MATH 487: Probability Theory

University of Nebraska-Lincoln, Fall 2017

Jacob Shiohira

December 13, 2017

The textbook used for this class is **asdf** by .

CHAPTER 1: PRELIMINARIES

Date: 8/22

The sample space Ω is the set of all outcomes of an experiment. For a few examples, we'll consider different types of sample spaces using descriptors such as finite, infinite, countable, and uncountable.

Finite and Countable

Example: Toss a coin twice - "H" for heads and "T" for tails,

$$\Omega = \Big\{(H,H),(H,T),(T,H),(T,T)\Big\}$$

The cardinality of Ω is then $|\Omega| = 4$. The sample space for this experiment is both finite **and** countable, so it is considered **discrete**.

Infinite and Countable

Example: Toss a coin until a heads appears. Let "H" represent heads and "T" represent tails. You can think of the outcomes as $\Big\{H,(T,H),(T,T,H),\ldots\Big\}$ where the first H occurs on the k^{th} toss, $k \in \mathbb{N}$. Thus, you can just refer to Ω as the set of natural numbers,

$$\mathbb{N} = \Big\{1,2,3,\ldots,k,k+1,k+2,\ldots\Big\}$$

The sample space for this experiment is infinite **yet** countable, so it is considered **discrete**.

Infinite and Uncountable

Example: A subject responds as quickly as possible. The outcome of the experiment is the latency of the response - time between the stimulus and response. If no time limit is imposed, then the sample space is,

$$\Omega = \Big(0,1\Big) = \Big\{t \in \mathbb{R} | 0 < t < \infty\Big\}$$

The sample space for this experiment is infinite **yet** countable, so it is considered **discrete**.

Any subset of $\mathbb{Z}$ is countable and any non-empty interval on $\mathbb{R}$ is not countable. Let Ω be a set and let A and B be subsets of Ω, then

$A \subseteq B$	A is a subset of B, possibly $A = B$
$A \subset B$ (or $A \not\subset B$)	A is a proper subset of B
$A \cup B$	A is a union of B, or $A + B$
$A \cap B$	A intersection B, or AB
$\overline{A}$	Complement of A in Ω, or $\Omega \setminus A$
$B - A$	Set of elements in B not in A

Consider some of the following properties of sets,

— $\emptyset$ is the empty set. By convention, $\emptyset$ is a subset of every set in Ω.

— A and B are disjoint if $A \cap B = \emptyset$.

— The set of all subsets of Ω is the power set, whose cardinality is equal to 2^{Ω}.

— The **Cartesian Product** of sets X, Y is $X \times Y = \left\{(x,y) | x \in X, y \in Y\right\}$. For example,

$$\mathbb{R} \times \mathbb{R} = \left\{(x,y) | x \in \mathbb{R}, y \in \mathbb{R}\right\} = \mathbb{R}^2.$$

$\mathbb{R}^3 \ldots \mathbb{R}^n$ are defined similarly.

We can represent **events** as subsets of sample spaces.

Example: Event A: Toss at least one heads of two tosses:

$$\Omega = \left\{(H,H),(H,T),(T,H),(T,T)\right\}$$
$$A = \left\{(H,H),(H,T),(T,H)\right\}$$

So, $A \subset \Omega$.

Date: 8/24

[Definition:] An **Indicator Function** is a function defined on a set Ω that indicates membership of an element in a subset of X of Ω. Let Ω be a non-empty set and X a subset of Ω. The indicator function on X is a function $I_x : \Omega \to \{0,1\}$ by,

$$I_x(w) = \begin{cases} 0 & w \in X \\ 1 & w \in \overline{X} (\text{or } w \notin X) \end{cases}$$

Properties of Indicator Functions:

— For subsets X, Y of Ω, $I_{X\cap Y} = I_X \cdot I_Y$, i.e. for every $w \in \Omega$, $I_{X\cap Y}(w) = I_X(w) \cdot I_Y(w)$.

Proof: Let $w \in X \cap Y$. Then, $w \in X$ and $w \in Y$. Therefore, by definition of an indicator function, $I_{X\cap Y}(w) = 1$, $I_X(w) = 1$, $I_Y(w) = 1$.
Hence, $I_{X\cap Y}(w) = I_X(w) \cdot I_Y(w)$ for $w \in X \cap Y$.

— $I_{X\cup Y} = I_X + I_Y - I_{X\cap Y}$ (if X, Y are disjoint, $I_{X\cup Y} = I_X + I_Y$)

— $I_{A^C} = 1 - I_A$

— If $X \subseteq Y$, then $I_X \subseteq I_Y$, i.e. $I_X(w) \leq I_Y(w)$ for every $w \in \Omega$.

Example Use of Indicator Function

Suppose random variable X has a probability distribution function given by,

$$f(x) = \begin{cases} 0, & x < -1 \\ 1 + x, & -1 \leq x < 0 \\ 1 - x, & 0 \leq x < 1 \\ 0, & x \geq 1 \end{cases}$$

We can then rewrite $f(x)$ as

$$f(x) = (1 + x)I_{[-1,0)}(x) + (1 - x)I_{[0,1)}(x),$$

or, more concisely,

$$f(x) = (1 + |x|)I_{[-1,1]}(x).$$

CHAPTER 2: SAMPLE SPACE AND EVENTS

Todo: Add chapter 2 material

CHAPTER 3: PROBABILITY AND AREA

Let Ω be a set, and $A_1, \ldots, A_n$ subsets of Ω. Then,

(i)

$$\underbrace{\left(A_1 \cup A_2 \cup \ldots \cup A_n\right)^C}_{\overline{\bigcup_{i=1}^{n} A_i}} = \underbrace{A_1^C \cap A_2^C \cap \ldots \cap A_n^C}_{\bigcap_{i=1}^{n} \overline{A_i}}$$

Proof of (i)**:** Let $w \in \left(A_1 \cup A_2 \cup \ldots \cup A_n\right)^C$. Thus, $w \in \overline{A_1}, w \in \overline{A_2}, \ldots, w \in \overline{A_n}$. Additionally, $w \in \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}$. So, $\left(A_1 \cup A_2 \cup \ldots \cup A_n\right)^C \subseteq \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}$. Let $w \in \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}$. Hence, $w \notin A_1, w \notin A_2, \ldots, w \notin A_n$. Thus, $w \notin A_1 \cup A_2 \cup \ldots \cup A_n$, i.e. $w \in \left(A_1 \cup A_2 \cup \ldots \cup A_n\right)^C$. Hence, $\overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n} \subseteq \left(A_1 \cup A_2 \cup \ldots \cup A_n\right)^C$.

We may then conclude that $\left(A_1 \cup A_2 \cup \ldots \cup A_n\right)^C = \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}$.

(ii) If $A_1, \ldots, A_n$ are mutually disjoint, ($A_i \cap A_j = \emptyset$, for $i \neq j$), then $(A_1 + A_2 + \ldots + A_n)B = A_1B + A_2B + \ldots + A_nB$ for any subset B of Ω.

Proof of (ii) is similar to proof of (i).

Note: (i) and (ii) extend without change in proof to arbitrary unions and intersections. In particular, the countable cases are

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^C = \bigcap_{i=1}^{\infty} \overline{A_i} \qquad \text{and} \qquad (\sum_{i=1}^{\infty} A_i)B = \sum_{i=1}^{\infty} A_iB$$

for subsets $B, A_1, A_2, \ldots, A_n$ of Ω.

[Definition] Let Ω be a set. Let $\mathcal{F}$ be a nonempty collection of subsets of Ω satisfying the axioms

[F1] For all $A, B \in \mathcal{F}$, $A \cup B \in \mathcal{F}$.

[F2] For each $A \in \mathcal{F}$, $\overline{A} \in \mathcal{F}$ also.

$\mathcal{F}$ is called a field of subsets of Ω.

Theorem 3.2: Let $\mathcal{F}$ be a field of subsets of Ω. Then,.

(i) $\Omega \in \mathcal{F}$

(ii) $\emptyset \in \mathcal{F}$

(iii) For all $A, B \in \mathcal{F}$, $A \setminus B = A\overline{B} \in \mathcal{F}$

$$\left.\begin{array}{ll} \text{(iv)} & \bigcup_{i=1}^{n} A_i \in \mathcal{F} \\ \text{(v)} & \bigcap_{i=1}^{n} A_i \in \mathcal{F} \end{array}\right\} \mathcal{F} \text{ is closed under finite union } \textbf{and} \text{ finite intersection}$$

Proof (i): Let $A \in \mathcal{F}$. By $[F2]$, $\overline{A} \in \mathcal{F}$. Hence, by $[F1]$, $\Omega = A + \overleftarrow{A} \in \mathcal{F}$.

(ii) We know that $A \cup \overline{A} = \Omega \in \mathcal{F}$. By $[F2]$, $\emptyset = \Omega^C \in \mathcal{F}$.

(iii) Let $A, B \in \mathcal{F}$. Then, by $[F2]$, $\overline{A}, \overline{B} \in \mathcal{F}$. By $[F1]$, $\overline{A} \cup B \in \mathcal{F}$. By $[F2]$, $A \cap \overline{B} = (\overline{A} \cup B)^C \in \mathcal{F}$. Thus, $A\overline{B} \in \mathcal{F}$.

Date: 8/29

CHAPTER 4: PROBABILITY MEASURES

Todo: Add chapter 4 material

CHAPTER 5: BASIC RULES OF PROBABILITY CALCULUS

Let Ω be a sample space, and $(\Omega, \mathcal{F}, \mathbb{P})$ a FAPs. Then, for any events $A, B, C \in \mathcal{F}$,

(i) $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$

Proof: $1 = \mathbb{P}(\Omega) = \mathbb{P}(A + \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A})$ by (iii)

(ii) $\mathbb{P}(\emptyset) = 0$

Proof: Set $A = \emptyset$ in (i). Then, $\overline{A} = \Omega$ and we get

$$1 = \mathbb{P}(\emptyset) + \mathbb{P}(\Omega) = \mathbb{P}(\emptyset) + 1 = 1.$$

So, $\mathbb{P}(\emptyset) = 0$.

(iii) If $A_1, \ldots, A_n$ are incompatible events in $\mathcal{F}$, then

$$\mathbb{P}\Big(\sum_{j=1}^{n} A_j\Big) = \sum_{j=1}^{n} \mathbb{P}(A_j)$$

Proof: (iii) has already been proven.

(iv) If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

Proof: Write B as $A + (B \setminus A) = A + B\overline{A}$. Hence, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A\overline{B}) \geq \mathbb{P}(A)$. So, the probability measure $\mathbb{P}$ is monotonic.

(v) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Proof: todo

(vi)

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) = \ & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ & - \mathbb{P}(AB) - \mathbb{P}(AC) - \mathbb{P}(BC) \\ & + \mathbb{P}(ABC)\end{aligned}$$

Proof: todo

(vii) $\mathbb{P}(A\overline{B}) = \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(AB)$

Proof: todo

We'll start to generalize (v) and (vi). Let $A_1, \ldots, A_4$ be events.

$$\begin{aligned}\mathbb{P}(A_1 \cup \cdots \cup A_4) = \ & \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_4) \\ & - [\mathbb{P}(A_1A_2) + \mathbb{P}(A_1A_3) + \mathbb{P}(A_1A_4) + \mathbb{P}(A_2A_3) + \mathbb{P}(A_2A_4) + \mathbb{P}(A_3A_4)] \\ & + [\mathbb{P}(A_1A_2A_3) + \mathbb{P}(A_1A_3A_4) + \mathbb{P}(A_1A_2A_4) + \mathbb{P}(A_2A_3A_4)] \\ & - \mathbb{P}(A_1A_2A_3A_4)\end{aligned}$$

(v), (vi), and the above example are instances of **Poincare's Identity**. To state it, well need the following notation,

> For $0 \le k \le n$, k, n integers, let $s(k,n)$ be sets of subsets of $\{1, 2, \ldots, n\}$ of cardinality k, e.g. $s(2,3) = \Big\{\{1,2\}, \{1,3\}, \{2,3\}\Big\}$.

Date: 9/12

Stirling's Approximation

$$n! \sim (2\Pi)^{\frac{1}{2}} n^{(n+\frac{1}{2})} e^{-n}, \text{ as } n \to \infty$$

Or, more specifically,

$$\frac{n!}{(2\Pi)^{\frac{1}{2}} n^{(n+\frac{1}{2})} e^{-n}} \sim 1, \text{ as } n \to \infty$$

Poincare's Identity: $s(k,n)$ is the set of subsets of size k of $\{1, \ldots, n\}$. For events $A_1, \ldots, A_n$,

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{l=1}^{n} (-1)^{l+1} \sum_{J \in s(l,n)} \mathbb{P}(\bigcap_{k \in J} A_k)$$

Example: Consider a tournament involving four teams. Let $a_1, \ldots, a_4$ be the finishing places of teams $1, \ldots, 4$ respectively, e.g. if teams 3 wins and team 1 comes in last, then $a_3 = 1$ and $a_1 = 4$. Given $\mathbb{P}(a_1 = 1) = \frac{1}{2}$, $\mathbb{P}(a_2 = 4) = \frac{1}{2}$, and $\mathbb{P}(a_3 < a_4) = \frac{1}{4}$,

$$\mathbb{P}(a_1 = 1 \text{ or } a_2 = 4) = \frac{3}{4}$$

Example: Find the maximum possible probability of the event D: either $a_1 = 1$ and $a_2 = 4$ or $a_3 < a_4$ or both:

$$\mathbb{P}(a_1 = 1 \text{ or } a_2 = 4) = \frac{3}{4}$$

We want the maximum possible probability of D given the four probabilities,

$$\begin{aligned}\mathbb{P}(D) = \mathbb{P}(AB \cup C) &= \mathbb{P}(AB) + \mathbb{P}(C) - \mathbb{P}(ABC) \\ &= \mathbb{P}(AB) + \frac{1}{4} - \mathbb{P}(ABC)\end{aligned}$$

We are given that $\mathbb{P}(A \cup B) = \frac{3}{4}$. We also then have $\mathbb{P}(A \cup B) = \mathbb{P}(AB) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$. So,

$$\begin{aligned}\mathbb{P}(AB) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{3}{4} \\ &= \frac{1}{4}\end{aligned}$$

Hence, $\mathbb{P}(D) = \frac{1}{4} + \frac{1}{4} - \mathbb{P}(ABC) \leq \frac{1}{2}$. So, the maximum probability of D is $\frac{1}{2}$.

Sampling with Replacement

No $A_1, \ldots, A_n$ be sets. Their cartesian product is

$$A_1 \times A_2 \times \cdots \times A_n = \bigtimes_{i=1}^{n} A_i = \left\{(a_1, \ldots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\right\}$$

We usually write $s \times s \times \cdots \times s$ (m copes of s) as $s^m = \left\{(s_1, s_2, \ldots, s_m) \mid s_j \in S, j = 1, \ldots, m\right\}$. These are ordered m-tuples. For finite sets $A_1, \ldots, A_n$ with cardinalities $| A_i |= K_i$ with $i = 1, \ldots, n$,

$$|A_1 \times \cdots \times A_n| = |A_i| \cdots |A_n| = K_1 \cdot K_2 \cdots K_n$$

A sample of size m from the set S is an element $(s_1, \ldots, s_m)$ of S^m.

