In this transcript, we will basically cover the textbook *Probability and Random Process by Geoffrey R Grimmett, David R Stirzaker* which used in course MATH2431 Honors Probability in HKUST. However, instead of using wording appeared in the textbook, I would use the wording that I can understand better personally. If you find any problems regarding this transcript, please contact me via `khliuae@connect.ust.hk`. Thanks a lot.

Contents

1 Sample space, σ-field, measure space, and probability

1. Initial Terminology
   (a) Experiments/ Trials: coin flipping, die rolling, lifetime of bulb
   (b) Outcome/ result: {H,T} {1,2,3,4,5,6}, $t \in [0, \infty)$
   (c) Probability: $\{\frac{1}{2}, \frac{1}{2}\}$, $\{\frac{1}{6}, ... \frac{1}{6}\}$,
2. Sample space Ω
   the set of all outcomes (denoted by ω) of an experiment, denoted as Ω, given an experiment
   $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$

coin flipping	$\Omega = \{H, T\}$
die rolling	$\Omega = \{1, 2, 3, 4, 5, 6\}$
lifetime of light bulb	$\Omega = [0, \infty)$
two coins flipping	$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$

3. Event E
   (a) event is a subset of the sample space Ω
   Remarks: all events are subsets of Ω, but not all subsets of Ω are events.

Die rolling $\Omega = 1, 2, 3, 4, 5, 6$	
"the outcome is even"	in words
"outcome is $2, 4, 6$"	in math
"the event $2, 4, 6$ occurs"	jargon

   (b) elementary events
   In an unknown number of die rollings, if the outcome $w = 2$ occurs, many events will occur, e.g., $\{2\}$, $\{2, 4, 6\}$, $\{1, 2, 3\}$,...
   (c) example: tossing coins until the first head turns up
   $\Omega = \{\omega_1, \omega_2, \omega_3, ...\}$, where ω_i denotes the outcome when the first $i - 1$ tosses are tails and the i th toss is a head. Let event A be that the first head occurs after an even number of tosses, i.e., $A = \{\omega_2, \omega_4, \omega_6, ...\} \subset \Omega$.
4. Set notations for events

$E \cup F$	E or F
$E \cap F$	E and F
E^c	Not E

5. Power set $2^\Omega, 0, 1^\Omega$
   Power set is the class of all subsets of Ω.
6. Field $\mathcal{F}$: a pre-version of σ-field
   a sub-collection of the set of all subsets of Ω, and satisfying:
   (a) if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$ (and thus $A \cap B \in \mathcal{F}$ by *De Morgan's Law*).
   (b) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
   (c) $\phi \in \mathcal{F}$, ($\Omega \in \mathcal{F}$ by (ii))

   To use layman languages to explain that,
   (a) that's a definition in modern probability
   (b) if we know $\mathbb{P}(A)$, then we have know that $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

(c) we must know $\mathbb{P}(\phi) = 0, \mathbb{P}(\Omega) = 1$

With properties (a), (b), (c) of a field $\mathcal{F}$, it follows that it's **closed under finite unions** (and thus interceptions by *De Morgan's Law*).

$$\text{if } A_1, A_2, A_3, ..., A_n \in \mathcal{F}, \text{ then } \bigcup_{i=1}^{\infty} A_i$$

Example: tossing coins until the first head turns up (revisited)
As discussed before, $\Omega = \{\omega_1, \omega_2, \omega_3, ...\}$, $A = \{\omega_2, \omega_4, \omega_6, ...\}$. A is an infinite countable union of of members of Ω and we require that $A \in \mathcal{F}$ in order to discuss its probability.

7. σ-field $\mathcal{F}$: a class of events of *interest*
   Why σ-field but not just power set? It is because we're only interested in certain events.
   σ-field $\mathcal{F}$ is a collection of set of subsets of Ω satisfying:

   (a) $\phi \in \mathcal{F}$

   (b) **closed under countably (finite/ infinite) unions** (and thus interceptions by *De Morgan's Law*):

   $$\text{if } A_1, A_2, ... \in \mathcal{F}, \text{ then } \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

   (c) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

   We observe that the only difference between a field and a σ-field is as follows:

Field	σ-field
closed under finite unions	closed under countably (finite/ infinite) unions

   Remark:

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

   This will be used in countable properties for probability measure later. Examples of σ-field

   (a) smallest σ-field associated with Ω is $\mathcal{F} = \{\phi, \Omega\}$
   Proof:
      i. $\phi \in \mathcal{F}$
      ii. $\phi, \Omega \in \mathcal{F} \implies \phi \cup \Omega = \Omega \in \mathcal{F}$
      iii. $\phi^c = \Omega \in \mathcal{F}, \Omega^c = \phi \in \mathcal{F}$

   (b) largest σ-field associated with Ω is $\mathcal{F}$ = power set of $\Omega = 2^{\Omega} = \{0, 1\}^c$
   Power set is the collection/class of all subsets of Ω.
   Proof:
      i. $\phi \in 2^{\Omega} = \mathcal{F}$
      ii. $\forall P_n \in 2^{\Omega}$ which Ω is finite set, $\bigcup_{i=0}^{\infty} P_n = \Omega \in \mathcal{F}$
      iii. $\forall P \in 2^{\Omega}, \exists P^c \in 2^{\Omega}$ since 2^{Ω} includes all the subsets of Ω.

