Ryan McCorvie
UCLA Math 275A
Exercises

Foundations

0.11 Let $\mathcal{A}$ be a collection of subset of Ω, and let $P(E)$ be a property of subsets $E \subset \Omega$. ($P(E)$ is true or false for every $E \subset \Omega$).

- $P(\emptyset)$ is true
- $P(E)$ is true for all $E \subset \mathcal{A}$
- if $E \subset \Omega$ is such that $P(E)$ is true, then $P(\Omega \setminus E)$ is true
- If $E_1, E_2, \dots \subset \Omega$ are such that $P(E_n)$ is true for all n then $P(\bigcup_n E_n)$ is true

Show that $P(E)$ is true for every $E \in \langle \mathcal{A} \rangle$

The collection $\mathcal{P} = \{E \subset \Omega : P(E) \text{ is true}\}$ is a σ-algebra. The items above show that $\emptyset \in \mathcal{P}$, that $\mathcal{P}$ is closed under complements, and that $\mathcal{P}$ is closed under countable unions. Furthermore, $\Omega \in \mathcal{P}$ since $\Omega = \Omega \setminus \emptyset$. Also for $E_1, E_2, \dots \in \mathcal{P}$, $\bigcap_n E_n \in \mathcal{P}$ since $\bigcap_n E_n = (\bigcup_n E_n^c)^c$.

Now the second item shows that $\mathcal{A} \subset \mathcal{P}$. Therefore $\langle A \rangle \subset \mathcal{P}$ since $\langle A \rangle$ is contained in every σ-algebra which contans $\mathcal{A}$ ■

0.15 Show that $\mathbb{R}^n$ with the Borel σ-algebra is the product of n copies $\mathbb{R}$ with the Borel σ-algebra

For any measurable set $M \subset \mathbb{R}$, the cylinder set $\{x \in \mathbb{R}^n : x_j \in M\}$ is measurable in $\mathbb{R}^n$. This is because the projection mapping $\pi_i : \mathbb{R}^n \to \mathbb{R}$ which maps $\pi_i(x) = x_i$ is continuous, and the inverse of the projection mapping is a cylinder. Since the cylinder sets generate the product σ-algebra, so the product σ-algebra is courser than the Borel σ-algebra on $\mathbb{R}^n$.

Conversely, since (a_i, b_i) is open in $\mathbb{R}$ the cylinder $C_i = \{x \in R^n : x_i \in (a_i, b_i)\}$ is in the product σ-algebra. Therefore the open box

$$(a_1, b_1) \times \cdots \times (a_n, b_n) = \bigcap_{i=1}^{n} C_i \tag{1}$$

is also in the product σ-algebra. However every open set $U \in \mathbb{R}^n$ can be written as the countable union of boxes with rational corner points– just take all such boxes contained in U. Therefore the product σ-algebra includes all of the open sets of $\mathbb{R}^n$, and these sets generate the Borel σ-algebra on $\mathbb{R}^n$. Therefore the Borel σ-algebra of $\mathbb{R}^n$ is courser than the product σ-algebra. We conclude the σ-algebras are the same. ■

0.17 Show that any continuous map from one topological space X to another Y is necessarily measurable (when one gives X and Y the Borel σ-algebra)

Let $f : X \to Y$ be a continous map. Let $\mathcal{M} = \{M \subset \mathbb{R} : f^{-1}(M) \text{ is measurable}\}$. First, $\emptyset \in \mathcal{M}$ since $f^{-1}(\emptyset) = \emptyset$. Furthermore since

$$X \setminus f^{-1}(M) = f^{-1}(Y \setminus M) \tag{2}$$

we conclude $Y \setminus M \in \mathcal{M}$ if $Y \in \mathcal{M}$ is, since the hypothesis implies $f^{-1}(M)$ is measurable, and hence $X \setminus f^{-1}(Y)$ is also. In short, $\mathcal{M}$ is closed under complements. Furthermore

$$\bigcup_n f^{-1}(M_n) = f^{-1}(\bigcup_n M_n) \tag{3}$$

so $\bigcup_n M_n \in \mathcal{M}$, since its its inverse is a measurable set. Finally, since f is continuous, for any open set $U \subset Y$, we know $f^{-1}(U)$ is open and hence measurable. Therefore $\mathcal{M}$ contains all open sets. By 0.11 it therefore contains the σ-algebra generated by open sets, namely the Borel σ-algebra. Thus $f^{-1}(M)$ is measurable in X for every M measurable in Y, and the function f is measurable. ∎

0.18 If $X_1 : \Omega \to R_1, \dots, X_n : \Omega \to R_n$ are measurable functions into measurable spaces $R_1, \dots, R_n$ show that the joint function $(X_1, \dots, X_n) : \Omega \to R_1 \times \cdots \times R_n$ given by $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$ is also measurable

In 0.17 we showed that $\mathcal{M} = \{M : f^{-1}(M) \text{ is measurable}\}$ is a σ-algebra for any function f. Thus to show the mapping above is measurable, it suffices to show the inverse of cylinder sets are measurable. However, the inverse of the cylinder set $C_i = \{x \in R_1 \times \cdots \times R_n : x_i \in E_i\}$ under the product map is the same as the inverse of E_i under X_i. This set is measurable by hypothesis because X_i is a measurable function. ∎

0.23 Prove the measure has a bunch of properties

1. (Monotonicity) Suppose $E \subset F$. Then by countable additivity and positivity

$$\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E) \tag{4}$$

Summarizing, $\mu(E) \leq \mu(F)$

2. (Subadditivity) Suppose $E_1, E_2, \dots$ are measurable (not necessarily disjoint) sets. Let $F_n = \bigcup_{m=1}^n E_m$ be the union of events up to n. Then the events $E_n' = E_n \setminus F_{n-1}$ are disjoint with $\bigcup_n E_n' = \bigcup_n E_n$. Furthermore, by monotonicity (property a) $\mu(E_n') \leq \mu(E_n)$. Therefore

$$\mu(\bigcup_n E_n) = \mu(\bigcup_n E_n') = \sum_n \mu(E_n') \leq \sum_n \mu(E_n) \tag{5}$$

Summarizing, for any events E_n, $\mu(\bigcup_n E_n) \leq \sum_n \mu(E_n)$.

3. (Continuity from below) Let $E_1 \subset E_2 \subset \dots$ be measurable. Using the construction from (2), the disjoint sets $E'_n = E_n \setminus E_{n-1}$ have the property that $E_n = \bigcup_{m=1}^n E_m = \bigcup_{m=1}^n E'_m$ and $\bigcup_{n=1}^\infty E_n = \bigcup_{n=1}^\infty E'_n$. Thus

$$\lim_n \mu(E_n) = \lim_n \sum_{m=1}^n \mu(E'_m) = \sum_{m=1}^\infty \mu(E'_m) = \mu(\bigcup_m E'_m) = \mu(\bigcup_m E_m) \tag{6}$$

Summarizing, $\lim_{n\to\infty} \mu(E_n) = \mu(\bigcup_n E_n)$

4. (Continuity from above) Suppose $E_1 \supset E_2 \supset \dots$ are measurable. Let $F_n = E_1 \setminus E_n$ and note $F_1 \subset F_2 \subset \dots$ satisfy the hypothesis of (3). Now $\bigcup_n F_n = E_1 \setminus \bigcap_n E_n$, so

$$\mu(E_1 \setminus \bigcap_n E_n) = \mu(\bigcup_n F_n) = \lim_n \mu(F_n) = \lim_n \mu(E_1 \setminus E_n) \tag{7}$$

Now $\mu(E_1) = \mu(E_1 \setminus E_n) + \mu(E_n)$, so if $\mu(E_1)$ is finite then so are $\mu(E_n)$ and $\mu(E_1 \setminus E_n)$, and the so the equation $\mu(E_1 \setminus E_n) = \mu(E_1) - \mu(E_n)$. Similarly $\mu(E_1 \setminus \bigcap_n E_n) = \mu(E_1) - \mu(\bigcap_n E_n)$. Therefore we can rewrite (7)

$$\mu(E_1) - \mu(\bigcap_n E_n) = \mu(E_1) - \lim_n \mu(E_n) \tag{8}$$

Canceling $\mu(E_1)$ gives the equation $\lim_n \mu(E_n) = \mu(\bigcap_n E_n)$.

if $\mu(E_1) = \infty$, its possible for this property to fail. Consider $E_n = [n, \infty)$. Then $E_1 \supset E_2 \supset \dots$. In this case $\mu(\bigcap_n E_n) = \mu(\emptyset) = 0$ however $\mu(E_n) = \infty$ for all n, so $\lim_{n\to\infty} \mu(E_n) = \infty$

∎

0.24 Let $(\Omega, \mathcal{B})$ be a measurable space.

- If $f : \Omega \to [-\infty, \infty]$ is a function taking values in the extended reals $[-\infty, \infty]$ show that f is measurable (with respect to the Borel σ-algebra) if and only if the sets $\{\omega \in \Omega : f(\omega) \leq t\}$ are measurable for all real t
- If $f, g : \Omega \to [-\infty, \infty]$ are functions, show that $f = g$ if and only if $\{\omega \in \Omega : f(\omega) \leq t\} = \{\omega \in \Omega : g(\omega) \leq t\}$ for all real t
- If $f_1, f_2, \cdots : \Omega \to [-\infty, \infty]$ are measurable show that $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf f_n$ are all measurable.

- In 0.17 we showed that $\mathcal{M} = \{M : f^{-1}(M) \text{ is measurable}\}$ is a σ-algebra. Thus, if we can show that all open sets are in $\mathcal{M}$ then we can conclude that f is measurable with respect to Borel measurable sets. By hypothesis the sets $[-\infty, t] \in \mathcal{M}$. Therefore so are $[-\infty, s) = \bigcup_n [-\infty, s - 1/n]$ and $(s, \infty] = [-\infty, s]^c$. Also, so is $(s, t) = [-\infty, s) \setminus [-\infty, t]$. But the sets (q_1, q_2) with $q_1, q_2 \in \mathbb{Q}$ along with $[-\infty, q)$ and $(q, \infty]$ form a countable basis for $[-\infty, \infty]$. Every open set $U \subset [-\infty, \infty]$ is the union of all such subsets which are contained in U, and there are only countably many such subsets. Therefore all open sets $U \subset [-\infty, \infty]$ are in $\mathcal{M}$.

- If $f = g$ then trivially the sets are equal. Conversely, swapping f and g if necessary, $f(\omega) > g(\omega) = t$ for some $\omega \in \Omega$. Then $\omega \notin f^{-1}([-\infty, t])$ but $\omega \in g^{-1}([-\infty, t])$ and the sets are not equal.

- Suppse $t \geq \sup_n f_n(\omega)$. Then also $t \geq f_n(\omega)$ since $\sup_n f_n \geq f_m$ for any m. Furthermore if $\sup_n f_n(\omega) > t$ then some $f_m(\omega) > t$ since otherwise t is an upperbound for all $f_n(\omega)$ and hence $\sup_n f_n(\omega) \leq t$. This shows that

$$\{\omega \in \Omega : \sup_n f_n(\omega) \leq t\} = \bigcap_n \{\omega \in \Omega : f_n(\omega) \leq t\} \tag{9}$$

Therefore $\{\omega \in \Omega : \sup_n f_n(\omega) \leq t\}$ is measurable, since its the countable union of measurable sets. By the first item, this shows $\sup_n f$ is measurable. Turning to the infimum

$$\{\omega \in \Omega : \inf_n f_n(\omega) \geq t\} = \bigcap_n \{\omega \in \Omega : f_n(\omega) \geq t\} \tag{10}$$

Thus the sets on the left hand side are measurable. Since the arguments in item 1 are symmetric with respect to the transformation $x \to -x$, this is sufficient to show $f = \inf f_n$ is measurable. Since $\liminf f_n = \sup_m \inf_{n \geq m} f_n$ and $\limsup f_n = \inf_m \sup_{n \geq m} f_n$, these functions are also measurable.

■

0.26 Let μ be a probability measure on the real line $\mathbb{R}$ (with the Borel σ-algebra). Define *Stieltjes measure function* $F : \mathbb{R} \to [0, 1]$ associated to μ by the formula

$$F(t) := \mu((-\infty, t]) = \mu(\{x \in \mathbb{R} : x \leq t\}) \tag{11}$$

Establish some properties

(i) F is non-decreasing, so that $F(s) \leq F(t)$ for $s < t$. This follows from the monotonicity property in 0.23 since $(-\infty, s] \subset (-\infty, t]$

(ii) Using continuity from below in 0.23, $\lim_{n\to\infty} F(t) = \lim_{n\to\infty} \mu((0, n]) = \mu((-\infty, \infty)) = 1$ for $n \in \mathbb{Z}$. The fact that F is non-decreasing lets us extend the limit to reals so that $\lim_{t\to\infty} F(t) = 1$ for $t \in \mathbb{R}$, since $F(t) \geq F(\lfloor t \rfloor)$. Thus $1 - F(t) < \epsilon$ for arbitrarily small ϵ for large $t > N \in \mathbb{Z}$ such that $1 - F(N) < \epsilon$, and N exists because the limit exists in $\mathbb{Z}$. Since this is a probability space, $\mu(E) \leq 1 < \infty$ for all measurable events. Therefore we can use the analogous argument with continuity from above to show that $\lim_{t\to-\infty} F(t) = \mu(\emptyset) = 0$

(iii) Let $s_k \downarrow s$ be any decreasing sequence which converges down to s. Then by continuity from above in 0.23 $\lim_{k\to\infty} F(s_k) = \mu(\bigcap_k (-\infty, s_k]) = \mu((-\infty, s]) = F(s)$. We can extend this to limits in $\mathbb{R}$ using the monotonicity of F, so $\lim_{t\to s+} F(t) = F(s)$.

■

0.28 Let X be a real random variable with cumulative distribution function F. For any real number t show that

$$\Pr(X < t) = \lim_{s \to t-} F(s) \tag{12}$$

and

$$\Pr(X = t) = F(t) - \lim_{s \to t-} F(s) \tag{13}$$

In particular, $\Pr(X = t) = 0$ for all t if and only if F is continuous.

This is the same argument as in 0.26iii. Let $t_k \uparrow t$ be a monotonic increasing sequence converging to t. Then using continuity from below

$$\lim_k F(t_k) = \lim_k \Pr((-\infty, t_k]) = \Pr((-\infty, t)) = \Pr(X < t) \tag{14}$$

since $\bigcup_k (-\infty, t_k] = (-\infty, t)$. From the monotonicity of the cumulative density function, this limit extends to the real numbers so $\lim_{s \to t-} F(s) = \Pr(X < t)$.

Furthermore

$$\Pr(X = t) = \Pr(\{X \le t\} \setminus \{X < t\}) = \Pr(X \le t) - \Pr(X < t) = F(t) - \lim_{s \to t-} F(s) \tag{15}$$

In 0.23 we showed that F is right continuous. From this equation, F is left continuous if and only if $\Pr(X = t) = 0$ for all t. ∎

0.29 (Skorokhod representation of a scalar variable) Let U have a uniform distribution. Thus the cumulative distribution function is given by

$$F_U(t) = \begin{cases} 0 & t \le 0 \\ t & 0 < t \le 1 \\ 1 & t > 1 \end{cases} \tag{16}$$

Let $F : \mathbb{R} \to [0, 1]$ be another cumulative distribution function. Show that

$$X^- := \sup\{y \in \mathbb{R} : F(y) < U\} \qquad X^+ := \inf\{y \in \mathbb{R} : F(y) \ge U\} \tag{17}$$

are random variables (i.e., they are measurable in any model Ω) and have cumulative distribution function F.

Basically $X = F^{-1}(U)$, except F need not be one-to-one. But its monotonic and right continuous, so its almost one-to-one. We just have to be careful with intervals where F is constant or points where F increases discontinuously.

This construction is attributed to Skorokhod, but it should not be confused with the Skorokhod representation theorem. It provides a quick way to generate a single scalar variable, but unfortunately it is difficult to modify this construction to generate multiple scalar variables, especially if they are somehow coupled to each other. ∎

0.30

■

Integration and Expectation

1.3

■

1.4

■

1.5

■

1.6 If $(\Omega, (p_\omega)_{\omega \in \Omega})$ is a discrete probability space (with the associated probability measure μ) and $f : \Omega \to [0, +\infty]$ is a function. Show that

$$\int_\Omega f \, d\mu = \sum_{\omega \in \Omega} f(\omega) p_\omega \tag{18}$$

On a discrete probability space, we can decompose all simple functions into supersimple functions $s = \sum_{\omega \in S} \delta_\omega s_\omega$ where $S \subset \Omega$ is a finite set which represents the support of s, and δ_ω is the indicator of the singleton set $1_{\{\omega\}}$ and $s_k \in [0, +\infty]$. Thus a simple function $s \leq f$ consists of some finite $S \subset \Omega$ and values s_ω for $\omega \in S$ such that $s_\omega \leq f(\omega)$. The expectation of such a simple function is

$$\text{Simp} \int s = \sum_{\omega \in S} s_\omega p_\omega \tag{19}$$

Its pretty clear that the supremum of all such expression is the sum

$$\int f = \sum_{\omega \in \Omega} f(\omega) p_\omega \tag{20}$$

since this sum is an upper bound for all of the Simp $\int s$ for $s \leq f$ and we can get arbitrarily close to the integral with a finite number of terms. ■

1.7

■

1.8

∎

1.9

∎

1.11

∎

1.19 Let $E_1, E_2, \ldots$ be a sequence of events with $\inf_n \Pr(E_n) > 0$. Show that with positive probability an infinite number of the E_n hold.

Let E be the event that infinitely many of the E_n occur. Then the indicators satisfy $1_E = \limsup_n 1_{E_n}$, since the right hand side is 1 for all samples ω where $1_{E_n}(\omega) = 1$ infinitely often and 0 when $1_{E_n}(\omega) = 0$ eventually. Taking complements, $1_{E^c} = \liminf_n 1_{E_n^c}$, so, by Fatou's lemma

$$\Pr(E^c) = \mathrm{E}\, 1_{E^c} = \mathrm{E} \liminf_n 1_{E_n^c} \leq \liminf_n \mathrm{E}\, 1_{E_n^c} = \liminf_n \Pr(E_n^c) \leq \sup_n \Pr(E_n^c) < 1 \tag{21}$$

Therefore $\Pr(E) > 0$

∎

1.20 Let $p_1, p_2, \cdots \in [0,1]$ be a sequence such that $\sum_{n=1}^{\infty} p_n = +\infty$. Show that there exists a sequence of events $E_1, E_2, \ldots$ modeled by some probability space Ω such that $\Pr(E_n) = p_n$ for all n such that almost surely infinitely many of the E_n occur.

Just take the events E_n to be independent. Let F_m be the event that none of the events E_n occur for $n \geq m$. Then by independence

$$\Pr(F_m) = \prod_{n \geq m} \Pr(E_n^c) = \prod_{n \geq m} (1 - p_n) \leq \exp\left(-\sum_{n \geq m} p_n\right) = \exp(-\infty) = 0 \tag{22}$$

Let $F = \cup_{m=1}^{\infty} F_m$ be the event that only finitely many of the E_n occur. Then $\Pr(F) \leq \sum_m \Pr(F_m) = 0$, and therefore E_n occurs infinitely often almost surely. ∎

1.26 (Scheffe's lemma) Let $X_1, X_2, \ldots$ be a sequence of absolutely integrable scalar random variables which converge almost surely to $X = \lim_n X_n$, another absolutely integrable scalar random variable. Prove that

$$\mathrm{E}|X_n - X| \to 0 \qquad \text{if and only if} \qquad \mathrm{E}|X_n| \to \mathrm{E}|X| \tag{23}$$

For any two numbers, the difference in their magnitudes is less than the magnitude of their difference. If they have the same sign, these quantities are the same, but if they have different signs, the latter is larger. In other words, $||a| - |b|| \le |a - b|$. From this it follows that

$$|\mathrm{E}|X_n| - \mathrm{E}|X|| \le \mathrm{E}||X_n| - |X|| \le \mathrm{E}|X_n - X| \tag{24}$$

If the right hand side tends to zero, the so does the left hand side, and therefore $\mathrm{E}|X_n| \to \mathrm{E}|X|$. This proves the "only if" part.

For the "if" part, define

$$X_n' = 1_{X_n<0} \max(X_n, -|X|) + 1_{X>0} \min(X_n, |X|) \qquad \text{and} \qquad X_n'' = X_n - X_n' \tag{25}$$

Now on any sample $\omega \in \Omega$, take n large enough so $\mathrm{sgn}(X_n(\omega)) = \mathrm{sgn}(X(\omega))$. In this case $|X(\omega) - X_n'(\omega)| = 0$ if $|X(\omega)| \ge |X_n'(\omega)|$ and $|X(\omega) - X_n'(\omega)| = |X(\omega) - X_n(\omega)|$ otherwise. In other words, for large enough n, $|X_n'(\omega) - X(\omega)| \le |X_n(\omega) - X(\omega)|$, and so $X_n'(\omega) \to X(\omega)$. This shows that $\lim_n X_n' = X$ almost surely. By dominated convergence it follows that $\mathrm{E}\, X_n' \to \mathrm{E}\, X$ since X_n' is dominated by $|X|$ by construction.

By the definition of X_n'', and since $|X_n'| \le |X_n|$ by construction, it follows that $|X_n''| = |X_n - X_n'| = |X_n| - |X_n'|$. Therefore

$$\mathrm{E}|X - X_n| \le \mathrm{E}|X - X_n'| + \mathrm{E}|X_n''| = \mathrm{E}|X - X_n'| + \mathrm{E}|X_n'| - \mathrm{E}|X_n| \tag{26}$$

The last two terms cancel in the limit since

$$\lim_n \mathrm{E}|X_n'| = \mathrm{E}|X| = \lim_n \mathrm{E}|X_n| \tag{27}$$

The first equality follows from the hypothesis and the second from the dominated convergence argument above. The first term in (26) is dominated in L^1 since $|X - X_n'| \le 2|X|$. Since $X_n' \to X$ almost surely, dominated convergence implies the first term tends to 0. Thus the left hand side tends to 0, as we wished to prove.

∎

1.35 LEet $f : \mathbb{R} \to [0, +\infty]$ be a measurable function with $\int_R f(x)\, dx = 1$. If one defines $m_f(E)$ for any Borel subset E of $\mathbb{R}$ by the formula

$$m_f(E) = \int_E f(x)\, dx \tag{28}$$

show that m_f is a probability measure on $\mathbb{R}$ with Stieltjes measure function $F(t) = \int_{-\infty}^t f(x)\, dx$. If X is a real random variable with probability distribution m_f. In this case X is absolutely continuous with respect to the Lesbegue measure and f is the Radon-Nikodym derivative which, in this context, is called the *probability density function* (pdf). Show that

$$\mathrm{E}\, G(X) = \int_{\mathbb{R}} G(x) f(x)\, dx \tag{29}$$

with either $G : \mathbb{R} \to [0, +\infty]$ is an unsigned measurable function or $G : \mathbb{R} \to \mathbb{C}$ is measurable with $G(X)$ absolutely integrable.

■

1.36 Let X be a random variable with probability desnity function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ of the standard normal distribution. Establish *Stein's identity*

$$\mathrm{E}\, XF(X) = \mathrm{E}\, F'(X) \tag{30}$$

whenever $F : \mathbb{R} \to \mathbb{R}$ is continuously differentiable function with F and F' both of polynomial growth. Use this identity to recursively establish the identities

$$\mathrm{E}\, X^k = \begin{cases} 0 & k \text{ odd} \\ (k-1)!! & k \text{ even} \end{cases} \tag{31}$$

Using integration by parts with $v' = xe^{-\frac{1}{2}x^2}$ and $u = F(x)$

$$\begin{aligned} \mathrm{E}\, XF(X) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} xF(x) e^{-\frac{1}{2}x^2}\, dx \\ &= -\frac{1}{\sqrt{2\pi}} F(x) e^{-\frac{1}{2}x^2} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} F'(x) e^{-\frac{1}{2}x^2} \\ &= \mathrm{E}\, F'(X) \end{aligned}$$

The assumption that $F \in C^2$ justifies integration by parts. The growth condition on F implies the boundary terms vanish at infinity. The grown condition on F' implies $F'(X)$ is integrable.

Taking $F(X) = X^{k-1}$ we get

$$\mathrm{E}\, X^k = (k-1)\, \mathrm{E}\, X^{k-2} \tag{32}$$

Since $\mathrm{E}\, X = 0$ and $\mathrm{E}\, X^2 = 1$, induction gives the formulas in the problem statement. ■

1.37 Let X be a real random variable with cumulative distribution function $F_X(x) = \Pr(X \le x)$. Show that

$$\mathrm{E}\, e^{tX} = \int_{\mathbb{R}} (1 - F_X(x)) te^{tx}\, dx \qquad \text{for all } t > 0 \tag{33}$$

If X is nonnegative show that

$$\mathrm{E}\, X^p = \int_0^{\infty} (1 - F_X(x)) px^{p-1}\, dx \tag{34}$$

These are applications of Fubini's theorem over the region $y \ge x$ either in $\mathbb{R}^2$ or in the first quadrant.

$$\begin{aligned} \mathrm{E}\, e^{tX} &= \int_{\mathbb{R}} e^{ty}\, \mu(dy) = \int_{\mathbb{R}} \int_{-\infty}^{y} te^{tx}\, dx\, \mu(dy) = \int_{\mathbb{R}} \int_{x}^{\infty} \mu(dy)\, te^{tx}\, dx \\ &= \int_{R} (1 - F_X(x)) te^{tx}\, dx \end{aligned} \tag{35}$$

Here we've used the assumption $t > 0$ with the fundamental theorem of calculus to make the identification

$$\int_{-\infty}^{y} te^{tx}\,dx = e^{tx}|_{-\infty}^{y} = e^{ty} \tag{36}$$

and the definition of the distribution to get

$$1 - F_X(x) = F_X(\infty) - F_X(x) = \mu((x,\infty)) = \int_x^\infty \mu(dx) \tag{37}$$

A similar calculation shows for non-negative X

$$\begin{aligned} \mathrm{E}\,X^p &= \int_0^\infty y^p\,\mu(dy) = \int_0^\infty \int_0^y px^p\,dx\,\mu(dy) = \int_0^\infty \int_x^\infty \mu(dy)\,px^p\,dx \\ &= \int_0^\infty (1 - F_X(x))px^p\,dx \end{aligned} \tag{38}$$

■

1.39 Let $f : \mathbb{C} \to \mathbb{R}$ be a convex function and let X be a complex random variable with X and $f(X)$ both absolutely integrable. Show that $$f(\mathrm{E}\,X) \le \mathrm{E}\,f(X) \tag{39}$$

The fact f is convex implies that for all $z, w \in \mathbb{C}$ and $t \in [0,1]$

$$f((1-t)z + tw) \le (1-t)f(z) + tf(w) \tag{40}$$

At any point a convex function has a supporting hyperplane, so we may write

$$f(z) \ge f(z_0) + c_1\,\mathrm{Re}(z - z_0) + c_2\,\mathrm{Im}(z - z_0) \tag{41}$$

Using linearity of expectations

$$\mathrm{E}\,f(X) \ge f(z_0) + c_1(\mathrm{Re}\,\mathrm{E}\,X - \Re z_0) + c_2(\mathrm{Im}\,\mathrm{E}\,X - \mathrm{Im}\,z_0) \tag{42}$$

Choosing $z_0 = \mathrm{E}\,X$ gives Jensen's inequality. ■

1.40 Show the expressions $\|X\|_p$ are non-decreasing in p for $p \in (0,\infty]$. In particular if $\|X\|_p$ is finite for some p then its automatically finite for all smaller values of p

For any constant c and $p \in (0,\infty)$ note that

$$\|c\|_p = (\mathrm{E}\,c^p)^{\frac{1}{p}} = c^{\frac{p}{p}} = c \tag{43}$$

Furthremore its clear the essential supremum of a constant is the constant so $\|c\|_\infty = c$. This shows that for a probability space $\|c\|_p = c$ for $p \in (0,\infty]$.

Let $p' > p$ for $p, p' \in (0, \infty)$ and let $r = p'/p > 1$. Then using Hölder's inequality with the functions 1 and X^p and exponents r and $s = r/(r-1)$.

$$\mathrm{E}|X|^p \le \||X|^p\|_r \, \|1\|_s = (\mathrm{E}|X|^{pr})^{\frac{1}{r}} = \left(\mathrm{E}|X|^{p'}\right)^{\frac{p}{p'}} \tag{44}$$

Taking $\frac{1}{p}$ powers of both sides gives $\|X\|_p \le \|X\|_{p'}$.

For $p' = \infty$, since $|X| \le \|X\|_\infty$ almost surely, the monotonicity of expectations implies

$$\|X\|_p = (\mathrm{E}|X|^p)^{\frac{1}{p}} \le (\mathrm{E}\|X\|_\infty^p)^{\frac{1}{p}} = \|\|X\|_\infty\|_p = \|X\|_\infty \tag{45}$$

Thus $\|X\|_p$ is monotonic for $p \in (0, \infty]$. ∎

1.41 Show that for any $X \in L^2$

$$\Pr(X \neq 0) \ge \frac{(\mathrm{E}|X|)^2}{\mathrm{E}(|X|^2)} \tag{46}$$

By Cauchy-Schwartz applied to $1_{X \neq 0}$ and $|X|$

$$\Pr(X \neq 0)\,\mathrm{E}|X|^2 = \mathrm{E}\, I_{X\neq 0}^2 \,\mathrm{E}|X|^2 \ge (\mathrm{E}\, 1_{X\neq 0}|X|)^2 = (\mathrm{E}|X|)^2 \tag{47}$$

We used the fact that $1_A^2 = 1_A$ for any indicator (it only takes the values 0 or 1) and the fact that $|X| = 0$ when $X = 0$ so $1_{X\neq 0}|X| = |X|$. ∎

1.42 Establish Minkowskis inequality

$$\|X + Y\|_p \le \|X\|_p + \|Y\|_p \tag{48}$$

First assume that X, Y are nonnegative and bounded, so that $0 \le X, Y \le K$ for some constant $K > 0$. By Hölder's inequality

$$\begin{aligned} \mathrm{E}(X\,|X+Y|^{p-1}) &\le \|X\|_p \,(\mathrm{E}|X+Y|^{q(p-1)})^{\frac{1}{q}} = \|X\|_p\,(\mathrm{E}|X+Y|^p)^{\frac{p-1}{p}} \\ &= \|X\|_p\,\|X+Y\|_p^{p-1} \end{aligned} \tag{49}$$

In the exponents I've used variations of $q = p/(p-1)$. Similarly

$$\mathrm{E}(Y\,|X+Y|^{p-1}) \le \|Y\|_p\,\|X+Y\|_p^{p-1} \tag{50}$$

Utilizing the fact that for these nonnegative random variables $X + Y = |X + Y|$

$$\begin{aligned} \|X+Y\|_p^p &= \mathrm{E}|X+Y|^p = \mathrm{E}(X\,|X+Y|^{p-1}) + \mathrm{E}(Y\,|X+Y|^{p-1}) \\ &\le (\|X\|_p + \|Y\|_p)\|X+Y\|^{p-1} \end{aligned} \tag{51}$$

Note that if $\|X+Y\|_p = 0$, then (48) is trivially true, and we may ignore that case. Furthermore, because X, Y are bounded by K, we know $X+Y$ is bounded by $2K$, so $\|X+Y\|_p < \infty$. Thus we may cancel $\|X+Y\|^{p-1}$ from both sides to yield (48).

