2) < \infty$. Then

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mu \text{ almost surely and in mean square.}$$

Theorem 33. Strong Law of Large Numbers (Grimmett and Stirzakker Theorem 7.5.1). Let $\{X_k\}$ be a sequence of (i) independent (ii) identically distributed random variables. Then if and only if (iii) $\mathbb{E}|X_k| < \infty$,

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{a.s.} \mu$$

Theorem 34. Strong Law of Large Numbers 1 (Kolmogorov) (Pesaran 8.8 Theorem 12). Let $\{X_k\}$ be a sequence of (i) independent random variables with (ii) $\mathbb{E}(X_k) = \mu_k < \infty$ and (ii) $\text{Var}(X_k) = \sigma_k^2$ such that (iii)

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$$

Then $\bar{X}_n - \bar{\mu}_n \xrightarrow{wp1} 0$. If the independence assumption (i) is replaced by a lack of correlation (i.e. $\text{Cov}(X_k, X_j) = 0, k \neq j$), the convergence of $\bar{X}_n - \bar{\mu}_n$ with probability one requires the stronger condition

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2 (\log k)^2}{k^2} < \infty$$

Theorem 35. Strong Law of Large Numbers 2 (Pesaran 8.8 Theorem 13) Suppose that X_1, X_2, \dots are (i) independent random variables, and that (ii) $\mathbb{E}(X_k) = 0$, (iii) $\mathbb{E}(X_k^4) \leq M \forall k$ where M is an arbitrary positive constant. Then

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{a.s.} 0$$

Theorem 36. Central Limit Theorem (Grimmett and Stirzaker theorem 5.10.4.) Let X_1, X_2, \dots be a sequence of independent identically distributed random variables with finite mean μ and finite non-zero variance σ^2 , and let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Theorem 37. (Grimmett and Stirzaker theorem 5.10.5.) Let X_1, X_2, \dots be independent random variables satisfying $\mathbb{E}(X_j) = 0$, $\text{Var}(X_j) = \sigma_j^2$, $\mathbb{E}|X_j^3| < \infty$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma(n)^3} \sum_{j=1}^n \mathbb{E}|X_j^3| = 0$$

where $\sigma(n)^2 = \text{Var}(\sum_{j=1}^n X_j) = \sum_{j=1}^n \sigma_j^2$. Then

$$\frac{1}{\sigma(n)} \sum_{j=1}^n X_j \xrightarrow{d} \mathcal{N}(0, 1)$$

Proof. See Loeve (1977, p. 287) and Grimmett and Stirzaker Problem 5.12.40. □

Lemma 38. Lindeberg's Condition: Let $\{X_k\}$ be a sequence of independent (not necessarily identically distributed) random variables with expectations μ_k and finite variances σ_k^2 . Let $s_n^2 = \sum_{k=1}^n \sigma_k^2$. If such a sequence of independent random variables X_k satisfies the condition

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}[(X_k - \mu_k)^2 \cdot \mathbf{1}_{\{|X_k - \mu_k| > \epsilon s_n\}}] = 0$$

for all $\epsilon > 0$ then the central limit theorem holds; that is, the random variables

$$Z_n = \frac{1}{s_n} \sum_{k=1}^n (X_k - \mu_k)$$

converge in distribution to $\mathcal{N}(0, 1)$ as $n \rightarrow \infty$.

1.7 The case of dependent and heterogeneously distributed observations (Pesaran 8.8)

Theorem 39. Central limit theorem for martingale difference sequences (Pesaran 8.8 Theorem 28). Let $\{x_t\}$ be a martingale difference sequence with respect to the information set Ω_t . Let $\bar{\sigma}_T^2 = \text{Var}(\sqrt{T}\bar{x}_T) = T^{-1} \sum_{t=1}^T \sigma_t^2$. If $\mathbb{E}(|x_t|^r) < K < \infty$ for any $r > 2$ and for all t , and

$$\frac{1}{T} \sum_{t=1}^T x_t^2 - \bar{\sigma}_T^2 \xrightarrow{p} 0$$

then $\sqrt{T}\bar{x}_T/\bar{\sigma}_T \xrightarrow{d} \mathcal{N}(0, 1)$.

1.8 Worked Examples from Math 505A Midterm 2

- (1) (a) **Fall 2010 Problem 1.** Let X_k , $k \geq 1$, be i.i.d. random variables with mean 1 and variance 1. Show that the limit

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_k^2}$$

exists in an appropriate sense, and identify the limit.