Example: Flip a coin (2 outcomes, H or T), throw one 6-sided die (6 outcomes), and pick a card from a standard deck (52 outcomes).

Question: How many distinct outcomes are there?
There are $2 \times 6 \times 52$ outcomes, i.e. points of the form $(H, 3, K - \text{Hearts})$. Here, we are sampling with replacement. This means that an entry may appear in more than one outcome.

Note that we are sampling with order here.

Example: Throw a die 3 times. There are $s^3 = 6 \times 6 \times 6 = 216$ possible outcomes - $(n_1, n_2, n_3) \in 1, 2, 3, 4, 5, 6^3$. We are sampling here with replacement and with order, i.e. $(1, 2, 3)$ and $(1, 3, 2)$ are distinct outcomes.

Observe that the number of ordered samples of size m, taken with replacement from a set S with cardinality $|S| = n$ is $|s \times s \times \cdots \times s| = n^m$.

Sampling without Replacement

There are $(m)_k = (m)(m-1)\cdots(m-k+1)$ ordered samples of size k from a set of size m, taken **without** replacement. More formal notation for representing sampling without replacement can be denoted as

$$(m)_k = \frac{(m)_k(m-k)!}{m-k} = \frac{m!}{(m-k)!}$$

Alternatively, if we sample with replacement, there would be m^k samples.

Date: 9/14

We can count ordered or unordered samples with or without replacement. If $|S| = n < \infty$, then the number of ordered samples of size m, taken **with** replacement is given by

$$n^m$$

The number of ordered samples taken **without** replacement is given by

$$(n)_m = (n)(n-1)\cdot(n-k+1) = \frac{n!}{(n-k)!}$$

For sufficiently large n, Stirling's Approximation will yield a close approximation of n factorial,

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}, \textbf{ as } n \to \infty$$

In other words,

$$\frac{n!}{\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}} \sim 1 \textbf{ as } n \to \infty$$

Among n people, let A be the event that at least two of the people share a birthday. The complementary event is $\overline{A}$: no two of the n people share a birthday. Then,

$$\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$$

and let $2 \leq n \leq 365$. Assume that

1. There are 365 distinct birthdays - excluding February 29$^{\text{th}}$.

2. All birthdays are equally as probable.

Assumption (2) allows us to use the counting measure:

$$\mathbb{P}(\overline{A}) = \frac{|\overline{A}|}{|\Omega|}.$$

$$|\Omega| = \text{ Number of ordered samples of size } n \text{ from a set size 365 taken with replacement } = 365^n$$

$$|\overline{A}| = \text{ Number of samples of size } n \text{ from a set of 365 birthdays taken without replacement } = \frac{365!}{(365-n)!}$$

Thus, $\mathbb{P}(\overline{A})$ is equivalent to

$$\frac{365!}{(365-n)!365^n}.$$

CHAPTER 7: Counting Subsets

Note: Sample sets require order while subsets only require identity
Let $|S| = n < \infty$. How many subsets of size m $(0 \leq m \leq n)$ has S? We'll label this number $\binom{n}{m}$ where

$$\binom{n}{m} = \frac{(n)_m}{m!} = \frac{n!}{m!(n-m)!}.$$

Since $\binom{n}{m}$ is the number of ways of selecting m objects from n objects, it is also called "n choose m", denoted by $n\mathrm{C}m$. More formally, it's also known as the **binomial coefficient**. Note the following properties of n choose m,

— By convention, if $m < 0$ or $n - m < 0$, we define $\binom{n}{m} = 0$.

— There is only one susbset of size 0 of a set S: $\emptyset$,

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{0!n!}$$

— S has only one subset of size n that is itself,

$$\binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!0!} = 1.$$

— The following equivalence holds,

$$\binom{n}{m} = \binom{n}{n-m}.$$

— For any numbers a, b,

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

The binomial coefficient is needed because it denotes how many $a^m b^{n-m}$ show up.

Theorem 7.4

For integers m, n ($0 \le m \le n$),

(i)

$$\binom{n}{m} = \binom{n}{n-m}$$

PROOF

$$\begin{aligned}\binom{n}{m} &= \binom{n}{n-m} \\ &= \frac{n!}{(n-m)!\big(n-(n-m)\big)!} \\ &= \frac{n!}{(n-m)!(m)!}\end{aligned}$$

(ii)

$$\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-1}{m} \text{ for } n \ge 1$$

PROOF: Suppose that $S = \{x_1, x_2, \ldots, x_n\}$. Partition the set of subsets of size m into two disjoint classes. Subsets of size m containers x_i. There are $\binom{n-1}{m-1}$ of these. Subsets of size m not containing x_i. There are $\binom{n-1}{m}$ of these.

(iii)

$$2^n = \sum_{m=0}^{n} \binom{n}{m} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{m}$$

PROOF: By the binomial theorem,

$$2^n = (1+1)^n = \sum_{m=0}^{n} \binom{n}{m}.$$

Or, you might show through induction that $|2^S| = 2^n$. And then, since $\binom{n}{m}$ is the number of subsets of size m,

$$|2^S| = \sum_{m=0}^{n} \binom{n}{m}.$$

So, $\binom{n}{m}$ is the number of ways of dividing S into disjoint subsets of sizes m and $n-m$ respectively. **Generally**, how many ways can we divide S into k disjoint subsets of sizes

$$m_1, m_2, \ldots, m_k$$

where $m_1 + m_2 + \ldots + m_k = n$? We can denote this number with the **multinomial coefficient**,

$$\binom{n}{m_1 m_2 \cdots m_k}.$$

Example: In how many ways can we partition $S = \{a, b, c, \ldots, z\}$ into $k = 4$ disjoint subsets of cardinalities of $12, 4, 3, 7$?
We can solve this problem with the multinomial coefficient,

$$\binom{n}{m_1 m_2 m_3 m_4} = \frac{n!}{m_k! m_{k-1}! \cdots m_1!}.$$

For each binomial coefficient, there was always an m_{k-i}, $i = 0, 1, 2, \ldots, k-1$.

Date: 9/19

Clarification of Multinomial Coefficient

$$\binom{n}{m_1 \cdots m_k} = \left(m_1, \ldots, m_k\right)! = \frac{n!}{m_k! m_{k-1}! \cdots m_1!}$$

Let $B_1, \ldots, B_k$ be k distinct bins. The number of ways of distributing n items among the bins so that B_1 holds m_1 items, B_2 holds m_2 items, ..., B_k holds m_k items.

Example: Players $P_1, \ldots, P_4$ are dealt bridge hands, that is, each player receives 13 cards from a well-shuffled deck. For $i = 1, \ldots, 4$, let A_i be the event that player P_i receives all 13 clubs. Clearly, $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \mathbb{P}(A_4)$. We'll compute $\mathbb{P}(A_1)$. The sample space is the set Ω of distributions of 52 cards among 4 players each receiving 13 cards. The sample space has cardinality,

$$|\Omega| = \binom{52}{13\ 13\ 13\ 13} = \frac{52!}{(13!)^4}.$$

The cardinality of A_1 is the number of ways in which 39 cards (clubs have all gone to P_1) can be dealt to P_2, P_3, P_4. Thus, $|A_1| = \binom{39}{131313} = \frac{39!}{(13!)^3}$. Using the counting measure on Ω, we get $\mathbb{P}(A_1) = \frac{|A_1|}{|\Omega|} \sim 1.56 \times 10^{-12}$.

If A is the event that one player (any one of $P_1, \ldots, P_4$) gets 13 clubs, then $A = \sum_{i=1}^{4} A_i$. Hence, $\mathbb{P}(A) = 4\mathbb{P}(A_1) \sim 6.24 \times 10^{-12}$.

A Congenerics of Discrete Distributions

Consider an experiment with two outcomes, e.g. tossing a coin. Here, $\Omega = \{T, H\}$. Or, guessing a correct card where $\Omega = \{S, F\}$ with s =success and f =failure. Let's idealize this sample space for such as $\Omega = \{0, 1\}$. Let $0 < \alpha < 1$. Let $p : \Omega \to [0, 1]$, by $p(0) = \alpha$ and $p(1) = 1 - \alpha$. p is called the **Bernoulli distribution** on Ω, with parameter α (or $1 - \alpha$).

Suppose next that we repeat the experiment from the previous example n times. We are interested in the number of k zeros in the n trials. We then take $\Omega = \{0, 1, \ldots, k, \ldots, n\}$.

We'll define a distribution p on Ω as follows,

$$p(0, \ldots, 0, 1, \ldots, 1) = \alpha^k (1 - \alpha)^{n-k}$$

assuming **independent** repititions. Therefore, $p(k$ zeros and $n - k$ ones) is equivalent to

$$\binom{n}{k} \alpha^k (1 - \alpha)^{n-k}$$

Thus, we define $p : \Omega \to [0, 1]$ by

$$p(k) = \binom{n}{k} \alpha^k (1 - \alpha)^{n-k}, \ k = 0, 1, 2, \ldots, n.$$

We should check that p is even a distribution. We see that $p(k) > 0$ for ever $k \in \Omega$. And,

$$\begin{aligned} \sum_{k \in \Omega} p(k) &= \sum_{k=0}^{n} \binom{n}{k} \alpha^k (1 - \alpha)^{n-k} \\ &= \Big[\alpha + 1 - \alpha\Big]^n \end{aligned}$$

p is the binomial distribution with parameters n, k meaning that there are **independent** and **repeated** trials with two possible outcomes.

Generalization

Instead of repeated, independent trials with 2 outcomes, consider repeated, independent trials with k outcomes.

> Suppose that a large urn holds balls of colors $c_1, c_2, \dots, c_k$. For $j = 1, \dots, k$, let α_j be the function of balls in the urn of color c_j. Then, $\alpha_j > 0$ for each j and $\sum_{j=1}^{k} \alpha_j = 1$. Sample n items from the urn with replacement. Record the color of the randomly chosen ball, then return the ball to the urn. [**Q**] What is the probability we get m_1 balls of color c_1, m_2 balls of color c_2, ..., m_k balls of color c_k (where $m_1 + m_2 + \cdots + m_k = n$)?
> The probability of drawing a ball of color j is α_j. There are $\binom{n}{m_1 m_2 \cdots m_k}$ ways of choosing m_j balls of color c_j, $j = 1, \dots, k$. Thus,
>
> $$\mathbb{P}(m_j \text{ balls of color } c_j,\ j = 1, \dots, k) = \binom{n}{m_1 \cdots m_k} \alpha_1^{m_1} \alpha_2^{m_2} \cdots \alpha_k^{m_k}.$$

Finally, define a sample space - let Ω be the set of k-tuples $m_1, \dots, m_k$ where $0 \le m_j \le n$ for $j = 1, 2, \dots, k$ and $\sum_{j=1}^{k} m_j = n$.

Geometric Distribution

Let's go back to the 2-outcome trial with outcomes 0 and 1, where $p(0) = \alpha$ and $p(1) = 1 - \alpha$.

> Repeat trials until 1 occurs. The outcome is the number of trials in which the 1 first appears. Let the sample space $\Omega = \mathbb{N}$, so 7 means that there were 6 zeros and 1 one. Assuming **independent** trials,
>
> $$p(0, 0, \dots, 0, 1) = \alpha^{k-1}(1 - \alpha).$$

We thus define $p : \Omega \to [0, 1]$ by $p(k) = \alpha^{k-1}(1 - \alpha)$ for $k = 1, 2, 3, \dots$ We should check that p is even a distribution. We see that $p(k) > 0$ for ever $k \in \Omega$. And,

$$\begin{aligned}
\sum_{k \in \Omega} p(k) &= \sum_{k=1}^{\infty} \alpha^{k-1}(1 - \alpha) \\
&= (1 - \alpha) \sum_{k=0}^{\infty} \alpha^{k}(1 - \alpha) \cdot \frac{1}{1 - \alpha} \\
&= 1.
\end{aligned}$$

p is therefore the **geometric distribution** with parameter α.

Date: 9/21

Discrete Distribution Functions

The **Poisson Distribution** shows up when we are counting "punctual" events - that is, events that occur in a very short period of time. Note that short is relative to the problem observation time. We'll divide the time period $(0,t]$ into n sub-intervals, each with length $\triangle t = \frac{t}{n}$, and endpoints $t_0 = 0, t_1 = \triangle t, \ldots, t_n = n \cdot \triangle t$. For large n, we will make the following assumptions,

— the probability of 2 or more occurrences in any sub-interval $(t_i, t_{i+1}]$ is negligible.

— the probability of exactly 1 occurrence in any sub-interval $(t_i, t_{i+1}]$ is approximately $\lambda \triangle t$ for a fixed $\lambda > 0$.

— [from the first 2 assumptions] the probability of no occurrence in $(t_i, t_{i+1}] \sim 1 - \lambda \triangle t - \mathbb{P}(\text{2 or more occurrences in any sub-interval})$

We also implicitly assume that the number of occurrences in different intervals are mutually independent. These assumptions allow us to approximate the probability that we detect exactly k occurrences of punctual events in the n sub-intervals by the binomial distribution with parameters n and $\lambda \triangle t$. So, for large n,

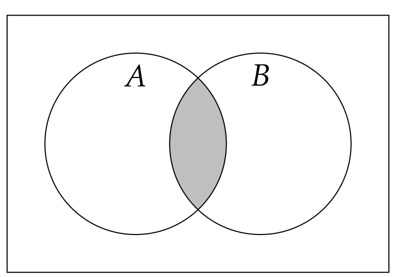
$$\mathbb{P}(\text{exactly } k \text{ occurrences in } (0,t]) \sim \binom{n}{k} (\lambda \triangle t)^k (1 - \lambda \triangle t)^{n-k}.$$

Conditional Probability

Choose a card at random from a 52 card deck,

$$\begin{aligned} A &: \text{ Card is a heart} \\ B &: \text{ Card is red} \end{aligned}$$

Clearly, $\mathbb{P}(A) = \frac{1}{4}$, but given B, what is the probability of $A = \frac{1}{2}$? We write $\mathbb{P}(A|B) = \frac{1}{2}$. $\mathbb{P}(A|B)$ is the conditional probability of A given B.