   (c) if A that is any subset of Ω, then $\mathcal{F} = \{\phi, A, A^c, \Omega\}$ is a σ-field.
   Proof:
      i. $\phi \in 2^{\Omega} = \mathcal{F}$

ii. $\phi \cup A \cup A^c \cup \Omega = \Omega \in \mathcal{F}$
$\phi \cup \Omega = \Omega \in \mathcal{F}$
$A \cup A^c = \Omega \in \mathcal{F}$...

iii. $\phi^c = \Omega \in \mathcal{F}$,
$A^c, (A^c)^c = \Omega \in \mathcal{F}$,
$\Omega^c = \phi \in \mathcal{F}$

8. Exercises for Section 1.2 in textbook

   (a) Let $\{A_i : i \in I\}$ be a collection of sets. Prove "De Morgan's Laws":

$$(\bigcup_i A_i)^c = \bigcap_i A_i^c, \ (\bigcap_i A_i)^c = \bigcup_i A_i^c.$$

   For $(\bigcup_i A_i)^c = \bigcap_i A_i^c$,
   suppose $x \in (\bigcup_i A_i)^c \iff x \notin \bigcup_i A_i \iff \forall i, x \notin A_i \iff \forall i, x \in A_i^c \iff x \in \bigcap_i A_i^c$.

   For $(\bigcap_i A_i)^c = \bigcup_i A_i^c$,
   suppose $x \in (\bigcap_i A_i)^c \iff x \notin \bigcap_i A_i \iff \exists i, x \notin A_i \iff \exists i, x \in A_i^c \iff x \in \bigcup_i A_i^c$

   (b) Let A and B belong to some σ-field $\mathcal{F}$. Show that $\mathcal{F}$ contains the sets $A \cap B$, $A \backslash B$, and $A \Delta B$.

   i. To show $A \cap B \in \mathcal{F}$,
   $A, B \in \mathcal{F} \implies A^c, B^c \in \mathcal{F} \implies A^c \cup B^c \in \mathcal{F} \iff A \cap B \in \mathcal{F}$.

   ii. To show $A \backslash B \in \mathcal{F}$, which is same as $A \cap B^c \in \mathcal{F}$,
   $A, B \in \mathcal{F} \implies A^c \in \mathcal{F} \implies A^c \cup B \in \mathcal{F} \implies (A^c \cup B)^c \in \mathcal{F} \iff A \cap B^c \in \mathcal{F}$.

   iii. To show $A \Delta B \in \mathcal{F}$, which $A \Delta B = (A \cup B) \backslash (A \cap B)$, that is also true by the facts shown above.

   (c) A conventional knock-out tournament (such as that at Wimbledon) begins with $2n$ competitions and has n rounds. There are no play-offs for the position $2, 3, ..., 2n - 1$, and the initial table of draws is specified. Give a concise description of the sample space of all possible outcomes. (TO BE COMPLETE)

9. Measurable space: $(\Omega, \mathcal{F})$

10. Probability (measure) $\mathbb{P}$
Given a measurable space $(\Omega, \mathcal{F})$, A probability on $(\Omega, \mathcal{F})$ is a function

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$

$$E \in \mathcal{F} \to \mathbb{P}(E)$$

,which is satisfying

   (a) $\mathbb{P}(\phi) = 0, \mathbb{P}(\Omega) = 1$

   (b) Countable additivity
   If $A_1, A_2, ... \in \mathcal{F}$ and they are *disjoint*, i.e., $A_i \cap A_j = \phi, \forall i \neq j$, then

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

   .

Collary: Countable additivity implies finite additivity.
Proof: Let $A_{k+1}, A_{k+2}, ... = \phi$. $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \mathbb{P}(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} \mathbb{P}(A_i)$
and finite additivity implies a lot of trivial things:
e.g., $\mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1 \implies \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
e.g., Let's say $A \subset B$, $\mathbb{P}(B) = \mathbb{P}(A \cup (B \backslash A)) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) > \mathbb{P}(A)$
e.g., Inclusion-exclusion formula (TO BE COMPLETE)

11. Probability space: $(\Omega, \mathcal{F}, \mathbb{P})$

12. General measure μ
Given a measurable space $(\Omega, \mathcal{F})$, a measure μ is a set function $\mu : \mathcal{F} \to [0, \infty]$, such that
    (a) $\mu(\phi) = 0$
    (b) countable additivity
$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$
e.g., Lebesgue measure: $\mu((a, b)) = b - a$, $\Omega = \mathbb{R}$.
e.g., Counting measure: $\mu(A) = |A|$, $\Omega = \mathbb{R}$.

13. Measure space: $(\Omega, \mathcal{F}, \mu)$
Caution: probability space is not part of measure space. They are in different areas of interests.