- (b) **Not included on midterm or final.** Let $(X_j)_{j \geq 1}$ be i.i.d. uniform on $(-1, 1)$. Let

$$Y_n = \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2 + \sum_{j=1}^n X_j^3}$$

Prove that $\lim_{n \rightarrow \infty} \sqrt{n}Y_n$ exists in an appropriate sense, and identify the limit.

Solution.

(a)

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_k^2} = \lim_{n \rightarrow \infty} \frac{n^{-1} \sum_{k=1}^n X_k}{n^{-1} \sum_{k=1}^n X_k^2}$$

Since X_1, X_2, \dots are i.i.d., $E(X_1^2) = \text{Var}(X_1) + (\mathbb{E}(X_1))^2 = 2 < \infty$, we have

$$n^{-1} \sum_{k=1}^n X_k \xrightarrow{a.s.} \mathbb{E}(X_1) = 1 \text{ as } n \rightarrow \infty$$

by Theorem 32 (Strong Law of Large Numbers). Also, X_1^2, X_2^2, \dots are clearly identically distributed, and are independent by Theorem 4.2.3 (“If X and Y are independent, then so are $g(X)$ and $g(Y)$.”). It is clear also that $\mathbb{E}(|X_1^2|) = \mathbb{E}(X_1^2) = \text{Var}(X_1) + \mathbb{E}(X_1)^2 = 1 + 1 = 2 < \infty$. Therefore by Theorem 33 (Strong Law of Large Numbers),

$$n^{-1} \sum_{k=1}^n X_k^2 \xrightarrow{a.s.} \mathbb{E}(X_1^2) = 2 \text{ as } n \rightarrow \infty$$

(From here I had two different ways of finishing the problem.)

- Because we have almost sure convergence in the numerator and denominator, by the Continuous Mapping Theorem (Theorem 28),

$$\lim_{n \rightarrow \infty} \frac{n^{-1} \sum_{k=1}^n X_k}{n^{-1} \sum_{k=1}^n X_k^2} = \frac{\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n X_k}{\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n X_k^2} \xrightarrow{a.s.} \boxed{\frac{1}{2}}$$

- Then, using one of Slutsky’s convergence theorems (Theorem 26: “If $x_t \xrightarrow{d} x$ and $y_t \xrightarrow{p} c$ where c is a finite constant, then $x_t/y_t \xrightarrow{d} x/c$, if $c \neq 0$.”), we have

$$\frac{n^{-1} \sum_{k=1}^n X_k}{n^{-1} \sum_{k=1}^n X_k^2} \xrightarrow{d} \frac{\mathbb{E}(X_1)}{\mathbb{E}(X_1^2)} = \frac{\mathbb{E}(X_1)}{\text{Var}(X_1) + \mathbb{E}(X_1)^2} = \frac{1}{1+1} = \frac{1}{2}$$

But then, by Theorem 15 (Theorem 7.2.4(a) in Grimmett and Stirzaker: “If $X_n \xrightarrow{d} c$ where c is constant, then $X_n \xrightarrow{p} c$.”), we have $\frac{n^{-1} \sum_{k=1}^n X_k}{n^{-1} \sum_{k=1}^n X_k^2} \xrightarrow{p} 1/2$.

(b) **(Not included on midterm or final.)**

$$Y_n = \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2 + \sum_{j=1}^n X_j^3} = \frac{n^{-1} \sum_{j=1}^n X_j}{n^{-1} \sum_{j=1}^n X_j^2 + n^{-1} \sum_{j=1}^n X_j^3}$$

Note that $\mathbb{E}(X_1) = 0, \mathbb{E}(X_1^2) = \text{Var}(X_1) + \mathbb{E}(X_1)^2 = (1 - (-1))^2/12 + 0^2 = 1/3, \mathbb{E}(X_1^3) = (1/2) \int_{-1}^1 x^3 dx = 0$. (We derived the formulae for the first three moments of a uniform distribution on Homework 4 problem 2(2).)

$$\Rightarrow \sqrt{n}Y_n = \frac{\sqrt{1/3}(\sum_{j=1}^n X_j - n\mathbb{E}(X_1))/\sqrt{n \cdot 1/3}}{n^{-1} \sum_{j=1}^n X_j^2 + n^{-1} \sum_{j=1}^n X_j^3}$$

By the Central Limit Theorem (Theorem 36),

$$\frac{\sum_{j=1}^n X_j - n\mathbb{E}(X_1)}{\sqrt{n \cdot 1/3}} \xrightarrow{d} \mathcal{N}(0, 1)$$

