MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 1

Readings:
(a) Notes from Lecture 1.
(b) Handout on background material on sets and real analysis (Recitation 1).

Supplementary readings:
[C], Sections 1.1-1.4.
[GS], Sections 1.1-1.3.
[W], Sections 1.0-1.5, 1.9.

Exercise 1.

(a) Let $\mathbb{N}$ be the set of positive integers. A function $f : \mathbb{N} \to \{0, 1\}$ is said to be *periodic* if there exists some N such that $f(n + N) = f(n)$, for all $n \in \mathbb{N}$. Show that the set of periodic functions is countable.

(b) Does the result from part (a) remain valid if we consider rational-valued periodic functions $f : \mathbb{N} \to \mathbb{Q}$?

Solution:

(a) For a given positive integer N, let A_N denote the set of periodic functions with a period of N. For a given N, since the sequence, $f(1), \cdots, f(N)$, actually defines a periodic function in A_N, we have that each A_N contains 2^N elements. For example, for $N = 2$, there are four functions in the set A_2:

$$f(1)f(2)f(3)f(4)\cdots = 0000\cdots; \quad 1111\cdots; \quad 0101\cdots; \quad 1010\cdots.$$

The set of periodic functions from $\mathbb{N}$ to $\{0, 1\}$, A, can be written as,

$$A = \bigcup_{N=1}^{\infty} A_N.$$

Since the union of countably many finite sets is countable, we conclude that the set of periodic functions from $\mathbb{N}$ to $\{0, 1\}$ is countable.

(b) Still, for a given positive integer N, let A_N denote the set of periodic functions with a period N. For a given N, since the sequence, $f(1), \cdots, f(N)$, actually defines a periodic function in A_N, we conclude that A_N has the same cardinality as $\mathbb{Q}^N$ (the Cartesian product of N sets of rational numbers). Since $\mathbb{Q}$ is countable, and the Cartesian product of finitely many countable sets is countable, we know that A_N is countable, for any given N. Since the set of periodic functions from $\mathbb{N}$ to $\mathbb{Q}$ is the union of $A_1, A_2, \cdots$, it is countable, because the union of countably many countable sets is countable.

Exercise 2. Let $\{x_n\}$ and $\{y_n\}$ be real sequences that converge to x and y, respectively. Provide a formal proof of the fact that $x_n + y_n$ converges to $x + y$.

Solution: Fix some $\epsilon > 0$. Let n_1 be such that $|x_n - x| < \epsilon/2$, for all $n > n_1$. Let n_2 be such that $|y_n - y| < \epsilon/2$, for all $n > n_2$. Let $n_0 = \max\{n_1, n_2\}$. Then, for all $n > n_0$, we have

$$|(x_n + y_n) - (x + y)| \leq |x_n - x| + |y_n - y| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which proves the desired result.

Exercise 3. We are given a function $f : A \times B \to \mathbb{R}$, where A and B are nonempty sets.

(a) Assuming that the sets A and B are finite, show that

$$\max_{x \in A} \min_{y \in B} f(x, y) \leq \min_{y \in B} \max_{x \in A} f(x, y).$$

(b) For general nonempty sets (not necessarily finite), show that

$$\sup_{x \in A} \inf_{y \in B} f(x, y) \leq \inf_{y \in B} \sup_{x \in A} f(x, y).$$

Solution:

(a) The proof rests on the application of the following simple fact: if $h(z) \leq g(z)$ for all z in some finite set Z, then

$$\min_{z \in Z} h(z) \leq \min_{z \in Z} g(z) \tag{1}$$

$$\max_{z \in Z} h(z) \leq \max_{z \in Z} g(z). \tag{2}$$

Observe that for all x, y,

$$f(x, y) \leq \max_{x \in A} f(x, y),$$

and Eq. (1) implies that for each x,

$$\min_{y\in B} f(x,y) \le \min_{y\in B}\max_{x\in A} f(x,y).$$

Now applying Eq. (2), let's take a maximum of both sides with respect to $x \in A$. Since the right-hand side is a number, it remains unchanged:

$$\max_{x\in A}\min_{y\in B} f(x,y) \le \min_{y\in B}\max_{x\in A} f(x,y),$$

which is what we needed to show.

(b) Along the same lines, we have the fact that if $h(z) \le g(z)$ for all $z \in Z$,

$$\inf_{z\in Z} h(z) \le \inf_{z\in Z} g(z) \tag{3}$$

$$\sup_{z\in Z} h(z) \le \sup_{z\in Z} g(z). \tag{4}$$

These follow immediately from the definitions of $\sup$ and $\inf$.

As before, we begin with

$$f(x,y) \le \sup_{x\in A} f(x,y),$$

for all x, y. By Eq. (3), for each x,

$$\inf_{y\in B} f(x,y) \le \inf_{y\in B}\sup_{x\in A} f(x,y),$$

and using Eq. (4),

$$\sup_{x\in A}\inf_{y\in B} f(x,y) \le \inf_{y\in B}\sup_{x\in A} f(x,y).$$

Exercise 4. A probabilistic experiment involves an infinite sequence of trials. For $k = 1, 2, \ldots$, let A_k be the event that the kth trial was a success. Write down a set-theoretic expression that describes the following event:

B: For every k there exists an ℓ such that trials $k\ell$ and $k\ell^2$ were both successes.

Note: A "set theoretic expression" is an expression like $\bigcup_{k>5}\bigcap_{\ell<k} A_{k+\ell}$.

Solution: $B = \bigcap_{k=1}^{\infty}\bigcup_{\ell=1}^{\infty}(A_{k\ell}\cap A_{k\ell^2})$.

Exercise 5. Let $f_n, f, g : [0,1] \to [0,1]$ and $a, b, c, d \in [0,1]$. Derive the following set theoretic expressions:

(a) Show that

$$\{x \in [0,1] \mid \sup_n f_n(x) \leq a\} = \bigcap_n \{x \in [0,1] \mid f_n(x) \leq a\},$$

and use this to express $\{x \in [0,1] \mid \sup_n f_n(x) < a\}$ as a countable combination (countable unions, countable intersections and complements) of sets of the form $\{x \in [0,1] \mid f_n(x) \leq b\}$.

(b) Express $\{x \in [0,1] \mid f(x) > g(x)\}$ as a countable combination of sets of the form $\{x \in [0,1] \mid f(x) > c\}$ and $\{x \in [0,1] \mid g(x) < d\}$.

(c) Express $\{x \in [0,1] \mid \limsup_n f_n(x) \leq c\}$ as a countable combination of sets of the form $\{x \in [0,1] \mid f_n(x) \leq c\}$.

(d) Express $\{x \in [0,1] \mid \lim_n f_n(x) \text{ exists}\}$ as a countable combination of sets of the form $\{x \in [0,1] \mid f_n(x) < c\}$, $\{x \in [0,1] \mid f_n(x) > c\}$, etc. (Hint: think of $\{x \in [0,1] \mid \limsup_n f_n(x) > \liminf_n f_n(x)\}$).

Solution: First observe the following set relations

$$[0,c) = \bigcup_{n=1}^{\infty} [0, c - \tfrac{1}{n}] \quad [0,c] = \bigcap_{n=1}^{\infty} [0, c + \tfrac{1}{n})$$
$$(c,1] = \bigcup_{n=1}^{\infty} [c + \tfrac{1}{n}, 1] \quad [c,1] = \bigcap_{n=1}^{\infty} (c - \tfrac{1}{n}, 1].$$

All conversions between strict and non-strict inequalities following from these relations and properties of the inverse image, i.e. homomorphism of arbitrary set operations. We will use the shorthand notation

$$\{f < a\} := \{x \in [0,1] \mid f(x) < a\}.$$

(a) Let $x \in \bigcap_n \{f_n \leq a\}$. Then, $f_n(x) \leq a$ for all $n \Longrightarrow \sup_n f_n(x) \leq a$, by definition of $\sup$ as a is an upper bound for $\{f_n(x)\}$. Therefore, as x was arbitrary,

$$\bigcap_{n=1}^{\infty} \{f_n \leq a\} \subset \{\sup_n f_n \leq a\}.$$

Let $x \in \{\sup_n f_n \leq a\}$. Then $\sup_n f_n(x) \leq a$ and for all n $f_n(x) \leq \sup_n(x) \leq a$. Therefore, as x was arbitrary,

$$\{\sup_n f_n \leq a\} \subset \mathop{}_{n=1}^{\infty} \{f_n \leq a\}.$$

Hence $\{\sup_n f_n \leq a\} = \bigcap_n \{f_n \leq a\}$. By De Morgan's this relation also implies

$$\{\sup_n f_n > a\} = \bigcup_{n=1}^{\infty} \{f_n \geq a\}.$$

Similar results hold for $\inf$.

Let $f = \sup_n f_n$. Using the above comment

$$\begin{aligned}
\{\sup_n f_n < a\} &= \{f < a\} \\
&= f^{-1}([0, a)) \\
&= \bigcup_{k=1}^{\infty} f^{-1} \;\; [0, a - \tfrac{1}{k}] \\
&= \bigcup_{k=1}^{\infty} \{\sup_n f_n \leq a - \tfrac{1}{k}\} \\
&= \bigcup_{k=1}^{\infty} \mathop{}_{n=1}^{\infty} \{f_n \leq a - \tfrac{1}{k}\}.
\end{aligned}$$

(b) Using countability and density of the rationals

$$\begin{aligned}
\{f > g\} &= \bigcup_{q \in \mathbb{Q}} \{f > q\} \cap \{q > g\} \\
&= \bigcup_{q \in \mathbb{Q}} \{f > q\} \cap \{q \geq g\} \\
&= \bigcup_{q \in \mathbb{Q}} \{f \geq q\} \cap \{q > g\}.
\end{aligned}$$

(c)

$$
\begin{aligned}
\{\limsup_{n\to\infty} f_n \le c\} &= \{\inf_{n\ge 1}\sup_{k\ge n} f_k \le c\} \\
&= \mathop{}_{m=1}^{\infty} \{\inf_{n\ge 1}\sup_{k\ge n} f_k < c + \tfrac{1}{m}\} \\
&= \mathop{}_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \{\sup_{k\ge n} f_k < c + \tfrac{1}{m}\} \\
&= \mathop{}_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcup_{\ell=1}^{\infty} \{\sup_{k\ge n} f_k \le c + \tfrac{1}{m} - \tfrac{1}{\ell}\} \\
&= \mathop{}_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcup_{\ell=1}^{\infty} \mathop{}_{k=n}^{\infty} \{f_k \le c + \tfrac{1}{m} - \tfrac{1}{\ell}\}.
\end{aligned}
$$

(d)

$$
\begin{aligned}
\{\lim_{n\to\infty} f_n \ \text{ exists}\} &= \{\liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n\} \\
&= \{\liminf_{n\to\infty} f_n < \limsup_{n\to\infty} f_n\}^c \quad (\liminf_{n\to\infty} f_n(x) \le \limsup_{n\to\infty} f_n(x)) \\
&= \left(\bigcup_{q\in\mathbb{Q}} \{\liminf_{n\to\infty} f_n < q\} \cap \{q < \limsup_{n\to\infty} f_n\}\right)^c \qquad (\text{part } b) \\
&= \mathop{}_{q\in\mathbb{Q}} \{\liminf_{n\to\infty} f_n \ge q\} \cup \{\limsup_{n\to\infty} f_n \le q\}.
\end{aligned}
$$

The sets $\{\liminf_{n\to\infty} f_n \ge q\}$ and $\{\limsup_{n\to\infty} f_n \le q\}$ can be expressed as countable combinations using part (c) and the fact that

$$
\begin{aligned}
-\limsup_{n\to\infty} f_n(x) &= -\inf_{n\ge 1}\sup_{k\ge n} f_k(x) \\
&= \sup_{n\ge 1}\inf_{k\ge n} (-f_k(x)) \\
&= \liminf_{n\to\infty} (-f_n(x)),
\end{aligned}
$$

i.e. $\{\liminf_{n\to\infty} f_n \ge q\} = \{\limsup_{n\to\infty}(-f_n) \le -q\}$. More specifically,

$$
\bigcap_{q\in\mathbb{Q}} \left[\left(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcup_{\ell=1}^{\infty} \bigcap_{k=n}^{\infty} \{f_k \ge c - \tfrac{1}{m} + \tfrac{1}{\ell}\}\right) \cup \left(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcup_{\ell=1}^{\infty} \bigcap_{k=n}^{\infty} \{f_k \le c + \tfrac{1}{m} - \tfrac{1}{\ell}\}\right)\right].
$$

Using one of the later two expressions of part (b), we can drop one of the outer intersections

$$\bigcap_{q\in\mathbb{Q}}\left[\left(\bigcup_{n=1}^{\infty}\bigcup_{\ell=1}^{\infty}\bigcap_{k=n}^{\infty}\{f_k \geq c+\tfrac{1}{\ell}\}\right)\cup\left(\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcup_{\ell=1}^{\infty}\bigcap_{k=n}^{\infty}\{f_k \leq c+\tfrac{1}{m}-\tfrac{1}{\ell}\}\right)\right]$$

or

$$\bigcap_{q\in\mathbb{Q}}\left[\left(\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcup_{\ell=1}^{\infty}\bigcap_{k=n}^{\infty}\{f_k \geq c-\tfrac{1}{m}+\tfrac{1}{\ell}\}\right)\cup\left(\bigcup_{n=1}^{\infty}\bigcup_{\ell=1}^{\infty}\bigcap_{k=n}^{\infty}\{f_k \leq c-\tfrac{1}{\ell}\}\right)\right].$$

Exercise 6. Let $\Omega = \mathbb{N}$ (the positive integers), and let $\mathcal{F}_0$ be the collection of subsets of Ω that either have finite cardinality or their complement has finite cardinality. For any $A \in \mathcal{F}_0$, let $\mathbb{P}(A) = 0$ if A is finite, and $\mathbb{P}(A) = 1$ if A^C is finite.

(a) Show that $\mathcal{F}_0$ is a field but not a σ-field.

(b) Show that $\mathbb{P}$ is finitely additive on $\mathcal{F}_0$; that is, if $A, B \in \mathcal{F}_0$, and A, B are disjoint, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

(c) Show that $\mathbb{P}$ is not countably additive on $\mathcal{F}_0$; that is, construct a sequence of disjoint sets $A_i \in \mathcal{F}_0$ such that $\cup_{i=1}^{\infty} A_i \in \mathcal{F}_0$ and $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) \neq \quad {}_{i=1}^{\infty}\, \mathbb{P}\left(A_i\right)$.

(d) Construct a decreasing sequence of sets $A_i \in \mathcal{F}_0$ such that $\cap_{i=1}^{\infty} A_i = \emptyset$ for which $\lim_{i\to\infty} \mathbb{P}(A_i) \neq 0$.

Solution:

(a) The empty set has zero cardinality, and therefore belongs to $\mathcal{F}_0$. Furthermore, if $A \in \mathcal{F}_0$, then either A or A^c has finite cardinality. It follows that either A^c or $(A^c)^c$ has finite cardinality, so that $A^c \in \mathcal{F}_0$.

Suppose that $A, B \in \mathcal{F}_0$. If both A and B are finite, then $A \cup B$ is also finite and belongs to $\mathcal{F}_0$. Suppose now that at least one of A or B is infinite. We have $A \cup B = (A^c \cap B^c)^c$. Since $A^c \cap B^c$ is finite, it follows that $A \cup B \in \mathcal{F}_0$. This shows that $\mathcal{F}_0$ is a field.

To see that $\mathcal{F}_0$ is not a σ-field, note that $\{2n\} \in \mathcal{F}_0$ for every $n \in \mathbb{N}$, but the set $\bigcup_{n=0}^{\infty}\{2n\}$, the set of even natural numbers, is not in $\mathcal{F}_0$.

(b) Let $A, B \in \mathcal{F}_0$ be disjoint. If both A and B are finite, then $\mathbb{P}\left(A \cup B\right) = 0 = \mathbb{P}(A) + \mathbb{P}(B)$. Suppose that either A or B (or both) is infinite. Since A and B are disjoint, we have $A \subset B^c$ and $B \subset A^c$. It follows that A and B cannot both be infinite. Therefore, $\mathbb{P}(A \cup B) = 1 = \mathbb{P}(A) + \mathbb{P}(B)$, and $\mathbb{P}$ is finitely additive.

(c) Note that $\{n\} \in \mathcal{F}_0$ and $\bigcup_{n\geq 1}\{n\} = \Omega$. However, $\mathbb{P}(\{n\}) = 0$ while $\mathbb{P}(\Omega) = 1$, hence $\mathbb{P}$ is not countably additive.

(d) Let $A_n = \{n, n+1, \ldots\}$. Then $(A_n)_{n\geq 1}$ forms a decreasing sequence of sets with $\bigcap_n A_n = \emptyset$. But $\mathbb{P}(A_n) = 1$ for all n, hence $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 2

Readings:

Notes from Lecture 2 and 3.

Supplementary readings:

[GS], Sections 1.4-1.7.

[C], Chapter 1.3

[W], Chapter 1.

Exercise 1. Consider a probabilistic experiment involving infinitely many coin tosses, and let $\Omega = \{0, 1\}^\infty$ (think of 0 and 1 corresponding to heads and tails, respectively). A typical element $\omega \in \Omega$ is of the form $\omega = (\omega_1, \omega_2, \ldots)$, with $\omega_i \in \{0, 1\}$.

As in the notes for Lecture 2, we define $\mathcal{F}_n$ as the σ-field consisting of all sets whose occurrence or nonoccurrence can be determined by looking at the result of the first n coin flips. The σ-field $\mathcal{F}$ for this model is defined as the smallest σ-field that contains all of the $\mathcal{F}_n$.

(a) Consider the event H consisting of all ω with the following property. There exists some time t at which the number of ones so far is greater than or equal to the number of zeros so far. Show that $H \in \mathcal{F}$.

(b) (Harder) Consider the set A of all ω for which the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega_i$$

exists. Show that $A \in \mathcal{F}$.

Note: This is important because, once we have also chosen a probability measure, it allows us to make statements about the probability that this limit (the long-term fraction of heads) exists.

Hint: The event A_x "the limit defined above exists and is equal to x" belongs to $\mathcal{F}$. However, this does not imply that $\bigcup_x A_x \in \mathcal{F}$ (why?). You need to find some other way of describing the event A in terms of unions, complements, etc., of events in the $\mathcal{F}_n$. For example, use the fact that a sequence converges if and only if it is a "Cauchy sequence."

Solution:

(a) Let $S_n = \{(\omega_1, \omega_2, \ldots) \mid \ {}_{i=1}^{n} \omega_i \geq \lceil n/2 \rceil\}$, i.e., S_n is the set of sequences where there are at least as many ones, in the first n entries as there are zeroes. Then,

$$H = \bigcup_{n=1}^{\infty} S_n.$$

(b) Let

$$a_n = \frac{1}{n} \sum_{i=1}^{n} \omega_i.$$

According to Cauchy criterion, the sequence $\{a_n\}$ converges if and only if for any positive integer r, there exists some positive integer N such that for any $n > m > N$,

$$|a_n - a_m| < 1/r.$$

For a pair of positive integers $n > m$, we define

$$A_{1/r,n,m} = \left\{ \omega : \ \frac{1}{n} \sum_{i=1}^{n} \omega_i - \frac{1}{m} \sum_{i=1}^{m} \omega_i \ < 1/r \right\} \in \mathcal{F}_n.$$

Thus,

$$A = \bigcap_{r=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{m=N}^{\infty} \bigcap_{n=m}^{\infty} A_{1/r,n,m} \in \mathcal{F}.$$

Exercise 2. Suppose that the events A_n satisfy $\mathbb{P}(A_n) \to 0$ and $\ {}_{n=1}^{\infty} \mathbb{P}(A_n^c \cap A_{n+1}) < \infty$. Show that $\mathbb{P}(A_n \text{ i.o.}) = 0$. *Note:* A_n i.o., stands for "A_n occurs infinitely often", or "infinitely many of the A_n occur", or just $\limsup_n A_n$. *Hint:* Borel-Cantelli.

Solution: Define the set

$$A = \limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

We wish to show $\mathbb{P}(A) = 0$. Now, $A \subseteq \cup_{m=n}^{\infty} A_m$ for all m, and by monotonicity of the measure, $\mathbb{P}(A) \leq \mathbb{P}(\cup_{m=n}^{\infty} A_m)$, for all n. In addition,

$$\begin{aligned} \bigcup_{m=n}^{\infty} A_m &= A_n \cup (A_{n+1} \setminus A_n) \cup (A_{n+2} \setminus A_{n+1}) \cup \cdots \\ &= A_n \cup (A_{n+1} \cap A_n^c) \cup (A_{n+2} \cap A_{n+1}^c) \cup \cdots . \end{aligned}$$

Therefore, by the union bound,

$$
\begin{aligned}
\mathbb{P}(A) &\leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \\
&\leq \mathbb{P}(A_n) + \sum_{m=n}^{\infty} \mathbb{P}(A_{m+1} \cap A_n^c).
\end{aligned}
$$

This holds for all n, and therefore it holds in the limit as n goes to infinity. But the limit of the final expression is zero, since $\mathbb{P}(A_n) \to 0$, and since $\sum_{n=1}^{\infty} \mathbb{P}(A_n^c \cap A_{n+1}) < \infty$.

Exercise 3. Consider one of our standard probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$, with $\Omega = (0,1]$, $\mathcal{F}$ – Borel and $\mathbb{P}$ – the Lebesgue measure. To every element $\omega \in \Omega$ we assign its infinite decimal representation. We disallow decimal representations that end with an infinite string of nines. Under this condition, every number has a unique decimal representation.

(a) Let A be the set of points in $(0,1]$ whose decimal representation contains at least one digit equal to 9. Find $\mathbb{P}[A]$.

(b) Let B be the set of points that have infinitely many 9's in the decimal representation. Find $\mathbb{P}[B]$. (Hint: Borel-Cantelli).

Solution: Part (a).

We will find the Lebesgue measure of A^c, the set of points in $(0,1]$ whose decimal representation contains no digit equal to 9. We can scale that set (by multiplying it with a real number) to obtain the set

$$A_0 = \frac{1}{10} A^c,$$

which is the set of points in $(0,1]$ whose decimal representation starts with a 0, and contains no digit equal to 9 afterwards. Since the set A_0 is just the same as A^c but scaled down by a factor of 10, we have that $\mathbb{P}(A_0) = \frac{1}{10}\mathbb{P}(A^c)$. Furthermore, we can do translations of that set to obtain analogous sets starting with different digits. In particular, let us define

$$A_k = \frac{k}{10} + A_0$$

as the set of points in $(0, 1]$ whose decimal representation starts with a k, and has no digit equal to 9 afterwards. Note that these sets are all disjoint, and that we have

$$A^c = \bigcup_{k=0}^{8} A_k.$$

Then, using the finite additivity property of measures, and the fact that the Lebesgue measure is invariant by translations, we obtain

$$\begin{aligned}
\mathbb{P}(A^c) &= \mathbb{P}\left(\bigcup_{k=0}^{8} A_k\right) \\
&= \sum_{k=0}^{8} \mathbb{P}(A_k) \\
&= \sum_{k=0}^{8} \mathbb{P}(A_0) \\
&= \sum_{k=0}^{8} \frac{1}{10}\mathbb{P}(A^c) \\
&= \frac{9}{10}\mathbb{P}(A^c).
\end{aligned}$$

This equality can only be true if $\mathbb{P}(A^c) = 0$, and thus $\mathbb{P}(A) = 1$.

Part (b). Let B_i be the event that there is a 9 in the i-th position of the expansion. These events are independent with $\mathbb{P}(B_i) = 1/10$, for all $i \geq 1$. Thus, we have

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i) = \infty.$$

Then, by Borel-Cantelli, we have

$$\mathbb{P}(B) = \mathbb{P}(\{B_i \ i.o.\}) = 1.$$

Exercise 4. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let A be an event (element of $\mathcal{F}$). Let $\mathcal{G}$ be collection of all events that are independent from A. Show that $\mathcal{G}$ need not be a σ-algebra.

Solution: $\mathcal{G}$ need not be a σ-algebra. For example, let X, Y be i.i.d., with $\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = 1/2$. Let Z be the mod two sum of X and Y, so that if $X = Y$, then $Z = 0$, and if $X \neq Y$, then $Z = 1$. Then pairwise, these three random variables are independent. Let A be the event $\{Z = 1\}$. Now, the events $B_1 = \{X = 1\}$, $B_2 = \{Y = 1\}$ are both independent of A. However, $B_1 \cap B_2$ is not independent of A.

Exercise 5. Let $A_1, A_2, \ldots$ and B be events.

(a) Suppose that $A_k \searrow A$, i.e. $A_k \supset A_{k+1}$ and $A = \cap_{k=1}^{\infty} A_k$. Assume B is independent of A_k. Show that B is independent of A.

(b) Suppose that A_1 is independent of B and also that A_2 is independent of B. Is it true that $A_1 \cap A_2$ is independent of B? Prove or give a counterexample.

Solution:

(a) The sequence of events $A_k \cap B$ is decreasing and converges to the event $A \cap B$. [To see this, note that $(\cap_{k\geq 1} A_k) \cap B = \cap_{k\geq 1}(A_k \cap B)$.] Using the continuity of probability measures in the first and last equalities below, and independence in the middle equality, we have

$$\mathbb{P}(A \cap B) = \lim_{k\to\infty} \mathbb{P}(A_k \cap B) = \lim_{k\to\infty} \mathbb{P}(A_k)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B).$$

(b) Consider two independent and fair coin tosses and let A_i be the event that the ith toss results in heads. Let B be the event that both tosses give the same result. It is easily checked that $\mathbb{P}(A_i \cap B) = \mathbb{P}(\{HH\}) = 1/4 = \mathbb{P}(A_i)\mathbb{P}(B)$, so that pairwise independence holds. On the other hand, $\mathbb{P}(B \mid A_1 \cap A_2) = 1 \neq \mathbb{P}(B)$. Thus, $A_1 \cap A_2$ and B are not independent.

Exercise 6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Show that function

$$d(A, B) \triangleq \mathbb{P}[A \triangle B]$$

satisfies the triangle inequality (i.e. $d(A, B) \leq d(A, C) + d(C, B)$ for any A, B, C).

Fun fact: Under this pseudo-metric any algebra is dense in the σ-algebra it generates. Thus, any event in a complicated σ-algebra (such as Borel) can be approximated arbitrarily well by events in a simple algebra (like finite unions of $[a, b)$).

Solution: The symmetric difference is $A \triangle B = (A \backslash B) \cup (B \backslash A)$

$$\begin{aligned}
A \triangle B &= (A \backslash B) \cup (B \backslash A) \\
&= (A \cap B^c) \cup (B \cap A^c) \\
&= (A \cap B^c \cap C) \cup (A \cap B^c \cap C^c) \cup (B \cap A^c \cap C) \cup (B \cap A^c \cap C^c) \\
&\subset (C \backslash B) \cup (A \backslash C) \cup (C \backslash A) \cup (B \backslash C) \\
&= (A \triangle C) \cup (C \triangle B).
\end{aligned}$$

Hence, by the union bound,

$$\mathbb{P}(A \triangle B) \leq \mathbb{P}(A \triangle C) + \mathbb{P}(C \triangle B).$$

Exercise 7. ***[Optional, not to be graded]*** Let $\Omega_1 \subset \Omega$ and let $\mathcal{C}$ be some collection of subsets of Ω. Let

$$\mathcal{C}_1 = \mathcal{C} \cap \Omega_1 \triangleq \{A \cap \Omega_1 : A \in \mathcal{C}\}$$

and denote by $\mathcal{F}_1$ ($\mathcal{F}$) the minimal σ-algebra on Ω_1 (Ω) generated by $\mathcal{C}_1$ ($\mathcal{C}$). Also define

$$\mathcal{F}_2 = \mathcal{F} \cap \Omega_1 \triangleq \{A \cap \Omega_1 : A \in \mathcal{F}\}.$$

$\mathcal{F}_2$ is called *a trace* of $\mathcal{F}$ on Ω_1. Show $\mathcal{F}_1 = \mathcal{F}_2$. (*Hint:* show that collection $\mathcal{G} = \{E \in \mathcal{F} : E \cap \Omega_1 \in \mathcal{F}_1\}$ is a monotone class.)

Solution: For a collection $\mathcal{D}$ and a space Ω let $\alpha_\Omega(\mathcal{D})$ denote the smallest algebra of sets in Ω containing $\mathcal{D}$.

Claim: $\alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1) = \alpha_\Omega(\mathcal{C}) \cap \Omega_1$.

By definition $\mathcal{C} \subset \alpha_\Omega(\mathcal{C})$ and therefore, $\mathcal{C} \cap \Omega_1 \subset \alpha_\Omega(\mathcal{C}) \cap \Omega_1$. The empty set $\phi = \phi \cap \Omega_1 \in \alpha_\Omega(\mathcal{C}) \cap \Omega_1$, as $\alpha_\Omega(\mathcal{C})$ is an algebra. Let $E \cap \Omega_1 \in \alpha_\Omega(\mathcal{C}) \cap \Omega_1$, then $(E \cap \Omega_1)^c = \Omega_1 \backslash (E \cap \Omega_1) = E^c \cap \Omega_1 \in \alpha_\Omega(\mathcal{C}) \cap \Omega_1$, as $E \in \alpha_\Omega(\mathcal{C})$ and $\alpha_\Omega(\mathcal{C})$ is an algebra. Let $E_1 \cap \Omega_1, E_2 \cap \Omega_1 \in \alpha_\Omega(\mathcal{C}) \cap \Omega_1$, then $(E_1 \cap \Omega_1) \cap (E_2 \cap \Omega_2) = (E_1 \cap E_2) \cap \Omega_1 \in \alpha_\Omega(\mathcal{C}) \cap \Omega_1$, as $E_1, E_2 \in \alpha_\Omega(\mathcal{C})$ and $\alpha_\Omega(\mathcal{C})$ is an algebra. Hence $\alpha_\Omega(\mathcal{C})$ is an algebra of sets in Ω_1 containing $\mathcal{C} \cap \Omega_1$, and by minimality of $\alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$, $\alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1) \subset \alpha_\Omega(\mathcal{C}) \cap \Omega_1$.