So, we derive a conditional probability that is normalized by $\mathbb{P}(B)$,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

Example: Throw a fair dice D_1 and D_2. The sample space is then

$$\begin{aligned}\Omega = \{ \ & (1,1) & \dots & \quad (1,6) \\ & \vdots & \ddots & \quad \vdots \\ & (6,1) & \dots & (6,6) \ \} \end{aligned}$$

Then, $|\Omega| = 36$. Let B be the event that D_2 shows an even number $(2,4,6)$. Then, let A be the event that the sum of the number is 7. We know that $\mathbb{P}(B) = \frac{1}{2}$. Then,

$$AB = \{(5,2),\ (3,4),\ (1,6)\}.$$

So, $\mathbb{P}(AB) = \frac{3}{36} = \frac{1}{12}$. Hence,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \frac{\frac{1}{12}}{\frac{1}{2}} = \frac{1}{6}$$

Date: 9/26

Let A, B be events, $\mathbb{P}(B) \neq 0$. Then, the conditional probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$

Example Consider a bridge game with players N, S, E, W.
Let A be the event that N receive exactly 10 spades. Let B be the event that S receive all of the hearts. Compute $\mathbb{P}(A \mid B)$.
Well, we know that

$$\begin{aligned}\mathbb{P}(B) &= \frac{\binom{39}{13\ 13\ 13}}{\binom{52}{13\ 13\ 13\ 13}} \\ &\sim 1.56 \times 10^{-12}\end{aligned}$$

What is $\mathbb{P}(AB)$? The counting measure $\mathbb{P}$ applies since distributions of 52% cards among the four players (13 each) are equaly probably. Therefore,

$$\mathbb{P}(B) = \frac{|AB|}{\binom{52}{13\ 13\ 13\ 13}}.$$

$$|AB| = \binom{13}{10}\binom{29}{3\ 13\ 13}.$$

Hence,

$$\begin{aligned}\mathbb{P}(A \mid B) &= \frac{\frac{\binom{13}{10}(13,13,13)!}{(13,13,13,13)!}}{\frac{(13,13,13)!}{(13,13,13,13)!}} \\ &= \frac{\binom{13}{10}\binom{39}{3\ 13\ 13}}{\binom{39}{13\ 13\ 13}} \\ &= \mathbb{P}(A \mid B)\end{aligned}$$

Theorem Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a FAPs. Then, for events A, B, C with $\mathbb{P}(C) \neq 0$,

(i) $\mathbb{P}(\frac{\Omega}{C}) = 1$
Proof

$$\mathbb{P}(\Omega|C) = \frac{\mathbb{P}(\Omega C)}{\mathbb{P}(C)} = 1 = \mathbb{P}(A|C) + \mathbb{P}(B|C). \tag{1}$$

(ii) $\mathbb{P}(A|C) \geq 0$
Proof Trivial

(iii) If $AB \neq 0$, then $\mathbb{P}(A + B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C)$.
Proof

$$\begin{aligned}\mathbb{P}((A+B)|C) &= \frac{\mathbb{P}((A+B)C)}{\mathbb{P}(C)}\\ &= \frac{\mathbb{P}(AC+BC)}{\mathbb{P}(C)}\\ &= \frac{\mathbb{P}(AC)}{\mathbb{P}(C)} + \frac{\mathbb{P}(BC)}{\mathbb{P}(C)}\\ &= \mathbb{P}(A|C) + \mathbb{P}(B|C)\end{aligned}$$

(iv) If $AC \neq 0$, then $\mathbb{P}(A|C) = 0$.

Proof

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(AC)}{\mathbb{P}(C)} = 0.$$

(v) If $\mathbb{P}(A) \neq 0$, then $\mathbb{P}(A|C) = \frac{\mathbb{P}(A)}{\mathbb{P}(C)}\mathbb{P}(C|A)$.

Proof

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(AC)}{\mathbb{P}(C)} \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = \mathbb{P}(C|A)\frac{\mathbb{P}(A)}{\mathbb{P}(C)}$$

For any $A \in \mathcal{F}$, define $\mathbb{P}_C(A) = \mathbb{P}(A|C)$.

By (i), $\mathbb{P}_C(\Omega) = 1$ ([K1 for a FAPs]). By (ii), $\mathbb{P}_C(A) \geq 0$ ([K2] for a FAPs). By (iii), $\mathbb{P}_C(A+B) = \mathbb{P}_C(A) + \mathbb{P}_C(B)$ for incompatible events A, B ([K3] for a FAPs).

Thus, $\mathbb{P}_C$ is itself a FAPM on $\mathcal{F}$, i.e. $(\Omega, \mathcal{F}, \mathbb{P}_C)$ is a FAPs. **Note:** If $(\Omega, \mathcal{F}, \mathbb{P})$ is a CAPs, where $\mathcal{F}$ is a σ-field, then we may replace (iii) by:

For incompatible events $\{A_i\}_{i=1}^{\inf}$,

$$\mathbb{P}(\sum_{i=1}^{\infty} A_i|C) = \sum \mathbb{P}(A_i|C).$$

In this case, $(\Omega, \mathcal{F}, \mathbb{P}_C)$ is a CAPs.
Note: In the FAPs case, by induction, ... for any finite ... :(

$$asdf$$

Theorem For events $A_1, \ldots, A_n$, $\mathbb{P}(A_i) \neq 0$, for $i = 1, \ldots, n$,

$$\mathbb{P}(A_1 \cdots A_n) = \mathbb{P}(A_1) \times \mathbb{P}(A_2|A_1) \times \cdots \times \mathbb{P}(A_n|A_1 \cdots A_{n-1}).$$

Proof

$$\begin{aligned}
\mathbb{P}(A_1\cdots A_n) &= \mathbb{P}(A_n|A_1\cdots A_{n-1})\times\mathbb{P}(A_1\cdots A_{n-1})\\
&= \mathbb{P}(A_n|A_{n-1}\cdots A_1)\times\mathbb{P}(A_{n-1}|A_{n-2}\cdots A_1)\times\mathbb{P}(A_{n-2}\cdots A_1)\\
&= \dots\\
&= \mathbb{P}(A_n|A_{n-1}\cdots A_1)\times\mathbb{P}(A_{n-1}|A_{n-2}\cdots A_1)\times\cdots\times\mathbb{P}(A_2|A_1)\mathbb{P}(A_1)
\end{aligned}$$

Date: 9/28

Theorem: Law of Total Probability
Let $(\Omega,\mathcal{F},\mathbb{P})$ be a FAPs, with events $\left\{H_1,\dots,H_n\right\}$ satisfying

(i) $\mathbb{P}(H_i)\neq 0$ for $i=1,\dots,n$

(ii) $H_i\cap H_j$ for $i\neq j$

(iii) $\sum_{i=1}^n H_i=\Omega$

Then, for any $A\in\mathcal{F}$,

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{j=1}^n \mathbb{P}(AH_j)\\
&= \sum_{j=1}^n \mathbb{P}(A|H_j)\mathbb{P}(H_j)
\end{aligned}$$

Proof Because $\sum_{j=1}^n H_j=\Omega$,

$$\begin{aligned}
\mathbb{P}(A)=\mathbb{P}(A\Omega) &= \mathbb{P}(A\sum_{j=1}^n H_j)\\
&= \mathbb{P}(\sum_{j=1}^n AH_j)\\
&= \sum_{j=1}^n \mathbb{P}(AH_j)\\
&= \sum_{j=1}^n \frac{\mathbb{P}(AH_j)}{\mathbb{P}(H_j)}\mathbb{P}(H_j)\\
&= \sum_{j=1}^n \mathbb{P}(A|H_j)\mathbb{P}(H_j)
\end{aligned}$$

Note: If you replace the FAPs with CAPs and replace the set of finite event $\{H_1, \ldots, H_n\}$ with a set of infinite events $\{H_1, \ldots, H_k, \ldots\}$, the **Law of Total Probability** still holds.
Next we will see that it's easy to define a scenario where the probability of multiplied events is less than the multiplication of probabilities of events. Additionally, we will see that it's easy to define a scenario where the multiplication of probabilities of events is less than the probability of multiplied events. Let $0 < \mathbb{P}(A) < 1$. Let $B = \overline{A}$. Then, $\mathbb{P}(A)\mathbb{P}(B) > 0$ (because $\mathbb{P}(B) > 0$ as well). But, $\mathbb{P}(AB) = \mathbb{P}(A\overline{A}) = \mathbb{P}(\emptyset) = 0 < \mathbb{P}(A)\mathbb{P}(B)$. Now, let $A \subset B$ with $0 < \mathbb{P}(A) < \mathbb{P}(B) < 1$. Then, $\mathbb{P}(AB) = \mathbb{P}(A)$. And $\mathbb{P}(A)\mathbb{P}(B) < \mathbb{P}(A) = \mathbb{P}(AB)$.

An event A "should be" independent of event B if $\mathbb{P}(A|B) = \mathbb{P}(A)$. In other words, if $\frac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \mathbb{P}(A)$, or $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition: Events $A, B \in \mathcal{F}$ are independent if

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B).$$

Consequences of Conditional Probability

- The Law of Total Probability
- Independent Events
- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a FAPs. Then, events A and B are independent if $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$.

Example Throw a fair die twice. Take as the sample space,

$$\Omega = \Big\{(1,1),(1,2),\ldots,(6,6)\Big\}.$$

Let $\mathcal{F} = 2^{\Omega}$. Let A be the event that there's an odd number on the first throw and B be the event that there is a 4 on the second throw. The probability measure $\mathbb{P}$ on $\mathcal{F}$ is the counting measure:

$$\mathbb{P}(C) = \frac{|C|}{36} \text{ for } C \in \mathcal{F}.$$

Because

$$A = \begin{Bmatrix} (1,1),(1,2),\ldots,(1,6) \\ (3,1),(3,2),\ldots,(3,6) \\ (5,1),(5,2),\ldots,(5,6) \end{Bmatrix}$$

and

$$B = \Big\{(1,4),(2,4),\ldots,(6,4)\Big\}.$$

$\mathbb{P}(A) = \frac{1}{2}$ and $\mathbb{P}(B) = \frac{1}{6}$ and because, $\mathbb{P}(AB) = \frac{1}{12} = \mathbb{P}(A)\mathbb{P}(B)$. So, A and B are independent.

A few facts about independence: Let $\mathbb{P}(A) = 0$. Then, for any other event B, $0 = \mathbb{P}(AB)$. But, $\mathbb{P}(A)\mathbb{P}(B) = 0$. Hence, A and B are independent. So, an event A with $\mathbb{P}(A) = 0$ is independent of all other events. Let A and B be independent. Then, $\overline{A}$ and B are also independent.

Proof

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B(A+\overline{A})) \\ &= \mathbb{P}(BA + B\overline{A}) \\ &= \mathbb{P}(BA) + \mathbb{P}(B\overline{A}))\end{aligned}$$

So, becasue A and B are independent, $\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(B\overline{A})$. Hence,

$$\mathbb{P}(B\overline{A}) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(B)\Big[1 - \mathbb{P}(A)\Big] = \mathbb{P}(B)\mathbb{P}(\overline{A}).$$

Corollary: If $\mathbb{P}(A) = 1$, then A is independent of all other events. Since $\mathbb{P}(A) = 1$, $\mathbb{P}(\overline{A}) = 0$.

Therefore, $\overline{A}$ is independent of B for any event B. Thus, $A = \overline{\overline{A}}$ is also independent of B.

If A and B are independent and $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. The proof is fairly trivial,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$

Definition: Let $\left\{A_\lambda\right\}_{\lambda\in\Lambda}$ be a collection of events. $\left\{A_\lambda\right\}_{\lambda\in\Lambda}$ is independent if

$$\mathbb{P}(A_{\lambda_1} \cdot A_{\lambda_2} \cdots A_{\lambda_n}) = \mathbb{P}(A_{\lambda_1}) \cdots \mathbb{P}(A_{\lambda_n})$$

for any finite subset $\lambda_1, \ldots, \lambda_n \in \Lambda$. In particular, pairwise independence,

$$\mathbb{P}(A_{\lambda_1} A_{\lambda_2}) = \mathbb{P}(A_{\lambda_1})\mathbb{P}(A_{\lambda_2})$$

for all pairs $\lambda_1, \lambda_2 \in \Lambda$ does not imply independence of $\left\{A_\lambda\right\}_{\lambda\in\Lambda}$.

Another consequence of conditional probability is **Bayes' Theorem**. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a FAPs. Let $\{H_1, H_2, \ldots, H_n\}$ be events in $\mathcal{F}$ satisfying the following,

(i) $H_i H_j = \emptyset$ for $1 \leq i \neq j \leq n$

(ii) $\sum_{i=1}^n H_i = \Omega$

(iii) $\mathbb{P}(H_i) \neq 0$, $i = 1, \ldots, n$.

Let $D \in \mathcal{F}$. Then, for each i $(1 \leq i \leq n)$,

$$\mathbb{P}(H_i \mid D) = \frac{\mathbb{P}(D \mid H_i)\mathbb{P}(H_i)}{\sum_{l=1}^n \mathbb{P}(D \mid H_l)\mathbb{P}(H_l)}.$$

Note: If $(\Omega, \mathcal{F}, \mathbb{P})$ is a CAPs, we may replace n with ∞.

Proof

$$\begin{aligned}
\mathbb{P}(H_i \mid D) &= \frac{\mathbb{P}(H_i D)}{\mathbb{P}(D)} \cdot \frac{\mathbb{P}(H_i)}{\mathbb{P}(H_i)} \\
&= \frac{\mathbb{P}(H_i D)}{\mathbb{P}(H_i)} \cdot \frac{\mathbb{P}(H_i)}{\mathbb{P}(D)} \\
&= \frac{\mathbb{P}(H_i D)\mathbb{P}(H_i)}{\mathbb{P}(D)} \\
&= \frac{\mathbb{P}(H_i D)\mathbb{P}(H_i)}{\sum_{l=1}^n \mathbb{P}(D \mid H_l)\mathbb{P}(H_l)}
\end{aligned}$$

by the **law of total probability**. Think of D as "data" or an "outcome," something observed and measured. Think of $H_1, \ldots, H_n$ as mutually exclusive hypotheses that would account for D. Bayes' Theorem uses conditional probability in the "usual direction" to calculate $\mathbb{P}(H_i \mid D)$. Or, more intuitively, given some event D that was observed, what is the probability that the hypothesis H_i holds. We compute the probability of a "cause" H_i given an "effect" D.

The numbers $\mathbb{P}(H_j)$ are the "a priori" probabilities and the numbers $\mathbb{P}(H_j \mid D)$ are the "a posteriori" probabilities.

Random Variables

Consider the experiment: Toss a fair die twice. Let $\Omega = \left\{(i,j) \mid 1 \le i,j \le 6\right\}$ be the sample space. Let $\mathcal{F} = 2^{\Omega}$ be our field of events. Define the two "random variables ":

X = number on the face of the die on the first throw
Y = sum of numbers from the two throws

Clearly, X has range $\left\{1,\ldots,6\right\}$ and Y has range $\left\{2,\ldots,12\right\}$. Think of X as a function on Ω,

$$X : \Omega \to \mathbb{R}$$

by $X(i,j) = i$. Likewise,

$$Y : \Omega \to \mathbb{R}$$

by $Y(i,j) = i + j$. Loosely speaking, a random variable on sample space Ω is a function $X : \Omega \to \mathbb{R}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a CAPs. So, Ω is a sample space, $\mathcal{F}$ is a σ-field of events, and $\mathbb{P}$ be a CAPM: For events $\{A_n\}_{n=1}^{\infty}$,

$$\mathbb{P}(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Loosely speaking, a random variable is a function $X : \Omega \to \mathbb{R}$.

Consider the experiment: Flip a coin n times. Suppose that we are interested only in the number of heads. We take as our sample space $\Omega = \{-, 1, \ldots, n\}$. Let X be the number of heads out of n flips. Then $X : \Omega \to \mathbb{R}$ by $X(k) = k$. So here, X is the identity function, or identity variable. X has range $\{0, 1, 2, \ldots, n\}$.

Consider the experiment: We record the number of occurrences of some "punctual event" in a unit time period. The sample space is $\Omega = \{0, 1, 2, \ldots, n\}$. Let N be the number of occurrences. Then, $N : \Omega \to \mathbb{R}$ by $N(k) = k$. Again, N is the identity function on Ω. X has range $\{0, 1, 2, \ldots, n\}$.

