Probability

STAT230

Jaiden Ratti

Prof. Erik Hintz

1239

Contents

1 Introduction to Probability

1.1 Definitions of Probability

Classical Definition of Probability

$\frac{\text{number of ways an event can occur}}{\text{total \# of possible outcomes}}$

*Provided things are equally likely.

Relative Frequency

The probability of an event is proportionate to number of times the event occurs in long repetitions. The probability of rolling a 2 is $\frac{1}{6}$ because after 6 million rolls, 2 will appear ~1 million times. (this is not very practical).

Subjective Definition

Persons belief on how likely something will happen (unclear since varies by person).

2 Mathematical Probability Models

2.1 Samples Spaces and Probability

Sample Spaces & Sets

Sample space S is a set of distinct outcomes of an experiment with the property that in a single trial, only one outcome will occur.

- Die: $S = \{1, 2, 3, 4, 5, 6\}$ or $S = \{\text{even, odd}\}$
- Number of coin flips until heads occurs: $S = \mathbb{N}^+$
- Waiting time in minutes until a sunny day. $S = [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$

Sample space is discrete if it is finite or countably infinite (one-to-one correspondence with $\mathbb{N}$). Otherwise non-discrete.

$\mathbb{N}$ is discrete (countably infinite).

Event is a subset of a sample space S.

A is an event if $A \subseteq S$ (A is a subset of S, A is contained in S).

- Die shows 6: $A = \{6\}$
- 20 or fewer tosses till heads. $A = \{1, \ldots, 20\}$
- $A = \{60, \ldots \infty\}$

Notation

1. Element of: $x \in A$ (x in A)
2. Union: $A \cup B$ $\{x | x \in A \text{ or } x \in B\}$
3. Intersection: $A \cap B$ $\{x | x \in A \text{ and } x \in B\}$
4. Complement: $A^{\complement} : \{x | x \in S, x \in A\} = A' = \overline{A}$
5. Empty: $\emptyset$
6. Disjoint $\implies A \cap B = \emptyset$

2 die rolled.

- $S = \{(x, y) : x, y \in \{(1, \ldots, 6)\}$
- $A = \{(x, y) : (x, y) \in S \wedge x + y = 7\}$
  - $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$

- $B^{\complement} = \{(x,y) : (x,y) \in S \wedge x+y < 4\}$
  - $B^{\complement} = \{(1,1),(1,2),(2,1)\}$
- $A \cap B^{\complement} = \emptyset$
- $A \cup B^{\complement} = \{A, B^{\complement}\}$

Probability

Let $\mathcal{S}$ denote the set of all events on a given sample S. Probability defined on $\mathcal{S}$ is defined as

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

1. Scale: $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. Additivity (infinite): $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

e.g. $A_1 = \{1\}, A_2 = \{2\}, A_3 = A_4 = \emptyset$

$P(1 \text{ or } 2) = P(A_1) + P(A_2)$

S (discrete) and $A \subset S$ is an event.

A is indivisible $\iff$ A is a simple point, otherwise compound.

e.g.

$S = \{1,2,3,4,5,6\}$
$A = \{1\} \leftarrow$ simple
$B = \{2,4,6\} \leftarrow$ compound.

Assign so that

1. $0 \leq P(a_i) \leq 1$
2. $\sum_i P(a_i) = 1$

$$\{P(a_i) : i = 1,2,3,\ldots\} \quad = \text{probability distribution}$$

$S = \{a_1, a_2, \ldots\}$ discrete, $A \subset S$ is an event.

$$P(A) = \sum_{a_i \in A} P(a_i)$$

A = Number is odd

$P(i) = \frac{1}{6}$ for $1,2,3,4,5,6$

$$\begin{aligned} P(A) &= P(1) + P(3) + P(5) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{3}{6} \\ &= \frac{1}{2} \end{aligned}$$

Sample space S is equally likely if probability of every outcome is the same.

$|A| = \#$ of elements in set.

$$1 = P(S) = \sum_{i=1}^{|S|} P(a_i) = P(a_i)|S|$$
$$P(a_i) = \frac{1}{|S|}$$

$$P(A) = \sum_{i:a_i \in A} P(a_i) = \frac{|A|}{|S|}$$

3 Probability and Counting Techniques

A and B are disjoint $A \cap B = \emptyset$, then

$$|A \cup B| = |A| + |B|$$

3.1 Addition and Multiplication Rules

Addition Rule

Sum larger than 8.

$B =$ Sum of Die larger than 8

$A_9 =$ Sum 9, $A_{10} =$ Sum 10, $A_{11} =$ Sum 11, $A_{12} =$ Sum 12.

$B = A_9 \cup A_{10} \cup A_{11} \cup A_{12}$ all A_j are disjoint.

$$\begin{aligned}|B| &= |A_9| + |A_{10}| + |A_{11}| + |A_{12}| \\ &= 4 + 3 + 2 + 1 \\ &= 10\end{aligned}$$

Thus,

$$|A| = |\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i|$$

Multiplication Rules

Ordered $k-$tuple, $(a_1, a_2, \ldots, a_k)$. n_1 choices for a_1, n_k choices for a_k etc.

$|A| = n_1 n_1 \ldots n_k = \prod_{i=1}^{k} n_i$

If there are p ways to do task 1, and q ways to do task 2, there are $p \times q$ ways to do tasks 1 and 2.

With replacement: When the object is selected, put it back after.

Without replacement: Once an object is picked, it stays out of the pool.

3.2 Counting Arrangements or Permutations

Factorials

n distinct objects, n factorial.

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$$

$0! = 1$ and $n! = n \cdot (n-1)!$

10 people standing next to each other, $10! = 3628800$ arrangements.

Out of 5 students, must choose 1 president and 1 secretary. $5 \times 4 = 20$.

Note, Stirling's Approximation.

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Permutations

n distinct objects, permutation of size k is an ordered subset of k individuals. (without replacement)

$$n^{(k)} = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

With replacement,

$$n^k = n(n)\dots(n) \quad k \text{ times}$$

n to the k factors.

Our example of president and secretary above can thus be seen as $5^{(2)} = 5 \times 4$.

3.3 Counting Subsets or Combinations

Question: If we have 5 members on a club, how can we select 2 to serve on a committee?

Here, order does not matter. Unlike a permutation, this is a combination.

In general, if we have k objects, there are $k!$ ways to reorder such objects. We can therefore get the combination count by dividing the permutation count $n^{(k)}$ by the number of ways of ordering the objects $k!$.

Definition

Given n distinct objects, a combination of size k is an unordered subset of k of the objects (without replacement). The number of combinations of size k taken from n objects is denoted $\binom{n}{k}$, which reads "n choose k, and

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

For our previous example we have

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{(n-k)!k!} \\ &= \frac{5!}{3!2!} \\ &= 10\end{aligned}$$

Properties of $\binom{n}{k}$

1. $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \geq 1$
2. $binomnk = \frac{n!}{k!(n-k)!} = \frac{n^{(k)}}{k!}$
3. $binomnk = \frac{n!}{(n-k)!k!} = \frac{n!}{(n-(n-k))!(n-k)!} = \binom{n}{n-k}$
4. If we define $0! = 1$, then $\binom{n}{0} = \binom{n}{n} = 1$
5. $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
6. Binomial Theorem $(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n$

Example

Group of 5 women and 7 men, a committee of 2 women and 3 men is formed at random. 2 of them men dislike each other. What is the probability they don't serve together?

First, get the size of the sample space $|S|$.

Pick 2 women from 5 women, $\binom{5}{2}$

Pick 3 men from 7 men, $\binom{7}{3}$

Thus $|S| = \binom{5}{2}\binom{7}{3}$

Consider the event $A = \{1$ and 2 do not serve in the committee together $\}$

Consider $A^{\complement} = \{$ 1 and 2 in the committee together$\}$

The size of the sample space $|A^{\complement}|$ is given by:

Pick 2 women from 5 women, $\binom{5}{2}$.

1 and 2 are in the committee already, we need to pick 1 man from the 5 left, $\binom{5}{1}$.

Thus,

$$P(A) = 1 - P(A^{\complement}) = 1 - \frac{|A^{\complement}|}{|S|} = 1 - \frac{\binom{5}{2}\binom{5}{1}}{\binom{5}{2}\binom{7}{3}} = \frac{\binom{5}{2}[\binom{7}{3} - \binom{5}{1}]}{\binom{5}{2}\binom{7}{3}}$$

3.4 Arrangements when Symbols are Repeated

Definition

Consider n objects of k types. Suppose n_1 objects of type 1, n_2 objects of type 2, and n_k objects of type k. Thus there are,

$$\frac{n!}{n_1!n_2!\dots n_k!}$$

distinguishable arrangements of n objects. This is the multinomial coefficient.

Letters of "SLEEVELESS" are arranged at random. What is the probability the word begins and ends with "S"?