Consider the set

$$\mathcal{D}_1 = \{E \in 2^\Omega \mid E \cap \Omega_1 \in \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)\}.$$

The collection $\mathcal{C} \subset \mathcal{D}_1$, as $\mathcal{C} \cap \Omega_1 \subset \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ by definition. The empty set $\phi \cap \Omega_1 = \phi \in \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$, as α_{Ω_1} is an algebra. Thus $\phi \in \mathcal{D}_1$. Let $E \in \mathcal{D}_1$, then $E^c \cap \Omega_1 = \Omega_1 \backslash (E \cap \Omega_1) = (E \cap \Omega_1)^c \in \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$, as $E \cap \Omega_1 \in \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ and $\alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ is an algebra. Thus $\mathcal{D}_1$ is closed under complements. Let $E_1, E_2 \in \mathcal{D}_1$, then $(E_1 \cap E_2) \cap \Omega_1 = (E_1 \cap \Omega_1) \cap (E_2 \cap \Omega_1) \in \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$, as $E_1 \cap \Omega_1, E_2 \cap \Omega_1 \in \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ and $\alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ is an algebra. Thus $\mathcal{D}_1$ is closed under intersections and $\mathcal{D}_1$ is an algebra of sets in Ω containing $\mathcal{C}$. Therefore, by minimality $\alpha_\Omega(\mathcal{C}) \subset \mathcal{D}_1$. By definition of $\mathcal{D}_1$, $\alpha_\Omega(\mathcal{C}) \cap \Omega_1 \subset \alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)$, which proves the claim.

Claim: For a collection of sets $\mathcal{D}$ and a space Ω, $\sigma_\Omega(\mathcal{D}) = \sigma_\Omega(\alpha_\Omega(\mathcal{D}))$.

By definition $\mathcal{D} \subset \alpha_\Omega(\mathcal{D}) \subset \sigma_\Omega(\mathcal{D})$, and by monotonicity of the $\sigma_\Omega(\cdot)$ operator, see recitation 2, $\sigma_\Omega(\mathcal{D}) \subset \sigma_\Omega(\alpha_\Omega(\mathcal{D})) \subset \sigma_\Omega(\sigma_\Omega(\mathcal{D})) = \sigma_\Omega(\mathcal{D})$. Thus $\sigma_\Omega(\mathcal{D}) = \sigma_\Omega(\alpha_\Omega(\mathcal{D}))$.

Combining the results of the two claims $\sigma_{\Omega_1}(\alpha_\Omega(\mathcal{C}) \cap \Omega_1) = \sigma_{\Omega_1}(\alpha_{\Omega_1}(\mathcal{C} \cap \Omega_1)) = \sigma_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ and $\sigma_\Omega(\alpha_\Omega(\mathcal{C})) = \sigma_\Omega(\mathcal{C})$. Therefore, it suffices to show that $\sigma_{\Omega_1}(\alpha_\Omega(\mathcal{C}) \cap \Omega_1) = \sigma_\Omega(\alpha_\Omega(\mathcal{C})) \cap \Omega_1$. By the monotone class theorem, as $\alpha_\Omega(\mathcal{C})$ is an algebra, this holds if and only if $\mu_{\Omega_1}(\alpha_\Omega(\mathcal{C}) \cap \Omega_1) = \mu_\Omega(\alpha_\Omega(\mathcal{C})) \cap \Omega_1$. Let $\mathcal{A} := \alpha_\Omega(\mathcal{C})$.

By definition $\mathcal{A} \subset \mu_\Omega(\mathcal{A})$ and therefore, $\mathcal{A} \cap \Omega_1 \subset \mu_\Omega(\mathcal{A}) \cap \Omega_1$. Let $\{E_n \cap \Omega_1\} \in \mu_\Omega(\mathcal{A}) \cap \Omega_1$, with $(E_n \cap \Omega_1) \subset (E_{n+1} \cap \Omega_1)$. The sequence $\{E_n\}$ may not be monotone however, $E'_n = \cup_{k=1}^n E_n$ is monotonic and by the monotonicity of $\{E_n \cap \Omega_1\}$, $(\cup_{k=1}^n E_k) \cap \Omega_1 = E_n \cap \Omega_1$, i.e. $E_n \cap \Omega_1 = E'_n \cap \Omega_1$. Since $\mu_\Omega(\mathcal{A})$ is a monotone class $E'_n \nearrow E \in \mu_\Omega(\mathcal{A})$. Therefore, $E_n \cap \Omega_1 \nearrow E \cap \Omega_1 \in \mu_\Omega(\mathcal{A}) \cap \Omega_1$, this follows since $\bigcup_{n=1}^\infty (E_n \cap \Omega_1) = (\bigcup_{n=1}^\infty E_n) \cap \Omega_1 = E \cap \Omega_1$. Similarly, let $\{E_n \cap \Omega_1\} \in \mu_\Omega(\mathcal{A}) \cap \Omega_1$, with $(E_n \cap \Omega_1) \supset (E_{n+1} \cap \Omega_1)$, and, by the construction given for increasing sets, WLOG $E_n \supset E_{n+1}$. Since $\mu_\Omega(\mathcal{A})$ is a monotone class $E_n \searrow E \in \mu_\Omega(\mathcal{A})$. Therefore, $E_n \cap \Omega_1 \searrow E \cap \Omega_1 \in \mu_\Omega(\mathcal{A}) \cap \Omega_1$, this follows since ${}_{n=1}^\infty (E_n \cap \Omega_1) = ({}_{n=1}^\infty E_n) \cap \Omega_1 = E \cap \Omega_1$. Hence $\mu_\Omega(\mathcal{A}) \cap \Omega_1$ is a monotone class of sets in Ω_1 containing $\mathcal{A} \cap \Omega_1$ and by minimality $\mu_{\Omega_1}(\mathcal{A} \cap \Omega_1) \subset \mu_\Omega(\mathcal{A}) \cap \Omega_1$.

Consider the set

$$\mathcal{D}_2 = \{E \in 2^\Omega \mid E \cap \Omega_1 \in \mu_{\Omega_1}(\mathcal{A} \cap \Omega_1)\}.$$

The algebra $\mathcal{A} \subset \mathcal{D}_2$, as $\mathcal{A} \cap \Omega_1 \subset \alpha_{\Omega_1}(\mathcal{A} \cap \Omega_1)$ by definition. Let $\{E_n\}$ be an increasing sequence of sets in $\mathcal{D}_2$, then $\{E_n \cap \Omega_1\}$ is an increasing sequence of sets in $\mu_{\Omega_1}(\mathcal{A} \cap \Omega_1)$, and as $\mu_{\Omega_1}(\mathcal{A} \cap \Omega_1)$ is a monotone class, $(E_n \cap \Omega_1) \nearrow (E \cap \Omega_1) \in \mu_{\Omega_1}(\mathcal{A} \cap \Omega_1)$. Therefore, $E_n \nearrow E \in \mathcal{D}_1$. A similar argument holds for a decreasing sequence of sets. Hence $\mathcal{D}_2$ is a monotone class of sets in Ω containing $\mathcal{A}$. Therefore, by minimality $\mu_\Omega(\mathcal{A}) \subset \mathcal{D}_2$. By definition of $\mathcal{D}_2$, $\mu_\Omega(\mathcal{A}) \cap \Omega_1 \subset \mu_{\Omega_1}(\mathcal{A} \cap \Omega_1)$.

Hence $\mu_\Omega(\mathcal{A}) \cap \Omega_1 = \mu_{\Omega_1}(\mathcal{A} \cap \Omega_1)$, and thusly, $\sigma_\Omega(\mathcal{C}) \cap \Omega_1 = \sigma_{\Omega_1}(\mathcal{C} \cap \Omega_1)$ as desired.

Exercise 8. ***[Optional, not to be graded]*** Let $\Omega = [0,1)$ and let $\mathcal{F}_0$ be the collection of finite unions $\cup_{i=1}^N [a_i, b_i)$ for $a_i, b_i \in [0,1]$. For any $A \in \mathcal{F}_0$, let $\mathbb{P}[A] = 1$ if one of the $b_i = 1$, and $\mathbb{P}[A] = 0$ otherwise. In Lectures we showed that $\mathcal{F}_0$ is an algebra but not a σ-algebra.

(a) Show that $\mathbb{P}$ is a non-negative (finitely) additive set-function on $\mathcal{F}_0$.

(b) Show that $\mathbb{P}$ is not countably additive on $\mathcal{F}_0$.

Solution:

(a) For all $A \in \mathcal{F}_0$, $\mathbb{P}[A] \in \{0, 1\}$. Thus $\mathbb{P}$ is non-negative.
Let $A_1, A_2 \in \mathcal{F}_0$ be disjoint. Then $A_1 = \bigcup_{i=1}^{N_1}[a_i^{(1)}, b_i^{(1)})$ and $A_2 = \bigcup_{j=1}^{N_2}[a_j^{(2)}, b_j^{(2)})$, where WLOG the intervals are ordered and non are empty $a_1^{(m)} < b_1^{(m)} < \ldots < a_{N_m}^{(m)} < b_{N_m}^{(m)}$. As $A_1 \cap A_2 = \emptyset$, $A_1 \cup A_2 = \bigcup_{k=1}^{N_1+N_2}[a_k^{(3)}, b_k^{(3)})$, where $a_k^{(3)} \in \{a_i^{(1)}, a_j^{(2)}\}$ and $b_k^{(3)} \in \{b_i^{(1)}, b_j^{(2)}\}$ are the results of interleaving the two collections of intervals and are again WLOG ordered. By construction $\mathbb{P}[A_1] = 1$ if and only if $b_{N_1}^{(1)} = 1$, $\mathbb{P}[A_2] = 1$ if and only if $b_{N_2}^{(2)} = 1$ and $\mathbb{P}[A_1 \cup A_2] = 1$ if and only if $b_{N_1+N_2}^{(3)} = 1$. Moreover, $b_{N_1+N_2}^{(3)} = 1$ if and only if either $b_{N_1}^{(1)} = 1$ or $b_{N_2}^{(2)} = 1$. Suppose $b_{N_1+N_2}^{(3)} = 1$ and WLOG assume $b_{N_1}^{(1)} = 1$, then, as A_1 and A_2 are disjoint, $b_{N_2}^{(2)} \neq 1$

$$\mathbb{P}(A_1 \cup A_2) = 1 = 1 + 0 = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Suppose $b_{N_1+N_2}^{(3)} \neq 1$ then neither $b_{N_1}^{(1)} = 1$ nor $b_{N_2}^{(2)} = 1$

$$\mathbb{P}(A_1 \cup A_2) = 0 = 0 + 0 = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

(b) Let $A_n = [0, 1 - \frac{1}{n})$. Then, for all n, $\mathbb{P}(A_n) = 0$. Moreover, $A_n \subset A_{n+1}$ and $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = [0, 1) \in \mathcal{F}_0$. Hence, by continuity of probability,

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} 0 = 0 \neq 1 = \mathbb{P}([0, 1)) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right),$$

and $\mathbb{P}$ is not countably additive.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 3

Readings:
Notes from Lecture 4 and 5.

Supplementary readings:
[GS], Sections 3.1-3.7.
[C], Section 2.1
[BT], for background on counting (which we don't cover) go through the last section of Ch. 1 of [BT], available at:
http://athenasc.com/Prob-2nd-Ch1.pdf.

Exercise 1. Suppose that $X_1, X_2, \ldots, X_n, \ldots$ are random variables defined on the same probability space. Show that $\max\{X_1, X_2\}$, $\sup_n X_n$, and $\limsup_{n\to\infty} X_n$ are random variables, using only Definition 1 in the notes for Lecture 4, and first principles, without quoting any other known facts about measurability.

Solution: For any $c \in \mathbb{R}$, we have

$$\{\omega : \max\{X_1(\omega), X_2(\omega)\} \leq c\} = \{\omega : X_1(\omega) \leq c\} \bigcap \{\omega : X_2(\omega) \leq c\} .$$

Since X_1 and X_2 are random variables, we know that the set on the left-hand side is measurable, i.e., $\max\{X_1, X_2\}$ is a random variable. We also have

$$\left\{\omega : \sup_n X_n(\omega) \leq c\right\} = \bigcap_n \{\omega : X_n(\omega) \leq c\} ,$$

and thus $\sup_n X_n$ is a random variable.

To show $\limsup_n X_n$ is a random variable, define $g_k(\omega) = \sup_{n\geq k} X_n(\omega)$. We have shown that these functions are measurable for all k. We now argue that

$$\left\{\omega : \limsup_n X_n(\omega) \geq c\right\} = \bigcap_k \{\omega : g_k(\omega) \geq c\},$$

which will immediately imply that $\limsup_n X_n$ is measurable. Indeed, suppose that ω belongs to the set on the right-hand side, i.e., $g_k(\omega) \geq c$, for all k. Since $\limsup X_n(\omega) = \lim_k g_k(\omega)$, and using the definition of $\limsup$, it follows that $\limsup X_n(\omega) \geq c$, and ω belongs to the set on the left-hand side.

Conversely, suppose that ω belongs to the set on the left-hand side. Then, $\lim_k g_k(\omega) =$
$\limsup_n X_n(\omega) \geq c$. However, $g_k(\omega)$ is a nonincreasing sequence of numbers, because the supremum is being progressively taken over smaller sets. It follows that $g_k(\omega) \geq c$ for all k, and ω belongs to set on the right-hand side.

Exercise 2. Given a distribution function F_X show that $F_X(x)$ is continuous at x_0 if and only if $\mathbb{P}(X = x_0) = 0$.

Solution: Since a distribution function is always right continuous, F_X is continuous at x_0 if and only if it is left-continuous at x_0, that is $\lim_{n\to\infty} F_X(x_n) = F_X(x_0)$ for any sequence $x_n \uparrow x_0$, or equivalently

$$\lim_{n\to\infty} \mathbb{P}(X^{-1}(-\infty, x_n]) = \mathbb{P}(X^{-1}(-\infty, x_0]).$$

The sequence of sets $\{X^{-1}(-\infty, x_n]\}$ is increasing, and by continuity of the probability measures $\mathbb{P}$ and action of the inverse image under unions

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(X^{-1}(-\infty, x_n]) &= \mathbb{P}(X^{-1}\left(\bigcup_{n=1}^{\infty}(-\infty, x_n]\right) \\ &= \mathbb{P}(X^{-1}(-\infty, x_0)) \\ &= \mathbb{P}(X^{-1}(-\infty, x_0]) - \mathbb{P}(X = x_0).\end{aligned}$$

Hence the desired equality is satisfied if and only if $\mathbb{P}(X = x_0) = 0$.

Exercise 3. The probabilistic method. A party of $n = 20$ people is gathered. A host selects some of the $n(n-1)/2$ pairs of people and introduces them to each other. Show that the host can do the introduction in such a way that for every group of 7 people there are at least two who are introduced to each other and there are at least two who are not.

Hint: Consider introducing people randomly and independently.

Note: The probabilistic method is a general method for proving existence: if you can prove that a randomly selected structure has certain desired properties with some positive probability (no matter how small), then a structure with these properties is guaranteed to exist.

Solution: Assume that all possible pairs are introduced independently with probability $1/2$. Let $\mathcal{S}$ be the set of all subsets of size 7 out of the 20 people at the party. For each $S \in \mathcal{S}$, let A_S be the event that there is a pair in S that was introduced and a pair in S that was not introduced. The probability of the complementary event is

$$\mathbb{P}(A_S^c) = 2\left(\frac{1}{2}\right)^{\binom{7}{2}} = \frac{1}{2^{20}},$$

as A_S^c occurs if either all of the $\binom{7}{2}$ pairs get introduced or non of the $\binom{7}{2}$ pairs get introduced.

By the probabilistic method it suffices to show that $A = \bigcap_{S\in\mathcal{S}} A_S$ has positive probability, as this event would have zero probability if no such configuration existed, or equivalently, the complementary event does not have unit probability. Applying a union bound

$$\mathbb{P}(A^c) = \mathbb{P}\left(\bigcup_{S\in\mathcal{S}} A_S^c\right) \leq \sum_{S\in\mathcal{S}} \mathbb{P}(A_S^c) = \frac{|\mathcal{S}|}{2^{20}} = \frac{\binom{20}{7}}{2^{20}} < 0.08 < 1,$$

as desired.

Exercise 4. Let X be a nonnegative integer random variable. Show that

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

Be careful in citing whatever results from the lecture notes are needed to justify the steps in your derivation.

Solution: By definition,

$$\begin{aligned} E[X] &= \sum_{a=1}^{\infty} a\mathbb{P}(X = a) \\ &= \sum_{a=1}^{\infty}\sum_{n=1}^{a} \mathbb{P}(X = a) \\ &= \sum_{n=1}^{\infty}\sum_{a=n}^{\infty} \mathbb{P}(X = a) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X \geq n), \end{aligned}$$

where the third inequality follows from Eq. (1) in Lecture 5, which allows us to interchange the order of summation in double sums of nonnegative numbers.

Exercise 5. Let F_1 and F_2 be two CDFs, and suppose that $F_1(t) < F_2(t)$, for all t. Assume that F_1 and F_2 are continuous and strictly increasing. Show that there exist random variables X_1 and X_2, with CDFs F_1 and F_2, respectively, defined on the same probability space such that $X_1 > X_2$. *Hint:* Think of simulating X_1 and X_2 using a common "random number generator".

Solution: Let $((0,1), \mathcal{B}, \lambda)$ be the Lebesgue probability space on $(0,1)$, where $\mathcal{B}$ is the Borel σ-algebra on $(0,1)$. By assumption F_i is strictly increasing and continuous on $\mathbb{R}$. Therefore, $F_i^{\ 1}$ exists and is strictly increasing on $(0,1)$. Let $X_i : ((0,1), \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$, $X_i = F_i^{\ 1}$. Then the X_i are random variables, as the F_i^{-1} are continuous which implies measurable, with CDFs

$$F_{X_i}(t) = \lambda(X_i^{-1}(-\infty, t]) = \lambda(F_i(-\infty, t]) = \lambda((0, F_i(t)]) = F_i(t).$$

Moreover, for $t \in (0,1)$

$$\begin{aligned} & t = F_1(F_1^{-1}(t)) < F_2(F_1^{-1}(t)) \\ \implies \quad & F_2^{-1}(t) < F_2^{-1}(F_2(F_1^{-1}(t))) = F_1^{-1}(t). \end{aligned}$$

Hence $X_1(\omega) = F_1^{-1}(\omega) > F_2^{-1}(\omega) = X_2(\omega)$ for all $\omega \in (0,1)$ and thusly, $X_1 > X_2$.

Exercise 6. Let $\{X_n\}$ be a sequence of independent non-negative random variables. Show that sequence X_n is almost surely bounded if and only if $\sum_{n=1}^{\infty} \mathbb{P}(X_n > c) < \infty$ for some c. (Hint: X_n a.s. bounded simply means $\mathbb{P}(\sup_n X_n = \infty) = 0$.)

Solution: Suppose that there exists such a c. Then, with probability 1, the set $S = \{n \mid X_n > c\}$ is finite. Then, $\sup_n X_n \le \max\{c, \max_{n \in S} X_n\} < \infty$, a.s.

Conversely, if if no such c exists, then $X_n > c$, i.o. By letting k range over the integers, we see that except for a countable union of zero measure sets, then for all k, there exists n_k such that $X_{n_k} > k$, so that $\sup_n X_n = \infty$, a.s.

(Alternate Solution). The event $\{\sup_n X_n = \infty\}$ can be expressed as

$$\{\sup_n X_n = \infty\} = \bigcap_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{X_k > m\}.$$

Suppose $\sup_n X_n(\omega) = \infty$. Suppose there exists natural numbers m_0 and n_0 so that $\sup_{k \ge n_0} \le m_0$, then

$$\begin{aligned} \sup_n X_n(\omega) &= \max\left\{ \max_{j=1,\dots,n_0-1} X_j(\omega), \sup_{k \ge n_0} X_k(\omega) \right\} \\ &\le \max\left\{ \max_{j=1,\dots,n_0-1} X_j(\omega), m_0 \right\} < \infty, \end{aligned}$$

a contradiction. Conversely suppose for all $M \in \mathbb{N}$ and for all $N \in \mathbb{N}$ there exists $k_N \geq N$ such that $X_{k_N}(\omega) > M$, then

$$\sup_n X_n(\omega) \geq X_{k_N} > M.$$

Therefore the $\sup$ is larger than any natural number and must be infinite. Hence the desired relation holds.

Rewriting the above expression

$$\left\{ \sup_n X_n = \infty \right\} = \bigcap_{m=1}^{\infty} \{X_n > m \text{ i.o. }\}.$$

Hence, for all $m \in \mathbb{N}$

$$\mathbb{P}\left(\sup_n X_n = \infty \right) \leq \mathbb{P}\left(X_n > m \text{ i.o. } \right\}. \tag{1}$$

Moreover, as the events $\{X_n > m\}$ are nested, $\{X_n > m+1\} \subset \{X_n > m\}$, by continuity of probability

$$\mathbb{P}\left(\sup_n X_n = \infty \right) = \lim_{m \to \infty} \mathbb{P}\left(X_n > m \text{ i.o } \right). \tag{2}$$

Suppose X_n is almost surely bounded. Suppose for all $c \in \mathbb{R}$

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n > c) = \infty.$$

then as the $\{X_n\}$ are independent by the Borel Cantelli Lemma, for all c,

$$\mathbb{P}(X_n > c \text{ i.o }) = 1.$$

In particular this holds for all $m \in \mathbb{N}$. Applying (2)

$$\mathbb{P}\left(\sup_n X_n = \infty \right) = \lim_{m \to \infty} \mathbb{P}\left(X_n > m \text{ i.o } \right) = \lim_{m \to \infty} 1 = 1,$$

a contradiction.

Conversely suppose there exists $c \in \mathbb{R}$ such that

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n > c) < \infty \quad \Longrightarrow \quad \mathbb{P}\left(X_n > c \text{ i.o} \right) = 0,$$

by Borel Cantelli. Therefore, there exists an $m_0 \in \mathbb{N}$, $m_0 \geq c$ and by monotonicity of $\mathbb{P}$ and (1)

$$\mathbb{P}(\sup_n X_n = \infty) \leq \mathbb{P}(X_n \geq m_0) \leq \mathbb{P}(X_n \geq c) = 0.$$

Hence X_n is almost surely bounded.

Exercise 7. (Another application of the probabilistic method.) Let G be an undirected graph with neither loops nor multiple edges, and write d_v for the degree of vertex v (i.e., the number of edges incident on v). An independent set is a set of vertices no pair of which is joined by an edge. Let $\alpha(G)$ be the size of the largest independent set of G. Use the probabilistic method to show that $\alpha(G) \geq \sum_v 1/(1+d_v)$. *Hint:* Order the nodes at random, and examine the nodes one at a time, putting them in the independent set as long as there are no conflicts with previously examined nodes. Find the expected value of the resulting set.

Solution: The result in this problem is known as Turan's theorem. Let V be the set of vertices of G and note $|V| = n$. Consider ordering the n nodes of a graph according to a random permutation. Nodes in this ordering have neighbors that come earlier, or later in the ordering.

Let I be the set of nodes whose neighbors all come after them in the randomly selected ordering. Note that I must be an independent set. We can write the random variable $|I|$ as the sum of indicator variables X_v, where X_v is 1 if and only if node v is in I, and zero otherwise.

Under a random ordering, the probability that a node v is in I, is at least the probability that a node is the first among it and all its d_v neighbors. This is $1/(d_v + 1)$. Therefore,

$$E(|I|) \geq \sum_{v \in V} \frac{1}{d_v + 1},$$

and therefore there must exist some ordering for which the size of I is at least as big as this. This gives a lower bound on $\alpha(G)$, and the result follows.

In addition, you need to be sure that you can solve elementary problems. As a check, make sure you are able to solve the next problem (not to be handed in).

Drill problem: At his workplace, the first thing Oscar does every morning is to go to the supply room and pick up one, two, or three pens with equal probability 1/3. If he picks up three pens, he does not return to the supply room again that day. If he picks up one or two pens, he will make one additional trip to the supply room, where he again will pick up one, two, or three pens with equal probability 1/3. (The number of pens taken in one trip will not affect the number of pens taken in any other trip.) Calculate the following:

(a) The probability that Oscar gets a total of three pens on any particular day.

(b) The conditional probability that he visited the supply room twice on a given day, given that it is a day in which he got a total of three pens.

(c) $\mathbb{E}[N]$ and $\mathbb{E}[N \mid C]$, where $\mathbb{E}[N]$ is the unconditional expectation of N, the total number of pens Oscar gets on any given day, and $\mathbb{E}[N \mid C]$ is the conditional expectation of N given the event $C = \{N > 3\}$.

(d) $\sigma_{N|C}$, the conditional standard deviation of the total number of pens Oscar gets on a particular day, where N and C are as in part (c).

(e) The probability that he gets more than three pens on each of the next 16 days.

(f) The conditional standard deviation of the total number of pens he gets in the next 16 days given that he gets more than three pens on each of those days.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 4

Readings:

(a) Notes from Lecture 6 and 7.

(b) [Cinlar] Sections I.4, I.5 and II.2

(c) [GS] Chapter 3

Exercise 1. Let N be a random variable that takes nonnegative integer values. Let $X_1, X_2, \ldots$, be a sequence of i.i.d. discrete random variables that have finite expectation and are independent from N. Use iterated expectations to show that the expected value of $\sum_{i=1}^{N} X_i$ is $\mathbb{E}[N]\mathbb{E}[X_1]$.

Solution:

$$\begin{aligned} E[\sum_{i=1}^{N} X_i] &= E[E[\sum_{i=1}^{N} X_i|N]] \\ &= E[NE[X_1]] \\ &= E[N]E[X_1] \end{aligned}$$

Exercise 2. Let X and Y be binomial with parameters (m, p) and (n, q), respectively.

(a) Show that if X is independent from Y, $m = n$, and $p = q$ then $X + Y$ is binomial. *Hint:* Use the interpretation of the binomial, not algebra.

(b) Does the conclusion of part (a) remain valid if $m \neq n$? If X and Y are not independent? If $p \neq q$?

(c) Show that if X and Y are independent, then

$$\mathbb{P}(X + Y = k) = \sum_{i=-\infty}^{\infty} p_X(i) p_Y(k - i).$$

(d) Use the result from part (c) to find the PMF of $X + Y$ where X and Y are independent Poisson random variables with parameters λ and μ, respectively. *Hint:* The "binomial theorem" states that

$$(a + b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.$$

Solution:

(a) X and Y can be constructed as sums of i.i.d. Bernoulli random variables. Thus, as long as $p = q$, then $X + Y$ can also be constructed as a sum of i.i.d. Bernoulli random variables, which means that it has a Binomial distribution.

(b) The conclusion of part (a) remains valid if $m \neq n$ using the same argument. If X and Y are not independent, it doesn't hold. For example, for $X = Y$ with $p \in (0, 1)$ and $n > 1$, we have that $X + Y$ only takes values in the even numbers, and not on odds. Thus, it cannot be binomial. If $p \neq q$, the conclusion also does not hold. For example, if $p = 1$, $q = 1/2$, and $n = m = 2$, we have that $\mathbb{P}(X + Y = 1) = \mathbb{P}(X + Y = 2) > \mathbb{P}(X + Y = 3)$, which cannot happen with a Binomial random variable.

(c) If X and Y are independent, then

$$\begin{aligned}\mathbb{P}(X + Y = k) &= \sum_{i=-\infty}^{\infty} p_{X|Y}(i|k - i) p_Y(k - i) \\ &= \sum_{i=-\infty}^{\infty} p_X(i) p_Y(k - i).\end{aligned}$$

(d) If X and Y are independent Poisson random variables with parameters λ and μ, respectively, we have

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{i=-\infty}^{\infty} p_X(i)p_Y(k-i) \\
&= e^{-\lambda}e^{-\mu}\sum_{i=0}^{k}\frac{\lambda^i\mu^{k-i}}{i!(k-i)!} \\
&= e^{-(\lambda+\mu)}\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda^i\mu^{k-i} \\
&= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!},
\end{aligned}$$

which is a Poisson random variable with parameter $\lambda+\mu$.

Exercise 3. A 4-sided die has its four faces labeled as a, b, c, d. Each time the die is rolled, the result is a, b, c, or d, with probabilities p_a, p_b, p_c, p_d, respectively. Different rolls are statistically independent. The die is rolled n times. Let N_a and N_b be the number of rolls that resulted in a or b, respectively. Find the covariance of N_a and N_b.

Solution: Let $\{X_k \mid k = 1, \dots, n\}$ be the i.i.d rolls of the dice. Using indicator functions, the number of rolls resulting in a particular outcome is

$$N_e = \sum_{k=1}^{n} \mathbb{1}\{X_n = e\},$$

with mean

$$E[N_e] = \sum_{k=1}^{n} E\left[\mathbb{1}\{X_n = e\}\right] = np_e.$$

The covariance between N_a and N_b is

$$\begin{aligned}\text{cov}(N_a, N_b) = E\left[\left(\sum_{i=1}^{n} (\mathbb{1}\{X_i = a\} - p_a)\right)\left(\sum_{j=1}^{n} (\mathbb{1}\{X_j = b\} - p_b)\right)\right] \\ = \sum_{k=1}^{n} E\left[(\mathbb{1}\{X_k = a\} - p_a)(\mathbb{1}\{X_k = b\} - p_b\})\right] \\ + \sum_{i=1}^{n} \sum_{j \neq i} E\left[(\mathbb{1}\{X_i = a\} - p_a)\left(\mathbb{1}\{X_j = b\} - p_b\right)\right].\end{aligned}$$

This later term is zero since X_i and X_j are independent for $i \neq j$. For this first term

$$\begin{aligned}&E\left[(\mathbb{1}\{X_k = a\} - p_a)(\mathbb{1}\{X_k = b\} - p_b\})\right] \\ &\quad = E\left[\mathbb{1}\{X_k = a\}\mathbb{1}\{X_k = b\}\right] - p_b E\left[\mathbb{1}\{X_k = a\}\right] - p_a E\left[\mathbb{1}\{X_k = b\}\right] + p_a p_b \\ &\quad = 0 - 2p_a p_b + p_a p_b = -p_a p_b.\end{aligned}$$

Hence

$$\text{cov}(N_a, N_b) = -np_a p_b.$$

Exercise 4. Suppose that X and Y are discrete random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. An elegant way of defining the conditional expectation of Y given X is as a random variable of the form $\phi(X)$ (where ϕ is a measurable function), such that

$$\mathbb{E}[\phi(X)g(X)] = \mathbb{E}[Yg(X)],$$

for all measurable functions g. In this problem, we will prove that this condition defines the conditional expectation uniquely; that is, if we also have

$$\mathbb{E}[\psi(X)g(X)] = \mathbb{E}[Yg(X)],$$

for every measurable function g, then $\phi(X)$ and $\psi(X)$ are almost surely equal, i.e., $\mathbb{P}(\phi(X) = \psi(X)) = 1$.

