1 Probability Theory

Chapter abstract:

Outline

1.1 Experiments and Sample Spaces

- Ω represents the set of all possible outcomes of an experiment. For example, if we roll a die, $\Omega = \{1,2,3,4,5,6\}$.
- $\mathcal{A}$ is a collection of subsets of Ω that satisfies:
  1. $\Omega \in \mathcal{A}$
  2. If $E \in \mathcal{A}$, then $\bar{E} \in \mathcal{A}$
  3. If $E_1, E_2, \ldots, E_n \in \mathcal{A}$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{A}$
- Additional concepts:
  - If $E, F \in \mathcal{A}$, then $E \cap F \in \mathcal{A}$

The simplest sigma algebra is $\mathcal{A} = \{\Omega, \emptyset\}$.

The Power set of a finite set $\Omega = \{a,b,c\}$:

$$2^{\Omega} = \{\{a,b,c\},\{a,b\},\{b,c\},\{a,c\},\{a\},\{b\},\{c\},\emptyset\}$$

The power set contains 2^n elements for a sample space with n elements.

1.2 Probabilities

A *probability measure*, P, is a function that assigns a real number to events in $\mathcal{A}$. For each $E \in \mathcal{A}$, we have $P(E) \in [0,1]$, and the following conditions hold:

- $P(\Omega) = 1$
- Additivity: For disjoint events $E_1, E_2, \ldots$,

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

- For two events E and F,

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Note, for three events:

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$$

1.3 Combinatory

1.3.1 Permutations

Permutations involve aligning n objects in n places. There are two key cases:

- **Distinct objects (ex. mother):** All n objects are distinct, and we can arrange them in $n!$ ways. For example, arranging 6 distinct objects in 6 places can be done in $P_n = n!$ ways. Ex. Olympics possible arrivals.

$$P_n = n!$$

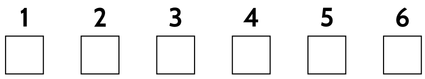


Figure 1.1: 6 boxes for permutation

- **Repetitive permutations (ex. mummy):** If some objects are repeated, we account for the repetition by dividing by the factorial of the count of the repeated objects. For example, arranging 5 objects where 3 are identical (mmm = 3 k1, u = 1k2, y = 1k3)can be done in:

$$P_n = \frac{n!}{k_1!k_2!k_3!\ldots} = \frac{5!}{3!1!1!}$$

1.3.2 Dispositions

Dispositions involve aligning objects in fewer places than the total number of objects. We consider two cases:

- **Simple dispositions:** This involves choosing and aligning k objects from n. For example, if we have 8 runners and we are selecting the top 3 positions, we compute:

$$D_{n,k} = \frac{n!}{(n-k)!} = \frac{8!}{(8-3)!}$$

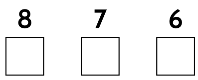


Figure 1.2: Simple dispositions

- **With repetition:** When repetition is allowed, the number of dispositions is given by:

$$D^*_{n,k} = n^k$$

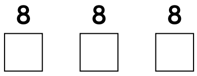


Figure 1.3: Disposition with repetition

1.3.3 Combinations

Combinations involve grouping objects, where the order does not matter. We calculate how many distinct subsets of k elements can be built out of n objects. For example, selecting 3 elements from a set can be done in:

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

Unlike permutations, combinations do not take into account the order of the elements. For example, a locker code is not a combination, but a permutation. Note the formula is just like the one for permutations but as here order does not matter we divide for the possible k positions that a single letter can take.

1.4 Examples

Example 1.
We are given a set of 6 green balls (G) and 4 red balls (R). If we pick 3 balls, we can explore the following cases:

Case 1: With repetition allowed
We calculate the probability of drawing 3 green balls as:

$$P(\text{3 greens}) = \frac{D^*_{6,3}}{D^*_{10,3}} = \frac{6^3}{10^3} = \frac{1}{6}$$

Case 2: Without repetition
If we pick 3 balls without repetition, the probability is calculated as:

$$P(\text{3 greens}) = \frac{D_{6,3}}{D_{10,3}} = \frac{6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8} = \frac{1}{6}$$

Case 3: All the three balls extracted together
In this case, we calculate the probability of drawing 3 green balls together:

$$P(\text{3 greens together}) = \frac{C_{6,3}}{C_{10,3}} = \frac{1}{6}$$

This is the same result as in Case 2, showing consistency across different experiment setups.