We may extend (48) to unbounded nonnegative random variables, applying the inequality to $\max(X, K)$ and $\max(Y, K)$ and then using the monotone convergence theorem as $K \to \infty$. To extend the inequality to random variables which take negative values, note that

$$\|X+Y\|_p \leq \||X|+|Y|\|_p \leq \||X|\|_p + \||Y|\|_p = \|X\|_p + \|Y\|_p \tag{52}$$

■

1.43 If $X \in L^2$ is non-negative and $\theta \in [0,1]$, establish the Paley-Zygmund inequality

$$\Pr(X > \theta \operatorname{E}(X)) \geq (1-\theta)^2 \frac{(\operatorname{E} X)^2}{\operatorname{E} X^2} \tag{53}$$

Suppose X is a nonnegative random variable. From the trivial inequalities

$$\operatorname{E}(X; X \leq K) \leq \operatorname{E}(K; X \leq K) \leq \operatorname{E} K = K \tag{54}$$

it follows that $\operatorname{E}(X; X > K) \geq \operatorname{E} X - K$. Therefore taking $K = \theta \operatorname{E} X$

$$\operatorname{E}(X; X > \theta \operatorname{E} X) \geq (1-\theta) \operatorname{E} X \tag{55}$$

By Cauchy-Schwartz

$$(\operatorname{E}(X; X > \theta \operatorname{E} X))^2 = (\operatorname{E} X 1_{X > \theta \operatorname{E} X})^2 \leq \operatorname{E} X^2 \operatorname{E} 1^2_{X > \theta \operatorname{E} X} = (\operatorname{E} X^2) \Pr(X > \theta \operatorname{E} X) \tag{56}$$

By combining these two ineqaulities, the Paley-Zygmund inequality follows

$$(1-\theta)^2 (\operatorname{E} X)^2 \leq (\operatorname{E} X^2) \Pr(X > \theta \operatorname{E} X) \tag{57}$$

Note that the equation in exercise 1.41 is a special case of Paley-Zygmund involving nonnegative random variable $|X|$ and $\theta = 0$. ■

1.44 Let X be a non-negative random variable that is almost surely bounded but not identically zero. Show that

$$\Pr\left(X \geq \frac{1}{2} \operatorname{E} X\right) \geq \frac{1}{2} \frac{\operatorname{E} X}{\|X\|_\infty} \tag{58}$$

Using boundary case the Hölder inequality for $p = \infty, q = 1$ for the functions $1_{X \geq \theta \operatorname{E} X}$ and X we get

$$\operatorname{E} X 1_{X > \theta \operatorname{E} X} \leq \|X\|_\infty \operatorname{E} 1_{X > \theta \operatorname{E} X} = \|X\|_\infty \Pr(X > \theta \operatorname{E} X) \tag{59}$$

Equation (55) gives a lower bound for the left hand side. We conclude

$$\Pr(X > \theta \operatorname{E} X) \geq (1-\theta)\frac{\operatorname{E} X}{\|X\|_\infty} \tag{60}$$

and the problem statement is the case when $\theta = \frac{1}{2}$. The same argument shows for $p^{-1} + q^{-1} = 1$

$$\Pr(X > \theta \operatorname{E} X) \geq (1-\theta)^q \frac{(\operatorname{E} X)^q}{\|X\|_p^q} \tag{61}$$

■

Product measures and independence

2.3 (Finite Products)

■

2.4

■

2.6 Show that for any collection of probability spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$ for $i \in A$ there is at most one product measure μ_A.

hint:Adapt the uniqueness argument in theorem 1 with the monotone class lemma ■

2.7 Let $\mu_1, \ldots, \mu_n$ be probability measures on $\mathbb{R}$. Let $F_1, \ldots, F_n : \mathbb{R} \to [0,1]$ be their Stieltjes measure functions. Show that $\mu_1 \times \cdots \times \mu_n$ is the unique probability measure on $\mathbb{R}^n$ whose Stietljes transform is the tensor product $(t_1, \ldots, t_n) \mapsto F_1(t_1) \ldots F_n(t_n)$ of $F_1, \ldots, F_n$

■

2.8 Extension Problem

■

2.9 Show that for any $(\Omega_i, \mathcal{F}_i)_{i \in A}$ and μ_B for finite $B \subset A$ as in the above extension problem there is at most one probability measure μ_A with the stated properties.

■

2.13 Kolmogorov extension theorem alternative form

■

2.15

■

2.18 A random variable X is independent of itself if and only if X is almost surely equal to a constant.

For any $A \subset R$,

$$\Pr(\{X \in A\}) \Pr(\{X \in A\} \cap \{X \in A\}) = \Pr(\{X \in A\}) \Pr(\{X \in A\}) \tag{62}$$

Thus $\Pr(X \in A) = 0$ or $\Pr(X \in A) = 1$. Thus if $A, B \subset R$ are disjoint, then at most one of $\{X \in A\}$ and $\{X \in B\}$ has positive probabilty, and that set is almost sure. For note, if $\Pr(X \in A) > 0$ then $\Pr(X \in B) = 1$ and similarly for $\Pr(X \in B)$. But then $\Pr(X \in A \cup B) = \Pr(\{X \in A\} \cup \{X \in B\}) = \Pr(X \in A) + \Pr(X \in B) = 2$ which is a contradiction. If R is a σ-compact complete metric space, then we can find a series of nesting sets $A_1 \supset A_2 \supset \ldots$ such that the diameter of $A_n \to 0$ with $\Pr(A_n) = 1$ and $\Pr(A_n^c) = 0$. By completeness $\bigcap_n A_n = \{x\}$ for some singleton with $\Pr(X = x) = 1$ ■

2.19 Show that a constant (deterministic) random variable is independent of any other random variable

Let N be a null event with $\Pr(N) - 0$. Then by monotonicity, any measurable event $N' \subset N$ is also null since $\Pr(N') \leq \Pr(N) = 0$. Suppose S is an almost sure event with $\Pr(S) = 1$. Then for any other event E, $\Pr(E) = \Pr(E \cap S) + \Pr(E \cap S^c) = \Pr(E \cap S)$ since $E \cap S^c$ is a subset of a null set.

Say $X : \Omega \to R$ is constant so that $X = c$ almost surely. For any $A \subset R$, if X is constant than $\Pr(X \in A) = 1$ or $\Pr(X \in A) = 0$ depending on whether $c \in A$ or not. Then for any other random variable $Y : \Omega \to S$ and $B \subset S$

$$\Pr(\{X \in A\} \cap \{Y \in B\}) = \begin{cases} 0 & = 0 \cdot \Pr(Y \in B) \text{ if } c \notin A \\ \Pr(Y \in B) & = 1 \cdot \Pr(Y \in B) \text{ if } c \in B \end{cases} \tag{63}$$

So the independence formula is satisfied for all A and B. ■

2.20 Let $X_1, \ldots, X_n$ be discrete random variables (i.e., each takes on contably many values). Show taht $X_1, \ldots, X_n$ are jointly independent if and only if

$$\Pr(\bigcap_{i=1}^{n} \{X_i = x_i\}) = \prod_{i=1}^{n} \Pr(X_i = x_i) \tag{64}$$

The "only if" is true because we can use the independence formula (16 in Tao's notes) on the events $\{X_i = x_i\}$ to arrive at the formula above. For the "if" part, it follows exercise 6 and the following

$$\begin{aligned}
\prod_{i=1}^{n} \Pr(X_i \in S_i) &= \prod_{i=1}^{n} \left(\sum_{x_i \in S_i} \Pr(X_i = x_i) \right) \\
&= \sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} \prod_{i=1}^{n} \Pr(X_i = x_i) \\
&= \sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} \Pr(X_1 = x_1, \dots, X_n = x_n) \\
&= \Pr(X_1 \in S_1, \dots, X_n \in S_n)
\end{aligned} \tag{65}$$

■

2.21 Let $X_1, \dots, X_n$ be real scalar random variables. Show that $X_1, \dots, X_n$ are jointly independent if and only if

$$\Pr(\bigcap_{i=1}^{n} \{X_i \leq x_i\}) = \prod \Pr(X_i \leq t_i) \tag{66}$$

The "only if" is true because we can use the independence formula (16) on the events $\{X_i \leq t_i\}$. For "if" note that the independence criterion says that, if $X_1, \dots, X_n$ are jointly independent, the joint probability is given by the product measure of the marginal probability measures of each random variable. The formula in the problem statement is just the formula in 2.7 for the Stieltjes transform of a product measure in probability terms. As that problem shows, the product measure is the unique measure satisfying the formula. ■

2.22 Let V be a finite dimensional vector space over a finite field F and let X be a random variable drawn uniformly from V. Let $\langle , \rangle \to F$ be a non-degenerate bilinear form on V and let $v_1, \dots, v_n$ be non-zero vectors V. Show that the random variables $\langle X, v_1 \rangle, \dots, \langle X, v_n \rangle$ are jointly independent if and only if the vectors $v_1, \dots, v_n$ are linearly independent.

Let $d = \dim V$ and let $q = |F|$, so that V has q^d elements. Let $v_1, \dots, v_d$ be any basis for V, so that any $v \in V$ can be written uniquely as $v = f_1 v_1 + \cdots + f_n v_n$ with $f_i \in F$. Then each coefficient f_i is uniform over F and the f_i are jointly independent. This follows from the fact that for any $f = F$, there are q^{d-1} elements of V with $f_i = f$, so the probability $f_i = f$ is $q^{d-1}/q^d = 1/q$. Similarly, if we fix any k coeffients, there are q^{d-k}, so the probability of of such an element is $q^{d-k}/q^d = 1/q^k$. This proves indepdendence for singleton sets, and the general case follows from 2.20.

If $v_1, \dots, v_n$ are linearly independent, extend them to a basis $v_1, \dots, v_d$. Let $V' = F^d$ be the canonical d-dimensional vector space over F, and define a linear transformation

$T : V \to V'$ by $X \mapsto (\langle X, v_1 \rangle, \ldots, \langle X, v_d \rangle)$. Since the v_i are linearly independent and $\langle , \rangle$ is non-degernate, T is one-to-one. Since V and V' have the same dimension, T is onto. Let $w_i = T^{-1}(e_i)$ where e_i is the d-tuple which is 0 except in the ith component, which is 1. Then the w_i are a basis of V which satisfies $X = \sum_{i=1}^{d} \langle X, v_i \rangle w_i$ for every X. From the argument above, the quantities $\langle X, v_i \rangle$ are jointly independent.

Conversely, if $c_1 v_1 + \cdots + c_n v_n = 0$ for some $c_1, \ldots, c_n$ not all 0 (without loss of generality, assume $c_1 \neq 0$), then

$$\langle X, v_1 \rangle = -c_1^{-1}(c_2 \langle X, v_2 \rangle + c_3 \langle X, v_3 \rangle + \cdots + c_n \langle X, v_n \rangle) \tag{67}$$

So given $\langle X, v_2 \rangle \ldots \langle X, v_n \rangle$, the quantity $\langle X, v_1 \rangle$ is constant rather than uniformly distributed over F. Therefore its not independent of the other values. ■

2.23 Given an example of three random variables X, Y, Z which are *pairwise independent* but not *jointly independent*.

Let X, Y be independent and uniformly distributed over $\mathbb{Z}_2$ and let $Z = X + Y$. This is the same as previous exercise letting $V = \mathbb{Z}_2 \times \mathbb{Z}_2$, $v_1 = (1, 0)$, $v_2 = (0, 1)$ and $v_3 = (1, 1)$. Pairwise independence follows from the "if" part, and the failure of joint independence follows from the "only if" part. ■

2.24 Let X be a random variable taking values on $\mathbb{R}^n$ with the Gaussian distribution, in the sense that

$$\Pr(X \in S) = \frac{1}{(2\pi)^{n/2}} = \int_S e^{-|x|^2/2}\, dx \tag{68}$$

(where $|x|$ denotes the Euclidean norm on $\mathbb{R}^n$) and let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^n$. Show that the random variables $X \cdot v_1, \ldots, X \cdot v_m$ are jointly independent if and only if the $v_1, \ldots, v_m$ are pairwise orthogonal TYPO

If the $v_1, \ldots, v_m$ are pairwise orthogonal, we can extend them to an orthonormal basis $v_1, \ldots, v_n$ and let H be the orthogonal transformation which translates the standard basis to this one. Letting $Y = H^{-1}X$ and note

$$\begin{aligned}
\Pr(Y \in S) &= \frac{1}{(2\pi)^{n/2}} \int_{HS} e^{-|x|^2/2}\, dx \\
&= \frac{1}{(2\pi)^{n/2}} \int_S e^{-|H^{-1}y|^2/2}\, |\det H^{-1}|\, dy \\
&= \frac{1}{(2\pi)^{n/2}} \int_S e^{-|y|^2/2}\, dy
\end{aligned} \tag{69}$$

The last line follows because $|\det H| = 1$ and $|X| = |HX|$ for orthogonal transformations. But this shows that Y has the same distribution as X. The components of multivariate

Gaussian are jointly independent since if $S = S_1 \times \cdots \times S_n$ then

$$\begin{aligned}
\Pr(X \in S) &= \frac{1}{(2\pi)^{n/2}} \int_{S_1 \times \cdots \times S_n} e^{-x_1^2 + \cdots + x_n^2/2}\, dx \\
&= \left(\frac{1}{\sqrt{2\pi}} \int_{S_1} e^{-x_1^2}\, dx_1 \right) \cdots \left(\frac{1}{\sqrt{2\pi}} \int_{S_n} e^{-x_n^2}\, dx_1 \right) \\
&= \Pr(X_1 \in S_1) \cdots \Pr(X_n \in S_n)
\end{aligned} \tag{70}$$

Since the components of Y are $(X \cdot v_1, \ldots, X \cdot v_n)$, so these quantities are also jointly orthogonal.

Conversely, suppose $Y_1 = X \cdot v_1, \ldots, Y_n = X \cdot v_n$ are jointly independent. Each is a Gaussian since is the linear combination of Gaussian random variables, with $\mathrm{E}\, Y_k = (\mathrm{E}\, X) \cdot v_k = 0$ and $\mathrm{Var}\, Y_k = |v_k|^2$. Since functions of independent random variables are also independent, we can rescale if necessary to get $|v_k| = 1$ without affecting the independence hypothesis. If $m < n$, exten to $v_1, \ldots, v_n$ where the $v_{m+1}, \ldots, v_n$ are chosen to be orthonormal and orthogonal to $v_1, \ldots, v_n$. (This can be done by extending to any basis then creating orthornomal vectors in a way similar to the Graham-Schmidt algorithm). Let P be any parallelpiped with edges parallel to $v_1, \ldots, v_n$ such that $v_k \cdot x$ ranges over some interval $[a_k, b_k]$.

$$\begin{aligned}
\frac{1}{(2\pi)^{n/2}} \int_P e^{-|x|^2/2}\, dx &= \Pr(X \in P) \\
&= \Pr(Y_1 \in [a_1, b_1]) \ldots \Pr(Y_n \in [a_n, b_n]) \\
&= \left(\frac{1}{\sqrt{2\pi}} \int_{a_1}^{b_1} e^{-u_1^2/2}\, du_1 \right) \cdots \left(\frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} e^{-u_n^2/2}\, du_n \right) \\
&= \frac{1}{(2\pi)^{n/2}} \int_P e^{-((v_1 \cdot x)^2 + \cdots + (v_n \cdot x)^2)/2}\, dx
\end{aligned} \tag{71}$$

Since this equation holds for all parallelpipeds P and the parallelpipeds generate the Borel σ-algebra, almost surely it must be the case that

$$\exp(-|x|^2/2) = \exp(-((v_1 \cdot x)^2 + \cdots + (v_n \cdot x)^2)/2) = \exp(-|Mx|^2) \tag{72}$$

where M is the matrix whose rows are given by $v_1, \ldots, v_n$. But this is possible for all x only if M is an isometry and $|x| = |Mx|$, which happens only if M is orthogonal. Thus $v_1, \ldots, v_m$ are orthonormal, and by our construction, this implies $v_1, \ldots, v_m$ are orthonormal. ∎

2.25

(i) Show that two events E, F are independent if and only if

$$\Pr(E \cap F) = \Pr(E)\Pr(F) \tag{73}$$

(ii) If E, F, G are events show that the condition $\Pr(E \cap F \cap G) = \Pr(E)\Pr(F)\Pr(G)$ is necessary but not sufficient to show that E, F, G are jointly independent.

(iii) Give (TYPO) an example of three events E, F, G that are pairwise independent but not jointly independent

(i) There are four cases to consider.

(a) $\Pr(E \cap F) = \Pr(E)\Pr(F)$ by hypothesis.

(b) Now note that $\Pr(E) = \Pr(E \cap F^c) + \Pr(E \cap F)$. Therefore $\Pr(E \cap F^c) = \Pr(E) - \Pr(E)\Pr(F) = \Pr(E)(1 - \Pr(F)) = \Pr(E)\Pr(F^c)$.

(c) Reversing the roles of F and E, $\Pr(E^c \cap F) = \Pr(E^c)\Pr(F)$

(d) $\Pr(E^c) = \Pr(E^c \cap F) + \Pr(E^c \cap F^c)$ so

$$\begin{aligned}\Pr(E^c \cap F^c) &= 1 - \Pr(E) - (1\Pr(E))\Pr(F^c) \\ &= (1 - \Pr(E))\Pr(F^c) \\ &= \Pr(E^c)\Pr(F^c)\end{aligned} \tag{74}$$

(ii) The condition is necessary since its the formula which corresponds to

$$\Pr(1_E = 1 \cap 1_F = 1 \cap 1_G = 1) = \Pr(1_E = 1)\Pr(1_F = 1)\Pr(1_G = 1) \tag{75}$$

However, this formula doesn't imply, for example, $\Pr(E \cap F) = \Pr(E)\Pr(F)$. Take a (convex) region with unit area. Choose some (convex) sub-region A with area equal to 1/27. From the remaining 26/27, choose three regions E_0, F_0, G_0 with areas equal to 1/3-1/27 = 8/27. The remaining region N has area 2/27. Choose a point uniformly and let event $E = E_0 \cup A$, and $F = F_0 \cup A$ and $G = G_0 \cup A$. Then $\Pr(E \cap F \cap G) = 1/27 = \Pr(E)\Pr(F)\Pr(G)$ but $\Pr(E \cap F) = \Pr(A) = 1/27$ as well. Thus the events are not independent.

(iii) Take $E = \{X = 1\}$ and $F = \{Y = 1\}$ and $G = \{X + Y = 1\}$ in problem

■

2.27 Let $\epsilon_1, \epsilon_2, \cdots \in \{0, 1\}$ be random variables that are indepenedent and identically distributed copies of the Bernoulli random variable with expectation 1/2, that is to say the $\epsilon_1, \epsilon_2, \ldots$ are jointly independent with $\Pr(\epsilon_i = 1) = \Pr(\epsilon_0) = 1/2$ for all i.

(i) Show that the random variable $\sum_{n=1}^{\infty} 2^{-n}\epsilon_i$ is uniformly distributed on the unit interval $[0, 1]$.

(ii) Show that the random variable $\sum_{n=1}^{\infty} 2 \times 3^{-n}\epsilon_i$ has the distribution of the Cantor measure.

■

2.28 Give an example of two square integrable real varaible X, Y which have vanishing $\mathrm{Cov}(X, Y)$ but are not independent.

Let $S_1 = \{(x, y) : x = y, -1 \leq x, y \leq 1\}, S_2 = \{(x, y) : x = -y, -1 \leq x, y \leq 1\}, S = S_1 \cup S_2$. Choose a point from S uniformly, let X be the x-coordinate of the point and Y be the y-coordinate. Then X and Y are uniformly distributed on $[-1, 1]$ (and hence have expectation 0). Also

$$\mathrm{Cov}(X, Y) = \int_S XY = \int_{S_1} XY + \int_{S_2} XY = \int_{-1}^{1} u^2\, du + \int_{-1}^{1} -u^2\, du = 0 \tag{76}$$

However, conditional on X, Y has the discrete uniform distribution on $\{X, -X\}$ rather than the uniform distribution over $[-1, 1]$. THerefore X and Y are not independent.

■

2.29

■

2.31 If $(X_i)_{i \in A}$ are a collection of random variables, show that $(X_i)_{i \in A}$ are jointly independent random variables if and only if $(\sigma(X_i))_{i \in A}$ are jointly independent σ-algebras

■

2.32 Let $X_1, X_2, \ldots$ be a sequence of random variables. Show that $(X_n)_{n=1}^{\infty}$ are jointly independent if and only if $\sigma(X_{n+1})$ is independent of $\sigma(X_1, \ldots, X_n)$ for all natural numbers n.

■

Weak and Strong Law of Large Numbers

3.2 Modes of Convergence Let X_n be a sequence of scalar random variables and let X be another random variable.

(i) If $X_n \to X$ almost surely, show that $X_n \to X$ in probability. Give a counterexample to show the converse does not necessarily hold.

(ii) Suppose that $\sum_n \Pr(|X_n - X| > \epsilon) < \infty$ for all $\epsilon > 0$. Show that $X_n \to X$ almost surely. Give a counterexample to show that the converse does not necessarily hold.

(iii) If $X_n \to X$ in probability. Show there is a subsequence X_{n_j} of the X_n such that $X_{n_j} \to X$ almost surely.

(iv) If X_n, X are absolutely integrable and $\mathrm{E}|X_n - X| \to 0$ as $n \to \infty$ show that $X_n \to X$ in probability. Give a counterexample to show thte converse does not necessarily hold.

(v) (Urysohn subsequence principle) Suppose that every X_{n_j} of X has a further subsequence $X_{n_{j_k}}$ that converges to X in probability. Show that X_n also converges in X in probability.

(vi) Does Urysohn's subsequence principle hold if "in probability" is replaced with "almost surely"?

(vii) If X_n converges in probabiity to X and $F : \mathbb{R} \to \mathbb{R}$ or $F : \mathbb{C} \to \mathbb{C}$ is continuous, show that $F(X_n)$ converges in probability to $F(X)$. More generally, if for each $i = 1, \ldots, k$ $X_n^{(i)}$ is a sequence of scalar random variables that converge in probability to $X^{(i)}$ and $F : \mathbb{R}^k \to \mathbb{R}$ or $F : \mathbb{C}^k \to \mathbb{C}$ is continuous, show that $F(X_n^{(1)}, \ldots, X_n^{(k)})$ converges in probability to $F(X^{(1)}, \ldots, X^{(k)})$

(viii) (Fatou's lemma for convergence in probability) If X_n are non-negative and converge in probability to X show that $\mathrm{E}\, X \le \liminf_{n\to\infty} \mathrm{E}\, X_n$.

(ix) (Dominated convergence in probabilty) If X_n converge in probabilitiy to X and one alomst surely has $|X_n| \le Y$ for all n and some absolutely integrable Y, show that $\mathrm{E}\, X_n$ converges to $\mathrm{E}\, X$

Lemma 1. *Let $E_1, E_2, \ldots$ be events. Then*

$$\Pr(E_k \text{ eventually}) \le \liminf_{n\to\infty} \Pr(E_k) \tag{77}$$

Proof. Let 1_{E_k} be the indicator for E_k. Then $I = \liminf_k 1_{E_k}$ is the indicator for the event that E_k happens eventually. For if $I(\omega) = 1$ then for large enough n $\omega \in E_n$ and the indicators are all 1. Conversely if $I(\omega) = 0$ then $\omega \notin E_k$ infinitely often. (Similarly $\limsup_k 1_{E_k}$ is the indicator for the event that E_k happens infinitely often). By Fatou's lemma

$$\Pr(E_k \text{ eventually}) = \mathrm{E}(\liminf_k 1_{E_k}) \le \liminf_k \Pr(E_k) \tag{78}$$

$\square$

(i) By the lemma,

$$\Pr(|X - X_n| \le \epsilon \text{ eventually}) \le \liminf_n \Pr(|X - X_n| \le \epsilon) \tag{79}$$

If $X_n \to X$ almost surely, the left hand side is 1, so the right hand side is also 1. Therefore $\limsup_n \Pr(|X - X_n| > \epsilon) = 0$. Since the liminf is at least 0 and is bound by the limsup, the two quantities are equal and the limsup is actually the limit. Therefore X_n converges in probability. TODO counterexample for converse

(ii) By Borel-Cantelli, for all ϵ, $|X_n - X| \ge \epsilon$ eventually with probability 1. But this is essentially the definition of almost sure convergence. TODO converse is false

(iii) Let $p_n = \Pr(|X - X_n| > \epsilon)$. By hypothesis, $p_n \to 0$ as $n \to \infty$, so we can find a subsequence such that $p_{n_k} \le 2^{-k}$. The X_{n_k} satisfy the hypothesis of (ii) and therefore converge almost surely.

(iv) Let $p_n = \Pr(|X - X_n| > \epsilon)$. There exists a subsequence such that $p_{n_k} \to \limsup p_n$. By hypothesis, there is a subsequence such that $p_{n_{k_j}} \to 0$. But subsequences of sequences which have a limit must converge to the same limit. Therefore $\limsup p_n = 0$ and hence (since $\liminf p_n \ge 0$ for a non-negative sequence) $\lim p_n = 0$. But this is precisely the condition that $X_n \to X$ in probability.

(v) Let $p_n = \Pr(|X - X_n| \ge \epsilon \text{ infinitely often})$. Using exactly the same reasoning on p_n as in (iv) $\lim p_n \to 0$.

(vi) Markov's inequality says that for every n and $\epsilon > 0$

$$\Pr(|X_n - X| > \epsilon) \le \frac{\mathrm{E}|X_n - X|}{\epsilon} \tag{80}$$

By assumption, the limit of the right hand side is 0, so the limit of the left hand side is also 0. TODO counterexample for converse

(vii) By continuity, for any $\epsilon > 0$ we can choose δ such that $|X - X_n| \le \delta$ implies that $|F(X) - F(X_n)| \le \epsilon$. Therefore $\{|F(X) - F(X_n)| > \epsilon\} \subset \{|X - X_n| > \delta\}$. Hence

$$\limsup_n \Pr(|F(X) - F(X_n| > \epsilon) \le \lim \Pr(|F(X) - F(X_n)| > \delta) = 0 \tag{81}$$

which shows that $F(X) \to F(X_n)$ in probability.

For $F : \mathbb{R}^n \to \mathbb{R}$ continuity implies that $\|X_n - X\|_2 = \sqrt{\sum_i (X^{(i)} - X_n^{(i)})^2} < \delta$ implies $|F(X) - F(X_n)| < \epsilon$. Thus if $|X^{(i)} - X_n^{(i)}| < \delta/\sqrt{k}$ for all n then $|F(X) - F(X_n)| < \epsilon$. Taking complements,

$$\{|F(X) - F(X_n)| \ge \epsilon\} \subset \bigcup_i \{|X^{(i)} - X_n^{(i)}| \ge \delta/\sqrt{k}\} \tag{82}$$

Therefore

$$\Pr(|F(X) - F(X_n)| \geq \epsilon) \leq \sum_i \Pr(|X^{(i)} - X_n^{(i)}| \geq \delta/\sqrt{k}) \tag{83}$$

The right hand side tends to 0 as $n \to \infty$ so $F(X_n) \to F(X)$ converges in probability.

(viii) For any sequence Y_n, since $X \to X_n$ in probability, for any $\epsilon > 0$ $\liminf_n 1_{|X > X_n| < \epsilon} = 1$ almost surely. That's because eventually $1_{|X > X_n| < \epsilon}$ with probability 1, and the liminf only depends on the tail values of Y_n. Therefore using the fact that X and X_n are close on $1_{|X - X_n| < \epsilon}$, Fatou's lemma, and the fact that X_n is non-negative.

$$\begin{aligned} \mathrm{E}\, X &= \mathrm{E} \liminf_n X 1_{|X - X_n| < \epsilon} \\ &\leq \mathrm{E} \liminf_n X_n 1_{|X - X_n| < \epsilon} + \epsilon \\ &\leq \liminf_n \mathrm{E}\, X_n 1_{|X - X_n| < \epsilon} + \epsilon \\ &\leq \liminf_n \mathrm{E}\, X_n + \epsilon \end{aligned} \tag{84}$$

Since this holds for all $\epsilon > 0$, the inequality in the problem statement holds

(ix) Let $I_n = 1_{|X - X_n| < \epsilon}$. Note $\lim_n I_n = 1$ almost surely because $X_n \to X$ in probability, and also $|X_n I_n - X I_n| < \epsilon$. Note

$$|\mathrm{E}\, X_n - \mathrm{E}\, X| \leq |\mathrm{E}\, X_n I_n - \mathrm{E}\, X_n| + |\mathrm{E}\, X I_n - \mathrm{E}\, X| + |\mathrm{E}\, X_n I_n - \mathrm{E}\, X I_n| \tag{85}$$

Now $|\mathrm{E}\, X_n I_n - \mathrm{E}\, X_n| \leq \mathrm{E}|(1 - I_n) X_n| \leq \mathrm{E}(1 - I_n) Y$. Now $(1 - I_n) Y \leq Y$ so by dominated convergence $E(1 - I_n) Y \to 0$ since the integrand converges almost surely to 0. Similarly $|\mathrm{E}\, X I_n - \mathrm{E}\, X| \leq \mathrm{E}|X|(1 - I_n)$. The integrand is dominated by $|X|$ and converges almost surely to 0, so the integral tends to 0. Finally $|\mathrm{E}\, X_n I_n - \mathrm{E}\, X I_n| \leq \mathrm{E}|X_n I_n - X I_n| \leq \epsilon$. But ϵ is arbitrary, and we conclude that $\limsup_n |\mathrm{E}\, X_n - \mathrm{E}\, X|$ is bound by an arbitrarily small number and therefore must be 0.

∎

3.3 Let $X_1, X_2, \ldots$ be a sequence of scalar random variables converging in probability to anothe random variable X.

(i) Suppose Y is independent of each X_i. Show Y is independent of X

(ii) Suppose the $X_1, X_2, \ldots$ are jointly independent. Show X is almost surely constant.

Its a little subtle to show, but convergence in probability implies convergence in distribution. For any constants c, r consider the ball centered at c with radius r. Note $|X_n - c| \geq r$ if $|X - c| \geq r + \epsilon$ and $|X - X_n| \leq \epsilon$. So, taking complements

$$\{|X_n - c| < r\} \subset \{|X - c| < r + \epsilon\} \cup \{|X - X_n| > \epsilon\} \tag{86}$$

So the probabilities satisfy

$$\begin{aligned}\lim_n \Pr(|X_n - c| < r) &\leq \Pr(|X - c| < r + \epsilon) + \lim_n \Pr(|X_n - X| > \epsilon) \\ &= \Pr(|X - c| < r + \epsilon)\end{aligned} \tag{87}$$

Similarly since $|X - c| \geq r - \epsilon$ if $|X_n - c| \geq r$ and $|X - X_n| \leq \epsilon$

$$\Pr(|X - c| < r - \epsilon) \leq \lim_n \Pr(|X_n - c| < r) \tag{88}$$

Letting $\epsilon \to 0$ in These inequalities, the limit from above/limit from below properties of probability measure imply that for the set $B_r(c)$, the ball of radius r centered at c.

$$\lim_n \Pr(X_n \in B_r(c)) = \Pr(X \in B_r(c)) \tag{89}$$

Let $\mathcal{F}$ be the collection of measurable sets A such that $\lim_n \Pr(X_n \in A) = \Pr(X \in A)$. This is a monotone class by the upward and downward continuity properties of probability. Its also an algebra (? is that right, what about intersections?). For example

$$\lim_n \Pr(X_n \in A^c) = 1 - \lim_n \Pr(X_n \in A) = 1 - \Pr(X \in A) = \Pr(X \in A^c) \tag{90}$$

so $\mathcal{F}$ is closed under complements.