(a) Prove that the following sets are $\mathcal{F}$-measurable: $\{\phi(X) > \psi(X)\}$ and, for any integer n, $A_n := \{\phi(X) > \psi(X) + 1/n\}$.

(b) Assume the contradiction $\mathbb{P}(\phi(X) = \psi(X)) < 1$ and use $g(x) = \mathbf{1}_{A_n}$ for some appropriate n to show that the conditional expectation is unique.

Solution:

(a) First, since X is discrete, $\phi(X)$ and $\psi(X)$ are random variables, without any further assumptions on ϕ, ψ. This implies that $\{\omega|\phi(X(\omega)) > \psi(X(\omega))\}$ and $\{\omega|\phi(X(\omega)) > \psi(X(\omega))\}$ are $\mathcal{F}$-measurable sets; indeed, since one can always fit a rational number between any two distinct real numbers, we have

$$\{\phi(X) > \psi(X)\} = \bigcup_{q \in Q} \{\phi(X) > q\} \bigcap \{q > \psi(X)\}$$

The set $\{\phi(X) > \psi(X) + a\}$ is also $\mathcal{F}$-measurable for any real a (and of course, for $1/n$) since, similarly,

$$\{\phi(X) > \psi(X) + a\} = \bigcup_{q \in Q} \{\phi(X) > q\} \bigcap \{q - a > \psi(X)\}$$

(b) We now proceed to the proof of uniqueness. Assume by contradiction that $\mathbb{P}(\phi(X) = \psi(X)) < 1$. Without loss of generality, we can assume $\mathbb{P}(\phi(X) < \psi(X)) > 0$. The sets

$$A_n = \{\omega|\phi(X)(\omega) + 1/n < \psi(X)(\omega)\}$$

form an increasing sequence of $\mathcal{F}$-measurable sets and

$$\{\omega|\phi(X)(\omega) < \psi(X)(\omega)\} = \bigcup_{n \geq 1} A_n.$$

By continuity of probability, there exists some n such that $\mathbb{P}(A_n) > 0$.

Now, define $g(x) = \mathbf{1}_{A_n}$. Its expectation is well-defined and by construction,

$$\begin{aligned} E[\phi(X)g(X)] &= E[\phi(X)1_{A_n}] \\ &\leq E[(\psi(X) - 1/n)1_{A_n}] \\ &= E[\psi(X)g(X)] - (1/n)P(A_n) \\ &< E[\psi(X)g(X)] \end{aligned}$$

which is a contradiction.

Exercise 5. A machine is refilled each morning with n portions of vanilla and chocolate ice creams each (a total of $2n$ portions). Customers arrive sequentially, each getting one of the ice creams independently with probability $1/2$. Consider the first moment when a customer receives an "out of order" message. Let X be the number of portions of the other type left at this moment, $0 \leq X \leq n$. Find the distribution of X.

Solution: If the unlucky customer ordered vanilla, then there are $X = m$ portions of the chocolate left, if there were $2n - m$ portions were given earlier, out of which n were vanilla. There are $\binom{2n-m}{n}$ sequences of orders that lead to that scenario, all happening with probability $2^{-(2n-m)}$. Thus, since vanilla and chocolate are interchangeable, we have

$$\mathbb{P}[X = m] = \binom{2n - m}{n} 2^{-(2n-m)}.$$

Exercise 6. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. (So, μ is a measure, but not necessarily a probability measure.) Let $g : \Omega \to \mathbb{R}$ be a nonnegative measurable function. Let $\{B_i\}$ be a sequence of disjoint measurable sets. Prove that

$$\int_{\cup_i B_i} g \, d\mu = \sum_{i=1}^{\infty} \int_{B_i} g \, d\mu.$$

(Be rigorous!)

Note: As an application, this exercise gives another rich source of probability measures. Namely, take f – a nonnegative measurable function on the real line with $\int_{\mathbb{R}} f(x)dx = 1$ (integral w.r.t. Lebesgue measure), and define a set-function $\mathbb{P}(A) = \int_A f dx$. The exercise shows that $\mathbb{P}(\cdot)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$. Function f is called the probability density function (PDF) of $\mathbb{P}$.

Solution: Let $B = \bigcup_{i=1}^{\infty} B_i$ and

$$g_k = g\mathbb{1}_{\cup_{i=1}^{k} B_i} = \sum_{i=1}^{k} g\mathbb{1}_{B_i},$$

where the last equality holds since the B_i are disjoint. Moreover, for any arbitrary countable collection of disjoint sets $\{U_n\}$

$$\mathbb{1}_{\bigcup_{n\in\mathbb{N}} U_n} = \sum_{n\in\mathbb{N}} \mathbb{1}_{U_n}.$$

Since g is nonnegative g_k is increasing to $g\mathbb{1}_B$, and therefore,

$$\begin{aligned}
\int_B g\,d\mu &= \int g\mathbb{1}_B\,d\mu \\
&= \int \lim_{k\to\infty} g_k\,d\mu \quad (g_k \nearrow g\mathbb{1}_B) \\
&= \lim_{k\to\infty} \int g_k\,d\mu \quad (\text{ Monotone convergence theorem}) \\
&= \lim_{k\to\infty} \int \sum_{i=1}^{k} g\mathbb{1}_{B_i}\,d\mu \\
&= \lim_{k\to\infty} \sum_{i=1}^{k} \int g\mathbb{1}_{B_i}\,d\mu \quad (\text{Linearity of integration}) \\
&= \sum_{i=1}^{\infty} \int_{B_i} g\,d\mu.
\end{aligned}$$

Exercise 7. ***[Optional, not to be graded]*** Let μ and ν be two finite measures on $(\mathbb{R}, \mathcal{B})$. Show that if

$$\int_{\mathbb{R}} f\,d\mu = \int_{\mathbb{R}} f\,d\nu$$

for all bounded continuous functions f then $\mu = \nu$. (*Hint:* write $\mathbb{1}_{(a,b)}(x)$ as an increasing limit of continuous functions.)

Note: This exercise shows that measure on Borel σ-algebra is uniquely characterized by its values on continuous functions. This is true on $\mathbb{R}$, $\mathbb{R}^n$ and any other topological space. Similar to how it is sufficient to know measures only on intervals $(-\infty, a)$ it is sufficient to consider only a handful of functions (such as all sines and cosines, or all exponents). This will be discussed later.

Solution: Let $a < b \in \mathbb{R}$. Consider the sequence of functions, defined for all $n > \frac{2}{b-a}$ and 0 otherwise,

$$f_n(x) = \begin{cases} 0 & x \leq a \\ n(x-a) & a < x < a + \frac{1}{n} \\ 1 & a + \frac{1}{n} \leq x \leq b - \frac{1}{n} \\ -n(x-b) & b - \frac{1}{n} < x < b \\ 0 & x \geq b \end{cases}.$$

By construction the $\{f_n\}$ are increasing, bounded and continuous, piecewise linear. Moreover, $\lim_{n\to\infty} f_n \nearrow \mathbb{1}_{(a,b)}$, this follows since

$$(a, a+\tfrac{1}{n}), (b-\tfrac{1}{n}, b) \to \emptyset \qquad [a+\frac{1}{n}, b-\frac{1}{n}] \to (a,b).$$

Therefore, by the monotone convergence theorem

$$\begin{aligned}
\mu(a,b) &= \int \mathbb{1}_{(a,b)}\, d\mu \\
&= \int \lim_{n\to\infty} f_n\, d\mu \quad (f_n \nearrow \mathbb{1}_{(a,b)} \text{ pointwise}) \\
&= \lim_{n\to\infty} \int f_n\, d\mu \quad (\text{Monotone converge theorem}) \\
&= \lim_{n\to\infty} \int f_n\, d\nu \quad (f_n \text{ bounded and continuous}) \\
&= \int \lim_{n\to\infty} f_n\, d\nu \quad (\text{Monotone convergence theorem}) \\
&= \int \mathbb{1}_{(a,b)}\, d\nu = \nu(a,b).
\end{aligned}$$

Similar to question 1 of homework 3, if two measures agree on a generating collection for a σ-algebra they agree on the entire σ-algebra. While the open intervals (a,b) have not been explicitly given as a generating collection for the Borel σ-algebra thus far, observe that any arbitrary closed interval can be written as the intersection of open intervals

$$[a,b] = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b + \frac{1}{n}).$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J — Fall 2018

Problem Set 5

Readings:
(a) Notes from Lectures 7-9.
(b) [Cinlar] Sections I.4-I.6

Exercise 1. The worker's union requests that all workers at a factory be given the day off if at least one worker has a birthday on that day. Otherwise workers agree to work 365 days a year. Management is to maximize the number of man-days worked per year. How many workers should they hire?

Solution: Management will maximize the expected number of man-days worked per year assuming that worker's birthdays are independent and identically distributed uniformly over the calender year. More specifically, given n workers, let $\{B_k \mid k = 1, \ldots, n\}$ be the birthday of the k-th worker and $D = 365$ be the number of calendar days, then, for all k and for all $d \in 1, \{\ldots, D\}$, $\mathbb{P}(B_k = d) = \frac{1}{D}$. Let $W_d(n)$ be an indicator random variable for whether or not the factory is open on day d and $W(n) = \sum_{d=1}^{D} W_d$ be the number of days worked. Then

$$\begin{aligned} &\{W_d(n) = 1\} = \{\text{No worker has a birthday on day } d\} \\ \implies \quad &\mathbb{P}(W_d(n) = 1) = \left(1 - \frac{1}{D}\right)^n, \end{aligned}$$

and the expected number of work days is

$$E\left[W(n)\right] = \left[\sum_{d=1}^{D} W_d(n)\right] = \sum_{d=1}^{D} E\left[W_d(n)\right] = \sum_{d=1}^{D} \left(1 - \frac{1}{D}\right)^n = D\left(1 - \frac{1}{D}\right)^n.$$

Hence the expected number of man-days worked is

$$nD\left(1 - \frac{1}{D}\right)^n.$$

Consider the ratio

$$r(n) = \frac{nD\left(1 - \frac{1}{D}\right)^n}{(n-1)D\left(1 - \frac{1}{D}\right)^{n-1}} = \frac{n}{n-1}\left(1 - \frac{1}{D}\right).$$

Then $r(n) \geq 1$ for $n \leq D$ and $r(n) < 1$ for $n > D$, and the optimal number of workers is $n = D = 365$.

Exercise 2. Let $\Omega = \mathbb{Z}_+$, $\mathcal{F} = 2^{\Omega}$. Complete construction of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and come up with a sequence of random variables X_n which is increasing a.e., but $\mathbb{E}[X_n]$ does not converge to $\mathbb{E}[X]$, where $X = \lim_n X_n$ a.e.

Solution: Consider the probability space $(\mathbb{N}, 2^{\mathbb{N}}, \mathbb{P})$ with $\mathbb{P}(k) = \frac{6}{\pi^2}\frac{1}{k^2}$. Let $X_n(k) = -\mathbb{1}\{k \geq n\}k$. Then $X_n \leq X_{n+1}$ and for all n

$$E\left[X_n\right] = \sum_{k=n}^{\infty} -k\frac{6}{\pi^2}\frac{1}{k^2} = -\frac{6}{\pi^2}\sum_{k=n}^{\infty}\frac{1}{k} = -\infty.$$

However, $\lim_{n\to\infty} X_n = 0$. Hence

$$\lim_{n\to\infty} E\left[X_n\right] = -\infty \neq 0 = E\left[\lim_{n\to\infty} X_n\right].$$

Exercise 3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probabilisty space and $X \geq 0$ a random variable. Show

$$\mathbb{E}[X] = \int_0^{\infty}(1 - F_X(x))dx\,,$$

where $F_X(x) = \mathbb{P}[X \leq x]$ is a CDF of X. (*Hint:* Fubini.)

Solution: We solve the problem for a more general case: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with μ σ-finite and $f \geq 0$ measurable, then

$$\int f(\omega)d\mu(\omega) = \int_0^{\infty} \mu[\{\omega : f(\omega) > x\}]dx,$$

Let $([0,\infty], \mathcal{B}, \lambda)$ be the Lebesgue measure space and $([0,\infty]\times\Omega, \mathcal{F}\times\mathcal{B}, \mu\times\lambda)$ the product measure space with $(\Omega, \mathcal{F}, \mu)$. Although not explicitly discussed in lecture, the construction of $([0,\infty], \mathcal{B})$ is very similar to that of $([0,\infty), \mathcal{B})$. Moreover, for a general topological space X, $\mathcal{B}(X)$ is defined as the smallest σ-algebra containing all the open sets. One way to explicitly generate the topology on $[0,\infty]$ is through the function $\tanh : [0,\infty] \to [0,1]$ with the continuous extensions $\tanh(\infty) = 1$. The open sets in $[0,\infty]$ are then the image of open sets in $[0,1]$ under $\tanh^{-1}$. As discussed in Lecture 9, the Lebesgue measure is σ-finite, and therefore the proofs for the probability measure and Fubini's theorem hold.

Consider the function $g : ([0,\infty] \times [0,\infty], \mathcal{B} \times \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$

$$g(x,y) = \mathbb{1}_{(x,\infty]}(y) = \begin{cases} 1 & x < y \\ 0 & \text{else} \end{cases}.$$

Let $B \in \mathcal{B}$,

$$g^{-1}(B) = \begin{cases} [0,\infty] \times [0,\infty] & \{0,1\} \in B \\ \{x \geq y\} & \{0\} \in B \text{ and } \{1\} \notin B \\ \{x < y\} & \{1\} \in B \text{ and } \{0\} \notin B \\ \emptyset & \text{else} \end{cases}.$$

As $\{x < y\}$ is open and $\mathcal{B} \times \mathcal{B}$ contains all open sets, $g^{-1}(B) \in \mathcal{B} \times \mathcal{B}$ in all cases. Hence g is $(\mathcal{B} \times \mathcal{B}, \mathcal{B})$ measurable.

Claim: Let h_1 be $(\mathcal{F}_1, \mathcal{G}_1)$ measurable and h_2 be $(\mathcal{F}_2, \mathcal{G}_2)$ measurable. The function $h(\omega_1, \omega_2) = (h_1(\omega_1), h(\omega_2))$ is $(\mathcal{F}_1 \times \mathcal{F}_2, \mathcal{G}_1 \times \mathcal{G}_2)$ measurable.

Let

$$\mathcal{L} = \{E \in \mathcal{G}_1 \times \mathcal{G}_2 \mid h^{-1}(E) \in \mathcal{F}_1 \times \mathcal{F}_2\}.$$

Let $B_k \in \mathcal{G}_k$, $h^{-1}(B_1 \times B_2) = (h_1^{-1}(B_1) \times h_2^{-1}(B_2)) \in \mathcal{F}_1 \times \mathcal{F}_2$ by measurability of h_1 and h_2.

$$\begin{aligned} f^{-1}(\emptyset) &= \{(\omega_1, \omega_2) \mid (h_1(\omega_1), h_2(\omega_2)) \in \emptyset\} \\ &= \{(\omega_1, \omega_2) \mid h_1(\omega_1) \in \emptyset \text{ or } h_2(\omega_2) \in \emptyset\} \\ &= \emptyset \times \emptyset = \emptyset. \end{aligned}$$

Thus, $\phi \in \mathcal{L}$. Let $\{E_k\} \in \mathcal{L}$. By properties of the inverse image $h^{-1}\left(\bigcup_{k=1}^{\infty} E_k\right) = \bigcup_{k=1}^{\infty} h^{-1}(E_k) \in \mathcal{F}_1 \times \mathcal{F}_2$ and $h^{-1}(E_k^c) = \left(h^{-1}(E_k)\right)^c \in \mathcal{F}_1 \times \mathcal{F}_2$. Hence $\mathcal{L}$ is a σ-algebra containing a generating p-system for $\mathcal{G}_1 \times \mathcal{G}_2$, and by minimality, $\mathcal{L} = \mathcal{G}_1 \times \mathcal{G}_2$.

In particular, the function $h : (\mathbb{R} \times \Omega) \to (\mathbb{R} \times \mathbb{R})$ $h(x, \omega) = (x, f(\omega))$ is $(\mathcal{B} \times \mathcal{F}, \mathcal{B} \times \mathcal{B})$ measurable. Therefore, the function $(\mathbb{R} \times \Omega) \mapsto (\mathbb{R} \times \mathbb{R})$

$$\mathbb{1}_{(x,\infty]}(f(\omega)) = (g \circ h)$$

is $(\mathcal{B} \times \mathcal{F}, \mathcal{B} \times \mathcal{B})$ measurable. The iterated integrals are

$$\int_\Omega \int_{[0,\infty]} \mathbb{1}_{(x,\infty]}(f(\omega))\, dx\, d\mu = \int_\Omega f(\omega)\, d\mu,$$

and

$$\begin{aligned} \int_{[0,\infty]} \int_\Omega \mathbb{1}_{(x,\infty]}(f(\omega))\, d\mu\, dx &= \int_{[0,\infty]} \int_\Omega \mathbb{1}_{f^{-1}(x,\infty]}(\omega)\, d\mu\, dx \\ &= \int_{[0,\infty]} \mu(f^{-1}(x,\infty])\, dx. \end{aligned}$$

Moreover, as the function is nonnegative Fubini's theorem applies and these are equal

$$\int_\Omega f(\omega)\, d\mu = \int_{[0,\infty]} \mu(f^{-1}(x,\infty])\, dx.$$

Exercise 4. Show that for integrable f

$$\left|\int f d\mu\right| \leq \int |f| d\mu.$$

Solution: By definition

$$\begin{aligned}
\left|\int f\, d\mu\right| &= \left|\int f_+\, d\mu - \int f_-\, d\mu\right| \\
&\leq \left|\int f_+\, d\mu\right| + \left|\int f_-\, d\mu\right| \quad \text{(Triangle Inequality)} \\
&= \int f_+\, d\mu + \int f_-\, d\mu \quad (f_+, f_- \geq 0) \\
&= \int |f|\, d\mu.
\end{aligned}$$

Exercise 5 (Weird integrable functions)**.** Let $\psi(x) = \frac{1}{\sqrt{x}} \mathbb{1}_{(0,1)}(x)$ and

$$F(x) = \sum_{n=1}^{\infty} 2^{-n} \psi(x - r_n)\,,$$

where $\{r_n\}$ is some enumeration of all rationals in $(0, 1)$. Show that $F(x)$ is a measurable non-negative function with

$$\int_{[0,1]} F d\lambda < \infty\,.$$

In particular, $F(x)$ is finite almost everywhere on $[0, 1]$, yet unbounded on every interval.

Solution: The function $\frac{1}{\sqrt{x}}$ is continuous on $(0, 1)$ and simple functions are measurable. Therefore, as continuous functions are measurable and the product of measurable functions is measurable, for all $r \in \mathbb{Q}$, $\psi(x - r)$ is measurable. The sum of measurable functions is measurable and thus

$$f_k := \sum_{n=1}^{k} 2^{-n} \psi(x - r_n)$$

is measurable. As $\psi \geq 0$, the $\{f_k\}$ are increasing and nonnegative. In particular, for all x, the sequence $\{f_k(x)\}$ is increasing and therefore

$$\lim_{n\to\infty} f_k(x)$$

exists. By definition, for all x,

$$F(x) = \lim_{n\to\infty} f_k(x).$$

Hence $f_k \to F$ pointwise and thus F is measurable and nonnegative. By the monotone convergence theorem

$$\begin{aligned}
\int_{[0,1]} F\,d\lambda &= \lim_{k\to\infty} \int_{[0,1]} \sum_{n=1}^{k} 2^{-n}\psi(x - r_n)\,d\lambda(x) \\
&= \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} \psi(x - r_n)\,d\lambda(x) \\
&=^{(a)} \sum_{n=1}^{\infty} 2^{-n} \int_{[-r,1-r]} \psi(x)\,d\lambda(x) \\
&= \sum_{n=1}^{\infty} 2^{-n} \int_{(0,1-r)} \frac{1}{\sqrt{x}}\,d\lambda(x) \\
&\leq \sum_{n=1}^{\infty} 2^{-n} \int_{(0,1)} \frac{1}{\sqrt{x}}\,d\lambda(x),
\end{aligned}$$

where (a) follows from a change of variables and (**??**). For all $m, n \in \mathbb{N}$

$$B_{m,n} = \left[\frac{m^2}{(n+1)^2}, \frac{m^2}{n^2}\right).$$

For a fixed m the $B_{m,n}$ are disjoint and

$$\bigcup_{n=m}^{\infty} B_{m,n} = [0,1), \tag{1}$$

and as $x^{-\frac{1}{2}}$ is decreasing

$$\frac{1}{\sqrt{x}}_{|B_{n,m}} \leq \frac{n+1}{m}. \tag{2}$$

Consider the sequence of functions

$$g_{m,k} = \sum_{n=m}^{k} \frac{n+1}{m} \mathbb{1}_{B_{m,n}}.$$

For all m, k $g_{m,k}$ is a simple functiona and thus measurable. Moreover, by (1) and disjoints of the $B_{m,n}$ these functions increase pointwise in k to a measurable function $g_m := \lim_{k\to\infty} g_{m,k}$ on $[0,1)$. Therefore, by the montone convergence theorem

$$\begin{aligned}
\int_{[0,1)} g_m \, dx &= \sum_{n=m}^{\infty} \int_{[0,1)} \frac{n+1}{m} \mathbb{1}_{B_{m,n}} \, dx \\
&= \sum_{n=m}^{\infty} \int_{[0,1)} \frac{n+1}{m} \left(\frac{m^2}{n^2} - \frac{m^2}{(n+1)^2} \right) \\
&= \sum_{n=m}^{\infty} \int_{[0,1)} m \frac{2n+1}{n^2(n+1)} \\
&\leq 2m \sum_{n=m}^{\infty} \frac{1}{n^2} \\
&= \frac{2}{m} + 2m \sum_{n=m+1}^{\infty} \frac{1}{n^2} \\
&\leq \frac{2}{m} + 2m \int_{m}^{\infty} \frac{1}{x^2} \, dx \\
&= \frac{2}{m} + 2.
\end{aligned}$$

By (2) $\frac{1}{\sqrt{x}} \leq g_m(x)$ for all $x \in [0,1)$. This provides

$$\int_{(0,1)} \frac{1}{\sqrt{x}} \, d\lambda(x) \leq \int_{[0,1)} \frac{1}{\sqrt{x}} \, d\lambda(x) \leq \int_{[0,1)} g_m \, d\lambda(x) \leq \frac{2}{m} + 2.$$

As this holds for all m, it holds in the limit. Therefore,

$$\int_{(0,1)} \frac{1}{\sqrt{x}} \, d\lambda(x) \leq 2.$$

Hence

$$\int_{[0,1]} F \, d\lambda \leq \sum_{n=1}^{\infty} 2 \cdot 2^{-n} < \infty$$

and F is finite almost everywhere. Moreover, as $\frac{1}{\sqrt{x}}$ is unbounded on $(0,1)$, F is unbounded on every interval, i.e. there is a rational r_n in every interval and the function $\psi(x - r_n)$ will be unbounded.

Exercise 6. For all n, let g_n and g be measurable functions. Suppose that $g_n \uparrow g$ and that $\int g_1^- \, d\mu < \infty$. Prove that $\int g_n \, d\mu \uparrow \int g \, d\mu$.

Solution: Let us decompose g and each of the g_n into a pair of nonnegative functions g^- and g^+, and g_n^- and g_n^+, such that $g = g^+ - g^-$ and $g_n = g_n^+ - g_n^-$. Since $g_n \uparrow g$, then we have that $g_n + g_1^-$ are nonnegative functions such that $g_n + g_1^- \uparrow g + g_1^-$. Then, using the fact that $\int g_1^- \, d\mu < \infty$ and the MCT, we have

$$\begin{aligned}
\int g_n d\mu &= \int g_n + g_1^- - g_1^- d\mu \\
&= \int g_n + g_1^- d\mu - \int g_1^- d\mu \\
&\uparrow \int g + g_1^- d\mu - \int g_1^- d\mu \\
&= \int g + g_1^- - g_1^- d\mu \\
&= \int g d\mu.
\end{aligned}$$

Exercise 7. (Differentiating under the integral sign)
Let $g : \mathbb{R}^2 \mapsto \mathbb{R}$ be a continuous function of two variables s and x. Furthermore, assume that the derivative $g'(s,x) = (\partial g/\partial s)$ exists for every s and x, is jointly measurable in (s,x) and is a continuous function of s for any fixed x. Assume $|g'(s,x)| \leq c$ for all s, x.

Let X be a random variable. Show that

$$\frac{\partial}{\partial s}\mathbb{E}[g(s,X)] = \mathbb{E}\left[\frac{\partial g}{\partial s}(s,X)\right].$$

Note: You can use the fact from elementary calculus that under our assumptions, $g(s,x) = g(0,x) + \int_0^s \frac{\partial g}{\partial s}(u,x)du$ for all x.

Solution: Taking expectation on both sides of the given identity,

$$\mathbb{E}[g(s,X)] = \mathbb{E}[g(0,X)] + \mathbb{E}\int_0^s \frac{\partial g}{\partial s}(u,x)du$$

Since the Lebesgue measure on $\mathbb{R}$ is σ-finite and $|\frac{\partial g}{\partial s}(u,x)| \leq c$, and c is integrable over $[0,s]$, Fubini theorem yields

$$\mathbb{E}\int_0^s \frac{\partial g}{\partial s}(u,x)du = \int_0^s \mathbb{E}\frac{\partial g}{\partial s}(u,x)du.$$

As a result, $\mathbb{E}[g(s,X)] = \mathbb{E}[g(0,X)] + \int_0^s \mathbb{E}\frac{\partial g}{\partial s}(u,x)du$ is differentiable with derivative at s equal to $\mathbb{E}\frac{\partial g}{\partial s}(s,x)$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 6

Readings:

(a) Notes from Lecture 10 and 11.

(b) [Grimmett-Stirzaker]: Section 4.1-4.10. Optionally, Section 4.11.

Exercise 1. The probabilistic method. Twelve per cent of the circumference of a circle is colored blue, the rest is red. Show that, irrespective of the manner in which the colors are distributed, it is possible to inscribe a regular octagon in the circle with all its vertices red.

Hint: The probabilistic method is a general method for proving existence: if you can prove that a randomly selected structure has certain desired properties with some positive probability (no matter how small), then a structure with these properties is guaranteed to exist.

Solution: We pick an inscribed regular octagon at random by choosing uniformly over the circle the position of a vertex. Let event V_i be the event that vertex i is red, $i = 1, \ldots, 8$. We are interested in showing that $\mathbb{P}(V_1 \cap \cdots \cap V_8) > 0$. Note that for any i, $\mathbb{P}(V_i) = 0.88$. Unfortunately, our events are not independent, so we cannot multiply probabilities. Instead, we have:

$$\begin{aligned}
\mathbb{P}(V_1 \cap \cdots \cap V_8) &= 1 - \mathbb{P}((V_1 \cap \cdots \cap V_8)^c) \\
&= 1 - \mathbb{P}(V_1^c \cup \cdots \cup V_8^c) \\
&\geq 1 - \sum_{i=1}^{8} \mathbb{P}(V_i^c) \\
&= 1 - 8 \times 0.12 = 0.04 > 0.
\end{aligned}$$

Exercise 2. Suppose X is a continuous random variable with a power law distribution. Namely there exists $c > 0$ and $\alpha > 0$ such that $\mathbb{P}(X > x) = \frac{c}{x}$, for every $x \geq c$. Consider the r-th moment of X, namely $\mathbb{E}[X^r]$, where $r > 0$ is any real value. Find necessary and sufficient conditions for r in terms of c and α for the r-th moment to be finite.

Solution: The pdf for X is

$$f_X(x) = \frac{d}{dx}F_X(x) = \alpha\, c^{\alpha} x^{-(1+\alpha)}.$$

Thus, we have

$$\mathbb{E}[X^r] = \int_c^{\infty} x^r \alpha c^{\alpha} x^{-(1+\alpha)} dx,$$

which is finite iff $r < \alpha$.

Exercise 3. We have a stick of unit length $[0, 1]$, and break it at X, where X is uniformly distributed on $[0, 1]$. Given the value x of X, we let Y be uniformly distributed on $[0, x]$, and let Z be uniformly distributed on $[0, 1-x]$. We assume that conditioned on $X = x$, the random variables Y and Z are independent. Find the joint PDF of Y and Z. Find $\mathbb{E}[X|Y]$, $\mathbb{E}[X|Z]$, and $\rho(Y, Z)$.

Solution: The conditional PDFs for $X|Y$ and $X|Z$ are

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{(1/x) 1_{\{y \leq x\}} 1_{x \in [0,1]}}{\log 1/y},$$

and

$$f_{X|Z}(x|z) = \frac{f_{Z|X}(z|x) f_X(x)}{f_Z(z)} = \frac{(1/(1-x)) 1_{z \leq 1-x} 1_{x \in [0,1]}}{\log 1/z}.$$

Therefore, the joint PDF of X, Y, Z is

$$\begin{aligned} f_{Y,Z,X}(y, z, x) &= f_{Y,Z|X}(y, z|x) f_X(x) = f_{Y|X}(y|x) f_{Z|X}(z|x) f_X(x) \\ &= \frac{1}{x}\frac{1}{1-x} 1_{\{y \leq x\}} 1_{\{z \leq 1-x\}}, \end{aligned}$$

and integrating the joint pdf of Y and Z is

$$\begin{aligned} f_{Y,Z}(y, z) &= \int_0^1 f_{Y,Z,X}(y, z, x) dx \\ &= \int_y^{1-z} \frac{1}{x(1-x)} dx \\ &= \log(1-z) - \log y + \log(1-y) - \log z \end{aligned}$$

when $y \leq 1-z$, and 0 otherwise.