Example 2.
Example 2: Ten Balls, Each with a Number from 1 to 10, Extract Three Balls without Repetition

1. What is the probability that the third ball is even? A priori, we just know that the probability of drawing an even ball at each trial is $\frac{1}{n}$.

2. What is the probability that the numbers of the extracted balls are in increasing order (i.e., $I < II < III$)? Ont of 10 elements, you can got 6 combinations for each 3 elements, of these 6, just one will be in increasing order out of 720 possible combinations, 120 will be in increasing order. Hence the potability will be $\frac{120}{720}$

Example 3.
In a group of n people, what is the probability that at least 2 people share the same birthday?

1. **What is the probability that 2 persons share the same birthday?**

The complement of this event is the probability that no one shares the same birthday. We can use the formula for simple dispositions (without repetition) to find the number of ways to assign k distinct birthdays of k people out of the possibility of dispositions with repetitions:

$$P(\overline{E}) = \frac{D_{365,n}}{D^*_{365,n}}$$

So, the probability that at least two people share the same birthday is:

$$P(E) = 1 - P(\overline{E}) = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - n)}{365^n}$$

2. **What is the probability that I share a birthday with one of these persons?** Probability everyone is disposed with repetition in any other day then mine, over probability of disposition with repetition on any other day:

$$P(\overline{E}) = \frac{D^*_{365-1}}{D^*_{365}} \to P(E) = 1 - \left(\frac{364}{365}\right)^n$$

As the number of people increases, the probability that at least one person shares a birthday with me increases as well.

Example 4: Dimostrazione

$$n \begin{cases} Letters \\ Envelopes \end{cases}$$

Let us assume there are two scenarios

$$\phi_1 = \text{probability that all the letters are in the right envelope} = \frac{1}{n!}$$
$$\phi_2 = \text{probability that no letter is in the right envelope} = \clubsuit$$

We have two complementary probabilities, N and $\bar{N}$, described as:

$$N = \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n$$
$$\bar{N} = E_1 \cup E_2 \cup \dots \cup E_n$$

$$\mathbb{P}(N) = 1 - \mathbb{P}(\bar{N}) = 1 - \mathbb{P}(E_1 \cup E_2 \cup \ldots \cup E_n) =$$
$$1 - \mathbb{P}(E_1) - \mathbb{P}(E_2) - \ldots - \mathbb{P}(E_n) + \mathbb{P}(E_1 \cap E_2) + \ldots + \mathbb{P}(E_{n-1} \cap E_n) - \ldots \tag{1.1}$$

$$1 - \left(\sum_{1 \le i \le n} \mathbb{P}(E_i) - \sum_{1 \le i \le j \le n} \mathbb{P}(E_i \cap E_j) + \sum_{1 \le i \le j \le k \le n} \mathbb{P}(E_i \cap E_j \cap E_k) - \ldots + (-1)^n \mathbb{P}(E_1 \cap E_2 \cap E_3 \cap \ldots E_n) \right) \tag{1.2}$$

Note that:

$$\mathbb{P}(E_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$
$$\mathbb{P}(E_i \cap E_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$
$$\mathbb{P}(E_i \cap E_j \cap E_k) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}$$
$$\mathbb{P}(E_1 \cap E_2 \cap E_3 \cap \ldots E_n) = \frac{0!}{n!} = 0$$

$$1 - \left(n \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{1}{n(n-1)} + \binom{n}{3} \cdot \frac{1}{n(n-1)(n-2)} - \ldots + (-1)^n \cdot 0 \right) \tag{1.3}$$

$$1 + \sum_{k-1}^{n} (-1)^k \binom{n}{k} \frac{(n-k)!}{n!} = 1 + \sum_{k=1}^{n} (-1)^k \frac{1}{k!} \tag{1.4}$$

$$\mathbb{P}(N) = \sum_{k=0}^{n} (-1)^k \frac{1}{k!} \xrightarrow{n \to \infty} e^{-1} \tag{1.5}$$

With

$$e^{\alpha} = \sum_{k=0}^{\infty} \frac{\alpha^t}{k!}$$

Example 5: Dimostrazione

Given n keys, define E_j = "enters home at trial j"

- **No key repetition**

$$\mathbb{P}(E_j) = \frac{D_{n-1,j-1} \cdot D_{1,1}}{D_{n,j}} = \frac{(n-1)!}{n!} = \frac{1}{n} \tag{1.6}$$

  This only holds true for $j = 1, 2, \ldots, n$, when the probability is *a priori*.