By the Law of Large Numbers (Theorem 33), since $\mathbb{E}(|X_1^2|) = \mathbb{E}(X_1^2) = 1/3 < \infty$,

$$\frac{1}{n} \sum_{j=1}^n X_j^2 \xrightarrow{a.s.} \mathbb{E}(X_1^2) = 1/3$$

By the Law of Large Numbers (Theorem 33), since $\mathbb{E}(|X_1^3|) = (1/2) \int_{-1}^1 |x^3| dx = \int_0^1 x^3 dx = 1/4 < \infty$,

$$\frac{1}{n} \sum_{j=1}^n X_j^3 \xrightarrow{a.s.} \mathbb{E}(X_1^3) = 0$$

In the denominator, since we have almost sure convergence, the regular rules of calculus/real analysis apply. That is, using the above results,

$$n^{-1} \sum_{j=1}^n X_j^2 + n^{-1} \sum_{j=1}^n X_j^3 \xrightarrow{a.s.} 1/3$$

Therefore

$$\sqrt{n}Y_n = \frac{\sqrt{1/3}(\sum_{j=1}^n X_j - n\mathbb{E}(X_1))/\sqrt{n \cdot 1/3}}{n^{-1} \sum_{j=1}^n X_j^2 + n^{-1} \sum_{j=1}^n X_j^3} \xrightarrow{d} \frac{\sqrt{1/3}}{1/3} \mathcal{N}(0, 1) = \boxed{\mathcal{N}(0, 3)}$$

- (2) **Fall 2010 Problem 2.** Fix $p \in (0, 1)$ and consider independent Poisson random variables $X_k, k \geq 1$ with

$$\mathbb{E}X_k = \frac{p^k}{k}$$

Verify that the sum $\sum_{k=1}^{\infty} kX_k$ converges with probability one and determine the distribution of the random variable $Y = \sum_{k=1}^{\infty} kX_k$.

Solution. Melike's solution (use for midterm): We have $\mathbb{E}[kX_k] = p^k$ and $\sum_{k=1}^{\infty} p^k = p/(1-p) < \infty$, and $\text{Var}(kX_k) = kp^k$ and

$$\sum_{k=1}^{\infty} kp^k = p \sum_{k=1}^{\infty} kp^{k-1} = p \frac{d}{dp} \sum_{k=1}^{\infty} p^k = p \frac{d}{dp} \frac{p}{1-p} = p \cdot \frac{(1-p) - p(-1)}{(1-p)^2} = \frac{p}{(1-p)^2} < \infty$$

Since the sequence $\{Y_k\}_{k \geq 1}$ is independent, by Kolmogorov's Two Series Theorem (Theorem 24: "Let X_1, X_2, \dots be independent random variables with $\mathbb{E}(X_n) = \mu_n$ and $\text{Var}(X_n) = \sigma_n^2$ such that $\sum_{n=1}^{\infty} \mu_n < \infty$ and $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. Then $\sum_{n=1}^{\infty} X_n$ converges in \mathbb{R} almost surely."), we conclude that $\sum_{k=1}^{\infty} kX_k$ converges almost surely.

To find the distribution of Y , let X be a Poisson random variable and consider its probability generating function:

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}$$

So $\mathbb{E}(s^{X_k}) = \exp\left(\frac{p^k}{k}(s-1)\right)$ and $\mathbb{E}(s^{kX_k}) = \mathbb{E}[(s^k)^{X_k}] = \exp\left(\frac{p^k}{k}(s^k-1)\right)$. Then define $Y_n = \sum_{k=1}^n kX_k$ and consider

$$\begin{aligned}
G_{Y_n}(s) &= \mathbb{E}(s^{Y_n}) = \mathbb{E}\left(\prod_{k=1}^n s^{kX_k}\right) = \prod_{k=1}^n \mathbb{E}(s^{kX_k}) = \prod_{k=1}^n \exp\left(\frac{p^k}{k}(s^k - 1)\right) = \exp\left(\sum_{k=1}^n \frac{p^k}{k}(s^k - 1)\right) \\
&= \exp\left(\sum_{k=1}^n \frac{(ps)^k}{k} - \sum_{k=1}^n \frac{p^k}{k}\right)
\end{aligned}$$

Now, by taking limits as $n \rightarrow \infty$ (since we are allowed to take limit inside of expectation here), we get

$$\begin{aligned}
G_Y(s) &= \mathbb{E}(s^Y) = \exp\left(\sum_{k=1}^{\infty} \frac{(ps)^k}{k} - \sum_{k=1}^{\infty} \frac{p^k}{k}\right) = \exp\left(\int \sum_{k=1}^{\infty} (ps)^{k-1} dp - \int \sum_{k=1}^{\infty} p^{k-1} dp\right) \\
&= \exp\left(\int \frac{1}{1-ps} dp - \int \frac{1}{1-p} dp\right) = \exp(-\log(1-ps) + \log(1-p)), \quad -1 \leq ps < 1 \text{ and } -1 \leq p < 1 \\
&= \frac{1-p}{1-ps}, \quad -1 \leq ps < 1
\end{aligned}$$