Moreover, the conditional expectations for $X|Y$ and $X|Z$ are

$$E[X|Y] = \int_0^1 x f_{X|Y}(x|y)dx = \int_0^1 \frac{1_{\{y\leq x\}}}{\log 1/y}dx = \frac{1-y}{\log 1/y} = \frac{y-1}{\log y},$$

and

$$\begin{aligned}
E[X|Z] &= \int_0^1 x f_{X|Z}(x|z)dx \\
&= \int_0^1 \frac{x}{1-x}\frac{1_{z\leq 1-x}}{\log 1/z}dx \\
&= \frac{1}{\log 1/z}\int_0^{1-z}\frac{x}{1-x}dx \\
&= \frac{-1+z-\log z}{\log 1/z} \\
&= \frac{1-z+\log z}{\log z}.
\end{aligned}$$

Finally, observe that X and $1-X$ are identically distributed, and therefore Y and Z are identically distributed. See lecture notes for Lecture 9 for the computation $E[Y] = 1/4$, $E[YZ] = 1/24$. It follows that $E[Z] = 1/4$, and thus,

$$\mathrm{cov}(Y,Z) = \frac{1}{24} - \frac{1}{4}\frac{1}{4} = -\frac{1}{48}.$$

Now we compute the variances. We have that

$$E[Y^2] = E[E[Y^2|X]] = E[(1/3)X^2] = \frac{1}{9},$$

where we used the fact that the uniform random variable on $[0,x]$ has square-expectation of $x^2/3$. Next,

$$\mathrm{var}(Y) = E[Y^2] - E[Y]^2 = \frac{1}{9} - \frac{1}{16} = \frac{7}{144}$$

Thus the correlation coefficient is

$$\rho(Y,Z) = \frac{\mathrm{cov}(Y,Z)}{\sqrt{\mathrm{var}(Y)\mathrm{var(Z)}}} = \frac{-1/48}{\sqrt{(7/144)(7/144)}} = -\frac{3}{7}.$$

Exercise 4. Assume that $X_1, \ldots, X_n$ are independent continuous random variables with common density function function f. Let $X^{(1)}, \ldots, X^{(n)}$ be the ordered statistics of $X_1, \ldots, X_n$. Namely, $X^{(1)}$ is the smallest of $X_1, \ldots, X_n$, $X^{(2)}$ is the second smallest, etc., and $X^{(n)}$ is the largest of them all. Establish that the joint distribution of $X^{(1)}, \ldots, X^{(n)}$ is given by the joint density

$$f_{X^{(1)},\ldots,X^{(n)}}(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n), \qquad x_1 < x_2 < \cdots < x_n,$$

and $f_{X^{(1)},\ldots,X^{(n)}}(x_1, \ldots, x_n) = 0$, otherwise. Use this to derive the densities for $\max_j X_j$ and $\min_j X_j$.

Solution: Define

$$g(x_1, \ldots, x_n) := n! f(x_1) f(x_2) \cdots f(x_n) 1_{x_1 < x_2 < \cdots < x_n}.$$

For $x_1 < x_2 < \cdots < x_n$

$$\begin{aligned}
&\mathbb{P}(X^{(1)} \le x_1, X^{(2)} \le x_2, \ldots, X^{(n)} \le x_n) \\
&\quad = n! P(X_1 \le x_1, X_1 < X_2 \le x_2, \ldots, X_{n-1} < X_n \le x_n), \\
&\quad = \int_{z_1 \le x_1, z_1 < z_2 \le x_2, \cdots, z_{n-1} < z_n \le x_n} n! f(z_1) f(z_2) \cdots f(z_n) \\
&\quad = \int_{z_1 \le x_1, z_2 \le x_2, \ldots, z_n \le x_n} g(z_1, \ldots, z_n),
\end{aligned}$$

where the first line follows by explicitly enumerating the $n!$ ways in which the event $\{X^{(1)} \le x_1, X^{(2)} \le x_2, \ldots, X^{(n)} \le x_n\}$ could occur; the second line follows since f is a density for each X_i, and the X_i are independent; and the third line follows by definition of g. This last equality implies that g is the joint density of the random variables $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$.

By definition, $\max_j X_j = X^{(n)}$ and the distribution for $X^{(n)}$ is obtained by integrating out the other variables. More specifically, the joint density of $X^{(2)}, \ldots, X^{(n)}$ is

$$\begin{aligned}
g(y_2, \ldots, y_n) &= \int_{-\infty}^{+\infty} g(y_1, y_2, \ldots, y_n) dy_1 \\
&= \int_{-\infty}^{y_2} n! f(y_1) f(y_2) \cdots f(y_n) 1_{y_1 < y_2 < \cdots y_n} dy_1 \\
&= n! F(y_2) f(y_2) \cdots f(y_n) 1_{y_2 < y_3 < \cdots < y_n}.
\end{aligned}$$

Similarly, the joint density of $X^{(3)}, \ldots, X^{(n)}$ is

$$\begin{aligned}
g(y_3, \ldots, y_n) &= \int_{-\infty}^{+\infty} n! F(y_2) f(y_2) \cdots f(y_n) 1_{y_2 < y_3 < \cdots y_n} dy_2 \\
&= n! \frac{1}{2} F(y_3)^2 f(y_3) f(y_4) \cdots f(y_n) 1_{y_3 < y_4 < \cdots y_n}.
\end{aligned}$$

After, $n-1$ such integrations

$$g(y_n) = \frac{n!}{(n-1)!} F(y_n)^n f(y_n) = nF(y_n)^{n-1} f(y_n).$$

Again, by definition $\min_j X_j = X^{(1)}$ and we proceed in a similar manor. The joint density of $X^{(1)}, \ldots, X^{(n-1)}$ is

$$\begin{aligned} g(y_1, \ldots, y_{n-1}) &= \int_{y_{n-1}}^{+\infty} n! f(y_1) f(y_2) \cdots f(y_n) 1_{y_1 < y_2 < \cdots y_n} dy_n \\ &= n! f(y_1) f(y_2) \cdots f(y_{n-1})[1 - F(y_{n-1})] 1_{y_1 < y_2 < \cdots y_{n-1}}, \end{aligned}$$

and of for $X^{(1)}, \ldots, X^{(n-2)}$

$$\begin{aligned} g(y_1, \ldots, y_{n-2}) &= \int_{y_{n-3}}^{+\infty} n! f(y_1) f(y_2) \cdots f(y_{n-1})[1 - F(y_{n-1})] 1_{y_1 < y_2 < \cdots y_{n-1}} dy_{n-1} \\ &= n! f(y_1) f(y_2) \cdots f(y_{n-2})[1 - F(y_{n-2})]^2 \frac{1}{2} 1_{y_1 < y_2 < \cdots y_{n-2}}. \end{aligned}$$

After $n-1$ such integrations, the resulting density is

$$g(y_1) = n[1 - F(y_1)]^{n-1} f(y_1).$$

Exercise 5. Let $X_1, \ldots, X_n$ be independent r.v. with $\mathrm{Exp}(\lambda)$ distribution. Consider $S_n = \sum_{1 \le j \le n} X_j$. The distribution of S_n is sometimes called *Erlang*.

(a) Establish that the density of S_n is $f_{S_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} \exp(-\lambda x)$.
(A Gamma distribution with an integer shape parameter n.)

(b) Consider the joint distribution of $S_1, S_2, \ldots, S_{n-1}$ given $S_n = x$. Establish that this joint distribution is the same as the joint distribution of $U^{(1)}, \ldots, U^{(n-1)}$, where $U^{(1)}, \ldots, U^{(n-1)}$ is the order statistics of $n-1$ independent r.v. with $U(0, x)$ distribution.

Solution:

(a) We proceed by induction. The case of $n = 1$ holds as this is just the exponential density. Assuming for n exponentials, the density of $n+1$ exponentials is given by convolution

$$\begin{aligned} f_{S_{n+1}}(z) &= \int_0^z f_{S_n}(t) f_{X_{n+1}}(z-t) dt \\ &= \int_0^z \frac{\lambda^n t^{n-1}}{(n-1)!} \lambda \exp(-\lambda(z-t)) dt \\ &= \frac{\lambda^{n+1} t^n}{n!} \end{aligned}$$

as desired.

(b) Let A be the event that $S_1 \leq S_2 \leq \cdots \leq S_n$. Then

$$\begin{aligned}
f_{S_1,\ldots,S_{n-1}|S_n}(s_1,\ldots,s_{n-1}|s_n) &= \frac{f_{S_1,\ldots,S_n}(s_1,\ldots,s_n)}{f_{S_n}(s_n)} \\
&= \frac{f_{X_1,\ldots,X_n}(s_1, s_2 - s_1 \ldots, s_n - s_{n-1})}{f_{S_n}(s_n)} \\
&= \frac{f_{X_1}(s_1)\cdot f_{X_2}(s_2 - s_1)\cdot\cdots\cdot f_{X_n}(s_n - s_{n-1})}{f_{S_n}(s_n)} \\
&= \frac{\lambda e^{-\lambda s_1}\cdot \lambda e^{-\lambda(s_2-s_1)}\cdot\cdots\cdot \lambda e^{-\lambda(s_n - s_{n-1})}}{\lambda^n s_n^{n-1} e^{-\lambda s_n}/(n-1)!}\mathbb{1}_A \\
&= \frac{\lambda^n e^{-\lambda s_n}}{\lambda^n s_n^{n-1} e^{-\lambda s_n}/(n-1)!}\mathbb{1}_A \\
&= \frac{(n-1)!}{s_n^{n-1}}\mathbb{1}_A.
\end{aligned}$$

Observing that $U(0, s_n)$ have density $f_U(u) = 1/s_n \mathbb{1}_{u\in(0,s_n)}$, we compute that

$$f_{U^{(1)},\ldots,U^{(n-1)}}(u_1,\ldots,u_{n-1}) = (n-1)! f_U^{n-1}(u)\mathbb{1}_A = \frac{(n-1)!}{s_n^{n-1}}\mathbb{1}_A,$$

giving the result.

Exercise 6. A needle of length $2s < 1$ unit is randomly tossed onto a quad-ruled sheet with horizontal and vertical lines spaced at 1 unit. Assuming the position and the angle of the needle are independent and uniform, find the average number of lines the needle intersects.

Solution: Consider the unit square $[0,1]\times[0,1]$ in the x, y plane. The center of the needle will be uniformly distributed within this square. Let X and Y be uniform random variables for the x and y coordinates, respectively. Moreover, let Θ be the angle the needles makes with the x-axis, where Θ is uniformly distributed between $[0, 2\pi]$. By symmetry it suffices to consider the angle uniformly distributed between $[0, \pi/2]$. Furthermore, as the $2s < 1$, the needle may not cross both horizontal and vertical lines simultaneously. Crossing one of these lines requires that the center is beyond either the horizontal or vertical midpoint. Therefore, by symmetry, it suffices to consider the needle distributed in one of the quadrants, e.g. the $[0, 1/2]\times[0, 1/2]$. By independence the resulting joint

PDF is

$$f_{X,Y,\Theta}(x,y,\theta) = \begin{cases} 2 \cdot 2 \cdot \frac{2}{\pi} & 0 \le x, y \le \frac{1}{2},\ 0 \le \theta \le \frac{\pi}{2} \\ 0 & \text{else} \end{cases}$$
$$= \begin{cases} \frac{8}{\pi} & 0 \le x, y \le \frac{1}{2},\ 0 \le \theta \le \frac{\pi}{2} \\ 0 & \text{else} \end{cases}.$$

The needle will cross the x-axis if $s \sin \Theta \ge X$ and similarly the needle will cross the y-axis if $s \cos \Theta \ge Y$. Therefore, the expected number of lines crossed is

$$\begin{aligned}
&E\left[\mathbb{1}\{X \le s\sin\Theta\} + \mathbb{1}\{Y \le s\cos\Theta\}\right] \\
&\quad = \int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} \left[\mathbb{1}\{x \le s\sin\theta\} + \mathbb{1}\{x \le s\cos\theta\}\right] \frac{8}{\pi}\, dx\, dy\, d\theta \\
&\quad = \frac{8}{\pi}\left[\frac{1}{2}\int_0^{\frac{\pi}{2}} \int_0^{s\sin\theta} dx\, d\theta + \frac{1}{2}\int_0^{\frac{\pi}{2}} \int_0^{s\cos\theta} dy\, d\theta\right] \\
&\quad = \frac{4}{\pi}\left[\int_0^{\frac{\pi}{2}} s\sin\theta\, d\theta + \int_0^{\frac{\pi}{2}} s\cos\theta\, d\theta\right] \\
&\quad = \frac{8s}{\pi}.
\end{aligned}$$

Hence the expected number of lines cross is $\frac{8s}{\pi}$.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 7

Readings:

Notes from Lectures 11-13.

[GS], Section 4.1-4.8 **and** 5.1-5.2

[Cinlar], Chapter IV.

Exercise 1. (Continuous-discrete Bayes rule) Let K be the number of heads obtained in six (conditionally) independent coins of a biased coin whose probability of heads is itself a random variable Z, uniformly distributed over $[0, 1]$. Find the conditional PDF of Z given K, and calculate $\mathbb{E}[Z \mid K = 2]$. You can use the following formula,

$$\int_0^1 y^\alpha (1-y)^\beta \, dy = \frac{\alpha! \, \beta!}{(\alpha+\beta+1)!},$$

known to be valid for positive integer α and β.

Solution: We have $\mathbb{P}(K = 2 \mid Z = z) = cz^2(1-z)^4$, where c is a normalizing constant. Using Bayes' rule, we have

$$f_Z(z|K=2) = \frac{P(K=2|Z=z)f_Z(z)}{P(K=2)} = \frac{z^2(1-z)^4}{\int_0^1 t^2(1-t)^4 \, dt} 1_{z\in[0,1]}$$

Thus,

$$\mathbb{E}[Z \mid K=2] = \frac{\int_0^1 z^3(1-z)^4 \, dz}{\int_0^1 t^2(1-t)^4 \, dt} = \frac{3}{8}.$$

Exercise 2. Let X and Y be independent exponential random variables with parameter 1. Find the joint density function of $U = X+Y$ and $V = X/(X+Y)$, and deduce that V is uniformly distributed on $[0, 1]$.

Solution: The transformation $x = uv$, $y = u - uv$ has Jacobian

$$J = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u.$$

Therefore, we have $|J| = |u|$, and thus $f_{U,V}(u, v) = ue^{-u}$ for $0 \leq u < \infty$, and $0 \leq v \leq 1$. Integrating with respect to u we see that we have $f_V(v) = 1$, and also that U, V are independent.

Exercise 3. A point (X, Y) is picked at random uniformly in the unit circle. Find the joint density of R and X, where $R^2 = X^2 + Y^2$.

Solution: We can make a change of variables, and use the Jacobian. We can also just compute this directly, as above, by finding the distribution function and differentiating. Using the convention that $\sqrt{r^2 - u^2} = 0$ when the argument of the square root becomes negative, we have

$$\begin{aligned} F(r, x) &= \mathbb{P}(R \leq r, X \leq x) = \frac{2}{\pi} \int_{-r}^{x} \sqrt{r^2 - u^2}\, du, \\ f(r, x) &= \frac{\partial^2 F}{\partial r \partial x} = \frac{2r}{\pi\sqrt{r^2 - x^2}}, \quad |x| < r < 1. \end{aligned}$$

Exercise 4. Let X_1, X_2, X_3 be independent random variables, uniformly distributed on $[0, 1]$.

a. What is the probability that three rods of lengths X_1, X_2, X_3 can be used to make a triangle? (That is, that the largest one is smaller than the sum of the other two.)

b. What is the probability distribution of the second largest X_k, i.e. $X^{(2)}$.

Solution:

a. Let $M = \max\{X_1, X_2, X_3\}$. The lengths X_1, X_2, X_3 form a triangle iff $M \leq X_1 + X_2 + X_3 - M$, i.e., the sum of any two sides is at least that of the third side. By symmetry, the probability that $M = X_i$ is the same for all

i, hence we have

$$
\begin{aligned}
\mathbb{P}(X_1, X_2, X_3 \text{ forms a triangle}) &= 3\mathbb{P}(X_1 \leq X_2 + X_3, X_2 \leq X_1, X_3 \leq X_1) \\
&= 3 \int_0^1 \int_{\{x_2+x_3 \geq x_1, x_2 \leq x_1, x_3 \leq x_1\}} dx_2 dx_3 dx_1 \\
&= 3 \int_0^1 \frac{x_1^2}{2} dx_1 \\
&= \frac{1}{2}.
\end{aligned}
$$

b. The joint PDF for the order statistics is

$$f(x^{(1)}, x^{(2)}, x^{(3)}) = n! f(x^{(1)}) f(x^{(2)}) f(x^{(3)}) \mathbb{1}_{x^{(1)} < x^{(2)} < x^{(3)}}.$$

Integrating out $x^{(1)}$ and $x^{(3)}$

$$
\begin{aligned}
f(x^{(2)}) &= \int_{\mathbb{R}} \int_{\mathbb{R}} 3! f(x^{(1)}) f(x^{(2)}) f(x^{(3)}) \mathbb{1}_{x^{(1)} < x^{(2)} < x^{(3)}} \, dx^{(1)} \, dx^{(3)} \\
&= 3! f(x^{(2)}) \left(\int_{x^{(2)}}^{\infty} f(x^{(1)}) \, dx^{(1)} \right) \left(\int_{-\infty}^{x^{(2)}} f(x^{(3)}) \, dx^{(3)} \right) \\
&= 3! f(x^{(2)})(1 - F(x^{(2)})) F(x^{(2)}).
\end{aligned}
$$

In particular, for X_k uniform

$$f(x^{(2)}) = 3! x^{(2)} (1 - x^{(2)}) \mathbb{1}_{[0,1]}(x^{(2)}).$$

Exercise 5. A stick is broken, at a location chosen uniformly at random. Find the average ratio of the lengths of the smaller and larger pieces.

Solution: WLOG assume the stick has unit length and by symmetry assume the small piece is distributed uniformly on $[0, \frac{1}{2}]$ with PDF

$$f_S(s) = \begin{cases} 2 & 0 \leq s \leq 1/2 \\ 0 & \text{else} \end{cases}.$$

Let $g : [0, 1/2] \to [0, 1]$, $g(x) = x/(1-x)$. This function has a well defined inverse $g^{-1}(x) = x/(1+x)$ and derivative $(g^{-1})'(x) = (1+x)^{-2}$. Let $X =$

$g(S)$, the ratio of the small to large piece. Using the formula from lecture 12, the resulting PDF is

$$
\begin{aligned}
f_X(x) &= f_S(g^{-1}(x))\frac{1}{|g'(g^{-1}(x))|} \\
&= f_S\left(\frac{x}{1+x}\right)|(g^{-1})'(x)| \\
&= 2(1+x)^{-2}\mathbb{1}_{[0,1]}(x).
\end{aligned}
$$

Therefore, the resulting expected ratio is

$$
E[X] = \int_0^1 x2\,(1+x)^{-2}\;dx = \log(4) - 1.
$$

Exercise 6. Let $X \sim \Gamma(a,c)$, $U, V \sim \Gamma(a, \sqrt{2c})$ and $Y \sim \mathcal{N}(0,1)$, all jointly independent. Compare the distribution of $U - V$ and $\sqrt{X}Y$. (*Hint:* compute MGFs using conditional expectation).

Solution: Let $N \sim \mathcal{N}(\mu, \sigma^2)$ and $G \sim \Gamma(a,c)$ there respective moment generating functions are

$$
\begin{aligned}
M_N(s) &= \exp\left(\mu s + \tfrac{\sigma^2}{2}s^2\right) \\
M_G(s) &= \int_0^\infty e^{sx}\frac{c^a x^{a-1}e^{-cx}}{\Gamma(a)}\,dx \\
&= \frac{c^a}{\Gamma(a)}\int_0^\infty x^{a-1}e^{-(c-s)x}\,dx \\
&= \quad 1-\frac{s}{c} \quad ^{-a}\;\frac{1}{\Gamma(a)}\int_0^\infty t^{a-1}e^{-t}\,dt \quad (c-s>0) \\
&= \quad 1-\frac{s}{c} \quad ^{-a} \qquad (c-s>0).
\end{aligned}
$$

Conditional on $X = x$, $\sqrt{x}Y \sim \mathcal{N}(0,x)$. Therefore, the MGF for $\sqrt{X}Y$ is

$$
\begin{aligned}
M_{\sqrt{X}Y}(s) &= E\left[\exp\left(s\sqrt{X}Y\right)\right] \\
&= E\left[E\left[\exp\left(s\sqrt{x}Y\right) \mid X = x\right]\right] \\
&= E\left[\exp\left(\frac{X}{2}s^2\right)\right] \\
&= E\left[\exp\left(\frac{s^2}{2}X\right)\right] \\
&= \left(1-\frac{s^2}{2c}\right)^{-a} \qquad \left(c - s^2/2 > 0\right).
\end{aligned}
$$

Similarly, the MGF for $U - V$ is

$$\begin{aligned}
M_{U-V}(s) &= E\left[\exp\left(s(U-V)\right)\right] \\
&= E\left[E\left[\exp\left(s(U-v)\right) \mid V=v\right]\right] \\
&= E\left[e^{-sv}E\left[\exp\left(sU\right) \mid V=v\right]\right] \\
&= E\left[e^{-sv}\left(1-\frac{s}{\sqrt{2c}}\right)^{-a}\right] \quad \left(\sqrt{2c}-s>0\right) \\
&= \left(1-\frac{s}{\sqrt{2c}}\right)^{-a} E\left[e^{-sv}\right] \quad \left(c-s^2/2>0\right) \\
&= \left(1-\frac{s}{\sqrt{2c}}\right)^{-a}\left(1-\frac{-s}{\sqrt{2c}}\right)^{-a} \quad \left(c-s^2/2>0\right. \\
&= \left(1-\frac{s^2}{2c}\right)^{-a} \quad \left(c-s^2/2>0\right).
\end{aligned}$$

Hence $\sqrt{X}Y$ and $U - V$ have the same MGF and thusly, the same distribution.

Exercise 7. Let $X, Y \sim \Gamma(1, c)$ be independent and $Z = X + Y$. Describe conditional distribution $P_{Y|Z}$. (Ideally, you want to describe it as a Markov kernel $K(z, dy)$, however, full credit will be given for just specifying the conditional pdf or cdf).

Solution: The PDF for a Gamma random variable $U \sim \Gamma(a, c)$ is

$$f_U(u) = \frac{1}{\Gamma(a)}c^a u^{a-1}e^{-cu}.$$

Moreover, for two random variables with the same scale parameter the shape parameters are additive. In particular, $Z = X + Y \sim \Gamma(2, c)$. Thus,

$$\begin{aligned}
f_Z(z) &= \frac{1}{\Gamma(2)}c^2 z^{2-1}e^{-cz}\mathbb{1}_{[0,\infty)}(z) = c^2 z e^{-cz}\mathbb{1}_{[0,\infty)}(z) \\
f_X(t) = f_Y(t) &= \frac{1}{\Gamma(1)}ce^{-ct} = ce^{-ct}\mathbb{1}_{[0,\infty)}(t).
\end{aligned}$$

The conditional distribution for $Y|Z$ is

$$\begin{aligned}
f_{Y|Z}(y \mid z) &= \frac{f_{Z|Y}(z \mid y) f_Y(y)}{f_Z(z)} \\
&= \frac{f_X(z-y) f_Y(y)}{f_Z(z)} \\
&= \frac{ce^{-c(z-y)} ce^{-cy}}{c^2 z e^{-cz}} \mathbb{1}_{[0,\infty)}(z-y) \mathbb{1}_{[0,\infty)}(y) \mathbb{1}_{[0,\infty)}(z) \\
&= \frac{1}{z} \mathbb{1}_{[0,\infty)}(y) \mathbb{1}_{[y,\infty)}(z).
\end{aligned}$$

Hence, the corresponding Markov Kernel is

$$K(z, dy) = f_{Y|Z}(y \mid z)\, dy = \frac{1}{z} \mathbb{1}_{[0,\infty)}(y) \mathbb{1}_{[y,\infty)}(z)\, dy.$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Problem Set 8

Readings:
Notes from Lecture 14 and 15.
[GS]: Section 4.9, 4.10, 5.7-5.9

Exercise 1. Let $\phi_A(t) = \mathbb{E}[e^{itA}]$ be a characteristic function of r.v. A.

(a) Find $\phi_X(t)$ if X is a Bernoulli(p) random variable.

(b) Suppose that $\phi_{X_n} = \cos(t/2^n)$. What is the distribution of X_n?

(c) Let $X_1, X_2, \ldots$ be independent and let $S_n = X_1 + \cdots + X_n$. Suppose that S_n converges almost surely to some random variable S. Show that $\phi_S(t) = \prod_{i=1}^{\infty} \phi_{X_i}(t)$.

(d) Evaluate the infinite product $\prod_{n=1}^{\infty} \cos(t/2^n)$. *Hint:* Think probabilistically; the answer is a very simple expression.

Solution:

(a)
$$\phi_X(t) = (1-p) + pe^{it}.$$

(b) Since
$$\phi_{X_n} = \frac{e^{it/2^n} + e^{it/2^n}}{2},$$
X_n has to be $1/2^n$ with probability $1/2$, and $-1/2^n$ with probability $1/2$.

(c) The interchange of limit and expectation can be justified by appealing to the dominated convergence theorem, since $|e^{itS}| = 1$ and $|e^{itS_n}| = 1$.

(d) The sum of the random variables X_n approaches a uniform random variable in $[-1, 1]$ almost surely. By part (c), the product of $\cos t/2^n$ is the characteristic function of $U[-1, 1]$, which is
$$\int_{-1}^{1} e^{itx} \frac{1}{2} dx = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t},$$
so
$$\prod_{n=1}^{\infty} \cos(\frac{t}{2^n}) = \frac{\sin t}{t}.$$

Exercise 2. Let X be a random variable with mean, variance, and moment generating function), denoted by $\mathbb{E}[X]$, $\text{var}(X)$, and $M_X(s)$, respectively. Similarly, let Y be a random variable associated with $\mathbb{E}[Y]$, $\text{var}(Y)$, and $M_Y(s)$. Each part of this problem introduces a new random variable Q, H, G, D. Determine the means and variances of the new random variables, in terms of the means, and variances of X and Y.

(a) $M_Q(s) = [M_X(s)]^5$.

(b) $M_H(s) = [M_X(s)]^3[M_Y(s)]^2$.

(c) $M_G(s) = e^{6s}M_X(s)$.

(d) $M_D(s) = M_X(6s)$.

Solution:

(a) The random variable Q is the sum of 5 independent random variables, each distributed as X. Thus, $\mathbb{E}[Q] = 5\mathbb{E}[X]$, $\text{var}(Q) = 5\text{var}(X)$.

(b) The random variable H is the sum of three independent random variables distributed as X, and another two independent random variables distributed as Y. Thus, $\mathbb{E}[H] = 3\mathbb{E}[X] + 2\mathbb{E}[Y]$, $\text{var}(H) = 3\text{var}(X) + 2\text{var}(Y)$.

(c) Multiplying a transform by e^{sa} corresponds to adding a to a random variable. Thus, $\mathbb{E}[G] = \mathbb{E}[X] + 6$, $\text{var}(G) = \text{var}(X)$.

(d) Replacing s by sa corresponds to replacing a random variable X by aX. Thus, $\mathbb{E}[D] = 6\mathbb{E}[X]$, $\text{var}(D) = 36\text{var}(X)$.

Exercise 3. A random (nonnegative integer) number of people K, enter a restaurant with n tables. Each person is equally likely to sit on any one of the tables, independently of where the others are sitting. Give a formula, in terms of the moment generating function $M_K(\cdot)$, for the expected number of occupied tables (i.e., tables with at least one customer).

Solution: Let D be the number of occupied tables. Let $X_1, \ldots, X_n$ be the respective indicator variables of each table, that is, $X_i = 1$ if there is at least one person at at table i, and $X_i = 0$ otherwise. Note that $D = X_1 + \cdots + X_n$. Thus we have:

$$\begin{aligned}
\mathbb{E}[D] &= \mathbb{E}[\mathbb{E}[D|K]] \\
&= \mathbb{E}[\mathbb{E}[X_1 + \cdots + X_n|K]] \\
&= n \cdot \mathbb{E}[\mathbb{E}[X_i|K] \\
&= n \cdot \mathbb{E}\Big[1 - \Big(\frac{n-1}{n}\Big)^K\Big] \\
&= n - n \cdot \mathbb{E}\Big[\Big(\frac{n-1}{n}\Big)^K\Big] \\
(\text{letting } s = \log((n-1)/n)) &= n - n \cdot \mathbb{E}[e^{sK}] \\
&= n - n \cdot M_K\big(\log((n-1)/n)\big).
\end{aligned}$$

Exercise 4. (Problem 7, Section 4.9, [GS]): Let the vector $X_r, 1 \le r \le n$ have a multivariate normal distribution with zero means and covariance matrix $V = (v_{ij})$. Show that, conditional on the event $\sum_{i=1}^n X_r = x$, $X_1 \stackrel{d}{=} N(a,b)$, where $a = (\rho s/t)x, b = s^2(1-\rho^2)$ and $s^2 = v_{11}, t^2 = \sum_{ij} v_{ij}, \rho = \sum_i v_{i1}/(st)$.

Solution: Let $S_n = \sum_{k=1}^n x_k$. Since the mapping $(X_1, X_2, \ldots, X_n) \to (X_1, S_n)$ is a linear mapping, and the family of multivariate normal distributions is closed under linear mappings, we find that (X_1, S_n) is a bivariate normal distribution. Furthermore, $E[X_1] = 0$, $E[S_n] = 0$, $\text{var}(X_1) = v_{1,1} = s^2$, var $(S_n) = \sum_{i,j} E[X_i X_j] = \sum_{i,j} v_{i,j} = t^2$, and cov $(X_1, S_n) = \sum_{k=1}^n v_{1,k}$. It follows from the definitions that the correlation of X_1 and S_n is $\left(\sum_{k=1}^n v_{1,k}\right)/(st)$. The desired result follows from the basic properties of bivariate normals proven in the lecture notes.

Exercise 5. Suppose that for every k, the pair (X_k, Y) has a bivariate normal distribution. Furthermore, suppose that the sequence X_k converges to X, almost surely. Show that (X, Y) has a bivariate normal distribution. *Hint:* First show that if X_k is a sequence of normally distributed random variables which converges to X almost surely, then X has to be normally distributed as well. Then use the "right" definition of the bivariate normal.