Sample space, $|S| = \frac{10!}{4!3!2!1!} = 12600$

Event, $|A| = \frac{8!}{1!4!2!1!} = \frac{40320}{48} = 840$

Thus $P(A) = \frac{840}{12600} = \frac{1}{15}$

3.5 Useful Series and Sums

Finite Geometric Series:

$$\sum_{i=0}^{n-1} t^i = 1 + t + t^2 + \dots + t^{n-1} = \frac{1 - t^n}{1 - t} \text{ if } t \neq 1$$

Infinite Geometric series if $|t| < 1$:

$$\sum_{x=0}^{\infty} t^x = 1 + t + t^2 + \dots = \frac{1}{1 - t}$$

Binomial Theorem (i), if n is a positive integer and t is any real number:

$$(1 + t)^n = 1 + \binom{n}{1}t^2 + \binom{n}{2}t^2 + \dots + \binom{n}{n}t^n = \sum_{x=0}^{n}\binom{n}{x}t^x$$

Binomial Theorem (ii), if n is not a positive integer but $|t| < 1$:

$$(1 + t)^n = \sum_{x=0}^{\infty}\binom{n}{x}t^x$$

Multinomial Theorem:

$$(t_1 + t_2 + \ldots t_k)^n = \sum \frac{n!}{x_1!x_2!\ldots x_k!} t_1^{x_1} t_2^{x_2} \ldots t_k^{x_k}$$

where the summation is over all non-negative integers $x_1, x_2, \ldots, x_k$ such that $\sum_{i=1}^{k} x_i = n$ where n is a positive integer.

Hypergeometric Identity:

$$\binom{a+b}{n} = \sum_{x=0}^{\infty} \binom{a}{x} \binom{b}{n-x}$$

Exponential Series:

$$e^t = \frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \ldots = \sum_{x=0}^{\infty} \frac{t^n}{n!}$$

for all t in the real numbers.

A related identity:

$$e^t = \lim_{n\to\infty} (1 + \frac{t}{n})^n$$

Series involving integers:

$$1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}$$
$$1^2 + 2^2 + 3^3 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$
$$1^3 + 2^3 + 3^3 + \ldots + n^3 = [\frac{n(n+1)}{2}]^2$$

4 Probability Rules and Conditional Probability

4.1 General Methods

Rules: $P(S) = \sum_{\text{all}i} P(a_i) = 1$

For any event $A, 0 \leq P(A) \leq 1$.

If A and B are two events with $A \subseteq B$, then $P(A) \leq P(B)$.

Fundamental Laws of Set Algebra

Commutativity

$$A \cup B = B \cup A \quad A \cap B = B \cap A$$

Associativity

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

Distributivity

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

De Morgan's Law

$$(\overline{A \cup B}) = \overline{A} \cap \overline{B}$$

The complement of a union is the intersection of the complements.

$$(\overline{A \cap B}) = \overline{A} \cup \overline{B}$$

The complement of an intersection is the union of the complements.

Applied for n events

$$\overline{(A_1 \cup A_2 \cup \ldots \cup A_n)} = \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}$$

4.2 Rules for Unions of Events

For arbitrary events A, B, C,

If A and B are disjoint ($A \cap B = \emptyset$) then,

$$P(A \cup B) = P(A) + O(B)$$

What if $A \cap B \neq \emptyset$?

Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The probability of the complement event is,

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

If A, B, and C are disjoint (mutually exclusive), then

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

Otherwise,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

In general,

$$P(\bigcup_{i=1}^{n} A_1) = \sum_i P(A) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) - \sum_{i<j<k<1} P(A_i A_j A_k A_1) + \ldots$$

Note, $P(A \cap B)$ is often written as $P(AB)$

4.3 Intersection of Events and Independence

Independent Events

Rolling die twice is an example of independent events. The outcome of the first doesn't affect the second.

Events are independent if and only if

$$P(A \cap B) = P(A)P(B)$$

$A =$"roll 6 on first" $B =$"roll 6 on second"

$$P(A \cap B) = P(\text{"both 6"}) = \frac{1}{36} = P(A)P(B)$$

Not independent. $C =$ First roll is 6, $D =$ first roll is even.

$P(C \cap D) = \frac{1}{6}$

$$P(C)P(D) = \frac{1}{12} \neq P(C \cap D)$$

A common misconception is that if A and B are mutually exclusive, then A and B are independent.

If A and B are independent, A and $\overline{B}$ are independent. Law of total probability.

$$\begin{aligned} &P(A \cap \overline{B}) \\ &\quad = P(A) - P(A \cap B) \\ &\quad = P(A) - P(A)P(B) \\ &\quad = P(A)[1 - P(B)] \\ &\quad = P(A)P(\overline{B}) \end{aligned}$$

We can also see,

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

4.4 Conditional Probability

$P(A|B)$ represents the probability that event A occurs, when we know that B occurs. This is the conditional probability of A given B.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0$$

If A and B are independent, then

$$\begin{aligned} &P(A \cap B) = P(A)P(B) \\ &P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A) \quad \text{provided } P(B) > 0 \end{aligned}$$

A and B are independent events if and only if either of the following statements is true

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$

Example

A = The sum of two die is 10
B = The first die is 6

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{36}}{\frac{3}{36}} = \frac{1}{3}$$

If $P(A) = 0$ or $P(B) = 0$, then A and B are independent.

Properties

- $P(B|B) = 1$
- $0 \leq P(A|B) \leq 1$
- If $A \subseteq C, P(A|B) \leq P(C|B)$
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$
- If A_1 and A_2 are disjoint: $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$
- $P(\overline{A}|B) = 1 - P(A|B)$

4.5 Product Rules, Law of Total Probability and Bayes' Theorem

Product Rules

For events A and B,

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

That means if we know $P(A|B)$ and $P(B)$; or $P(B|A)$ and $P(A)$, we can find $P(A \cap B)$.

More events:

- $P(A \cap B) = P(A)P(B|A)$
- $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$
- $P(A \cap B \cap C \cap D) = P(A)P(B|A)P(C|A \cap B)P(D|A \cap B \cap C)$

Law of Total Probability

Let $A_1, A_2, \ldots, A_k$ be a partition of the sample space S into disjoint (mutually exclusive) events, that is

$$A_1 \cup A_2 \cup \ldots \cup A_k = S \quad \text{and} \quad A_i \cap A_j = \emptyset \text{ if } i \neq j$$

Let B be an arbitrary event in S. Then

$$\begin{aligned} P(B) &= P(BA_1) + P(BA_2) + \ldots + P(BA_k) \\ &= \sum_{i=1}^{k} P(B|A_i)P(A_i) \end{aligned}$$

A common way in which this is used is that

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

since A and $\overline{A}$ partition S.

Bayes Theorem

Suppose A and B are events defined on a sample space S. Suppose also that $P(B) > 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|\overline{A})P(\overline{A}) + P(B|A)P(A)}$$

Proof:

$$\begin{aligned} \frac{P(B|A)P(A)}{P(B|\overline{A})P(\overline{A}) + P(B|A)P(A)} &= \frac{P(AB)}{P(\overline{A}B) + P(AB)} \\ &= \frac{P(AB)}{P(B)} \\ &= P(A|B) \end{aligned}$$

5 Discrete Random Variables

5.1 Random Variables and Probability Functions

Definition

A random variable is a function that assigns a real number to each point in a sample space S. Often a random variable is abbreviated with RV or rv.

Definition

The values that a random variable can take on are called the range of the random variable. We often denote the range of a random variable X by $X(S)$.

Example

If we roll a 6-sided dice, our sample space is $S = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$ and if we define $X =$ sum of die rolls, the range is $\{2,3,\ldots,11,12\}$

Definitions

We say that a random variable is discrete if it takes values in a countable set (finite or countably infinite).

We say that a random variable is continuous if it takes values in some interval of real numbers, e.g. $[0,1],(0,\infty),\mathbb{R}$.

Important: Don't forget that a random variable can be infinite and discrete.

Definition

Let X be a discrete random variable with range A. The probability function (p.f.) of X is the function

$$f(x) = P(X = x), \quad \text{defined for all } x \in A$$

The set of pairs $\{(x, f(x)) : x \in A\}$ is called the probability distribution of X.

A probability function has two important properties:

1. $0 \leq f(x) \leq 1$ for all $x \in A$
2. $\sum_{\text{all } x\in A} f(x) = 1$

Example

A random variable X has a range $A = \{0,1,2\}$ with $f(0) = 0.19, f(1) = 0.2k^2, f(2) = 0.8k^2$, what values of k makes $f(x)$ a probability function.

From our rules,

$$\begin{aligned} &0.19 + 0.2k^2 + 0.8k^2 = 0.19 + k^2 = 1 \\ \implies &k^2 = 0.81 \\ &k = \pm 0.9 \end{aligned}$$

Definition The Cumulative Distribution Function (CDF) of a random variable X is

$$F(x) = P(X \leq x), \text{ define for all } x \in \mathbb{R}$$

We use the shorthand that $X \sim F$ if X has CDF F. Here,

$$P(X \leq x) = P(\{a \in S : X(a) \leq x\})$$

where $\{X \leq a\}$ is the event that contains all outcomes with $X(a) \leq x$.