Consider the response time experiment: We record the response time to some stimulus. The sample space is the set of positive times. $\Omega = (0, \infty) = \{t \in \mathbb{R} \mid t > 0\}$. Let T be the response time. Then $T : \Omega \to \mathbb{R}$ by $T(t) = t$. So, the random variable T is also the identity function on Ω. T has range $(0, \infty)$.

Note: Random variables are often identity functions on sample spaces. So, why discern between some random variable as a function such that $X : \Omega \to \mathbb{R}$ by $X(k) = k$. Can we just use the definition "*A random variable is a variable whose values represent an experiment outcomes*"? For the moment, let's ignore the formal definition of random variable as a function on Ω, and just think of random variables simply as variables whose values represent an experiment's outcomes.

Example: Let X be a random variable with range $\{x_1, x_2\}$. Let $\mathbb{P}(X = x_1) = \alpha$ and $\mathbb{P}(X = x_2) = 1 - \alpha$, where $0 < \alpha < 1$. X is a Bernoulli-distributed with parameter α. In an empirical situation, consider a weighted coin, $\mathbb{P}(H) = .6$ and $\mathbb{P}(T) = .4$. Let

$$X = \begin{cases} 1 & \text{if } H, \\ 0 & \text{if } T \end{cases}$$

Then, $\mathbb{P}(X = 1) = .6$ and $\mathbb{P}(X = 0) = .4$. X is the Bernoulli distributed parameter with parameter 0.6 (or 0.4).

Example: Consider a sequence of n identical, independent trials, each with two possible outcomes (Success/Failure, Head/Tail, or 0/1), or Bernoulli trials. Let's use S (success) and F (failure). Suppose that $\mathbb{P}(S) = \alpha$ and $\mathbb{P}(F) = 1 - \alpha$, where $0 < \alpha < 1$. Let X be the number of S's in the n trials. Consider $\mathbb{P}(S \cdots S \cdot F \cdots F) = \mathbb{P}(S) \cdots \mathbb{P}(S) \cdot \mathbb{P}(F) \cdots \mathbb{P}(F)$ with k S's and $n - k$ F's. Then, we see that the sequence is equal to $\alpha^k (1-\alpha)^{n-k}$. There are exactly $\binom{n}{k}$ sequences of S's and F's with probability $\alpha^k (1 - \alpha)^{n-k}$. Therefore,

$$\mathbb{P}(X = k) = \binom{n}{k} \alpha^k (1 - \alpha)^{n-k}, \; k = 0, 1, 2 \ldots, n.$$

Thus, we see that $p(k)$ is just equal to the binomial distribution with parameters n, α. So, X is binomially distributed random variable with parameters n and α. $X \sim \text{bin}(n, \alpha)$.

Example: Throw a die until a quincunx appears. Let X be the number of throws. X is a random variable with range $\{1,2,3,\dots,n,\dots\} = \mathbb{N}$. By the independence of throws $\mathbb{P}(X = n) = \left(\frac{5}{6}\right)^{n-1}\frac{1}{6}$, where p is the geometric distribution function with parameter $\frac{1}{6}$. X is geometrically distributed with parameter $\frac{1}{6}$.

Example: An urn contains N balls, b of them black and w of them white. Choose m balls without replacement. Let X be the number of black balls in the m chosen. There are $\binom{b+w}{m}$ ways to choose m balls from $b+w$ balls. There are $\binom{b}{k}\binom{w}{m-k}$ ways of choosing m balls, with exactly k of the black. Therefore,

$$\mathbb{P}(X = k) = \frac{\binom{b}{k}\binom{w}{m-k}}{\binom{b+w}{m}} = \frac{\binom{b}{k}\binom{N-b}{m-k}}{\binom{N}{m}}$$

The hypergeometric distribution, with parameters N, b, m. X is a hypergeometrically distributed random variable.

Date: 10/10

Note: We are now pulling from information in Grinstead and Snell, $v2.2$ as a result of our former book not being equally available for all students.

Continuous Random Variable

Experiment Choose a number in $[0,1]$. Let X be the random variable representing the chosen value. *Classical View:* X is a variable assuming values in the range $[0,1]$. *Modern View:* $\Omega = [0,1]$ is the sample space. X is a real-valued function on Ω, in particular,

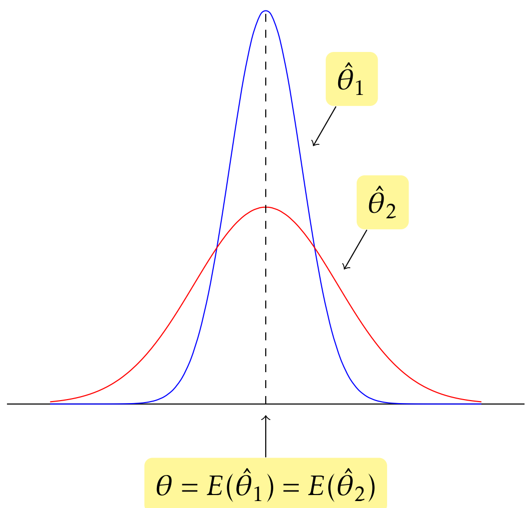
$$X : \Omega \to \mathbb{R}$$

by $X(\omega) = \omega$. That is, X is the identity function on Ω. For any $x \in [0,1]$, $\mathbb{P}(X = x) = 0$. So, there is no probability distribution mapping $p : \Omega \to [0,1]$ such that $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$ for this experiment.

$$f(x) = \begin{cases} 1 & \text{for } 0 < X < 1, \\ 0 & \text{for } X > 0,\ X < 1. \end{cases}$$

Thus, for $0 \le a < b \le 1$, $\mathbb{P}(a \le X \le b) = b - a = \int_a^b f(x)dx$. $f(x)$ is called a probability density function (pdf) for the random variable X. Since the range $[0,1]$ of X is not discrete, we say that X is a continuous random variable. Note every random variable can be classified as discrete **or** continuous. But nearly all of the important examples involve random variables that are either discrete with probability distributions or continuous with probability density functions.

Some Important Examples If X is a random variable with a pdf of the form $f_x = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$. Then, X is normally distributed with parameters μ and σ. The normal pdf f_x is also called the Gaussian distribution.



If T is a random variable with a pdf of the form

$$f_T(t) = \begin{cases} 0 & \text{for } t < 0, \\ \alpha e^{-\alpha t} & \text{for } t \geq 0. \end{cases}$$

where $\alpha > 0$, then T is an exponentially distributed random variable with parameter α. $\mathbb{P}(a \leq T \leq b) = \int_a^b \alpha e^{-\alpha t} dt = \left[-e^{-\alpha t} \right]_a^b = -e^{-a\alpha} - e^{-b\alpha}$.

If a random variable X with pdf

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x-m)^2}$$

where $x \in \mathbb{R}$, $m \in \mathbb{R}$ is Cauchy distributed with parameter m. The probability from a to b on a Cauchy distribution is found by $\int_a^b f_X(x)dx = \mathbb{P}(a \leq X \leq b)$.

Fix a and b, $a < b$. Consider the experiment of choosing a number X in the interval (a, b). Then, X has pdf

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b, \\ 0 & \text{for } x < a, x > b. \end{cases}$$

X is **uniformly** distributed on (a, b). Two properties shared by all pdfs $f(x)$ include

(i) $f(x) \geq 0$ for all x

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$.

When doing computations with normally distributed random variables, it helps to recall a result of Liouville's: For $a > 0$,

$$\int_{-\infty}^{\infty} e^{-ax^2}\, dx = \sqrt{\frac{\pi}{a}}.$$

Clearly, for any $\mu \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} e^{-a(x-\mu)^2}\, dx = \sqrt{\frac{\pi}{a}}.$$

This is also valid for any $a \in \mathbb{C}$ for which $\mathrm{Re}\{a\} < 0$.

Definition: Random Variables$*$ X and Y are independent if $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$, for subset A and B of $\mathbb{R}$. $*$ We should say "Jointly distributed random variables X and Y." This means that X and Y are defined for the sample range and sample space. A family $\{X_\lambda\}_{\lambda \in \Lambda}$ of (jointly distributed) random variables is independent if $\mathbb{P}(X_{\lambda_1} \in A_1, X_{\lambda_2} \in A_2, \ldots, X_{\lambda_n} \in A_n) = \mathbb{P}(X_{\lambda_1} \in A_1) \cdot \mathbb{P}(X_{\lambda_2} \in A_2) \cdots \mathbb{P}(X_{\lambda_n} \in A_n)$ for all finite subsets $\lambda_1, \lambda_2, \ldots, \lambda_n$ of Λ and subsets $A_1, \ldots, A_n$ of $\mathbb{R}$.

Let X be a random variable. The cumulative distribution function (cdf) of X is $F_X(x) = \mathbb{P}(X \le x)$. Consider the Bernoulli variable X, where X takes values 0 and 1, with $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$.
This is the cdf of the discrete random variable X.

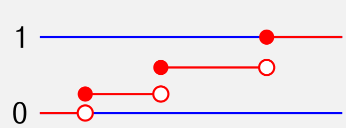
Date: 10/12

For any random variable X, the cdf is $F_X(x) = \mathbb{P}(X \le x)$. Suppose that X is the Bernoulli variable assuming values 0 and 1, with probabilities α and $1 - \alpha$.

$$\mathbb{P}(X = 0) = \alpha,\ \mathbb{P}(X = 1) = 1 - \alpha$$

Example Let N be a Poisson-distributed random variable with parameter $\lambda > 0$. Then, N assumes values in the range $\{0, 1, 2, \ldots\}$ with

$$\mathbb{P}(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$



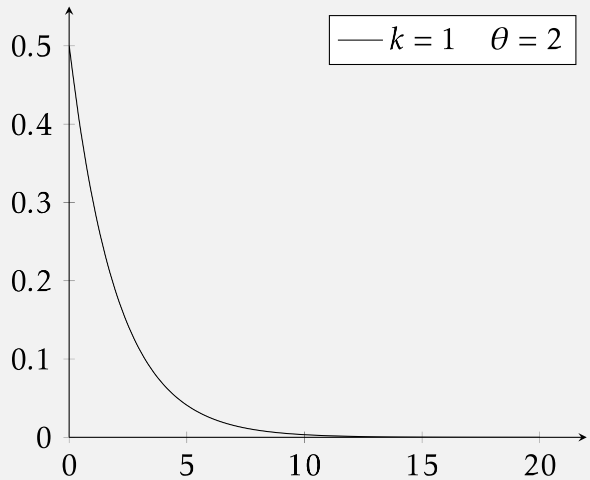
Let f be the continuous random variable X. Let $F(x) = \mathbb{P}(X \le x)$ be the cdf. For convenience, assume that $f(x)$ is continuous, except, possibly at finitely many points. Then,

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(u)\,du.$$

And by the Fundamental Theorem of Calculus, $F'(x) = f(x)$, where $f(x)$ is continuous.

Example Let T be exponentially distributed with parameter $\alpha > 0$. So, T has pdf

$$f(t) = \begin{cases} 0 & \text{for } t < 0, \\ \alpha e^{-\alpha t} & \text{for } t \geq 0. \end{cases}$$



What about the cdf?

$$F(t) = \mathbb{P}(T \leq t) = \int_{-\infty}^{t} f(s)ds.$$

Clearly, $F(t) = 0$ for $t < 0$. For $t \geq 0$,

$$F(t) = \int_{0}^{t} f(x)ds = 10e^{-\alpha t}.$$

Example Let X be a Cauchy-distributed with random variable with $m = 0$. Hence, X has pdf

$$f(x) = \frac{1}{\pi} + \frac{1}{1+x^2}.$$

The CDF is $F(x) = \int_{-\infty}^{\infty} f(u)du$, which is equal to

$$\begin{aligned}
&= \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+u^2}du \\
&= \frac{1}{\pi}\Big[\arctan(x) - \arctan(-\infty)\Big] \\
&= \frac{1}{\pi}\Big[\arctan(x)\frac{\pi}{2}\Big].
\end{aligned}$$

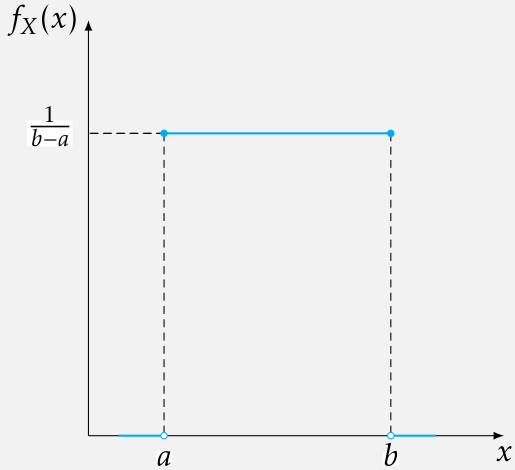
$F(x)$, the CDF of the Cauchy variable X.

Properties of CDFs: $F(x) = \mathbb{P}(X \leq x)$

1. $\lim_{x\to-\infty} F(x) = 0$,
2. $\lim_{x\to\infty} F(x) = 1$,
3. $F(x)$ is non-decreasing,
4. $F(x)$ is right-continuous, i.e. $\lim_{h\to 0^+} F(x)$ at every x.

Example Let U be a uniformly distributed random variable on the interval (a,b). So, U has the pdf

$$f(u) = \begin{cases} \frac{1}{b-a} & \text{for } a < u < b, \\ 0 & \text{for } u \le a, \ u \ge b. \end{cases}$$



The CDF is then

$$F(u) = \mathbb{P}(U \le u) = \int_{-\infty}^{u} f(v)dv = \begin{cases} 0 & \text{for } u \le a, \\ \frac{u-a}{u-b} & \text{for } a < u < b, \\ 1 & \text{for } u \ge b. \end{cases}$$

$\mathbb{P}(X \le x)$. We assign a probability to the event $(X \le x)$. Think of X as a function on a sample space Ω,

$$X : \Omega \to \mathbb{R}.$$

For any x, we'll write $(X \le x)$ for the set $\left\{\omega \in \Omega | X(\omega) \le x\right\}$. **Question:** For a CAPs $(\Omega, \mathcal{F}, \mathbb{P})$, is the subset $(X \le x)$ an event, i.e. is $(X \le x \in \mathcal{F}$? (Here $\mathcal{F}$ is the σ-field of events.)
Example: Consider the sample space $\Omega = \{1,2,3,4,5\}$. Let $A = \{1,2\}$, $B = \{3,4\}$, and $C = \{5\}$. Let $\mathcal{F}$ be the field (actually a σ-field) of events,

$$\left\{\emptyset, A, B, C, A+B, B+C, A+C, \Omega\right\}.$$

Let $X : \Omega \to \mathbb{R}$ by $X(\omega) = \omega$. The subset $(X \le 3)$ is $\{1,2,3\} \notin \mathcal{F}$. Hence, $\mathbb{P}(X \le 3)$ is undefined for the CAPS $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a CAPs. So, $\mathcal{F}$ is a σ-field of subsets of Ω. A random variable is a function $X : \Omega \to \mathbb{R}$ such that for every $x \in \mathbb{R}$, $(X \le x) \in \mathcal{F}$. So, X is an $\mathcal{F}$-measurable function.