Solution: For any $a, b \in \Re$, $aX_k + bY \xrightarrow{as} aX + bY \;\Rightarrow\; aX_k + bY \xrightarrow{d} aX + bY$. Let $Z_k = aX_k + bY$ be a sequence of normal random variables with variance σ_k^2. As X_k and Y have zero mean, their corresponding characteristic functions are

$$\phi_k(t) = \exp\left(-\frac{\sigma_k^2}{2}t^2\right).$$

Suppose, for all t, $\phi_k(t) \to \phi(t)$ for some continuous function $\phi(t)$. Consider the sequence $\phi_k(1) = \exp\left(-\frac{\sigma_k^2}{2}\right)$. If this sequence convergences then the composition with any continuous function converges. In particular, $-2\log\phi_k(1) = \sigma_k^2$ converges.

Suppose $\sigma_k^2 \to \infty$, then

$$\phi(t) = \begin{cases} 1 & t = 0 \\ 0 & \text{else} \end{cases}.$$

By the inversion theorem, the PDF of the resulting random variable is one over the entire real line, but this does not integrate to one, a contradiction. Therefore, the $\sigma_k^2 \to \sigma^2$ for some $\sigma^2 \in [0, \infty)$ and $Z = aX + bY$ is normal, i.e. $aX + bY \sim \mathcal{N}(0, \sigma^2)$. Hence, (X, Y) has a bivariate normal distribution.

Exercise 6. Suppose that $X, Z_1, \ldots, Z_n$ have a multivariate normal distribution, and X has zero mean. Furthermore, suppose that $Z_1, \ldots, Z_n$ are independent. Show that $\mathbb{E}[X \mid Z_1, \ldots, Z_n] = \sum_{i=1}^n \mathbb{E}[X \mid Z_i]$. Is this result true without the multivariate normal example? (Prove or give a counterexample.)

Solution: Solution: Let $Z = (Z_1, \ldots, Z_n)$, the multivariate normal. By the conditional expectation formula for multivariate normals

$$\mathbb{E}[X \mid Z_1, \ldots, Z_n] = V_{XZ} V_{ZZ}^{-1} (Z - \mu_Z)$$

Note that V_{XZ} is a $1 \times n$ row vector, V_{ZZ} is an $n \times n$ matrix, and $Z - \mu_Z$ is an $n \times 1$ column vector. Now, by independence of Z_i, V_{ZZ} is the diagonal matrix whose i'th diagonal entry is $\text{var}(Z_i)$. Thus, we can rewrite the right hand side above as,

$$\begin{aligned} \mathbb{E}[X \mid Z_1, \ldots, Z_n] &= \sum_{i=1}^n V_{XZ_i}(Z_i - \mu_{Z_i})/\text{var}(Z_i) \\ &= \sum_{i=1}^n E[X|Z_i]. \end{aligned}$$

Exercise 7. Let $Y_1, \ldots, Y_n$ be independent N(0,1) random variables, and let $X_j = \sum_{r=1}^n c_{jr} Y_r$, for some constants c_{jr}. Show that

$$\mathbb{E}[X_j \mid X_k] = \left(\frac{\sum_r c_{jr} c_{kr}}{_r\, c_{kr}^2} \right) X_k.$$

Solution: If (X, Y) are jointly normal with means μ_X, μ_Y respectively, and cov$(X, Y) = V_{XY}$, and var$(Y) = V_{YY}$, then $\mathbb{E}[X|Y] = \mu_X + V_{XY}V_{YY}^{-1}(Y - \mu_Y)$. Note that in our case, since the Y_r are independent with zero mean, the X_i are also zero mean. Then, we have:

$$
\begin{aligned}
\mathbb{E}[X_j|X_k] &= \mu_{X_j} + V_{jk}V_{kk}^{-1}(X_k - \mu_2) \\
&= \mathbb{E}[X_jX_k]\mathbb{E}[X_kX_k]^{-1}X_k \\
&= \left(\frac{{}^{n}_{r_1,r_2=1}\, c_{jr_1}c_{kr_2}\mathbb{E}[Y_{r_1}Y_{r_2}]}{\sum_{r=1}^{n} c_{kr}^2\mathbb{E}[Y_r 2] + {}_{r_1 \neq r_2}\, c_{kr_1}c_{kr_2}\mathbb{E}[Y_{r_1}Y_{r_2}]} \right) X_k \\
&= \left(\frac{\sum_r c_{jr}c_{kr}}{{}_r\, c_{kr}^2} \right) X_k.
\end{aligned}
$$

Exercise 8. [Optional, not for grade] Let X, Y be i.i.d. with finite second moments. Suppose that $X+Y$ and $X-Y$ are independent. Show that they must be Gaussian. (*Hint:* Derive a second order differential equation on $\phi_X(t)$.)

Solution: Let $\phi(t) = \phi_X(t) = \phi_Y(t)$. Using both independence relations and properties of characteristic functions

$$\begin{aligned}
\phi((a+b)t)\phi((a-b)t) &= \phi_{(a+b)X}(t)\phi_{(a-b)Y}(t) \\
&= \phi_{(a+b)X+(a-b)Y}(t) \quad (X \perp\!\!\!\perp Y) \\
&= \phi_{a(X+Y)+b(X-Y)}(t) \\
&= \phi_{a(X+Y)}(t)\phi_{b(X-Y)}(t) \quad (X+Y \perp\!\!\!\perp X-Y) \\
&= \phi_{aX}(t)\phi_{aY}(t)\phi_{bX}(t)\phi_{-bY}(t) \quad (X \perp\!\!\!\perp Y) \\
&= \phi(at)^2\phi(bt)\phi(-bt).
\end{aligned}$$

In other words, as a and b were arbitrary,

$$\phi(t+u)\phi(t-u) = \phi(t)^2\phi(u)\phi(-u).$$

By assumption X and Y have finite second moment and thus ϕ is twice continuously differentiable. Differentiating both sides with respect to u

$$\phi'(t+u)\phi(t-u) - \phi(t+u)\phi'(t-u) = \phi(t)^2\left[\phi'(u)\phi(-u) - \phi(u)\phi'(-u)\right],$$

and

$$\begin{aligned}
&\phi''(t+u)\phi(t-u) - \phi'(t+u)\phi'(t-u) - \phi'(t+u)\phi'(t-u) + \phi(t+u)\phi''(t-u) \\
&= \phi(t)^2\left[\phi''(u)\phi(-u) - \phi'(u)\phi'(-u) - \phi'(u)\phi'(-u) + \phi(u)\phi''(-u)\right].
\end{aligned}$$

Evaluating at $u = 0$

$$\begin{aligned}
2\left(\phi''(t)\phi(t) - \phi'(t)^2\right) &= \phi(t)^2 2\left(\phi''(0)\phi(0) - \phi'(0)^2\right) \\
&= \phi(t)^2 2\left(\phi''(0) - \phi'(0)^2\right).
\end{aligned}$$

The resulting differential equation is

$$\phi''(t)\phi(t) - \phi'(t)^2 - (\phi''(0) - \phi'(0)^2)\phi(t)^2 = 0.$$

Consider the test function

$$f(t) = e^{ic_1 t + c_2 t^2}.$$

The first and second derivate for f are

$$f'(t) = (ic_1 + 2c_2 t)e^{ic_1 t + c_2 t^2}$$
$$f''(t) = 2c_2 e^{ic_1 t + c_2 t^2} + (ic_1 + 2c_2 t)^2 e^{ic_1 t + c_2 t^2}$$

and evaluating at zero

$$f'(0) = ic_1 \quad f''(0) = 2c_2 - c_1^2.$$

Therefore

$$\begin{aligned} &f''(t)f(t) - f'(t)^2 - \left(f''(0) - f'(0)^2\right) f(t)^2 \\ &= \left(2c_2 e^{ic_1 t + c_2 t^2} + (ic_1 + 2c_2 t)^2 e^{ic_1 t + c_2 t^2}\right) e^{ic_1 t + c_2 t^2} \\ &\quad - (ic_1 + 2c_2 t)^2 e^{2(ic_1 t + c_2 t^2)} - (2c_2 - c_1^2 + c_1^2) e^{2(ic_1 t + c_2 t^2)} \\ &= \left(2c_2 + (ic_1 + 2c_2 t)^2 - (ic_1 + 2c_2 t)^2 - 2c_2\right) e^{2(ic_1 t + c_2 t^2)} \\ &= 0. \end{aligned}$$

Hence $f(t)$ satisfies the differential equation and

$$\phi(t) = e^{ic_1 t + c_2 t^2}.$$

From properties of characteristic functions

$$\phi'(0) = iE[X] \quad \phi''(0) = -E[X^2] = -\text{var}(X) - E[X]^2.$$

Thus

$$c_1 = E[X] \quad c_2 = -\text{var}(X)/2.$$

Therefore,

$$\phi(t) = e^{iE[X]t - \frac{\text{var}(X)}{2} t^2}$$

and by the inversion theorem X and Y are Gaussion.

Exercise 9. [Optional, not for grade] (Problem 20 in p. 142, Section 4.14 of [GS]): Suppose that X and Y are independent and identically distributed, and not necessarily continuous random variables. Show that $X + Y$ cannot be uniformly distributed on $[0, 1]$.

Solution: Suppose X and Y are i.i.d. and $Z = X + Y$ is uniform on $[0,1]$. Then $X, Y \in [0, 1/2]$ almost everywhere, or else $Z \notin [0,1]$ with positive probability. Let $z \in [0, 1/4]$. Then, as $X, Y \geq 0$ a.e.,

$$z = \mathbb{P}(Z \leq z) \leq \mathbb{P}(X \leq z, Y \leq z) = \mathbb{P}(X \leq z)\mathbb{P}(Y \leq z) = \mathbb{P}(X \leq z)^2,$$

and similarly, as $X, Y \leq 1/2$ a.e.,

$$z = \mathbb{P}(Z > 1 - z) \leq \mathbb{P}(X > 1/2 - z, Y > 1/2 - z) \leq \mathbb{P}(X > 1/2 - z)^2.$$

Combining provides

$$\mathbb{P}(X \leq z),\, \mathbb{P}\left(X > 1/2 - z\right) \geq \sqrt{z}. \tag{1}$$

Moreover, noting both the upper and lower a.e. bounds of X and Y

$$\begin{aligned}
2z &= \mathbb{P}\left(1/2 - z < Z \leq 1/2 + z\right) \\
&\geq \mathbb{P}\left(X > 1/2 - z, Y \leq z\right) + \mathbb{P}\left(Y > 1/2 - z, X \leq z\right) \\
&= 2\mathbb{P}\left(X > 1/2 - z\right)\mathbb{P}(X \leq z) \\
&\geq 2\sqrt{z}\sqrt{z} \\
&= 2z.
\end{aligned}$$

Thus all inequalities are equalities, and in particular

$$\mathbb{P}\left(X > 1/2\right)\mathbb{P}(X \leq z) = z,$$

and, in conjunction with (1), this implies that (1) holds with equality, namely

$$\mathbb{P}(X \leq z) = \mathbb{P}\left(X > \frac{1}{2} - z\right) = \sqrt{z}. \tag{2}$$

Therefore,

$$\begin{aligned}
\mathbb{P}\left(X, Y \leq \frac{1}{8}, X + Y > \frac{1}{8}\right) &\geq \mathbb{P}\left(\frac{1}{16} < X \leq \frac{1}{8}, \frac{1}{16} < Y \leq \frac{1}{8}\right) \\
&= \mathbb{P}\left(\frac{1}{16} < X \leq \frac{1}{8}\right)^2 \\
&= \left(\mathbb{P}\quad \frac{1}{8}\right) - \mathbb{P}\left(\frac{1}{16}\right)\Bigg)^2 \\
&= \left(\sqrt{\frac{1}{8}} - \sqrt{\frac{1}{16}}\right)^2 \\
&> 0.
\end{aligned}$$

This provides

$$\begin{aligned}
\frac{1}{8} &= \mathbb{P}\left(Z \leq \frac{1}{8}\right) \\
&= \mathbb{P}\left(X \leq \frac{1}{8}, Y \leq \frac{1}{8}\right) - \mathbb{P}\left(X, Y \leq \frac{1}{8}, X + Y > \frac{1}{8}\right) \\
&< \mathbb{P}\left(X \leq \frac{1}{8}, Y \leq \frac{1}{8}\right) = \frac{1}{8},
\end{aligned}$$

a contradiction.

6.436J / 15.085J Fundamentals of Probability
Fall 2018

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 9

Readings:

Notes from Lectures 16-19.

[GS], Section 7.1-7.6

[Cinlar], Chapter III

Exercise 1. We study convergence of algebraic operations:

(a) Show that

$$X_n \xrightarrow{\text{i.p.}} X, Y_n \xrightarrow{\text{i.p.}} Y \quad \Rightarrow \quad X_n Y_n \xrightarrow{\text{i.p.}} XY$$

(*Hint:* reduce to $\xrightarrow{\text{a.s.}}$.)

(b) Show, however, that

$$X_n \xrightarrow{\text{d}} X, Y_n \xrightarrow{\text{d}} Y \quad \not\Rightarrow \quad X_n Y_n \xrightarrow{\text{d}} XY$$

(c) Assume $X_n \perp\!\!\!\perp Y_n$ and $X \perp\!\!\!\perp Y$. Show that then

$$X_n \xrightarrow{\text{d}} X, Y_n \xrightarrow{\text{d}} Y \quad \Rightarrow \quad X_n Y_n \xrightarrow{\text{d}} XY$$

(*Hint:* reduce to $\xrightarrow{\text{a.s.}}$.)

Solution:

(a) Writing the product as a sum of squares

$$X_n Y_n = \frac{(X_n + Y_n)^2 - (X_n - Y_n)^2}{4}.$$

Therefore, by the triangle inequality,

$$|X_n Y_n| \leq \frac{|X_n + Y_n|^2 + |X_n - Y_n|^2}{4}.$$

Thus

$$\mathbb{P}\left(|X_n Y_n| > \varepsilon\right) \leq \left(\frac{|X_n + Y_n|^2}{4} > \frac{\varepsilon}{2}\right) + \left(\frac{|X_n - Y_n|^2}{4} > \frac{\varepsilon}{2}\right),$$

and it suffices to show that convergence in probability is closed under scalar multiplication, addition and squares. Addition is given in the lecture notes, scalar multiplication follows since, for all $c \geq 0$,

$$\mathbb{P}\left(c|X| > \varepsilon\right) \mathbb{P}\left(|X| > \varepsilon/c\right),$$

and squaring follows since

$$\mathbb{P} \quad |X|^2 > \varepsilon \quad = \mathbb{P} \quad |X| > \sqrt{\varepsilon} \quad ,$$

i.e. by choose ε appropriately.

(b) Consider the probability space $([0,1], \mathcal{B}, \lambda)$. Let

$$X_n = \begin{cases} \mathbb{1}_{[0,1/2]} & n \text{ odd} \\ \mathbb{1}_{(1/2,1]} & n \text{ even} \end{cases} \qquad Y_n = \begin{cases} \mathbb{1}_{[0,1/2]} & n \text{ even} \\ \mathbb{1}_{(1/2,1]} & n \text{ odd} \end{cases}.$$

Then for all n, X_n and Y_n are Bernoulli $1/2$ random variables and thusly converge in distribution to a Bernoulli $1/2$ random variable. However, $X_n Y_n \equiv 0$ and the product of Bernoulli random variables is not identically zero.

(c) As $X_n \to X$ and $Y_n \to Y$

$$\begin{aligned}
\lim_{n\to\infty} F_{X_n,Y_n}(x,y) &= \lim_{n\to\infty} F_{X_n}(x) F_{Y_n}(y) \quad \text{(Independence)} \\
&= \left(\lim_{n\to\infty} F_{X_n}(x)\right)\left(\lim_{n\to\infty} F_{Y_n}(y)\right) \quad \text{(Independence)} \\
&=^{(a)} F_X(x) F_Y(y) \\
&= F_{X,Y}(x,y),
\end{aligned}$$

where (a) holds at all continuity points of F_X and F_Y or equivalent of $F_{X,Y}$. Therefore, $(X_n, Y_n) \to (X,Y)$ in distribution. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function and q_n, q and p_n, p be the quantile functions for the $\{X_n\}$, X, the $\{Y_n\}$ and Y respectively, i.e. $q_n \to q$ almost everywhere $p_n \to p$ almost everywhere and

$$X_n \sim q_n(U) \quad X \sim q(U) \qquad Y_n \sim p_n(V) \quad Y \sim p(V)$$

for independent random variables U and V. Therefore, $q_n(U)p_n(V) \to$

$q(U)p(V)$ almost everywhere and by the BCT

$$\begin{aligned}
\lim_{n\to\infty} E\left[f(X_nY_n)\right] &= \lim_{n\to\infty} E\left[f(q_n(U)p_n(V))\right] \\
&= E\left[\lim_{n\to\infty} f(q_n(U)p_n(V))\right] \quad \text{(BCT)} \\
&= E\left[f(q(U)p(V))\right] \quad (f \text{ continuous }) \\
&= E\left[f(XY)\right].
\end{aligned}$$

Hence $X_nY_n \to XY$ in distribution.

Exercise 2 (Metrization of convergence in probability)**.** Define a pseudo-metric on the space of random-variables:

$$d(X,Y) \triangleq \mathbb{E} \quad \frac{|X-Y|}{1+|X-Y|} \quad .$$

Show $X_n \overset{\text{i.p.}}{\to} X$ iff $d(X_n, X) \to 0$.

Solution: Let $\varepsilon > 0$ and WLOG assume $\varepsilon < 1$. Then,

$$\{|X-Y| \geq \varepsilon\} = \left\{\frac{|X-Y|}{1+|X-Y|} \geq \frac{\varepsilon}{1+\varepsilon}\right\},$$

and

$$\left\{\frac{|X-Y|}{1+|X-Y|} \geq \varepsilon\right\} = \left\{|X-Y| \geq \frac{\varepsilon}{1-\varepsilon}\right\}$$

By Markov's inequality

$$\begin{aligned}
\mathbb{P}\left(|X-Y| \geq \varepsilon\right) &= \mathbb{P}\left(\frac{|X-Y|}{1+|X-Y|} \geq \frac{\varepsilon}{1+\varepsilon}\right) \\
&\leq \frac{1+\varepsilon}{\varepsilon} E\left[\frac{|X-Y|}{1+|X-Y|}\right] \\
&= \frac{1+\varepsilon}{\varepsilon} d(X,Y).
\end{aligned}$$

Therefore, convergence in the metric implies convergence in probability. For the other direction let

$$Z = \frac{|X-Y|}{1+|X-Y|}$$

and

$$Z_\varepsilon = \varepsilon \mathbb{1}\{|Z| < \varepsilon\} + \mathbb{1}\{|Z| \geq \varepsilon\}.$$

Since $Z \leq 1$, then $Z \leq Z_\varepsilon$ and

$$\begin{aligned} d(X,Y) &= E\,[Z] \\ &\leq E\,[Z_\varepsilon] \\ &= E\,[\varepsilon \ \ \{|Z| < \varepsilon\} + \ \ \{|Z| \geq \varepsilon\}] \\ &\leq \varepsilon + \mathbb{P}\,(|Z| \geq \varepsilon) \\ &= \varepsilon + \mathbb{P}\left(|X - Y| \geq \frac{\varepsilon}{1-\varepsilon}\right). \end{aligned}$$

Hence convergence in probability implies convergence in the metric.

Exercise 3 (20 pts)**.** Prove Cauchy criterions for convergence a.s. and i.P.:

(i) Show that X_n converges almost surely iff

$$\forall \epsilon > 0 \quad \mathbb{P}[\sup_{k \geq 0} |X_{n+k} - X_n| > \epsilon] \to 0 \quad n \to \infty$$

(ii) Show that X_n converges in probability iff

$$\forall \epsilon > 0 \quad \sup_{k \geq 0} \mathbb{P}[|X_{n+k} - X_n| > \epsilon] \to 0 \quad n \to \infty$$

Solution:

(i) Equivalently, for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ so that for all $n \geq N$

$$\mathbb{P}\left(\sup_{k \geq 0} |X_{n+k} - X_n| > \varepsilon\right) < \varepsilon.$$

Let $\varepsilon > 0$. Suppose X_n converges almost surely to some random variable X. Then, there exists a measurable set E with $\mathbb{P}(E) < \varepsilon$ so that $X_n \Rightarrow X$ uniformly on E^c. Therefore, there exists $N \in \mathbb{N}$ so that for all $n \geq N$ and for all $\omega \in E^c$

$$|X_n(\omega) - X(\omega)| \leq \varepsilon.$$

Hence,

$$\mathbb{P}\left(\sup_{k \geq 0} |X_{n+k} - X_n| > \varepsilon\right) = \mathbb{P}(E) < \varepsilon.$$

Conversely suppose the Cauchy criterion is satisfied. For all k, $\sup_{k \geq 0} |X_{n+k} - X_n| \geq |X_{n+k} - X_n|$. Therefore, for all k,

$$\mathbb{P}\,(|X_{n+k} - X_n| > \varepsilon) \leq \mathbb{P}\left(\sup |X_{n+k} - X_n| > \varepsilon\right).$$

Hence

$$\sup_{k\geq 0}\mathbb{P}\left(|X_{n+k}-X_n|>\varepsilon\right)\leq \mathbb{P}\left(\sup_{k\geq 0}|X_{n+k}-X_n|>\varepsilon\right),$$

and thusly this condition is stronger than the condition in part (ii). Therefore, by part (ii) $\{X_n\}$ converges in probability to a random variable X. The set of points where X_n does not converge to X is

$$\{X_n \not\to X\} = \bigcup_{\varepsilon>0}\bigcap_{n=1}^{\infty}\{\sup_{k\geq n}|X_k - X|>\varepsilon\},$$

and by continuity of probability

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty}\left\{\sup_{k\geq n}|X_k-X|>\varepsilon\right\}\right) = \lim_{n\to\infty}\mathbb{P}\left(\sup_{k\geq n}|X_k-X|>\varepsilon\right).$$

Thus it suffices to show that for all $\varepsilon>0$ there exists an $N\in\mathbb{N}$ so for all $n\geq N$

$$\mathbb{P}\left(\sup_{k\geq n}|X_k-X|>\varepsilon\right)<\varepsilon.$$

By the triangle inequality

$$|X_k-X|\leq |X_k-X_n|+|X_n-X|,$$

and thus

$$\begin{aligned}\sup_{k\geq n}|X_k-X| &\leq \sup_{k\geq n}|X_k-X_n|+|X_n-X|\\ &\leq \sup_{k\geq 0}|X_{n+k}-X_n|+|X_n-X|.\end{aligned}$$

Choose N_1 to satisfy the Cauchy criterion for $\varepsilon/2$ and N_2 for the convergence in measure for $\varepsilon/2$. Let $N=\max\{N_1,N_2\}$, then for all $n\geq N$

$$\mathbb{P}\left(\sup_{k\geq n}|X_k-X|>\varepsilon\right)\leq \mathbb{P}\left(\sup_{k\geq 0}|X_{n+k}-X_n|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(|X_n-X|>\frac{\varepsilon}{2}\right)<\varepsilon$$

as desired.

(ii) Suppose that the sequence satisfies the Cauchy criterion. Then, there exists a subsequence $\{X_{n_k}\}$ such that, if $E_k = \{|X_{n_k} - X_{n_{k+1}}| \geq 2^{-k}\}$, then $\mathbb{P}(E_k) \leq 2^{-k}$. Let $F_k = \cup_{j=k}^{\infty} E_j$. For $\omega \notin F_k$, and $i \geq j \geq k$, we have

$$|X_{n_j}(\omega) - X_{n_i}(\omega)| \leq \sum_{l=j}^{i-1} |X_{n_{l+1}}(\omega) - X_{n_l}(\omega)| \leq 2^{1-j}$$

by the definition of the subsequence. This means that $\{X_{n_k}\}$ is pointwise Cauchy on F_k^c. Let $F = \cap_{k=1}^{\infty} F_k = \limsup E_k$, and note that $\mathbb{P}(F) = 0$. Let us define a random variable X such that on $X(\omega) = 0$ for all $\omega \in F$, and $X(\omega) = \lim X_{n_k}(\omega)$ for all $\omega \notin F$. Then $X_{n_k} \to X$ a.s. and thus $X_{n_k} \to X$ i.p. Finally, we have

$$\{|X_n - X| \geq \epsilon\} \subset \{|X_{n_k} - X_n| \geq \epsilon/2\} \cup \{|X_{n_k} - X| \geq \epsilon/2\},$$

and the right-hand side can be made arbitrarily small with large k and n, and thus we have convergence in probability of X_n to X.

Conversely, for $a > 0$ and $m, n \in \mathbb{N}$, let $E_n(a) = \{\omega : |X_n(\omega) - X(\omega)| \geq a\}$ and $F_{m,n}(a) = \{\omega : |X_m(\omega) - X_n(\omega)| \geq a\}$. Then, for every $a > 0$ and every m, n, we have

$$F_{m,n}(a) \subset E_m(a/2) \cup E_n(a/2).$$

In fact, if ω is neither in $E_m(a/2)$ nor in $E_n(a/2)$, then $|X_m(\omega) - X(\omega)| < a/2$ and $|X_n(\omega) - X(\omega)| < a/2$, so that $|X_m(\omega) - X_n(\omega)| \leq |X_m(\omega) - X(\omega)| + |X_n(\omega) - X(\omega)| < a/2 + a/2 = a$ so $|X_m(\omega) - X(\omega)| < a$, which shows $\omega \notin F_{m,n}(a)$. In virtue of the monotonicity and the subadditivity of the measure, we also have

$$0 \leq \mathbb{P}(F_{m,n}(a)) \leq \mathbb{P}(E_m(a/2) \cup E_n(a/2)) \leq \mathbb{P}(E_m(a/2)) + \mathbb{P}(E_n(a/2)). \tag{1}$$

Fix $a > 0$. Since $X_n \to X$ in probability, for every $\epsilon > 0$ there exists k such that

$$n > k \Rightarrow \mathbb{P}(E_n(a/2)) < \epsilon/2. \tag{2}$$

Then, for $m > k$ and $n > k$, it follows from (1) and (2) that

$$\mathbb{P}(F_{m,n}(a)) < \epsilon/2 + \epsilon/2 = \epsilon$$

which means $lim_{m,n\to\infty} \mathbb{P}(F_{m,n}(a)) = 0$. Since this holds for every $a > 0$, it follows that the Cauchy criterion is satisfied.

Exercise 4. Let $\{X_n\}$ be a sequence of random variables defined on the same probability space.

(a) Show $\mathbb{E}[|X_n - X|] \to 0$ implies $X_n \overset{\text{i.p.}}{\to} X$.

(b) Suppose that $X_n \overset{\text{i.p.}}{\to} 0$ and that for some constant c, we have $|X_n| \le c$, for all n, with probability 1. Show that

$$\lim_{n\to\infty} \mathbb{E}[|X_n|] = 0.$$

(c) Suppose that each X_n can only take the values 0 and 1 and, that $\mathbb{P}(X_n = 1) = 1/n$.

  (i) Give an example in which we **have** almost sure convergence of X_n to 0.

  (ii) Give an example in which we **do not have** almost sure convergence of X_n to 0.

Solution:

(a) Follows from Markov's inequality

$$\mathbb{P}(|X_n - 0| \ge \epsilon) \le \frac{E[|X_n|]}{\epsilon},$$

Therefore, if $E[|X_n|]$ approaches 0, then X_n approaches 0 in probability.

(b) Fix $\epsilon > 0$ and define a new random variable X_n^ϵ as follows. We have $X_n^\epsilon = \epsilon$ whenever $|X_n| \le \epsilon$, and $X_n^\epsilon = c$ whenever $|X_n| > \epsilon$. Then, it is always true that $|X_n| \le X_n^\epsilon$ and therefore

$$E[|X_n|] \le E[X_n^\epsilon] = \epsilon P(|X_n| \le \epsilon) + cP(|X_n| > \epsilon)$$

Taking limits as n goes to infinity, we get

$$\lim_n E[|X_n|] \le \epsilon,$$

and since this holds for all $\epsilon > 0$, we get $\lim_n E[|X_n|] = 0$.

(c) (i) Consider the Lebesgue probability space $([0,1], \mathcal{B}, \lambda)$. Let $X_n = \mathbb{1}_{[0,\frac{1}{n}]}$. For all $\omega \in (0,1]$ there exists $n \in N$ such that $1/n < \omega$. Thus $X_n \overset{\text{a.s.}}{\to} 0$ and for all n $\mathbb{P}(X_n = 1) = E[X_n] = 1/n$.

(ii) Let $\{X_n\}$ be independent. Then $\{X_n = 1\}$ occurs infinitely often by the Borel-Cantelli Lemma, as

$$\sum_{i=1}^{\infty} P(X_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

and the events $\{X_n = 1\}$ are independent. Hence X_n cannot converge to 0.

Exercise 5. Let $X_1, X_2, \ldots$ be i.i.d. exponential random variables with parameter $\lambda = 1$. Let $S_n = X_1 + \cdots + X_n$. Let $a > 1$. What is the Chernoff upper bound for $\mathbb{P}(S_n \geq na)$?

Solution: To use Chernoff's bound as stated in the lecture notes, we need to work with random variables that have mean 0. Let $Y_i = X_i - 1$, and $S'_n = Y_1 + \cdots + Y_n$. Then

$$P(S_n \geq na) = P(S'_n \geq n(a-1)).$$

Now the moment generating function of the X_i is $1/(1-s)$, so Y_i has moment generating function $e^{-s}/(1-s)$. We must optimize

$$\sup_{s \geq 0} s(a-1) - \log \frac{e^{-s}}{1-s},$$

The optimum occurs at $s = (a-1)/a$, and equals $a - 1 - \log(a)$. So,

$$P(S_n \geq na) \leq e^{-n(a-1-\log(a))} = a^n e^{-n(a-1)}.$$

Exercise 6. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, uniformly distributed on the interval $[0,1]$. For n odd, let M_n be the median of $X_1, X_2, \ldots, X_n$, i.e. the $(\frac{n+1}{2})$ order statistic $X^{(\frac{n+1}{2})}$. Show that M_n converges to 1/2, in probability.

Solution: Fix some $\epsilon > 0$. We will show that $\mathbb{P}(M_n > 1/2 + \epsilon)$ converges to zero. By symmetry, this will also imply that $\mathbb{P}(M_n < 1/2 - \epsilon)$ also converges to zero, and will establish the desired convergence.

Let N_n be the number of X_is ($i = 1, \ldots, n$) for which $X_i > 1/2 + \epsilon$. If $M_n > 1/2 + \epsilon$, then $N_n/n > 1/2$. But $\mathbb{E}[N_n/n] = 1/2 - \epsilon$, so that $\mathbb{P}(N_n/n > 1/2) \to 0$, by the weak law of large numbers.