- **With key repetition**

$$\mathbb{P}(E_j) = \frac{D^*_{n-1,j-1} \cdot \overbrace{1}^{D_{1,j}}}{D^*_{n,j}} = \frac{(n-1)^{j-1}}{n^j} = \left(1 - \frac{1}{n}\right)^{j-1} \cdot \frac{1}{n} \tag{1.7}$$

  Where the latter element of the equation is a series to 1.

1.5 Independent Events

Stochastic Independence

Definition 1: Stochastic Independence
Given the probability space $(\Omega, A, \mathbb{P})$, two events $E, F \in A$ are said to be **stocastically independent** if $\mathbb{P}(E) \cap \mathbb{P}(F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$

$$\mathbb{P}(E|F) = \mathbb{P}(E) \iff \mathbb{P}(F|E) = \mathbb{P}(F) \tag{1.8}$$

Conditional Stochastic Independence

Definition 2: Conditional Stochastic Independence
Given the probability space $(\Omega, A, \mathbb{P})$ and the events A, B, F, **Conditional Stochastic Independence** can be characterised as:

$$\mathbb{P}(A \cap B|F) = \mathbb{P}(A|F) \cdot \mathbb{P}(B|F) \tag{1.9}$$

Warning : Stochastic Independence **does not imply** Conditional Stochastic Independence, and vice versa.

1.6 Conditional Probability

In the probability space $(\Omega, A, \mathbb{P})$, conditional probability can be written as

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \tag{1.10}$$

with $\mathbb{P}(F) > 0$

Conditional Probability as Intersection of Events

$$\begin{aligned} \mathbb{P}(E \cap F) =& \mathbb{P}(F) \cdot \mathbb{P}(E|F) \\ \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) =& \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_2 \cap E_1) \cdots \mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \end{aligned} \tag{1.11}$$

Law of Total Probability

Given $E_1,\ E_2,\ \cdots\ E_n$ partitions of Ω and the conditions

1. $E_i \cap E_j = \emptyset \forall i \neq j$
2. $E_1 \cup E_2 \cup \cdots \cup E_n = \Omega$

We can state that:

$$\mathbb{P}(F) = \sum_{i=1}^{n} \mathbb{P}(F|Ei)\mathbb{P}(E_i) \tag{1.12}$$

Bayes Formula

Given

$$\mathbb{P}(E_n|F) = \frac{\mathbb{P}(F|E_n)\mathbb{P}(E_n)}{\sum_{i=1}^{n} \mathbb{P}(F|Ei)\mathbb{P}(E_i)} \to \mathbb{P}(A|P)\frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{1.13}$$

Ex.1 Conditional Proabbility

Probability	Damaged	
A	0.5	0.06
B	0.3	0.04
C	0.2	0.08

Table 1.1: Example Table

$$\begin{aligned} \mathbb{P}(D) &= \mathbb{P}(D|A)\cdot\mathbb{P}(A) + \mathbb{P}(D|B)\cdot\mathbb{P}(B) + \mathbb{P}(D|C)\cdot\mathbb{P}(C) \\ &= 0.06\cdot 0.5 + 0.04\cdot 0.3 + 0.08\cdot 0.2 = 0.06 + 0.012 + 0.016 = 0.088 \end{aligned} \tag{1.14}$$

$$\mathbb{P}(B|D) = \frac{\mathbb{P}(D|B)\mathbb{P}(B)}{\mathbb{P}(D)} = \frac{\mathbb{P}(B\cap D)}{\mathbb{P}(D)} = \frac{0.04\cdot 0.3}{0.088} = \frac{0.012}{0.088} = 0.136 \tag{1.15}$$

Ex.2 Conditional Probability

$$\mathbb{P}(t_+|ill) = 0.98$$

$$\mathbb{P}(t\_|not\ ill) = 0.99$$

Given $\mathbb{P}(ill) = p$

$$\to \mathbb{P}(ill|t_+) = \frac{\mathbb{P}(t_+|ill)\mathbb{P}(ill)}{\mathbb{P}(t_+)} = \frac{0.98\cdot p}{\mathbb{P}(t_+)} = \frac{0.98\cdot p}{0.98\cdot p + 0.01\cdot(1-p)} \tag{1.16}$$