In general, for any $x \in \mathbb{R}$

$$F(x) = P(X \leq x) = \sum_{u \leq x} P(X = u) = \sum_{u\leq x} f(u)$$

The CDF satisfies that

1. $0 \leq F(x) \leq 1$
2. $F(x) \leq F(y)$ for $x < y$ (and we say $F(x)$ is a non-decreasing function of x)
3. $\lim_{x\to-\infty} F(x) = 0$, and $\lim_{x\to\infty} F(x) = 1$
4. $\lim_{x\to a^+} F(x) = F(a)$ (right continuous)

If X takes value on $a_1 < a_2 < \ldots < a_n < \ldots$, we can get probability function from CDF:

$$f(a_i) = F(a_i) - F(a_{i-1})$$

In general, we have

$$P(a < X \leq b) = F(b) - F(a)$$

Note, we often use

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(x \leq 0)$$

Definition

Two random variables X and Y are said to have the same distribution if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$. We denote this by

$$X \sim Y$$

Note, X and Y having the same distribution does not mean $X = Y$.

5.2 Discrete Uniform Distribution

Setup

Suppose the range of X is $\{a, a+1, \ldots, b\}$ where a and b are integers and suppose all values are equally probable. Then X has a Discrete Uniform Distribution on the set $\{a, a+1, \ldots, b\}$. The variables a and b are called the parameters of the distribution.

Illustrations

If X is the number obtained when a die is rolled, then X has a discrete Uniform distribution with $a = 1$ and $b = 6$.

Probability Function

There are $b - a + 1$ values in the set $\{a, a+1, \ldots, b\}$ so the probability of each value must be $\frac{1}{b-a+1}$ in order for $\sum_{x=a}^{b} f(x) = 1$. Therefore

$$f(x) = P(X = x) = \begin{cases} \frac{1}{b-a+1} & \text{for } x = a, a+1, \ldots, b \\ 0 & \text{otherwise} \end{cases}$$

Example

Suppose a fair die is thrown once and let X be the number on the face. Find the cumulative distribution function of X.

Solution

This is an example of a Discrete Uniform distribution on the set $\{1, 2, 3, 4, 5, 6\}$ having $a = 1, b = 6$ and probability function

$$f(x) = P(X = x) = \begin{cases} \frac{1}{6} & \text{for } x = 1, 2, \ldots, 6 \\ 0 & \text{otherwise} \end{cases}$$

The cumulative distribution function is $F(x) = P(X \leq x)$,

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{[x]}{6} & \text{for } 1 \leq x < 6 \\ 1 & \text{for} x \geq 6 \end{cases}$$

where $[x]$ is the largest integer less than or equal to x.

Example

Let X be the largest number when a die is rolled 3 times. First find the cumulative distribution function, and then find the probability function of X.

Solution

This is another example of a distribution constructed from the Discrete Uniform.

$$S = \{(1,1,1),(1,1,2),\ldots,(6,6,6)\}$$

The probability that the largest number is less than or equal to x is,

$$F(x) = \frac{x^3}{6^3}$$

for $x = 1,2,3,4,5,6$. Here is the CDF for all real values of x:

$$F(x) = P(X \leq x) = \begin{cases} \frac{[x]^3}{216} & \text{for } 1 \leq x < 6 \\ 0 & \text{for } x < 1 \\ 1 & \text{for } x \geq 6 \end{cases}$$

To find the p.f. we may use the fact that $x \in \{1,2,3,4,5,6\}$ we have $P(X = x) = P(X \leq x) - P(X < x)$ so that

$$\begin{aligned} f(x) &= F(x) - F(x-1) \\ &= \frac{x^3 - (x-1)^3}{216} \\ &= \frac{[x-(x-1)][x^2 + x(x-1) + (x-1)^2]}{216} \\ &= \frac{3x^2 - 3x + 1}{216} \quad \text{for } x = 1,2,3,4,5,6 \end{aligned}$$

5.3 Hypergeometric Distribution

Setup

Consider a population of N objects, of which r are considered "successes" (S) and the remaining $N - r$ are considered "failures" (F).

Suppose that a subset of size n is chosen at random from the population without replacement

We say that the random variable X has a hypergeometric distribution if X denotes the number of successes in the subset (shorthand: $X \sim hyp(N,r,n)$).

- N: Number of objective
- r: Number of successes
- n: Number of draws

Illustrations

Drawing 2 balls without replacement from a bag with 3 blue balls and 4 red balls. Let X denote the number of blue balls drawn. Then

$$X \sim hyp(7,3,2)$$

Drawing 5 cards from a deck of cards. Let X denote the number of aces. Then

$$X \sim hyp(52,4,5)$$

Probability Function

Suppose $X \sim hyp(N,r,n)$

$$f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \quad max\{0, n-(N-r)\} \leq x \leq min\{r,n\}$$

$x \leq min\{r,n\}$

- $x \leq n$: the number of successes drawn cannot exceed the number drawn
- $x \leq r$: we have at most r success

$x \geq max\{0, n-(N-r)\}$

- $x \geq 0$ obviously
- $x \geq n-(N-r)$: if n exceeds the number of failures $N-r$, we have at least $n-(N-r)$ of successes.

We can verify the probability function of the hypergeometric distribution sums to 1.

$$\sum_{\text{all } x} f_X(x) = \sum_{\text{all } x} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{\text{all } x} \binom{r}{x}\binom{N-r}{n-x} = \frac{\binom{r+N-r}{n}}{\binom{N}{n}} = 1$$

Example

Consider drawing a 5-card hand at random from a standard 52-card deck. What is the probability that the hand contains at least 3 K's?

X : the number of K in hand.
Success type: 13 K's Failure type: Not K cards
$N = 52 \quad r = 13 \quad n = 5$
$X \sim hyp(52, 13, 5)$

$$\begin{aligned} P(X \geq 3) &= P(X=3) + P(X=4) + P(X=5) \\ &= \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} + \frac{\binom{13}{4}\binom{39}{1}}{\binom{52}{5}} + \frac{\binom{13}{5}\binom{39}{0}}{\binom{52}{5}} \\ &= 0.00175 \end{aligned}$$

5.4 Binomial Distribution

Definition

A Bernoulli trial with probability of success p is an experiment that results in either a success or failure, and the probability of success is p.

Setup

Consider an experiment in which n Bernoulli trials are independently formed, each with probability of success p. X denotes the number of successes from n trials.

$$X \sim Binomial(n, p)$$

Illustrations

Flip a coin independently 20 times, let X denote the number of heads observed. Then

$$X \sim Binomial(20, 0.5)$$

Drawing 2 balls with replacement from a bag with 3 blue balls and 4 red balls. Let X denote the number of blue balls drawn.

$$X \sim Binomial(2, \frac{3}{7})$$

Assumptions:

1. Trials are independent
2. The probability of success, p, is the same in each Bernoulli trial

Probability Function

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \ldots, n$$

Proof that $\sum_{\text{all} x} f(x) = 1$ for $0 < p < 1$:

$$\begin{aligned}
\sum_{x=0}^{n} f(x) &= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \\
&= (1-p)^n \sum_{x=0}^{n} \binom{n}{x} \left(\frac{p}{1-p}\right)^x \\
&= (1-p)^n \left(1 + \frac{p}{1-p}\right)^n \\
&= (1-p)^n \left(\frac{1-p+p}{1-p}\right)^n \\
&= 1^n = 1
\end{aligned}$$

If N is very large, and we keep the number of success $r = pN$ where $p \in (0,1)$. We choose a relatively small n without replacement from N.

Let $X \sim hyp(N, r, n)$ and $Y \sim Binomial(n, p)$. Then

$$P(X = k) \approx P(Y = k)$$

The approximation is good if N and r are large compared to n.

5.5 Negative Binomial Distribution

Setup

Consider an experiment in which Bernoulli trials are independently performed, each with probability of success p, until exactly k successes are observed. Then if X denotes the number of failures before observing k successes, we say that X is Negative Binomial with parameters k and p.

$$X \sim NB(k, p)$$

Let Y be the number of trials until exactly k successes are observed. We have $Y = X + k$.

Illustrations

Flip a coin until 5 heads are observed, and let X denote the number of tails observed. Then

$$X \sim NB(5, 0.5)$$

Probability Function

$$f(x) = \binom{x+k-1}{k-1} p^k (1-p)^k, \quad x = 0, 1, 2, \ldots$$

Proof it is valid

$$\begin{aligned}
\sum_{x=0}^{\infty} f(x) &= \sum_{x=0}^{\infty} \binom{-k}{x} (-1)^x p^k (1-p)^k \\
&= p^k \sum_{x=0}^{\infty} \binom{-k}{x} [(-1)(1-p)]^x \\
&= p^k [1 + (-1)(1-p)]^{-k} \quad \text{if } 0 < p < 1 \\
&= p^k p^{-k} \\
&= 1
\end{aligned}$$

5.6 Geometric Distribution

Setup

Geometric distribution is a special case of the Negative Binomial where we stop after the first success $k = 1$.

$$X \sim Geo(p)$$

Probability Function

$$f(x) = (1-p)^x p, \quad x = 0, 1, 2, 3, \ldots$$

Geometric Distribution: CDF
For an integer x, the CDF of X is

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{t=0}^{x} (1-p)^x p \\ &= p \sum_{t=0}^{x} (1-p)^x \\ &= p \frac{1-(1-p)^{x+1}}{1-(1-p)} \\ &= 1-(1-p)^{x+1} \end{aligned}$$

So $F(x) = 1 - (1-p)^{[x]+1}$ for $x \geq 0$, and 0 otherwise.