Exercise 7. [Optional, not to be graded] Show that for every $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B})$ there exist a sequence $\mathbb{P}_{X_n} \xrightarrow{\text{d}} \mathbb{P}_X$ such that every $\mathbb{P}_{X_n}$ has a continuous, bounded, infinitely-differentiable PDF. Steps:

(i) Show $X_\epsilon = X + \epsilon Z \xrightarrow{\text{d}} X$ as $\epsilon \to 0$.

(ii) Let $X \perp\!\!\!\perp Z$ and $Z \sim \mathcal{N}(0,1)$. Show that CDF of X_ϵ is continuous (*Hint:* BCT) and differentiable (*Hint: Fubini*) with derivative

$$f_{X_\epsilon}(a) = \mathbb{E}\left[f_Z\left(\frac{a-X}{\epsilon}\right)\frac{1}{\epsilon}\right].$$

(iii) Show that $a \mapsto f_{X_\epsilon}(a)$ is continuous.

(iv) [Optional] Conclude the proof (*Hint:* derivatives of f_Z are uniformly bounded on $\mathbb{R}$).

Solution:

(i) Let Z be a random variable defined on the same probability space as X. Let $\delta > 0$ and WLOG assume $\delta < 1$, as $\{\varepsilon|Z| \geq x\} \subset \{\varepsilon|Z| \geq y\}$ for $x \geq y$, Therefore,

$$\begin{aligned}
\mathbb{P}\left(\varepsilon|Z| \geq \delta\right) &= \mathbb{P}\left(Z \leq -\delta/\varepsilon\right) + \mathbb{P}\left(Z \geq \delta/\varepsilon\right) \\
&\leq^{(a)} \mathbb{P}\left(Z \leq -\delta/\varepsilon\right) + \mathbb{P}\left(Z < \delta^2/\varepsilon\right) \\
&= F_Z\left(-\frac{\delta}{\varepsilon}\right) + 1 - F_Z\left(\frac{\delta^2}{\varepsilon}\right) \\
&\to 0 + 1 - 1 = 0,
\end{aligned}$$

where (a) follows since $\delta^2 < \delta$ for $\delta < 1$. Therefore, $\varepsilon Z \to 0$ in probability. Thus, as convergence in probability is closed under addition, $X + \varepsilon Z \to X$ in probability and thusly $X + \varepsilon Z \to X$ in distribution.

(ii) For any measurable function g, as $X \perp\!\!\!\perp Z$,

$$\int g(\gamma)\,\mathbb{P}_{X_\varepsilon}(d\gamma) = \int\int g(\alpha+\beta)\,\mathbb{P}_X(d\alpha)\,\mathbb{P}_{\varepsilon Z}(d\beta),$$

where

$$\mathbb{P}_{\varepsilon Z}(d\beta) = \frac{d}{d\beta}F_Z\left(\frac{\beta}{\varepsilon}\right) = \frac{1}{\varepsilon}f_Z\left(\frac{\beta}{\varepsilon}\right)\lambda(d\beta)$$

and this integration makes sense and can be interchanged by Fubini's Theorem. Letting $g = \mathbb{1}_{(-\infty,z]}$

$$
\begin{aligned}
F_{X_\varepsilon}(z) &= \int \quad {}_{(-\infty,z]} \, d\mathbb{P}_{X_\varepsilon} \\
&= \int\int \mathbb{1}_{(-\infty,z]}(\alpha+\beta) f_Z\left(\frac{\beta}{\varepsilon}\right)\frac{1}{\varepsilon}\,\lambda(d\beta)\,\mathbb{P}_X(d\alpha) \\
&= \int\int \quad {}_{(-\infty,z]}(\gamma) f_Z\left(\frac{\gamma-\alpha}{\varepsilon}\right)\frac{1}{\varepsilon}\,\lambda(d(\gamma-\alpha))\,\mathbb{P}_X(d\alpha) \\
&= \int\int \quad {}_{(-\infty,z]}(\gamma) f_Z\left(\frac{\gamma-\alpha}{\varepsilon}\right)\frac{1}{\varepsilon}\,\lambda(d\gamma)\,\mathbb{P}_X(d\alpha) \quad \text{(Shift invariance)} \\
&= \int_{(-\infty,z]}\int f_Z\left(\frac{\gamma-\alpha}{\varepsilon}\right)\frac{1}{\varepsilon}\,\mathbb{P}_X(d\alpha)\,\lambda(d\gamma) \\
&= \int_{(-\infty,z]} E\left[f_Z\left(\frac{\gamma-X}{\varepsilon}\right)\frac{1}{\varepsilon}\right]\,\lambda(d\gamma).
\end{aligned}
$$

Let

$$
f_{X_\varepsilon}(a) := E\left[f_Z\left(\frac{a-X}{\varepsilon}\right)\frac{1}{\varepsilon}\right].
$$

Thus

$$
F_{X_\varepsilon}(z) = \int_{-\infty}^{z} f_{X_\varepsilon}(\gamma)\,\lambda(d\gamma).
$$

From part (iii) f_{X_ε} is continuous and therefore this integral agrees with the Riemann integral. Hence, by the fundamental theorem of calculus, $F_{X_\varepsilon}(z)$ is differential with derivative

$$
\frac{d}{dz}F_{X_\varepsilon}(z) = f_{X_\varepsilon}(z).
$$

(iii) Limits and integration can be interchanged using the bounded convergence theorem.

(iv) Same as part (iii).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Problem Set 10

Readings:
Notes from Lecture 18, 19 and 20.
[Cinlar] Chapter III.
[GS] Section 7.5, 7.10, 8.1-8.3.

Exercise 1. Let $S_n = \sum_{j=1}^{n} X_j$ be a sum of independent random variables X_j with $|X_j| \leq 1$ almost surely. Show that S_n converges in probability if and only if it converges almost surely (to a finite value).
(Hint: See how the case $\sum \mathrm{var}[X_j] = \infty$ was treated in the converse part of Kolmogorov-Khintchine in Lecture 19.)

Solution: Note that a.s. $\Rightarrow$ i.p. is instantaneous. So we only need i.p. $\Rightarrow$ a.s.

Let $S_n \to S$ in probability to a finite probability distribution; first, let's establish that we can WLOG assume that $E[X_j] = 0$ for all j. If $\sum_{j=1}^{\infty} E[X_j]$ converges, we can simply replace X_j with $\hat{X}_j := X_j - E[X_j]$ (the original S_n's converge if and only if the new $\hat{S}_n$'s converge). But $\sum_{j=1}^{\infty} E[X_j]$ cannot diverge because $|X_j| \leq 1$ and the X_j's are independent.

Now we show a.s. convergence via the Cauchy criterion

$$\lim_{n\to\infty} P[\sup_{k\geq 1} |S_{n+k} - S_n| > \varepsilon]$$

Since the X_i's are independent and mean-0, we can apply Kolmogorov's Inequality. In particular,

$$S_{n+k} - S_n = \sum_{j=n+1}^{n+k} X_j$$

so we get from Kolmogorov's Inequality that

$$P\Big[\sup_{k\geq 1} \Big| \sum_{j=n+1}^{n+k} X_j \Big| \geq \epsilon\Big] \leq \frac{2}{\varepsilon^2} \sum_{j=n+1}^{\infty} \mathrm{var}[X_j]$$

We'd like this to converge to 0 as $n \to \infty$ (then by the Cauchy criterion we're done); that happens if $\sum_{j=1}^{\infty} \mathrm{var}[X_j] = \sum_{j=1}^{\infty} E[X_j^2] < \infty$.

Now we borrow the trick from the lecture notes, as described in the hint. Suppose $\sum_{j=1}^{\infty} E[X_j^2] = \infty$ (so we will want to show a contradiction); then note that because $|X_j|$ almost surely we know that

$$\sum_{j=1}^{\infty} E[|X_j|^3] \leq \sum_{j=1}^{\infty} E[X_j^2]$$

Therefore, we can apply the CLT for non-identical variables (defining $D_n := \sum_{j=1}^{n} \mathrm{var}[X_j]$) and get

$$\frac{S_n}{\sqrt{D_n}} \to Z \sim N(0,1) \text{ in distribution}$$

But now let us fix some $t > 0$. Then

$$P[S_n > t] = P[S_n/\sqrt{D_n} > t/\sqrt{D_n}]$$

Since $t/\sqrt{D_n} \to 0$ and $S_n/\sqrt{D_n} \to Z$ (in distribution), we get that $P[S_n > t] \to 1/2$ as $n \to \infty$ for all $t > 0$, which means that S_n cannot converge in probability to any finite-valued random variable. This is a contradiction, and we're done.

Exercise 2. Let $\{X_n\}$ be a sequence of identically distributed random variables, with finite variance. Suppose that $\mathrm{cov}(X_i, X_j) \leq \alpha^{|i-j|}$, for every i and j, where $|\alpha| < 1$. Show that the sample mean $(X_1 + \cdots + X_n)/n$ converges to $\mathbb{E}[X_1]$, in probability.

Solution: Let

$$Y_n = \frac{\sum_{k=1}^{n} X_k}{n},$$

with $E[Y_n] = E[X_1]$. By Chebyshev's inequality,

$$\mathbb{P}\left(|Y_n - E[X_1]| \geq \varepsilon\right) = \mathbb{P}\left(|Y_n - E[Y_n]| \geq \varepsilon\right) \leq \frac{\mathrm{var}(Y_n)}{\varepsilon^2}.$$

The variance of Y_n is

$$\begin{aligned}\mathrm{var}(Y_n) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{cov}(X_i, X_j) \leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha^{|i-j|} \leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\alpha|^{|i-j|} \\ &\leq \frac{1}{n^2}(2n)\sum_{k=0}^{n-1}|\alpha|^k = \frac{2}{n}\frac{1-|\alpha|^n}{1-|\alpha|} \to 0.\end{aligned}$$

Hence $Y_n \to E[X_1]$ in distribution, which implies (since $E[X_1]$ is a constant) that it converges in probability as well.

Exercise 3. Given an i.i.d. sequence $X_n, n \geq 1$ with $\sigma^2 \triangleq \mathrm{var}(X_1) < \infty$, the CLT states that

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\sum_{1\leq i\leq n} X_i - n\mathbb{E}[X_1]}{\sigma n^{\alpha}} \leq x\Big) = \Phi(x) \triangleq \int_{-\infty}^{x} \frac{e^{\frac{-t^2}{2}}}{\sqrt{2\pi}} dt,$$

when $\alpha = 1/2$. Compute the limit above for every $\alpha > 0$ and every x.

Solution: Let

$$Y_n = \frac{1}{\sigma}\left(\sum_{i=1}^{n} X_n - n\mathbb{E}[X_1]\right).$$

For all x, n and α, we have that

$$\mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) = \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq x n^{\alpha-1/2}\right) = \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq \frac{x}{n^{1/2-\alpha}}\right)$$

There are two cases of interest

1. Let $0 < \alpha < 1/2$. Suppose $x \geq 0$. Then the sequence $\left\{\frac{x}{n^{\frac{1}{2}-\alpha}}\right\}$ monotonically decreases to zero. Then for every ϵ there exists N such that for all $n > N$, $0 \leq \frac{x}{n^{1/2-\alpha}} \leq \epsilon$. By monotonicity of probability, for all $n > N$,

$$\mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq 0\right) \leq \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq \frac{x}{n^{1/2-\alpha}}\right) \leq \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq \epsilon\right).$$

Taking a liminf and a limsup and applying the CLT

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq \varepsilon\right) = \Phi(\varepsilon)$$
$$\liminf_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{\alpha}} \leq x\right) \geq \lim_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq 0\right) = \Phi(0).$$

Letting epsilon tend to zero, these bound coincide by continuity of Φ. Hence the limit exists and equals $\frac{1}{2}$. Similarly, for $x < 0$ the sequence $\{\frac{x}{n^{\frac{1}{2}-\alpha}}\}$ monotonically increases to zero and yields the bounds

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) \leq \Phi(0) \qquad \liminf_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{\alpha}} \leq x\right) \geq \Phi(-\varepsilon).$$

2. Let $\alpha > 1/2$. Suppose $x = 0$. Then, by the CLT

$$\mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq xn^{\alpha-1/2}\right) = \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq 0\right) \to \Phi(0) = 1/2.$$

Suppose $x > 0$. Then the sequence $\left\{\frac{x}{n^{\frac{1}{2}-\alpha}}\right\}$ monotonically increases to infinity. Therefore, for every $M \in \mathbb{R}$ there exists $N > 0$ such that for all $n > N$ $xn^{\alpha-1/2} > M$ and

$$\mathbb{P}\left(\frac{Y}{n^{1/2}} \leq xn^{\alpha-1/2}\right) \geq \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq M\right)$$

Thus

$$\liminf_{n\to\infty} \mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) \geq \liminf_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{\alpha}} \leq M\right) = \Phi(M)$$

and this bound holds in the limit. Therefore,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) = 1.$$

Similarly, suppose $x < 0$. Then the sequence $\left\{\frac{x}{n^{\frac{1}{2}-\alpha}}\right\}$ monotonically decreases to infinity. Therefore, for every $M \in \mathbb{R}$ there exists $N > 0$ such that for all $n > N$ $xn^{\alpha-1/2} < M$ and

$$\mathbb{P}\left(\frac{Y}{n^{1/2}} \leq xn^{\alpha-1/2}\right) \leq \mathbb{P}\left(\frac{Y_n}{n^{1/2}} \leq M\right)$$

Thus

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{\alpha}} \leq M\right) = \Phi(M)$$

and this bound holds in the limit. Therefore,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Y}{n^{\alpha}} \leq x\right) = 0.$$

In summary

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Y_n}{n^{\alpha}} \leq x\right) = \begin{cases} \frac{1}{2} & 0 < \alpha < \frac{1}{2} \\ \Phi(x) & \alpha = \frac{1}{2} \\ 0 & \alpha > \frac{1}{2} \text{ and } x < 0 \\ \frac{1}{2} & \alpha > \frac{1}{2} \text{ and } x = 0 \\ 1 & \alpha > \frac{1}{2} \text{ and } x > 0 \end{cases}.$$

Exercise 4. Show that given an i.i.d. sequence $X_n, n \geq 1$ with mean μ, variance σ^2, while $(\sum_{1\leq i\leq n} X_i - \mu n)/(\sqrt{n}\sigma) \to N(0,1)$ in distribution, it is not the case that the same sequence converges in probability. (Hint: Cauchy criterion)

Solution: By the Cauchy criterion if $Y_n \to Y$ in probability then

$$\mathbb{P}\left(|Y_{2n} - Y_n| \geq \varepsilon\right) \to 0,$$

and contrapositively, if $\mathbb{P}\left(|Y_{2n} - Y_n| \geq \varepsilon\right) \not\to 0$ then Y_n cannot converge in probability. Let

$$Z_n = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu).$$

Observe that

$$\begin{aligned} Z_{2n} - Z_n &= \frac{1}{\sqrt{2n}\sigma}\sum_{i=n+1}^{2n}(X_i - \mu) + \frac{1}{\sqrt{2n}\sigma}\sum_{i=1}^{n}(X_i - \mu) - \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu) \\ &= \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=n+1}^{2n}(X_i - \mu)\right) + \left(\frac{1}{\sqrt{2}} - 1\right)\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu)\right) \end{aligned}$$

and relabeling terms, as the X_k are i.i.d

$$Z_{2n} - Z_n = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i' - \mu)\right) + \left(\frac{1}{\sqrt{2}} - 1\right)\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu)\right),$$

where the $\{X_k\}$ and $\{X_k'\}$ are independent. By the CLT, the first term converges in distribution to an $N(0, \frac{1}{2})$ and the second term converges in distribution to an $N\left(0\left(\frac{1}{\sqrt{2}} - 1\right)^2\right)$. Hence, $Z_{2n} - Z_n$ converges in distribution to an $N(0, \frac{1}{2} + \left(\frac{1}{2} - 1\right)^2)$, as these two terms are independent, and thusly cannot converge in probability to zero.

Exercise 5. Give an example of:

1. Independent zero-mean X_j's such that $\sum \text{var} X_j$ diverges but

$$S_n = \sum_{k=1}^{n} X_j \tag{1}$$

converges almost surely.

2. Independent zero-mean X_j taking values in $[-1, 1]$ such that $X_j \overset{\text{a.s.}}{\to} 0$ but S_n does not converge almost surely.

Solution:

1. Let
$$X_n = \begin{cases} 0 & \text{w.p. } 1 - \frac{1}{n^2} \\ \sqrt{n} & \text{w.p. } \frac{1}{2n^2} \\ -\sqrt{n} & \text{w.p. } \frac{1}{2n^2} \end{cases}.$$
Then
$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq 0) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$
As the X_n are independent, by the partial inverse to the Borel-Cantelli lemma, $\{X_n \neq 0\}$ only finitely often almost everywhere. Therefore, almost everywhere S_n is a sum of a finite number of terms and thus converges. Hence S_n converges almost everywhere. However,
$$\sum_{n=1}^{\infty} \text{var}(X_n) = \sum_{n=1}^{\infty} 2n \frac{1}{2n^2} = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

2. Let
$$X_n = \begin{cases} \frac{1}{\sqrt{n}} & \text{w.p } \frac{1}{2} \\ -\frac{1}{\sqrt{n}} & \text{w.p } \frac{1}{2} \end{cases}.$$
Then
$$\sum_{n=1}^{\infty} E\left[|X_n|^4\right] = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$
Hence $X_n \to 0$ almost surely by lecture 18 proposition 1. However,
$$\sum_{n=1}^{\infty} \text{var}(X_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$
and thusly S_n cannot converge almost surely by theorem 2 of lecture 19.

Exercise 6. Let $\{X_n\}$ be a sequence of nonnegative integrable random variables and X an integrable random variable. Suppose $X_n \overset{\text{a.s.}}{\to} X$ and $\mathbb{E}[X_n] \to \mathbb{E}[X]$. Show that the family $\{X_n, n = 1, \ldots\}$ is uniformly integrable. Conclude that $X_n \overset{L_1}{\to} X$, i.e.
$$\mathbb{E}[|X_n - X|] \to 0$$
(Thus, $Y_n \overset{\text{a.s.}}{\to} Y$ is u.i. iff $\mathbb{E}[|Y_n|] \to \mathbb{E}[|Y|]$.)

Solution: As the X_n and consequently X are nonnegative the absolute values in the definitions of uniform integrability can be ignored. For all $b > 0$ with $\mathbb{P}_X(b) = 0$, $X_n \mathbb{1}\{X_n \leq b\} \to X\mathbb{1}\{X \leq b\}$ almost everywhere, as a probability measure can only have countable many items this condition can be ignored, i.e. a $b' > b$ always exists. By the bounded convergence theorem $E[X_n \mathbb{1}\{X_n \leq b\}] \to E[X\mathbb{1}\{X \leq b\}]$. Therefore, as $E[X_n] \to E[X]$, by linearity of integration $E[X_n \mathbb{1}\{X_n > b\}] \to E[X\mathbb{1}\{X > b\}]$.

By assumption all of the X_n and X are integrable. Thus they are individually uniformly integrable and any finite collection is also uniformly integrable, i.e. taking a max over a finite collection of integers. Let $\varepsilon > 0$. There exists $b_1 > 0$ such that $E[X\mathbb{1}\{X > b_1\}] < \varepsilon$. There exists an $N \in \mathbb{N}$ such that for all $n \geq N$

$$|E[X_n \mathbb{1}\{X_n > b_1\}] - E[X\mathbb{1}\{X > b_1\}]| < \varepsilon.$$

Moreover, as any finite collection is uniformly integrable, there exists a $b_2 > 0$ such that

$$\sup_{1 \leq n < N} E\left[X_n \mathbb{1}\{X_n > b_2\}\right] < \varepsilon.$$

Let $b \geq \max\{b_1, b_2\}$ then

$$\begin{aligned}\sup_n E\left[X_n \mathbb{1}\{X_n > b\}\right] &\leq \sup_{1 \leq n < N} E\left[X_n \mathbb{1}\{X_n > b\}\right] + \sup_{n \geq N} E\left[X_n \mathbb{1}\{X_n > b\}\right] \\ &< \varepsilon + E\left[X\mathbb{1}\{X > b_1\}\right] + \varepsilon \\ &< 3\varepsilon.\end{aligned}$$

Hence $\{X_n\}$ is uniformly integrable. As $X_n \to X$ almost everywhere implies $X_n \to X$ in probability, the result follows from theorem 1 of lecture 19.

Exercise 7. Let $N(\cdot)$ be a Poisson process with rate λ. Find the covariance of $N(s)$ and $N(t)$.

Solution: **Solution.** Suppose $s \leq t$. We have that

$$\begin{aligned}E[N(s)N(t)] &= E[N(s)(N(s) + (N(t) - N(s)))] \\ &= E[N(s)^2] + E[N(s)(N(t) - N(s))] \\ &= E[N(s)^2] + E[N(s)]E[(N(t) - N(s))] \\ &= \lambda s + \lambda^2 s^2 + \lambda s \lambda(t - s) \\ &= \lambda s + \lambda^2 st\end{aligned}$$

so that

$$\operatorname{cov}(N(s), N(t)) = E[N(s)N(t)] - E[N(s)]E[N(t)] = \lambda s + \lambda^2 st - \lambda s \lambda t = \lambda s.$$

Keeping in mind that we assumed $s \leq t$, we instead write this in the general case as

$$\operatorname{cov}(N(s), N(t)) = \lambda \min(s, t).$$

Exercise 8. Based on your understanding of the Poisson process, determine the numerical values of a and b in the following expression and explain your reasoning.

$$\int_t^{\infty} \frac{\lambda^5 \tau^4 e^{-\lambda\tau}}{4!}\, d\tau = \sum_{k=a}^{b} \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Solution: The left-hand side is the probability that an Erlang random variable of order 5 and rate λ is larger than t, i.e., the probability of at most 4 arrivals over an interval of length t. The right-hand side is the probability that the number of arrivals in a Poisson process with rate λ, over an interval of length t, is between a and b (inclusive). Thus, $a = 0$ and $b = 4$.

Exercise 9. ***(practice problem, not for grade)***

(a) Shuttles depart from New York to Boston every hour on the hour. Passengers arrive according to a Poisson process of rate λ per hour. Find the expected number of passengers on a shuttle. (Ignore issues of limited seating.)

(b) Now, and for the rest of this problem, suppose that the shuttles are not operating on a deterministic schedule, but rather their interdeparture times are exponentially distributed with rate μ per hour, and independent of the process of passenger arrivals. Find the PMF of the number shuttle departures in one hour.

(c) Let us define an "event" in the terminal to be either the arrival of a passenger, or the departure of a shuttle. Find the expected number of "events" that occur in one hour.

(d) If a passenger arrives at the gate, and sees 2λ people waiting, find his/her expected time to wait until the next shuttle.

(e) Find the PMF of the number of people on a shuttle.

Solution: Solution:

(a) The number of people that arrive within an hour is Poisson-distributed with parameter λ, and its expected value is λ.

(b) If the interarrival times for the shuttles are exponentially distributed, then shuttle departures form a Poisson process of rate μ. Thus, the number of departures in one hour has a Poisson PMF with parameter μ.

(c) Here, we are merging two independent Poisson processes, which results in a Poisson process of rate $\mu + \lambda$. Therefore, the expected number of "events" occurring in one hour will be $\mu + \lambda$.

(d) The number of people waiting conveys some information on the time since the last departure. On the other hand, because of memorylessness of the exponential distribution, this number is independent from the time until the next departure. Thus, the expected waiting time is just $1/\mu$, irrespective of how many people are waiting.

(e) Every event at the airport has probability $\lambda/(\lambda + \mu)$ of being a passenger arrival ("failure") and probability $\mu/(\lambda + \mu)$ of being a shuttle departure ("success"). Furthermore, different events are independent. The number of passengers on a shuttle is the number of failures until the first success and is distributed as $K - 1$, where K is a geometric random variable with parameter $\mu/(\lambda+\mu)$. Thus, the PMF of the number of people on the shuttle is

$$\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right), \qquad k=0,1,\ldots$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Problem Set 11

Readings:
Notes from Lecture 21,22
Chapter 7 of Bertsekas and Tsitsiklis "Introduction to Probability"
For *stopping times*: [Cinlar] Chapter V.1.
[GS] Chapter 6

Exercise 1. A particle performs a random walk on the vertex set of a finite connected undirected graph G, which for simplicity we assume to have neither self-loops nor multiple edges. At each stage it moves to a neighbor of its current position, each such neighbor being chosen with equal probability. If G has η edges, show that the stationary distribution is given by $\pi_v = d_v/(2\eta)$, where d_v is the degree of each vertex v.

Solution: One way to do this problem is to simply check that the proposed solution satisfies the defining equations: $\pi P = \pi$, and $\sum_v \pi_v = 1$ (we can see immediately that we have nonnegativity). We have:

$$\begin{aligned} \sum_v \pi_v &= \sum_v \frac{d_v}{2\eta} \\ &= \frac{1}{2\eta}\sum_v d_v \\ &= 1, \end{aligned}$$

since the sum of the degrees is twice the number of edges (each edge increases the sum of the degrees by exactly 2). Similarly, we can show that $\pi P = \pi$. Let us define δ_{vu} to be 1 if vertices u and v are adjacent, and 0 otherwise. Then, we have:

$$\begin{aligned} \sum_v \pi_v P_{vu} &= \frac{1}{2\eta}\sum_v d_v \left(\frac{1}{d_v}\delta_{vu}\right) \\ &= \frac{1}{2\eta}\sum_v \delta_{vu}. \end{aligned}$$

But $\sum_v \delta_{vu}$ is the number of edges incident to node u, that is, $\sum_v \delta_{vu} = d_u$. Therefore we have:

$$\sum_v \pi_v P_{vu} = \frac{1}{2\eta} d_u = \frac{d_u}{2\eta} = \pi_u.$$

Exercise 2. A particle performs a random walk on a bow tie $ABCDE$ drawn on Figure 1, where C is the knot. From any vertex, its next step is equally likely to be to any neighbouring vertex. Initially it is at A. Find the expected value of:

(a) The time of first return to A.

(b) The number of visits to D before returning to A.

(c) The number of visits to C before returning to A.

(d) The time of first return to A, given that there were no visits to E before the return to A.

(e) The number of visits to D before returning to A, given that there were no visits to E before the return to A.

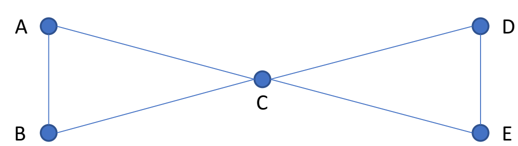

Figure 1: A bow tie graph.

Solution: First, we can compute that the steady state distribution is $\pi_A = \pi_B = \pi_D = \pi_E = 1/6$, and $\pi_C = 1/3$. We can do this either by solving a system of linear equations (as usual) or just use our result from the first problem above.

(a) By the result from class, and on the handout, we have: $t_A = 1/\pi_A = 6$. Alternatively, we can solve the following system of equations (observe than t_A appears in only one equation):

$$\begin{aligned}
t_A &= \frac{1}{2}(t_B+1) + \frac{1}{2}(t_C+1) \\
t_B &= \frac{1}{2} + \frac{1}{2}(t_C+1) \\
t_C &= \frac{1}{4} + \frac{1}{4}(t_B+1) + \frac{1}{4}(t_D+1) + \frac{1}{4}(t_E+1) \\
t_D &= \frac{1}{2}(t_C+1) + \frac{1}{2}(t_E+1) \\
t_E &= \frac{1}{2}(t_C+1) + \frac{1}{2}(t_D+1).
\end{aligned}$$

(b) By the result from the handout on Markov Chains, we know that

$$\pi_D = \frac{\mathbb{E}[\#\text{ transitions to } D \text{ in a cycle that starts and ends at } A]}{\mathbb{E}[\#\text{ transitions in a cycle that starts and ends at } A]},$$

from which we find that the quantity we wish to compute is $6\pi_D = 1$.

(c) Using the same method as in part (b), we find the answer to be $6\pi_C = 2$.

(d) We let $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot|X_0 = i)$, and let T_j be the time of the first passage to state j, and let $\nu_i = \mathbb{P}_i(T_A < T_E)$. Then, as we obtained the equations above, that is, by conditioning on the first step, we have

$$\begin{aligned}
\nu_A &= \frac{1}{2}\nu_B + \frac{1}{2}\nu_C \\
\nu_B &= \frac{1}{2} + \frac{1}{2}\nu_C \\
\nu_C &= \frac{1}{4} + \frac{1}{4}\nu_B + \frac{1}{4}\nu_D \\
\nu_D &= \frac{1}{2}\nu_C.
\end{aligned}$$

Solving these, we find: $\nu_A = 5/8, \nu_B = 3/4, \nu_C = 1/2, \nu_D = 1/4$. Now we can compute the conditional transition probabilities, which we call τ_{ij}. We have:

$$\begin{aligned}
\tau_{AB} &= \mathbb{P}_A(X_1 = B|T_A < T_E) \\
&= \frac{\mathbb{P}_A(X_1 = B)\mathbb{P}_B(T_A < T_E)}{\mathbb{P}_A(T_A < T_E)} \\
&= \frac{\nu_B}{2\nu_A} = \frac{3}{5}.
\end{aligned}$$

Similarly, we find: $\tau_{AC} = 2/5, \tau_{BA} = 2/3, \tau_{BC} = 1/3, \tau_{CA} = 1/2, \tau_{CB} = 3/8, \tau_{CD} = 1/8, \tau_{DC} = 1$. Now we have essentially reduced to a problem like part (a). We can compute the conditional expectation by solving a system of linear equations using the new transition probabilities:

$$\begin{aligned}
\tilde{t}_A &= 1 + \frac{3}{5}\tilde{t}_B + \frac{2}{5}\tilde{t}_C \\
\tilde{t}_B &= 1 + \frac{2}{3}(1) + \frac{1}{3}\tilde{t}_C \\
\tilde{t}_C &= 1 + \frac{1}{2}(1) + \frac{3}{8}\tilde{t}_B + \frac{1}{8}\tilde{t}_D \\
\tilde{t}_D &= 1 + \tilde{t}_C.
\end{aligned}$$

Solving these equations, yields $\tilde{t}_A = 14/5$.