Ex.3 Conditional Probability

Image we toss two dices

$$\Omega = \{(1,1),(1,2),(1,3)...\}$$

What's $\mathbb{P}(E)$, given E : "sum is 8" and that F = first toss = 5?

$$\begin{aligned} \mathbb{P}(E|F) &= \frac{1}{6} \\ \mathbb{P}(E) &= \frac{5}{36} \end{aligned} \tag{1.17}$$

Two events E and F are positively correlated wgenever $\mathbb{P}(E|F) > \mathbb{P}(E) \to \mathbb{P}(E \cap F) > \mathbb{P}(E) \cdot \mathbb{P}(F)$

1.7 Random Variables

Given the sample space of tossing a dice $\Omega = \{1,2,3,4,5,6\}$, we can define:

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

Remark 1.
In the equation above:
- if Ω is finite $\to$ it is a random variable
- if Ω is discrete $\to$ it is **not** a random variable

Properties to be considered a random variable:

- $\forall t \in \mathbb{R}$

  $\{\omega \in \Omega : X(\omega) \leq t\} = E_t \in A$

  With X being **Borel measurable**
- $\mathbb{P}(E_t) = \mathbb{P}(\omega \in \Omega : X(\omega) \leq t) = \mathbb{P}(X(\omega) \leq t) = \mathbb{P}(X \leq t) 0 F_x(t)$
- $F_x : \mathbb{R} \to \mathbb{R}$, with $F_x(t)$ being the distribution function of the variable.

  $\lim_{t \to -\infty} F(t) = 0$

  $\lim_{t \to \infty} F(t) = 1$
  - If X is discrete, the graph will possess the following properties:
    1. Stepwise
    2. Non-decreasing
    3. F_x continuous from the right
  - If X is absolutely continuous
    1. F is continuous

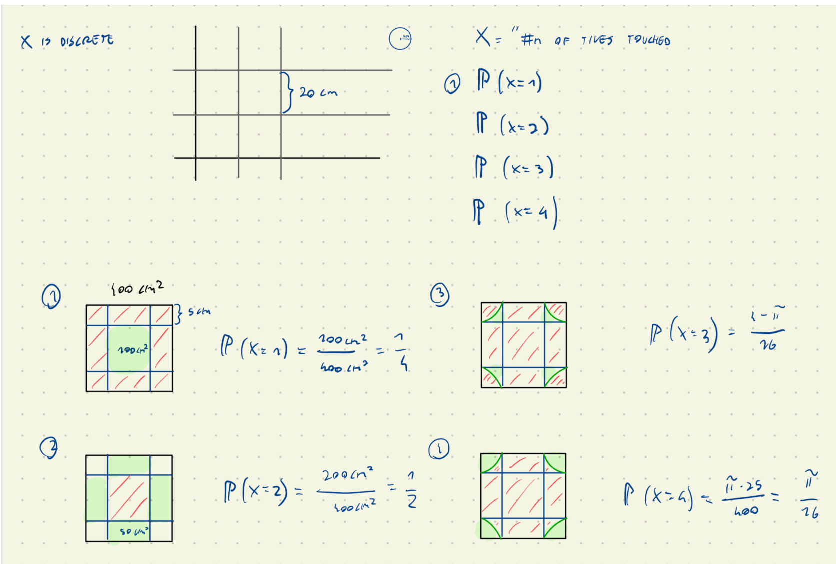


Figure 1.4: Visual representation of an example where X is a **discrete** random variable

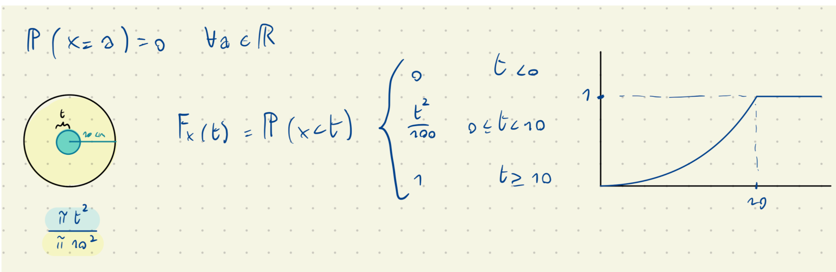


Figure 1.5: Visual representation of an example where X is an **absolutely continuous** random variable

2 Probability Functions

Chapter abstract: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse augue est, porttitor commodo velit a, tristique pharetra ante. Mauris pretium ante at lorem suscipit porttitor. Sed neque tortor, lacinia a aliquam quis, molestie tincidunt nisi. Vivamus congue cursus iaculis. Aenean id massa convallis, sodales metus a, imperdiet velit. In metus erat, suscipit vel mollis sed, tincidunt at ante. In hac habitasse platea dictumst. Cras malesuada mollis odio, eget mattis mauris tincidunt a.