Let X be a random variable. So, for every x, $(X \leq x) \in \mathcal{F}$. Observe that $(x < X) = (X \leq x)$. So, because $\mathcal{F}$ is closed under complementation, $(x < X)$ is an event. What about $(X < x)$? Well,

$$(X < x) = \bigcup_{n=1}^{\infty} (X \leq x - \frac{1}{n}) \in \mathcal{F}$$

because $\mathcal{F}$ is closed under countable union. By another complementation argument,

$$(x \leq X) \in \mathcal{F}. \tag{2}$$

By similar arguments, $(a < X \leq b)$, $(a < X \leq b)$, and $(a < X \leq b)$ lie in $\mathcal{F}$ also. So, for any interval I, $(X \in I) \in \mathcal{F}$. For $(a < X \leq b)$, $(a < X \leq b) = \Big((X \leq b) \in \mathcal{F} \cap (X > a) \in \mathcal{F}\Big) \in \mathcal{F}$ also.

Date: 10/19

We have a CAPs $(\Omega, \mathcal{F}, \mathbb{P})$. A random variable X is a function,

$$X : \Omega \to \mathbb{R}$$

satisfying the condition $\{\omega \in \Omega \mid X(\omega) \leq x\}$ for ever $x \in \mathbb{R}$ Because $(X \leq x) \in \mathcal{F}$ for every $x \in \mathbb{R}$, X is called $\mathcal{F}$-measurable. So, a random variable is an $\mathcal{F}$-measurable function on Ω. Let X be a random variable. We showed that for every interval I, $\{\omega \in \Omega \mid X(\omega) \in I\} = (X \in I) \in \mathcal{F}$. Let $A \subset \mathbb{R}$ be a set that could be "manufactured" by countably many union, intersection, and complementation operations, starting with intervals. Then, by the countable additivity of P, the face that $\mathcal{F}$ is a σ-field, and the $\mathcal{F}$-measurability of X, $(X \in A) \in \mathcal{F}$.

The set of all subsets A of $\mathbb{R}$ that can be constructed in this manner is itself a σ-field. It is called the Borel σ-field. So, for every $A \in \mathcal{B}$, the Borel σ-field, $(X \in A) \in \mathcal{F}$. Alternative **definition**: $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if $(X \in A) \in \mathcal{F}$ for every Borel subset A of $\mathbb{R}$. A rigorous definition of the Borel σ-field,

(i) Let $\left\{\mathcal{F}_\lambda\right\}_{\lambda \in \Lambda}$ be a family of σ-fields. Then, $\bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda$ is itself a σ-field.

(ii) Let $\left\{\mathcal{B}_\lambda\right\}_{\lambda \in \Lambda}$ be the class of all σ-fields of subsets of $\mathbb{R}$ that include all intervals of the form $(-\infty, x]$.

(iii) Thus, $\bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda$ is itself a σ-field that contains all subsets of $\mathbb{R}$ of the form $(-\infty, x]$. In fact, $\bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda$ is the smallest σ-field that contains these intervals. We define $\mathcal{B} = \bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda$. Thus, the Borel σ-field is the smallest σ-field that contains all sets of the form $(-\infty, x] \subset \mathbb{R}$.

Grinstead and Snell 4.1: Consider a discrete sample space Ω. Recall that a distribution function (df) m on Ω is a function

$$m : \Omega \to [0,1]$$

such that $\sum_{\omega \in \Omega} m(\omega) = 1$. Let X be a random variable on Ω (so, X is discrete) with range $\{x_i\}_{i \in I}$ where I is a countable index set. The probability mass function (pmf) $p : \{x_i\}_{i \in I} \to [0,1]$ by $p(x_i) = \mathbb{P}(X = x_i)$. Hence,

$$p(x_i) = \mathbb{P}(X = x_i) = \sum_{\omega \in (X = x_i)} m(\omega).$$

Throw a fair die twice. Thus, $\Omega = \left\{(1,1),(1,2),\ldots,(6,6),\right\}$. So, $|\Omega| = 36$, and we use the distribution function $m : \Omega \to [0,1]$ by $m(\omega) = \frac{1}{36}$. Define X to be the sum of the two throws. Thus, for $\omega = (\omega_1, \omega_2)$, $X(\omega) = \omega_1 + \omega_2$. The pmf of X is

$$\begin{aligned}
p(2) &= \mathbb{P}(X = 2) = \frac{1}{36} \\
p(3) &= \mathbb{P}(X = 3) = m(1,2) + m(2,1) = \frac{1}{18} \\
p(4) &= \mathbb{P}(X = 4) = m(1,3) + m(2,2) + m(3,1) = \frac{1}{12} \\
&\vdots \\
p(12) &= \mathbb{P}(X = 12) = m(6,6) = \frac{1}{36}
\end{aligned}$$

For discrete sample spaces and random variables, we have distribution functions an probability mass functions. For continuous random variables, we have probability distribution functions. For all random variables, we have cumulative distribution functions.

Let X and Y be jointly distributed random variables, i.e. random variables on the same CAPs $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that X and y are discrete, with ranges $\{x_i\}$ and $\{y_j\}$ respectively. The joint pmf of X and Y is $p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j) = \sum_{\omega \in (X = x_i, Y = y_j)} m(\omega)$. The conditional probability that $X = x_i$, given $Y = y_j$ is

$$\frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)}.$$

The joint probability mass function of X and Y is $p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$. Let p_X and p_Y be the probability mass functions of X and Y respectively. Then,

$$\begin{aligned}
p_X = \mathbb{P}(X = x_i) &= \sum_j \mathbb{P}(X = x_i, Y = y_j) \\
&= \sum_j p(x_i, y_j)
\end{aligned}$$

and,

$$\begin{aligned}p_Y = \mathbb{P}(Y = y_j) &= \sum_i \mathbb{P}(X = x_i, Y = y_j)\\ &= \sum_j p(x_i, y_j)\end{aligned}$$

In this context, p is the joint probability mass function and p_X and p_Y are the marginal probability mass functions. Let $p_{X|Y}(x_i|y_j) = \mathbb{P}(X = x_i|Y = y_j)$. Then, $p_{X|Y}(x_i|y_j) = \frac{p(x_i,y_j)}{p_Y(y_j)}$. $p_{X|Y}$ is the conditional probability mass function of X given Y.

If X and Y are independent, then $p(x_i, y_j) = p_X(x_i)p_Y(y_j)$. Why? Well,

$$\begin{aligned}p(x_i, y_j) &= \mathbb{P}(X = x_i, Y = y_j)\\ &= \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)\\ &= p_X(x_i)p_Y(y_j).\end{aligned}$$

Let X and Y be jointly distributed, continuous random variables. The joint probability density function of X and Y is the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\mathbb{P}(X \le a, Y \le b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)dx\,dy.$$

More generally,

$$\mathbb{P}(X \in A, Y \in B) = \int_B\int_A f(x,y)dx\,dy.$$

Let a and a be the probability dis

Date: 10/24

Let X be a discrete random variable with range $\{x_i\}$. The pmf of X is

$$\mathbb{P}_X(x) = \begin{cases}\mathbb{P}(X = x) & \text{for } x \text{ in the range of } X,\\ 0 & \text{if } a = 0 \text{for } x \text{ not in the range of } X.\end{cases} \tag{3}$$

Thus, $p_X(x_i) = \mathbb{P}(X = x_i)$. If X and Y are discrete (jointly distributed) random variables, then their joint probability mass function is $\mathbb{P}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$. The marginal probability mass functions are then $\mathbb{P}_X x_i = \sum_j \mathbb{P}(x_i, y_j)$ and $\mathbb{P}_Y(y_j) = \sum_i \mathbb{P}(x_i, y_j)$. The conditional probability mass function of X given $Y = y_j$ is

$$p_{X|Y}(x_i|y_j) = \frac{p(x_i, y_j}{p_Y(y_j)}.$$

Thus $p(1,4) = \mathbb{P}(X = 1, Y = 4) = .1$. The marginal probability mass function of Y is

$$\begin{aligned}
\mathbb{P}_Y = \mathbb{P}(Y = 2) &= \mathbb{P}(X = 0, Y = 2) + \mathbb{P}(X = 1, Y = 2) + \mathbb{P}(X = 2, Y = 2) \\
&= \mathbb{P}(X = 0, Y = 2) + \mathbb{P}(X = 1, Y = 2) + \mathbb{P}(X = 2, Y = 2) \\
&= .35
\end{aligned}$$

Similarly, $P\mathbb{P}_Y(3) = .35$ and $P\mathbb{P}_Y(4) = .4$. Now, what is $p_{X|Y}(x|2)$?

$$p_{X|Y}(0|2) = \frac{.15}{.35} = \frac{3}{7} \qquad p_{X|Y}(1|2) = \frac{.1}{.35} = \frac{2}{7} \qquad p_{X|Y}(2|2) = \frac{.1}{.35} = \frac{2}{7}$$

Let $\{X_\lambda\}_{\lambda\in\Lambda}$ be a set of jointly distributed random variables. The X_λ are independent if for every finite subset $\lambda_1, \ldots, \lambda_n$ of Λ, $\mathbb{P}(X_{\lambda_1} \in B_1, \ldots, X_{\lambda_n} \in B_n) = \mathbb{P}(X_{\lambda_1} \in B_1) \cdots \mathbb{P}(X_{\lambda_n} \in B_n)$. Back to the case of discrete, independent random variables X and Y, with ranges $\{x_i\}$ and $\{y_j\}$ (subsets with 1 point each). Then,

$$\begin{aligned}
p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j) &= \mathbb{P}(X \in \{x_i\}, Y \in \{y_j\}) \\
&= \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j) \\
&= \mathbb{P}_X(x_i) \cdot \mathbb{P}_Y(y_j).
\end{aligned}$$

Continuous Conditional Probability

Let X and Y be continuous random variables with joint probability distribution function $f(x,y)$. So,

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_c^d \int_a^b f(x,y)dx\, dy$$

More generally, let $B \subset \mathbb{R}^2$. Then,

$$\mathbb{P}((X,Y) \in B) = \iint_B f(x,y)dA.$$

Clearly, $\iint_{\mathbb{R}^2} f(x,y)dA = 1$.
Conditional probability distribution functions: We'd like t define $f_{X|Y}(x|y)$, the conditional probability distribution function of X given $Y = y$. How do we do this, given $\mathbb{P}(Y = y) = 0$ for a continuous variable Y? Start with $\mathbb{P}(a < X \leq b | Y = y)$. Assume that this expression is well-defined in some sense. Then, it should be that

$$
\begin{aligned}
\mathbb{P}(a < X \le b | Y = y) &= \lim_{h \to 0^+} \mathbb{P}(a < X \le b | y \le Y \le y + h) \\
&= \lim_{h \to 0^+} \frac{\mathbb{P}(a < X \le b, y \le Y \le y + h)}{\mathbb{P}(y \le Y \le y + h)} \\
&= \lim_{h \to 0^+} \frac{\frac{1}{h} \int_a^b \int_y^{y+h} f(x,\eta) d\eta \ dx}{\frac{1}{h} \int_y^{y+h} f_Y(\eta) d\eta} \\
&= \frac{\int_a^b f(x,y) dx}{f_Y(y)} \\
&= \int_a^b \frac{f(x,y)}{f_Y(y)} dx.
\end{aligned}
$$

Thus, $\frac{f(x,y)}{f_Y(y)} = f_{X|Y}(x|y)$ is the conditional probability distribution function of X given $Y = y$. Clearly, $f_{X|Y}(x|y)$, (as a function of x) is non-negative, and

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_{X|Y}(x|y) dx &= \int_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)} dx \\
&= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x,y) dx \\
&= \frac{f_Y(y)}{f_Y(y)} = 1
\end{aligned}
$$

Thus, $f_{X|Y}(x|y)$ at fixed y, is a pdf, as a function of x. We have the law of total probability in the continuous case

$$
\begin{aligned}
\mathbb{P}(a < X \le b) = \mathbb{P}(a < X \le b, -\infty < Y < \infty) &= \int_a^b \int_{-\infty}^{\infty} f(x,y) dy \ dx \\
&= \int_a^b \int_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)} f_Y(y) dy \ dx \\
&= \int_a^b \int_{-\infty}^{\infty} f_{X|Y}(x|y) \cdot f_Y(y) dy \ dx \\
&= \int_{-\infty}^{\infty} \int_a^b f_{X|Y}(x|y) dx \cdot f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \mathbb{P}(a < X \le b | Y = y) \cdot f_Y(y) dy
\end{aligned}
$$

Suppose now that X and Y are independent random variables. Then, $\mathbb{P}(a < X \le b, c < Y \le d) = \mathbb{P}(a < X \le b) \cdot \mathbb{P}(c < Y \le d)$. Thus,

$$\int_a^b \int_c^d f(x,y)dy\ dx = \int_a^b f_X(x)dx \int_c^d f_Y(y)dy = \int_a^b \int_c^d f_X(x)\cdot f_Y(y)dy\ dx.$$

Thus, $f(x,y) = f_X(x)\cdot f_Y(y)$. This extends in the obvious way to independent, continuous random variables $X_1,\ldots,X_n$. For any jointly distributed random variables $X_1,\ldots,X_n$, the joint cumulative distribution function is $F(x_1,\ldots,x_n) = \mathbb{P}(X_1 \le x_1,\ldots,X_n \le x_n)$. Suppose that the x_j are continuous, with joint probability distribution function $f(x_1,\ldots,x_n)$. Then,

$$\begin{aligned} F(x_1,\ldots,x_n) &= \mathbb{P}(X_1 \le x_1,\ldots,X_n \le x_n) \\ &= \int_{-\infty}^{x_n}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_1} f(u_1,\ldots,u_n)du_1\ldots du_n \end{aligned}$$

Thus, $f(x_1,\ldots,x_n) = \frac{\partial^n}{\partial x_1\partial x_2\cdots\partial x_n}F(x_1,\ldots,x_n)$.

Let $X_1,\ldots,X_n$ be random variables with cumulative distribution function $F(x_1,\ldots,x_n)$ and marginal cumulative distribution functions $f_{X_i}(x_i)$. If $f(x_1,\ldots,x_n)$ is the joint probability distribution function, then $\frac{\partial^n}{\partial x_1\partial x_2\cdots\partial x_n}F(x_1,\ldots,x_n) = f(x_1,\ldots,x_n)$. Finally, if $X_1,\ldots,X_n$ are independent, then

$$F(x_1,\ldots,x_n) = \mathbb{P}(X_1 \le x_1,\ldots,X_n \le x_n) = \mathbb{P}(X_1 \le x_1)\cdots\mathbb{P}(X_n \le x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_n).$$

Example: Let X and Y be continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} x^2 + xy + \frac{5}{4}y^2 & 0 < x < 1 \text{ and } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Say we want to find $\mathbb{P}(X < Y)$. Then, $\mathbb{P}(X < Y) = \mathbb{P}\Big((X,Y) \in T\Big)$. Formally, we know that this is equivalent to

$$\begin{aligned} \mathbb{P}\Big((X,Y) \in T\Big) &= \int_0^1 \int_{x=0}^{x=y} (x^2 + xy + \frac{5}{4}y^2) dx\, dy \\ &= \int_0^1 (\frac{1}{3}y^3 + \frac{1}{2}y^3 + \frac{5}{4}y^3) dy \\ &= \frac{25}{12}\int_0^1 y^3 dy \\ &= \frac{25}{48}. \end{aligned}$$

Find the marginal probability distribution function f_Y of Y. We see that

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y) dx \\ &= \int_{-\infty}^{\infty} (x^2 + xy + \frac{5}{4}y^2) dx \\ &= \frac{1}{3} + \frac{1}{2}y + \frac{5}{4}y^2, \text{ for } 0 < y < 1. \end{aligned}$$

Hence, at fixed $y \in (0,1)$, the conditional probability distribution function of X, given $Y = y$ is.