14. Continuity of probability measure
Recall: continuity of a function $f : \mathbb{R} \to \mathbb{R}$
f is continuous at some point x, if $\forall x_n, x_n \to x, n \to \infty$, we have
$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x)$$
Similarly, we say a set function μ is continuous, if $\forall A_n$ with $\lim_{n\to\infty} A_n = A$, we have
$$\lim_{n\to\infty} \mu(A_n) = \mu(\lim_{n\to\infty} A_n) = \mu(A).$$
Definition: set limit ($\lim_{n\to\infty} A_n = A$)
Recall: limit of numbers x_n, ...where $n \geq 1$
$$\limsup x_n = \lim_{m\uparrow\infty} \sup_{n\geq m} x_n = \lim_{m\uparrow\infty} c_m$$
$$\liminf x_n = \lim_{m\uparrow\infty} \inf_{n\geq m} x_n = \lim_{m\uparrow\infty} b_m$$
We say x_n is convergent in $\mathbb{R} \cup \{\pm\infty\}$ if $\limsup x_n = \liminf x_n$.
Given $A_1, A_2, ..., A_n, ... \in \mathcal{F}$,
$$\limsup\nolimits_{n\to\infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many n}\} = A_n \text{ infintely often}$$
Proof: Let $LHS = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$, $RHS = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many n}\} \implies \exists$,
If $\omega \in RHS \implies w \in \bigcup_{n=m}^{\infty} A_n, \forall m \geq 1 \implies \omega \in \bigcap_{n=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = LHS$
If $\omega \notin RHS \implies \exists m_0, \text{s.t. } \omega \notin \bigcap_{n=m}^{\infty} A_n \forall m \geq m_0 \implies \omega \notin \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = LHS$
$$\liminf\nolimits_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for all but finitely many } n\} = A_n \text{ all but finitely often}$$

Remarks: $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$
Definition: Set Limit $\lim_{n\to\infty} A_n$
We say A_n converges if $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$
Example: If $A_1 \subseteq A_2 \subseteq A_3 ...$

$$\begin{aligned}
\limsup_{n\to\infty} A_n &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n \\
&= \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} A_n \\
&= \bigcup_{n=1}^{\infty} A_n \\
\liminf_{n\to\infty} A_n &= \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n \\
&= (A_1 \cap A_2 \cap A_3 \cap ...) \cup (A_2 \cap A_3 \cap ...) \cup (A_3 \cap A_4 \cap ...) \cup ... \\
&= A_1 \cup A_2 \cup A_3 \cup ... \\
&= \bigcup_{n=1}^{\infty} A_n
\end{aligned}$$

As a result, $\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n \implies \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$
Example: If $A_1 \supseteq A_2 \supseteq A_3 \supseteq ...$,

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

Proof: by De Morgan's Law, If $A_1 \subseteq A_2 \subseteq A_3 \subseteq ...$, then $A_1^c \supseteq A_2^c \supseteq A_3^c \supseteq ....$
Let $B_i = A_i^c, \forall i \geq 1 \implies A_1^c \supseteq A_2^c \supseteq A_3^c \supseteq ... = B_1 \supseteq B_2 \supseteq B_3 \supseteq ...$,

$$\begin{aligned}
\lim_{n\to\infty} B_n &= \lim_{n\to\infty} A_n^c \\
&= (\lim_{n\to\infty} A_n)^c \\
&= (\bigcup_{n=1}^{\infty} A_n)^c \\
&= (\bigcap_{n=1}^{\infty} A_n^c) \\
&= \bigcap_{n=1}^{\infty} B_n
\end{aligned}$$

Remark:

$$\text{For } \mathbf{1}_{A(\omega)} = \begin{cases} 1, \text{if } \omega \in A \\ 0, \text{if } \omega \in A^c \end{cases} \tag{1}$$

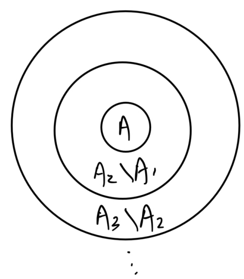
$\limsup_{n\to\infty} = \{\omega \in \Omega : \limsup_{n\to\infty} \mathbf{1}_{A(\omega)} = 1\}$
$\limsup_{n\to\infty} = \{\omega \in \Omega : \liminf_{n\to\infty} \mathbf{1}_{A(\omega)} = 1\}$

Theorem: Continuity of probability measure $\mathbb{P}$
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $A_1, A_2, A_3, ... \in \mathcal{F}$ such that $\lim_{n\to\infty} A_n (= A)$ exists. Then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\lim_{n\to\infty} A_n) = \mathbb{P}(A)$$

Proof: (for the special case where $A_1 \subseteq A_2 \subseteq A_3 ...$)
Notice that in this special case, $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$, so

$$\mathbb{P}(\lim_{n\to\infty} A_n) = \mathbb{P}(\bigcup_{n=1}^{\infty} A_n)$$



Let $B_n = A_n \backslash A_{n-1}$, $A_0 = \phi$

$$\begin{aligned}
\mathbb{P}(\lim_{n\to\infty} A_n) &= \mathbb{P}(\bigcup_{n=1}^{\infty} A_n) \\
&= \mathbb{P}(\bigcup_{n=1}^{\infty} B_n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(B_n) \text{ (countable additivity)} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}(B_n) \\
&= \lim_{N\to\infty} \mathbb{P}(\bigcup_{n=1}^{N} B_n) \\
&= \lim_{N\to\infty} \mathbb{P}(A_n)
\end{aligned}$$