Since we know $\Pr(X = k) = \frac{G_X^{(k)}(0)}{k!}$, we have

$$\begin{aligned}
G_Y(s) &= \frac{1-p}{1-sp}, \quad G'_Y(s) = \frac{p(1-p)}{(1-sp)^2}, \quad G''_Y(s) = \frac{2p^2(1-p)}{(1-sp)^3}, \quad G^{(3)}_Y(s) = \frac{3 \cdot 2p^3(1-p)}{(1-sp)^3}, \dots \\
G^{(k)}_Y(s) &= \frac{k!p^k(1-p)}{(1-sp)^k} \text{ for } k = 0, 1, 2, \dots
\end{aligned}$$

So we have

$$\Pr(Y = k) = (1-p)p^k, \quad k = 0, 1, 2, \dots$$

$$= \Pr(G_1(1-p) = k+1) = \Pr(G_1(1-p) - 1 = k)$$

which means $Y \sim G_1(1-p) - 1$.

(3) Spring 2017 Problem 3.

- (a) Consider the sequence $\{X_k, k \geq 1\}$ of random variables such that X_1 is uniform on $(0, 1)$ and, given X_k , the distribution of X_{k+1} is uniform on $(0, CX_k)$, where $\sqrt{3} < C < 2$.
- (i) For $n \geq 1$, compute the conditional expectation $\mathbb{E}(X_{n+1}^r | X_n)$.
 - (ii) For $n \geq 1$, compute $\mathbb{E}(X_n^r)$.
 - (iii) Show that $\lim_{n \rightarrow \infty} X_n = 0$ in ℓ_1 and with probability one, but not in ℓ_2 .
 - (iv) Investigate the same questions for all other values of $C > 0$.
- (b) Let $a > 0$, let $X_n, n \geq 1$ be i.i.d. random variables that are uniform on $(0, a)$, and let $Y_n = \prod_{k=1}^n X_k$. Determine, with a proof, all values of a for which $\lim_{n \rightarrow \infty} Y_n = 0$ with probability one.

Solution.

- (a) (i) We have that $X_{n+1} \mid X_n \sim U(0, CX_n)$. Therefore

$$\begin{aligned}\mathbb{E}(X_{n+1}^r \mid X_n) &= \frac{1}{CX_n} \int_0^{CX_n} x^r dx = \frac{1}{CX_n} \cdot \frac{x^{r+1}}{r+1} \Big|_0^{CX_n} = \frac{C^r X_n^r}{r+1} \\ \implies \mathbb{E}(X_{n+1}^r) &= \mathbb{E}[\mathbb{E}(X_{n+1}^r \mid X_n)] = \frac{C^r}{r+1} \cdot \mathbb{E}(X_n^r) \\ \implies \boxed{\mathbb{E}(X_{n+1}^r \mid X_n) &= \frac{C^r}{r+1} X_n^r}\end{aligned}$$

- (ii) Note that $\mathbb{E}(X_1^r) = \int_0^1 x^r dr = 1/(r+1)$. Therefore

$$\mathbb{E}(X_{n+1}^r) = \frac{C^r}{r+1} \cdot \mathbb{E}(X_n^r) = \left(\frac{C^r}{r+1}\right)^n \cdot \mathbb{E}(X_1^r) = \boxed{\left(\frac{C^r}{r+1}\right)^n \cdot \frac{1}{r+1}}$$

- (iii) We would like to show that $X_n \xrightarrow{w.p.1} 0$ and that $X_n \xrightarrow{1} 0$, but that the same result does not follow for the ℓ_2 norm.