Example

The number of times I have to roll 2 dice before I get snake eyes.

$$X \sim Geo(\frac{1}{36})$$

5.7 Poisson Distribution (from Binomial)

As $n \to \infty$ and $p \to 0$, the binomial function approaches

$$f(x) \approx e^{-\mu} \frac{\mu^x}{x!}$$

for $\mu = np$

Setup

Let $\mu = np$. Then if n is large, and p is close to zero,

$$\binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\mu} \frac{\mu^x}{x!}$$

Probability Function

A random variable X has a Poisson distribution with parameter μ ($X \sim Poission(\mu)$) if

$$f(x) = e^{-\mu} \frac{\mu^x}{x!}, \quad x = 0, 1, 2, 3, \ldots$$

5.8 Poisson Distribution from Poisson Process

A process satisfying the following 3 conditions is called a Poisson process.

1. Independence: The number of occurrences in non-overlapping intervals are independent. For $t > s, (X_t - X_s)$ and X_s are independent.
2. Individuality or Singularity: Events occur singly, not in clusters. $P(2 \text{ or more events in } (t, t+\Delta t)) = o(\Delta t)$ as $\Delta t \to 0$.
3. Homogeneity or Uniformity: Events occur at a uniform rate λ (in events per unit of time). $P(\text{one event in } (t, t+\Delta t)) = \lambda\Delta_t + o(\Delta t)$ as $\Delta t \to 0$.

Little o

A function $g(\Delta t)$ is $o(\Delta t)$ as $\Delta t \to 0$ if

$$\lim_{\Delta t \to 0} \frac{g(\Delta t)}{\Delta t} = 0$$

Poisson Distribution

If X_t is a Poisson counting process with a rate of λ per unit. Then,

$$X_t \sim Poisson(\lambda t)$$

and

$$f_t(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad \text{for } x = 0, 1, 2, \ldots$$

Examples

Suppose that trucks arrive at a receiving dock with an average arrival rate of 3 per hour. What is the probability exactly 5 trucks will arrive in a two-hour period?

We have $\lambda = 3, t = 2$.

$$P(X_t = 5) = \frac{e^{-(3\times 2)}(3\times 2)^5}{5!}$$

5.9 Combining Models

Website hits for a given website occur according to a Poisson process with a rate of 100 hits per minute. A second is a break if there are no hits in that second.

What is the probability of a break in any given second?

$P(A) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{\frac{-100}{60}} = 0.189$

Compute the probability of observing exactly 10 breaks in 60 consecutive seconds.

$p = P(A) = 0.189$, for $X \sim Binomial(60, p)$

$$P(X = 10 = \binom{60}{10} p^{10}(1-p)^{60-10} = 0.124$$

Compute the probability that one must wait for 30 seconds to get 2 breaks.

Let Y be the seconds we wait for 2 breaks. We want $P(Y = 30)$.

Negative Binomial Distribution: $Y - 1 \sim NB(2, p)$

$$P(Y = 30) = P(Y - 1 = 29) = \binom{29}{1} p^2(1-p)^{30-2} = 0.00295$$

6 Expected Value and Variance

6.1 Summarizing Data on Random Variables

Median: A value such that half the results are below it and half above it, when the results are arranged in numerical order.

Mode: Value which occurs most often.

Mean: The mean of n outcomes $x_1, \ldots, x_n$ for a random variable X is $\sum_{i=1}^{n} \frac{x_i}{n}$

6.2 Expectation of a Random Variable

Let X be a discrete random variable with $range(X) = A$ and probability function $f(x)$. The expected value of X is given by

$$E(X) = \sum_{x \in A} x f(x)$$

Suppose X denotes the outcome of one fair six sided die roll. Compute $E(X)$.

We know $X \sim \mathcal{U}(1, 6)$: that is, $f(x) = \frac{1}{6} \quad x = 1, 2, \ldots, 6$.

And so,

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \ldots + 6 \times \frac{1}{6} = 3.5$$

Possible Range

Suppose X is a random variable satisfying $a \leq X \leq b$ for all possible values of X. We have,

$$a \leq E(x) \leq b$$
$$a = a \sum_{x \in A} f(x) = \sum_{x \in A} a f(x) \leq \sum_{x \in A} x f(x) \leq \sum_{x \in A} b f(x) = b \sum_{x \in A} f(x) = b$$

Expectation of $g(X)$

Let X be a discrete random variable with $range(X) = A$ and probability function $f(x)$. The expected value of some function $g(X)$ of X is given by

$$E[g(X)] = \sum_{x \in A} g(x) f(x)$$

Proof

$$\begin{aligned} E[g(X)] &= \sum y \in B y F_Y(y) \\ &= \sum_{y \in B} y \sum_{x \in D_y} f(x) \\ &= \sum_{y \in B} \sum_{x in D_y} g(x) f(x) \\ &= \sum_{x \in A} g(x) f(x) \end{aligned}$$

Linearity Properties of Expectation

For constants a and b,

$$E[a g(X) + b] = a E[g(X)] + b$$

Proof

$$
\begin{aligned}
E[ag(X)+b] &= \sum_{\text{all } x}[ag(x)+b]f(x) \\
&= \sum_{\text{all } x}[ag(x)f(x)+bf(x)] \\
&= a\sum_{\text{all } x} g(x)f(x) + b\sum_{\text{all } x} f(x) \\
&= aE[g(X)]+b
\end{aligned}
$$

Thus it is also easy to show

$$E[ag_1(X)+bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$$

The expectation of a sum is the sum of the expectations.

6.3 Means and Variances of Distributions

Expected value of a Binomial random variable

Let $X \sim Binomial(n,p)$. Find $E(X)$.

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_{x=0}^{n} x\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)![(n-1)-(x-1)]!}pp^{x-1}(1-p)^{(n-1)-(x-1)} \\
&= np(1-p)^{n-1}\sum_{x=1}^{n}\binom{n-1}{x-1}\left(\frac{p}{1-p}\right)^{x-1}
\end{aligned}
$$

Let $y = x-1$

$$
\begin{aligned}
&= np(1-p)^{n-1}\sum_{y=0}^{n-1}\binom{n-1}{y}\left(\frac{p}{1-p}\right)^{y} \\
&= np(1-p)^{n-1}\left(1+\frac{p}{1-p}\right)^{n-1} \\
&= np(1-p)^{n-1}\frac{(1-p+p)^{n-1}}{(1-p)^{n-1}} \\
&= np
\end{aligned}
$$

Example

If we toss a coin $n = 100$ times, with probability $p = 0.5$ of a head, we'd expect

$$E[X] = 100 \times 0.5 = 50 \text{ heads}$$

Hypergeometric Distribution

If $X \sim hyp(N, r, n)$, then $E[X] = n\frac{r}{N}$

Geometric Distribution

If $X \sim Geo(p)$, then $E[X] = \frac{1-p}{p}$

Negative Binomial

If $X \sim NB(k, p)$, then $E[X] = \frac{(1-p)}{p}k$

Poisson Distribution

If $X \sim Poi(\mu)$, then $E[X] = \mu$

Variances of Distribution

Using the expected value is one way of predicting the value of a random variable. But we might want to know "how likely will observed data be exactly (or close to) the expected value?"

One may wonder how much a random variable tends to deviate from its mean. Suppose $E[X] = \mu$

Expected deviation:

$$E[(X - \mu)] = E[X] - \mu = 0$$

Expected absolute deviation:

$$E[|X - \mu|] = \sum_{x \in A} |X - \mu| f(x)$$

Expected squared deviation:

$$E[(X - \mu)^2] = \sum_{x \in A} (X - \mu)^2 f(x)$$

Variance

The variance of a random variable X is denoted $Var(X)$, and is defined by

$$\sigma^2 = Var(X) = E[(X - E[X])^2]$$

or sometimes,

$$Var(X) = E(X^2) - [E(X)]^2$$

We derive this by

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2E[X]X + (E[X])^2] \\ &= E[X^2] - 2E[X]E[X] + (E[X])^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

Suppose X satisfies $P(X = 0) = \frac{1}{2} = P(X = 1)$. What is $Var(X)$?