2 Probability Generating function

1. Definition: generating function
Let's say we have a sequence $a_i = \{a_0, a_1, a_2, ...\}$, we can define a generating function $G_a(s)$ for a_i.

$$G_a(s) = \sum_{i=0}^{\infty} a_i s^i$$

We could get back a_i from $G_a(s)$ **in most but not all cases** (only when the derivatives and countable sum are inter-changable).

$$a_i = \frac{G_a^{(i)}(0)}{i!}$$

counter-example:

$$a_1 = \sin nx,\ a_n = \frac{\sin nx}{n} - \frac{\sin (n-1)x}{n-1} \text{ for } n = 2, 3, ...$$

2. Convolution
Suppose we have two sequence a_i, b_i,

$$a_i = \{a_0, a_1, a_2, ...\}, b_i = \{b_0, b_1, b_2, ...\}$$

we then define another sequence c_i,

$$c_n = a_n * b_n = \sum_{i=0}^{n} a_i b_{n-i}$$

$$c_i = \{a_0 b_n, a_0 b_n + a_1 b_{n-1}, a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2}\}$$

Then , we claim that $G_c(s) = G_a(s) G_b(s)$.
Proof:

$$\begin{aligned}
G_c(s) &= \sum_{i=0}^{\infty} c_n s^n \\
&= \sum_{n=0}^{\infty} \sum_{i=0}^{n} a_i b_{n-i} s^n \\
&= \sum_{n=0}^{\infty} \sum_{i=0}^{n} a_i b_{n-i} s^i s^{n-i} \\
&= \sum_{i=0}^{\infty} \sum_{n=i}^{\infty} a_i b_{n-i} s^i s^{n-i} \\
&= \sum_{i=0}^{\infty} a_i s^i \sum_{n=i}^{\infty} b_{n-i} s^{n-i} \\
&= \sum_{i=0}^{\infty} a_i s^i \sum_{n-i=0}^{\infty} b_{n-i} s^{n-i} \\
&= \sum_{i=0}^{\infty} a_i s^i \sum_{j=0}^{\infty} b_j s^j = G_a(s) G_b(s)
\end{aligned}$$

3. Definition: Probability Generating Function
   The probability generating function of a discrete random variable X taking non-negative integer value is given by
   $$G_X(s) = \mathbb{E}(s^X) = \sum_{i=0}^{\infty} f_X(i)s^i$$
   Power series (review of calculus II)
   Let's say we a power series $f(s) = \sum_{n=0}^{\infty} a_n s^n$, how to find its radius of convergence R?
   (a) use root test (definition)
   $$\begin{aligned} \lim_{n\to\infty} \sqrt[n]{|a_n s^n|} &< 1 \\ \lim_{n\to\infty} \sqrt[n]{|a_n|}|s| &< 1 \\ |s| = \frac{1}{\lim_{n\to\infty} \sqrt[n]{|a_n|}} &:= R \end{aligned}$$
   (b) use ratio test (informal)
   $$\begin{aligned} \lim_{n\to\infty} \left|\frac{a_{n+1}s^{n+1}}{a_n s^n}\right| &< 1 \\ \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right||s| &< 1 \\ |s| < \frac{1}{\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|} &:= R \end{aligned}$$
   Theorem If R is the radius of convergence of $G_a(s) = \sum_{n=0}^{\infty} a_n s^n$, then
   (a) $G_a(s)$ converges absolutely for all $|s| < R$ and diverges for all $|s| > R$.
   (b) for $|s| < R$, $G_a(s)$ is differentiable or integratable for any fixed number of times term by term (i.e. derivative and countable summation is interchangable).
   (c) for $R \neq 0$, and $G_a(s) = G_b(s)$ for $|s| < R'$ for some $0 < R' < R$, then $a_n = b_n$ for all n.

   Question Why do we care about $s = 1$?
   Because we can find the moment $\mathbb{E}X$ using $G'_X(1)$, however, (b) in above theorem haven't cover the case when $s = 1$. (we still don't know if we can inter-change derivative and countable summation when $s < R = 1$)
   This could be solved by applying abel theorem as following.