Outline

2.1 Probability Density Function | $f(x)$

$$\begin{aligned}\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) &= \overbrace{F_x(b)}^{CDF} - F_x(a) \\ &= \int_a^b \underbrace{f(x)}_{PDF} \, dx\end{aligned} \tag{2.1}$$

Note : Relationship between Cumulative Distribution Function and Probability Density Function:

$$f_x(x) = \frac{d}{dx} F_x(X) \iff F_x(t) = \int_{-\infty}^{t} f_x(x) dx$$

2.1.1 Properties of the Probability Density Function

1. **Non Negativity**: the density is never negative

$$f_x(x) \geq 0 \; \forall x \in \mathbb{R} \tag{2.2}$$

2. **Normalisation**: the area below the curve of the density function is always equal to 1

$$\int_{-\infty}^{\infty} f_x(x) dx = 1 \tag{2.3}$$

Example 4.

$$f_x(x) = \begin{cases} cx^2 & \text{if } -1 < x \leq 2 \\ 0 & \text{a.e.} \end{cases} \tag{2.4}$$

$$1 = \int_{-\infty}^{\infty} f_x(x)dx = \underbrace{\int_{-\infty}^{-1} f_x(x)dx}_{=0} + \int_{-1}^{2} f_x(x)dx + \underbrace{\int_{2}^{\infty} f_x(x)dx}_{=0} = \int_{-1}^{2} f_x(x)dx =$$

$$\int_{-1}^{2} cx^2 dx = \left[c\frac{x^3}{3}\right]_{-1}^{2} = \frac{c}{3}(8+1) = 3c \to 3c = 1 \to c = \frac{1}{3}$$

Given the function $c \cdot x^2$, the constant c that allows the function to respect the properties of a *probability density function* is $c = \frac{1}{3}$. The final form of function 2.5 is hence:

$$f_x(x) = \begin{cases} \frac{1}{3}x^2 & \text{if } -1 < x \leq 2 \\ 0 & \text{a.e.} \end{cases}$$

Example 5.
Imagine there is a traffic light, where the green light lasts for 20' and the red light lasts for 40'. Define X as the waiting time. (Note: X is neither discrete nor continuous).

$$F_x(t) = \begin{cases} 0 & \text{if } t < 0 \\ \clubsuit & \text{if } 0 \leq t < 40 \\ 1 & \text{if } t \geq 40 \end{cases} \tag{2.5}$$

$$\clubsuit = \mathbb{P}(x \leq t) = \mathbb{P}(x \leq t|R) \cdot \mathbb{P}(R) + \mathbb{P}(x \leq t|G) \cdot \mathbb{P}(G) =$$

$$\frac{t}{40} \cdot \overbrace{\frac{40}{20+40}}^{prob.\ light=R} + 1 \cdot \overbrace{\frac{20}{20+40}}^{prob.\ light=G} =$$

$$\clubsuit = \frac{1}{3} \cdot \frac{t+20}{60}$$

2.2 Expected Value

The expected value of a probability function can be described as:

$$E(x) = \int_0^{+\infty} (1 - F_x(t))dt - \int_{-\infty}^{0} F_x(t)dt \begin{cases} \sum x_i p_x(I) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{+\infty} x f_x(x)dx & \text{if } X \text{ is continuous} \end{cases} \tag{2.6}$$

Example 6.
When X = *number of tiles*:

$$E(x) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{4-\pi}{16} + 4 \cdot \frac{\pi}{16} = 2 + \frac{\pi}{16}$$

Example 7.
When X = *distance from centre*:

$$E(x) = \int_{-\infty}^{+\infty} x f_x(x) dx$$

Example 8.
When X = *waiting time at the traffic light*:

$$E(x) = \int_{-\infty}^{+\infty} x f_x(x) dx$$

fine pagine

3 Discrete Probability Distributions

Chapter abstract: This chapter provides an overview of the most common discrete probability distributions. We start by introducing the Bernoulli distribution, which models a single coin toss. We then move on to the binomial distribution, which models the number of successes in a fixed number of Bernoulli trials. We also introduce the geometric and negative binomial distributions, for the number of trials needed to get the first success and a fixed number of successes, respectively. Finally, we introduce the hypergeometric distribution, which models the number of successes in a sample drawn without replacement from a finite population.