$E[X] = 0.5 \quad \text{and} \quad E[X^2] = 0.5$
$Var(X) = E[X^2] - E[X]^2 = 0.25$

For all random variables X, $Var(X) \geq 0$. $Var(X) = 0$ if and only if $P(X = E[X]) = 1$. $E[X^2] \geq (E[X])^2$

Standard Deviation

The standard deviation of a random variable X is denoted $SD(X)$, and defined by

$$\sigma = SD(X) = \sqrt{Var(X)}$$

Variance of a Linear Transformation

If a and b are constants, and $Y = aX + b$, then

$$Var(Y) = a^2 Var(X)$$

the constant b does not affect anything. $SD(Y) = aSD(X)$

Variances of distributions

$$X \sim Binomial(n,p) \to Var(X) = np(1-p)$$
$$X \sim hyp(N,r,n) \to Var(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$
$$X \sim Geo(p) \to Var(X) = \frac{(1-p)}{p^2}$$
$$X \sim NB(k,p) \to Var(X) = \frac{(1-p)}{p^2}k$$
$$X \sim Poi(\mu) \to Var(X) = \mu$$
$$X \sim \mathcal{U}(a,b) \to Var(X) = \frac{(b-a+1)^2-1}{12}$$

7 Continuous Random Variables

7.1 General Terminology and Notation

Definition

A random variable X is said to be continuous if its range is an interval $(a,b) \subset \mathbb{R}$. X is continuous if it can take any value between a and b.

We can't use $f(x) = P(X = x)$ with continuous distributions because $P(X = x) = 0$.

Suppose X is a crv with range $[0,4]$. We now use integrals instead of sums.

$$\int_0^4 f(x)dx = 1$$

Probability Density Function

We say that a continuous random variable X has probability density function $f(x)$ if

$$f(x) \geq 0 \quad \forall x \in \mathbb{R}$$
$$\int_{-\infty}^{\infty} f(x)dx = 1$$
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

For a four number spinner, if X is where the spinner stops, we can define our pdf as:

$$f(x) = \begin{cases} 0.25 & \text{if } 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

We can see that this satisfies $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_0^4 0.25dx + \int_4^{+\infty} 0dx = 4 \times 0.25 = 1$$

Spinner Example

$P(1 \leq X \leq 2)$

$$\begin{aligned} P(1 \leq X \leq 2) &= \int_1^2 f(x)dx \\ &= \int_1^2 0.25dx \\ &= 0.25x|_1^2 \\ &= 0.25 \cdot 2 - 0.25 \cdot 1 = 0.25 \end{aligned}$$

<u>Definition</u>

The support of a *pdf* $f(x)$ is defined as

$$supp(f) = \{x \in \mathbb{R} : f(x) \neq 0\}$$

<u>Example</u>

Suppose that X is a continuous random variable with probability density function

$$f(x) = \begin{cases} cx(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute c so that this is a valid pdf.

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx = \int_0^1 cx(1-x)dx = 1 \\ \implies c\int_0^1 x - x^2 dx = 1 \\ \implies x\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1 \\ \implies c\left[\frac{1}{2} - \frac{1}{3}\right] = 1 \\ \implies c = 6 \end{aligned}$$

Compute $P(X > \frac{1}{2})$

$$\begin{aligned} P(X > \frac{1}{2}) &= \int_{1/2}^1 6x(1-x)dx \\ &= 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{1/2}^1 \\ &= 6\left(\left[\frac{1}{2} - \frac{1}{3}\right] - \left[\frac{1/4}{2} - \frac{1/8}{3}\right]\right) \\ &= 6\left[\frac{1}{6} - \frac{1}{12}\right] \\ &= \frac{1}{2} \end{aligned}$$

Note

$$P(X = a) \neq f(x)$$
$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < x \leq b)$$

<u>Definition</u>

The Cumulative Distribution Function of a random variable X is

$$F(x) = P(X \leq x)$$

If X is continuous with pdf $f(x)$, then

$$F(x) = \int_{-\infty}^{x} f(u)du$$

By the fundamental theorem of calculus

$$f(x) = \frac{d}{dx}F(x)$$

Properties:

1. $F(x)$ is defined for all real x
2. $F(x)$ is a non-decreasing function of x for all real x
3. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
4. $P(a < X \leq b) = F(b) - F(a)$
5. Since $P(X = a) = 0$, we have $P(a \leq X \leq b) = F(b) - F(a)$

Spinner example,

$$F(x) = \int_{-\infty}^{x} f(u)du = \int_{0}^{x} 0.25dx = 0.25x$$

Thus,

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.25x & 0 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

We can then get the pdf by

$$\begin{aligned} f(x) &= \frac{d}{dx}F(x) \\ &= \frac{d}{dx}0.25x = 0.25 \end{aligned}$$

General approach to find $F(x)$ from $f(x)$

- Treat each piece of $f(x)$ separately
- Note $F(x) = 0$ for $x <$ the minimum value in the support of $f(x)$
- Note $F(x) = 1$ for $x \geq$ the maximum value in the support of $f(x)$
- Find $F(x) = \int_{-\infty}^{x} f(u)du$

General approach to find $f(x)$ from $F(x)$

- Treat each piece of $F(x)$ separately
- Find $f(x) = \dfrac{d}{dx}F(x)$

Note

$$P(a \leq X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx$$

Quantile

Suppose X is a continuous random variable with CDF $F(x)$. The p^{th} quantile of X is the value $q(p)$ such that

$$P(X \leq q(p)) = p$$

If $p = 0.5$, then $q(0.5)$ is the median of X. We can find a given quantile by solving $F(x) = p$ for x.

Change of Variables

What if we want to find the CDF or pdf of a function of X? For the spinner example, the winning is inverse of the point we spin. Hence, the winning is $Y = \frac{1}{X}$, and we want to find

$$F_Y(y) = P(Y \leq y)$$

Specific example

$$P(Y \leq 2) = P(X^{-1} \leq 2) = P(X \geq 0.5) = \frac{7}{8}$$

How can we generalize this approach to $P(Y \leq y)$

The process

1. Find the range of values of Y
2. Write the CDF of Y as a function of X
3. Use $F_X(x)$ to find $F_Y(y)$
4. Differentiate $F_Y(y)$ if we want the pdf of Y, $f_Y(y)$

Spinner example:

Since $X \in [0, 4]$, we know $Y \in [\frac{1}{4}, \infty]$

Write the CDF of Y as a function of X. Let $y \in [\frac{1}{4}, \infty)$.

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P\left(\frac{1}{X} \leq y\right) \\ &= P\left(X \geq \frac{1}{y}\right) \\ &= 1 - P\left(X < \frac{1}{y}\right) \\ &= 1 - F_X\left(y^{-1}\right) \end{aligned}$$

Then use $F_X(x)$ to find $F_Y(y)$.

$F_Y(y) = 1 - F_X(Y^{-1})$, and we know

$$\begin{aligned} F_X(x) &= \frac{x}{4} \quad x \in [0, 4] \\ F_X(y^{-1}) &= \frac{y^{-1}}{4} = \frac{1}{4y} \end{aligned}$$

Thus,

$$F_Y(y) = \begin{cases} 0 & y < \frac{1}{4} \\ 1 - \frac{1}{4y} & y \geq \frac{1}{4} \end{cases}$$

Differentiate $F_Y(y)$ if we want the pdf of Y, $f_Y(y)$. We have $F_Y(y) = 1 - \frac{1}{4y}$ for $y \geq \frac{1}{4}$, so the pdf is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{4y^2}, \quad y \geq \frac{1}{4}$$

Expectation, Mean, and Variance for Continuous Random Variables

If X is a continuous random variable with *pdf* $f(x)$, and $g : \mathbb{R} \to \mathbb{R}$, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Thus,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ Var(X) = E([X - E(X)]^2) &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx \end{aligned}$$

Note that the shortcut still holds.

Example

$$f(x) = \begin{cases} 6x(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = 0 + \int_0^1 x \cdot 6x(1-x) dx + 0 \\ &= 6\left[\frac{1}{3} - \frac{1}{4}\right]_0^1 \\ &= 0.5 \end{aligned}$$

Now solve for $Var(X)$

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} g(x) f(x) dx \\ E(X^2) &= 0 + \int_0^1 x^2 6x(1-x) dx + 0 \\ &= \int_0^1 (6x^3 - 6x^4) dx \\ &= 6\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 \\ &= 6\left[\frac{1}{4} - \frac{1}{5}\right] = 0.3 \end{aligned}$$

Thus,

$$Var(X) = E[X^2] - (E[x])^2 = 0.3 - 0.25 = 0.05$$

7.2 Continuous Uniform Distribution

Definition

We say that X has a continuous uniform distribution on (a, b) if X takes values in (a, b) (or $[a, b]$) where all subintervals of a fixed length have the same probability.

X has pdf

$$f(x) = \begin{cases} \frac{1}{b-1} & x \in (a, b) \\ 0 & \text{otherwise} \end{cases}$$

then the CDF is

$$F(x) = \begin{cases} 0 & x < a \\ \int_a^x \frac{1}{b-a} du = \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Continuous Uniform: Mean and Variance

$$E(X) = \frac{(a+b)}{2}$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{(b-a)} \left(\frac{x^3}{3}|_a^b\right) = \frac{b^3 - a^3}{3(b-a)} \\ &= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{b^2 + ab + a^2}{3} \end{aligned}$$

$$\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 \\
&= \frac{4b+4ab+4a^2-3b^2-6ab-3a^2}{12} \\
&= \frac{b^2-2ab+a^2}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

7.3 Exponential Distribution

Consider cars arriving following a Poisson process. The number of cars arriving in t minutes follows $Poi(\lambda t)$

Let $X =$ the time you wait before you see the first car, in minutes. This is a crv.