4. Abel Theorem
   If $a_n \geq 0 \ \forall n$, $G_a(s)$ is its generating function having the radius of convergence $R = 1$, then additionally if $\sum_{n=0}^{\infty} a_n$ converges in $\mathbb{R} \cup \{\infty\}$, then we have
   $$\lim_{s\to 1^-} G_a(s) = \lim_{s\to 1^-} \sum_{n=0}^{\infty} a_n s^n = \sum_{n=0}^{\infty} a_n \lim_{s\to 1^-} s^n = \sum_{n=0}^{\infty} a_n$$

5. Probability generating functions of some typical random variables

Random variables	Generating functions	interval of s
$Be(p)$	$G_X(s) = 1 - p + ps$	$\mathbb{R}$
$Bin(p)$	$G_X(s) = (1 - p + ps)^n$	$\mathbb{R}$
$Poisson(\lambda)$	$G_X(s) = e^{\lambda(s-1)}$	$\mathbb{R}$
$Geom(p)$	$G_X(s) = \frac{ps}{1-s(1-p)}$	$\lvert s\rvert < \frac{1}{1-p}$

6. Moment and probability generating functions
   If X has probability generating function $G_X(s)$ then
   (a) $\mathbb{E}X = \lim_{s\to 1^-} G'(s) := G'(1)$
   (b) $\mathbb{E}[X(X-1)(X-2)...(X-(k-1))] = \mathbb{E}\frac{X!}{(X-k)!} = \lim_{s\to 1^-} G^k(s) := G^k(1)$

   Remark

$$\begin{aligned} Var(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\ &= \mathbb{E}X(X-1) + \mathbb{E}(X) - (\mathbb{E}X)^2 \\ &= G''(1) + G'(1) - (G'(1))^2 \end{aligned}$$

7. Sum of independent random variables
   Theorem if $X \perp\!\!\!\perp Y$, then $G_{X+Y}(s) = G_X(s)G_Y(s)$
   Proof

$$f_{X+Y}(z) = f_X * f_Y(z) = \sum_{x=0}^{z} f_X(x) f_Y(n-x)$$

$$\begin{aligned} G_{X+Y}(z) &= \sum_{n=0}^{\infty} f_{X+Y}(z) z^n \\ &= \sum_{n=0}^{\infty}\sum_{x=0}^{z} f_X(x) f_Y(n-x) z^x z^{n-x} \\ &= \sum_{x=0}^{\infty}\sum_{z=x}^{\infty} f_X(x) f_Y(n-x) z^x z^{n-x} \\ &= \sum_{x=0}^{\infty} f_X(x) z^x \sum_{z=x}^{\infty} f_Y(n-x) z^{n-x} \\ &= G_x(z) G_Y(z) \end{aligned}$$

   Caution: its converse may not be true!

8. Sum of random number of random variables
   Theorem
   If $X_1, X_2, X_3, ..., X_N$ (i.i.d) with common probability generating function $G_X(s)$ and $N \geq 0$ is a random variable independent of all X_i's with probability generating function $G_N(s)$. Let

$$T = X_1 + X_2 + ... + X_N$$

   , then

$$G_T(s) = G_N(G_X(s))$$

Proof

$$\begin{aligned}G_T(s) &= \mathbb{E}s^T \\ &= \mathbb{E}(\mathbb{E}(s^T|N)) \\ &= \sum_{n=0}^{\infty}\mathbb{E}(s^T|N=n)\mathbb{P}(N=n) \\ &= \sum_{n=0}^{\infty}\mathbb{E}(s^{X_1+X_2+...+X_n})\mathbb{P}(N=n) \\ &= \sum_{n=0}^{\infty}(G_X(s))^n\mathbb{P}(N=n) \\ &= G_N(G_X(s))\end{aligned}$$

Example Sum of a Poisson number of independent Bernouli is still a Poisson.
A hen lays N eggs where $N \sim Poisson(\lambda)$. Each egg hatches with probability p independently. Let K be the number of chicks. Then,

$$K = X_1 + X_2 + ... + X_N$$

, where $X_i \sim Be(p)$ is the indicator random variable indicating where the i-th egg hatches or not, i.e.,

$$X_i = 1, \text{if i-th egg hatches}, X_i = 0, \text{ otherwise}$$

Therefore we can just apply the formula,

$$\begin{aligned}G_K(s) &= G_N(G_X(s)) \\ &= e^{\lambda(G_X(s)-1)} \\ &= e^{\lambda(1-p+ps-1)} \\ &= e^{\lambda p(1-s)}\end{aligned}$$

Therefore, we could deduce that $K \sim Poisson(\lambda p)$

9. Multivariate extension: joint probability generating function

$$G_{X_1,X_2}(s_1,s_2) = \mathbb{E}{s_1}^{X_1}{s_2}^{X_2} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}s_1^i s_2^j f_{X_1,X_2}(i,j)$$

$$f_{X_1,X_2}(i,j) = \frac{1}{i!j!}\frac{\partial^{i+j}}{\partial s_1^i \partial s_2^j}G_{X_1,X_2}(s_1,s_2)|_{(s_1,s_2)=(0,0)}$$

Theorem $X_1 \perp\!\!\!\perp X_2 \iff G_{X_1,X_2}(s_1,s_2) = G_{X_1}(s_1)G_{X_2}(s_2)$

10. Application 1: recurrence/ transience of 1D simple random walk
Let $S_n = \sum_{i=0}^{n} X_i$, $S_0 = 0$, $n \geq 0$, where S_n is the location at n-step, X_i is representing the i-th step such that $\mathbb{P}(X_i = 1) = p, \mathbb{P}(X_i = -1) = 1 - p := q$.
Let $T_0 := \min\{i \geq 0 : S_i = 0\}$, representing the first return time to the origin (i.e., how many steps to take for return to the origin?).