In any case

(i) By the independence of Y and X_n, and the the above property for limits of distributions

$$\begin{aligned}\Pr(X \in A, Y \in B) &= \lim_{n\to\infty} \Pr(X_n \in A, Y \in B) \\ &= \lim_{n\to\infty} \Pr(X_n \in A) \Pr(Y \in B) \\ &= \Pr(X \in A) \Pr(Y \in B)\end{aligned} \tag{91}$$

Since A and B are arbitrary, this shows that X and Y are independent.

(ii) For any $n \geq 1$, we have $X = \lim_{k\to\infty} X_{n+k}$ in probability, which, by mutual independence, means that X is independent of $X_1, X_2, \ldots, X_n$. Since X is indenpendent of every finite subcollection of $X_1, X_2, \ldots$, this means that X is independent of the limit. This shows X is independent of itself, and, hence, that X is a constant (by 2.18)

■

3.9 If X is geometric distribution with parameter p for some $0 < p \leq 1$ then X has mean $\frac{1}{p}$ and variance $\frac{1-p}{p}$

First note these identities, found by differentiating the geometric series term-by-term

$$S_1(x) = \sum_{k \geq 0} kx^{k-1} = \frac{d}{dx} \sum_{k \geq 0} x^k = \frac{d}{dx} \frac{1}{1-x} = \frac{1}{(1-x)^2} \tag{92}$$

$$S_2(x) = \sum_{k \geq 0} k(k-1)x^{k-1} = \frac{d}{dx} \sum_{k \geq 0} kx^{k-1} = \frac{d}{dx} \frac{1}{(1-x)^2} = \frac{2}{(1-x)^3} \tag{93}$$

For X, discrete with $\Pr(X = k) = p_k = (1-p)^{k-1}p$, we use the above identities to get

$$\mathrm{E}\, X = \sum_k k(1-p)^{k-1} p = pS_1(1-p) = \frac{1}{p} \tag{94}$$

$$\mathrm{E}\, X^2 = \sum_k k^2 (1-p)^{k-1} p = p(1-p)S_2(1-p) + pS_1(1-p) = \frac{2-p}{p^2} \tag{95}$$

$$\mathrm{Var}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 = \frac{1-p}{p^2} \tag{96}$$

■

3.10 (Second Borel-Cantelli lemma) Let $E_1, E_2, \ldots$ be a sequence of jointly independent events. If $\sum_{n=1}^{\infty} \Pr(E_n) = \infty$ show that almost surely an infinite number of the E_n hold simultaneously.

Consider $S_n = \sum_{i=1}^{n} 1_{E_i}$.

$$s_n := \mathrm{E}\, S_n = \sum_{i=1}^{n} \Pr(E_i) \tag{97}$$

$$v_n := \mathrm{Var}\, S_n = \sum_{i=1}^{n} \mathrm{Var}(1_{E_i}) = \sum_{i=1}^{n} \Pr(E_i) - \Pr(E_i)^2 \leq s_n \tag{98}$$

Therefore by Chebyshev's inequality

$$\Pr\left(|S_n - s_n| \geq \frac{1}{2} s_n \right) \leq \frac{4v_n}{s_n^2} \leq \frac{4}{s_n} \to 0 \tag{99}$$

However, the event that $\lim_{n\to\infty} S_n$ is finite is a subset of the event that $S_n \leq s_n/2$ eventually. If $p_n = \Pr(S_n \leq s_n/2)$ then the probability this inequality holds eventually is bound by $\liminf_{n\to\infty} p_n$, which is 0 by the above inequality. Thus, almost surely S_n is infinite and, hence, an infinite nuber of the E_k occur. ■

3.11 (Infinite Monkey Theorem) Let $X_1, X_2, \ldots$ be iid random variabls drawn uniformly from a finite alphabet A. Show that almost surely, every finite word $a_1 a_2 \ldots a_k$ appears infinitely often in the string $X_1 X_2 X_2 \ldots$

Let T_n be the event that $X_{nk+1} = a_1, X_{nk+2} = a_2, \dots, x_{n(k+1)} = a_k$. The T_n are all independent and $\Pr(T_n) = |A|^{-k} > 0$. Hence $\sum_n \Pr(T_n) \to \infty$ and therefore by the second Borel-Cantelli theorem, infinitely many of the T_n occur. ∎

3.12 (Triangular Arrays) Let $(X_{i,n})_{i,n\in\mathbb{N};i\le n}$ be a triangular array of scalar random variables $X_{i,n}$ such that for each n the row $X_{1,n}, \dots, X_{n,n}$ is a collection of independent random variables. For each n, we form the partial sums

$$S_n = X_{1,n} + \cdots + X_{n,n} \tag{100}$$

(i) *(Weak law)* If all the X_n have mean μ and $\sup_{i,n}|X_{i,n}|^2 < \infty$, show that S_n/n converges in probability to μ

(ii) *(Strong law)* If all the $X_{i,n}$ have mean μ and $\sup_{i,n}|X_{i,n}|^4 < \infty$ show that S_n/n converges almos surely to μ

(i) Let $M = \sup_{i,n} \mathrm{E}|X_{i,n}|^2$. Note $\operatorname{Var} X_{i,n} = \mathrm{E}\, X_{i,n}^2 - \mu^2 \le M$ as well, so $\operatorname{Var} S_n \le nM$. By Chebyshev's lemma

$$\Pr(|S_n/n - \mu| > \epsilon) \le \frac{M}{\epsilon^2 n} \to 0 \tag{101}$$

This shows that $S_n/n \to \mu$ in probability.

(ii) For any $q > p$, set $r = q/p > 1$l and use Hölder's inequality to find

$$\mathrm{E}\, X^p \le (\mathrm{E}(X^p)^r)^{1/r}(\mathrm{E}\, 1^s)^{1/s} = (\mathrm{E}\, X^q)^{1/r} \qquad \text{where } \frac{1}{r} + \frac{1}{s} = 1 \tag{102}$$

Hence if $p < q$ then $\|X\|_p \le \|X\|_q$. Now supposing for $p \in \mathbb{N}$ we have $\sup_i \|X_i\|_p < \infty$. Then also we have for any constant $a \in \mathbb{R}$, $\sup_i \|X_i + a\|_p < \infty$. This follows from

$$\mathrm{E}(X_i + a)^p = \sum_{k=0}^{p} \binom{p}{k} a^{p-k}\, \mathrm{E}\, X_i^k \le \sum_{k=0}^{n} \binom{p}{k} a^{p-k} \|X_i\|_p^k = (\|X_i\|_p + a)^p \tag{103}$$

Hence $\sup_i \|X_i + a\|_p \le \sup_i \|X_i\|_p + a < \infty$.

Let $N = \sup_{i,n} \mathrm{E}|X_{i,n} - \mu|^4 < \infty$. Let $Z_{i,n} = X_{i,n} - \mu$, and let's compute using Markov's inequality

$$\begin{aligned}
\Pr(|\sum_i X_{i,n} - \mu n| \ge \epsilon n) &\le \frac{\mathrm{E}(\sum_i X_{i,n} - \mu n)^4}{\epsilon^4 n^4} \\
&= \frac{1}{\epsilon^4 n^4}\left(\sum_{i\ne j} \operatorname{Var} X_{i,n} \operatorname{Var} X_{i,n} + \sum_k \mathrm{E}(X_{i,k})\right) \qquad (104) \\
&\le \frac{1}{\epsilon^4 n^4}\left(\binom{n}{2} M + nN\right) = O(n^{-2})
\end{aligned}$$

Thus by Borel-Cantelli, almost surely at most finitely many of the inequalities $|\sum_i X_{i,n}/n - \mu| \geq \epsilon$ are satisfied, which implies that $\sum_i X_{i,n}/n \to \mu$ as $n \to \infty$ almost surely.

■

3.13 An *Erdös-Renyi graph* (V, E) on n vertices is a random variable on the set of graphs such that the events that the edge $\{i, j\} \in E$ are jointly independent with probability p. For each n let (V_n, E_n) be an Erdös-Renyi graph with $p = \frac{1}{2}$.

(i) If $|E_n|$ is the number of edges, show that $|E_n|/\binom{n}{2} \to 1/2$ almost surely

(ii) If $|T_n|$ is the number of triangles in (V_n, E_n) show that $|T_n|/\binom{n}{3} \to 1/8$ in probability.

(iii) Show that in fact $|T_n|/\binom{n}{3} \to 1/8$ almost surely.

(i) Consider the triangular array $X_{\{i,j\},n}$ where the index runs over all subsets of $\{1, 2, \ldots, n\}$ of 2 elements $\{i, j\}$ and represents the event $\{i, j\} \in E_n$. From the definition of the Erdös-Renyi graph the $X_{\{i,j\},n}$ are independent with $\mathrm{E}\, X_{\{i,j\},n} = 1/2$. Furthermore $\mathrm{E}|X_{\{i,j\},n}|^p = 1$ for any $p > 1$ since its a binomial random variable. Hence by exercise 12, the row-wise averages converge to 1/2 almost surely. Since $|E_n| = \sum_{\{i,j\} \subset E_n} X_{\{i,j\},n}$, the law of large numbers is just the statement $|T_n|/\binom{n}{2} \to 1/2$

(ii) Consider the triangular array $Y_{\{i,j,k\},n}$ where the index runs over all subsets of $\{1, 2, \ldots, n\}$ of 2 elements $\{i, j\}$ and represents the event $\{\{i, j\}, \{j, k\}, \{k, i\}\} \subset E_n$. (In words, $X_{\{i,j,k\},n}$ is the event that the graph (V_n, E_n) contains the triangle with edges $\{i, j\}$ and $\{j, k\}$ and $\{k, i\}$). Clearly we have

$$Y_{\{i,j,k\},n} = X_{\{i,j\},n} X_{\{j,k\},n} X_{\{k,i\},n} \tag{105}$$

From this and independence, its immediate that $\mathrm{E}\, Y_{\{i,j,k\},n} = (1/2)^3 = 1/8$. Now $|T_n| = \sum_{\{i,j,k\} \subset [n]\}} Y_{\{i,j,k\},n}$ so let's compute $\mathrm{E}|T_n|^2$. By linearity this the sum of all expressions of the form $\mathrm{E}\, Y_{\{i,j,k\},n} Y_{\{i',j',k'\},n}$. Let's consider the cases

1) $\{i, j, k\}$ and $\{i', j', k'\}$ are disjoint. In this case

$$\begin{aligned} \mathrm{E}\, Y_{\{i,j,k\},n} Y_{\{i',j',k'\},n} &= \mathrm{E}\, X_{\{i,j\},n} X_{\{j,k\},n} X_{\{k,i\},n} X_{\{i',j'\},n} X_{\{j',k'\},n} X_{\{k',i'\},n} \\ &= (1/2)^6 = 1/64 \end{aligned} \tag{106}$$

There are $\binom{n}{3}\binom{n-3}{3}$ such cases.

2) If $|\{i, j, k\} \cap \{i', j', k'\}| = 1$. As a representative case, let $i = i'$ with the rest distinct. Even though the triangles share a vertex, all of the edges are disjoint, hence

$$\begin{aligned} \mathrm{E}\, Y_{\{i,j,k\},n} Y_{\{i,j',k'\},n} &= \mathrm{E}\, X_{\{i,j\},n} X_{\{j,k\},n} X_{\{k,i\},n} X_{\{i,j'\},n} X_{\{j',k'\},n} X_{\{k',i\},n} \\ &= (1/2)^6 = 1/64 \end{aligned} \tag{107}$$

There are $3\binom{n}{3}\binom{n-3}{2}$ such cases.

3) If $|\{i,j,k\} \cap \{i',j',k'\}| = 2$. As a representative case, let $i = i'$ and $j = j'$ with the rest distinct. The triangles share one vertex

$$\begin{aligned} \mathrm{E}\, Y_{\{i,j,k\},n} Y_{\{i,j,k'\},n} &= \mathrm{E}\, X^2_{\{i,j\},n} X_{\{j,k\},n} X_{\{k,i\},n} X_{\{j,k'\},n} X_{\{k',i\},n} \\ &= (1/2)^5 = 1/32 \end{aligned} \tag{108}$$

There are $3\binom{n}{3}\binom{n-3}{1}$ such cases.

4) If $|\{i,j,k\} \cap \{i',j',k'\}| = 3$. The triangles are the same

$$\begin{aligned} \mathrm{E}\, Y^2_{\{i,j,k\},n} &= \mathrm{E}\, X^2_{\{i,j\},n} X^2_{\{j,k\},n} X^2_{\{k,i\},n} \\ &= (1/2)^3 = 1/8 \end{aligned} \tag{109}$$

There are $\binom{n}{3}$ such cases.

From this we compute

$$\begin{aligned} \operatorname{Var}|T_n| &= \mathrm{E}\,|T_n|^2 - (\mathrm{E}|T_n|)^2 \\ &= \frac{1}{64}\binom{n}{3}\binom{n-3}{3} + \frac{3}{64}\binom{n}{3}\binom{n-3}{2} + \frac{3}{32}\binom{n}{3}\binom{n-3}{1} + \frac{1}{8}\binom{n}{3} - \frac{1}{64}\binom{n}{3}^2 \\ &= \frac{1}{64}\binom{n}{3}\left(\binom{n-3}{3} + 3\binom{n-3}{2} - \binom{n}{3}\right) + O(n^4) \\ &= O(n^4) \end{aligned} \tag{110}$$

Essentially the cancelation comes from the identity $\binom{n}{k} = \sum_{l=0}^{k}\binom{n-m}{k-l}\binom{m}{l}$. We get an extra term of cancelation because triangles with a common vertex are still independent. Hence from Chebyshev's inequality

$$\Pr\left(\left||T_n|/\binom{n}{3} - \frac{1}{8}\right| \geq \epsilon\right) \leq \operatorname{Var}|T_n|/\binom{n}{3}^2 \epsilon^2 = O(n^{-2}) \tag{111}$$

The right hand side tends to 0, so $|T_n|/\binom{n}{3} \to 1/8$ in probability

(iii) Since $\sum n^{-2}$ is summable, we can apply Borel-Cantelli to the probability (111) to show only almost surely only finitely many of the inequalities are violated. Hence, $|T_n|/\binom{n}{3} \to 1/8$ almost surely.

∎

3.14 For each n let $A_n = (a_{ij,n})_{1\le i,j\le n}$ be a random $n\times n$ matrix (i.e., a random variable taking values in the space $\mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$ of $n\times n$ matricies) such that the entries $a_{ij,n}$ of A_n are jointly independen in i,j and take values in $\{-1,+1\}$ with probability of $1/2$ each. (This is called a *random sign matrix*). We do not assume independence for the sequence $A_1, A_2, \ldots$.

(i) Show that the random variables $\operatorname{tr} A_n A_n^*/n^2$ are deterministically equal to 1

(ii) Show that for any natural number k the quantities $\operatorname{E}\operatorname{tr}(A_nA_n^*)^k/n^{k+1}$ are bounded uniformly in n.

(iii) if $\|A_n\|_{op}$ denotes the operator norm of A_n and $\epsilon > 0$ show that $\|A_n\|_{op}/n^{1/2+\epsilon}$ converges almost surely to 0 and that $\|A_n\|_{op}/n^{1/2-\epsilon}$ diverges almost surely to infinity. (Use the spectral theorem to relate $\|A_n\|_{op}$ with $\operatorname{tr}(A_nA_n^*)^k$)

(i) If $A_n = (a_{ij})$ and $B = (b_{ij})$ then $\operatorname{tr} AB = \sum_{ij} a_{ij}b_{ij}$. When $B = A^*$ then $b_{ij} = aji$ so $\operatorname{tr} AA^* = \sum_{ij} a_{ij}^2$. Since $a_{ij} = \pm 1$ each term in the sum is 1 and there are n^2 terms so $\operatorname{tr} A_nA_n^*/n^2 = 1$.

(ii) Let the rows of A_n be given by $X_{i,n}$. Then $X_{i,n}$ is a random vector with each component equal to ± 1. The matrix $A_nA_n^*$ is given by $(\langle X_{i,n}, X_{j,n}\rangle)$ where $\langle X, Y\rangle$ is the inner product of vectors X and Y. If $M = (m_{ij})$ then by induction we can show

$$\operatorname{tr} M^k = \sum_{1\le i_1,\ldots,i_k\le n} m_{i_1,i_2}m_{i_2,i_3}\cdots m_{i_{k-1},i_k}m_{i_k,i_1} \tag{112}$$

(In words, the subscripts are taken cyclically in pairs). Therefore we have

$$\begin{aligned}
\operatorname{E}(\operatorname{tr} AA^*)^k &= \sum_{1\le i_1,\ldots,i_k\le n} \operatorname{E}\langle X_{i_1,n}, X_{i_2,n}\rangle\langle X_{i_2,n}, X_{i_3,n}\rangle\cdots\langle X_{i_k,n}, X_{i_1,n}\rangle \\
&= \sum_{1\le i_1,\ldots,i_k\le n} \operatorname{E}\left(\sum_{1\le j_1\le n} a_{i_1j_1}, a_{i_2j_1}\right)\left(\sum_{1\le j_1\le n} a_{i_2j_2}a_{i_3j_2}\right)\cdots \\
&\qquad \left(\sum_{1\le j_n\le n} a_{i_kj_k}a_{i_1j_k}\right) \\
&= \sum_{\substack{1\le i_1,\ldots,i_k\le n\\ 1\le j_1,\ldots,j_k\le n}} \operatorname{E} a_{i_1j_1}a_{i_2j_1}a_{i_2j_2}a_{i_3j_2}\cdots a_{i_kj_k}a_{i_1j_k}
\end{aligned} \tag{113}$$

We consider this sum term by term. In a given term, we can group the various a_{ij} factors on. If any of them occurs an odd number of times, then the expectation of that term is 0 since, after grouping, the factors are independent and $\operatorname{E} a_{ij}^p = (-1)^p(1/2) + 1^p(1/2) = 0$. If all of them occurs an even number of times, then the term is 1. There appear to be n^{2k} terms, but we will show that all but $O(n^{k+1})$ are 0.

The indices i, j occur in a particular pattern which we can visualize as a cycle on a bipartite multi-graph. One part of the vertices consists of the values $\{1, \ldots, n\}$

which correspond to the i's and the other part consists of the values $\{1, \ldots, n\}$ which correspond to the j's, and an edge occurs in this graph if the factor a_{ij} occurs in the term. Thus each edge in the multigraph must occur an even number of times, and we can group those edges in pairs. For example, for any choice of $i_1, i_2, \ldots, i_k$ and j such that $j_1 = j_2 = \cdots = j_k = j$ then the term is of the desired form. Similarly for any choice $j_1, j_2, \ldots, j_k$ and i such that $i_1 = i_2 = \cdots = i_k = i$ then the term is of the desired form. (In terms of the multigraph, these cases correspond to a star pattern eminating from the single value on one part of the vertices).

We can group terms by their "cyclic shape" on the bipartate multigraph. That is, we can replace the the distinct values of the vertices i and the distinct values of j with any other collection of distinct values (such that there are the same number of distinct values) and trace the same cyclic path through the new vertices. If there are d_1 distinct i's and d_2 distinct j's in a given "cyclic shape" then there are $\binom{n}{d_1}\binom{n}{d_2} = O(n^{d_1+d_2})$ terms with the same shape. The number of shapes is a function of k not of n so as $n \to \infty$ this corresponds to a constant factor we need not concern ourselves with.

So to prove that $\mathrm{E}(\operatorname{tr} AA^*)^k / n^{k+1}$ is bounded uniformly in n we need show that $d_1 + d_2 \leq k + 1$. The number of distinct vertices on a path of length $2k$ is at most $k+1$. However since each edge must be paired with another edge in order that a term be non-zero, the second time an edge appears we are revisiting a vertex previously counted. Thus in the path at least k of vertices are revisted by the path. Thus at most $k+1$ vertices are visited on the cycle corresponding to this shape. The bound is tight, as shown by the "star" examples given above.

(iii) The operator norm satisfies

$$\|A\|_{op} = \sup_{\|v\|=1} \|Av\| = \sup_{\|v\|=1} \langle Av, Av\rangle^{1/2} = \sup_{\|v\|=1} \langle A^*Av, v\rangle^{1/2} \tag{114}$$

Now A^*A is symmetric and positive definite, so we can diagonalize the matrix with an orthogonal basis. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues and $v_1, \ldots, v_n$ the corresponding orthogonal eigenbasis. If $v = \sum_i a_i v_i$ then $\|v\|^2 = \sum_i a_i^2 = 1$ and $\|Av\|^2 = \sum_i \lambda_i a_i^2$. Clearly this is maximized when $a_1 = 1$ and $a_i = 0$ for $i > 1$. Thus $\|A\|_{op} = \lambda_1^{1/2}$.

Furthermore note that the eigenvalues of $(A^*A)^k$ are $\lambda_1^k, \ldots, \lambda_n^k$. Therefore $\operatorname{tr}(A^*A)^k = \sum_i \lambda_i^k$ so

$$\begin{aligned} \lim_{k\to\infty} (\operatorname{tr}(A^*A)^k)^{1/k} &= \lim_{k\to\infty} \lambda_1 (1 + (\lambda_2^k + \cdots + \lambda_n^k)/\lambda_1^k)^{1/k} \\ &\leq \lambda_1 \lim_{k\to\infty} \exp(\sum_{i=2}^{n} (\lambda_i/\lambda_1)^k / k) \\ &= \lambda_1 \end{aligned} \tag{115}$$

Since we also have the simple inequality $\operatorname{tr}(A^*A)^k \geq \lambda_1^k$,

$$\lim_{k\to\infty} (\operatorname{tr}(A^*A)^k)^{1/k} = \lambda_1 = \|A\|_{op}^2 \tag{116}$$

Hence

$$
\begin{aligned}
\Pr(\|A_n\|_{op}/n^{1/2+\epsilon} \geq \theta) &= \Pr(\|A_n\|_{op}^{2k} \geq \theta^{2k} n^{k+2k\epsilon}) \\
&\leq \mathrm{E}\|A_n\|_{op}^{2k}/\theta^{2k} n^{k+2k\epsilon} \\
&\leq \mathrm{E}(\mathrm{tr}\, A_n^* A_n)^k/\theta^{2k} n^{k+2k\epsilon} \\
&\leq C_k/\theta^{2k} n^{2k\epsilon-1}
\end{aligned}
\tag{117}
$$

If we choose k such that $k\epsilon > 1$ then the quantity on the right is summable. By Borel-Cantelli, almost surely only finitely many of the inequalities $\|A_n\|_{op}/n^{1/2+\epsilon} \geq \theta$. Since $\theta > 0$ is arbitrary, $\|A_n\|_{op}/n^{1/2+\epsilon} \to 0$ almost surely.

Let $T_{k,n} = \mathrm{tr}(A_n^* A_n)^k$. To prove the divergence statement, first let's bound $\mathrm{E}\, T_{k,n}^2 = \mathrm{E}(\mathrm{tr}(A_n^* A_n)^k)^2$. The terms of $(\mathrm{tr}(A_n^* A_n)^k)^2$ correspond to the union of two length $2k$ cycles in the bipartate multigraph. In terms with non-zero expectation, each edge appears an even number of times. If the two cycles overlap, then we can rearrange the order of the edges to get a length $4k$ cycle. This is a non-zero term in $\mathrm{E}\, T_{2k,n} = \mathrm{E}\,\mathrm{tr}(A_n^* A_n)^{2k} = C_{2k} n^{2k+1}$. If the cycles are disjoint, then it corresponds to an ordered pair of cycles, which is a term in $(\mathrm{E}\, T_{k,n})^2 = (\mathrm{E}\,\mathrm{tr}(A_n^* A_n)^k)^2 = C_k^2 n^{2k+2}$. Hence

$$
\mathrm{E}(\mathrm{tr}(A_n^* A_n)^k)^2 \leq (\mathrm{E}\,\mathrm{tr}(A_n^* A_n)^k)^2 + \mathrm{E}\,\mathrm{tr}(A_n^* A_n)^{2k} = C_{2k} n^{2k+1} + C_k^2 n^{2k+2} \tag{118}
$$

By the Paley-Zygmund inequality

$$
\begin{aligned}
\Pr(T_{k,n} \geq \theta_n \mathrm{E}\, T_{k,n}) &\geq (1-\theta_n)^2 \frac{(\mathrm{E}\, T_{k,n})^2}{\mathrm{E}\, T_{k,n}^2} \\
&\geq (1-\theta_n)^2 \frac{C_k^2 n^{2k+2}}{C_{2k} n^{2k+1} + C_k n^{2k+2}}
\end{aligned}
\tag{119}
$$

If $\theta_n \to 0$ then $\liminf_{n\to\infty} \Pr(\mathrm{tr}(A_n^* A_n)^k \geq \theta_n C_k n^{k+1}) = 1$. In particular, since the A_n are independent, by the second Borel-Cantelli lemma the inequality occurs infinitely often.

Choose $\alpha > 1$ and $\theta_n = \alpha^k K n^{-1-2\epsilon k}$ (for large enough n since we need $\theta_n < 1$ to apply Paley-Zygmund)

$$
\begin{aligned}
&\Pr(T_{k,n} \geq \alpha^k K n^{-1-2\epsilon k}\, \mathrm{E}\, T_{k,n} \text{ i.o.}) \leq \\
&\qquad \Pr(\|A_n\|/n^{1/2-\epsilon} > K \text{ i.o.}) + \Pr((\mathrm{tr}(A_n^* A_n)^k)^{1/k}/\|A_n\|^2 > \alpha \text{ i.o.})
\end{aligned}
\tag{120}
$$

We've shown the probability on the left is 1. Letting $k \to \infty$, since $\|A_n\|^{2k}/\,\mathrm{tr}(A_n^* A_n)^k \to 1$, the second term on the right tends to 0. (Is that right???). Hence $\|A_n\|/n^{1/2}$ is arbitrarily large infinitely often, and we conclude it diverges.

∎

3.15 The *Cramér random model for the primes* is a random subset $\mathcal{P}$ of the natural numbers with $1 \notin \mathcal{P}$, $2 \in \mathcal{P}$ and the events $n \in \mathcal{P}$ for $n = 3, 4, \ldots$ being jointly independent with $\Pr(n \in \mathcal{P}) = \frac{1}{\log n}$ (the restriction $n \geq 3$ is to ensure that $\frac{1}{\log n} \leq 1$). Its a simple yet convincing probabilistic model for the primes $\{2, 3, 5, 7, \ldots\}$ which can provide heuristic confirmations for conjectures in analytic number theory. Let $\pi(x) = |\{n \leq x : n \in \mathcal{P}\}|$

(i) (Probabilitic prime number theorem) Prove that

$$\frac{\pi(x)}{x/\log x} \to 1, \qquad \text{almost surely as } x \to \infty \tag{121}$$

(ii) (Probabilistic Riemann hypothesis) Show that if $\epsilon > 0$ then the quantity

$$\frac{\pi(x) - \int_2^x \frac{dt}{\log t}}{x^{1/2+\epsilon}} \to 0, \qquad \text{almost surely as } x \to \infty \tag{122}$$

(iii) (Probabilistic twin prime conjecture) Show that almost surely there are an infinite number of elements p of $\mathcal{P}$ such that $p+2$ also lies in $\mathcal{P}$

(iv) (Probabilistic Goldbach conjecture) Show that almost surely all but finitely many natural numbers n are expressible as the sum of two elements of $\mathcal{P}$

(i) First we show that for a constant $c > 0$

$$\left| \sum_{k=3}^{x} \frac{1}{\log k} - \int_3^x \frac{du}{\log u} \right| \leq c \tag{123}$$

This follows because if we define the series $b_n = 1/\log n - \int_n^{n+1} du/\log u$, each term is positive and the sum converges. The series b_n is summable because

$$\begin{aligned} b_n &= \frac{1}{\log n} - \int_n^{n+1} \frac{du}{\log u} \\ &\leq \frac{1}{\log n} - \frac{1}{\log(n+1)} = \frac{\log(1+n^{-1})}{\log n \log(n+1)} \\ &\leq \frac{1}{n(\log n)^2} \end{aligned} \tag{124}$$

Hence by the integral test $\int_a^\infty du/u(\log u)^2 = \int_{\exp a}^\infty v^{-2}\, dv < \infty$ where we made ths substitution $v = \log u$

Next we show that $\frac{x}{\log x} / \int_3^x \frac{du}{\log u} \to 1$. If we integrate by parts we get

$$\int \frac{dx}{\log x} = \frac{x}{\log x} - \int \frac{dx}{(\log x)^2} \tag{125}$$

By L'Hospital's rule $\int \frac{dx}{(\log x)^2} / \int \frac{dx}{\log x} = 1/\log x \to 0$.

Now we make the standard argument using Markov's inequality

$$\Pr\left(|\pi(x) - \sum_{k=3}^{x} 1/\log k| \geq \epsilon \frac{x}{\log x}\right) \leq \frac{\mathrm{E}(\pi(x) - \sum_{k\leq x} 1/\log k)^4}{\epsilon^4 (x/\log x)^4} \tag{126}$$

The numerator is of the form $\mathrm{E}(\sum_{k\leq x} Z_k)^4$ where $Z_k = E_k - \mathrm{E}\, E_k$ and $E_k = \{k \in \mathcal{P}\}$. Hence $\mathrm{E}\, Z_k = 0$ and the Z_k are independent. Therefore, expanding $\mathrm{E}(\sum_{k\leq x} Z_k)^4$ we are left with terms $\mathrm{E}\, Z_i^2 Z_j^2 = \mathrm{Var}\, X_i \,\mathrm{Var}\, X_j$ for $i \neq j$ and also $\mathrm{E}\, Z_k^4 \leq \mathrm{E}\, Z_k^2 = \mathrm{Var}\, Z_k$. (Note $|Z_k| \leq 1$ so $\mathrm{E}\, Z_k^q \leq \mathrm{E}\, Z_k^p$ for $p \leq q$). Thus we can bound

$$\begin{aligned} \mathrm{E}(\pi(x) - \sum_{k\leq x} 1/\log k)^4 &\leq \sum_{i\neq j} \mathrm{Var}\, Z_i \,\mathrm{Var}\, Z_j + \sum_k \mathrm{Var}\, Z_k \\ &\leq \left(\sum_{k\leq x} \mathrm{Var}\, Z_k\right)^2 + \sum_{k\leq x} \mathrm{Var}\, Z_k \\ &= O((x/\log x)^2) \end{aligned} \tag{127}$$

The last line follows since $\mathrm{Var}\, Z_k = \mathrm{Var}\, E_k = \frac{1}{\log k} - \frac{1}{(\log k)^2}$ and from the fact $\frac{x}{\log x} = \int \left(\frac{1}{\log x} - \frac{1}{(\log x)^2}\right) dx$. Therefore

$$\Pr\left(\left|\frac{\pi(x)}{x/\log x} - 1\right| \geq \epsilon\right) \leq \frac{(\log x)^2}{\epsilon^4 x^2} \tag{128}$$

The expression on the right is summable, so at most finitely many of the events on the left happen, which implies that $\pi(x)/(x/\log x) \to 1$ almost surely. (We elided a step replacing $\sum_{k\leq x} 1/\log k$ with $x/\log x$ but this is justified by the asyptotic equivalence proved at the begining).