Outline

3.1 Bernoulli Distribution

Bernoulli Scheme can describe a series of coin tossing experiments, or the extraction of a ball from an urn with two colors, if there is replacement.

Definition 3: Bernoulli Scheme

A Bernoulli random variable is a random variable that takes the value 1 with probability p and the value 0 with probability $1-p$. We denote a Bernoulli random variable as $X \sim \text{Bern}(p)$.

We observe a Bernoulli Scheme when we have:

- A sequence of n independent trials.
- Each trial has two possible outcomes: success or failure.
- The probability of success of a single trial is constant (p).

$$x = \begin{cases} 0 & 1-p \\ 1 & p \end{cases}$$

$$X \sim \text{Bern}(p)$$

What is the expected value, and the variance?

$$\begin{aligned} E[X] &= 0q + 1p = p \\ E[X^2] &= 0^2q + 1^2p = p \\ Var[X] &= E[X^2] - E[X]^2 = p - p^2 = p(1-p) = pq \end{aligned} \tag{3.1}$$

If, in any probability space there is an event, A linked to an indicator function $\mathbb{1}$, then the indicator function of A is a Bernoulli random variable with parameter p.

$$\mathbb{1}_A \sim \text{Bern}(p)$$

3.2 Binomial Distribution

Definition 4: Binomial Distribution
The binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of Bernoulli trials.
The binomial distribution is characterized by two parameters:

- n - the number of trials.
- p - the probability of success in each trial.

We can define the random variable X as the number of successes in n bernoulli trials, with values $0 \leq k \leq n$. The probability of getting k successes in n trials is given by the binomial distribution.

$$X \sim \text{Bin}(n, p)$$

The probability density function of X is given by[1] [2]:

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \tag{3.2}$$

[1] The binomial coefficient $\binom{n}{k}$ is the number of ways to choose k successes in n trials. It is equivalent to writing:

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

[2]

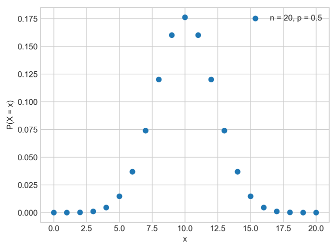


Figure 3.1: probability density function of a binomial random variable with p = 0.2 and n = 20.

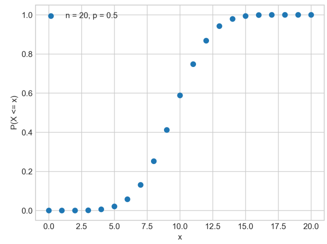


Figure 3.2: Cumulative distribution function.

What is the link between the binomial and the Bernoulli distribution?
Defining Y_i as the i^{th} Bernoulli random variable, we can say that the sum of n independent Bernoulli random variables is a binomial random variable.

$$X = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n, p) \tag{3.3}$$

Example 9.
Adam and Barbara are playing table tennis. In a single game, Barbara wins with probability of 0.6. If they play a series of 7 games, what is the probability that Barbara wins exactly 5 games? Assume that the outcome of each game is independent.

We can model the number of games Barbara wins as a binomial random variable $X \sim \text{Bin}(7, 0.6)$. The probability of Barbara winning exactly 5 games is given by:

$$P(X = 5) = \binom{7}{5} 0.6^5 0.4^2$$

Example 10.
In the previous setting, what is the probability that Barbara wins at least 5 games?

The probability of Barbara winning at least 5 games is given by TEST:

$$P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7)$$

Example 11.
In the previous setting, what is the probability of Barbara winning the last 2 games, given that Barbara won 4 and lost 3 games?

We need to find the probability that the last two games were wins given that Barbara has 4 wins and 3 losses in total.

The possible scenarios are:

- 2 wins, 3 losses in the first 5 games, then 2 wins in the last 2 games.
- 3 wins, 2 losses in the first 5 games, then 1 win and 1 loss in the last 2 games.
- 4 wins, 1 loss in the first 5 games, then 0 wins in the last 2 games.