Remark Y is "defective" random variable $\iff \mathbb{P}(Y = \infty) > 0$

Question When is T_0 "defective"? It suffices to find $\mathbb{P}(T_0 = \infty)$ for different p and q.
Solution
Let $f_0(n) := \mathbb{P}(S_1 \neq 0, S_2 \neq 0, S_3 \neq 0, ..., S_n = 0) = \mathbb{P}(T_0 = n)$ for $n \geq 1$.
Then, define its generating function $F_0(n)$

$$\begin{aligned} F_0(n) &= \sum_{n=1}^{\infty} f_0(n)s^n \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} f_0(n)s^n \\ &\text{Since } \infty \text{ is not measured by the random variable } T_0 \\ F_0(1) &= \lim_{N\to\infty} \sum_{n=1}^{N} f_0(n) = 1 - \mathbb{P}(T_0 = \infty) \end{aligned}$$

Also, let $p_0(n) = \mathbb{P}(S_n = 0) = \binom{n}{\frac{n}{2}} p^{n/2} q^{n/2}$ for even n and $p_0(0) = \mathbb{P}(S_0 = 0) = 1$ since initial location is 0.
Similarly, define its generating function $P_0(s) := \sum_{n=1}^{\infty} p_0(n)s^n = (1 - 4pqs^2)^{-\frac{1}{2}}$ after some calculations
Theorem

(a) $P_0(s) = 1 + P_0(s)F_0(s)$
(b) $P_0(s) = (1 - 4pqs^2)^{-1/2}$ (known)
(c) $F_0(s) = 1 - (1 - 4pqs^2)^{1/2}$ (easy by (a) and (b))

Proof of (a)
Let $A_n = \{S_n = 0\}$ (i.e., at n-th step, it returns to origin),
$B_k = \{S_1 \neq 0, S_2 \neq 0, ..., S_{k-1} \neq 0, S_k = 0\}$ (i.e., at k-th step, it returns to origin first time).
Note that $p_0(n) = \mathbb{P}(A_n)$, $f_0(k) = \mathbb{P}(B_n)$.
By law of total probability,

$$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(A_n|B_k)\mathbb{P}(B_k)$$

Let's consider

$$\begin{aligned} \mathbb{P}(A_n|B_k)\mathbb{P}(B_k) &= \mathbb{P}(S_n = 0|S_1 = 0, S_2 = 0, ..., S_k = 0) \\ &= \mathbb{P}(S_n = 0|S_k = 0) \\ &= \mathbb{P}(S_{n-k} = 0) \\ &= \mathbb{P}(A_{n-k}) \\ &= p_0(n-k) \end{aligned}$$

Therefore, we have

$$\begin{aligned} p_0(n) &= \mathbb{P}(A_n) \\ &= \sum_{k=1}^{n} \mathbb{P}(A_n|B_k)\mathbb{P}(B_k) \\ &= \sum_{k=1}^{n} p_0(n-k)f_0(k), \text{ for } n >= 1 \\ p_o(0) &= 1 \end{aligned}$$

As a result, we have

$$\begin{aligned}
P_0(s) &= \sum_{n=0}^{\infty} p_0(n)s^n \\
&= p_0(0) + \sum_{n=1}^{\infty}\sum_{k=1}^{n} p_0(n-k)f_0(k)s^n \\
&= 1 + \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} p_0(n-k)f_0(k)s^n \\
&= 1 + \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} p_0(n-k)f_0(k)s^k s^{n-k} \\
&= 1 + P_0(s)F_0(s)
\end{aligned}$$

Conclusion

case 1

If $p = q = \frac{1}{2}$, then $F_0(1) = 1-(1-4(1/2)(1/2)(1)^2)^{1/2} = 1$. Therefore, $\mathbb{P}(T_0 = \infty) = 1 - F_0(1) = 0$. In this case, we call this a recurrence since T_0 is not "defective".

case 2

If $p \neq q$, then $F_0(1) < 1$. $\mathbb{P}(T_0 = \infty) > 0$. We call this a transient since T_0 is "defective".

3 Characteristic Function

1. definition: The characteristic function of a given random variable X is given by

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx = \int_{-\infty}^{\infty} \cos itx f_X(x)dx + i\int_{-\infty}^{\infty} \sin itx f_X(x)dx$$

4 Convergence of Random Variables

1. Modes of convergence of random variables

   (a) Almost surely convergence

   $$X_n \xrightarrow{a.s.} X \iff \mathbb{P}(\lim_{n\to\infty} X_n(\omega) \to X, \forall\omega \in \Omega) \iff \mathbb{P}(|X_n - X| \geq \epsilon, i.o.) = 0$$

   Remark: Almost surely convergence can be proved by B.C.I and disproved by B.C.II.