CDF

$$\begin{aligned}
F(x) &= P(X \leq x) \\
&= P(\text{time to 1st event } > x) \\
&= 1 - P(\text{no event occurs in } (0,x)) \\
&= 1 - P(Y_x = 0)
\end{aligned}$$

where $Y_x \sim Poi(\lambda x)$ is the number of events in $(0,x)$.

Thus,

$$F(x) = 1 - \frac{e^{-\lambda x}(\lambda x)^0}{0!} = 1 - e^{-\lambda x}$$

Taking the derivative gives the pdf:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 \quad x \leq 0 \end{cases}$$

Exponential Distribution

Let $\theta = \frac{1}{\lambda}$ provide an alternate parameterization of the exponential distribution.

Definition

We say that X has an exponential distribution with parameter θ ($X \sim exp(\theta)$) if the density of X is

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x > 0 \\ 0 \quad x \geq 0 \end{cases}$$

The CDF becomes $F(x) = 1 - e^{x/\theta}$ for $x \geq 0$.

In general, for any Poisson process with rate λ, the time between events will follow an exponential distribution $exp(\theta = 1/\lambda)$

Now we want to compute $E[X]$ and $Var(X)$. We will need to solve

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \frac{1}{\theta} e^{-x/\theta} dx$$

and

$$E[X^2] = \int_0^\infty x^2 \frac{1}{\theta} e^{-x/\theta} dx$$

We can use the Gamma function.

Definition

The Gamma function, $\Gamma(\alpha)$ is defined for all $\alpha > 0$ as

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Note

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- If $a \in \mathbb{Z}^+$, then $\Gamma(\alpha) = (\alpha - 1)!$

The Gamma function tells us that if α is a positive integer, then

$$\int_0^\infty y^{\alpha-1} e^{-y} dy = (\alpha - 1)!$$

If we write $y = \frac{x}{\theta}$, then $dx = \theta dy$ and

$$\begin{aligned} E[X] &= \int_0^\infty x \frac{1}{\theta} e^{-x/\theta} dx \\ &= \int_0^\infty y e^{-y} \theta dy \\ &= \theta \int_0^\infty y^1 e^{-y} dy \end{aligned}$$

Because

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy = (\alpha - 1)!$$

$$\begin{aligned} E[X] &= \theta \int_0^\infty y^1 e^{-y} dy \\ &= \theta\Gamma(2) \\ &= \theta(1!) = \theta \end{aligned}$$

We can use the same trick with Variance to see that $Var(X) = \theta^2$

Memoryless Property

$$P(X > t + x | X > t) = P(X > x)$$
$$P(a \le X - t \le b | X > t) = P(a \le X \le b)$$

7.4 Computer Generated Random Numbers

Transform simulated observations from $U \sim \mathcal{U}(0,1)$ to obtain observations from X with CDF F. Find a function h such that $X = h(U)$.

$$\begin{aligned} F(x) &= P(X \le x) \\ &= P(h(U) \le x) \\ &= P(U \le h^{-1}(x)) \quad \text{assuming } h \text{ is strictly increasing} \\ &= h^{-1}(x) \quad \text{given the CDF of } \mathcal{U} \sim \mathcal{U}(0,1) \end{aligned}$$

Assuming F is continuous and strictly increasing, our result $F(\cdot) = h^{-1}(\cdot)$ implies

$$h = F^{-1}$$

Example

Consider the CDF of the $geo(p)$ distribution

$$F(x) = \begin{cases} 1 - (1-p)^{[x]+1} & x \geq 0 \\ 0 \quad x < 0 \end{cases}$$

Find a transformation h so that if $U \sim \mathcal{U}(0,1), X = h(u)$ has CDF F.

CDF is not continuous nor strictly increasing so we have to use the generalized inverse $F^{-1}(u) = inf\{x; F(x) \geq u\}$. For a given $0 \leq u \leq 1$:

$$\begin{aligned} F(x) &\geq u \\ 1 - (1-p)^{[x]+1} &\geq u \\ (1-p)^{[x]+1} &\leq 1 - u \\ ([x]+1)\log(1-p) &\leq \log(1-u) \\ [x] &\geq \frac{\log(1-u)}{\log(1-p)} - 1 \end{aligned}$$

The smallest x that satisfies the equation above is

$$x = \left\lceil \frac{\log(1-u)}{\log(1-p)} - 1 \right\rceil$$

Thus $h(u) = inf\{x; F(x) \geq u\} = \left\lceil \frac{\log(1-u)}{\log(1-p)} - 1 \right\rceil$

7.5 Normal Distribution

Characteristics: Symmetric about a mean value, more concentrated around the mean than the tails (and unimodal).

Definition

X is said to have a normal distribution (or Gaussian distribution) with mean μ and variance σ^2 if the density of X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

or $X \sim \mathcal{N}(\mu, \sigma^2)$, or $X \sim G(\mu, \sigma)$.

Properties:

1. Symmetric about its mean: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $P(X \leq \mu - t) = P(X \geq \mu + t)$
2. Density of unimodal: Peak is at μ. The mode, median, and mean are the same μ.
3. Mean and Variance are the parameters: $E(X) = \mu$, and $Var(X) = \sigma^2$

A classic problem, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then what is the value of

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = ???$$

This integral is weird.

Definition

We say that X is a standard normal random variable if $X \sim \mathcal{N}(0,1)$. The probability density function is:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

The CDF is

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy$$

Standard Normal Tables (Z-Tables). Values of the function $\phi(x)$ are tabulated in Standard Normal Tables.

We use symmetry:

$$\begin{aligned} & P(|Z| \le c) = 0.2 \\ \implies & P(Z \le -c) + P(Z \ge c) = 0.8 \\ \implies & P(Z \ge c) = 0.4 \\ \implies & P(Z \le c) = 0.6 \end{aligned}$$

<u>Standardization</u>

<u>Theorem</u>

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then defining

$$Z = \frac{X - \mu}{\sigma}$$

we have $Z \sim \mathcal{N}(0,1)$.

An important consequence of $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right)$$

where $Z \sim \mathcal{N}(0,1)$

The process to find $P(Z \le x)$ when $X \sim \mathcal{N}(\mu, \sigma^2)$:

1. Compute $\frac{x-\mu}{\sigma}$
2. Use standard normal tables to find $P\left(Z \le \frac{x-\mu}{\sigma}\right)$
3. This equals $P(X \le x)$

$$P(X > x) = P\left(Z > \frac{x-\mu}{\sigma}\right)$$
$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

where $Z \sim \mathcal{N}(0,1)$.

Let $X \sim \mathcal{N}(\mu, \sigma)$. Find a such that $P(X \le a) = 0.6$; this is the 60th percentile of Z.

$$P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = 0.6 \implies \frac{a-\mu}{\sigma} = \Phi^{-1}(0.6)$$

So $a = \sigma\Phi^{-1}(0.6) + \mu$

8 Multivariate Distributions

8.1 Basic Terminology and Techniques

So far we've only considered univariate distributions.

Suppose that X and Y are discrete random variables defined on the sample space. The joint probability function of X and Y is

$$f(x, y) = P(\{X = x\} \cap \{Y = y\}) x \in X(S), y \in Y(S)$$

a shorthand is,

$$f(x, y) = P(X = x, Y = y)$$

For a collection of n discrete random variables, $X_1, \ldots, X_n$, the joint probability function is defined as

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

For example if we rolled three dice, and let X_i denote the result on the i^{th} die, we'd have $f(x_1, x_2, x_3) = \frac{1}{216}$

Properties:

1. $0 \leq f(x, y) \leq 1$
2. $\sum_{x,y} f(x, y) = 1$

Computing probability from the joint probability function:

Let A be a subset of (x, y) values that (X, Y) could take. Then

$$P((X, Y) \in A) = \sum_{(x,y) \in A} f(x, y)$$

Definition

Suppose that X and Y are discrete random variables with joint probability function $f(x, y)$. The marginal probability function of X is

$$f_X(x) = P(X = x) = \sum_{y \in Y(S)} f(x, y), \quad x \in X(S)$$

Similarly, the marginal distribution of Y is

$$f_Y(y) = P(Y = y) = \sum_{x \in X(S)} f(x, y), \quad y \in Y(S)$$

Definition

Suppose that X and Y are discrete random variables with joint probability function $f(x, y)$ and marginal probability functions $f_X(x)$ and $f_Y(y)$. X and Y are said to be independent random variables if and only if

$$f(x, y) = f_X(x) f_Y(y), \quad \forall x \in X(S), y \in Y(S)$$

This is the same as saying

$$P(X = x, Y = y) = P(X = x) P(Y = y), \quad \forall x, y$$

Rolling two 6-sided dice with $X =$ the outcome on the first die and $Y =$ the outcome on the second die:

$$f(x, y) = \frac{1}{36}, \quad x, y \in \{1, 2, 3, 4, 5, 6\}$$

but we also know

$$f_X(x) = \frac{1}{6}, f_Y(y) = \frac{1}{6} \quad x, y \in \{1, 2, 3, 4, 5, 6\}$$

and so $f(x, y) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = f_X(x) f_Y(y)$ and we have independence.