(ii) Defining $Z_k = E_k - \mathrm{E}\, E_k$ like in (i), for $n \in \mathbb{N}$ let's analyze the expression

$$\mathrm{E}(\pi(x) - \sum_{k\leq x} 1/\log k)^n = \mathrm{E}(\sum_{k\leq x} Z_k)^n = \sum_{i_1,i_2,\ldots,i_n=1}^{x} \mathrm{E}\, Z_{i_1} Z_{i_2} \ldots Z_{i_n} \tag{129}$$

A partition of n is a sequence $n_1 \leq n_2 \leq n_l$ such that $n = n_1 + \cdots + n_l$ and $n_k \in \mathbb{N}^+$. Each in the sum above corresponds to a partition of n generated by grouping the repetitions in the i_k. Going the other way, given a partition of n with l terms, and a selection of l indices $\{i_1, \ldots, i_l\} \subset \{1, \ldots, x\}$ we get a unique term on the right. Thus we will group the terms on the right by their partition. Thus tuples $(i_1, \ldots, i_n) \in [x]^n$ are in one-to-one correspondence with a partition $(n_1, \ldots, n_l)$ and an ordered subset $\{i_1, \ldots, i_l\} \subset [x]$ (in other words, the set of i_k corresponding to the partition must be distinct).

If the partition of n contains 1 as a term, then the expectation of that term is 0. Thus we consider only partitions where each term is 2 or greater. (When we did this analysis for $n = 4$ that left us with terms $4 = 2 + 2$ and $4 = 4$, since other terms corresponding to the partitions $4 = 3+1$ and $4 = 1+1+1+1$ equal zero). From the bound $\mathrm{E}\, Z_i^k \leq \mathrm{E}\, Z_i^2$, we sum over all terms corresponding to a particular partition and get the following bound

$$\begin{aligned}
\sum_{\substack{3\leq i_1,\ldots,i_l\leq x\\ i_k \text{ distinct}}} X_{i_1}^{n_1}\ldots X_{i_l}^{n_l} &\leq \sum_{\substack{3\leq i_1,\ldots,i_l\leq x\\ i_k \text{ distinct}}} X_{i_1}^{2}\ldots X_{i_l}^{2} \\
&\leq \sum_{3\leq i_1,\ldots,i_l\leq x} X_{i_1}^{2}\ldots X_{i_l}^{2} \qquad (130)\\
&= \left(\sum_{3\leq k\leq x} X_k^2\right)^l = O((x/\log x)^l)
\end{aligned}$$

For fixed n, there are a finite number of partitions, so asymptotically $\mathrm{E}(\sum_{k\leq x} Z_k)^n$ equals $O((x/\log x)^l)$ where l is the length of the longest partition of n such that every term is at least 2. Clearly $l = \lfloor n/2 \rfloor$.

For any $\epsilon > 0$ choose n such that $2n\epsilon > 1$. Then using Markov's inequality, for large enough x,

$$\Pr\left(|\pi(x) - \sum_{k=3}^{x} 1/\log k| \geq \lambda x^{1/2+\epsilon}\right) \leq \frac{C(x/\log x)^n}{\lambda^n x^{n+2n\epsilon}} = \frac{C'}{(\log x)^n x^{2n\epsilon}} \qquad (131)$$

The expression on the right is summable so by Borel-Cantelli almost surely only finitely many of the events on the left occur. Since $\sum_{k=3}^{x} 1/\log k - \int_3^x du/\log u$ is bound by a constant and since $\lambda > 0$ we have for all $\epsilon > 0$

$$\frac{\pi(x) - \int_3^x du/\log u}{n^{1/2+\epsilon}} \to 0, \qquad \text{almost surely} \qquad (132)$$

(iii) Consider the event $T_p = \{p \in \mathcal{P} \text{ and } p+2 \in \mathcal{P}\}$. Clearly $\Pr(T_p) = 1/(\log p \log(p+2))$. Furthermore if $p' \neq p$ and $p' \neq p+2$ then T_p and $T_{p'}$ are independent. Thus the events $T_4, T_8, T_{12}, \ldots$ are jointly independent. Note that by comparison with $\sum_k 1/4k$

$$\sum_{k\geq 1} \Pr(T_{4k}) = \sum_{k\geq 1} \frac{1}{\log(4k)\log(4k+2)} = \infty \qquad (133)$$

By the second Borel-Cantelli almost surely infinitely many of the events T_{4k} occur.

(iv) Let $\mathcal{G} = \{n : n = p_1 + p_2 \text{ where } p_2, p_2 \in \mathcal{P}\}$. Then we can write

$$\begin{aligned}
\Pr(n \notin \mathcal{G}) &= \Pr\left(\bigcap_{3 \le k \le \lfloor n/2 \rfloor} \{k \notin \mathcal{P}\} \cup \{n-k \notin \mathcal{P}\}\right) \\
&\le \prod_{k=3}^{\lfloor n/2 \rfloor} \left(1 - \frac{1}{\log k \log(n-k)}\right) \\
&\le \left(1 - \frac{1}{\log(n/2)^2}\right)^{\lfloor n/2 \rfloor - 2} \\
&\le C \exp\left(-\frac{n/2}{\log(n/2)^2}\right) \\
&\le C \exp(-n^{1/2})
\end{aligned} \tag{134}$$

(The inequalities above hold for large enough n, not necessarily for every n). Now $\sum \exp(n^{1/2})$ converges (the integral test results an instance of the gamma function). Hence almost surely only finitely many $n \notin \mathcal{G}$ or, all large enough n are in $\mathcal{G}$

■

3.16 (Hardy-Ramanujan theorem) Let $x \ge 100$ be a natural number (so that in particular $\log\log x \ge 1$) and let n be a natural number drawn uniformly from $\{1, \ldots, x\}$. Assume Merten's theorem

$$\sum_{p \le x} \frac{1}{p} = \log\log x + O(1) \tag{135}$$

for all $x \ge 100$ where the sum is over primes up to x.

(i) Show that the random variable $\sum_{p \le x^{1/10}} 1_{p|n}$ (where $1_{p|n}$ is 1 when p divides n and 0 otherwise) has mean $\log\log x + O(1)$ and variance $O(\log\log x)$.

(ii) If $\omega(n)$ denotes the number of distinct prime factors of n show that $\omega(n)/\log\log n \to 1$ in probability as $x \to \infty$. More precisely show taht

$$\frac{\omega(n) - \log\log n}{g(n)\sqrt{\log\log n}} \to 0, \qquad \text{in probability} \tag{136}$$

whenever $g : \mathbb{N} \to \mathbb{R}$ is a function with $g(n) \to \infty$ as $n \to \infty$.

(i) For fixed x, the number of n that are divisible by $p < x^{1/10}$ is $\lfloor x/p \rfloor = x/p + O(1)$ and hence the probability is $p^{-1} + O(x^{-1})$. Let $\theta(n) = \sum_{p \le x^{1/10}} 1_{p|n}$. We conclude that

$$\begin{aligned} \mathrm{E}\,\theta(n) &= \mathrm{E}\sum_{p\le x^{1/10}} 1_{p|n} \\ &= \sum_{p\le x^{1/10}} (\frac{1}{p} + O(x^{1/10})) \\ &= \log\log(x^{1/10}) + O(1) + O(x^{-9/10}) \\ &= \log\log x + O(1) \end{aligned} \tag{137}$$

Let p and q be distinct primes. By unique factorization, $pq \mid n$ iff $p \mid n$ and $q \mid n$. Therefore $p \mid n$ and $q \mid n$ are (approximately) independent events since

$$(\mathrm{E}\,1_{p|n})(\mathrm{E}\,1_{q|n}) \approx (pq)^{-1} \approx \mathrm{E}\,1_{pq|n} = \mathrm{E}\,1_{p|n}1_{q|n} \tag{138}$$

Here the approximate equality $\approx$ indicates that equality holds plus an error of $O(x^{-1})$. In particular this means that $\mathrm{Cov}(1_{p|n}, 1_{q|n}) = O(x^{-1})$. Thus

$$\mathrm{Var}(\sum_{p\le x^{1/10}} 1_{p|n}) = \sum_{p\le x^{1/10}} \frac{1}{p} - \frac{1}{p^2} + O(x^{-8/10}) = O(\sum_{p\le x^{1/10}} \mathrm{E}\,1_{p|n}) = O(\log\log x) \tag{139}$$

(We're using the crude bound $|\sum p^{-1} - p^{-2}| \le 2\sum p^{-1}$).

(ii) Now $\omega(n) = \sum_{p\le n} 1_{p|n}$. Therefore

$$\Pr\left(\left|\frac{\omega(n) - \log\log n}{g(n)\log\log n}\right| > \epsilon\right) \le \frac{\mathrm{E}(\omega(n) - \log\log n)^2}{\epsilon^2 n^2 (\log\log n)^2} \tag{140}$$

Now without loss of generality we may assume $n > x^{1/10}$ since for any event E_x we have

$$\Pr(E_x) \le \Pr(E_x, n > x^{1/10}) + \Pr(n \le x^{1/10}) = \Pr(E_x, n > x^{1/10}) + x^{-9/10} \tag{141}$$

As $x \to \infty$ the second term tends to 0. If $n > x^{1/10}$ then

$$\log\log x - \log\log n \le \log\log x - \log\log x^{1/10} \le \log 10 = O(1) \tag{142}$$

Thus the difference is bounded by a constant independent of x. From the discussion above, $\omega(n) - \theta(n) = \sum_{x^{1/10}<p\le n} 1_{p|n}$ is independent of $\theta(n)$. Furthermore

$$\begin{aligned} \mathrm{E}(\omega(n) - \theta(n))^2 &\le 2\,\mathrm{E}(\omega(n) - \theta(n)) \\ &= 2(\log\log x - \log\log x^{1/10}) + O(1) = O(1) \end{aligned} \tag{143}$$

$$\begin{aligned} \mathrm{E}(\omega(n) - \log\log n)^2 &= \mathrm{E}(\omega(n) - \theta(n) + \theta(n) - \log\log n)^2 \\ &= \mathrm{E}(\theta(n) - \log\log n)^2 + \mathrm{E}(\theta(n) - \log\log n) + O(1) \\ &= O(\log\log x) \end{aligned} \tag{144}$$

Using this estimate in (140) we get

$$\Pr\left(\left|\frac{\omega(n) - \log\log n}{g(n)\log\log n}\right| > \epsilon\right) \leq O(g(n)^{-1}) \to 0 \tag{145}$$

This shows the expression (136) converges to 0 in probability

∎

3.17(Shannon entropy) Let A be a finite non-empty set of some cardinality $|A|$ and let X be a random variable taking values in A. Define the Shannon entropy

$$H(X) := -\sum_{x\in A}\Pr(X = x)\log\Pr(X = x) \tag{146}$$

(i) Show that $0 \leq H(X) \leq \log|A|$. When does the upper bound hold?

(ii) Let $\epsilon > 0$ and $n \in \mathbb{N}$. Let $X_1, \ldots, X_n$ be n iid copies of X, thus $\vec{X}_n := (X_1, \ldots, X_n)$ is a random variable taking values in A^n and the distribution μ_X is a probability measure on A^n. Let $\Omega_n \subset A^n$ denote the set

$$\Omega_n := \{\vec{x} \in A^n : \exp(-(1+\epsilon)nH(X)) \leq \mu_{\vec{X}_n}(\{\vec{x}\}) \leq \exp(-(1-\epsilon)nH(X))\} \tag{147}$$

Show that if n is sufficiently large then

$$\Pr(\vec{X}_n \in \Omega_n) \geq 1 - \epsilon \tag{148}$$

and

$$\exp((1-2\epsilon)nH(X)) \leq |\Omega_n| \leq \exp((1+2\epsilon)nH(X)) \tag{149}$$

Roughly speaking, $\vec{X}_n$ is in practice concentrated in a set of size about $\exp(nH(X))$ and is roughly uniformly distributed on that set.

(i) The function $\log x$ is concave downward so $\mathrm{E}\log f(x) \leq \log \mathrm{E} f(x)$. In the case of entropy, $f(x) = \Pr(X = x)^{-1}$, so

$$\begin{aligned} H(x) &= -\sum_x \Pr(X = x)\log(\Pr(X = x)) \\ &= \sum_x \Pr(X = x)\log(\Pr(X = x)^{-1}) \\ &\leq \log(\sum_x \Pr(X = x)\Pr(X = x)^{-1}) = \log(|A|) \end{aligned} \tag{150}$$

When $\Pr(X = x) = |A|^{-1}$ then $H(X) = \sum_x \Pr(X = x)\log(|A|) = \log(|A|)$ and the upper bound holds.

(ii) Define the random variable $Y_n = \log\Pr(X_n)$. Using the crude bound $\log x \leq x - 1 < x$ we find

$$\mathrm{E}\,|Y_n|^p \leq \mathrm{E}\Pr(X)^p = \sum_{x\in A}\Pr(X = x)^{p+1} \leq \left(\sum_{x\in A}\Pr(X = x)\right)^{p+1} = 1 \tag{151}$$

Now

$$\frac{\log \Pr(\vec{X}_n = x)}{n} = \frac{1}{n}\sum_{i=1}^{n} \log \Pr(X_i = x_i) = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{152}$$

Thus by the law of large numbers, $\frac{\log \Pr(\vec{X}_n = x)}{n} \to \mathrm{E}\, Y = -H(X)$ almost surely. From convergence in probability we have for any $\epsilon > 0$

$$\Pr\left(\left|\frac{\log \Pr(\vec{X}_n = x)}{n} + H(X)\right| < \epsilon H(X)\right) \to 1 \tag{153}$$

But the left hand side is precisely $\Pr(\vec{X}_n \in \Omega_n) > 1 - \epsilon$ for large enough n.

For any $\lambda < 1$, we have $\lambda \le \Pr(\vec{X}_n \in \Omega_n) \le 1$ for large enough n. Crudely bounding $\Pr(\vec{X}_n \in \Omega_n) = \mathrm{E}\, 1_{\Omega_n}$

$$|\Omega_n| p_{\min} \le \mathrm{E}\, 1_{\Omega_n} \le 1 \qquad \Rightarrow \qquad |\Omega_n| \le 1/p_{\min} = \exp((1+\epsilon)nH(X)) \tag{154}$$

and also

$$|\Omega_n| p_{\max} \ge \mathrm{E}\, 1_{\Omega_n} \ge \lambda \qquad \Rightarrow \qquad |\Omega_n| \ge \lambda/p_{\max} = \lambda \exp((1-\epsilon)nH(X)) \tag{155}$$

Let $\lambda \to 1$ to get the desired inequality. This is a little bit better than the problem statement.

■

3.18 Let $X_1, X_2, \ldots$ be iid copies of an unsigned random variable X with infinite mean and write $X_n = X_1 + \cdots + X_n$. Show that S_n/n diverges to infinity in probability.

For $\alpha > 0$ consider the sequence of iid random variables $X_n^{(\alpha)} = \min(X_n, \alpha)$. For all $p > 1$ we have $\mathrm{E}|X_n^{(\alpha)}|^p \le \alpha^p < \infty$, so by the fourth-moment law of large numbers $S_n^{(\alpha)}/n \to \mathrm{E}\, X^{(\alpha)}$ almost surely. Therefore we have the simple inequality

$$\liminf_n S_n/n \ge \liminf S_n^{(\alpha)}/n = \mathrm{E}\, X^{(\alpha)} \tag{156}$$

By monotone convergence the right hand side tends to infinity as $\alpha \to \infty$, which shows S_n/n diverges.

■

3.19 With the notation of the above analysis of the St Petersburg paradox, show that $\frac{S_n}{n \log_2 n}$ is almost surely unbounded. (Hint: show $X_n/n \log_2 n$ is unbounded and use second Borel Cantelli)

SUppose X has the St Petersburn distribution. We calculate

$$\Pr(X \geq m) = \sum_{k \geq \lceil \log_2 m \rceil} 2^{-k} = 2^{\lceil \log_2 m \rceil + 1} \tag{157}$$

Since $\lceil \log_2 m \rceil > \log_2 m - 1$ we have $\Pr(X \geq m) > 1/m$. Hence

$$\sum \Pr(X_n \geq Mn \log_2 n) > \sum_n \frac{1}{Mn \log_2 n} = \infty \tag{158}$$

The sum diverges by the integral test: $\int dx/x \log x = \log \log x$. Since the X_n are independent, by the second Borel-Cantelli lemma infinitely many of the events must occur. Therefore for any $M > 0$

$$\limsup_n \frac{S_n}{n \log_2 n} \geq \limsup \frac{X_n}{n \log_2 n} \geq M \tag{159}$$

and we conclude that the limsup is unbounded. ∎

3.20 A real random variable X is said to have a standard *Cauchy distribution* if it has the probability density function $x \mapsto \frac{1}{\pi} \frac{1}{1+x^2}$.

(i) Verify standard Cauchy distributions exist

(ii) Show a real random variable with the standard Cauchy distribution is not absolutely integrable

(iii) If $X_1, X_2, \ldots$ are iid copies of a random variable X with the standard Cauchy distribution, show that $\frac{|X_1| + \cdots + |X_n|}{n \log n}$ converges in probability to $\frac{2}{\pi}$ but is almost surely unbounded.

(i) The random variable exists by the standard Skorokhod construction, providing this is actually a density fucntion. But it is since

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \frac{1}{\pi}(\arctan(\infty) - \arctan(-\infty)) = 1 \tag{160}$$

(ii) Note that

$$\int_{-\infty}^{\infty} \frac{|x|\, dx}{\pi(1+x^2)} = 2 \int_0^{\infty} \frac{x\, dx}{\pi(1+x^2)} = \left. \frac{\log(1+x^2)}{\pi} \right|_0^{\infty} = \infty \tag{161}$$

(iii) First we calculate the bound

$$\Pr(X > m) = \int_m^{\infty} \frac{du}{\pi(1+x^2)} \leq \int_m^{\infty} \frac{du}{\pi x^2} = -\left. \frac{1}{\pi x} \right|_m^{\infty} = \frac{1}{\pi m} \tag{162}$$

As an aside, this shows Cauchy random variables are weak L^1. With an analogous calculation we can show for $m > 1$ that $\Pr(X > m) \geq 2/\pi m$. Since the pdf is an even function this gives the bound

$$\sum_n \Pr(|X_n| \geq Mn \log n) \geq \frac{4}{\pi M} \sum_n \frac{1}{n \log n} = \infty \tag{163}$$

Let $S_n = |X_1| + \cdots + |X_n|$. Therefore for any $M > 0$, by the second Borel-Cantelli

$$\limsup_n \frac{S_n}{n \log n} \geq \limsup \frac{X_n}{n \log n} \geq M \tag{164}$$

and the random variable.

Let $Y_{\leq n} = \min(|X|, n)$ and $Y_> = \max(|X|, n)$ where X is a Cauchy random variable. Then

$$\begin{aligned} \operatorname{E} Y_{\leq n} &= \int_0^n \frac{2x\,dx}{\pi(1+x^2)} = \frac{2 \log n}{\pi} \\ \operatorname{Var} Y_{>n} \leq \operatorname{E} Y_{>n}^2 &= \int_0^n \frac{2x^2\,dx}{\pi(1+x^2)} \leq \frac{2}{\pi} \int_0^n dx = \frac{2n}{\pi} \end{aligned} \tag{165}$$

Therefore

$$\Pr(|\sum_{i=1}^n |X_i| - m| \geq \lambda) \leq \Pr(|\sum_{i=1}^n Y_{i,\leq n} - m| \geq \lambda/2) + \Pr(\sum_{i=1}^n Y_{i,>n} \geq \lambda/2) \tag{166}$$

Using Markov's inequality and independence, if $m = n \operatorname{E} Y_{\leq n} = \frac{2}{\pi} n \log n$, then

$$\Pr(|\sum_{i=1}^n Y_{i,\leq n} - m| \geq \lambda/2) \leq \frac{8n^2}{\pi \lambda^2} \tag{167}$$

Using a crude union bound

$$\Pr(\sum_{i=1}^n Y_{i,>n} \geq \lambda/2) \leq n \Pr(Y_{>n} \geq \lambda/2) \leq \frac{2n}{\pi \max(\lambda/2, n)} \tag{168}$$

Hence choosing $\lambda = \epsilon n \log n$ we get for large enough n

$$\Pr\left(\left|\frac{|X_1| + \cdots + |X_n|}{n \log n} - \frac{2}{\pi}\right| \geq \epsilon\right) \leq \frac{8}{\pi \epsilon^2 (\log n)^2} + \frac{4}{\pi \epsilon \log n} \to 0 \tag{169}$$

■

3.21 (Weak law for triangular arrays) Let $(X_{i,n})_{i,n\in\mathbb{N}:1\le i\le n}$ be a triangular array of random variabes, with the variables $X_{1,n},\dots,X_{n,n}$ jointly independent for eacn n. Let M_n be a sequence going to infinity and write $X_{i,n,\le} := X_{i,n}1_{|X_{i,n}|\le M_n}$ and $\mu_n := \sum X_{i,n,\le}$. Assume that

$$\sum_{i=1}^{n} \Pr(|X_{i,n}| > M_n) \to 0 \tag{170}$$

and

$$\frac{1}{M_n^2}\sum_{i=1}^{n} \mathrm{E}\,|X_{i,n,\le}|^2 \to 0 \tag{171}$$

as $n\to\infty$. Show that

$$\frac{X_{1,n}+\cdots+X_{n,n}-\mu_n}{M_n} \to 0 \tag{172}$$

in probability

Let $L_n = \bigcap_{i=1}^n \{|X_{i,n}| \le M_n\}$ be the event that all the $X_{i,n}$ are "small" and let E_n be any sequence of events. If wish to show $\Pr(E_n)\to 0$, but it suffices to show that $\Pr(E_n\cap L_n)\to 0$. That's because by the union bound

$$\begin{aligned}\Pr(E_n) &\le \Pr(E_n\cap L_n) + \sum_{i=1}^{n}\Pr(E_n\cap\{|X_{i,n}|>M_n\})\\ &\le \Pr(E_n\cap L_n) + \sum_{i=1}^{n}\Pr(|X_{i,n}|>M_n)\end{aligned} \tag{173}$$

The second term on the right tends to 0 by assumption, so we only need to prove the first term on the right tends to 0.

Of course in order to show (172) we wish to consider the event $E_n = \{|S_n-\mu_n| > \epsilon M_n\}$. Let $E_{n,\le} = \{|S_{n,\le}-\mu_n| > \epsilon M_n\}$ where $S_{n,\le} = X_{1,n,\le}+\cdots+X_{n,n,\le}$. Note that if $|X_{i,n}|\le M_n$ for $i=1,\dots,n$ then $S_n = S_{n,\le}$. Hence $E_n\cap L_n = E_{n,\le}\cap L_n \subset E_{n,\le}$, and it suffices to show that $\Pr(E_{n,\le})\to 0$ as $n\to\infty$.

The coup de grace comes from Chebyshev's inequality. Let $\mu_{i,n} := \mathrm{E}\,X_{i,n,\le}$ and $Y_{i,n} = X_{i,n,\le}-\mu_{i,n}$. Then $S_{n,\le}-\mu_n = \sum_{i=1}^n Y_{i,n}$ and

$$\Pr\left(\left|\sum_{i=1}^{n} Y_{i,n}\right| > \epsilon M_n\right) \le \frac{\mathrm{E}(\sum_{i=1}^n Y_{i,n})^2}{\epsilon^2 M_n^2} \tag{174}$$

However, the $Y_{i,n}$ are independent and mean zero, and $\mathrm{E}\,X_{i,n}^2 = \mu_{i,n}^2 + \mathrm{E}\,Y_{i,n}^2$, so

$$\mathrm{E}\left(\sum_{i=1}^{n} Y_{i,n}\right)^2 = \sum_{i=1}^{n}\mathrm{E}\,Y_{i,n}^2 \le \sum_{i=1}^{n}\mathrm{E}\,X_{i,n}^2 \tag{175}$$

Hence

$$\Pr(|S_{n,\le}-\mu_n| > \epsilon M_n) \le \frac{1}{\epsilon^2 M_n^2}\sum_{i=1}^{n}\mathrm{E}\,X_{i,n}^2 \to 0 \tag{176}$$

We conclude that $\Pr(|S_n-\mu_n|/M_n > \epsilon)\to 0$ for all ϵ and therefore $(S_n-\mu_n)/M_n \to 0$ in probability. ∎

3.23 Let $X_1, X_2, \ldots$ be iid copies of a real random variable X.

(i) Show that for every real number t one has almost surely that

$$\frac{1}{n}\left|\{1 \le i \le n : X_i \le t\}\right| \to \Pr(X \le t) \tag{177}$$

and

$$\frac{1}{n}\left|\{1 \le i \le n : X_i < t\}\right| \to \Pr(X < t) \tag{178}$$

as $n \to \infty$

(ii) Establish the *Glivenko-Cantelli theorem*: almost surely one has

$$\frac{1}{n}\left|\{1 \le i \le n : X_i \le t\}\right| \to \Pr(X \le t) \tag{179}$$

uniformly in t as $n \to \infty$.

(i) Consider the random variables $Y_i = 1_{X_i \le t}$ and $Y_i' = 1_{X_i < t}$. The averages in the problem statement in part (i) are precisely $\frac{1}{n}\sum_i Y_i$ and $\frac{1}{n}\sum_i Y_i'$ which, by the strong law of large numbers, converge almost surely to $\mathrm{E}\, Y = \Pr(X \le t)$ and $\mathrm{E}\, Y' = \Pr(X < t)$ respectively.

(ii) We want to show that pointwise convergence implies uniform convergence, so this is a sort of analog of Dini's theorem. Let $\lfloor x \rfloor_m = \lfloor mx \rfloor / m$ be the greatest multiple of $\frac{1}{m}$ which is less than or equal to x. Let $F_X : t \mapsto \Pr(X \le t)$ be the cumulative distribution function and let $G_{n,X} : t \mapsto \frac{1}{n}\sum_{i=1}^n 1_{X_i \le t}$ be the empirical distribution function. All of these functions are monotonically nondecreasing in t and bounded between 0 and 1.

Note that for all $x \in \mathbb{R}$, $|x - \lfloor x \rfloor_m| \le \frac{1}{m}$. On the other hand if $|x - y| \le \frac{1}{m}$ then $|\lfloor x \rfloor_m - \lfloor y \rfloor_m| \le \frac{1}{m}$ since x and y are in the same bucket or adjacent buckets. Fixing m for the moment, for $k = 0, \ldots, m$ let $t_k = \inf_t \lfloor F_X(t) \rfloor_m = \frac{k}{m}$ be the transition point where $\lfloor F_X(t) \rfloor_m$ increses from $\frac{k-1}{m}$ to $\frac{k}{m}$. Some care is needed for $k = 0$ and $k = m$ since it may be that $F_X(t) > 0$ or $F_X(t) < 1$ for all $t \in \mathbb{R}$. We can define F_X and $G_{n,X}$ on the extended real line $[-\infty, \infty]$ by identifying $F_X(-\infty) = \lim_{t \to -\infty} F_X(t)$ and $F_X(\infty) = 1 = \lim_{t \to \infty} F_X(t)$ and similarly for $G_{n,X}$. With these conventions, for $t \in [t_k, t_{k+1})$, $\lfloor F_X(t) \rfloor_m = \frac{k}{m}$ and for no other points.

Since $G_{n,X}$ converges to F_X almost surely at each point $t \in \mathbb{R}$, it almost surely converges at any finite collection of points. Thus we may almost surely choose n large enough so that $|F_X(t_k) - G_{n,X}(t_k)| \le \frac{1}{m}$ for $k = 0, \ldots, m$. Therefore for n large enough

$$|\lfloor F_X(t_k) \rfloor_m - \lfloor G_{n,X}(t_k) \rfloor_m| \le \frac{1}{m} \qquad \text{for } k = 0, \ldots, m \tag{180}$$

Hence for $t \in [t_k, t_{k+1})$

$$\frac{k-1}{m} = \lfloor F_X(t_k) \rfloor_m - \frac{1}{m} \le \lfloor G_{n,X}(t_k) \rfloor_m \le \lfloor G_{n,X}(t) \rfloor_m \tag{181}$$

and

$$\lfloor G_{n,X}(t) \rfloor_m \le \lfloor G_{n,X}(t_{k+1}) \rfloor_m \le \lfloor F_X(t_{k+1}) \rfloor_m + \frac{1}{m} = \frac{k+2}{m} \tag{182}$$

Since $F_X(t) = \frac{k}{m}$ on the interval $[t_k, t_{k+1})$, we have a bound for this interval

$$|\lfloor F_X(t) \rfloor_m - \lfloor G_{n,X}(t) \rfloor_m| \le \frac{2}{m} \tag{183}$$

Since n was choosen to be close at all of the interval endpoints simultaneously, this shows means that (183) holds for all $t \in [-\infty, \infty]$. Hence $|F_X(t) - G_{n,X}(t)| \le \frac{3}{m}$ for all t. By choosing m large enough this inequality can be made arbitrarily small, showing that, almost surely, $G_{n,X}$ converges uniformly to F_X.

∎

3.24 (Lack of strong law for triangular arrays) Let X be a random variable taking values in the natural numbers with $\Pr(X = x) = \frac{1}{\zeta(3)} \frac{1}{n^3}$ where $\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3}$ (this is an example of the *zeta distribution*).

(i) Show that X is absolutely integrable

(ii) LEt $(X_{i,n})_{i,n \in \mathbb{N}: 1 \le i \le n}$ be independent copies of X. Show that the random variables $S_n = \frac{X_{1,n} + \cdots + X_{n,n}}{n}$ are almost surely unbounded.

(i) This follows from

$$\mathrm{E}|X| = \sum_{n=1}^{\infty} n \Pr(X = n) = \sum_{n=1}^{\infty} \frac{n}{\zeta(3) n^3} = \frac{\zeta(2)}{\zeta(3)} < \infty \tag{184}$$

(ii) Note that $\Pr(X \ge m) \approx \int_m^{\infty} \frac{dx}{\zeta(3) x^3} = O(m^{-2})$. Therefore, for large enough n

$$\Pr(S_n/n \ge M) \ge \sum_{i=1}^{n} \Pr(X_{i,n}/n \ge M) \ge \sum_{i=1}^{n} \frac{C}{M^2 n^2} = \frac{C}{M^2 n} \tag{185}$$

Therefore, since the S_n are independent, and since $\sum_{i=1}^{\infty} \Pr(S_n/n > M) = \infty$, the second Borel-Cantelli lemma implies that infinitely many of the events must occur. But this is the same as saying that $\limsup_{n\to\infty} S_n/n \ge M$ for any real $M > 0$, or S_n/n is unbounded.