   (b) Convergence in rth-mean

   $$X_n \xrightarrow{r} X \iff \mathbb{E}(|X_n - X|^r) \to 0, r \geq 1$$

   Alternatively, we have proof $\mathbb{E}(|X_n - X|^r)^{\frac{1}{r}} \to 0$.
   Remark:$X_n \xrightarrow{r} X \implies X_n \xrightarrow{l} X$ for $1 \leq l < r$

   (c) Convergence in probability

   $$X_n \xrightarrow{\mathbb{P}} X \iff \mathbb{P}(|X_n - X| \geq \epsilon) \to 0$$

   (d) Convergence in distribution

   $$X_n \xrightarrow{D} X \iff \mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$$

2. Ways to proof convergences

   (a) Almost surely convergence

      i. Directly show for 'almost' all ω in Ω,

      $$\mathbb{P}(\lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$$

      ii. (Seldom) Show

      $$\forall\epsilon > 0, \mathbb{P}(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$$

      iii. Use Borel-Cantelli Lemma I
      Show that

      $$\forall\epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty$$

      , then we can apply B.C. I to show that

      $$\mathbb{P}(\limsup_{n\to\infty} |X_n - X| > \epsilon) = 0$$

      iv. (As complement of (iii))
      If we find $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) = \infty$,
      show that for $\epsilon > 0$, let $B_n = \{|X_n - X| < \epsilon\}$,

      $$X_n \xrightarrow{a.s.} X \iff \lim_{n\to\infty} \mathbb{P}(\bigcap_{m=n}^{\infty} B_m) = 1$$

Proof:
For all $\epsilon > 0$, let $A_n = \{|X_n - X| \geq \epsilon\}$.

$$\begin{aligned}
X_n \xrightarrow{a.s.} X &\iff \mathbb{P}(A_n, i.o.) = 0 \\
&\iff \mathbb{P}(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n) = 0 \\
&\iff \mathbb{P}((\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n)^c) = 1 \\
&\iff \mathbb{P}(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c) = 1 \\
&\iff \lim_{n\to\infty} \mathbb{P}(\bigcap_{n=m}^{\infty} A_n^c) = 1 \\
&\iff \lim_{n\to\infty} \mathbb{P}(\bigcap_{n=m}^{\infty} B_n) = 1 \qquad (B_n := A_n^c)
\end{aligned}$$

Remark: If X_i's are independent, then

$$X_n \xrightarrow{a.s.} X \iff \lim_{n\to\infty} \prod_{n=m}^{\infty} \mathbb{P}(B_n) = 1$$

(b) Convergence in r-th mean

i. Directly show $\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0$ or $\lim_{n\to\infty}(\mathbb{E}|X_n - X|^r)^{\frac{1}{r}} = 0$

ii. Useful formulas

- Triangular inequality

$$\begin{aligned}
(\mathbb{E}|X_n|^r)^{\frac{1}{r}} = (\mathbb{E}|X_n - X + X|^r)^{\frac{1}{r}} &\leq (\mathbb{E}|X_n - X|^r)^{\frac{1}{r}} + (\mathbb{E}|X|^r)^{\frac{1}{r}} \\
\iff (\mathbb{E}|X_n|^r)^{\frac{1}{r}} - (\mathbb{E}|X|^r)^{\frac{1}{r}} &\leq (\mathbb{E}|X_n - X|^r)^{\frac{1}{r}}
\end{aligned}$$

- Hölder's inequality
  For $1 < p, q < \infty, \dfrac{1}{p} + \dfrac{1}{q} = 1$,

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

(c) Convergence in probability

i. Directly show for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0$ or $\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1$

ii. (Partial converse theorem) Show that $X_n \xrightarrow{D} c$ which implies $X \xrightarrow{P} c$.

Proof:
Suppose $X_n \xrightarrow{D} c$, then for all $\epsilon > 0$, $\lim_{n\to\infty} F_{X_n}(c - \epsilon) = 0$ and $\lim_{n\to\infty} F_{X_n}(c + \epsilon) = 1$.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) &= \lim_{n\to\infty} \mathbb{P}(X_n - c \geq \epsilon) + \lim_{n\to\infty} \mathbb{P}(X_n - c \leq -\epsilon) \\
&= \lim_{n\to\infty} \mathbb{P}(X_n \geq \epsilon + c) + \lim_{n\to\infty} \mathbb{P}(X_n \leq -\epsilon + c) \\
&= \lim_{n\to\infty} (1 - \mathbb{P}(X_n < \epsilon + c)) + \lim_{n\to\infty} \mathbb{P}(X_n \leq -\epsilon + c) \\
&= \lim_{n\to\infty} (1 - F_{X_n}(c + \epsilon)) + \lim_{n\to\infty} F_{X_n}(c - \epsilon) \\
&= (1 - 1) + 0 \\
&= 0
\end{aligned}$$

Thus, $X \xrightarrow{P} c$.

iii. Useful properties
- If $\omega \in A \implies \omega \in B$, then
$$\mathbb{P}(A) \subseteq \mathbb{P}(B)$$
- Triangular inequality
$|X+Y| \leq |X| + |Y|$,
$|X| = |X-Y+Y| \leq |X-Y| + |Y|$
- Markov's inequality
$$\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}(g(X))}{g(\epsilon)}$$