(e) We can use the conditional transition probabilities above, to reduce to a problem essentially like that in part (b). Let N be the number of visits to D. Then, denoting by η_i the expected value of N given that we start at i, and that $T_A < T_E$, we have the equations:

$$\begin{aligned}
\eta_A &= \frac{3}{5}\eta_B + \frac{2}{5}\eta_B \\
\eta_B &= 0 + \frac{1}{3}\eta_C \\
\eta_C &= 0 + \frac{3}{8}\eta_B + \frac{1}{8}(1 + \eta_D) \\
\eta_D &= \eta_C.
\end{aligned}$$

Solving, we obtain: $\eta_A = 1/10$.

Exercise 3. Let $(\Omega, \mathcal{F}) = (\mathbb{R}^\infty, \mathcal{B}^\infty)$, $X_k(\omega) = \omega_k$, $k \in \mathbb{N}$, be the canonical coordinate functions and $\{\mathcal{F}_k\}$ a filtration of $\mathcal{F}$. Recall that a filtration is a sequence of increasing σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ contained in $\mathcal{F}$, $\mathcal{F}_k \subset \mathcal{F}$. We say that τ is a stopping time of the filtration $\{\mathcal{F}_k\}$ if

(a) τ is a positive integer

(b) for every $k \geq 1$ we have $\{\tau \leq k\} \in \mathcal{F}_k$

Let $\tau : \Omega \to \mathbb{N}$ be $(\mathcal{F}, \mathcal{B})$ measurable. Show that τ is a stopping of $\{\mathcal{F}_k\}$ if and only if for every $\omega, \omega' \in \Omega$ and for every $n \geq 1$

$$\tau(\omega) = n,\ X_k(\omega) = X_k(\omega') \quad \forall\, 1 \leq k \leq n \quad \Rightarrow \quad \tau(\omega') = n. \tag{1}$$

Solution: A positive integer valued random variable τ is a stopping time if and only if $\{\tau = n\} \in \mathcal{F}_n$ for all n. The forward direction follows from $\{\tau \leq n\} = \bigcup_{k=1}^{n}\{\tau = k\}$ and the reverse direction follows from $\{\tau = n\} = \{\tau \leq n\} \backslash \{\tau \leq n-1\}$. The relation $\omega \overset{n}{\sim} \omega'$ if

$$X_k(\omega) = X_k(\omega') \quad 1 \leq k \leq n$$

is an equivalence relation, i.e. reflexive, symmetric, and transitive. For all $E \subset \Omega$ define

$$[E]_n = \{\omega \in \Omega \mid \exists\, \omega' \in E \text{ s.t. } \omega' \overset{n}{\sim} \omega\}.$$

Condition 1 is equivalent to $[\{\tau = n\}]_n \subset \{\tau = n\}$. Therefore, it suffices to show that, for all n, $\{\tau = n\} \in \mathcal{F}_n$ if and only if $[\{\tau = n\}]_n \subset \{\tau = n\}$.

Suppose τ is a stopping time. Let

$$\mathcal{D} = \{E \subset \Omega \mid [E]_n \subset E\}.$$

By definition, $\mathcal{D}$ contains the empty set and sets of the form $X_j^{-1}(B)$ for $B \subset \mathbb{R}$ and $1 \leq j \leq n$. Moreover, let $\{E_j\} \in \mathcal{D}$, then

$$\left[\bigcup_{j=1}^{\infty} E_j\right]_n = \bigcup_{j=1}^{\infty} [E_j]_n \qquad \left[\bigcap_{j=1}^{\infty} E_j\right]_n \subset \bigcap_{j=1}^{\infty} [E_j]_n \,,$$

and therefore, $\mathcal{D}$ is a monotone class. Let

$$\mathcal{C} = \{X_j^{-1}(B) \mid B \in \mathcal{B},\ 1 \leq j \leq n\}.$$

Then, the minimal algebra containing $\mathcal{C}$ $\alpha(\mathcal{C})$ is the set of finite unions of finite intersections of sets of the form $X_j^{-1}(B)$ or $X_j^{-1}(B)^c$. As the inverse image respects complements and $\mathcal{D}$ is closed under intersections and unions, $\mathcal{D}$ contains $\alpha(\mathcal{C})$ and by the monotone class theorem $\mathcal{D} \supset \sigma(\mathcal{C}) = \mathcal{F}_n$. Hence $\{\tau = n\} \in \mathcal{D}$.

Conversely, suppose that condition 1 is satisfied. By definition, $[\{\tau = n\}]_n \supset \{\tau = n\}$ and thusly $[\{\tau = n\}]_n = \{\tau = n\}$. Therefore, Ω decomposes as a union of equivalence classes $\Omega = \bigcup_{\alpha \in I} U_\alpha$, for some indexing set I where $[U_\alpha]_n = U_\alpha$ for all α and $U_\alpha \cap U_\beta = \emptyset$ for $\alpha \neq \beta$. For each $\alpha \in I$ choose a representative $\omega_\alpha \in U_\alpha$. Let $f : \Omega \to \Omega$ with $f_{|U_\alpha} \equiv \omega_\alpha$. To show that f is measurable it suffices to check on a generating collection. Let $S \subset \mathcal{N}$ be a finite set and $B = \prod_{s \in S} B_s$ with $B_s \in \mathcal{B}(\mathbb{R})$, then $f^{-1}(B) = \bigcap_{k=1}^{n} X_k^{-1}(X_k(B)) \in \mathcal{F}_n$ since $X_k(B)$ is either B_k or $\emptyset$ and X_k is measurable. Therefore, f is $(\mathcal{F}_n, \mathcal{F})$ measurable and, as $[\{\tau = n\}]_n = \{\tau = n\}$ and τ is $(\mathcal{F}, \mathcal{B})$ measurable, $\{\tau = n\} = f^{-1}(\{\tau = n\}) \in \mathcal{F}_n$. Hence τ is a stopping time.

Exercise 4. Let τ be a stopping time of a filtration $\mathcal{F}_n$. Recall that the σ-algebra $\mathcal{F}_\tau$ of "past until τ" is defined as

$$\mathcal{F}_\tau = \{E : E \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \forall n\}$$

Show that for every random variable V measurable with respect to $\mathcal{F}_\tau$ there exists a stochastic process $\{G_n, n = 1, \ldots\}$, with G_n measurable with respect to $\mathcal{F}_n$, such that

$$V = G_\tau \,.$$

(Hint: First consider simple V).

Solution: Let V be a random variable measurable with respect to $\mathcal{F}_\tau$. Then V decomposes as

$$V = V\mathbb{1}\{V > 0\} + V\mathbb{1}\{V = 0\} - (-V)\mathbb{1}\{V < 0\} = V_+ - V_-.$$

Let $G_n = V\mathbb{1}\{\tau \leq n\}$. Then $G_\tau = V$ and

$$G_n = V_+\mathbb{1}\{\tau \leq n\} - V_-\mathbb{1}\{\tau \leq n\}.$$

As random variables are closed under addition and scalar multiplication, it suffices to show that G_n is measurable with respect to $\mathcal{F}_n$ for positive V. If $V > 0$ then $G_n \geq 0$. Let $x \geq 0$. Then

$$\{G_n > x\} = \{V\mathbb{1}\{\tau \leq n\} > x\} = \{V > x\} \cap \{\tau \leq n\} \in \mathcal{F}_n$$

since V is measurable with respect to $\mathcal{F}_\tau$. As $\{(x, \infty)\}$ is a generating p-system for the Borel sigma algebra on the real numbers, G_n is measurable with respect to $\mathcal{F}_n$.

Exercise 5. *(Cover time of C_n)* For a MC with state space $\mathcal{X}$ we define τ_{cov} to be the first time that every element of $\mathcal{X}$ was visited. The covering time $t_{cov} = \max_{x \in \mathcal{X}} \mathbb{E}^x[\tau_{cov}]$. Consider a MC that is a simple random walk on an n-cycle: it moves with probability $1/2$ to one of the neighbors each time. Show that $t_{cov}(n) = \frac{n(n-1)}{2}$ (Lovász'93). (Hint: Let τ_n be the first time a simple random walk on $\mathbb{Z}$ started at 0 visits n distinct states. Relate to t_{cov} and gambler's ruin.)

Solution: Clearly, by symmetry, it does not matter what vertex we start from. Let us define σ_k to be the first time that at least k distinct vertices have been visited; obviously $\sigma_1 = 0$. We now note that $t_{cov} = \mathbb{E}[\sigma_n]$; we can also telescope these like so:

$$\sigma_n = (\sigma_n - \sigma_{n-1}) + (\sigma_{n-1} - \sigma_{n-2}) + \cdots + (\sigma_2 - \sigma_1)$$

(note that we omit the "$\cdots + \sigma_1$" because it's just 0). This of course means that $t_{cov} = \sum_{k=1}^{n-1} \mathbb{E}[\sigma_{k+1} - \sigma_k]$ (by linearity).

Now let us examine what the situation is like at time σ_k for $k < n$. We have k visited vertices, which obviously are contiguous (and so form a path); furthermore, X_{σ_k} must be at an endpoint of the path since by definition of σ_k, it must be the first visit we made to this vertex.

Now we ask: how long from then until σ_{k+1}? Well, we have a Gambler's Ruin problem: exiting either end of the path of visited vertices gives us a new one. To be precise, it's a Gambler's Ruin starting with 1 dollar and ending either with 0 dollars or $k+1$ dollars; we know that the expected number of steps for this is $j(k+1-j)$ where $j=1$, which gives k steps. Therefore,

$$\mathbb{E}[\sigma_{k+1} - \sigma_k] = k$$

Plugging this in to the above, we get

$$t_{cov} = \sum_{k=1}^{n-1} \mathbb{E}[\sigma_{k+1} - \sigma_k] = \sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$$

Exercise 6. *(Last visited vertex of C_n)* Consider a simple random walk X_t on an n-cycle C_n and let τ_{cov} be the first time that every vertex was visited. Show that given that $X_0 = v$ the distribution of $X_{\tau_{cov}}$ is uniform on $\{v\}^c$. (Hint: Notice that to have $X_{\tau_{cov}} = k$ the random walk should visit the states $k-1$ and $k+1$ before k.)
Fun fact: cycles and cliques are the only graphs with this property (Lovász-Winkler'93).

Solution: Fix a vertex x; let σ_x be the first time that a *neighbor* of x is visited. For $x \neq v$, obviously a neighbor of x must be visited before x is (keeping in mind that v itself could be this neighbor). Let $u = X_{\sigma_u}$ (the first neighbor visited) and w be the other neighbor, which by definition has not been visited by time σ_x.

Now note that if x is visited before w, then x cannot be the last vertex, i.e. $X_{\tau_{cov}} \neq x$; but if w is visited before x, then *every* other vertex must have also been visited before x since there is no way to get from u to w without either passing through x or passing through literally every other vertex.

Finally, note that this is simply a Gambler's Ruin problem - where the gambler starts with 1 dollar (since u is next to x) and wins if he gets to $n-1$ dollars (since w is the target). The probability of winning is just $\frac{1}{n-1}$. Since this holds regardless of what x is (provided $x \neq v$ of course) we get that every non-v vertex has an equal probability of being the final vertex.

(**Sanity check:** The probabilities should sum up to 1, which they do because there are $n-1$ non-starting vertices, each with $\frac{1}{n-1}$ probability of being the last visited.)

Exercise 7. Let B_k be iid with law $\mathbb{P}[B_k = +1] = p = 1 - \mathbb{P}[B_k = -1]$. Answer the following:

- Let $X_n = B_n B_{n+1}$, $n \geq 0$. Is it Markov? If yes, find its transition kernel.

- Let $Y_n = \frac{1}{2}(B_n - B_{n-1})$, $n \geq 1$. Is it Markov? If yes, find its transition kernel.

- Let $Z_n = |\sum_{k=1}^{n} B_k|$, $n \geq 1$. Is it Markov? If yes, find its transition kernel.

- If $\{V_i, i \geq 0\}$ is a Markov process with state space $\mathcal{X}$, and E_j are some subsets of $\mathcal{X}$, is it true that

  $$\mathbb{P}[V_n \in E_n | V_{n-1} \in E_{n-1}, V_{n-2} \in E_{n-2}, \ldots, V_0 \in E_0] = \mathbb{P}[V_n \in E_n | V_{n-1} \in E_{n-1}],$$

  provided that $\mathbb{P}[V_{n-1} \in E_{n-1}, \ldots, V_0 \in E_0] > 0$?

- Suppose that $P(x, y)$ is a kernel of an irreducible Markov chain. If $P(\cdot, x_1) = P(\cdot, x_2)$ show that $\pi(x_1) = \pi(x_2)$, where π is a stationary distribution. What if the chain is not irreducible?

Solution:

1) It is not Markov (a couple exceptions, listed at the end). Let $p = 0.99$, and consider $\mathbb{P}[X_3 = 1 | X_2 = -1]$. Note that $X_2 = -1$ means either $B_2 = -1$ and $B_3 = 1$ or vice versa; and (given no other information) these two cases are equally probable. So no matter what B_4 happens to be, $\mathbb{P}[X_3 = 1 | X_2 = -1] = 1/2$. But now suppose that we add the information that $X_1 = -1$ as well. If $X_1 = X_2 = -1$, then we have one of the following two cases:

1. $(B_1, B_2, B_3) = (-1, 1, -1)$;
2. $(B_1, B_2, B_3) = (1, -1, 1)$.

Note that the second case is vastly more probable than the first; therefore,

$$\mathbb{P}[X_3 = 1 | X_2 = -1,\ X_1 = -1] > 1/2$$

(we could calculate it precisely using Baye's Theorem, but we don't really need to go to the trouble). Therefore $\{X_n\}$ does not satisfy the Markov property.

(**Remark:** The exceptions are when $p = 1/2$ or, if we'll allow such a thing, $p = 0$ or 1.)

2) Same as for 1 - a counterexample can be easily constructed, so it is not Markovian.

3) Yes it is Markov, although this is far from obvious. We'll be using the *reflection principle* to see this. First, note that if $Z_n = 0$, then $Z_{n+1} = 1$ for

sure, so that $P(0,1) = 1$; also note that Z_n can never move except by 1, so $P(i,j) = 0$ for all $|i-j| \neq 1$.

Now let's start with the difficult part. Since

$$Z_n = \sum_{k=1}^{n} B_n$$

it is obvious that $P(i,j) = 0$ if $j \neq i-1, i+1$. Furthermore, we can easily see that $P(0,1) = 1$ (and that this obviously does not depend on the history), and that Z_n can never be negative. Now we just have to examine $P(i,i+1)$ (noting that $P(i,i-1) = 1 - P(i,i+1)$.

We define $W_n := \sum_{k=1}^{n} B_k$. Now note that if we know whether W_n is positive or negative, we could immediately determine $\mathbb{P}[Z_{n+1} = Z_n + 1]$ – it would be p if $W_n > 0$, and $1-p$ if $W_n < 0$ – and therefore the transition probabilities would only be determined by the current position Z_n.

Now suppose that $Z_k = z_k$ for all $k = 0, 1, \ldots, n$, and $z_n = \ell$ (the current state). Then we can define a *possible history* of W_k's as a sequence $\mathbf{w} = (w_0, w_1, \ldots, w_n)$ such that

- $w_k \in \{-z_k, z_k\}$ (so that $|w_k| = z_k$) for all k;
- $|w_k - w_{k-1}| = 1$ for all $k = 1, 2, \ldots, n$.

Define S to be the set of all such sequences (and obviously it is finite); define

$$S_- := \{\mathbf{w} \in S : w_n = -\ell\} \text{ and } S_+ := \{\mathbf{w} \in S : w_n = \ell\}$$

Note the following:

- this is a partition of S – every $\mathbf{w} \in S$ is in exactly one of S_-, S_+;
- $|S_-| = |S_+|$ because for any $\mathbf{w} \in S_-$, we have $-\mathbf{w} \in S_+$ (this is the "reflection" we were talking about). So let's call

  $$m := |S_-| = |S_+|$$

- for any $\mathbf{w} \in S_-$, we have $\frac{n-\ell}{2}$ increments (corresponding to $B_k = 1$) and $\frac{n+\ell}{2}$ decrements (corresponding to $B_k = -1$), and for any $\mathbf{w} \in S_+$, we have $\frac{n+\ell}{2}$ increments and $\frac{n-\ell}{2}$ decrements. Therefore,

  $$\mathbb{P}[W_k = w_k \text{ for all } k \leq n] = \begin{cases} p^{\frac{n+\ell}{2}}(1-p)^{\frac{n-\ell}{2}} & \text{if } \mathbf{w} \in S_+ \\ p^{\frac{n-\ell}{2}}(1-p)^{\frac{n+\ell}{2}} & \text{if } \mathbf{w} \in S_- \end{cases}$$

  Note that this only depends on the value of w_n.

Note, therefore, that

$$\begin{aligned}
\mathbb{P}[Z_k = z_k \text{ for } k \le n] &= \sum_{\mathbf{w}\in S} \mathbb{P}[W_k = w_k \text{ for } k \le n] \\
&= \sum_{\mathbf{w}\in S_+} \mathbb{P}[W_k = w_k \text{ for } k \le n] + \sum_{\mathbf{w}\in S_-} \mathbb{P}[W_k = w_k \text{ for } k \le n] \\
&= \sum_{\mathbf{w}\in S_+} p^{\frac{n+\ell}{2}}(1-p)^{\frac{n-\ell}{2}} + \sum_{\mathbf{w}\in S_-} p^{\frac{n-\ell}{2}}(1-p)^{\frac{n+\ell}{2}} \\
&= m\big(p^{\frac{n+\ell}{2}}(1-p)^{\frac{n-\ell}{2}} + p^{\frac{n-\ell}{2}}(1-p)^{\frac{n+\ell}{2}}\big) \\
&= m(p(1-p))^{\frac{n-\ell}{2}}\big(p^\ell + (1-p)^\ell\big)
\end{aligned}$$

Now we can apply Bayes' Theorem (remember that $z_n = \ell$ here):

$$\begin{aligned}
\mathbb{P}[W_n = \ell \mid Z_k = z_k \text{ for } k \le n] &= \mathbb{P}[\{W_k\} \in S_+ \mid Z_k = z_k \text{ for } k \le n] \\
&= \frac{\mathbb{P}[\{W_k\} \in S_+] \cdot \mathbb{P}[Z_k = z_k \mid \{W_k\} \in S_+]}{\mathbb{P}[Z_k = z_k \text{ for } k \le n]} \\
&= \frac{\mathbb{P}[\{W_k\} \in S_+]}{\mathbb{P}[Z_k = z_k \text{ for } k \le n]} \\
&= \frac{m(p(1-p))^{\frac{n-\ell}{2}}\big(p^\ell\big)}{m(p(1-p))^{\frac{n-\ell}{2}}\big(p^\ell + (1-p)^\ell\big)} \\
&= \frac{p^\ell}{p^\ell + (1-p)^\ell}
\end{aligned}$$

(part of this was noting that $\mathbb{P}[Z_k = z_k \mid \{W_k\} \in S_+] = 1$ by definition of S_+). Note that this depends *only* on the value of $Z_n = \ell$, and not on any other Z_k's or even on n – so therefore we can conclud that it is *Markovian*!.

Now we have to compute the transition kernel. We have:

$$\mathbb{P}[W_n = -\ell \mid Z_k = z_k \text{ for } k \le n] = 1 - \frac{p^\ell}{p^\ell + (1-p)^\ell} = \frac{(1-p)^\ell}{p^\ell + (1-p)^\ell}$$

Therefore (letting $z_n = \ell$ below), we get $\mathbb{P}[Z_{n+1} = \ell + 1 \mid Z_k = z_k \text{ for } k \le n]$

$$\begin{aligned}
&= \mathbb{P}[Z_{n+1} = \ell + 1 \text{ and } W_n = \ell \mid Z_k = z_k \text{ for } k \le n] \\
&\quad + \mathbb{P}[Z_{n+1} = \ell + 1 \text{ and } W_n = -\ell \mid Z_k = z_k \text{ for } k \le n]
\end{aligned}$$

Dealing with each piece here on its own, we get:

$$\begin{aligned}
&\mathbb{P}[Z_{n+1}=\ell+1 \text{ and } W_n=\ell \,|\, Z_k=z_k \text{ for } k\le n] \\
&=\mathbb{P}[Z_{n+1}=\ell+1 \,|\, W_n=\ell \text{ and } Z_k=z_k \text{ for } k\le n]\cdot\mathbb{P}[W_n=\ell \,|\, Z_k=z_k \text{ for } k\le n] \\
&=p\cdot\mathbb{P}[W_n=\ell \,|\, Z_n=\ell]=\frac{p^{\ell+1}}{p^\ell+(1-p)^\ell}
\end{aligned}$$

An analogous computation for the other piece gives

$$\mathbb{P}[Z_{n+1}=\ell+1 \text{ and } W_n=-\ell \,|\, Z_k=z_k \text{ for } k\le n]=\frac{(1-p)^{\ell+1}}{p^\ell+(1-p)^\ell}$$

We then finally put all of this together to get

$$P(\ell,\ell+1)=\mathbb{P}[Z_{n+1}=\ell+1 \,|\, Z_n=\ell]=\frac{p^{\ell+1}+(1-p)^{\ell+1}}{p^\ell+(1-p)^\ell}$$

(and of course $P(\ell,\ell-1)=1-\frac{p^{\ell+1}+(1-p)^{\ell+1}}{p^\ell+(1-p)^\ell}$). As noted at the very top, we have also $P(0,1)=1$ and $P(i,j)=0$ for all $j\neq i+1,i-1$.

(**Remark:** A common error was to assume that because the answer differs depending on the history of the B_k's, it cannot be Markov. But when evaluating whether the Z_n's are Markov, you cannot look at the history of the B_k's, only on the history of the Z_n's.)

4) Not always. An easy example is a random walk on a 6-cycle (labeled in order a,b,c,d,e,f) with uniformly-randomly-chosen starting point V_0; let $E_n=\{a\}$ and $E_{n-2}=\{d\}$ and $E_{n-1}=\mathcal{X}$ (the rest of the E_k don't matter, but if we want to feel better about ourselves we can set them to $\mathcal{X}$ as well). Then

$$\mathbb{P}[V_n\in E_n \,|\, V_{n-1}\in E_{n-1}]=\mathbb{P}[V_n=a]=1/6$$

because the condition $V_{n-1}\in\mathcal{X}$ says nothing. But of course if $V_{n-2}\in E_{n-2}$ (i.e. $V_{n-2}=d$), there's no way that $V_n=a$ since you can't reach it in time. So

$$\mathbb{P}[V_n\in E_n \,|\, V_k\in E_k \text{ for all } k<n]=0\neq 1/6$$

5) This follows easily from the equation $\pi^T=\pi^T P$. If the chain is not irreducible, that does not alter the previous statement, so it remains true.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J — Fall 2018

Problem Set 12 (OPTIONAL)

Readings:
Notes from Lectures 23-26.
[Cinlar], Chapter V.1-V.4
[Grimmett-Stirzaker], Chapter 6.1-6.6, Chapter 12.1-12.6

Exercise 1. Consider two irreducible Markov chains K_1 and K_2 on $\mathbb{Z}_+$ whose only jumps are of the form $n \to \{n-1, n, n+1\}$. Suppose

$$K_1(i, \{i+1\}) \geq K_2(j, \{j+1\}) \qquad \forall i, j \geq 0$$

and

$$K_1(i, \{i-1\}) \leq K_2(i, \{i-1\}) \qquad \forall i \geq 0.$$

Show that there exists a coupling such that $X_n \geq Y_n$ a.s.. Conclude that if Y is transient, then so is X.

Solution: Start the two Markov chains in some arbitrary state i_0. Evolve X_n according to K_1 and Y_n as follows. Suppose $X_n = i$. Let

$$p_{-1} = K_1(i, \{i-1\}) \qquad p_0 = K_1(i, \{i\}) \qquad p_1 = K_1(i, \{i+1\}),$$

$$q_{-1} = K_2(i, \{i-1\}) \qquad q_0 = K_2(i, \{i\}) \qquad q_1 = K_2(i, \{i+1\}).$$

By assumption $p_1 \geq q_1$ and $p_{-1} \leq q_{-1}$. There are two cases $p_0 \geq q_0$ and $p_0 \leq q_0$. Define $\mathbb{P}(\underline{X}_{n+1}) := (\mathbb{P}(X_{n+1} = i+1), \mathbb{P}(X_{n+1} = i), \mathbb{P}(X_{n+1} = i-1))$ and $\mathbb{P}(\underline{Y}_{n+1}) := (\mathbb{P}(Y_{n+1} = i+1), \mathbb{P}(Y_{n+1} = i), \mathbb{P}(Y_{n+1} = i-1))$. Let $\mathbb{P}(\underline{Y}_{n+1}) = A\mathbb{P}(\underline{X}_{n+1})$ where the corresponding matrices for the two case are respectively

$$\begin{bmatrix} \frac{q_1}{p_1} & 0 & 0 \\ 0 & \frac{q_0}{p_0} & 0 \\ 1 - \frac{q_1}{p_1} & 1 - \frac{q_0}{p_0} & 1 \end{bmatrix} \qquad \begin{bmatrix} \frac{q_1}{p_1} & 0 & 0 \\ \frac{q_0 - p_0}{p_1} & 1 & 0 \\ \frac{q_{-1} - p_{-1}}{p_1} & 0 & 1 \end{bmatrix}.$$

Exercise 2. Consider a Bernoulli process $(X_n, n \geq 1)$. For each of the integer valued random variables T below determine whether it is a stopping time or not. In case T is a stopping time describe the corresponding sequence of functions $h_n = h_n(x_1, \ldots, x_n)$ determining T.

(a) T is the first time n such that $\sum_{1\leq i\leq n} X_i = 2$. Namely, $T = \min\{n : {}_{1\leq i\leq n} X_i = 2\}$. If T is indeed a stopping time describe the corresponding sequence of functions h_n.

(b) $T = \max(10, \min\{n : \sum_{1\leq i\leq n} X_i = 2\})$.

(c) T is the first time n such that $\sum_{1\leq i\leq n} X_i = \sum_{n+1\leq i\leq 2n} X_i$.

(d) T is the first time n such that $\sum_{1\leq i\leq n/2} X_i = \sum_{n/2+1\leq i\leq n-1} X_i$.

Solution:

(a) Yes, it is a stopping time.

$$h_n(x_1, \dots, x_n) = \mathbb{1}(x_1 + \cdots + x_{n-1} < 2, x_1 + \cdots x_n = 2).$$

(b) Yes, it is a stopping time. $h_n(x_1, \dots, x_n) =$

$$\mathbb{1}(n = 10, x_1 + \cdots + x_{10} \geq 2) + \mathbb{1}(x_1 + \cdots + x_{n-1} < 2, x_1 + \cdots x_n = 2, n > 10).$$

(c) No, it is not a stopping time because it depends on the future (entries in $n+1, \dots, 2n$).

(d) Yes it is a stopping time.

$$h_n(x_1, \dots, x_n) = \mathbb{1}\left(\sum_{i=1}^{\lfloor k/2 \rfloor} x_i \neq \sum_{\lfloor k/2 \rfloor + 1}^{k-1} x_i \quad k < n-1, \sum_{i=1}^{\lfloor n/2 \rfloor} x_i = \sum_{\lfloor n/2 \rfloor + 1}^{n-1} x_i\right)$$

Exercise 3. Let X_1, X_2, X_3 be independent exponential random variables with mean 1. Let

$$\alpha = \mathbb{P}(X_1 > X_2 + X_3).$$

(a) Find α, without calculating any integrals.

(b) Find the probability that the largest of the three random variables X_1, X_2, X_3 is larger than the sum of the other two. [You can express your answer in terms of the constant α from part (a).]

Solution:

(a) Consider two independent rate one Poisson processes, say A and B. Let X_2 be the time until the first arrival of A, let X_3 be the time between the first and second arrival of A, and let X_1 be the time of the first arrival of B. Then α is the probability that A has two arrivals before B. By Poisson splitting/merging, we can instead consider a single rate two Poisson process, where we assign arrivals to A or B i.i.d. with probability $1/2$. Thus α is the probability that the first two arrivals of the merged process both go to A, thus $\alpha = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

(b) As the events $\{X_1 > X_2 + X_3\}$, $\{X_2 > X_1 + X_3\}$, and $\{X_3 > X_1 + X_2\}$ are disjoint, and there union is the event that the largest X_i is bigger than the sum of the smaller three, we have

$$
\begin{aligned}
&\mathbb{P}(\{X_1 > X_2 + X_3\} \cup \{X_2 > X_1 + X_3\} \cup \{X_3 > X_1 + X_2\}) \\
&\qquad = \mathbb{P}(X_1 > X_2 + X_3) + \mathbb{P}(X_2 > X_1 + X_3) + \mathbb{P}(X_3 > X_1 + X_2) \\
&\qquad = 3\mathbb{P}(X_1 > X_2 + X_3) \\
&\qquad = 3\alpha.
\end{aligned}
$$

where second equality follows by symmetry as the variables are i.i.d.

Exercise 4. Fast and slow customers arrive at a 24 hour store according to independent Poisson processes, each with rate 1 per minute. Fast customers stay in the bookstore for 1 minute, slow customers stay in the store for 2 minutes.

(a) What is the PMF of the total number of customer arrivals during a one minute interval?

(b) Find the variance of the number of customers in the store at 3 p.m.

(c) At 3 p.m., there is only one customer present in the store.

  (i) What is the probability, β, that the customer is a fast one?

  (ii) What is the PDF that this customer will depart before a new customer arrives? [You may express your answer in terms of the constant β from part (i). Also, you may leave your answer as a formula involving integrals – you do not have to evaluate the integrals.]

Let N_t be the number of fast customer arrivals during $[0, t]$.

(d) Does $(N_{2t} - N_t)/t$ converge in probability, as $t \to \infty$? With probability 1? If yes, to what? Outline a rigorous justification for your answers. You can start with t integer-valued and then argue for $t \in \Re$.

(e) Find (approximately) a time k such that

$$\mathbb{P}(N_k \geq 100) \approx 0.758.$$

Note that if Z is a standard normal random variable, then $\mathbb{P}(Z \leq 0.7) = 0.758$. [You do not need to be rigorous in deriving your answer. You may leave your answer in the form of an equation for k, which you do not need to solve numerically.]

Solution:

(a) By Poisson merging, the total number of arrivals is a Poisson process with a rate of two customers per minute, so in an one minute window there are $\mathrm{Pois}(2)$ arrivals.