∎

3.28 (Kronecker lemma) Let $\sum_{n=1}^{\infty} a_n$ be a convergent series of real numbers and let $0 < b_1 \le b_2 \le \ldots$ be in increasing sequence with $b_n \to \infty$ as $n \to \infty$. Show that $\frac{1}{b_n} \sum_{i=1}^{n} a_i b_i \to 0$ as $n \to \infty$

Let $S_n = \sum_{k=1}^n a_k$ and let $s = \lim_{n\to\infty} S_n$. Using summation-by-parts and defining $b_0 = 0$, we see

$$\sum_{i=k}^{n} a_k b_k = S_n b_n - \sum_{k=1}^{n-1} S_k(b_{k+1} - b_k) \tag{186}$$

Let m be such that $|S_n - s| < \epsilon$ for all $n > m$. Then

$$\left| s(b_n - b_m) - \sum_{k=m}^{n-1} S_k(b_{k+1} - b_k) \right| \le \epsilon \sum_{k=m}^{n-1} (b_{k+1} - b_k) = \epsilon(b_n - b_m) \tag{187}$$

Adding the terms related to m (which, for fixed m, correspond to constants) and dividing both sides by b_n we see for large enough n

$$\begin{aligned} \left| s - \frac{1}{b_n} \sum_{k=1}^{n-1} S_k(b_{k+1} - b_k) \right| &= \left| s - \frac{1}{b_n} \sum_{k=m}^{n-1} S_k(b_{k+1} - b_k) \right| \\ &\qquad + \left| \frac{1}{b_n} \sum_{k=1}^{m-1} S_k(b_{k+1} - b_k) \right| + \left| \frac{s b_m}{b_n} \right| \\ &< 3\epsilon \end{aligned} \tag{188}$$

Since ϵ is arbitrary, this shows $\frac{1}{b_n} \sum_{k=1}^{n-1} S_k(b_{k+1} - b_k) \to s$. Dividing both sides of (186) by b_n and taking limits as $b_n \to \infty$ shows

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=k}^{n} a_k b_k = \lim_{n\to\infty} S_n - s = 0 \tag{189}$$

∎

3.29 (Kolmogorov three series theorem, one direction) Let $X_1, X_2, \ldots$ be a sequence of jointly independent real random variables, and let $A > 0$. Suppose that $\sum_{n=1}^\infty \Pr(|X_n| > A)$ and $\sum_{n=1}^\infty \operatorname{Var}(X_n 1_{|X_n|<A})$ are absolutely convergent and that $\sum_{n=1}^\infty \mathrm{E}\, X_n 1_{|X_n| \le A}$ is convergent. Show that $\sum_{n=1}^\infty X_n$ is almost surely convergent.

Let $Y_n = X_n 1_{|X_n|<A} - \mathrm{E}\, X_n 1_{|X_n|<A}$. Then $\mathrm{E}\, Y_n = 0$ and $\operatorname{Var} Y_n = \mathrm{E}\, Y_n^2 = \operatorname{Var} X_n 1_{|X_n|<A}$. We can write

$$\sum_{n=1}^{N} X_n = \sum_{n=1}^{N} Y_n + \sum_{n=1}^{N} \mathrm{E}\, X_n 1_{|X_n|<A} + \sum_{n=1}^{N} X_n 1_{|X| \ge A} \tag{190}$$

By theorem 26, $\sum_{n=1}^\infty Y_n$ is almost surely convergent because each term has mean zero and $\sum_{n=1}^\infty \operatorname{Var} Y_n < \infty$. The second term is convergent by assumption. The third term almost surely has finitely many terms by Borel Cantelli, since $\Pr(|X_n| > A)$. Therefore it is also almost surely convergent. Therefore the event that $\sum_{n=1}^\infty X_n$ is convergent is an almost sure event, since the intersection of a finite number of almost sure events is almost sure (in this case, the three events that the terms above converge). ∎

3.30 (Cheap law of the iterated logarithm) Let $X_1, X_2, \ldots$ be a sequence of jointly independent real random variables of mean zero and bounded variance (so $\sup_n \mathrm{E}\, X_n^2 < \infty$). Write $S_n := X_1 + \cdots + X_n$. Show that $S_n / n^{1/2} (\log n)^{1/2+\epsilon}$ converges almost surely to 0 as $n \to \infty$ for any $\epsilon > 0$. (Hint use theorem 26 and the Kronecker lemma).

Let $B = \sup_n \mathrm{E}\, X_n^2$. Let $Y_n := X_n / n^{1/2} (\log n)^{1/2+\epsilon}$. Note that $\mathrm{E}\, Y_n = 0$ and

$$\sum_{n=1}^{\infty} \operatorname{Var} Y_n \leq B \sum_{n=1}^{\infty} \frac{1}{n(\log n)^{1+2\epsilon}} < \infty \tag{191}$$

The sum is convergent by the integral test since $\int \frac{dx}{x(\log x)^{1+2\epsilon}} = \frac{1}{(\log x)^{2\epsilon}}$. Therefore by theorem 26, the following sum is convergent almost surely

$$\sum_{n=1}^{\infty} Y_n = \sum_{n=1}^{\infty} \frac{X_n}{n^{1/2}(\log n)^{1/2+\epsilon}} < \infty \text{ a.s.} \tag{192}$$

Applying Kroneker's lemma with $b_n = n^{1/2}(\log n)^{1/2+\epsilon}$ we get

$$\lim_{n=0} \frac{X_1 + \cdots + X_n}{n^{1/2}(\log n)^{1/2+\epsilon}} \to 0 \text{ a.s.} \tag{193}$$

∎

3.31 Let $X_1, X_2, \ldots$ be iid copies of an absolutely integrable random variable X with mean μ. Show that the averages $\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$ converges in L^1 to μ. That is to say

$$\mathrm{E} \left| \frac{S_n}{n} - \mu \right| \to 0 \tag{194}$$

as $n \to \infty$

WLOG, replacing X_n with $X_n - \mu$, we can take $\mu = 0$. For $\epsilon > 0$, let $A_{\epsilon,n} = \{|S_n/n| > \epsilon\}$. By the weak law of large numbers, we can take n so large that $\Pr(A_{\epsilon,n}) < \epsilon$.

$$\begin{aligned} \mathrm{E}|S_n/n| &= \mathrm{E}_{A^c_{\epsilon,n}} |S_n/n| + \mathrm{E}_{A_{\epsilon,n}} |S_n/n| \\ &\leq \Pr(A^c_{\epsilon,n})\epsilon + \mathrm{E}|S_n/n| \Pr(A_{\epsilon,n}) \\ &\leq \epsilon + \epsilon\, \mathrm{E}|S_n/n| \to 0 \end{aligned} \tag{195}$$

∎

3.32 A scalar random variable X is said to be *weak* L^1 if one has

$$\sup_{t>0} t \Pr(|X| \geq t) < \infty \tag{196}$$

Markov's inequality implies that every absolutely integrable random variable is in weak L^1, but the converse is not true. (e.g. random variables with the Cauchy distribution are weak L^1 but not absolutely integrable). Show that if $X_1, X_2, \ldots$ are copies of an unsigned weak L^1 random variable then there exist quantities $a_n \to \infty$ such that S_n / a_n converges in probability to 1, where $S_n = X_1 + \ldots X_n$. (Thus: there is a weak law of large numbers for weak L^1 random variables and a strong law for strong L^1 random variables)

WLOG assume that X is not strong L^1, since in that case we have already proved a weak law of large numbers with $a_n = n$. Let $X_{\leq n} = \min(X, n)$ and $X_{>n} = \max(X, n)$. From the weak L^1 property we have $\Pr(X_{>n} > m) \leq C/\max(n, m)$. Let $\mu_n = \mathrm{E}\, X_{\leq n} \leq n$. Note that μ_n is an increasing sequence with $\lim_{n\uparrow\infty} \mu_n = \mathrm{E}\, X = \infty$ by monotonic convergence. Furthermore

$$\operatorname{Var} X_{\leq n} \leq \mathrm{E}\, X_{\leq n}^2 \leq n\, \mathrm{E}\, X_{\leq n} = n\mu_n \tag{197}$$

Let $a_n = n\mu_n$. Clearly $a_n \to \infty$ as $n \to \infty$. Using the same argument as the St Petersburg problem we calculate

$$\Pr(|X_{1,\leq n} + \dots X_{n,\leq n} - n\mu_n| > \epsilon a_n/2) \leq \frac{4 \operatorname{Var} X_{\leq n}}{\epsilon^2 n^2 \mu_n^2} \leq \frac{4}{\epsilon^2 n \mu_n} \to 0 \tag{198}$$

and for n large enough

$$\Pr(X_{1,>n} + \dots X_{n,\leq n} > \epsilon a_n/2) \leq \frac{nC}{\max(\epsilon a_n/2, n)} = \frac{2C}{\epsilon \mu_n} \to 0 \tag{199}$$

Now we can bound the probability of the event

$$\Pr(|S_n - a_n| \geq \epsilon a_n) \leq \Pr(|S_{n,\leq n} - a_n| > \epsilon a_n/2) + \Pr(S_{n,>n} > \epsilon a_n/2) \tag{200}$$

and we have shown the terms on the right tend to 0 in the limit $n \to 0$. Hence $\Pr(|S_n/n - 1| > \epsilon) \to 0$ for every ϵ, or $S_n/a_n \to 1$ in probability. ∎

Central Limit Theorem

> **4.5** For any natural number n let X_n be a discrete random variable drawn uniformly from $\{0/n, 1/n, \dots, (n-1)/n\}$ and let X be the continuous random variable drawn uniformly from $[0, 1]$. Then X_n converges in distribution to X. (A continuous random variable can emerge as the limit of discrete random variables)

If G is continuous and compactly supported then it is uniformly continuous. Choose δ such that $|G(x) - G(y)| < \epsilon$ whenever $|x - y| < \delta$. Then choose $N > \delta^{-1}$ so that each interval $[k/n, (k+1)/n]$ has length at most δ. Then for $n > N$ we have

$$\left| \int_{k/n}^{(k+1)/n} G(x)\, dx - \frac{1}{n} G(k/n) \right| \leq \int_{k/n}^{(k+1)/n} |G(x) - G(k/n)|\, dx = \frac{\epsilon}{n} \tag{201}$$

Summing these inequalities over k we get $|\mathrm{E}\, G(X) - \mathrm{E}\, G(X_n)| \leq \epsilon$ for all $n > N$. This proves convergence of the expectations. Since G is arbitrary, this proves vague convergence. ∎

4.7 (Portmanteau theorem) Show the following are equivalent

(i) $\mathrm{E}\, g(X_n) \to \mathrm{E}\, g(X)$ for every continuous compactly supported g

(ii) $\limsup_{n\to\infty} \Pr(X_n \in K) \le \Pr(X \in K)$ for all closed sets $K \subset \mathbb{R}$

(iii) $\liminf_{n\to\infty} \Pr(X_n \in U) \ge \Pr(X \in U)$ for all open sets $U \subset \mathbb{R}$

(iv) For any Borel set E whose topological boundary ∂E is such that $\Pr(X \in \partial E) = 0$ one has

$$\lim_{n\to\infty} \Pr(X_n \in E) = \Pr(X \in E) \tag{202}$$

- (i)⇒(ii) For closed K let $i_{\epsilon,K}$ be a function such that $i_{\epsilon,K}(x) = 1$ whenever $x \in K$ and $i_{\epsilon,K}(x) = 0$ whenever $d(x, K) \ge \epsilon$. (Here $d(x, K) = \inf_{y\in K} |x - y|$) This exists by Urysohn's lemma, or just take $i_{\epsilon,K}(x) = \max(1 - d(x, K)/\epsilon, 0)$. Then by (i)

$$\limsup_{n\to\infty} \Pr(X_n \in K) = \limsup_{n\to\infty} \mathrm{E}\, 1_K(X_n) \le \limsup_{n\to\infty} \mathrm{E}\, i_{\epsilon,K}(X_n) = \mathrm{E}\, i_{\epsilon,K}(X) \tag{203}$$

By bounded convergence we can take the limit $\epsilon \to 0$ to get

$$\limsup_{n\to\infty} \Pr(X_n \in K) \le \lim_{\epsilon\to 0} \mathrm{E}\, i_{\epsilon,K}(X) = \mathrm{E}\, 1_K(X) = \Pr(X \in K) \tag{204}$$

- (ii) ⇔ (iii) Taking complements, statement (ii) is the same as

$$\begin{aligned} 1 - \liminf_{n\to\infty} \Pr(X_n \in U) &= \limsup_{n\to\infty} 1 - \Pr(X_n \in U) \\ &= \limsup_{n\to\infty} \Pr(X_n \in U^c) \\ &\le \Pr(X \in U^c) \\ &= 1 - \Pr(X \in U) \end{aligned} \tag{205}$$

Rearranging, this is the same as $\liminf_{n\to\infty} \Pr(X_n \in U) \ge \Pr(X \in U)$. Going from (iii) to (ii) is the same argument in reverse.

- (ii) and (iii) ⇒ (i) By the linearity of expectation and the fact $g(x) = \max(g(x), 0) - |\min(g(x), 0))|$, its sufficient to prove (i) for nonnegative g. We can approximate g by $g_n^l(x) = 2^{-n}\lfloor 2^n g(x)\rfloor$ and $g_n^u(x) = 2^{-n}\lceil 2^n g(x)\rceil$. Each of these uniformly approximates g so, for example, $|g - g_n^l| \le 2^{-n}$. Furthermore we can write

$$g_n^l(x) = \sum_{k=1}^{\infty} 2^{-n} 1_{g(x)\ge k2^{-n}}(x) \qquad \text{and} \qquad g_n^u(x) = \sum_{k=0}^{\infty} 2^{-n} 1_{g(x)>k2^{-n}}(x) \tag{206}$$

Since g is bounded, only finitely many of the terms in each of the above sums is non-zero. Therefore

$$\begin{aligned}
\limsup_{n\to\infty} \mathrm{E}\, g(X_n) &\leq \limsup_{n\to\infty} \mathrm{E}(2^{-m} + g_m^l(X_n)) \\
&= 2^{-m} + \sum_{k=0}^{\infty} 2^{-m} \limsup_{n\to\infty} \Pr(X_n \in g^{-1}([k2^{-m}, \infty))) \\
&\leq 2^{-m} + \sum_{k=0}^{\infty} 2^{-m} \Pr(X \in g^{-1}([k2^{-m}, \infty))) \\
&= 2^{-m} + \mathrm{E}\, g_m^l(X) \\
&\leq 2^{-m} + \mathrm{E}\, g(X)
\end{aligned} \tag{207}$$

Let $m \to \infty$ to get $\limsup_{n\to\infty} \mathrm{E}\, g(X_n) \leq \mathrm{E}\, g(X)$. By a similar argument using g_n^u we find $\liminf_{n\to\infty} \mathrm{E}\, g(X_n) \geq \mathrm{E}\, g(X)$. Hence the limit exists and equals $\mathrm{E}\, g(X)$

- (ii) and (iii) $\Rightarrow$ (iv) First note that by assumption $\Pr(X \in \operatorname{int} E) = \Pr(X \in \operatorname{cl} E) = \Pr(X \in E)$. By (ii)

  $$\liminf \Pr(X_n \in E) \geq \liminf \Pr(X_n \in \operatorname{int} E) \geq \Pr(X \in \operatorname{int} E) = \Pr(X \in E) \tag{208}$$

  And by (iii)

  $$\limsup \Pr(X_n \in E) \leq \limsup \Pr(X_n \in \operatorname{cl} E) \leq \Pr(X \in \operatorname{cl} E) = \Pr(X \in E) \tag{209}$$

  This shows that $\limsup \Pr(X_n \in E) = \liminf \Pr(X_n \in E)$ so the limit exists and equals $\Pr(X \in E)$

- (iv) $\Rightarrow$ (ii) Let $K_\epsilon = \{x : d(x, K) \leq \epsilon\}$. The sets $\partial K_\epsilon = \{x : d(x, K) = \epsilon\}$ are disjoint for different values of ϵ. Therefore at most countably many have $\Pr(\partial K_\epsilon) > 0$ (for example, at most n have $\Pr(\partial K_\epsilon) \geq 1/n$). Hence we may find a sequence $\epsilon_n \downarrow 0$ such that $\Pr(\partial K_{\epsilon_n}) = 0$. Then we have

  $$\limsup_{n\to\infty} \Pr(X_n \in K) \leq \limsup_{n\to\infty} \Pr(X_n \in K_{\epsilon_m}) = \Pr(X \in K_{\epsilon_m}) \tag{210}$$

  Since the K_{ϵ_m} are nested and $K = \cap_{m=1}^{\infty} K_{\epsilon_m}$, taking the limit $m \to \infty$ we have

  $$\limsup_{n\to\infty} \Pr(X_n \in K) \leq \Pr(X \in K) \tag{211}$$

■

4.9 (DeMoivre-Laplace theorem) Let X be a Bernoulli random variable taking values in $\{0,1\}$ with $\Pr(X=0)=\Pr(X=1)=1/2$. Thus X has mean $1/2$ and variance $1/4$. Let $X_1, X_2, \ldots$ be iid copies of X and write $S_n := X_1 + \cdots + X_n$.

i Show that S_n takes values in $\{0,\ldots,n\}$ with $\Pr(S_n = i) = 2^{-n}\binom{n}{i}$ (this is an example of a *binomial distribution*)

ii Assume Stirling's formula

$$n! = (1+o(1))\sqrt{2\pi n}\, n^n e^{-n} \tag{212}$$

where $o(1)$ is a function of n that goes to zero as $n \to \infty$. Without using the central limit theorem show that

$$\Pr(a \le 2\sqrt{n}(\frac{S_n}{n} - \frac{1}{2}) \le b) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx \tag{213}$$

(i) We have $S_n = i$ iff exactly i of the n variables X_k are equal to 1 and $n-i$ are equal to 0. Any particular vector of $(X_1,\ldots,X_n) = (x_1,\ldots,x_n)$ has probability 2^{-n}. Of all the vectors, $\binom{n}{i}$ have exactly i entries equal to 1, so $\Pr(S_n = i) = 2^{-n}\binom{n}{i}$

(ii) Using Stirling's approximation, with $n = k + k'$

$$\begin{aligned} 2^{-n}\binom{n}{k} &= 2^{-n}\frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi k}\, k^k e^{-k}\sqrt{2\pi k'}\, k'^{k'} e^{-k'}} \\ &= \frac{1}{\sqrt{2\pi}}\exp(-n\log 2 + (\log n - \log k - \log k')/2)\times \\ &\qquad \exp(-n\log n - k\log k - k'\log k') \end{aligned} \tag{214}$$

Now when $k = \frac{1}{2}(n + a\sqrt{n})$ then $k' = \frac{1}{2}(n - a\sqrt{n})$ and

$$\begin{aligned} \log k &= \log\frac{n}{2} + \frac{a}{\sqrt{n}} - \frac{a^2}{2n} + O(\frac{1}{n^{3/2}}) \\ \log k' &= \log\frac{n}{2} - \frac{a}{\sqrt{n}} - \frac{a^2}{2n} + O(\frac{1}{n^{3/2}}) \end{aligned} \tag{215}$$

Hence our expressions have enormous amounts of cancelation, and after a bit of algebra we find

$$\frac{1}{2}(\log n - \log k - \log k') = \log\frac{2}{\sqrt{n}} + O(\frac{1}{n}) \tag{216}$$

and

$$n\log n - k\log k - k'\log k' = n\log 2 - 2a^2 + O(\frac{1}{\sqrt{n}}) \tag{217}$$

By considering that $1 = \Delta k = \frac{\sqrt{n}}{2}\Delta a$, we can consider the factor $\frac{2}{\sqrt{n}}$ to be Δa, pulling this together equation (214) becomes

$$2^{-n}\binom{n}{k} = (1+o(1))\frac{1}{\sqrt{2\pi}}e^{-a^2/2}\Delta a \tag{218}$$

So summing over $k_a = \lfloor \frac{1}{2}(n + a\sqrt{n}) \rfloor$ to $k_b = \lfloor \frac{1}{2}(n + b\sqrt{n}) \rfloor$, we get a Riemann sum which converges to the desired integral

$$\sum_{k=k_a}^{k_b} 2^{-n} \binom{n}{k} = o(1) + \sum_{k=k_a}^{k_b} \frac{1}{\sqrt{2\pi}} e^{2n(k/n-1/2)^2} \frac{2}{\sqrt{n}} \to \int_a^b \frac{\exp(u^2/2)\, du}{\sqrt{2\pi}} \tag{219}$$

■

4.10 Let $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ be sequences of real random variables and let X, Y be further real random variables.

(i) If X is deterministic, show that X_n converges in distribution to X if and only if X_n converges in probability to X

(ii) Suppose that X_n is independent of Y_n for each n and X is independent of Y. Show that $X_n + iY_n$ converges in distribution to $X + iY$ if and only if X_n converges in distribution to X and Y_n converges in distribution to Y. What happens if independence is dropped? (hint: prop 4 or Stone-Weierstrass)

(iii) If X_n converges in distribtion to X show that for every $\epsilon > 0$ there exists $K > 0$ such that $\Pr(|X_n| > K) < \epsilon$ for all sufficiently large n. (That is to say, X_n is a *tight sequnces* of random variables.)

(iv) Show that X_n converges in distribution to X if and only if after extending the probability space model if necessary, one can find copies $Z_1, Z_2, \ldots$ and Z of $X_1, X_2, \ldots$ and X respectively such that Z_n converges almost surely to Z. (Hint: Skorokhod representation)

(v) If $X_1, X_2, \ldots$ converges in distribution to X and $F : \mathbb{R} \to \mathbb{R}$ is continuous, show that $F(X_1), F(X_2), \ldots$ converges in distribution to $F(X)$. Generalize this claim to the case when X takes values in an arbitrary locally compact Hausdorff space.

(vi) (Slutsky's theorem) If X_n converges in distribution to X and Y_n converges in probability to a deterministic limit Y, show that $X_n + Y_n$ converges in distribution to $X + Y$ and $X_n Y_n$ converges in distribution to XY. (Use (iv) or (iii) to control some error terms). This statement combines well with (i). What happens if Y is not assumed to be deterministic?

(vii) (Fatou lemma) If $G : \mathbb{R} \to [0, \infty)$ is continuous and X_n converges in distribution to X show that

$$\liminf_{n\to\infty} \mathrm{E}\, G(X_n) \geq \mathrm{E}\, G(X) \tag{220}$$

(viii) (Bounded convergence) If $G : \mathbb{R} \to \mathbb{R}$ is continuous and bounded and X_n converges in distribution to X show that $\lim_{n\to\infty} \mathrm{E}\, G(X_n) = \mathrm{E}\, G(X)$.

(ix) (Dominated convergence) If X_n converges in distribution to X and there is an absolutely integrable Y such that $|X_n| \leq Y$ almost surely for all n show that $\lim_{n\to\infty} \mathrm{E}\, X_n = \mathrm{E}\, X$

(i) Let $X = x$ a constant. The distribution function F_X has one discontinity at x and other wise is constant and equal to 0 for $t < x$ and constant and equal to 1 for $t > x$. Hence, for any $\epsilon > 0$ the points $x + \epsilon$ and $x - \epsilon$ are points of continuity of the distribution F_X, and by proposition 4,

$$\Pr(X_n \leq x - \epsilon) = F_{X_n}(x - \epsilon) \to F_X(x - \epsilon) = 0 \tag{221}$$

Similarly

$$\Pr(X_n \leq x + \epsilon) = F_{X_n}(x + \epsilon) \to F_X(x + \epsilon) = 1 \tag{222}$$

Hence

$$\lim_{n\to\infty} \Pr(|X_n - x| > \epsilon) \leq \lim_{n\to\infty} \Pr(X_n \leq x - \epsilon) + \lim_{n\to\infty} \Pr(X_n > x + \epsilon) = 0 \tag{223}$$

This shows that $X_n \to X$ in probability.

(ii) TODO

(iii) Note that if $\Pr(|X| > K) < \epsilon$, then for any $K' > K$ its also the case that $\Pr(|X| > K') < \epsilon$. Take any sequence $\epsilon_n > 0$ such that $\epsilon_n \downarrow 0$. Since X has at most countably many points of discontinuity, for each ϵ_n we may find a K_{ϵ_n} with $\Pr(|X| > K_{\epsilon_n}) < \epsilon_n$ such that $\pm K_{\epsilon_n}$ is not a point of discontinuity for F_X. For any $\epsilon > 0$ we can find $\epsilon_n < \epsilon$, so we may associate $K_\epsilon = K_{\epsilon_n}$ which has the property $\pm K_\epsilon$ are not points of discontinuity.

Thus for $\epsilon > 0$, by propsition 4, $\Pr(X_n \leq -K_\epsilon) \to \Pr(X \leq -K_\epsilon)$ and $\Pr(X \geq K_\epsilon) \to \Pr(X \geq K_\epsilon)$. Hence for n large enough, the left and right tail probabilities for X_n are within ϵ of the corresponding probabilities for X.

$$\Pr(|X_n| > K_\epsilon) \leq \Pr(|X| > K_\epsilon) + 2\epsilon = 3\epsilon \tag{224}$$

This shows that X_n is tight.

(iv) Let $\omega \sim U(0,1)$ have a uniform distribution on $[0,1]$. Then by Skorokhod's construction define $Z_n = \sup\{z \in \mathbb{R} : F_{X_n}(z) < \omega\}$ and $Z = \sup\{z \in \mathbb{R} : F_X(z) < \omega\}$ have distributions the same distributions as the corresponding X variables, so $Z_n \sim X_n$ and $Z \sim X$. Suppose $\omega = F_X(z)$ and also that z is a point of continuity for F_X. Then $Z = z$ and also, by proposition 4, $Z_n \to z$. TODO finish this

(v) We'll just do the generic case since the argument is identical. Let $F : R \to S$ where S is a locally compact Hausdorff space. Let $G : S \to \mathbb{R}$ be a continuous function with compact support. Note that $G \circ F : R \to \mathbb{R}$ is continuous with compact support, and hence $\mathrm{E}\, G(F(X_n)) \to \mathrm{E}\, G(F(X))$. Since G is arbitrary, this shows that $F(X_n) \to F(X)$ in distribution.

(vi) By (ii) we can find a probability space where $Z_n \sim X_n$, and $Z \sim Z$ and $Z_n \to Z$ in probability. By (i) we know that $Y_n \to Y$ in probability. By 3.2(vii) this means that for any continuous function $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ we have $F(Z_n, Y_n) \to F(Z, Y)$ in probability. Hence $F(Z_n, Y_n) \to F(Z, Y)$ converges in distribution. Let $F(x, y) = x + y$ or $F(x, y) = xy$ to prove the theorem.

I guess I'm surprised that one of these problems isn't to show that convergence in probability implies convergence in distribution. Here's a short proof. Suppose $X_n \to X$ in probability. For continuous g we have $g(X_n) \to g(X)$ in probability. Any compactly supported g is bounded so by 3.2(ix) we have $\mathrm{E}\, g(X_n) \to \mathrm{E}\, g(X)$. This shows $X_n \to X$ in distribution.

(vii) Take $H_m : [0, \infty) \to [0, \infty)$ to be a sequence of continuous compactly supported functions which satisfy $H_m(x) \leq x$ everywhere and $H_n(x) = x$ on $[0, m]$. We've constructed the H_m to truncate the extreme values of F but to converge pointwise to the identity. Then $H_n \circ G$ is continuous and compactly supported and hence

$$\liminf_{n\to\infty} \mathrm{E}\, G(X_n) \geq \liminf_{n\to\infty} \mathrm{E}\, H_m(G(X_n)) = \mathrm{E}\, H_m(G(X)) \tag{225}$$

By Fatou's lemma $\liminf_{m\to\infty} \mathrm{E}\, H_m(G(X)) \geq \mathrm{E}\, G(X)$. Combining these statements we get the desired result.

(viii) By the linearity of expectation, and by considering $G = G_+ - G_-$ where $G_+ = \max(G, 0)$ and $G_- = -\min(G, 0)$, it sufficies to prove the statement when G is non-negative. Let $G(x) \leq M$. Then applying (vii) to $M - G$

$$\liminf_n \mathrm{E}(M - G(X_n)) \geq \mathrm{E}(M - G(X)) \qquad \Rightarrow \qquad \limsup G(X_n) \leq \mathrm{E}\, G(X) \tag{226}$$

Combining this with another application of (vii) we get

$$\mathrm{E}\, G(X) \geq \limsup_n \mathrm{E}\, G(X_n) \geq \liminf_n \mathrm{E}\, G(X_n) \geq G(X) \tag{227}$$

Hence all the expressions are equal, which shows the limit exists and has the value $\mathrm{E}\, G(X)$

(ix) Note that $Z_n^- = Y - X_n$ converges in distribution to $Y - X$ and $Z_n^+ = Y + X_n$ converges in distribution to $Y + X$. Since $|X_n| \leq Y$ both Z_n^+ and Z_n^- are non-negative, so by Fatou's lemma

$$\liminf_n \mathrm{E}(Y - X_n) \geq \mathrm{E}(Y - X) \qquad \Rightarrow \qquad \liminf_n \mathrm{E}(Y + X_n) \geq \mathrm{E}(Y + X) \tag{228}$$

Using linearity and the fact $\mathrm{E}\, Y$ is a constant, this shows $\limsup_n \mathrm{E}\, X_n \leq \mathrm{E}\, X \leq \liminf_n \mathrm{E}\, X_n$. Since the opposite inequality is trivial, the lim sup and lim inf are equal and equal to $\mathrm{E}\, X$.

∎

4.13 (Probabilistic interpretation of convolution) Let $f, g : \mathbb{R} \to [0, +\infty]$ be measurable functions with $\int_{\mathbb{R}} f(x)\, dx = \int_{\mathbb{R}} g(x)\, dx = 1$. Define the convolution $f * g$ of f and g to be

$$f * g := \int_{\mathbb{R}} f(y) g(x - y)\, dy \tag{229}$$

Show that if X, Y are independent real random variables with probability density functions f, g respectively, then $X + Y$ has probability density function $f * g$.

Let $S = X + Y$ and consider $\mathrm{E}\, h(S)$ for any continuous compactly supported function h. This is given by

$$\mathrm{E}\, h(S) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(x+y) f(x) g(y)\, dx\, dy \tag{230}$$

Making a change of variables $s = x + y$ and $t = y$ so that $\frac{\partial(s,t)}{\partial(x,y)} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$ we get

$$\mathrm{E}\, h(S) = \int_{\mathbb{R}} \int_{\mathbb{R}} h(s) f(s-t) g(t)\, ds\, dt = \int_{\mathbb{R}} h(s) f * g(s)\, ds \tag{231}$$

Since h is arbitrary, this shows that $f * g$ is the pdf for S. ∎

4.14 (Lindeberg central limit theorem) Let k_n be a sequence of natural numbers going to infinity in n. For each natural number n let $X_{1,n}, \ldots, X_{k_n,n}$ be jointly independent real random variables of mean zero and finite variance. (We do not require the random variables $(X_{1,n}, \ldots, X_{k_n,n})$ to be jointly independent in n or even to be modeled by a common probability space). Let σ_n be defined by

$$\sigma_n^2 := \sum_{i=1}^{k_n} \mathrm{Var}(X_{i,n}) \tag{232}$$

and assume that $\sigma_n > 0$ for all n.