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} 0 & \text{for } x \leq 0,\ x \geq 1, \\ \frac{x^2+xy+\frac{5}{4}y^2}{\frac{1}{3}+\frac{1}{2}y+\frac{5}{4}y^2} & \text{for } 0 < x < 1. \end{cases}$$

Expected Value

Let X be a discrete random variable on a discrete sample space Ω. The expected value or expectation of X (or the mean of X) is the average value of X, taken over Ω. The $E(X) = \sum_{\omega \in \Omega} X(\omega)m(\omega)$, where m is the distribution function on Ω. **Note:** m is valid as a distribution function here because the sample space is discrete. A **more useful** expression is based on the probability mass function $p_X(x_i)$. Let $\{x_i\}$ be the range of X. Then, the expected value of X is $\mathrm{E}(X) = \sum_{\omega\in\Omega} X(\omega)m(\omega) = \sum_i \sum_{X=x_i} X(\omega)m(\omega) = \sum_i \sum_{X=x_i} m(\omega) = \sum_i \mathbb{P}(X = x_i) = \sum_i x_i p_X(x_i)$. Here, the constraint on the second summation is really " for every i, $X = x_i$", which is equivalent to $\{\omega \in \Omega | X(\omega) = x_i\}$.

Suppose now that X is a continuous random variable with probability distribution function $f_X(x)$. In this case, a simple limiting argument yields the formula $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_x(x)dx$. Should the sum or integral diverge, the random variable X is said to have no expected value.

Example: Let X be a Bernoulli variable with range $\{x_1, x_2\}$ and probability mass function $p_X(x_1) = \alpha_1$, $p_X(x_2) = \alpha_2$, $\alpha_1 + \alpha_2 = 1$. Then, X has mean $\mathrm{E}(X) = \alpha_1 x_1 + \alpha_2 x_2$.

Example: We are flipping a coin until a tail occurs. Let $P(H) = p$, $P(T) = 1 - p$, $0 < p < 1$. Let N be the number of the trial in which the T occurs. Then, N is geometrically distributed:

$$p_N(n) = P(N = n) = p^{n-1}(1-p),\ n = 1, 2, 3, \ldots$$

Then, $\mathrm{E}(N) = \sum_{n=1}^{\infty} n p_N(n) = \sum_{n=1}^{\infty} n p^{n-1}(1-p) = (1-p)\frac{d}{dp}\sum_{n=0}^{\infty} p^n = (1-p)\frac{d}{dp}\frac{1}{1-p} = \frac{1}{1-p}$.
Example: Let X be Poisson-distributed with parameter $\lambda > 0$. So, X has range $\{0, 1, 2, 3, \ldots\}$ and probability mass function $f_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}$. Then, $\mathrm{E}(X) = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda} \cdot e^{-\lambda} = \lambda$. So, $\mathrm{E}(X) = \lambda$ is the mean occurrence rate.

Let Y be Cauchy-distributed with parameters $m = 0$. Y has probability distribution function $f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{1+(y-m)^2} = \frac{1}{1+y^2} \cdot \frac{1}{\pi}$, when $m = 0$. Hence

$$\begin{aligned}
\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{1+y^2}dy \\
&= \lim_{a\to\infty,\, b\to\infty} \frac{1}{\pi}\int_0^a \frac{y}{1+y^2}dy + \frac{1}{\pi}\int_{-b}^{\infty} \frac{y}{1+y^2}dy \\
&= lim_{a\to\infty}\frac{1}{\pi}\int_0^{\infty} \frac{y}{1+y^2}dy \\
&= lim_{a\to\infty}\frac{1}{2\pi}\ln(1+a^2) = \infty.
\end{aligned}$$

So, the integral diverges, and we conclude that $\mathrm{E}(X)$ does not exist for the Cauchy distribution.

Let X be normaly distributed with parameters μ, σ. So, X has the probability distribution function $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$, $\mu \in \mathbb{R}$, $\sigma > 0$. $\mathrm{E}(X) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$. Set $z = \frac{x-\mu}{\sigma}$. So, $dx = \sigma dz$, and the integral becomes:

$$\frac{1}{\sqrt{2\pi\sigma^2}}\Big[\int_{-\infty}^{\infty}\sigma z e^{-\frac{z^2}{2}\cdot\sigma dz}+\mu\int_{-\infty}^{\infty}\sigma z e^{-\frac{z^2}{2}}\cdot\sigma dz\Big]$$

The first term is equal to 0, and the second term is equal to $\sigma\mu\sqrt{2\pi}$ from Liouville's formula.

Date: 10/31

For X discrete with range $\{x_i\}$, $\mathrm{E}[X]=\sum_\Omega = X(\omega)m(\omega)=\sum_i x_i p_x(x_i)$, where m is the df on Ω and p_x the probability mass function of x. For a continuous x, $\mathrm{E}[X]=\int_{-\infty}^{\infty} x f_X(x)\,dx$ where f_X is the probability distribution function of X.

Properties of expectations: Let X be a random variable and $g:\mathbb{R}\to\mathbb{R}$ be continuous. Then, $Y=g(X)$ is also a random variable. What is $\mathrm{E}[X]=\mathrm{E}[g(X)]$? Suppose that X is discrete with probability mass function p_X. Let m be a df on Ω. Then, by definition, $\mathrm{E}[Y]=\sum_\Omega g(X(\omega))m(\omega)=\sum_i\sum_{x=x_i} g(X(\omega))m(\omega)=\sum_i\sum_{x=x_i} g(x_i)m(\omega)=\sum_i g(x_i)\sum_{x=x_i} m(\omega)=\sum_i g(x_i)\mathbb{P}(X=x_i)=\sum_i g(x_i)p_X(x_i)$.

For a continuous random variable X, with probability distribution function f_X, $\mathrm{E}[g(X)]=\int_{-\infty}^{\infty} g(x)f_X(x)\,dx$.

Let T be exponentially distributed with probability distribution function

$$f_T(t)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}0 & \text{for } t<0, \text{ wher } \alpha>0.\\ \alpha e^{-\alpha t} & \text{for } t\geq 0.\end{cases}$$

Then, we see that

$$\begin{aligned}\mathrm{E}[T]=\int_{-\infty}^{\infty} t f_T(t)\,dt=\int_0^\infty \alpha t e^{-\alpha t}\,dt&\\ &=\int_0^\infty t d\Big(e^{-\alpha t}\Big)=\Big[-te^{-\alpha t}\Big]_0^\infty+\int_0^\infty e^{-\alpha t}\,dt\\ &=\left[\frac{e^{-\alpha t}}{-\alpha}\right]_0^\infty=\frac{1}{\alpha}.\end{aligned}$$

Let $g(u)=e^{-u}$. The expected value of the random variable $Y=g(T)=e^{-T}$ is $\mathrm{E}[Y]=\mathrm{E}[g(T)]=\int_{-\infty}^{\infty} g(t)f_T(t)\,dt=\int_0^\infty e^{-t}e^{-\alpha t}\,dt\cdot\alpha=\alpha\int_0^\infty e^{-(1+\alpha)t}dt=\frac{\alpha}{1+\alpha}$.

Next, let c be any constant and $g:\mathbb{R}\to\mathbb{R}$ by $g(u)=cu$. Then (for X discrete), $\mathrm{E}[cX]=\mathrm{E}[g(X)]=\sum_i g(x_i)f_X(x_i)=\sum_i cx_i f_X(x_i)=c\sum_i x_i f_X(x_i)=c\mathrm{E}[X]$. The same result is valid for the continuous case: $\mathrm{E}[cX]=c\mathrm{E}[X]$.

Let $g : \mathbb{R}^2 \to \mathbb{R}$ be continuous. Let X and Y be jointly distributed random variables. What is $\mathrm{E}\Big[g(X,Y)\Big]$? Let $Z = g(X,Y)$. Assume that X and Y are discrete with ranges $\{x_i\}$ and $\{y_i\}$ respectively and joint probability mass function f. Then, $\mathrm{E}[Z] = \mathrm{E}[g(X,Y)] = \sum_\Omega g(X(\omega), Y(\omega))m(\omega)$. Let A_{ij} be the event $(X = x_i, Y = y_j)$. [**Note**: A_{ij} is just the intersection of the events $X = x_i$ and $Y = y_j$ because both are Borel subsets the real number line, and thus subsets of Ω]. Then, by $\mathrm{E}[Z] = \mathrm{E}[g(X,Y)]$, $\mathrm{E}[g(X,Y)] = \sum_i \sum_j \sum_{A_{ij}} g(X(\omega), Y(\omega))m(\omega) = \sum_i \sum_j g(x_i, y_j) \sum_{A_{ij}} m(\omega) = \sum_i \sum_j g(x_i, y_j)\mathbb{P}(m(\omega) = \sum_i \sum_j g(x_i, y_j) p(x_i, y_j)$.

In the continuous case, $\mathrm{E}\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dxdy$, where f is the joint probability distribution function. Special case: $g(u,v) = u + v$. Then, $\mathrm{E}\Big[g(X,Y)\Big] = \mathrm{E}\Big[X + Y\Big] = \sum_\Omega \Big[X(\omega) + Y(\omega)\Big] m(\omega) = \sum_\Omega X(\omega)m(\omega) + \sum_\Omega Y(\omega)m(\omega) = \mathrm{E}\Big[X\Big] + \mathrm{E}\Big[Y\Big] = \sum_i x_i p_X(x_i) + \sum_j y_j p_Y(y_j)$. The same holds in the case of continuous variables. Combine two results: Let X and Y be jointly distributed random variables, and a and b constants. Then, $\mathrm{E}\Big[aX + bY\Big] = a\mathrm{E}[X\Big] + b\mathrm{E}[Y\Big]$.

Let X and Y be independent, jointly distributed random variables. Then, $\mathrm{E}\Big[XY\Big] = \mathrm{E}\Big[X\Big] \cdot \mathrm{E}\Big[Y\Big]$. Suppose that X and Y are continuous, with joint probability distribution function f. Let $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(u,v) = uv$. Then, $\mathrm{E}\Big[XY\Big] = \mathrm{E}\Big[g(X,Y)\Big] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy$.

Date: 11/7

Let X be binomially distributed with parameters n, p: $P(X = k) = p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, 2, \ldots, n$. We already know that $\mathrm{E}(X) = np$. Then,

$$
\begin{aligned}
\mathrm{E}(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} (k^2 - k + k) \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} (k^2 - k) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \left[\sum_{k=2}^{n} k(k-1) \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \right] + np \\
&= \left[n(n-1)p^2 \sum_{k=2}^{n} k(k-1) p^{k-2} \frac{(n-2)!}{(n-k)!k!} (1-p)^{n-k} \right] + np \\
&= \left[n(n-1)p^2 \sum_{k=2}^{n} k(k-1) p^2 \frac{(n-2)!}{(k-2)!\big[(n-2)-(k-2)\big]!} p^{k-2} (1-p)^{(n-2)-(k-2)} \right] + np, n-2 = m, k-2 = l \\
&= \left[n(n-1)p^2 \sum_{l=0}^{m} \binom{m}{l} p^l (1-p)^{m-l} \right] + np \\
&= n(n-1)p^2 + np
\end{aligned}
$$

So, $\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p)$. Another method to approach this is to define the indicator, or "counter" variables $X_1, \ldots, X_n$ by

$$
X_j = \begin{cases} 1 & \text{if "H" on flip } j, \\ 0 & \text{if "T" on flip } j. \end{cases}
$$

The X_j above are iid (independent, identically distributed) random variables.

Chebyshev's Inequality: Let X be a random variable with $\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) = \sigma^2$. For $\epsilon > 0$, what is $\mathbb{P}(|x - \mu| \geq \epsilon)$? Clearly this depends on X. Chebyshev's inequality is an estimate of this probability: $\mathbb{P}(|x - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$.

Proof Let's take (WLOG) X to be continuous with pdf f_X. Then,

$$\begin{aligned}\sigma^2 = \mathrm{V}(X) &= \mathrm{E}(\Big[(X-\mu)^2\Big]) \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx \\ &\geq \int_{|x-\mu|\geq\epsilon} (x-\mu)^2 f_x(x)dx \\ &\geq \int_{|x-\mu|\geq\epsilon} \epsilon^2 f_x(x)dx \\ &= \epsilon^2 \int_{|x-\mu|\geq\epsilon} f_x(x)dx \\ &= \epsilon^2 \mathbb{P}(|x-\mu|\geq\epsilon)\end{aligned}$$

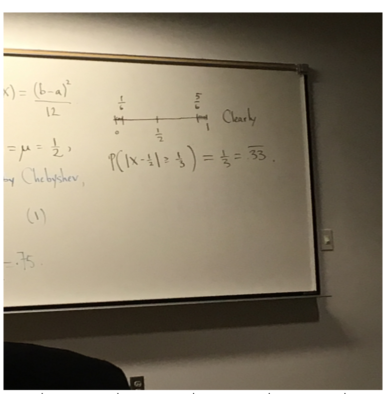
Just how good is Chebyshev's inequality at estimating $\mathbb{P}(|x-\mu|\geq\epsilon)$? Let X be uniformly distributed on (a,b): So X is continuous, with pdf

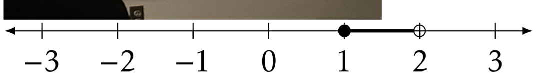
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b, \\ 0 & \text{otherwise.}\end{cases}$$

Then, $\mathrm{E}(X) = \frac{b+a}{2}$ and $\mathrm{V}(X) = \frac{(b-a)^2}{12}$. So, with $a = 0$, $b = 1$, $\mathrm{E}(X) = \mu = \frac{1}{2}$, and $\mathrm{V}(X) = \sigma^2 \frac{1}{12}$. So, by Chebyshev,

$$\mathbb{P}(|x-\frac{1}{2}|\geq\epsilon) \leq \frac{1}{12\epsilon^2}.$$

Thus, $\mathbb{P}(|x-\frac{1}{2}|\geq\frac{1}{3}) \leq \frac{9}{12} \sim 0.75$.





Let Z be a standard-normal random variable...