(b) A fast customer will be in the store at 3pm iff he arrived between 2:59pm and 3:00pm, because they only stay for a minute. Thus there will be $\mathrm{Pois}(1)$ fast customers in the store. Likewise a slow customer will be in the store iff he arrived between 2:58pm and 3:00pm, thus there will be $\mathrm{Pois}(2)$ slow customers in the store. As the sum of $\mathrm{Pois}(\lambda)$ and $\mathrm{Pois}(\mu)$ independent is $\mathrm{Pois}(\lambda + \mu)$, the total number of customers in the store at 3:00pm will be $\mathrm{Pois}(3)$. This random variable has variance 3.

(c) (i) By Poisson merging, we can instead suppose that we drew $\mathrm{Pois}(3)$ total customers to be in the store, and then for each customer, with probability $1/3$ assigned them as "fast" and with probability $2/3$ assigned them as slow, i.d.d. and independent of the total number of customers. Thus by independence of the number of customers and their classification, conditional on the total number of customers being one, the probability the customer will be fast is $\beta = 1/3$.

(ii) Let Z be the amount of time that the customer in the store will remain before departing. Recall that for a Poisson process conditioned to have k arrivals in $[0, t]$, the time of each arrival is i.i.d. Uni$(0, t)$. Thus the distribution of the arrival time is Uni$(0, 1)$ when the customer was fast and Uni$(0, 2)$ when the customer was slow. Thus the distribution of Z is Uni$(0, 1)$ if the customer was fast and Uni$(0, 2)$ if the customer was slow. Given that $Z = z$, the probability that we have a new arrival before the customer departs is $\mathbb{P}(\mathrm{Exp}(2) < z)$. Thus by total probability,

$$\mathbb{P}(\text{arrival before departure}) = \beta \int_0^1 \exp(-2z)dz + (1-\beta)\int_0^2 \frac{1}{2}\exp(-2z)dz.$$

(d) Assume t is integer. Let $X_k = N_k - N_{k-1}$ for k integer, so X_k are i.i.d Pois(1). We have that $N_{2t} - N_t =_d N_t = \ {}^{t}_{k=1} X_k$ by Lecture 21 page 5 property b. Thus for all $\epsilon > 0$,

$$\lim_{t\to\infty} \mathbb{P}\left(\left|\frac{N_{2t}}{t} - \frac{N_t}{t} - 1\right| > \epsilon\right) = \lim_{t\to\infty} \mathbb{P}\left(\left|\frac{1}{t}\sum_{i=1}^{t} X_i - 1\right| > \epsilon\right) = 0,$$

by the WLLN. This shows the result for integer t. The assumption that t was integer was not essential. Suppose we take $t \to \infty$ along $t_n = cn$ for any $c > 0$, where n is integer. Then we can divide N_t into intervals of length c making $\tilde{X}_k \sim \text{Pois}(c)$ be i.i.d. so then

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{N_{2t_n}}{t_n} - \frac{N_{t_n}}{t_n} - 1\right| > \epsilon\right) &= \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{cn}\sum_{i=1}^{n} \tilde{X}_i - 1\right| > \epsilon\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum^{n} \tilde{X}_i - c\right| > \epsilon c\right) = 0\end{aligned}$$

by the WLLN.

(e) As in the previous problem for integer k, $N_k = \sum_{i=1}^{k} X_i$, where $X_i \sim$ Pois(1) are i.i.d., and thus $\mathbb{E}[X_1] = \text{var}(X_1) = 1$. Thus by the CLT,

$$\begin{aligned}\mathbb{P}(N_k \geq 100) = \mathbb{P}\left(\sum_{i=1}^{k} X_i \geq 100\right) &= \mathbb{P}\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{k}(X_i - 1) \geq \frac{100-k}{\sqrt{k}}\right) \\ &\approx \mathbb{P}\left(\text{N}(0,1) \geq \frac{100-k}{\sqrt{k}}\right).\end{aligned}$$

Thus we need k that solves

$$\frac{100-k}{\sqrt{k}} = -0.7.$$

Exercise 5. Let S be the set of arrival times in a Poisson process on $\mathbb{R}$ (i.e., a process that has been running forever), with rate λ. Each arrival time in S is displaced by a random amount. The random displacement associated with each element of S is a random variable that takes values in a finite set. We assume that the random displacements associated with different arrivals are independent and identically distributed. Show that the resulting process (i.e., the process whose arrival times are the displaced points) is a Poisson process with rate λ. (We expect a proof consisting of a verbal argument, using known properties of Poisson processes; formulas are not needed.)

Solution: Let $\{v_1, \ldots, v_m\}$ be the set of values that our perturbations can take, and let $\{p_1, \ldots, p_m\}$ be the probabilities of each perturbation outcome. As the perturbation are independent of the Poisson process, by Poisson splitting this is equivalent to m independent Poisson processes, with rate λp_i for $i = 1, \ldots, m$, where the ith process has every point translated by v_i. But by the stationarity property of Poisson process (property b, page 5 of lecture 21), the distribution of the number of points on any interval for each of the i independent and translated processes will be the same as if we did not apply the translation, since the length of the interval determines the distribution of the number of points. Finally, we can apply Poisson merging to the untranslated processes (which are still independent), to recover a Poisson process with rate λ.

Exercise 6.

(a) Consider a Markov chain with several recurrent classes $R_l, 1 \le l \le L$. For each $l = 1, 2, \ldots, L$ consider the distribution $\pi_l = (\pi_x^l, 1 \le x \le N)$ on the entire states space defined as follows. For each $x \in R_l$, $\pi_x = 1/\mu_x$, where μ_x is the mean recurrence time of state x, and $\pi_y = 0$ for all $y \notin R_l$. Show that π_l is a stationary distribution.

(b) Suppose π and μ are stationary distribution of some Markov chain X_n. Prove that any convex combination $\nu = \lambda\pi + (1-\lambda)\mu$ is also a stationary distribution. Namely $\lambda \in [0, 1]$ and ν is defined by $\nu_i = \lambda\pi_i + (1-\lambda)\mu_i$.

From parts (a) and (b) we conclude that a Markov chain with several recurrence classes has infinitely many stationary distribution. One can show that every stationary distribution can be obtained in the way described by (b), namely as a convex combination $\sum_{1 \le l \le L} \lambda_l \pi_l$ of stationary distributions described in part (a).

Solution:

(a) Let P be the transition matrix of X_n. Fix some recurrent class. WLOG, assume that the states of the Markov chain are $\{1, \ldots, n\}$ and the recurrent class contains states $\{1, \ldots, k\}$, with $k < n$. Let Q the top left k by k block of P. In particular, let P have the block decomposition

$$P = \begin{bmatrix} Q & 0 \\ R & S \end{bmatrix}$$

where the top right corner is all zeros because the states $\{1, \ldots, k\}$ never transition outside of $\{1, \ldots, k\}$.

Create a new Markov chain that only contains the states $\{1, \ldots, k\}$, and the edges internal to the recurrent class, with the same transition probabilities, (so with transition matrix Q). First, we observe that as the set of states is a recurrent class, every state in the class can only point to other states in the class, thus the transition probabilities leaving each state sum to one and our new structure is indeed a Markov chain. Next, as the new Markov chain has a single recurrent class, it has a unique stationary distribution $\pi = (\pi_1, \ldots, \pi_k)$ satisfying $\pi' Q = \pi'$. By theorem 1 from lecture 23, $\pi_i = 1/\mu_i$, where μ_i is the mean recurrence time. (The assumption of aperiodicity is only needed to show that the transient distribution converges to the steady state distribution.) However, the mean recurrence time of state i in the new Markov chain will be same as the mean recurrence time of state i in the original Markov chain, as by a coupling argument we can make there recurrence times equal surely. Let $\bar{\pi}$ be an n dimensional vector such that

$$\bar{\pi}_i = \begin{cases} \pi_i & i \leq k \\ 0 & i > k. \end{cases}$$

Then $\bar{\pi}$ is of the form in the claim, so it suffices to prove that $\bar{\pi}' P = \bar{\pi}'$. Let Q_i be the ith column of Q, R_i be the ith column of R, S_i be the ith column of S. Then the ith column of P for $i \leq k$ is (Q_i, R_i), so

$$\bar{\pi}' P_i = \pi' Q_i + 0 R_i = \pi_i = \bar{\pi}_i$$

and the ith column of P for $i > k$ is given by $(0, S_i)$, so

$$\bar{\pi}' P_i = \pi' 0 + 0 S_i = 0 = \bar{\pi}_i.$$

Thus $\bar{\pi} P = \bar{\pi}$, so $\bar{\pi}$ is stationary. This shows the claim.

(b) Let P be the transition matrix for X_n. As π and μ are stationary, we have that $\pi' P = \pi'$ and $\mu' P = \mu'$. Thus for $\lambda \in [0, 1]$, if $\nu = \lambda\pi + (1 - \lambda)\mu$, then by the linearity,

$$\nu' P = (\lambda\pi + (1 - \lambda)\mu)' P = \lambda\pi' P + (1 - \lambda)\mu' P = \lambda\pi' + (1 - \lambda)\mu' = \nu',$$

thus ν is stationary.

Exercise 7. Consider a Markov chain $\{X_k\}$ on the state space $\{1, \ldots, n\}$, and suppose that whenever the state is i, a reward $g(i)$ is obtained. Let R_k be the total reward obtained over the time interval $\{0, 1, \ldots, k\}$, that is, $R_k = g(X_0) + g(X_1) + \cdots + g(X_k)$. For every state i, let

$$m_k(i) = \mathbb{E}[R_k \mid X_0 = i],$$

and

$$v_k(i) = \text{var}(R_k \mid X_0 = i)$$

be the mean and variance, respectively of R_k, conditioned on the initial state being equal to i.

(a) Find a recursion that given the values of $m_k(1), \ldots, m_k(n)$ allows the computation of $m_{k+1}(1), \ldots, m_{k+1}(n)$.

(b) Find a recursion that given the values of $v_k(1), \ldots, v_k(n)$ allows the computation of $v_{k+1}(1), \ldots, v_{k+1}(n)$. *Hint:* The following formula (the "law of total variance") may be useful:

$$\text{var}(X) = \mathbb{E}[\text{var}(X \mid Y)] + \text{var}(\mathbb{E}[X \mid Y]).$$

Here, $\text{var}(X \mid Y)$ stands for the variance of the conditional distribution of X given Y, and is itself a random variable because it is a function of Y.

Solution:

(a) We have $m_{k+1}(i) = \mathbb{E}[R_{k+1}|X_0 = i]$. Using the total expectation theorem, we have:

$$\begin{aligned}
m_{k+1}(i) &= \mathbb{E}[g(X_0) + \cdots + g(X_{k+1})|X_0 = i] \\
&= g(i) + \sum_k p_{ij}\mathbb{E}[g(X_1) + \cdots + g(X_{k+1})|X_1 = j] \\
&= g(i) + \sum_j p_{ij} m_k(j).
\end{aligned}$$

(b) Let $Q = g(X_1) + \cdots + g(X_{k+1})$, so that $R_{k+1} = g(X_0) + Q$ Using the law of total variance, and noting that adding a constant does not affect variance, we have:

$$\begin{aligned}
\text{var}(R_{k+1} \mid X_0 = i) &= \text{var}(Q \mid X_0 = i) \\
&= \text{var}(\mathbb{E}[Q \mid X_0 = i, X_1]) + \mathbb{E}[\text{var}(Q \mid X_0 = i, X_1)].
\end{aligned}$$

Let us consider the first term in the final sum above: The random variable $\mathbb{E}[Q \mid X_0 = i, X_1]$ takes the value $\mathbb{E}[Q \mid X_1 = j] = m_k(j)$ with probability p_{ij}. Given that $X_0 = i$, its mean is thus

$$\sum_j p_{ij} m_k(j) = m_{k+1}(i) - g(i),$$

and therefore its variance is

$$\sum_j p_{ij}(g(i) + m_k(j) - m_{k+1}(i))^2.$$

For the second term, notice that $\mathrm{var}(Q \mid X_0 = i, X_1)$ is equal to $v_k(j)$ whenever X_1 happens to be j. Thus,

$$\mathbb{E}[\mathrm{var}(Q \mid X_0 = i, X_1)] = \sum_j p_{ij} v_k(j).$$

Putting these two terms together, we find:

$$\mathrm{var}(R_{k+1} \mid X_0 = i) = \sum_j p_{ij}(g(i) + m_k(j) - m_{k+1}(i))^2 + \sum_j p_{ij} v_k(j).$$

Exercise 8. Consider a discrete-time, finite-state Markov chain $\{X_t\}$, with states $\{1, \ldots, n\}$, and transition probabilities p_{ij}. States 1 and n are absorbing, that is, $p_{11} = 1$ and $p_{nn} = 1$. All other states are transient. Let A_1 be the event that the state eventually becomes 1. For any possible starting state i, let $a_i = \mathbf{P}(A_1 \mid X_0 = i)$ and assume that $a_i > 0$ for every $i \neq n$. Conditional on the information that event A_1 occurs, is the process X_n necessarily Markov? If yes, provide a proof, together with a formula for its transition probabilities. If not, provide a counterexample.

Solution: The answer is yes. Let B be an event of the form

$$B = \{X_0 = i_0, X_1 = i_1, \ldots, X_{t-1} = i_{t-1}\}.$$

It suffices to show that the transition probability $\mathbb{P}(X_{t+1} = j \mid X_t = i, A_1, B)$ is unaffected by the past history (the event B). We have

$$\mathbb{P}(X_{t+1} = j \mid X_t = i, A_1, B) = \frac{\mathbb{P}(X_{t+1} = j, A_1 \mid X_t = i, B)}{\mathbb{P}(A_1 \mid X_t = i, B)}.$$

By the Markov property of the process $\{X_t\}$ (the future is independent of the past, given the present), we have

$$\mathbb{P}(X_{t+1} = j, A_1 \mid X_t = i, B) = \mathbb{P}(X_{t+1} = j, A_1 \mid X_t = i),$$

and

$$\mathbb{P}(A_1 \mid X_t = i, B) = \mathbb{P}(A_1 \mid X_t = i),$$

from which the desired result follows.

Furthermore,

$$\begin{aligned}
\mathbb{P}(X_{t+1} = j \mid X_t = i, A_1) &= \frac{\mathbb{P}(X_{t+1} = j, A_1 \mid X_t = i)}{\mathbb{P}(A_1 \mid X_t = i)} \\
&= \frac{\mathbb{P}(A_1 \mid X_t = i, X_{t+1} = j)\mathbb{P}(X_{t+1} = j \mid X_t = i)}{\mathbb{P}(A_1 \mid X_t = i)} \\
&= \frac{p_{ij} a_j}{a_i}.
\end{aligned}$$

Exercise 9. A certain production device is in one of two states at any time: operational or repair. The operation time of the device is a random variable which is uniformly distributed over the integers $1, 2, \dots, n$. The repair time has a deterministic value m which is a positive integer. The operation mode and repair mode alternate and the process continues indefinitely.

1. For every $1 \le i \le n$, and $t \ge 0$, let $X_t = i$ if the system is operational at time t and has been operational continuously for i time units (that is it was operational at times $t - i + 1, \dots, t$, but was in the repair mode at time $t - i$). For every $n + 1 \le i \le n + m$, let $X_t = i$ if the system is in the repair mode at time t and the repair mode began at time $t - i + 1$. Show that X_t is a Markov chain. Identify the transition rates for this M.c., its transient and recurrent states and steady state distributions. How many are there?

2. Suppose we observe the system in steady state. What is the likelihood that at the time of the observation the system is operational and has been operational at least $(2/3)n$ time units?

Solution:

1. The Markov chain has been drawn in Figure 1. We can justify the transition probabilities out of state i for $1 \le i \le n$ as follows. If $X \sim \text{Uni}(\{1, \dots, n\})$, then X conditional on $X \ge i$ is $\text{Uni}(\{i, i+1, \dots, n\})$.

A simple computation shows this. Thus as $|\{i, \ldots, n\}| = n - i + 1$, with probability $1/(n - i + 1)$ the machine will fail in the next time period. As the dynamics of X_n are described by the one step transition probabilities from each state, X_n is a Markov chain. As there is the path $1, 2, \ldots, n, n + 1, \ldots, n + m, 1$, all the states are in the same recurrent class.

If $n = 1$, the MC is periodic, and not very interesting. From now on, we assume $n > 1$. Then the MC is aperiodic (this takes some work, but intuitively, if we can have loops of size $m + 1$ or $m + 2$, then for all $n > (m + 2)^2$, you can be in state one). From the steady state equation $\pi' P = \pi'$, we have

$$\begin{aligned}
\pi_1 &= \pi_{n+m} \\
\pi_2 &= \left(\frac{n-1}{n}\right)\pi_1 \\
\pi_3 &= \left(\frac{n-2}{n-1}\right)\pi_2 = \frac{n-2}{n}\pi_1 \\
&\vdots \\
\pi_i &= \left(\frac{n-i+1}{n-i+2}\right)\pi_{i-1} = \frac{n-i+1}{n}\pi_1 && i = 2, \ldots, n \\
\pi_{n+1} &= \frac{1}{n}\pi_1 + \frac{1}{n-1}\pi_2 + \cdots + \frac{1}{2}\pi_{n-1} + \pi_n \\
\pi_{n+2} &= \pi_{n+1} \\
&\vdots \\
\pi_{n+i} &= \pi_{n+i-1} && i = 2, \ldots, m
\end{aligned}$$

We can ignore the equation for π_{n+1} as we have a redundant equation, and we note that from the final $m - 2$ equations and the first equation, we have that $\pi_1 = \pi_{n+i}$ for all $i = 1, \ldots, m$. Now using that the probabilities must sum to one, we have

$$1 = \pi_1 + \sum_{i=2}^{n} \frac{n-i+1}{n}\pi_1 + m\pi_1 = \pi_1\left(m + 1 + \frac{n-1}{2}\right),$$

thus for $i = 1, \ldots, m$,

$$\pi_1 = \frac{1}{m + 1 + \frac{n-1}{2}} = \pi_{n+i}.$$

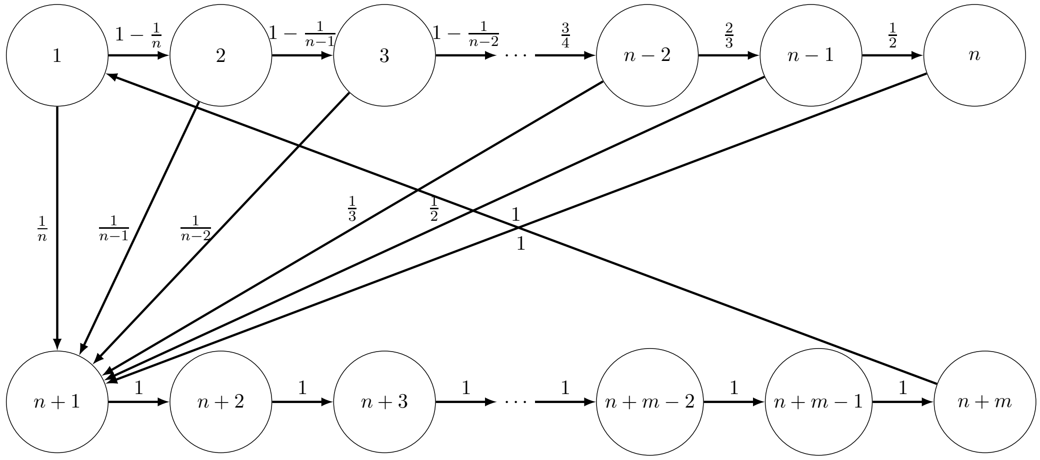

Figure 1: The Markov chain for problem 11. In the top row, the machine is in operation, while in the bottom row, the machine is undergoing repair.

and for $i = 2, \ldots, n$,

$$\pi_i = \frac{1}{m + 1 + \frac{n-1}{2}} \frac{n - i + 1}{n}$$

2.

$$\mathbb{P}\left(X_\infty \in \left\{\frac{2n}{3}, \ldots, n\right\}\right) = \sum_{i=2n/3}^{n} \pi_i = \pi_1 \sum_{i=2n/3}^{n} \frac{n-i+1}{n} = \pi_1 \left(\frac{n}{18} + \frac{1}{n} + \frac{1}{2} \right)$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Midterm

Boilerplate:

- No collaboration
- No internet, laptops, or cellphones
- Closed books, Closed notes.
- Cheat sheet allowed.
- Total: 100 pts

Exercise 1 (20 pts)**.** In honor of Laos' 65th independence day, some questions about independence.

1. Suppose A is independent of B, B is independent of C, and C is independent of A. Is $A \cap C$ necessarily independent of B?

   Prove or give counterexample

2. What if, in the above part, we also have that $B \cap C$ is independent of A?

   Prove or give counterexample

3. Suppose $A_1 \supset A_2 \supset A_3 \supset \ldots$ and $A_n \to A$, and suppose that B is independent of every A_n. Is B necessarily independent of A?

   Prove or give counterexample

4. Suppose A is independent of *itself*. What does this say about $\mathbb{P}(A)$?

   Don't laugh - this actually happens and is crucial to prove certain very strong theorems, which we will go over later!

Bonus: No points, but a gold star for anyone who knows what country Laos gained independence from!

Exercise 2 15pts **.** Let $A_1, \ldots, A_n$ be events. Let $X(\omega)$ be the number of events that occurred when ω was the elementary outcome. Show

$$\sum_{1 \le i < j \le n} \mathbb{P}[A_i \cap A_j] = \mathbb{E}\left[\frac{X(X-1)}{2}\right] .$$

Use this to prove

$$\mathbb{P}[\cup_{i=1}^{n} A_i] \geq \sum_{i=1}^{n} \mathbb{P}[A_i] - \sum_{1\leq i<j\leq n} \mathbb{P}[A_i \cap A_j]\,.$$

Hint: rewrite this as $\mathbb{E}[f(X)] \geq 0$.)

Exercise 3 15pts **.** Let X be a random variable taking values on non-negative integers $\mathbb{Z}_+ = \{0, 1, \ldots\}$. It has the following amazing property: There is a constant $c > 0$ such that for any bounded function $f : \mathbb{Z}_+ \to \mathbb{R}$ we have

$$\mathbb{E}[f(X)] = c\mathbb{E}[Xf(X-1)]\,.$$

Note: c does not depend on f.) Find distribution of X.

Exercise 4 20pts **.** Let X, Y be two independent standard normal random variables $\mathcal{N}(0, 1)$. Let $Z = X^2 + Y^2$. Recall you don't need to prove it) that Z has pdf $f_Z(z) = \frac{1}{2}e^{-z/2}$, i.e. $Z \sim \mathrm{Exp}(1/2)$.

1. Show that if U is independent of Z and uniform on $[0, 2\pi)$ then $\sqrt{Z}\sin mU$ is standard normal for any positive integer m.
2. Show that $T = \frac{2XY}{\sqrt{X^2+Y^2}}$ is standard normal. Hint: use polar coordinates

Exercise 5 30pts **.** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. Bernoulli random variables coin tosses , such that $\mathbb{P}(X_1 = H) = p \in (0, 1)$. Let

$$L_n = \max\{m \geq 0 : X_n = H,\ X_{n+1} = H,\ \ldots,\ X_{n+m-1} = H,\ X_{n+m} = T\}$$

be the length of the run of heads starting from the n-th coin toss. Prove that

$$\limsup_{n\to\infty} \frac{L_n}{\log(n)} = \frac{1}{\log(1/p)} \qquad \text{a.s..} \tag{1}$$

Steps:

1. Show that events $\{L_n \geq r\}, \{L_{n+r} \geq r\}, \{L_{n+2r} \geq r\}, \ldots$ are jointly independent.
2. Show that for any random variables Z_n

   $$\mathbb{P}[Z_n > \beta\text{-i.o.}] = 0 \quad \Rightarrow \quad \limsup Z_n \leq \beta \text{ a.s.}$$

   and

   $$\mathbb{P}[Z_n > \beta\text{-i.o.}] = 1 \quad \Rightarrow \quad \limsup Z_n \geq \beta \text{ a.s.}$$

3. Prove 1 . Hint: Borel-Cantelli

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Final Exam

Boilerplate:

- No collaboration
- No internet
- Closed books, Closed notes.
- Two (2-sided) cheat sheets allowed.
- Total: 100 pts
- **Partial credit will be given** (but please write clearly).

Exercise 1 (10 pts)**.** (Cautious Gambler's ruin) A gambler starts with $k \in [0, n]$ dollars and at each time step either skips a turn, or bets and wins 1 dollar, or bets and loses 1 dollar – all three cases happening with equal probability, independently across time. If he gets to n dollars, he stops and we say he "won". If he gets to 0 dollars he also stops and we say he is "ruined".

1. Show that eventually he must win or be ruined.
2. Find the probability that he wins.

Exercise 2 (20 pts)**.** Suppose $A_1, A_2, \dots$ are independent events with $\mu_n = \sum\limits_{i=1}^{n} \mathbb{P}(A_i) \to \infty$ as $n \to \infty$. Let

$$X_n := \frac{1}{\mu_n} \sum_{i=1}^{n} 1_{A_i}$$

1. Prove $X_n \overset{\text{i.p.}}{\to} 1$
2. Prove $X_n \overset{L_1}{\to} 1$. (Hint: $\mathbb{E}[|V|] \le \sqrt{\mathbb{E}[V^2]}$. What do you know about the variance of the sum of independent RVs?)
3. Prove $X_n \overset{\text{a.s.}}{\to} 1$ (Hint: first, let subsequence n_k be such that $(k-1)^4 \le \mu_{n_k} \le k^4$. What can you say about $\{|X_{n_k} - 1| > \frac{1}{k}\text{–i.o.}\}$?)

Exercise 3 (15 pts)**.** Two players are playing the following game. At time $t \ge 1$ both players generate random moves: player A's move is A_t and player B's move is B_t. The moves are iid and independent of each other. If $\sum_{s=1}^{t} A_s \ge \sum_{s=1}^{t} B_s$ then player A is declared a winner at step t (and we set $X_t = 0$), otherwise we say player B is a winner at step t (and set $X_t = 1$).

1. Let A_t be ± 2 with equal probability and B_t be ± 1 with equal probability. Find $\lim_{t\to\infty} \mathbb{P}[X_t = 1]$.
2. Now suppose both A_t and B_t are ± 1 with equal probability (and independently of each other). Find $\lim_{t\to\infty} \mathbb{P}[X_t = 1]$.
3. In the setting from part 2 show that X_t almost surely does not converge. (Hint: How many times do two symmetric random walks on $\mathbb{Z}$ meet?)

Exercise 4 (10pts)**.** Consider a finite-state homogeneous Markov chain with transition matrix $P(i, j)$. Suppose that for a certain state T we have $P(T, T) = 1$ and $P(i, T) = \epsilon > 0$ for all $i \neq T$. Let $X_0, X_1, \ldots, X_n$ be a trajectory of this Markov chain, started from some distribution $X_0 \sim \pi$ with $\pi(T) < 1$. Show that conditioned on $X_n \neq T$ the law of the sequence $X_0, \ldots, X_n$ is still a Markov chain. Find its transition matrix $\tilde{P}$ and initial distribution $\tilde{\pi}$ (i.e. $\tilde{\pi}$ is the law of X_0 given $\{X_n \neq T\}$). (Hint: you need to compute $\mathbb{P}[X_0 = a_0, \ldots, X_n = a_n | X_n \neq T]$ and factorize it.)

Exercise 5 (10 pts)**.** Two friends are observing iid sequence $X_i \sim \mathrm{Ber}(p)$ with unknown $p \in [0, 1]$, which they are trying to learn from the observations. They decide to record their observations as a running sum $S_t = \sum_{i=1}^{t} X_i$. Having observed n samples they start arguing. One says that they can write down the value S_n and forget $S_1, S_2, \ldots, S_{n-1}$, since it won't help in determining p. The other one argues that there might be some useful information in the trajectory $S_1, \ldots, S_{n-1}$ that will help learn p better. Who is right? (Hint: find $\mathbb{P}[S_1 = a_1, \ldots S_{n-1} = a_{n-1} | S_n = a_n]$ as a function of p.)

Exercise 6 (15pts)**.** Let X_i be independent with $\mathbb{P}[X_i = \frac{1}{p_i}] = 1 - \mathbb{P}[X_i = 0] = p_i$. Let $M_t = \prod_{i=1}^{t} X_i$, and $M_0 = 1$. Denote $a = \prod_{i=1}^{\infty} p_i$.

1. Find $\mathbb{E}[M_t]$.
2. Show that M_t converges almost surely to a random variable M_∞ and find its distribution. (Hint: you may want to consider cases of $a = 0$ and $a > 0$ separately).
3. If $a = 0$ is collection $\{M_t, t = 0, 1, \ldots\}$ uniformly integrable? (Hint: compute $\mathbb{E}[M_\infty]$)
4. If $a > 0$ is collection $\{M_t, t = 0, 1, \ldots\}$ uniformly integrable?

Exercise 7 (20 pts)**.** Two drunks walk along a street with n blocks (and $n+1$ intersections labeled $0, 1, \ldots, n$), starting at locations a and b (where a, b have the same parity, i.e. $a = b \bmod 2$). Two bars are located at 0 and n. Before he reaches a bar, each drunk performs a random walk, moving either left or right at every step with probability $1/2$ (independent of past steps); when a drunk arrives at a bar, he stops walking and goes in. If the two drunks meet they keep moving together ever after (still with probability 1/2 left-right).

What is the probability that the two drunks meet (either at a bar or while walking)? Consider three cases:

1. Their random walks are independent.
2. At each step they either both move right or both move left with equal probability. (If one of them is already captured by a bar, then only the remaining one keeps moving randomly.)
3. At each step they either both towards each other, or away from each other with equal probability. (If one of them is already captured by a bar, then only the remaining one keeps moving randomly.)

Hint: Think about union and intersection of events "left drunk goes to right bar", "right drunk goes to left bar".

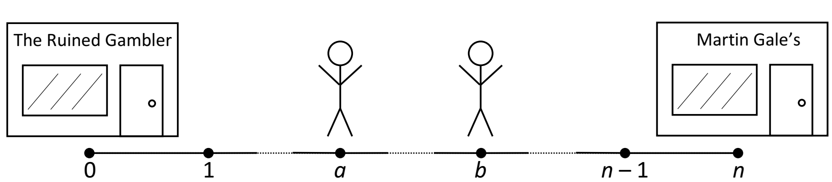

Figure 1: The drunks at the start of the process.