(i) If one assumes the *Lindeberg condition* that

$$\frac{1}{\sigma_n^2} \sum_{i=1}^{k_n} \mathrm{E}(|X_{i.n}^2 1_{|X_{i,n}| > \epsilon\sigma_n}|) \to 0 \tag{233}$$

as $n \to \infty$ for any $\epsilon > 0$ then show that the random variables $\frac{X_{1,n} + \cdots + X_{k_n,n}}{\sigma_n}$ converge in distribution to a random variable with the normal distribution $N(0,1)$.

(ii) Show that the Lindeberg condition implies the *Feller condition*

$$\frac{1}{\sigma_n^2} \max_{1 \le i \le k_n} \mathrm{E}|X_{i,n}|^2 \to 0 \tag{234}$$

as $n \to \infty$

Let $\sigma_{i,n}^2 = \mathrm{E}\, X_{i,n}^2$ and $\sigma_n^2 = \sum_{i=1}^{k_n} \sigma_{i,n}^2$. Without loss of generality, scaling each $X_{i,n}$ by σ_n, we may assume that $\sigma_n = 1$ for all n. By Stone-Weirestrass, we can approximate arbitrary continuous compactly supported G pointwise with a smooth compactly supported G. Thus we may assume all derivatives of G exist and are bounded.

We're going to use the truncation method, so let $\mu_{i,n} := \mathrm{E}\, X_{i,n} 1_{|X_{i,n}| \le \epsilon}$ and define

$$X_{i,n}^{\le} := X_{i,n} 1_{|X_{i,n}| \le \epsilon} - \mu_{i,n} \qquad X_{i,n}^{>} := X_{i,n} 1_{|X_{i,n}| > \epsilon} + \mu_{i,n} \tag{235}$$

Note that $X_{i,n} = X^{\leq}_{i,n} + X^{>}_{i,n}$. First let's bound the error caused by trunction. Define

$$T_{i,n} := X_{1,n} + \cdots + X_{i,n} + X^{\leq}_{i+1,n} + \ldots X^{\leq}_{k_n,n} \tag{236}$$

Then since $T_{i+1} = T_i + X^{>}_{i,n}$, taking the Taylor expansion with remainder to the second order

$$\begin{aligned} |\operatorname{E} G(T_{i+1}) - \operatorname{E} G(T_i)| &= \left| \operatorname{E} G'(S_i) \operatorname{E} X^{>}_{i,n} + \frac{1}{2} G''(\xi(S_i, X^{>}_{i,n})) \operatorname{E}(X^{>}_{i,n})^2 \right| \\ &\leq \sup |G''| \operatorname{E}(X^{>}_{i,n})^2 \end{aligned} \tag{237}$$

Considering the telescoping sum

$$|\operatorname{E} G(X_{1,n} + \ldots X_{k_n,n}) - \operatorname{E} G(X^{\leq}_{1,n} + \cdots + X^{\leq}_{1,n})| \leq \sum_{i=0}^{k_n-1} |\operatorname{E} G(T_{i+1}) - \operatorname{E} G(T_i)| \tag{238}$$

By the above, this is bound by $O(\sum_{i=1}^{k_n} \operatorname{E}(X^{>}_{i,n})^2)$ (where the implied constant depends on the supremum of G''). Now $\operatorname{E} {X^{>}_{i,n}}^2 \leq \operatorname{E} X^2_{i,n} 1_{|X_{i,n}|>\epsilon}$ since the variance of a random variable is bound by the second moment. However, $\sum_i \operatorname{E} X^2_{i,n} 1_{|X_{i,n}|>\epsilon} \to 0$ by the Lindeberg condition, which shows the error introduced by truncation tends to zero.

Let $N_{i,n}$ for each $i = 1, \ldots, k_n$ be a sequence of mutually independent random variables with distribution $N(0,1)$ which are also mutually independent of all the $X_{i,n}$. Working now with the truncated series, define

$$Z_{i,n} := X^{\leq}_{1,n} + \cdots + X^{\leq}_{i-1,n} + \sigma_{i+1,n} N_{i+1,n} + \cdots + \sigma_{k_n,n} N_{k_n,\sigma} \tag{239}$$

Suppressing the indices i and n for a moment, taking the Taylor expasion to the third order we get a bound

$$\begin{aligned} |\operatorname{E} G(Z + X^{\leq}) - \operatorname{E} G(Z + \sigma N)| =& \big| \operatorname{E} G'(Z)(\operatorname{E} X^{\leq} - \operatorname{E} \sigma N) \\ &+ \frac{1}{2} \operatorname{E} G''(Z)(\operatorname{E} {X^{\leq}}^2 - \sigma^2 \operatorname{E} N^2) \\ &+ \frac{1}{6}(\operatorname{E} G'''(\xi_1) {X^{\leq}}^3 - \operatorname{E} G'''(\xi_2) N^3) \big| \\ \leq& C |\operatorname{E} {X^{\leq}}^2 - \sigma^2| + C' \operatorname{E} |X^{\leq}|^3 + C'' \sigma^3 \end{aligned} \tag{240}$$

The constants here depend only on G and not on the X's or N's. Using the same type of telescoping sum as above, considering that $Z_{0,n} = \sum_i \sigma_{i,n} N_{i,n}$ and $Z_{k_n,n} = \sum_i X^{\leq}_{i,n}$ we get a bound

$$\begin{aligned} &\left| \operatorname{E} G\left(\sum_i X^{\leq}_{i,n} \right) - \operatorname{E} G\left(\sum_i \sigma_{i,n} N_{i,n} \right) \right| \leq \\ &\qquad C \sum_i \left| \operatorname{E} {X^{\leq}_{i,n}}^2 - \sigma^2_{i,n} \right| + C' \sum_i \operatorname{E} |X^{\leq}_{i,n}|^3 + C'' \sum_i \sigma^3_{i,n} \end{aligned} \tag{241}$$

We'll show that each of these three terms tends to 0 or can be made arbitrarily small. First, again supressing subscripts, note that

$$\mathrm{E}\, X^{\leq 2} = \mathrm{E}\, X^2 1_{|X|\leq\epsilon} - \mu^2 \qquad\qquad \mathrm{E}\, X^{>2} = \mathrm{E}\, X^2 1_{|X|>\epsilon} - \mu^2 \tag{242}$$

Hence,

$$\mathrm{E}\, X^{\leq 2} + \mathrm{E}\, X^{>2} = \sigma^2 - 2\mu^2 \qquad \Rightarrow \qquad |\sigma^2 - \mathrm{E}\, X^{\leq 2}| \leq 2\mu^2 + \mathrm{E}\, X^{>2} \tag{243}$$

By Cauchy-Schwarz $\mu^2 \leq \mathrm{E}\,|X 1_{|X|>\epsilon}| \leq \mathrm{E}\,\mathrm{E}\, X^2 1_{|X|>\epsilon}$. So we have the bound

$$\sum_i \left| \sigma_{i,n}^2 - \mathrm{E}\, X^{\leq 2} \right| \leq 3 \sum_i \mathrm{E}\, X_{i,n}^2 1_{|X_{i.n}|>\epsilon} \tag{244}$$

The right hand side tends to 0 by assumption– this is the same as the Lindeberg condition.

Next, note that since $|\mu| = |\,\mathrm{E}\, X 1_{|X|\leq\epsilon}| \leq \epsilon$ we also have

$$|X^{\leq}| \leq |X 1_{|X|\leq\epsilon}| + |\mu| \leq 2\epsilon \tag{245}$$

and hence

$$\sum_i \mathrm{E}\,|X_{i,n}^{\leq}|^3 \leq 2\epsilon \sum_i \mathrm{E}\,|X_{i,n}^{\leq}|^2 \leq 2\epsilon \sum_i \sigma_{i,n}^2 = 2\epsilon \tag{246}$$

This can be made arbitrarily small by choosing the truncation parameter.

Finally

$$\sum_i \sigma_{i,n}^3 \leq (\max_{i=1,\ldots,k_n} \sigma_{i,k}) \sum_i \sigma_{i,n}^2 = \max_{i=1,\ldots,k_n} \sigma_{i,k} \tag{247}$$

Thus we can prove this term tends to 0 if we can prove the Feller condition. To show this, first observe

$$\mathrm{E}\, X^2 = \mathrm{E}\, X^2 1_{|X|<\epsilon} + \mathrm{E}\, X^2 1_{|X|\geq\epsilon} \leq \epsilon^2 + \mathrm{E}\, X^2 1_{|X|\geq\epsilon} \tag{248}$$

Majorizing $\mathrm{E}\, X_{i,n}^2 1_{|X_{i,n}|\geq\epsilon}$ by the sum of all such terms we get

$$\max_i \sigma_{i,n}^2 \leq \epsilon^2 + \sum_{i=1}^{k_n} \mathrm{E}\, X_{i,n}^2 1_{|X_{i,n}|\geq\epsilon} < 2\epsilon^2 \tag{249}$$

for any n large enough that the Lindeberg condition implies the sum is less than ϵ^2. Since ϵ is arbitrary, this shows that $\max_i \sigma_{i,n}^2 \to 0$ as $n \to \infty$. ∎

4.15 (Weak Barry-Esséen theorem) Let $X_1, \ldots, X_n$ be iid copies of a real random variable X of mean zero, unit variance and finite third moment.

(i) Show that

$$\mathrm{E}\, G(\frac{X_1 + \cdots + X_n}{\sqrt{n}}) = \mathrm{E}\, G(N) + O(n^{-1/2}\,\mathrm{E}|X|^3 \sup_{x\in\mathbb{R}} |G'''(X)|) \tag{250}$$

(ii) Show that

$$\Pr(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \le 1) = \Pr(N \le t) + O(n^{-1/2}\,\mathrm{E}|X|^3)^{1/4} \tag{251}$$

for any $t \in \mathbb{R}$ with the implied constant absolute

(i) Extend the probability space by adding iid $\mathcal{N}(0,1)$ random variables $N_1, N_2, \ldots$ and N. Let $S_{n,k} = X_1 + \cdots + X_k + N_{k+1} + \cdots + N_n$. Let $Z_k = S_{n,k} - X_k = S_{n,k-1} - N_k$. Using the Taylor expansion for G we have

$$\begin{aligned} G(n^{-1/2}S_{n,k}) - G(n^{-1/2}S_{n,k-1}) =& G(Z_k) - G(Z_k) + \\ & \frac{1}{\sqrt{n}} G'(Z_k)(X_k - N_k) + \\ & \frac{1}{2n} G''(Z_k)(X_k^2 - N_k^2) + \\ & \frac{1}{6n^{3/2}}(G'''(a)X_k^3 - G'''(b)N_k^3) \end{aligned} \tag{252}$$

Here $a \in [0, n^{-1/2}\sum_i S_{n,k}]$ is the value for the remainder term, which depends on the value $S_{n,k}$, and similarly with b. Now take expectations of both sides conditional on the values of $X_1, \ldots, X_{k-1}$ and $N_{k+1}, \ldots, N_n$. The first line is trivially 0, the second is zero by independence and because $\mathrm{E}\, X_k = \mathrm{E}\, N_k = 0$. The third line by independence and because 0 because $\mathrm{Var}\, X_k = \mathrm{Var}\, N_k = 1$. Thus, using the tower law to take expectations of the conditional expectations, we can bound

$$|\mathrm{E}\, G(n^{-1/2}S_{n,k}) - \mathrm{E}\, G(n^{-1/2}S_{n,k-1})| \le \frac{1}{6} n^{-3/2} \sup |G'''(x)| (\mathrm{E}|X_k|^3 + \mathrm{E}|N_k|^3) \tag{253}$$

Note that $\mathrm{E}|X_k|^3 \ge (\mathrm{Var}\, X_k)^{3/2} = 1$ and $\mathrm{E}|N_k|^3 = 2\sqrt{2/\pi}$ is just a constant, so $\frac{1}{6}\mathrm{E}|X_k|^3 + \mathrm{E}|N_k|^3 \le \frac{1}{6}\left(1 + \sqrt{\frac{8}{\pi}}\right)\mathrm{E}|X_k|^3$ for some absolute constant c. Summing over the n terms $|\mathrm{E}\, G(n^{-1/2}S_{n,k}) - \mathrm{E}\, G(n^{-1/2}S_{n,k-1})|$, and noting that $n^{-1/2}\sum_{i=1}^n N_i \sim N$

$$\left| \mathrm{E}\, G\left(n^{-1/2}\sum_{i=1}^n X_i\right) - \mathrm{E}\, G(N) \right| \le O(n^{-1/2} \sup |G'''(x)|\, \mathrm{E}|X|^3) \tag{254}$$

(ii) Let ϕ be a non-negative, C^∞ "bump function" with support on $[-1, 1]$ which satisfies $\int_{-1}^1 \phi(x)\,dx = 1$. For $\lambda > 0$ we can scale $\phi_\lambda(x) = \lambda^{-1}\phi(\lambda^{-1}x)$ to get a narrower bump function supported on $[-\lambda, \lambda]$ where all the other properties (in particular, the total integral) are the same.

Then we can define an approximate step function $G_{\lambda,t}(x) = \int_{-\infty}^{x} \frac{1}{\lambda}\phi\left(\frac{x-t}{\lambda}\right)\, dx$. Note that $G_{\lambda,t}(x) = 0$ for $x \le t - \lambda$ and $G_{\lambda,t}(x) = 1$ for $x \ge t + \lambda$. Also $G_{\lambda,t}$ is increasing and $G_{\lambda,t} \in C^\infty$ with $|G'''_{\lambda,t}(x)| \le c/\lambda^3$ for some constant c.

Thus since $1_{s \le t} \le G_{t,t+\lambda}(x) \le 1_{x \le t+2\lambda}$,

$$\begin{aligned}\Pr(n^{-1/2}S_n \le t) &\le \mathrm{E}\, G_{\lambda,t+\lambda}(n^{-1/2}S_n) \\ &\le \mathrm{E}\, G_{\lambda,t+\lambda}(N) + \text{error} \\ &\le \Pr(N \le t + 2\lambda) + \text{error}\end{aligned} \tag{255}$$

Here the error term $O(\mathrm{E}|X|^3/\sqrt{n})$. By a similar comparison

$$\Pr(N \le t - 2\lambda) \le \Pr(n^{-1/2}S_n \le t) + \text{error} \tag{256}$$

Now $\Pr(N \le t + 2\lambda) - \Pr(N \le t) \le 2\lambda/\sqrt{2pi}$ (since the pdf of a normal distribution is peaked at $1/\sqrt{2\pi}$). Similarly $\Pr(N \le t) - \Pr(N \le t - 2\lambda) \le 2\lambda/\sqrt{2\pi}$. Therefore

$$|\Pr(n^{-1/2}S_n \le t) - \Pr(N \le t)| \le \sqrt{\frac{2}{\pi}}\lambda + c\frac{\mathrm{E}|X|^3}{\sqrt{n}\lambda^3} \tag{257}$$

Here the constant c is absolute and doesn't depend on any of the parameters. The λ which minimizes $a\lambda + b/\lambda^3$ is $\lambda = (3b/a)^{1/4}$ with minimum value

$$\left(3^{\frac{1}{4}} + 3^{-\frac{3}{4}}\right) a^{\frac{3}{4}} b^{\frac{1}{4}} \tag{258}$$

This gives us an error term $O((n^{-1/2}\,\mathrm{E}|X|^3)^{1/4})$

■

4.17 (Kolmogorov three-series theorem, converse direction) Let $X_1, X_2, \ldots$ be a sequence of jointly indepdendent real random variables with the property that the series $\sum_{n=1}^{\infty} X_n$ is almost surely convergent (i.e., the partial sums are almost surely convergent) and let $A > 0$

(i) Show that $\sum_{n=1}^{\infty} \Pr(|X_n| > A)$ is finite

(ii) Show that $\sum_{n=1}^{\infty} \mathrm{Var}(X_n 1_{|X_n| \le A})$ is finite

(iii) Show that the series $\sum_{n=1}^{\infty} \mathrm{E}\, X_n 1_{|X_n| \le A}$ is convergent

(i) By the second Borel-Cantelli and the independence of the X_n, if $\sum_{n=1}^{\infty} \Pr(|X_n| > A)$ is infinite than almost surely infinitely many of the events $|X_n| > A$ occur. However, if infinitely many $|X_n| > A$, then either infinitely many $X_n > A$ or infinitely many $X_n < -A$. Thus either $\limsup_n X_n \le A$ or $\liminf_n X_n \le -A$. Either way this contradicts the fact that $\lim X_n \to 0$ whenever $\sum_n X_n$ converges.

(ii) Let $Y_{i,n} = X_i 1_{|X_i|\leq A} - \mathrm{E}\, X_i 1_{|X_i|\leq A}$. Note that $\mathrm{Var}\, Y_{i,n} = \mathrm{Var}\, X_i 1_{|X_i|\leq A}$. Assume that $\sigma_n^2 = \sum_{i=1}^n \mathrm{Var}\, Y_{i,n} \to \infty$ as $n \to \infty$. Note that $Y_{i,n}$ bounded by $2A$, since each term in its definition is bounded by A from truncation. Choosing n large enough that $\epsilon\sigma_n > 2A$, we must have $Y_{i,n} 1_{|Y_{i,n}|>\epsilon\sigma_n} = 0$ for all i. The Lindeberg condition follows directly from this observation. Therefore $\sum_{i=1}^{k_n} Y_{i,n}/\sigma_n$ converges in distribution to $N(0,1)$.

On the other hand, note that $L = \lim_{n\to\infty} \sum_{i=1}^n (X_i 1_{|X_i|<A} - \mu_i)//\sigma_n$ is independent of the value of any finite collection of the X_i since the denominator tends to infinity. Since the X_i are mutually independent, Kolmogorov's 0-1 law implies that L is a constant, contradicting the fact it has a $N(0,1)$ distribution.

Therefore σ_n^2 must be bounded as $n \to \infty$. Since the quantity is increasing, it must converge. Hence $\sum_i^\infty \mathrm{Var}(X 1_{|X|\leq A})$ is finite.

(iii) Let $Y_n = X_n 1_{|X_n|\leq A} - \mathrm{E}\, X_n 1_{|X_n|\leq A}$. Note $\mathrm{E}\, Y_n = 0$, the Y_n are jointly independent, and $\sum_{n=1}^\infty \mathrm{E}\, Y_n^2 = \sum_{n=1}^\infty \mathrm{Var}(X_n 1_{|X_n|\leq A})$, which converges almost surely by the (ii). Therefore by theorem 3.26 $\sum_{n=1}^\infty Y_n$ converges almost surely. Futhermore $\sum_{n=1}^\infty X_n 1_{|X_n|\leq A}$ converges almost surely, since by (i) and Borel-Cantelli, almost surely it omits a finite number of terms. Note that

$$\sum_{n=1}^N \mathrm{E}\, X_n 1_{|X_n|\leq A} = \sum_{n=1}^N X_n 1_{|X_n|\leq A} - \sum_{n=1}^N Y_n \tag{259}$$

The right hand side converges on the intersection of two almost sure events (namely, that each of the terms converges), thus the left hand side converges almost surely.

■

4.19 Show that the normal $N(\mu, \sigma^2)$ has characteristic function $\phi(t) = e^{it\mu} e^{-\sigma^2 t^2/2}$

First we take the case $X \sim \mathcal{N}(0,1)$ For any bounded C^1 function f, when $X \sim \mathcal{N}(0,1)$ we have $\mathrm{E}\, f'(X) = \mathrm{E}\, XF(X)$. This follows from integration by parts

$$\begin{aligned} \mathrm{E}\, f'(X) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(x) e^{-x^2/2}\, dx \\ &= \left.\frac{f(x)e^{-x^2/2}}{\sqrt{2\pi}}\right|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \left(-xe^{x^2/2}\right)\, dx \\ &= \mathrm{E}\, X f(X) \end{aligned} \tag{260}$$

Therfore, for $f(x) = e^{itx}$ we get

$$\frac{d}{dt}\phi_X(t) = i\,\mathrm{E}\, Xe^{itX} = i^2\,\mathrm{E}\, te^{itX} = -t\phi_X(t) \tag{261}$$

The solution to this differential equation is $\phi_X(t) = Ce^{-t^2/2}$ for some constant C. Since $\phi_X(0) = 1$ the constant $C = 1$.

For the general case $X = \mu + \sigma N$ for some $N \sim \mathcal{N}(0,1)$. Hence

$$\phi_X(t) = \mathrm{E}\, e^{it(\mu+\sigma N)} = e^{i\mu t}\phi_N(\sigma t) = e^{i\mu t - \sigma^2 t^2/2} \tag{262}$$

■

4.20 Let $\lambda > 0$ and let X be a Poisson random variable with intensity λ, thus X takes values in the non-negative integers with $\Pr(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. Show that $\phi_X(t) = \exp(\lambda(e^{it} - 1))$

$$\mathrm{E}\, e^{itX} = \sum_{n=0}^{\infty} e^{itn}\frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^{it}\lambda)^n}{n!} = \exp(-\lambda)\exp(e^{it}\lambda) = \exp(\lambda(e^{it}-1)) \tag{263}$$

■

4.21 Let X be uniformly distributed in some interval $[a,b]$ show that $\phi_X(t) = \frac{e^{itb}-e^{ita}}{it(b-a)}$ for all non-zero t.

$$\mathrm{E}\, e^{itX} = \int_a^b e^{itx}\frac{dx}{b-a} = \frac{e^{itb} - e^{itb}}{it(b-a)} \tag{264}$$

■

4.22 Let $x_0 \in \mathbb{R}$ and $\gamma > 0$ and let X be a Cauchy random variable with parameters x_0, γ which means that X is a real random variable with probability density function $\frac{\gamma}{\pi((x-x_0)^2+\gamma^2)}$. Show that $\phi_X(t) = e^{ix_0 t}e^{-\gamma|t|}$ for all $t \in \mathbb{R}$

See Williams 16.3 ■

4.23 (Riemann-Lebesgue lemma) Show that if X is a real random variable that has an absolutely integrable probability density function f then $\phi_X(t) \to 0$ as $t \to \infty$. Note the claim fails if X doesn't have a probability density function.

Given $\epsilon > 0$, since continuous compactly supported functions are dense in L^1 so we can find a function $q \in C_0$ such that $\int_{\mathbb{R}}|p(x) - q(x)|\,dx < \epsilon$. Since q is compactly supported, there is some N such that $\operatorname{supp} g \subset [-N, N]$. Since q is uniformly continuous, we can find a step function $s(x) = \sum_{i=k}^{n} c_k 1_{[a_k,b_k)}$ which uniformly approximates q pointwise. That is for all $x \in \mathbb{R}$, $|q(x) - s(x)| < \epsilon/2N$. Then we have $\int_{\mathbb{R}}|g(x) - s(x)|\,dx < 2N(\epsilon/2N) = \epsilon$. Hence

$$\left|\int_{\mathbb{R}} e^{ixt}p(x)\,dx - \int_{\mathbb{R}} e^{ixt}s(x)\,dx\right| \le \int_{\mathbb{R}}|p(x) - s(x)| \le 2\epsilon \tag{265}$$

This shows we can find a step function s such that $|\widehat{f}(t) - \widehat{s}(t)| \le \epsilon$ uniformly in t.

For any single step $1_{[a,b)}$

$$\int_{\mathbb{R}} e^{itx} 1_{[a,b)}(x) = \int_a^b e^{itx} = \frac{e^{ibt} - e^{iat}}{it} \tag{266}$$

In particular note that $|\widehat{1_{[a,b)}}| \leq 2t^{-1}$. Therefore for a step function $s(x) = \sum_{k=1}^n c_k 1_{[a_k,b_k)}$ we have

$$|\widehat{s}(t)| \leq t^{-1} \sum_{k=1}^n |c_k| \tag{267}$$

Therefore $|\widehat{s}(t)| \to 0$ as $|t| \to \infty$. Since $|\phi_X(t) - \widehat{s}(t)| \leq \epsilon$ uniformly, this shows that $\phi_X(t) \to 0$ as $|t| \to \infty$. ∎

4.24 Show the characteristic function ϕ_X of a real random variable is in fact uniformly continuous on its domain.

$$|\phi(t+\delta) - \phi(t)| = |\mathrm{E}\, e^{itX}(e^{i\delta X} - 1)| \leq \mathrm{E}|e^{i\delta X} - 1| \tag{268}$$

Choose N so large that $\Pr(|X| > N) < \epsilon$. Since e^{iNx} is continuous at zero and $e^0 = 1$, we can choose δ small enough such that $|e^{i\delta X} - 1| < \epsilon$ whenever $|X| \leq N$. To estimate the tail note that $|e^{i\delta X} - 1| \leq 2$, so

$$\mathrm{E}_{|X|>N}|e^{i\delta X} - 1| \leq 2\Pr(|X| > N) \leq 2\epsilon \tag{269}$$

Hence, for δ small enough independent of t we have

$$|\phi(t+\delta) - \phi(t)| \leq 3\epsilon \tag{270}$$

This shows that $\phi_X(t)$ is uniformly continuous. ∎

4.25 Let X be a real random variable with finite k^{th} moment for some $k \geq 1$. Show that ϕ_X is k times continuously differentiable with

$$\frac{d^j}{dt^j}\phi_X(t) = i^j\, \mathrm{E}\, X^j e^{itX} \tag{271}$$

for all $0 \leq j \leq k$. Conclude in particular that the partial Taylor expansion

$$\phi_X(t) = \sum_{j=0}^k \frac{(it)^j}{j!}\, \mathrm{E}\, X^j + o(|t|^k) \tag{272}$$

where $o(|t|^k)$ is a quantity that goes to zero as $t \to 0$ times $|t|^k$.

We will justify differentiating under the integral sign. First consider the finite difference

$$\frac{\phi_X(t_h) - \phi_X(t)}{h} = \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} \, dm \tag{273}$$

where m is the measure on $\mathbb{R}$ which corresponds to the distribution of X. Note there is a bound

$$|e^{\theta i} - 1| \leq |\theta| \tag{274}$$

since $|e^{\theta i} - 1|$ is the distance of the chord on the unit circle from 1 to $e^{i\theta}$ and θ is the arc length from 1 to $e^{i\theta}$. Therefore the integrand is dominated by $|hx|/h = |x|$. By assumption $\mathrm{E}|X| < \infty$, so by dominated convergence

$$\lim_{h\to\infty} \frac{\phi_X(t_h) - \phi_X(t)}{h} = \int_{\mathbb{R}} e^{itx} \left. \frac{de^{itx}}{dt} \right|_{t=0} dm = \int_{\mathbb{R}} e^{itx} ix \, dm = i \,\mathrm{E}\, e^{itX} X \tag{275}$$

To compute $\frac{d^k}{dt^k}\phi_X(t)$, proceed by induction using the measure $x^{k-1}\, dm$ instead of dm. Dominated convergence still applies since $\mathrm{E}|X|^k < \infty$ by assumption, so we can differentiate under the integral sign.

$$\frac{d^k}{dt^k}\phi_X(t) = \frac{d}{dt}(i^{k-1}\,\mathrm{E}\,X^{k-1}e^{itX}) = i^{k-1}\,\mathrm{E}\left(X^{k-1}\frac{d}{dt}e^{itX}\right) = i^k\,\mathrm{E}\,X^k e^{itX} \tag{276}$$

In particular note $\phi_X^{(k)}(0) = i^k\,\mathrm{E}\,X^k$ so using Taylor's theorem with remainder we get the expansion in the problem statement. ∎

> **4.26** Let X be a real random variable and assume that it is *subgaussian* in the sense that there exist constants $C, c \geq 0$ such that
>
> $$\Pr(|X| \geq t) \leq Ce^{-ct^2} \tag{277}$$
>
> for all $t \in \mathbb{R}$. (A bounded random variable is subgaussian, as is any Gaussian random variable). In this case, rigorously establish
>
> $$\phi_X(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\,\mathrm{E}\,X^k \tag{278}$$
>
> and show that the series converges locally uniformly in t

(Sketch) Use the formula $\mathrm{E}|X|^k = \int_0^\infty kt^{k-1}\Pr(|X| \leq t)\,dt$ to get an upper bound for the moments of X using the subgaussian bound. We find that $\mathrm{E}|X|^k \leq \Gamma(k/2)$ approximately. Since $\Gamma(k/2)/k! \leq 1/(k/2)!$, the Taylor series is convergent for all x. This is because for $k > k_0 = 4x^2$, the terms $t^k/(k/2)! \leq A2^{k_0}2^{-k}$, and, excluding a finite number of terms, we can dominate the tail terms by a geometric series. In fact if we take $k_0 = 8x^2$ then for all $|t'| \leq \sqrt{2}|t|$ we can dominate the terms by a geometric series $A'2^{k_0}2^{-k}$ uniformly. This shows for every t, the Taylor series converges uniformly on a compact set containing t, which is the definition of local uniform convergence. ∎

4.29 (Levy's continuity theorem, full version) Let X_n be a sequence of real valued random variables. Suppose that ϕ_{X_n} converges pointwise to a limit ϕ. Show the following are equivalent.

(i) ϕ is continuous at 0

(ii) X_n is a tight sequence (as in Exercise 4.10(iii))

(iii) ϕ is the characteristic function of a real valued random variable X (possibly after extending the sample space)

(iv) X_n converges in distribution to some real valued random variable X (possibly after extending the sample space)

- (i) $\Rightarrow$ (ii) Consider the average value of ϕ_{X_n} and ϕ on a small interval $(-\delta, \delta)$

$$A_n(\delta) = \frac{1}{2\delta}\int_{-\delta}^{\delta} \phi_{X_n}(t)\,dt \qquad A(\delta) = \frac{1}{2\delta}\int_{-\delta}^{\delta} \phi(t)\,dt \tag{279}$$

First, since $|\phi_{X_n}| \le 1$ for all n, by bounded convergence we have $A_n(\delta) \to A(\delta)$ as $n \to \infty$. Furthermore, by the continuity of ϕ at 0, since $\phi(0) = \mathrm{E}\, e^0 = 1$, by choosing δ small enough $|1 - A(\delta)| \le \epsilon$ for any $\epsilon > 0$.

Now these integrals (279) can be regarded as $(2\delta)^{-1}\int_{\mathbb{R}} \hat{G}_\delta(t)\phi_{X_n}(t)\,dt$ where $\hat{G}_\delta(t) = 1_{|t|\le\delta}(t)$. The Fourier pair corresponding to $\hat{G}_\delta$ satisfies

$$G_\delta(x) = \frac{2\sin(\delta x)}{x} \tag{280}$$

so by identity (7) in the text, $A_n(\delta) = (2\delta)^{-1}\,\mathrm{E}\, G_\delta(X_n)$. Therefore we can make the simple estimate

$$\begin{aligned} |A_n(\delta)| &\le \left|\int_{\mathbb{R}} \frac{\sin(\delta x)}{\delta x}\,d\mu_n\right| \le \left|\int_{|x|<T} \frac{\sin(\delta x)}{\delta x}\,d\mu_n\right| + \left|\int_{|x|\ge T} \frac{\sin(\delta x)}{\delta x}\,d\mu_n\right| \\ &\le \Pr(|X| < T) + \frac{1}{T\delta}\Pr(|X| \ge T) \end{aligned} \tag{281}$$

Hence

$$1 - A_n \ge \left(1 - \frac{1}{T\delta}\right)\Pr(|X_n| \ge T) \tag{282}$$

Choosing $T = \delta/2$ gives

$$\Pr(|X_n| \ge 2/\delta) \le 2|1 - A_n| \tag{283}$$

Now$|1 - A| \le \epsilon$ for small enough δ. Also for some N_δ, for any $n \ge N_\delta$ we have $|A - A_n| \le \epsilon$, so we have $\Pr(|X_n| \ge 2/\delta) \le 4\epsilon$, which can be made arbitrarily small. This shows that X_n is tight.