(d) Convergence in distribution

i. Directly show $\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \lim_{n\to\infty} \mathbb{P}(X \leq x)$

ii. For non-negative integer valued discrete random variables
If $X_1, X_2, X_3, ...$ are non-negative integer valued random variables, then
$$X_n \xrightarrow{D} X \iff \lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x)$$

Proof:

A. Proving $X_n \xrightarrow{D} X \implies \lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x)$
Suppose $X_n \xrightarrow{D} X \implies \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$,
Note: since X_n's are non-negative valued, then $F_{X_{n(x)}}$ is continuous in $x \in \mathbb{R}\backslash\{0, 1, 2, 3, ...\}$.
$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X_n = x) &= \lim_{n\to\infty} (\mathbb{P}(X_n = x + \frac{1}{2}) - \mathbb{P}(X_n = x - \frac{1}{2})) \\
&= \lim_{n\to\infty} F_{X_n}(x + \frac{1}{2}) - \lim_{n\to\infty} F_{X_n}(x - \frac{1}{2}) \\
&= F_X(x + \frac{1}{2}) - F_X(x - \frac{1}{2}) \\
&= \mathbb{P}(X = x)
\end{aligned}$$
Thus, $\lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x)$.

B. Proving $\lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x) \implies X_n \xrightarrow{D} X$
Suppose $\lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x)$,
$$\begin{aligned}
\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty} \mathbb{P}(X_n \leq x) \\
&= \lim_{n\to\infty} \sum_{i=0}^{\lfloor \frac{1}{x} \rfloor} \mathbb{P}(X_n = i) \\
&= \sum_{i=0}^{\lfloor \frac{1}{x} \rfloor} \mathbb{P}(X = x) \\
&= \mathbb{P}(X \leq x) \\
&= F_X(x)
\end{aligned}$$
Thus, we have $X_n \xrightarrow{D} X$.

3. Relationships between modes of convergence

(a) $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{\mathbb{P}} X$

(b) $X_n \xrightarrow{r} X \implies X_n \xrightarrow{\mathbb{P}} X$

(c) $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{D} X$

4. Borel-Cantelli Lemma
Let's say A_n is a sequence of event in $\mathcal{F}$.

(a) B.C. I
If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.
Proof:

$$\begin{aligned}
\mathbb{P}(\limsup_{n\to\infty} A_n) &= \mathbb{P}(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m) \\
&= \lim_{n\to\infty} \mathbb{P}(\bigcup_{m=n}^{\infty} A_m) \\
&\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_n) \\
&= 0 && \text{(by the assumption in the if statement)}
\end{aligned}$$

(b) B.C. II
If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and A_n's are independent, then $\mathbb{P}(\limsup_{n\to\infty}) = 1$.
proof:
We want to show that $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.
This is same as showing that

$$\begin{aligned}
\mathbb{P}((\limsup_{n\to\infty} A_n)^c) &= \mathbb{P}((\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m)^c) \\
&= \mathbb{P}(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_n^c) \\
&= \lim_{m\to\infty} \mathbb{P}(\bigcap_{n=m}^{\infty} A_n^c) \\
&= \lim_{m\to\infty} \lim_{r\to\infty} \mathbb{P}(\bigcap_{n=m}^{r} A_n^c) \\
&= \lim_{m\to\infty} \lim_{r\to\infty} \prod_{n=m}^{r} \mathbb{P}(A_n^c) \text{ (since } A_n\text{'s are independent)} \\
&= \lim_{m\to\infty} \lim_{r\to\infty} \prod_{n=m}^{r} (1 - \mathbb{P}(A_n)) \\
&\leq \lim_{m\to\infty} \prod_{n=m}^{\infty} e^{-\mathbb{P}(A_n)} \\
&= \lim_{m\to\infty} e^{\sum_{n=m}^{\infty} \mathbb{P}(A_n)} \\
&= \lim_{m\to\infty} e^{-\infty} \\
&= 0
\end{aligned}$$

Remark: B.C. II is a partial converse statement of B.C. I.

5. Different versions of Law of Large Numbers

   (a) WLLN

   If we have a sequence of random variables $X_1, X_2, ...$ (i.i.d.), with $\mathbb{E}(X_1) = \mu < \infty$, then let

   $$S_n = X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i$$

   By WLLN, we have

   $$\frac{S_n}{n} \xrightarrow{D/\mathbb{P}} \mu$$

   (b) L^2-WLLN

   If we have a sequence of random variables $X_1, X_2, ...$ where X_i's are uncorrelated with $\mathbb{E}(X_i) = \mu$ *(common mean)*, $Var(X_i) \leq c < \infty$ *(bound variance)* for all i, then let

   $$S_n = X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i$$

   By L^2-WLLN, we have

   $$\frac{S_n}{n} \xrightarrow{2} \mu$$

   (c) SLLN

   If we have a sequence of random variables $X_1, X_2, ...$ (i.i.d.), with $\mathbb{E}(X_1) = \mu$ and $\mathbb{E}(|X_1|) < \infty$, then let

   $$S_n = X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i$$

   By SLLN,

   $$\frac{S_n}{n} \xrightarrow{a.s.} \mu$$