This abstracts to size n.

Definition

The conditional probability function of X given $Y = y$ is denoted $f_{X|Y}(x|y)$, and is defined to be

$$f_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{f_Y(y)}$$

What about functions of random variables?

Suppose that

$$h : \mathbb{R}^2 \to \mathbb{R}$$

For jointly distributed random variables X and Y, $\mathcal{U} = h(X, Y)$ is a random variable. If X and Y have joint p.f. $f(x, y)$, then the probability function of $\mathcal{U}$ is given by:

$$f_{\mathcal{U}} = P(\mathcal{U} = t) = \sum_{(x,y):h(x,y)=t} f(x, y)$$

- If $X \sim Binomial(n, p)$ and $Y \sim Binomial(m, p)$ independently, then $X + Y \sim Binomial(n + m, p)$
- If $X \sim Poi(\mu_1)$ and $Y \sim Poi(\mu_2)$ independently, then $X + Y \sim Poi(\mu_1 + \mu_2)$

8.2 Multinomial Distribution

Definition

Multinomial Distribution: Consider an experiment in which:

1. Individual trials have k possible outcomes, and the probabilities of each individual outcome are denoted $p_i, i = 1, \ldots, k$, so that $p_1 + p_2 + \cdots + p_k = 1$
2. Trials are independently repeated n times, with X_i denoting the number of times outcome i occurred, so that $X_1 + X_2 + \cdots X_k = n$

If $X_1, \ldots, X_k$ have multinomial distribution with parameters n and $p_1, \ldots, p_k$, then their joint probability function is

$$f(x_1, \ldots x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

where $x_1, \ldots, x_k$ satisfy $x_1 + \cdots + x_k = n, x_i \geq 0$.

The terms

$$\frac{n!}{x_1!x_2! \cdots x_k!}, \quad \text{for } x_1 + \cdots + x_k = n$$

are called multinomial coefficients.

If $X_1, \ldots, X_k$ have joint Multinomial distribution with parameters n and $p_1, \ldots, p_k$, then

$$X_i \sim Bin(n, p_i)$$

Also

$$\sum X_i \sim Bin\left(n, \sum p_i\right)$$

8.3 Expectation for Multivariate Distributions: Covariance and Correlation

Recall,

$$E[X] = \sum xf(x) \quad E[X] = \int xf(x)dx$$

Definition

Suppose X and Y are jointly distributed random variables with joint probability function $f(x,y)$. Then for a function $g : \mathbb{R}^2 \to \mathbb{R}$,

$$E(g(X,Y)) = \sum_{(x,y)} g(x,y)f(x,y)$$

More generally, if $g : \mathbb{R}^n \to \mathbb{R}$, and $X_1, \ldots, X_n$ have joint p.f. $f(x_1, \ldots, x_n)$, then

$$E(g(X_1, \ldots, X_n)) = \sum_{(x_1,\ldots,x_n)} g(x_1, \ldots, x_n)f(x_1, \ldots, x_n)$$

Linear Properties of Expectation

$$E(X+Y) = \sum_x xf_X(x) + \sum_y yf_Y(y) = E(X) + E(Y)$$

and in general,

$$E[a \cdot g_1(X,Y) + b \cdot g_2(X,Y)] = a \cdot E(g_1(X,Y)) + b \cdot E(g_2(X,Y))$$

Independence gives a useful - but simplistic - way of describing relationships between variables.

Definition

If X and Y are jointly distributed, then $Cov(X,Y)$ denotes the covariance between X and Y. It is defined by

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

or the shortcut,

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Properties of Covariance

- Positive $Cov(X,Y) \implies Y$ increases as X increases
- Negative $Cov(X,Y) \implies Y$ decreases as X increases
- The larger the absolute value of $Cov(X,Y)$, the stronger the relationship is
- $Cov(X,X) = Var(X)$. $\quad Cov(X,X) = E[XX] - E[X]E[X] = E[X^2] - E[X]^2$
- $Cov(X,c) = 0$ for any constant c.
- $Cov(Y,X) = Cov(X,Y) = E[(Y - \mu_Y)(X - \mu_X)]$

Theorem

If X and Y are independent, then $Cov(X,Y) = 0$.

Proof

$$\begin{aligned} Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] &= \sum_{\text{all } y}\left[\sum_{\text{all } x}(x-\mu_X)(y-\mu_Y)f_1(x)f_2(y)\right] \\ &= \sum_{\text{all } y}\left[(y-\mu_Y)f_2(y)\sum_{\text{all } x}(x-\mu_X)f_1(x)\right] \\ &= \sum_{\text{all } y}\left[(y-\mu_Y)f_2(y)E(X-\mu_X)\right] \\ &= \sum_{\text{all } y} 0 = 0 \end{aligned}$$

X and Y are independent $\implies Cov(X,Y)=0$.

It is important to know that the converse is false. If $Cov(X,Y)=0$ then X and Y are not necessarily independent.

$Cov(X,Y)=0 \iff E[XY]=E[X]E[Y]$

X and Y are uncorrelated if $Cov(X,Y)=0$, but that does not mean they are independent.

Independence $\implies Cov(X,Y)=0 \implies E[XY]=E[X]E[Y]$
Independence $\implies E[g_1(X)g_2(Y)]=E[g_1(X)]E[g_2(Y)]$

Definition

The correlation of X and Y is denoted $corr(X,Y)$, and is defined by

$$\rho = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

with $-1 \leq \rho \leq 1$.

Correlation measures the strength of the linear relationship between X and Y.

Properties:

- If $p \approx +1$, then X and Y will have an approximately positive linear relationship
- If $p \approx -1$, then X and Y will have an approximately negative linear relationship
- If $p \approx 0$, then X and Y are said to be uncorrelated.

We say that X and Y are uncorrelated if $Cov(X,Y)=0$ (or $corr(X,Y)=0$).

X and Y are independent $\implies$ X and Y are uncorrelated.

Once again, the converse is not true.

8.4 Mean and Variance of a Linear Combination of Random Variables

Suppose $X_1,\ldots,X_n$ are jointly distributed random variables with joint probability function $f(x_1,\ldots,x_n)$. A linear combination of the random variables $X_1,\ldots,X_n$ is a random variable of the form

$$\sum_{i=0}^{n} a_i X_i$$

where $a_1,\ldots,a_n \in \mathbb{R}$

Examples

$$S_n = \sum_{i=1}^{n} X_i \quad a_i = 1, \quad 1 \leq i \leq n$$
$$\bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i \quad a_i = \frac{1}{n}, \quad 1 \leq i \leq n$$

Expected Value of a Linear Combination:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

Mean of sample mean:

$$E[\bar{X}] = \sum_{i=1}^{n} \frac{1}{n} E[X_i] = \frac{1}{n} n\mu = \mu$$

Covariance of linear combinations:

Two useful results:

$$Cov(aX + bY, cU + dV) = acCov(X, U) + adCov(X, V) + bcCov(Y, U) + bdCov(Y, V)$$

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$$

Variance of linear combination

When $Cov(X, Y) = 0$

$Var(X + Y) = Var(X) + Var(Y)$

$Var(X - Y) = Var(X) + Var(Y)$

When $Cov(X, Y) > 0$

$Var(X + Y) > Var(X) + Var(Y)$

$Var(X - Y) < Var(X) + Var(Y)$

In the case of two random variables X and Y, and constants a and b, we have:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

If $a = b = 1$

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

If $a = 1, b = -1$

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

In general, let $X_1, X_2, \ldots, X_n$ be random variables, and write $Var(X_i) = \sigma_i^2$, then:

$$\begin{aligned} Var\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j Cov(X_i, X_j) \end{aligned}$$

If $X_1, X_2, \ldots, X_n$ are independent, then $Cov(X_i, X_j) = 0$ (for $i \neq j$) and so

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

Variance of sample mean

In general, we have for $X_1, X_2, \ldots, X_n$ independent random variables all with variance σ^2

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

this holds for any set of independent random variables that have the same variance.

8.5 Linear Combinations of Independent Normal Random Variables

Linear transformation of a normal random variable:

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$ where a and b are constants, then

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

The linear transformation of a normal random variable is still normal.

Linear combination of 2 independent normal random variables:

Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independently, and let a and b be constants, then

$$aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Linear combination of independent normal random variables is still normal.

Let $X_1, X_2, \ldots, X_n$ be independent $\mathcal{N}(\mu, \sigma^2)$ random variables. Then

$$S_n \equiv \sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

and

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$$

The IQs of UWaterloo Math students are normally distributed with mean 120 and variance 100. The probability that the average IQ in a class of 25 students is between 118 and 123 is:

$$X_i \sim \mathcal{N}(120, 100) \implies \bar{X} \sim \mathcal{N}(120, 100/25)$$
$$\implies P(118 \leq \bar{X} \leq 123) = P\left(\frac{118-120}{2} \leq \frac{\bar{X}-120}{2} \leq \frac{123-120}{2}\right) = P(-1 \leq Z \leq 1.5)$$

8.6 Indicator Random Variables

If $X \sim Binomial(n, p)$ we can think of X in the following way:

Observe the first trial, and see if it succeeds. We set X_1 to be 1 if it succeeds and 0 if it fails.