- (ii) $\Rightarrow$ (iii) and (iv): By theorem 11 (Prokhorov's theorem), there is a subsequence n_j and a random variable X (possibly on an extended probability space from the $X_1, X_2, \dots$) such that $X_{n_j} \to X$ in distribution. By theorem 27, $\phi_{X_{n_j}} \to \phi_X$ pointwise. However, $\phi_{X_{n_j}} \to \phi$ pointwise, since ϕ_{n_j} is a subsequence of ϕ_n and every subsequence has the same limit. Therefore $\phi = \phi_X$ and, by theorem 27 in the other direction, $X_n \to X$ in distribution.
- (iii) $\Leftrightarrow$ (iv) by Theorem 27.
- (iii) $\Rightarrow$ (i) by exercise 4.24

∎

4.31 (Esséen concentration inequality) Let X be a random variable taking values in $\mathbb{R}$. Then for any $r > 0, \epsilon > 0$ show that

$$\sup_{x_0 \in \mathbb{R}} \Pr(|X - x_0| \le r) \le C_\epsilon r \int_{t \in \mathbb{R}: |t| < \epsilon/r} |\phi_X(t)| \, dt \tag{284}$$

for some constant C_ϵ depending only on ϵ. The left hand side (as well as higher dimensional analogues) is known as the small ball probability of X at radius r.

First note that if we can establish the inequality for $x_0 = 0$ we have established it for all x_0 since $\phi_{X+x_0}(t) = e^{ix_0}\phi_X(t)$, and the right side is unchanged under this transformation. Essentially this inequality follows directly from the scaling identity $H(\hat{t}/r) = rH(rx)$ and $\mathrm{E}\, H(X) = \int_{\mathbb{R}} \hat{H}(t)\phi_X(t)\, dt$ for a well-chosen G. We want $\hat{H}$ to be compactly supported and H to be nonnegative. One choice is $H = G^2$ and $\hat{H} = \hat{G} * \hat{G}$ where G is the function used in problem 4.29(i). Scaling t and the function, we can write $\hat{H}(t) = (1 - |t|)^+$ and $H(t) = 2\sin^2(x/2)/x^2 = 2(1 - \cos x)/x^2$. For $\epsilon \in (0, 2\pi)$, since H is peaked around 0 and non-negative,and since $\hat{H}(t) \le 1_{\{|u| \le 1\}}(t)$

$$H(\epsilon) \Pr(|X| \le \epsilon) \le \mathrm{E}\, H(X) = \int_{\mathbb{R}} \hat{H}(t)\phi_X(t)\, dt \le \int_{|t| \le 1} |\phi_X(t)|\, dt \tag{285}$$

If we scale X by ϵ/r and set $C_\epsilon = 1/\epsilon H(\epsilon)$ this becomes the desired relation

$$\Pr(|X| \le r) \le C_\epsilon r \int_{|t| \le \epsilon/r} |\phi(t)|\, dt \tag{286}$$

∎

4.32 (Fourier identities) Let X, Y be independent real random variables. Then

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) \tag{287}$$

for all $t \in V$. Also for any scalar c one has

$$\phi_{cX}(t) = \phi_X(\bar{c}t) \tag{288}$$

and more generally, for any linear transformation $T : V \to V$ one has

$$\phi_{TX}(t) = \phi_X(T^* t) \tag{289}$$

For the first identity observe

$$\phi_{X+Y}(t) = \mathrm{E}\, e^{it(X+Y)} = \mathrm{E}\, e^{itX} e^{itY} = \mathrm{E}\, e^{itX}\, \mathrm{E}\, e^{itY} = \phi_X(t)\phi_Y(t) \tag{290}$$

For the second (and I think this is somewhat conventional, what's the characteristic function of a complex argument supposed to mean?)

$$\phi_{cX}(t) = \mathrm{E}\, e^{i\langle t, cX\rangle} = \mathrm{E}\, e^{i\langle \bar{c}t, X\rangle} \tag{291}$$

For the third

$$\phi_{cX}(t) = \mathrm{E}\, e^{i\langle t, TX\rangle} = \mathrm{E}\, e^{i\langle T^* t, X\rangle}$$

∎

4.34 (Vector-valued central limit theorem) Let $\vec{X} = (X_, \dots, X_d)$ be a random variable taking values in $\mathbb{R}^d$ with finite second moment. Define the covariance matrix $\Sigma(\vec{X})$ to be the $d \times d$ matrix Σ whose ij^{th} entry is the covariance $\mathrm{E}(X_i - \mathrm{E}(X_i))(X_j - \mathrm{E}(X_j))$.

(i) Show that the covariance matrix is positive semi-definite real symmetric.

(ii) Conversely, given any positive definite real symmetric $d \times d$ matrix Σ and $\mu \in \mathbb{R}^d$ show that the multivariate normal distribution $N(\mu, \Sigma)_{\mathbb{R}^d}$ given by the absolutely continuous measure

$$\frac{1}{\sqrt{(2\pi)^d \det \Sigma}} e^{-(x-\mu)\cdot\Sigma^{-1}(x-\mu)/2}\, dx \tag{292}$$

has mean μ and covariance matrix Σ and has a characteristic function

$$\Phi(t) = e^{i\mu\cdot t} e^{-t\cdot\Sigma t/2} \tag{293}$$

How would one define the normal distribution $N(\mu, \Sigma)_{\mathbb{R}^d}$ if Σ degenerated to be merely positive semi-definite instead of positive definite?

(iii) If $\vec{S}_n := \vec{X}_1 + \cdots + \vec{X}_n$ is the sum of n iid copies of $\vec{X}$ show that $\frac{\vec{S}_n - n\mu}{\sqrt{n}}$ converges in distribution to $N(0, \Sigma(X))_{\mathbb{R}^d}$.

(i) The symmetry of Σ is evident since

$$\Sigma_{ij} = \mathrm{E}(X_i - \mathrm{E}\,X_i)(X_j - \mathrm{E}\,X_j) = \mathrm{E}(X_j - \mathrm{E}\,X_j)(X_i - \mathrm{E}\,X_i) = \Sigma_{ji} \tag{294}$$

Now let $a_1, \ldots, a_d \in \mathbb{R}$ let $Y = \sum_{i=1}^{d} a_i X_i$. Then

$$\operatorname{Var} Y = \sum_{i,j=1}^{d} a_i a_j \operatorname{Cov}(X_i, X_j) = a \cdot \Sigma a \tag{295}$$

Since $\operatorname{Var} Y \geq 0$, this shows Σ is positive semi-definite,

(ii) First consider $\mu = 0$ and diagonal Σ

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & & \ddots & \\ 0 & & & \sigma_d^2 \end{pmatrix} \tag{296}$$

Then consider the absolutely continuous measure for $(X_1, \ldots, X_d)$ where component $X_i \sim \mathcal{N}(0, \sigma_i)$ and all of the components are independent. Then the pdf is given by

$$\left(\frac{1}{\sqrt{2\pi}\sigma_1} \exp(-x_1^2/2\sigma_1^2) \right) \cdots \left(\frac{1}{\sqrt{2\pi}\sigma_d} \exp(-x_d^2/2\sigma_d^2) \right) \tag{297}$$

which has the form (292) since $\det \Sigma = \prod \sigma_i^2$, and $\Sigma^{-1} = \operatorname{diag}(1/\sigma_1, \ldots, 1/\sigma_d)$ and since $x \cdot \Sigma^{-1} x = \sum_{i=1}^{d} x_i^2/\sigma_i^2$.

For this special case, the characteristic function is separable

$$\begin{aligned} \Phi(t) = \mathrm{E}\, e^{it \cdot X} &= \int_{\mathbb{R}^d} \left(\frac{e^{it_1 x_1 - x_1^2/2\sigma_1^2}}{\sqrt{2\pi}\sigma_1} \right) \cdots \left(\frac{e^{it_d x_d - x_d^2/2\sigma_d^2}}{\sqrt{2\pi}\sigma_d} \right) \\ &= \mathrm{E}\, e^{it_1 X_i} \cdots \mathrm{E}\, e^{it_d X_d} \\ &= e^{-\sigma_1^2 x_1^2/2} \cdots e^{-\sigma_d^2 x_d^2/2} \\ &= e^{x \cdot \Sigma x} \end{aligned} \tag{298}$$

For the general case, note that any positive semidefinite symmetric $\Sigma = U^* D U$ for a orthogonal matrix U and a diagonal matrix D, where the entries in D are non-negative. Here $*$ represents the adjoint (transpose) operator which for orthogonal U has the property $U^{-1} = U^*$. Then let $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_d)$ be a radom variable as described above with $\mathcal{N}(0, D)$, and let $X = U\tilde{X} + \mu$. Clearly $\mathrm{E}\, X = \mu$ and $\operatorname{Cov} X = U^* \operatorname{Cov} \tilde{X} U = U^* D U = \Sigma$. Since we can write $\tilde{X} = U^*(X - \mu)$, the pdf of X can be written

$$\frac{1}{(\sqrt{2\pi})^d \det D} e^{-(U^*(x-\mu)) \cdot D^{-1}(U^*(x-\mu))/2} \tag{299}$$

Note that $\Sigma^{-1} = UD^{-1}U^*$ and $\det \Sigma = (\det U)^{-1} \det D \det U = \det D$ so this expression is actually the same as (292). Furthermore by 4.32 and the translation identity

$$\Phi_{U^*X+\mu}(t) = e^{i\mu t}\Phi_X(Ut) = e^{i\mu t}e^{-(U^*t)\cdot(DUt)/2} = e^{i\mu t}e^{-t\cdot\Sigma t/2} \tag{300}$$

From this expression its also clear that $\left.\frac{\partial\Phi}{\partial t_i}\right|_0 = i\mu_i$ and $\left.\frac{\partial^2\Phi}{\partial t_i \partial t_j}\right|_0 = -\Sigma_{ij}$, so the entries in μ and Σ do really correspond to the mean and covariances of X.

(iii) By the multivariate Taylor expansion we can write

$$\Phi_{\vec{X}}(t) = 1 + i\mu t - \frac{1}{2}t\cdot\Sigma t + o(|t|^2) = e^{i\mu t - t\cdot\Sigma t/2 + o(|t|^2)} \tag{301}$$

Therefore

$$\Phi_{(\vec{S}_n - n\mu)/\sqrt{n}}(t) = e^{i\sqrt{n}\mu}(\Phi_{\vec{X}}(t/\sqrt{n}))^n = \exp(-t\cdot\Sigma t/2 + no(n^{-1})) \tag{302}$$

The expression converges to $\exp(-t\cdot\Sigma t/2)$. By Levy's continuity theorem, this means that $\frac{\vec{S}_n - n\mu}{\sqrt{n}}$ converges in distribution to the multivariate normal $\mathcal{N}(0,\Sigma)$.

■

4.35 (Complex central limit theorem) Let X be a complex random variable of mean $\mu \in \mathbb{C}$ whose real and imaginary parts have variance $\sigma^2/2$ and covariance 0. Let $X_1, \dots, X_n \sim X$ be iid copies of X. Show that as $n \to \infty$ the normalized sums $\frac{\sqrt{n}}{\sigma}\left(\frac{X_1+\cdots+X_n}{n} - \mu\right)$ converges in distribution to the standard complex gaussian $N(0,1)_{\mathbb{C}}$ defined as the measure μ on $\mathbb{C}$ with

$$\mu(S) := \frac{1}{\pi}\int_S e^{-|z|^2}\,dz \tag{303}$$

for Borel sets $S \subset \mathbb{C}$ where dz the Lebesgue measure on $\mathbb{C}$ (identified as $\mathbb{R}^2$ in the usual fashion)

Note that the variance of X is defined such that $\operatorname{Var} X = \sigma^2$ as a complex variable in the sense

$$\mathrm{E}|X-\mu|^2 = \mathrm{E}(\Re X - \mu_1)^2 + \mathrm{E}(\Im X - \mu_2)^2 = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2 \tag{304}$$

Considering X to be a vector in $\mathbb{R}^2$, the covariance is given by

$$\operatorname{Cov} X = \Sigma = \begin{pmatrix} \frac{\sigma^2}{2} & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix} \tag{305}$$

Exercise 4.34 shows that the pdf of the normalized sums in $\mathbb{R}^2$ is given by

$$\frac{1}{\sqrt{(2\pi)^2\cdot(1/2)^2}}e^{-(2x_1^2+2x_2^2)/2} = \frac{1}{\pi}e^{-|x|^2} \tag{306}$$

■

4.36 Use characteristic functions and the truncation argument to give an alternate proof of the Lindeberg central limit theorem (Theorem 14)

??? I think this is the same as 4.41, see that solution ■

4.38 Show the error terms in theorem 37 are sharp (up to constants) when X is a signed Bernoulli random variable

Let $X = +1$ or $X = -1$ with probability 1/2. Now consider $\Pr(S_{2n}/\sqrt{2n} \leq 0)$. By symmetry, for any $k \in \mathbb{N}$, we have $\Pr(S_{2n} = -k) = \Pr(S_{2n} = k)$. Therefore

$$\Pr(S_{2n}/\sqrt{2n} \leq 0) = \frac{1}{2} + \frac{1}{2}\Pr(S_{2n} = 0) = \Pr(N \leq 0) + \frac{1}{2}\Pr(S_{2n} = 0) \tag{307}$$

where $N \sim N(0,1)$. By Stirling's approximation $n! = O(\sqrt{n}(n/e)^n)$ we have

$$\Pr(S_{2n} = 0) = \binom{2n}{n} 2^{-2n} = O\left(\frac{\sqrt{2n}(2n/e)^{2n}}{(\sqrt{n}(n/e)^n)^2} 2^{-2n}\right) = O(n^{-1/2}) \tag{308}$$

This shows, up to a constant, the error in theorem 37 cannot be improved. ■

4.39 Let X_n be a sequene of real random variables which converge in distribution to a real random variable X and let Y_n be a sequence of real random variables which converge in distribution to a real random variable Y. Suppose that for each n, X_n and Y_n are independent and suppose also that X and Y are independent. Show that $X_n + Y_n$ converges in distribution to $X + Y$.

Note that $\phi_{X_n+Y_n}(t) = \phi_{X_n}(t)\phi_{Y_n}(t)$ by independence. Clearly $\phi_{X_n+Y_n}(t) \to \phi_X(t)\phi_Y(t)$ pointwise. Also, $\phi_X(t)\phi_Y(t)$ is evidently the characteristic function for $X + Y$, so it must be that $X_n + Y_n$ converges in distribution for $X + Y$. In fact we can say something stronger. If we don't know that X and Y are independent a priori, the continuity of $\phi_X(t)$ and $\phi_Y(t)$ at 0 imply that $\phi_X(t)\phi_Y(t)$ is continuous at 0. Therefore there is a random variable Z such that $X_n + Y_n$ converges to Z in distribution. Furthermore, $\phi_Z(t) = \phi_X(t)\phi_Y(t)$, so by inspection the distribution of Z is given by $X' + Y'$ where $X' \sim X$ and $Y' \sim Y$ and X' and Y' are independent. ■

4.40 Let $X_1, X_2, \ldots$ be iid copies of an absolutely integrable random variable X of mean zero.

(i) In this part we assume that X is *symmetric* which means that X and $-X$ have the same distribution. Show that for any $t > 0$ and $M > 0$

$$\Pr(X_1 + \cdots + X_n \geq t) \geq \frac{1}{2}\Pr(X_1 1_{|X_1| \leq M} + \cdots + X_n 1_{|X_n| \leq M} \geq t) \tag{309}$$

(ii) If X is symmetric and $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges in distribution to a real random variable Y, show that X has finite variance.

(iii) Generalize (ii) by removing the hypothesis that X is symmetric

(i) Note that $S_n = X_1 + \cdots + X_n$ can be written $S_n = A_n + B_n$ where

$$A_n = X_1 1_{|X_1| \le M} + \cdots + X_n 1_{|X_n| \le M} \tag{310}$$
$$B_n = X_1 1_{|X_1| > M} + \cdots + X_n 1_{|X_n| > M} \tag{311}$$

Clearly $\{S_n \ge t\} \supset \{A_n \ge t, B_n \ge 0\}$. Now B_n is a symmetric random variable since the X_i are symmetric. Consequently, $\Pr(B_n > 0) = \Pr(B_n < 0)$ and hence $\Pr(B_n \ge 0) \ge \frac{1}{2}$.

Furthermore, A_n and B_n are independent. This is because the X_i's which contribute to the value of A_n are disjoint from the X_i's which contribute to the value of B_n. In more detail, given $I \subset \{1, \ldots, n\}$, if we condition $|X_i| \le M$ for $i \in I$ and $|X_i| > M$ for $i \notin I$, then A_n depends only on X_i for $i \in I$ and B_n depends only on X_i for $i \notin I$. So conditionally, A_n and B_n are independent. However, this same argument applies for all subsets $I \subset \{1, \ldots, n\}$ so A_n and B_n are unconditionally indepdent.

Therefore

$$\Pr(S_n \ge t) \ge \Pr(A_n \ge t, B_n \ge 0) = \Pr(A_n \ge t)\Pr(B_n \ge 0) \ge \frac{1}{2}\Pr(A_n \ge t) \tag{312}$$

(ii) For any fixed M $\operatorname{Var} X1_{|X| \le M} \le M^2$ is finite. By the central limit theorem, $A_n/\sqrt{n} \to \mathcal{N}(0, \sigma^2)$ in distribution where $\sigma^2 = \operatorname{Var} X1_{|X| \le M}$. In particular for $\epsilon > 0$, for large enough n, if we let $N \sim \mathcal{N}(0,1)$

$$\Pr\left(\frac{X_1 1_{|X_1| \le M} + \cdots + X_n 1_{|X_n| \le M}}{\sqrt{n}} \ge t\right) \ge \Pr\left(N \ge \frac{t}{\sigma}\right) - \epsilon \tag{313}$$

By the monotone convergence theorem, $\operatorname{Var} X1_{|X| \le M} \uparrow \operatorname{Var} X$ as $M \uparrow \infty$, and this holds even if $\operatorname{Var} X = \infty$. So, if X does not have finite variance, as $M \to \infty$, then $\sigma \to \infty$ and, for any fixed t, $\Pr(N \ge t/\sigma) \to \frac{1}{2}$. Thus may choose M and n large enough so that the expression above is at least $\frac{1}{4}$.

If $(X_1 + \cdots + X_n)/\sqrt{n}$ converges to a random variable Y in distribution then, except for at most countably many t,

$$\Pr\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \ge t\right) \to \Pr(Y \ge t) \tag{314}$$

Choose t large enough so that $\Pr(Y \ge t) \le \epsilon$ and n large enough so that $\Pr(X_1 + \cdots + X_n \ge t\sqrt{n}) \le 2\epsilon$. Then

$$\begin{aligned} 2\epsilon &\ge \Pr(X_1 + \ldots X_n \ge t\sqrt{n}) \\ &\ge \frac{1}{2}\Pr(X_1 1_{|X_1| \le M} + \cdots + X_n 1_{|X_n| \le M} \ge t\sqrt{n}) \\ &\ge \frac{1}{8} \end{aligned} \tag{315}$$

This is a contradiction since we can make ϵ arbitrarily small with our chose of t, but this inequality says ϵ is at least $\frac{1}{16}$.

(iii) For arbitrary X we can consider $\tilde{X} = X - X'$ where X' has the same distribution as X and is independent. Now $-\tilde{X} = X' - X$ evidently has the same distribution as $\tilde{X}$ since its the difference of independent random variables whose distribution is the same as X. Hence $\tilde{X}$ is symmetric.

Suppose that $(X_1 + \cdots + X_n)/\sqrt{n} \to Y$ in distribution for some random variable Y. Then by exercise 4.39, we have $(\tilde{X}_1 + \ldots \tilde{X}_n)/\sqrt{n} \to Y - Y'$ where Y' has the same distribution as Y and is independent of it. By (ii) this implies that $\tilde{X}$ has finite variance. By direct calculation $\operatorname{Var} \tilde{X} = \operatorname{Var} X + \operatorname{Var} X' = 2 \operatorname{Var} X$ so we conclude $\operatorname{Var} X$ is finite.

∎

4.41

(i) If X is a real random variable of mean zero and variance σ^2 and t is a real number, show that

$$\phi_X(t) = 1 + O(\sigma^2 t^2) \tag{316}$$

and that

$$\phi_X(t) = 1 - \frac{1}{2}\sigma^2 t^2 + O(t^2 \operatorname{E} \min(|X|^2, t|X|^3)) \tag{317}$$

(ii) Establish the pointwise inequality

$$|z_1 \cdots z_n - z_1' \cdots z_n'| \le \sum_{i=1}^{n} |z_i - z_i'| \tag{318}$$

whenever $z_1, \ldots, z_n, z_1', \ldots, z_n'$ are in the complex disk $\{z \in \mathbb{C} : |z| \le 1\}$

(iii) Suppose that for each n, $X_{1,n}, \ldots, X_{n,n}$ are jointly independent real random variables of mean zero and finite variance obeying the uniform bound

$$|X_{i,n}| \le \epsilon_n \sqrt{n} \tag{319}$$

for all $i = 1, \ldots, n$ and some ϵ_n going to zero as $n \to \infty$, and obeying the variance bound

$$\sum_{i=1}^{n} \operatorname{Var}(X_{i,n}) \to \sigma^2 \tag{320}$$

as $n \to \infty$ for some $0 < \sigma < \infty$. If $S_n := X_{1,n} + \cdots + X_{n,n}$ use (i) and (ii) to show that

$$\phi_{S_n/\sqrt{n}}(t) \to e^{-\sigma^2 t^2/2} \tag{321}$$

as $n \to \infty$ for any given t

(iv) Use (iii) and a truncation argument to give an alternative proof of the Lindeberg central limit theorem (theorem 14)

(i) From Taylor's theorem with remainder

$$e^{ix} = \sum_{k=0}^{n} \frac{i^k}{k!}x^k + \frac{i^{n+1}}{n!}\int_0^x (x-u)^n e^{iu}\,du \tag{322}$$

This immediately leads to a bound

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{i^k}{k!}x^k \right| \le \frac{|x|^{n+1}}{(n+1)!} \tag{323}$$

Using this inequality for $n-1$ we can get a bound

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{i^k}{k!}x^k \right| \le \left| e^{ix} - \sum_{k=0}^{n-1} \frac{i^k}{k!}x^k \right| + \frac{|x|^n}{n!} \le \frac{2|x|^n}{n!} \tag{324}$$

Since by assumption $\mathrm{E}\,X = 0$ and $\mathrm{E}\,X^2 = \sigma^2$, we have $\mathrm{E}(1 + itX - \frac{1}{2}t^2X^2) = 1 - \frac{1}{2}t^2\sigma^2$. Therefore we can use the above to bound the characteristic function, for an absolute constant C

$$\left| \phi_X(t) - 1 + \frac{1}{2}t^2\sigma^2 \right| \le \mathrm{E}\,C\min(t^2|X|^2, t^3|X|^3) \tag{325}$$

(ii) It sufficies to prove the case $n = 2$ since the general case follows by induction.

$$z_1z_2 - z_1'z_2' = (z_1 - z_1')z_2 + (z_2 - z_2')z_1' \tag{326}$$

Taking the absolute value of both sides, using the triangle inequality, and noting $|z_2| \le 1$ and $|z_1'| \le 1$ we have

$$|z_1z_2 - z_1'z_2'| \le |z_1 - z_1'| + |z_2 - z_2'| \tag{327}$$

(iii) I'm just going to prove the Lindeberg central limit theorem rather than the problem as stated

(iv) I'm just going to prove the Lindeberg central limit theorem rather than the problem as stated

Let $S_n = X_{1,n} + \cdots + X_{k_n,n}$ where $\mathrm{E}\,X_{i,n} = 0$ and $\sigma_{i,n}^2 = \mathrm{E}\,X_{i,n}^2$ and $\sum_{i=1}^{k_n} \sigma_{i,n}^2 = 1$. (If the sum of the variances is not one, we can divide every term $X_{i,n}$ by this quantity to satisfy the hypothesis of the theorem). We wish to show that for all t

$$\left| \phi_{S_n}(t) - e^{-t^2/2} \right| \to 0 \tag{328}$$

Then by Levy's continuity theorem, this shows that S_n converges to $N(0,1)$ in distribution. Noting $\phi_{S_n}(t) = \prod_{i=1}^{k_n} \phi_{X_{i,n}}(t)$ and $e^{-t^2/2} = \prod_{i=1}^{k_n} e^{-\sigma_{i,n}^2 t^2/2}$, by (ii) it suffices to prove

$$\sum_{i=1}^{n} |\phi_{X_{i,n}}(t) - e^{-\sigma_{i,n}^2 t^2/2}| \to 0 \qquad \text{as } n \to \infty \tag{329}$$

First note that by the inequality in (i)

$$\begin{aligned}\sum_{i=1}^{k_n}\left|\mathrm{E}\,\phi_{X_{i,n}}(t)-1+\frac{1}{2}\sigma_{i,n}^2t^2\right| &\le \sum_{i=1}^{k_n}\mathrm{E}\,|tX_{i,n}|^3 1_{|X_{i,n}|\le\epsilon}+\sum_{i=1}^{k_n}\mathrm{E}\,|tX_{i,n}|^2 1_{|X_{i,n}|>\epsilon}\\ &\le \epsilon t^2\sum_{i=1}^{k_n}\sigma_{i,n}^2+\sum_{i=1}^{k_n}\mathrm{E}\,|tX_{i,n}|^2 1_{|X_{i,n}|>\epsilon}\end{aligned} \tag{330}$$

The first term is just ϵt^2 and the second tends to 0 as $n \to \infty$ by the Lindeberg condition. Since ϵ is arbitrary this shows the expression tends to 0. Next from the bound $|e^x - 1 + x| \le x^2$ we get

$$\sum_{i=1}^{k_n}\left|e^{-\sigma_{i,n}^2t^2/2}-1+\frac{1}{2}\sigma_{i,n}^2t^2\right| \le t^4\sum_{i=1}^{k_n}\sigma_{i,n}^4 \le t^4\max_i\sigma_{i,n}^2\sum_{i=1}^{k_n}\sigma_{i,n}^2 = t^4\max_i\sigma_{i,n}^2 \tag{331}$$

As shown in 4.14, the Lindeberg condition implies the Feller condition, so this term also tends to 0 as n tends to infinity. ∎

4.42 (Subgaussian random variables) Let X be a real random variable. Show that the following statements are equivalent.

(i) There exists a $C, c > 0$ such that $\Pr(|X| \ge t) \le Ce^{-ct^2}$ for all $t > 0$

(ii) There exists a $C' > 0$ such that $\mathrm{E}|X|^k \le (C'k)^{k/2}$ for all $k \ge 1$

(iii) There exists a $C'', c'' > 0$ such that $\mathrm{E}\, e^{tX} \le C''e^{c''t^2}$ for all $t \in \mathbb{R}$

Furthermore, if (i) holds for some C, c then (ii) holds for C' depends only on C, c and similarly for any of the other implications. Variables obeying (i), (ii)or (iii) are called *subgaussian*. The function $t \mapsto \mathrm{E}\, e^{tX}$ is known as the *moment generating function* of X; it is of course closely related to the characteristic function.

- (iii) ⇒ (i): Using the Markov inequality for $\theta > 0$

$$\Pr(X \ge t) = \Pr(e^{\theta X} \ge e^{\theta t}) \le \frac{\mathrm{E}\, e^{\theta X}}{e^{\theta t}} \le C''\exp(c''\theta^2 - t\theta) \le C''\exp(-t^2/2c'') \tag{332}$$

In the last inequality, we maximized the expression over all θ to find $\theta = t/2c''$. Note that if X satisfies the inequality in (iii) then $-X$ does as well since $\mathrm{E}\, e^{t(-X)} = \mathrm{E}\, e^{(-t)X} \le C''e^{c''t^2}$. Therefore using (332) for $-X$

$$\Pr(X \le -t) = \Pr(-X \ge t) \le C''\exp(-t^2/2c'') \tag{333}$$

Adding this equation to (332) gives $\Pr(|X| \ge t) \le 2C''\exp(c''t^2/2)$, so let $C = 2C''$ and $c = 1/2c''$ to get the desired relation.

- (i) $\Rightarrow$ (ii): Use the relation

$$\mathrm{E}|X|^k = \int_0^\infty k u^{k-1} \Pr(|X| \geq u)\, du \leq kC \int_0^\infty u^{k-1} e^{-cu^2}\, du = \frac{kC}{2} c^{-k/2} \Gamma\left(\frac{k}{2}\right) \tag{334}$$

  The first equality comes from integration by parts of $\mathrm{E}\, X^k = \int x^k \Pr(X \in dx)$. There's a simple inequality $\Gamma(x) \leq x^{x-1}$ for $x > 1$ so for $k > 2$

$$\mathrm{E}|X|^k \leq C \left(\frac{k}{2c}\right)^{k/2} \tag{335}$$

  Choosing $C' \geq \max(C^2, 1)/2c$ and large enough so the inequality is satisfied for $k = 1$ and $k = 2$, we get the desired relation.

- (ii) $\Rightarrow$ (iii) By Stirling's approximation, where exist constants C_1 and C_2 such that for all k, $C_1 \leq k!/k^{k+1/2}e^{-k} \leq C_2$. For example, we can take $C_1 = \sqrt{2\pi}$ and $C_2 = \sqrt{2\pi} e^{\frac{1}{12}}$. Therefore for even k.

$$\frac{k^{k/2}}{k!} \leq \frac{k^{k/2}}{C_1 k^{k+1/2} e^{-k}} = \frac{2^{(k+1)/2} e^{-k/2}}{C_1 (k/2)^{(k+1)/2} e^{-k/2}} \leq \frac{C_2 2^{(k+1)/2} e^{-k/2}}{C_1 (k/2)!} = A \frac{b^{k/2}}{(k/2)!} \tag{336}$$

  where $A = e^{\frac{1}{12}} \sqrt{2}$ and $b = 2/e$ are absolute constants.

  Turning to the moment generating function

$$\begin{aligned}
\mathrm{E}\, e^{tX} &\leq \mathrm{E}\, e^{tX} + e^{-tX} \\
&= \sum_{k \text{ even}} \frac{2t^k\, \mathrm{E}\, X^k}{k!} \leq \sum_{k \text{ even}} \frac{2t^k (C'k)^{k/2}}{k!} \\
&\leq 2A \sum_{k \text{ even}} \frac{(C'bt^2)^{k/2}}{(k/2)!} = 2A \exp(C'bt^2)
\end{aligned} \tag{337}$$

  So the desired relationship holds wtih $C'' = 3A$ and $c'' = C'b$.

∎

4.43 Use the truncation method to show that in order to prove the central limit theorem (theorem 8) it suffices to do so in the case when the underlying random variable X is bounded (and, in particular, subgaussian)

Let X be a random variable with mean 0 and standard deviation σ. For $K > 0$, let $\mu = \mathrm{E}\, X 1_{|X| \leq K}$ and consider $X^{\leq} = X 1_{|X| \leq K} - \mu$ and $X^{>} = X 1_{|X| \leq K} + \mu$ so that $X = X^{\leq} + X^{>}$ and $\mathrm{E}\, X^{\leq} = \mathrm{E}\, X^{>} = 0$.