Sums of Independent Random Variables

Let f and g be pdfs. The convolution integral is

$$(f \cdot g)(x) = \int_{-\int}^{\int} f(x-y)g(y) \;\; dy.$$

Let $x - y = \xi$. So, $d\xi = -dy$, and hence

$$\begin{aligned}(f \cdot g)(x) &= -\int_{\infty}^{-\infty} f(\xi)g(x-\xi)d\xi \\ &= \int_{-\infty}^{\infty} g(x-\xi)f(\xi)d\xi \\ &= (g \cdot f)(x)\end{aligned}$$

So, $\cdot$ is commutative. **Special Case**: Suppose that $f(x) = g(x) = 0$ for $x < 0$. Then,

$$\begin{aligned}(f \cdot g)(x) &= \int_{-\infty}^{\infty} f(x-y)g(y)dy \\ &= \int_{0}^{\infty} f(x-y)g(y)dy \\ &= \int_{0}^{x} f(x-y)g(y)dy.\end{aligned}$$

Let X and Y be independent random variables with pdfs f_X and f_Y respectively. Let $Z = X + Y$ have pdf f_Z with cdf f_Z. Let f_X and f_Y be the cdfs of X and Y.

$$\begin{aligned}F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(X + Y \leq z) &= \int_{-\infty}^{\infty} \mathbb{P}(X + Y \leq z | Y = y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X + y \leq z | Y = y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X \leq z - y | Y = y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}(X \leq z - y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} F_X(z - y) f_Y(y)dy.\end{aligned}$$

So, $f_Z(z) = \frac{d}{dz}F_z(z) = \frac{d}{dz}\int_{-\infty}^{\infty} F_x(z-y)f_y(y)dy = \int_{-\infty}^{\infty} F_x(z-y)f_y(y)dy = (f_x \cdot f_y)(z)$. In other words, $f_z(z)$ is equal to the convolution, $(f_x \cdot f_y)(z)$.

Date: 11/9

Sums of Independent Random Variables

Recall that if X and Y are continuous and independent, with probability density function f_x and f_y then $Z = X + Y$ has probability density function

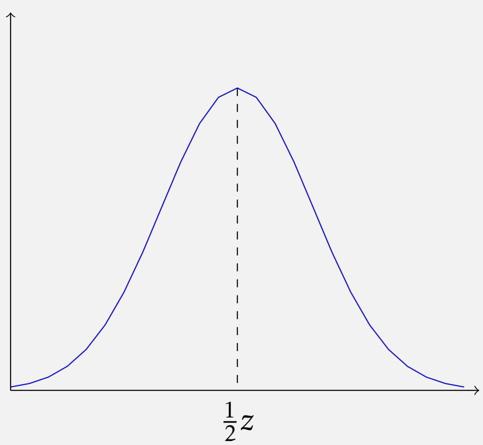
$$f_z(z) = (f_x \cdot f_y)(z) = \int_{-\infty}^{\infty} f_x(z-y)f_y(y)dy.$$

The convolution is associative: $(f \cdot g) \cdot h = f \cdot (g \cdot h)$. We'll just write $f \cdot g \cdot h$ when grouping is not an issue. We may thus extend the convolution to any set $f_1, \ldots, f_n$ of probability density functions. So, if $X_1, \ldots, X_n$ are continuous, independent random variables with probability density functions $f_1, \ldots, f_n$ respectively, then $z = X_1 + \ldots + X_n$ has probability density function $f_z(z) = (f_1 \cdot \cdots \cdot f_n)(z)$.

Let X and Y be normally distributed with mean 0 and variance 1. Let them be independent. Then $Z = X + Y$ is... ? Well,

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 0 \qquad \mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) = 2$$

$$\begin{aligned}
f_z(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(z-y)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{y^2}{2}}\, dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[(z-y)^2+y^2\right]} dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[z^2-2yz+2y^2\right]} dy \\
&= \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2-yz)} dy \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} e^{\frac{1}{4}z^2} \int_{-\infty}^{\infty} e^{-(y^2-yz+\frac{1}{4}z^2)} dy \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}z^2} \int_{-\infty}^{\infty} e^{-(y-\frac{1}{2}z)^2} dy
\end{aligned}$$



Note that as you vary z, you do not change the value of the above integral.

Let T be exponentially distributed with parameter $\lambda > 0$. So, T has probability density function

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

Thus, the $\mathrm{E}(T) = \int_{-\infty}^{\infty} tf(t)dt = \frac{1}{\lambda}$, and $\mathrm{V}(T) = \frac{1}{\lambda^2}$.

Let $T_1, \ldots, T_n$ be the service times for customers $1, \ldots, n$, respectively. Assume that the T_j are iid exponentially distributed random variables, each wiit parameter λ. Thus, in this context, $\frac{1}{\lambda} = \mathrm{E}(T)$ is the mean service time. Hence, λ is the mean service rate (number of customers served per unit time). λ is called the rate parameter. $S_n = T_1 + \cdots + T_n$ is time taken to serve n customers. How is S_n distributed? Let $g_n(t)$ be the probability distribution function of S_n. So, $g_1(t) = f(t)$, $g_2(t) = (f \cdot f)(t)$, and $g_n(t) = (f \cdot \cdots \cdot f)(t)$. For $t \geq 0$,

$$\begin{aligned}
g_2(t) &= \int_{-\infty}^{\infty} f(t-s)f(s)ds \\
&= \int_0^t f(t-s)f(s)ds \\
&= \lambda^2 \int_0^t e^{-\lambda(t-s)}e^{-\lambda s}ds \\
&= \lambda^2 \int_0^t dse^{-\lambda t} \\
&= \lambda^2 te^{-\lambda t}.
\end{aligned}$$

So, $T_1 + T_2 = S_2$ has probability distribution function

$$g_2(t) = \begin{cases} \lambda^2 te^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The probability distribution function of $S_3 = T_1 + T_2 + T_3$ is (for $t \geq 0$)

$$\begin{aligned}
g_3(t) = (f \cdot f \cdot f)(t) = f \cdot (f \cdot f)(t) = (f \cdot g_2)(t)& \\
&= \int_0^t \lambda e^{-\lambda(t-s)} \cdot \lambda^2 sds = e^{-\lambda t} \int_0^t sds \\
&= \frac{\lambda^3}{2} t^2 e^{-\lambda t}.
\end{aligned}$$

Therefore, s_3 has probability distribution function

$$g_3(t) = \begin{cases} \frac{\lambda^3}{2} t^2 e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

By induction, $g_n(t)$, the probability distribution function of $s_n = T_1 + \cdots + T_n$ is

$$g_n(t) = \begin{cases} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

$g_n(t)$ is the probability distribution function of the Erlang distribution, with rate parameter λ and shape parameter n. The gamma function is $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, defined for $x > 0$. For an integer $n > 0$, $\Gamma(n) = (n-1)!$. So, for $t \geq 0$, $g_n(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}$. Let $\alpha > 0$. The probability distribution function for the gamma distribution, with rate parameter λ and shape parameter α is

$$\begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

Date: 11/16

$$g_1(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

Then, we have $g_2(t) = (g_1 \cdot g_1)(t), \ldots, g_n(t) = (g_1 \cdots g_1)(t) = (g_1 \cdot g_{n-1})(t)$. This generalizes to

$$g_n(t) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

For β not necessarily an integer, we define the gamma density with rate parameter $\lambda > 0$ and shape $\beta > 0$,

$$g_n(t) = \begin{cases} \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

Let $T_1, \ldots, T_n$ be service times for customers in a queuing system. we follow the simplest model $(m/m/1)$ of queues. The T_k is identically and independently exponentially distributed random variables (iid's). Let λ be their common rate parameter. Then, $\mathrm{E}(T_k) = \frac{1}{\lambda}$ and $\mathrm{V}(T_k) = \frac{1}{\lambda^2}$ for $k = 1, \ldots, n$. The time required to serve all n customers is $S_n = T_1 + \cdots + T_n$. Therefore, S_n is Erlang/gamma distributed - its probability distribution function is $g_n(t)$.

Hence, since T_k is iid, $\mathrm{E}(S_n) = n\mathrm{E}(T_1) = \frac{n}{\lambda}$ and $\mathrm{V}(S_n) = n\mathrm{V}(T_1) = \frac{n}{\lambda^2}$.

Consider a small café with one barista and iid exponential service times with rate parameter $\lambda = .4$/min. Let A_n be the event that, upon your arrival, there are n customers in line ahead of you. Let T be your unconditional total time in the system. Thus, your total (waiting + service) in the queuing system is

$$S_{n+1} = T_1 + \cdots + T_n + T_{n+1}$$

where $T_1, \ldots, T_n$ are the service times of those ahead of you, and T_{n+1} your service time*. Hence, your time S_{n+1}, given A_n is gamma distributed with rate parameter $\lambda = .4$/min and shape parameter $n+1$. So,

$$\mathbb{P}(T > t \mid A_n) = \mathbb{P}(S_{n+1} > t) = \int_t^\infty g_{n+1}(s)ds = \frac{.4^{n+1}}{n!}\int_t^\infty s^n e^{-.4s}ds.$$

In R, this can be written as

$$\text{pgamma}(t, n{+}1.4, lower.tail = FALSE) = \text{pgamma}(t, n{+}1.4, lower = F) = 1 - \text{pgamma}(t, n{+}1.4).$$

For example,

$$\mathbb{P}(T > 10 \mid A_3) = \mathbb{P}(S_4 > 10) = \text{pgamma}(t0, 4.4, lower.tail = FALSE) = 0.43347.$$

* Conditional time in the system, given A_n.

What is the unconditional distribution of T?

For queuing in equilibrium, the probability that the system state (number of customers in line at the café) is $(1-\rho)\rho^n$ where $\rho = \frac{\mu}{\lambda}$ where μ is the mean arrival wait,

$$
\begin{aligned}
\mathbb{P}(T > t) &= \sum_{n=0}^{\infty} \mathbb{P}(T > t \mid A_n)\mathbb{P}(A_n) \qquad \text{(Law of Total Probability)} \\
&= \sum_{n=0}^{\infty} \mathbb{P}(S_{n+1} > t)(1-\rho)\rho^n \\
&= \sum_{n=0}^{\infty} \frac{\lambda^{n+1}}{n!} \int_t^{\infty} s^n e^{-\lambda s} ds (1-\rho)\rho^n \\
&= (1-\rho)\lambda \int_t^{\infty} \sum_{n=0}^{\infty} \frac{(\lambda\rho s)^n}{n!} e^{-\lambda s} ds \\
&= (1-\rho)\lambda \int_t^{\infty} e^{\lambda\rho s} e^{-\lambda s} ds \\
&= (1-\rho)\lambda \int_t^{\infty} e^{-\lambda(1-\rho)s} ds. \\
&= (1-\rho)\lambda \left[\frac{e^{-\lambda(1-\rho)s}}{-\lambda(1-\rho)} \right]_{s=t}^{s=\infty}.
\end{aligned}
$$

Thus, T is exponentially distributed with parameter $\lambda(1-\rho)$. So, your expected (unconditional) total time in the system is $\mathrm{E}(T) = \frac{1}{\lambda(1-\rho)}$.

Rayleigh Distribution

Let X and Y be independent, standard normal variables. So,

$$
f_X(x) = f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.
$$

Then, how is X^2 distributed? Well, $\mathbb{P}(X^2 \le r) = F_{X^2}(r) = \mathbb{P}(-\sqrt{r} \le X \le \sqrt{r}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{r}}^{\sqrt{r}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{r}} e^{-\frac{x^2}{2}}$. Hence, the probability distribution function of X^2

$$
\begin{aligned}
f_{X^2}(r) = F_{X^2}^{'}(r) &= \frac{2}{\sqrt{2\pi}} e^{-\frac{r}{2}} \cdot \frac{1}{2\sqrt{r}} \\
&= \frac{1}{\sqrt{2\pi r}} e^{-\frac{r}{2}},
\end{aligned}
$$

for $r > 0$.

$$\begin{aligned}
\Gamma(\frac{1}{2}) = \int_0^\infty t^{\frac{1}{2}-1}e^{-t}dt &= \int_0^\infty \frac{e^{-t}}{\sqrt{t}}dt \\
&= \int_0^\infty \frac{e^{-u^2}}{u}2u \ \ du \\
&= 2\int_0^\infty e^{-u^2}du \\
&= 2\int_{-\infty}^\infty e^{-u^2}du \\
&= \sqrt{\pi}.
\end{aligned}$$

The gamma density with rate parameter $\lambda = \frac{1}{2}$ and shape parameter $\beta = \frac{1}{2}$ is

$$g(r) = \frac{(\frac{1}{2})^{\frac{1}{2}} r^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})} e^{-\frac{1}{2}r} = \frac{1}{\sqrt{2\pi r}} e^{-\frac{1}{2}r},$$

for $r > 0$.
Thus, X^2 and Y^2 are gamma distributed with $\lambda = \frac{1}{2}$ and $\beta = \frac{1}{2}$.

Date: 11/28

7.1 The (weak) Law of Large Numbers

Let $\left\{x_k\right\}_{k=1}^\infty$ be i.i.d. random variables, with means $\mathrm{E}(x_k) = \mu$ and variance $\mathrm{V}(x_k) = \sigma^2$, $k = 1, 2, 3, \ldots$ where μ and σ^2 are finite. Let $S_n = X_1 + \ldots + X_n$. How is $\frac{S_n}{n}$ distributed for "large" n? Consider the degenerate *probability mass function* ,

$$p(x) = \begin{cases} 1 & \text{if } x = \mu, \\ 0 & \text{otherwise..} \end{cases}$$

In some sense, it should be that $\frac{S_n}{n}$ (approximately) has *probability mass function* p for n "large". This statement is the assertion of the *Weak Law of Large Numbers*,

$$\text{For } \epsilon > 0, \lim_{n\leftarrow\infty} P(|\tfrac{S_n}{n} - \mu| < \epsilon) = 1.$$

Proof The expected value can be calculated as follows,

$$\mathrm{E}(\frac{S_n}{n}) = \frac{1}{n}\mathrm{E}(S_n) = \frac{1}{n}n\mu = \mu.$$

The variance can be calculated as follows,

$$\mathrm{V}(\frac{S_n}{n}) = \frac{1}{n^2}\mathrm{V}(S_n) = \frac{1}{n^2}nV(X_1) = \frac{1}{n}\sigma^2.$$

Then,

$$P(|\frac{S_n}{n} - \mu| < \epsilon) + P(|\frac{S_n}{n} - \mu| \geq \epsilon) = 1.$$

By Chebysev, $P(|\frac{S_n}{n} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$. By plugging this into the equation directly above, letting $n \to \infty$,

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - \mu| < \epsilon) = 1.$$

Example Coin flips: Let $\left\{x_k\right\}_{k=1}^{\infty}$ be a sequence of i.i.d. Bernoulli random variables with ranges $\{0,1\}$ and common *probability mass function* $p(0) = p(1) = .5$. Think of X_k as the "head-counting" variable for a sequence of fair coin flips. So,

$$X_k = \begin{cases} 0 & \text{if "Tail" on flip } k, \\ 1 & \text{if "Head" on flip } k. \end{cases}$$

So, $S_n = \sum_{k=1}^{n} X_k$ is the number of heads observed in n flips. We know that $\mathrm{E}(X_k) = .5$ and that $\mathrm{V}(X_k) = .25$. So, by the *Weak Law of Large Numbers*, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - .5| < \epsilon) = 1.$$

Since $\mathrm{V}(\frac{S_n}{n}) \to 0$ as $n \to \infty$, the support of the probability mass function (or probability density function) of $\frac{S_n}{n}$ is collapsing to the single point μ. Consider $\frac{S_n}{\sqrt{n}}$. Then, $\mathrm{V}(\frac{S_n}{\sqrt{n}}) = \frac{1}{n}\mathrm{V}(S_n) = \frac{1}{n}n\mathrm{V}(X_1) = \sigma^2$. Since $\mathrm{V}(\frac{S_n}{\sqrt{n}})$ does not go to 0, can we say anything about the asymptotic distribution of $\frac{S_n}{\sqrt{n}}$? Note that $\mathrm{E}(\frac{S_n}{\sqrt{n}}) = \sqrt{n}\mu \to \infty$ as $n \to \infty$. Hence in considering the asymptotic behavior of $\frac{S_n}{\sqrt{n}}$, we have to make one more adjustment. We "center" the X_k by subtracting their means μ. While we're at it, we may as well normalize variances to 1. In summary, we consider the "standardized" random variables

$$X_k^* = \frac{X_k - \mu}{\sigma}.$$

The expected value is then $\mathrm{E}(X_k^*) = \frac{1}{\sigma}\mathrm{E}(X_k - \mu) = 0$ and the variance is $\mathrm{V}(X_k^*) = \mathrm{V}(\frac{X_k - \mu}{\sigma}) =$

$\frac{1}{\sigma^2}\mathrm{V}(X_k-\mu)=1$. Let $S_n^* = X_1^* + \ldots + X_n^*$.