Observe the second trial, and see if it succeeds. We set X_2 to be 1 if it succeeds and 0 if it fails.

Observe the n^{th} trial, and see if it succeeds. We set X_n to be 1 if it succeeds and 0 if it fails.

Sum the successful events to find X

$$X = \sum_{i=1}^{n} I_i$$

X is a linear combination of random variables.

Definition

Let A be an event which may possibly result from an experiment. We say that $\mathbf{1}_A$ is the indicator random variable of the event A. $\mathbf{1}_A$ is defined by:

$$\mathbf{1}_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Covariance of Indicator Random Variables:

Suppose we have two events, A and B, and we define

$$\mathbf{1}_A = 1 \text{ if } A \text{ occurs, and } \mathbf{1}_A = 0 \text{ otherwise}$$
$$\mathbf{1}_B = 1 \text{ if } B \text{ occurs, and } \mathbf{1}_B = 0 \text{ otherwise}$$

How do we find $Cov(\mathbf{1}_A, \mathbf{1}_B)$

For general, X, Y

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

so for indicator variables we have

$$Cov(\mathbf{1}_A, \mathbf{1}_B) = E[\mathbf{1}_A \times \mathbf{1}_B] - E[\mathbf{1}_A]E[\mathbf{1}_B]$$

We know that $E[\mathbf{1}_A] = P(A)$ and $E[\mathbf{1}_B] = P(B)$, so we just have to find

$$\begin{aligned}
&E[\mathbf{1}_A \times \mathbf{1}_B] \\
&= 1 \times P(A \cap B) + 0 \times (1 - P(A \cap B)) \\
&= P(A \cap B)
\end{aligned}$$

If A and B are events, then

$$Cov(\mathbf{1}_A, \mathbf{1}_B) = P(A \cap B) - P(A)P(B)$$

$\mathbf{1}_A$ and $\mathbf{1}_B$ are independent $\iff$ $\mathbf{1}_A$ and $\mathbf{1}_B$ are uncorrelated.

9 Central Limit Theorem and Moment Generating Functions

9.1 Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are independent random variables from the same distribution, with mean μ and variance σ^2, then as $n \to \infty$, the shape of the probability histogram for the random variable $S_n = \sum_{i=1}^n X_i$ approaches the shape of a $\mathcal{N}(n\mu, n\sigma^2)$ probability density function.

The cumulative distribution function of the random variable

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

approaches the $\mathcal{N}(0,1)$ cumulative distribution function. Similarly, the cumulative distribution function of

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the $\mathcal{N}(0,1)$ cumulative distribution function.

In other words, if n is large:

$$S_n = \sum_{i=1}^n X_i$$

has approximately a $\mathcal{N}(n\mu, n\sigma^2)$ distribution, and

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$$

has approximately a $\mathcal{N}(\mu, \sigma^2/n)$ distribution.

Example

Roll a 6-sided die 1000 times and record each result. If the die is a fair die, estimate the probability that the sum of the die rolls is less than 3400.

Let X_i be the dot for the $i-$th roll. We want the probability $P(\sum_{i=1}^{1000} X_i < 3400)$. Without CLT this is very difficult.

Using CLT:

Seems reasonable to assume independence, each X_i is a discrete $\mathcal{U}(1,6)$ random variable, and $n = 1000$ is large.

Mean and variance

$$\mu = \frac{a+b}{2} = 3.5 \quad \sigma^2 = \frac{(b-a+1)^2-1}{12} = \frac{35}{12}$$
$$E[S_n] = n\mu = 3.5n \quad Var(S_n) = n\sigma^2 = \frac{35}{12}n$$

Apply CLT (Standardization)

$$\frac{S_n - 3.5n}{\sqrt{\frac{35}{12}n}} \to \mathcal{N}(0,1)$$

Use the Z-table

$$P(S_n < 3400) = P\left(\frac{S_n - 3.5n}{\sqrt{\frac{35}{12}n}} < \frac{3400 - 3.5n}{\sqrt{\frac{35}{12}n}}\right)$$
$$\approx P(Z < -1.852) = 0.032$$

Note:

- CLT doesn't hold if μ and/or σ^2 don't exist
- Accuracy depends on the size of n and the actual distribution of the X_i
- CLT works for any distribution of X_i

Normal approximation to binomial

Note that $S_n = \sum_{i=1}^{n} X_i \sim Binomial(n,p)$ if all $X_i \sim Binomial(1,p)$ are independent.

Then if $S_n \sim Binomial(n,p)$ then for large n, the random variable

$$Z = \frac{S_n - np}{\sqrt{np(1-p)}}$$

has approximately a $\mathcal{N}(0,1)$.

Continuity Correction

We need to be careful with CLT when working with discrete random variables. For example we've computed $P(15 \le S_n \le 20)$. But we know that X can't take non-integer values. Therefore, $P(14.5 \le S_n \le 20.5)$ gives us a better estimate. This is called continuity correction.

We should apply this when approximating discrete distributions using the CLT, and not continuous distributions.

- $P(a \le X \le b) \to P(a - 0.5 \le X \le b + 0.5)$
- $P(X \le b) \to P(X \le b + 0.5)$
- $P(X < b) \to P(X < b - 0.5)$
- $P(X \ge a) \to P(X \ge a - 0.5)$
- $P(X > a) \to P(X > a + 0.5)$
- $P(X = x) \to P(x - 0.5 \le X \le x + 0.5)$

If $X \sim Poisson(\mu)$, then the cumulative distribution function of

$$Z = \frac{X - \mu}{\sqrt{\mu}}$$

approaches that of a $\mathcal{N}(0,1)$ random variable as $\mu \to \infty$

9.2 Moment Generating Functions

This is the third type of function that uniquely determines a distribution.

Definition

The moment generating function (MGF) of a random variable X is given by

$$M_X(t) = E(e^{tX}), \quad t \in \mathbb{R}$$

If X is discrete with probability function $f(x)$, then

$$M_X(t) = \sum_x e^{tx} f(x), \quad t \in \mathbb{R}$$

If X is continuous with density $f(x)$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx, \quad t \in \mathbb{R}$$

Properties:

$$M_X(t) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \frac{t^3 E(X^3)}{3!} + \dots$$

So long as $M_X(t)$ is defined in a neighbourhood of $t = 0$

$$\frac{d}{dt^k} M_X(0) = E(X^k)$$

Suppose X has a $Binomial(n, p)$ distribution. Then its moment generating function is

$$\begin{aligned} M(t) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

Therefore,

$$\begin{aligned} M'(t) &= npe^t(pe^t + 1 - p)^{n-1} \\ M''(t) &= npe^t(pe^t + 1 - p)^{n-1} + n(n-1)p^2 e^{2t}(pe^t + 1 - p)^{n-2} \end{aligned}$$

so

$$\begin{aligned} E(X) &= M'(0) = np \\ E(X^2) &= M''(0) = np + n(n-1)p^2 \\ Var(X) &= E(X^2) - E(X)^2 = np(1-p) \end{aligned}$$

What is the Moment Generating Function for $\mathbf{I}_A$?

The distribution of $\mathbf{I}_A$ is

$$\begin{aligned} P(\mathbf{I}_A = 1) &= P(A), \quad P(\mathbf{I}_A = 0) = 1 - P(A) \\ M_{\mathbf{I}_A}(t) &= E[e^{t\mathbf{I}_A}] \\ &= e^{t \times 0} P(\mathbf{I}_A = 0) + e^{t \times 1} P(\mathbf{I}_A = 1) \\ &= 1 - P(A) + e^t P(A) \end{aligned}$$

What is the Moment Generating Function for $X \sim \mathcal{U}(a, b)$?

When $t = 0$, we have $M_X(t) = E[e^{tX}] = E[e^{0\times X}] = E[1] = 1$

When $t \neq 0$

$$\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_a^b e^{tx} f(x)dx \\
&= \int_a^b e^{tx} \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \int_a^b e^{tx} dx \\
&= \frac{1}{b-a} \frac{e^{tx}}{t} |_a^b \\
&= \frac{e^{bt} - e^{at}}{t(b-a)}
\end{aligned}$$

Theorem (Uniqueness Theorem)

Suppose that random variables X and Y have MGF's $M_X(t)$ and $M_Y(t)$ respectively. If $M_X(t) = M_Y(t)$ for all t, then X and Y have the same distribution.

Theorem

Suppose that X and Y are independent and each have moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of $X + Y$ is

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E\left(e^{tX}\right) E\left(e^{tY}\right) = M_X(t)M_Y(t)$$

We can use this to prove the following.

Suppose that $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and X and Y are independent.

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

9.3 Multivariate Moment Generating Functions

Definition

The joint moment generating function of two random variables, X and Y is

$$M(s,t) = E\left(e^{sX+tY}\right)$$

And so if X and Y are independent

$$M(s,t) = E\left(e^{sX}\right) E\left(e^{tY}\right) = M_X(s)M_Y(t)$$