Approximate continuous compactly supported G pointwise by a smooth function $\tilde{G}$ such that $|G(x) - \tilde{G}(x)| \leq \epsilon$ everywhere (this is possible, e.g., by Stone-Wierestrass). Therefore

$$|\,\mathrm{E}\, G(X) - \mathrm{E}\, \tilde{G}(X)| \leq \epsilon \Pr(X \in \operatorname{supp} G) \leq \epsilon \tag{338}$$

Thus if we can prove the convergence in distribution on smooth, compactly supported G, the result extends to all continuous compactly supported G.

Now let $X_1, X_2, \dots$ be iid copies of X and define $X_k^{\leq}$ and $X_k^{>}$ as above. Suppose Z_k is independent of $X_k^{>}$, then

$$G(Z_k + \frac{1}{\sqrt{n}} X_k^{>}) = G(Z_k) + \frac{1}{\sqrt{n}} G'(Z_k) X_k^{>} + \frac{1}{2n} G''(\xi(Z_k, X_k^{>})) X_k^{>2} \tag{339}$$

where $\xi(Z_k, X_k^{>})$ gives the value of the remainder term. Therefore

$$|\operatorname{E} G(Z_k + \frac{1}{\sqrt{n}} X_k^{>}) - \operatorname{E} G(Z_k)| \leq O(n^{-1} \operatorname{E} X^{>2}) \tag{340}$$

(the implied constant depends on the maximum value of G'', which is bounded because G'' is compactly supported).

TODO finish this argument

■

4.46 (Converse direction of moment continuity theorem) Let X_n be a sequence of uniformly subgaussian random variables (this there exist $C, c > 0$ such that $\Pr(|X_n| \geq t) \leq Ce^{-ct^2}$ for all $t > 0$ and all n, and suppose X_n converges in distribution to a limit X. Show that for any $k = 0, 1, 2, \dots$, $\operatorname{E} X_n^k$ converges pointwise to $\operatorname{E} X^k$.

We will use the dominated convergence theorem with the formula

$$\operatorname{E}|X_n|^k = \int_0^\infty k u^{k-1} \Pr(|X_n| \geq u)\, du \tag{341}$$

The integrands are uniformly bounded by $Cku^{k-1}e^{-cu^2}$, which is integrable on $[0, \infty)$. Except for countably many u (which is a set of measure zero under the Lesbegue measure), the integrand converges pointwise to $ku^{k-1} \Pr(|X| \geq u)$ since X_n converges to X in distribution. Thus

$$\operatorname{E}|X_n|^k \to \int_0^\infty k u^{k-1} \Pr(|X| \geq u)\, du = \operatorname{E}|X|^k \tag{342}$$

■

4.47 (Chernoff bound) Let $X_1, \dots, X_n$ be iid copies of a real random variable X of mean zero and unit variance, which is subgaussian in the sense of exercise 42. Write $S_n := X_1 + \cdots + X_n$

(i) Show that there exist $c'' > 0$ such that $\operatorname{E} e^{tX} \leq e^{c''t^2}$ for all $t \in \mathbb{R}$. Conclude that $\operatorname{E} e^{tS_n/\sqrt{n}} \leq e^{c''t^2}$ for all $t \in \mathbb{R}$

(ii) Conclude the Chernoff bound

$$\Pr\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq \lambda\right) \leq Ce^{-c\lambda^2} \tag{343}$$

for some $C, c > 0$ and all $\lambda > 0$ and all $n \geq 1$

(i) Since X is subgaussian, we know that $\mathrm{E}\, e^{tX} \leq C' e^{c't^2}$, so the key is to show that for mean 0 and unit variance, this inequality actually holds with $C' = 1$ (possibly with a different c'). Without loss of generality, $C' > 1$ since otherwise the inequality already holds with $C' = 1$. Now for $|t| \geq 1$ note that

$$C' e^{c't^2} \leq e^{(c' + \log C')t^2} \tag{344}$$

so the assertion holds for large enough t.

For $|t| < 1$, we will show that

$$e^{tX} = \sum_{k=0}^{\infty} \frac{t^k X^k}{k!} \leq 1 + tX + 3t^2 \left(\sum_{k \text{ even}} \frac{X^k}{k!} \right) \tag{345}$$

Now we justify the inequality on the right. When $X < 0$, this follows because all the odd terms are negative, so we can drop the terms with $k \geq 3$. The remaining terms are all positive, $t^2 \geq t^k$, and we can loosen the inequality by adding some additional terms and increasing some positive constants to get the above. When $X > 0$, the terms are all positive, so we can use $t^2 \geq t^k$, and focus only on $\sum_{k=2}^{\infty} X^k/k!$. Note that by the arithmetic-geometric inequality, we can compare the odd terms to the even terms for $X > 0$,

$$\frac{X^{k+1}}{(k+1)!} + \frac{X^{k-1}}{(k-1)!} \geq \left(2\sqrt{\frac{k}{k+1}} \right) \frac{X^k}{k!} \geq \frac{X^k}{k!} \tag{346}$$

So replacing each odd term with the sum of its neighbors we get

$$\sum_{k=2}^{\infty} \frac{X^k}{k!} \leq 2\frac{X^2}{2} + 3 \sum_{\substack{k \geq 4 \\ k \text{ even}}} \frac{X^k}{k!} \leq 3 \sum_{k \text{ even}} \frac{X^k}{k!} \tag{347}$$

We've shown that for $|t| < 1$,

$$e^{tX} \leq 1 + tX + \frac{3}{2} t^2 (e^{-X} + e^X) \tag{348}$$

Taking expectations, using the fact $\mathrm{E}\, X = 0$ and X is subgaussian, we find

$$\mathrm{E}\, e^{tX} \leq 1 + 3C' e^{c'}\, t^2 \leq \exp(3C' e^{c'}\, t^2) \tag{349}$$

So if we take $c'' = \max(c' + \log C', 3C'e^{c'})$ then we have an inequality $\mathrm{E}\, e^{tX} \leq e^{c''t^2}$ for all t.

The moment generating function $f_X(t) = \mathrm{E}\, e^{tX}$ satisfies many of the same functional relationships as the characteristic function $\phi_X(t)$. In particular, if X and Y are independent, then $f_{X+Y}(t) = f_X(t) f_Y(t)$ and $f_{\alpha X}(t) = f_X(\alpha t)$ for any $\alpha \in \mathbb{R}$. Therefore $f_{S_n/\sqrt{n}}(t) = f_X(t/\sqrt{n})^n$ and for subgaussian X this gives an inequality

$$f_{S_n/\sqrt{n}}(t) \leq e^{nc''(t/\sqrt{n})^2} = e^{c''t^2} \tag{350}$$

(ii) By (i) the moment generating functions of the sums $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ are *uniformly* subgaussian with the same coefficient $c''$for all n. Thus, by 4.42, can translate this into an equivalent set of uniform inequalities, since the constants in any one subgaussian inequality depend absolutely on the constants in other inequalities. Hence, there are some constants $C, c > 0$ independent of n such that

$$\Pr\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq \lambda\right) \leq Ce^{-c\lambda^2} \tag{351}$$

for all $\lambda > 0$.

∎

4.48 (Erdös-Kac theorem) For any natural number $x \geq 100$ let n be a natural number drawn uniformly at random from the natural numbers $\{1, \ldots, x\}$ and let $\omega(n)$ denote the number of distint prime factors of n.

(i) Show that for any $k = 0, 1, 2, \ldots$ one has

$$\mathrm{E}\left(\frac{\omega(n) - \log\log x}{\sqrt{\log\log x}}\right)^k \to 0 \tag{352}$$

as $x \to \infty$ if k is odd and

$$\mathrm{E}\left(\frac{\omega(n) - \log\log x}{\sqrt{\log\log x}}\right)^k \to \frac{k!}{2^{k/2}(k/2)!} \tag{353}$$

as $x \to \infty$ if k is even.

(ii) Establish the Erdös-Kac theorem

$$\frac{1}{x}|\{n \leq x : a \leq \frac{\omega(n) - \log\log x}{\sqrt{\log\log x}} \leq b\}| \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx \tag{354}$$

as $x \to \infty$ for any fixed $a < b$. Informally, the Erdös-Kac theorem asserts that $\omega(n)$ behaves like $\mathcal{N}(\log\log n, \log\log n)$ for "random" n. Note this refines the Hardy-Ramanujan theorem 3.16.

∎

Variants of the Central Limit Theorem

5.5 (Khintchine Inequality)

(i) For any non-negative reals $a_1, \ldots, a_n$ and any $0 < p < \infty$ show that

$$\mathrm{E}\left|\sum_{i=1}^{n} \epsilon_i a_i\right|^p \leq C_p \left(\sum_{i=1}^{n} |a_i|^2\right)^{p/2} \tag{355}$$

for some constant C_p depending only on p. When $p = 2$ show that one can take $C_2 = 1$ and that equality holds.

(ii) With the hypothesis in (i) obtain the matching lower bound

$$\mathrm{E}\left|\sum_{i=1}^{n} \epsilon_i a_i\right|^p \geq c_p \left(\sum_{i=1}^{n} |a_i|^2\right)^{p/2} \tag{356}$$

for some $c_p > 0$ depending only on p.

(iii) For any $0 < p < \infty$ and any functions $f_1, \ldots, f_n \in L^p(X)$ on a measure space $X = (X, \mathcal{X}, \mu)$ show that

$$\mathrm{E} \| \sum_{i=1}^{n} \epsilon_i f_i \|_{L^p(X)}^p \leq C_p \| (\sum_{i=1}^{n} |f_i|^2)^{1/2} \|_{L^p(X)}^p \tag{357}$$

and

$$\mathrm{E} \| \sum_{i=1}^{n} \epsilon_i f_i \|_{L^p(X)}^p \geq c_p \| (\sum_{i=1}^{n} |f_i|^2)^{1/2} \|_{L^p(X)}^p \tag{358}$$

with the same constants c_p, C_P as in (i) and (ii). When $p = 2$ show that one can take $C_2 = c_2 = 1$ and equality holds.

(iv) (Marcinkiewicz-Zygmund theorem) Let X, Y be measure spaces and let $1 < p < \infty$ and suppose $T : L^p(X) \to L^p(Y)$ is a linear operator obeying the bound

$$||Tf||_{L^p(Y)} \leq A||f||_{L^p(X)} \tag{359}$$

for all $f \in L^p(X)$ and some finite A. Show that for any finite sequence $f_1, \ldots, f_n \in L^p(X)$ one has the bound

$$\left\| \left(\sum_{i=1}^{n} |Tf_i|^2\right)^{1/2} \right\|_{L^p(Y)} \leq C'_p A \left\| \left(\sum_{i=1}^{n} |f_i|^2\right)^{1/2} \right\|_{L^p(X)} \tag{360}$$

for some constant C'_p depending only on p.

(v) By using Gaussian sums in place of random signs, show that one can take the constant C'_p in (iv) to be 1.

(i) In the case $p = 2$ we get the elementary calculation

$$\mathrm{E}\left(\sum_{i=1}^{n} \epsilon_i a_i\right)^2 = \sum_{i,j=1}^{n} \mathrm{E}\,\epsilon_i \epsilon_j a_i a_j \tag{361}$$

Considering the terms on the right, if $i \neq j$ then the term is 0, otherwise the term is a_i^2, so the sum is $\sum_{i=1}^{n} a_i^2$ as desired.

Let $S_n = \epsilon_i a_i$. Since the range of $\epsilon_i a_i$ is in $[-a_i, a_i]$, the hypothesis of the Höeffding inequality applies with $\left(\sigma^{(n)}\right)^2 = \sum_{i=1}^{n}(2a_i)^2$ and hence

$$\Pr\left(|S_n| \geq 4\lambda \sum_{i=1}^{n} a_i^2\right) \leq 2\exp(-2\lambda^2) \tag{362}$$

Thus we can calculate

$$\begin{aligned}
\mathrm{E}\,|S_n|^p &= \int_0^\infty p x^{p-1} \Pr(|S_n| \geq x)\, dx \\
&\leq \int_0^\infty 2p x^{p-1} \exp(-x^2/2(a_1^2 + \cdots + a_n^2)^{1/2})\, dx \\
&= p 2^{-p/2} \Gamma\left(\frac{p}{2}\right) \left(\sum_{i=1}^{n} a_i^2\right)^{p/2}
\end{aligned} \tag{363}$$

This is the desired inequality with $C_p = p2^{-p/2}\Gamma(p/2)$

Here's an alternative solution: In the case $p < 2$ we can use Jensen's inequality with the concave downward function $f(x) = |x|^{p/2}$ to find

$$\mathrm{E}(\sum_{i=1}^{n} \epsilon_i a_i)^p \leq \left(\mathrm{E}(\sum_{i=1}^{n} \epsilon_i a_i)^2\right)^{p/2} = (\sum_{i=1}^{n} a_i^2)^{p/2} \tag{364}$$

so the inequality holds with $C_p = 1$.

In the case $p > 2$ it suffices to consider the case when p is an even integer, since for arbitrary p suppose the inequality is true for $p < 2k$ with $k \in \mathbb{N}$. Then let $S = \sum_{i=1}^{n} \epsilon_i a_i$. Using the inequality in 1.40

$$(\mathrm{E}\,|S|^p)^{1/p} \leq (\mathrm{E}\,|S|^{2k})^{1/2k} = C_{2k}^{1/2k} (\sum_{i=1}^{n} a_i^2)^{1/2} \tag{365}$$

Thus we can take $C_p = C_{2k}^{p/2k}$.

When p is an even integer, we get

$$\mathrm{E}(\sum_{i=1}^{n} \epsilon_i a_i)^p = \sum_{i_1,\ldots,i_p=1}^{n} \mathrm{E}\,\epsilon_{i_1} \cdots \epsilon_{i_p} a_{i_1} \cdots a_{i_p} \tag{366}$$

Consider the tuple $(i_1, \ldots, i_p)$. If any value appears an odd number of times, then this term is 0. Thus its possible to pair each index i_k with another index $i_{k'}$ such that $i_k = i_{k'}$. In this case the product $\epsilon_{i_1} \cdots \epsilon_{i_p}$ is identically 1. On the other hand consider the deterministic value

$$(\sum_{j=1}^{n} a_j^2)^{p/2} = \sum_{j_1,\ldots,j_p=1}^{n} a_{j_1}^2 \cdots a_{j_{p/2}}^2 \tag{367}$$

We may map each tuple $(i_1, \ldots, i_p)$ where the indices come in pairs to a tuple $(j_1, \ldots, j_p)$ such that $a_{i_1} \cdots a_{i_p} = a_{j_1}^2 \cdots a_{j_{p/2}}^2$. For example, use a greedy algorithm on the paired index, so let $j_1 = i_1$ and delete the smallest i_k such that $i_k = i_1$. Then repeat for the smallest remaining index. The mapping is clearly onto since $(j_1, j_1, j_2, j_2, \ldots, j_{p/2}, j_{p/2})$ maps to $(j_1, \ldots, j_{p/2})$. Also, a given tuple $(j_1, \ldots, j_{p/2})$ has finitely many pre-images. By considering the mechanism of the greedy algorithm, an upper bound is given by $(p-1)!!$. That is, there are at most $p-1$ indices to associate with i_1, and $p-3$ to associate with the next undeleted index (since we remove 1, the pair of 1, and the next undeleted index), and so on. This bound lets us take $C_p = (p-1)!!$, since the terms of (366) can be collected up to correspond to terms in (367) according to the greedy algorithm, and this gives at most $(p-1)!!$ copies of each term. Hence C_p times (367) bounds (366).

(ii) Note that by Hölder's inequality for $p > 1$, let q satisfy $p^{-1} + q^{-1} = 1$,

$$\begin{aligned} \sum_{i=1}^{n} a_i^2 = \mathrm{E}(\sum_{i=1}^{n} \epsilon_i a_i)^2 &\leq \left(\mathrm{E}\,|\sum_{i=1}^{n} \epsilon_i a_i|^p\right)^{1/p} \left(\mathrm{E}\,|\sum_{i=1}^{n} \epsilon_i a_i|^q\right)^{1/q} \\ &\leq \left(\mathrm{E}\,|\sum_{i=1}^{n} \epsilon_i a_i|^p\right)^{1/p} \left(C_q^{1/q}(\sum_{i=1}^{n} a_i^2)^{1/2}\right) \end{aligned} \tag{368}$$

Thus the desired inequality holds with $c_p = C_q^{-p/q}$. When $p = 1$ the inequality is an equality with $c_p = 1$. TODO the case when $p < 1$.

(iii) For simple functions $f_1, \ldots, f_n$ this follows from parts (i) and (ii) and the linearity of expectation. We can write $f_i(x) = a_{i,j} 1_{A_j}$ where the sets $A_1, \ldots, A_m$ are a disjoint common refinement of all of the measurable sets defining the f_i. Then we have

$$\begin{aligned}
\mathrm{E}\int_X \left|\sum_{i=1}^n \epsilon_i f_i\right|^p &= \mathrm{E}\int_X \left|\sum_{i=1}^n \epsilon_i \sum_{j=1}^m a_{i,j} 1_{A_j}\right|^p \\
&= \sum_{j=1}^m \mu(A_j)\,\mathrm{E}\left|\sum_{i=1}^n \epsilon_i a_{i,j}\right|^p \\
&\le C_p \sum_{j=1}^m \mu(A_j) \left(\sum_{i=1}^n a_{i,j}^2\right)^{p/2} \\
&= C_p \int_X \left|\left(\sum_{i=1}^n f_i^2\right)^{1/2}\right|^p
\end{aligned} \tag{369}$$

This proves the statement for simple functions. For arbtrary functions, approximate $f_1, \ldots, f_n$ by simple functions until each side of the inequality is within ϵ of its true value.

(iv) Stringing together the previous inequalities

$$\begin{aligned}
\left\|\left(\sum_{i=1}^n |Tf_i|^2\right)^{1/2}\right\|^p &\le \frac{1}{c_p}\,\mathrm{E}\left\|\sum_{i=1}^n \epsilon_i T f_i\right\|^p_{L^p(Y)} \\
&= \frac{1}{c_p}\,\mathrm{E}\left\|T\left(\sum_{i=1}^n \epsilon_i f_i\right)\right\|^p_{L^p(Y)} \\
&\le A^p \frac{1}{c_p}\,\mathrm{E}\left\|\sum_{i=1}^n \epsilon_i f_i\right\|^p_{L^p(X)} \\
&\le \frac{A^p C_p}{c_p}\left\|\left(\sum_{i=1}^n |f_i|^2\right)^{1/2}\right\|^p_{L^p(X)}
\end{aligned} \tag{370}$$

Taking pth roots, the desired inequality holds with $C_p' = (C_p/c_p)^{1/p}$

(v) Following the logic above, all we really need to show is that when $\epsilon \sim N(0,1)$ then

$$\mathrm{E}\left|\sum_{i=1}^n \epsilon_i a_i\right| = C_p' \left(\sum_{i=1}^n a_i^2\right)^{p/2} \tag{371}$$

In other words, we have Khintchine-like relationship, but its an equality, so the constants, $c_p = C_p = C_p^{\mathrm{normal}}$, are equal. But this is just a simple property of normal distributions. If $\epsilon_i \sim N(0,1)$ then $\sum_{i=1}^n a_i \epsilon_i \sim N(0, \sum a_i^2)$. The pth moment of $N(0,\sigma^2)$ is $C_p^{\mathrm{normal}}\sigma^p$ where

$$C_p^{\mathrm{normal}} = \mathrm{E}\,N^p = \frac{2^{p/2}}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right) \tag{372}$$

where $N \sim N(0,1)$. Since the analogs of c_p and C_p are equal for Gaussians, the analog of C'_p in part (iv) is 1.

■

5.6 Let $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. copies of a Bernoulli random variable ϵ drawn uniformly from $\{-1, +1\}$.

(i) Show that $\mathrm{E}\, e^{t\epsilon} \leq \exp(t^2/2)$ for any real t

(ii) Show that for any real numbers $a_1, \ldots, a_n$ and any $\lambda > 0$ we have

$$\Pr\left(\left| \sum_{i=1}^{n} \epsilon_i a_i \right| > \lambda (\sum_{i=1}^{n} a_i^2)^{1/2} \right) \leq 2e^{-\lambda^2/2} \tag{373}$$

(i) First note that

$$\mathrm{E}\, e^{t\epsilon} = \frac{1}{2}e^{t} + \frac{1}{2}e^{-t} = \cosh t = \sum_{n \text{ even}} \frac{t^n}{n!} \tag{374}$$

and also

$$e^{t^2/2} = \sum_{m=0}^{\infty} \frac{t^{2m}}{2^m m!} = \sum_{n \text{ even}} \frac{t^n}{n!!} \tag{375}$$

where $n!! = n(n-2)(n-4)\cdots 2$. Since $n!! \leq n!$ its clear that, comparing the expressions term-by-term, $\mathrm{E}\, e^{t\epsilon} \leq e^{-t^2/2}$ as desired.

(ii) Let $S_n = \sum_{i=1}^{n} \epsilon_i a_i$ and let's compute the moment generating function

$$\mathrm{E}\, e^{tS_n} = \prod_{i=1}^{n} \mathrm{E} \exp(t a_i \epsilon_i) \leq \prod_{i=1}^{n} \exp(-t^2 a_i^2/2) = \exp(-\frac{1}{2}t^2 \sum_{i=1}^{n} a_i^2) \tag{376}$$

Let $\sigma_n^2 = \sum_{i=1}^{n} a_i^2$. By Markov's inequality and the symmetry of S_n, for any $t > 0$,

$$\begin{aligned} \Pr(|S_n| > \lambda\sigma_n) &= 2\Pr(e^{tS_n} > e^{t\lambda\sigma_n}) \\ &\leq 2\left(\mathrm{E}\, e^{tS_n}\right)/e^{t\lambda\sigma_n} \\ &\leq 2\exp\left(-t\lambda\sigma_n + t^2\sigma_n^2/2\right) \end{aligned} \tag{377}$$

Minimizing over t to get the tightest inequality gives $t = \lambda/\sigma_n$ in which case the inequality becomes

$$\Pr(|S_n| > \lambda\sigma_n) \leq 2\exp(-\lambda^2/2) \tag{378}$$

as desired.

■

5.10 Establish the conclusion of Theorem 9 directly from explicit computation of the probabilities $\Pr(S_n = k)$ in the case when each $X_{j,n}$ takes values in $\{0, 1\}$ with $\Pr(X_{j,n} = 1) = \lambda/n$ for some fixed $\lambda > 0$.

Using the formula for the binomial distribution

$$\Pr(S_n = k) = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \tag{379}$$

Now for fixed k

$$\frac{n!}{(n-k)!n^k} = \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right) \to 1 \qquad \text{as} \qquad n \to \infty \tag{380}$$

Also

$$\left(1-\frac{\lambda}{n}\right)^{n-k} = \left(1-\frac{\lambda}{n}\right)^{n} \Big/ \left(1-\frac{\lambda}{n}\right)^{k} \to e^{-\lambda}/1 \qquad \text{as} \qquad n \to \infty \tag{381}$$

So we get

$$\Pr(S_n = k) \to e^{-\lambda}\frac{\lambda^k}{k!} \qquad \text{as} \qquad n \to \infty \tag{382}$$

as desired. ∎

5.11 Suppose we replace the hypothesis (iii) in Theorem 9 with the alternative hypothesis

$$\lambda_n := \sum_{j=1}^{n} \Pr(X_{j,n} = 1) \to \infty \text{ as } n \to \infty \tag{383}$$

while leaving hypothesis (i) and (ii) unchanged. Show that $(S_n - \lambda_n)/\sqrt{\lambda_n}$ converges in distribution to the normal distribution $N(0,1)$

As in the proof of theorem 9 let $S_n = X_{1,n} + \cdots + X_{n,n}$ so that

$$\phi_{(S_n-\lambda_n)/\sqrt{\lambda_n}}(t) = e^{-it\sqrt{\lambda_n}}\phi_{S_n/\sqrt{\lambda_n}} = e^{-it\sqrt{\lambda_n}}\prod_{i=1}^{n}(1 - p_{i,n} + p_{i,n}e^{it/\sqrt{\lambda_n}}) \tag{384}$$

Now

$$\begin{aligned} 1 + p_{i,n}(e^{it/\sqrt{\lambda_n}} - 1) &= \exp(p_{i,n}(e^{it/\sqrt{\lambda_n}} - 1)) + O(p_{i,n}^2(e^{it/\sqrt{\lambda_n}} - 1)^2) \\ &= \exp(p_{i,n}(e^{it} - 1)) + O(p_{i,n}^2 t^2/\lambda_n) \end{aligned} \tag{385}$$

Hence

$$\prod_{i=1}^{n}(1 - p_{i,n} + p_{i,n}e^{it/\sqrt{\lambda_n}}) = \exp\left(\sum_{i=1}^{n} p_{i,n}(e^{it/\sqrt{\lambda_n}} - 1)\right) + O\left(\sum_{i=1}^{n} p_{i,n}^2 t/\lambda_n\right) \tag{386}$$

Since $\sum_i p_{i,n}^2/\lambda_n \le (\sup_i p_{i,n})\sum_i p_{i,n}/\lambda_n$, by assumption (ii) this error term tends to 0 pointwise for all t. Finally we calculate

$$\begin{aligned} \phi_{(S_n-\lambda_n)/\sqrt{\lambda_n}}(t) &= \exp(\lambda_n e^{it/\sqrt{\lambda_n}} - \lambda_n - it\sqrt{\lambda_n}) + o(1) \\ &= \exp(-t^2/2 + O(t^3\lambda_n^{-1/2})) + o(1) \\ &\to \exp(-t^2/2) \end{aligned} \tag{387}$$

By Levy's continuity theorem, this shows that the quantity $(S_n - \lambda_n)/\sqrt{\lambda_n}$ converges in distribution to $N(0,1)$. ∎

5.12 For each $\lambda > 0$ let P_λ be a Poisson random variable with intensity λ. Show that as $\lambda \to \infty$ the random variables $(P_\lambda - \lambda)/\sqrt{\lambda}$ converge in distribution to the normal distribution $N(0,1)$. Discuss how this is consistent with Theorem 9 and the previous exercise.

Its a straightforward calculation with characteristic functions. Let $S = (P_\lambda - \lambda)/\sqrt{\lambda}$. Then

$$\begin{aligned}\phi_S(t) &= e^{-it\sqrt{\lambda}}\phi_P(t/\sqrt{\lambda}) \\ &= \exp(-it\sqrt{\lambda} + \lambda(e^{it/\sqrt{\lambda}} - 1)) \\ &= \exp(-t^2/2 + O(t^3/\sqrt{\lambda}))\end{aligned} \tag{388}$$

Thus as $\lambda \to \infty$, $\phi_S(t) = \phi_N(t)$ where $N \sim N(0,1)$. Thus by Levy's continuity theorem, $S \to N$ in distribution.

Note that P_λ is divisible in the sense that $P_{\lambda_1} + P_{\lambda_2} = P_{\lambda_1+\lambda_2}$ and thus we may consider $P_\lambda = \sum_{i=1}^n P^{(i)}_{\lambda/n}$, in which case the $P^{(i)}$ satisfy the assumptions of 5.11. ∎

5.14 Let Y and Y' be non-degenerate real random variables. Suppose that a random variable X lies in the basin of attraction of both Y and Y'. Then there exists $a > 0$ and real b such that $Y' = aY + b$

∎

5.15 Let X lie in the basin of attraction for a non-degenerate law Y

(i) Show that for any iid copies $Y_1, \dots, Y_k$ of Y there exists a $c_k > 0$ and $d_k \in \mathbb{R}$ such that

$$Y_1 + \cdots + Y_k = c_k Y + d_k \tag{389}$$

Also show that

$$c_k Y_1 + c_l Y_2 = c_{k+l} Y + d_{k+l} - d_l \tag{390}$$

for all natural numbers $k, l \in \mathbb{N}$

(ii) Show that the c_k are strictly increasing with $c_{kl} = c_k c_l$ for all $k, l \in \mathbb{N}$. Also show that $d_k l + c_k d_l = d_{kl}$ for all $k, l \in \mathbb{N}$

(iii) Show that there exists $\alpha > 0$ such that $c_k = k^\alpha$ for all k

(iv) If $\alpha = 1$ and Y_1, Y_2 are iid copies of Y show that

$$\frac{k}{k+l} Y_1 + \frac{l}{k+l} Y_2 = Y + \theta_{k,l} \tag{391}$$

for all natural numbers k, l and some bounded real $\theta_{k,l}$. Then show that Y has a stable law in this case.

(v) If $\alpha \neq 1$ show that $d_k = \mu(k - k^\alpha)$ for some real μ and all k. Then show that Y has a stable law in this case.

■

5.16 (Classification of stable laws) Let Y be a non-degenerate stable law, then Y lies in its own basin of attraction and one can then define c_k, d_k, α as in the preceding exercise.

(i) If $\alpha \neq 1$ and μ is as in part (v) of the preceding exercise, show that $\phi_Y(t)^k = \phi_Y(k^\alpha t)e^{i\mu(k-k^\alpha)}$ for all t and k. Then show that

$$\phi_Y(t) = \exp(it\mu - |ct|^\alpha(1 - i\beta\,\mathrm{sgn}(t))) \tag{392}$$

for some real c, β

(ii) Now suppose $\alpha = 1$. Show that $d_{k+l} = d_k + d_l + O(k+l)$ for all k, l (where the implied constant in the $O()$ notation is allowed to depend on Y). Conclude that $d_k = O(k\log k)$ for all k

(iii) We continue to assume $\alpha = 1$. Show that $d_k = -\beta k \log k$ for some real number k. Then use the estimate from part (ii) to show that β does not actually depend on k_0.

(iv) We continue to assume $\alpha = 1$. Show that $\phi_Y(t)^k = \phi(kt)e^{-i\beta k\log k}$ for all t and k. Then show that

$$\phi_Y(t) = \exp(it\mu - |ct|(1 - i\beta\,\mathrm{sgn}(t)\log t)) \tag{393}$$

for all t and some real c.

Its possible to determine which choices of μ, c, α, β are actually achievable by some random variable Y but we will not do so here.

■

5.17 Let X be a real random variable which is symmetric (that is, $X \sim -X$) and obeys the distribution identity

$$\Pr(|X| \geq x) = L(x)\frac{2}{\pi x} \tag{394}$$

for all $x > 0$ which $L : (0, +\infty) \to (0, +\infty)$ is a function which is *slowly varying* in the sense that $L(cx)/L(x) \to 1$ as $x \to \infty$ for all $c > 0$

(i) Show that

$$\mathrm{E}(e^{itX}) = 1 - |t| + o(|t|) \tag{395}$$

as $t \to 0$ where $o(|t|)$ denotes a quantity such that $o(|t|)/|t| \to 0$ as $t \to 0$.

(ii) Let $X_1, X_2, \ldots$ be iid copies of X. Show that $\frac{X_1+\cdots+X_n}{n}$ converge in distribution to a copy of the standard Cauchy distribution (i.e. to a random variable with probability density function $x \mapsto \frac{1}{\pi}\frac{1}{1+x^2}$)

■