Question: How does $\frac{S_n^*}{\sqrt{n}}$ behave as $n \to \infty$? Bear in mind, $\mathrm{E}(\frac{S_n^*}{\sqrt{n}})=0$ and $\mathrm{V}(\frac{S_n^*}{\sqrt{n}})=\frac{1}{n}n=1$.

Well, let the X_k be i.i.d. exponentials with $\lambda=2$. Using R, we generate 1000 samples values and plotted the histogram. Numerical evidence suggests that for n large, $\frac{S_n^*}{\sqrt{n}} \sim N(0,1)$, approximately.

We prove (sort of) that this is indeed the case, by means of moment generating functions. The moment generating function (MGF) $g_x(t)$ of a random variable X is $g_X(t)=\mathrm{E}(e^{tX})$.

Example Let X be Poisson distributed, with mean λ. Then,

$$g_X(t)=\mathrm{E}(e^{tX})=\sum_{n=0}^{\infty} e^{tn}\frac{\lambda^n}{n!}e^{-\lambda}=e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^t\lambda)^n}{n!}=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}.$$

Example Let X be Gaussian distributed, with mean 0 and variance 1. Then,

$$\begin{aligned}
g_X(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tn}e^{-\frac{x^2}{2}}dx\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx+t^2)}dx\cdot e^{\frac{1}{2}t^2}\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)}dx\cdot e^{\frac{1}{2}t^2}\\
&= \frac{\sqrt{2\pi}}{\sqrt{2\pi}}e^{\frac{1}{2}t^2}, \text{ using Liouville's formula}\\
&= e^{\frac{1}{2}t^2}
\end{aligned}$$

Skipping some technical details,

$$X \text{ and } Y \text{ have the same distribution} \leftrightarrow g_X(t)=g_Y(t).$$

The **Central Limit Theorem** (CLT) asserts that $\frac{S_n^*}{\sqrt{n}}$ is asymptotically standard-normal as $n\to\infty$. We will prove this by showing that

$$g_{\frac{S_n^*}{\sqrt{n}}}(t) \longrightarrow e^{\frac{1}{2}t^2} \text{ as } n\to\infty.$$

We will need these properties:

— Let X and Y be independent with MGF's g_X and g_Y, respectively. Then, $g_{X+Y}(t)=g_X(t)g_Y(t)$.
  Proof: $g_{X+Y}(t)=\mathrm{E}(e^{t(X+Y)})=\mathrm{E}(e^{tX}e^{tY})=\mathrm{E}(e^{tX})\cdot\mathrm{E}(e^{tY})=g_X(t)g_Y(t)$.

— Let X have MGF $g_X(t)$. Then, for any constant c, $g_{cX}(t) = g_X(ct)$.
Proof: $g_{cX}(t) = \mathrm{E}(e^{t(cX)}) = \mathrm{E}(e^{(ct)X}) = g_X(ct)$.

— Let $g(t)$ be the MGF of X_k^*. The X_k^* are i.i.d. and hence

$$g_{\frac{S_n^*}{\sqrt{n}}}(t) = g_n(t) = g_{\frac{X_1^*+\dots+X_n^*}{\sqrt{n}}}(t) = g_{X_1^*+\dots+X_n^*}(\frac{t}{\sqrt{n}}) = g\Big(\frac{t}{\sqrt{n}}\Big)^n$$

Date: 12/5

CLT Example: Roll a fair die 500 times. So, $n = 500$. Let S_{500} be the number of 1's. Use the CLT to estimate $\mathbb{P}(80 \leq S_{500} \leq 110)$. Then, let

$$X_k = \begin{cases} 1 & \text{if "1" on roll } k, \\ 0 & \text{otherwise on roll } k \end{cases}$$

for $k = 1,\dots,500$. Assume that X_k are independent. Thus, they are i.i.d. with means $\mu = \mathrm{E}(X_k) = \frac{1}{6}$ and variances $\mathrm{V}(X_k) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$. Therefore, $D(X_k) = \sigma = \frac{\sqrt{5}}{6}$. So, $n = 500$, $\mu = \frac{1}{6}$, $\sigma = \frac{\sqrt{5}}{6}$, and $S_{500} = \sum_{k=1}^{500} X_k$. So,

$$\begin{aligned} \mathbb{P}(80 \leq S_{500} \leq 110) &= \mathbb{P}(\frac{80 - n\mu}{\sqrt{n}\sigma} \leq \frac{S_n}{\sqrt{n}} \leq \frac{110 - n\mu}{\sqrt{n}\sigma}) \\ &= \mathbb{P}(-.894 \leq \frac{S_n}{\sqrt{n}} \leq .7155) \sim \Phi(7.155) - \Phi(-.894) \\ &= \mathrm{pnorm}(7.155, 0, 1) - \mathrm{pnorm}(-.894, 0, 1) = .8144. \end{aligned}$$

Note that S_{500} is actually binomial with parameters

and

. So, in R, $\mathbb{P}(80 \leq S_{500} \leq 110) = \mathrm{pnorm}(10, 500, 1/6) - \mathrm{pnorm}(80, 500, 1/6) = .628$.

Example The "lifespan" of a certain type of industrial drill bit is the number of hours of continuous use before the bit must be discarded. For one type of drillbit, the lifespan is exponentially distributed with mean $\frac{1}{\lambda} = 5.2$ hour - λ is the rate parameter. Let $T_1 + \dots + T_{100}$ be the lifespan of 100 such bits. We assume them to be i.i.d. exponentials. Let $S_{100} = T_1 + \dots + T_{100}$. Use the CLT to estimate $\mathbb{P}(550 \leq S_{100})$. Observe that $\mathrm{E}(S_{100}) = 520$. Here, $n = 100$, $\mu = \mathrm{E}(T_k) = 5.2$, and $\mathrm{V}(T_k) = 5.2^2$, so that $\sigma = 5.2$ also. Hence, $\mathbb{P}(550 \leq S_{100}) = \mathbb{P}(\frac{550-100\cdot 5.2}{10\cdot 5.2} \leq \frac{S_{100}}{10})$. The distribution is approximately standard normal for "large" n.

We see that $\mathbb{P}(\frac{550-100\cdot 5.2}{10\cdot 5.2} \leq \frac{S_{100}}{10}) = \mathbb{P}(\frac{30}{52} \leq \frac{S_{100}}{10}) \sim 1 - \Phi(\frac{15}{26}) = .282$. So, S_{100} is actually gamma (or Erlang) distributed with shape parameter $n = 100$ and rate parameter $\lambda = \frac{1}{5.2}$. So, $\mathbb{P}(550 \leq S_{100}) = \frac{\lambda^100}{99!} \int_{550}^{\infty} t^{99} e^{-\lambda t} dt = 1 - \mathrm{pgamma}(550, 100, \lambda)$. We subtract from 1 here, but we can also specify that *tail=false* in the *pgamma* function in *R*.

Brownian Motion Consider a "Brownian particle" moving random but continuously on the x-axis, with initial position $x = 0$. Let $t > 0$ be fixed. Let B_t be the random variable representing the position of the particle on the x-axis at time t. How is B_t distributed? We'll start with a "random walk" on $[0, t]$. Divide $[0, t]$ into n sub-intervals, each of duration $\triangle t = \frac{t}{n}$. Let $t_0 = 0$, $t_1 = \triangle t$, $t_2 = 2\triangle t$, ..., $t_n = n\triangle t$ be the endpoints of the sub-intervals. In each sub-interval, the particle moves a distance $\triangle x$, either to the left or to the right. Define the variables $\xi_1, \ldots, \xi_n$ by

$$\xi_k = \begin{cases} 1 & \text{if particle moves to the right in subinterval k,} \\ -1 & \text{if particle moves to the left in subinterval k.} \end{cases}$$

Let $x_k = \xi_k \triangle x$. So, the position of the particle at time t is $x_1 + \ldots + x_n = S_n$. Assume that the xi_k are i.i.d. with $\mathbb{P}(\xi_k = 1) = \mathbb{P}(\xi_k = -1) = .5$. The random walk is "symmetric". With additional stipulations, we take $\mathbb{P}(a \le B_t \le b) = \lim_{n\to\infty} \mathbb{P}(a \le S_n \le b)$.

$$\mathbb{P}(a \le S_n \le b) = \mathbb{P}(a \le (\xi_1 + \cdots + \xi_n)\triangle x \le b) = \mathbb{P}(\frac{a}{\triangle x} \le \xi_1 + \cdots + \xi_n \le \frac{b}{\triangle x}).$$

Observe that the ξ_k are i.i.d. with mean $\mathrm{E}(\xi_k) = 0 = \mu$ and variances $\mathrm{V}(\xi_k) = 1$ ($\sigma = 1$). Thus, $\xi_1 + \cdots + \xi_n = S_n$. So, the equation above becomes

$$\mathbb{P}(a \le S_n \le b) = \mathbb{P}(\frac{a}{\triangle x} \le S_n \le \frac{b}{\triangle x}) = \mathbb{P}(\frac{a}{\sqrt{n}\triangle x} \le \frac{S_n}{\sqrt{n}} \le \frac{b}{\sqrt{n}\triangle x}).$$

We know that $n\triangle = t$. The diffusion coefficient $D = \frac{\triangle x^2}{2\triangle t}$. Probabilists usually set $D = \frac{1}{2}$, so that $\frac{\triangle x^2}{\triangle t} = 1$, or $\frac{\triangle x}{\sqrt{\triangle t}} = 1$. So, the above equation becomes

$$\mathbb{P}(a \le S_n \le b) = \mathbb{P}(\frac{a}{\sqrt{n}\triangle x} \le \frac{S_n}{\sqrt{n}} \le \frac{b}{\sqrt{n}\triangle x}) = \mathbb{P}(\frac{a}{\sqrt{n}\triangle x \frac{\triangle x}{\sqrt{\triangle t}}} \le \frac{S_n}{\sqrt{n}} \le \frac{b}{\sqrt{n}\triangle x \frac{\triangle x}{\sqrt{\triangle t}}}) = \mathbb{P}(\frac{a}{\sqrt{t}} \le \frac{S_n}{\sqrt{n}} \le \frac{b}{\sqrt{t}}).$$

Therefore,

$$\begin{aligned} \mathbb{P}(a \le S_n \le b) &= \lim_{n\to\infty} \mathbb{P}(a \le S_n \le b) \\ &= \lim_{n\to\infty} \mathbb{P}(\frac{a}{\sqrt{t}} \le S_n \le \frac{b}{\sqrt{t}}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{t}}}^{\frac{b}{\sqrt{t}}} e^{-\frac{z^2}{2}} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{u^2}{2t}} \frac{du}{\sqrt{dt}} \\ &= \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-\frac{u^2}{2t}} du \end{aligned}$$

Hence, B_t is Gaussian with mean 0 and variance t.

A **stochastic process** is a collection of random variables. We usually study stochastic processes of the form $\left\{X_n\right\}_{n=0}^{\infty} = \left\{X_0, X_1, \ldots\right\}$ or $\left\{X_t\right\}_{t\geq 0}$. n and t usually represent discrete and continuous time. Thus, X_3 represents the "state" of the process at time $n = 3$ or $t = 3$. We showed that the position B_t of the Brownian particle was a Gaussian variable with mean 0 and variance t. $\left\{B_t\right\}_{t\geq 0}$ is a stochastic process with $\mathbb{P}(B_0 = 0) = 1$. Brownian motion is seen as important in statistical mechanics, quantum mechanics, stochastic differential equations, and finance.

We assume that Brownian Motion (the Wiener Process) has **independent** and **stationary** increments. Clearly, B_s and B_t are not independent. But, it is reasonable to assume that $B_t - B_s$ and $B_s - B_r$ are independent. This is the assumption of independent increments. For any times $t_1, \ldots, t_n$, $B_{t_1+s} - B_{t_1}, ldots, B_{t_n+s} - B_{t_n}$ have the same distribution. What is that distribution? Well, $B_{t_k+s} - B_{t_k}$ has the same distribution as $B_{0+s} - B_0 = B_s$, which is Gaussian with mean 0 and variance s. Thus, for any t, the increment $B_{t+s} - B_t$ is Gaussian, mean 0, variance s.

Let $t_0 = 0$, $x_0 = 0$, and then $0 < t_1 < t_2$. Let I_1 and I_2 be intervals on the x-axis. What is $\mathbb{P}(B_{t_1} \in I_1, B_{t_2} \in I_2)$?

First, the particle must reach a point $x_1 \in I_1$ at time t_1. From x_1, it must reach a point $x_2 \in I_2$, in time period of length $t_2 - t_1$. So, by independent and stationary increments,

$$\mathbb{P}(B_{t_1} \in I_1, B_{t_2} \in I_2) = \frac{1}{\sqrt{2\pi(t_1 - t_0)}} \frac{1}{\sqrt{2\pi(t_2 - t_1)}} \int_{I_2} \int_{I_1} e^{\frac{-(x_1 - x_0)^2}{2t_1}} e^{\frac{-(x_2 - x_1)^2}{2(t_2 - t_1)}} dx_1 dx_2$$

We also know there are different types of Brownian motion,

— Geometric Brownian Motion: $X_t = Ae^{B_t}$

— Integrated Brownian Motion: $X_t = \int_0^t B_s ds$

— Brownian Bridge: $X_t = B_t - \frac{t}{T} B_T$

How many punctual events occur in the period $(0, t]$? Let the number be N_t. $\mathbb{P}(N_0) = 0$. $\{N_t\}_{t\geq 0}$ is called the Poisson process. How is N_t distributed? Divide $(0, t]$ into n sub-intervals, each of length $\Delta \frac{t}{n}$. Assume that in each sub-interval, the probability $\mathbb{P}(N_t = k)$? Let λ be the mean occurrence rate per unit time. Then, $\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$. Customers arrive at a cafe according to a Poisson process with rate parameter $\lambda = .25/\text{min}$. What is the probability that between 5 and 8 customers arrive during the period 12:00 and 12:32?

$$
\begin{aligned}
\mathbb{P}(5 \leq N_{32} \leq 8) &= \sum_{k=5}^{8} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\
&= \sum_{k=5}^{8} \frac{(8^k}{k!} e^{-8} \\
&= .493
\end{aligned}
$$

$$
\int_0^\infty f(x) dx.
$$