- **Convergence with probability one:** We seek to show that $\Pr(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0\}) = 1$. By Markov's Inequality (Lemma 3), we have

$$\Pr(|X_n| \geq a) \leq \frac{\mathbb{E}(X_n)}{a} \quad \forall a > 0$$

$$\iff \Pr(|X_n| \geq a) \leq \left(\frac{C^1}{1+1}\right)^{n-1} \cdot \frac{1}{1+1} \cdot \frac{1}{a} = \left(\frac{C}{2}\right)^{n-1} \cdot \frac{1}{2a} \quad \forall a > 0$$

Since $\sqrt{3} < C < 2$, $\sqrt{3}/2 < C/2 < 1$. Since $X_n \in [0, CX_{n-1}]$, $X_n \geq 0$, so $|X_n| = X_n$. Therefore we have

$$\Pr(\lim_{n \rightarrow \infty} |X_n| \geq a) = \Pr(\lim_{n \rightarrow \infty} X_n \geq a) \leq \lim_{n \rightarrow \infty} \left(\frac{C}{2}\right)^{n-1} \cdot \frac{1}{2a} = 0 \quad \forall a > 0$$

Since $|X_n| \geq 0$, this implies that $\Pr(\lim_{n \rightarrow \infty} X_n = 0) = \Pr(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0\}) = 1$, so by the Borel-Cantelli Lemma (Theorem 23), X_n converges to 0 with probability 1.

- **Convergence in ℓ_1 norm:** We seek to show that $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|) = 0$. Since $X_n \in [0, CX_{n-1}]$, $X_n \geq 0$, so $|X_n| = X_n$. Therefore

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \lim_{n \rightarrow \infty} \left(\frac{C}{2}\right)^{n-1} \cdot \frac{1}{2}$$

Since $\sqrt{3} < C < 2$, $\sqrt{3}/2 < C/2 < 1$, so $C/2 < 1$. Therefore we have

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|) = \lim_{n \rightarrow \infty} \left(\frac{C}{2}\right)^{n-1} \cdot \frac{1}{2} = 0$$

so X_n converges to 0 in 1st mean.

- **Convergence in ℓ_2 norm:** We seek to show that $\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|^2) \neq 0$. We have

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|^2) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n^2) = \lim_{n \rightarrow \infty} \left(\frac{C^2}{3}\right)^{n-1} \cdot \frac{1}{3}$$

Since $\sqrt{3} < C < 2$, $3/3 < C^2/3 < 4/3$, so $C^2/3 > 1$. Therefore we have

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|^2) = \lim_{n \rightarrow \infty} \left(\frac{C^2}{3} \right)^{n-1} \cdot \frac{1}{3} = \infty \neq 0$$

so X_n does not converge to 0 in 2nd mean.

- (iv) From the above, it is clear that for convergence with probability one or in 1st mean we require $0 < C/2 < 1$ and for convergence in second mean we require $0 < C^2/3 < 1$. For $0 < C < \sqrt{3}$, we see that X_n would converge to zero in 2nd mean since this would imply that $0 < C^2/3 < 1$. It would also still converge to 0 in 1st mean (and with probability 1) since we would have $(0 < C/2 < \sqrt{3}/2 < 1)$.

For $C = \sqrt{3}$, X_n would still converge to 0 with probability one and in 1st mean for the same reasons. However, it would not converge in 2nd mean because we would have

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n|^2) = \lim_{n \rightarrow \infty} \left(\frac{\sqrt{3}^2}{3} \right)^{n-1} \cdot \frac{1}{3} = \frac{1}{3} \neq 0$$

For $C \geq 2$, it would diverge in all three cases, since in this case $C/2 \geq 2/2 = 1$ and $C^2/3 \geq 4/3 > 1$.

- (b) **Probably won't be on midterm.** Note that

$$\lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} \prod_{k=1}^n X_k = 0 \iff \log(Y_n) = \log\left(\prod_{k=1}^n X_k\right) = \sum_{k=1}^n \log(X_k) \rightarrow -\infty$$

Note that

$$\begin{aligned} \mathbb{E}[\log(Y_n)] &= \mathbb{E}\left(\sum_{k=1}^n \log(X_k)\right) = \sum_{k=1}^n \mathbb{E}[\log(X_k)] = \sum_{k=1}^n \mathbb{E}[\log(X_1)] = \sum_{k=1}^n \int_0^a (\log(x)/a) dx \\ &= \sum_{k=1}^n \frac{1}{a} [x \log x - x]_0^a = \sum_{k=1}^n \frac{a \log a - a}{a} = \sum_{k=1}^n (\log(a) - 1) = n(\log(a) - 1) \end{aligned}$$

As $n \rightarrow \infty$ we have

$$\mathbb{E}[\log(Y_n)] = \begin{cases} -\infty & a < e \\ 0 & a = e \\ \infty & a > e \end{cases}$$

Since $\mathbb{E}[\log(Y_n)] \rightarrow \infty$ for $a < e$, we have $\lim_{n \rightarrow \infty} Y_n = 0$ for $a < 3$. Therefore

$$\boxed{\lim_{n \rightarrow \infty} Y_n = \lim_{n \rightarrow \infty} \prod_{k=1}^n X_k = 0 \iff a < e.}$$