The only favorable scenario is the first one.

Number of favorable outcomes (2 wins in the last three games): $\binom{5}{2} = 10$

Number of possible outcomes: $\binom{7}{4} = 35$

Therefore, the probability is:

$$P(\text{2 wins in the last 2 games} \mid \text{4 wins and 3 losses}) = \frac{10}{35}$$

3.3 Geometric Distribution

If X is the number of trials needed to get the first success in a sequence of Bernoulli trials.

$$P(X = k)$$

1. For the first success to occur on the k-th trial, the first $k-1$ trials must all result in failures. The probability of failure on each trial is $1-p$. Therefore, the probability that the first $k-1$ trials all fail is $(1-p)^{k-1}$.

2. The k-th trial must be a success. The probability of a success on this trial is p.

Hence, the probability that the first success occurs on the k-th trial is given by:

$$P(X = k) = (1-p)^{k-1}p$$

Definition 5: Geometric Distribution
The geometric distribution is a probability distribution that models the number of Bernoulli trials needed to get the first success.
The geometric distribution is characterized by a single parameter:

- p - the probability of success in each trial.

$$E[X] = \sum_{k=1}^{+\infty} k(q)^{k-1}p = \frac{1}{p} \qquad Var[X] = \frac{q}{p^2}$$

3.4 Negative Binomial

If we fix the number of successes desired, and we model the number of trials needed to get the desired number of successes, we get the negative binomial distribution.

Definition 6: Negative Binomial Distribution
The negative binomial distribution is a probability distribution that models the number of Bernoulli trials needed to get a fixed number of successes.
The negative binomial distribution is characterized by two parameters:

- n - the number of successes desired.
- p - the probability of success in each trial.

$X \sim \text{NegBin}(r, p)$

We have n successes in k trials, the number of failures is therefore $k-n$. Since the last trial must be a success, the number of possible arrangements of outcomes is $\binom{k-1}{n-1}$. We can then write the probability as:

$$P(X = k) = \binom{k-1}{n-1} p^n q^{k-n}$$

3.5 Hypergeometric Distribution

Definition 7: Hypergeometric Distribution
The hypergeometric distribution is a probability distribution that models the number of successes in a sample of size n drawn without replacement from a finite population of size N that contains K successes. It is characterized by three parameters:

- N - the population size.
- K - the number of successes in the population.
- n - the sample size.

$$X \sim \text{Hypergeom}(N, K, n)$$

The probability density function of X is given by:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Its expected value is:

$$E[X] = n\frac{K}{N}$$

4 Continuous Probability Distributions

Chapter abstract: This chapter is dedicated to continuous probability distributions. We start by introducing the Poisson distribution, which models the number of events occurring in a fixed interval of time or space. We then move on to the exponential distribution, which models the time until the first event occurs in a Poisson process. We also introduce the gamma distribution, which models the sum of n exponential random variables. Finally, we introduce the uniform distribution, which models the probability of all outcomes in an interval being equally likely.

Outline

4.1 Poission Distribution

This is the continuous time equivalent of a bernoulli random variable. It is used to model the number of events occurring in an interval of time or space.

An example is the number of phone calls received by a person. At any given time, the number of calls received is either zero or one, modeled by a Poisson distribution.

Fixing the time interval $[0,T]$, $X =$ number of events in $[0,T]$ is a Poisson random variable[3].

Definition 8: Poisson Distribution
The Poisson distribution is a probability distribution that models the number of events occurring in a fixed interval of time or space.
The Poisson distribution is characterized by a single parameter:

- λ - the average rate of events occurring in the interval

$$\lambda = \frac{\text{number of arrivals}}{\text{time interval}}$$

The probability of having 1 arrival in a time interval is given by:

$$P(X=1) \approx \lambda \Delta t$$

$$P(X=1) = \lambda \Delta t + o(\Delta t)$$

The probability density function of a Poisson random variable is given by:

$$P(X=k) = \frac{e^{-\lambda T}(\lambda T)^k}{k!}$$

3

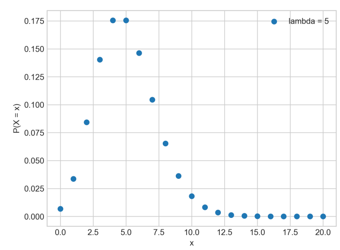


Figure 4.1: probability density function of a Poisson random variable with $\lambda = 5$.

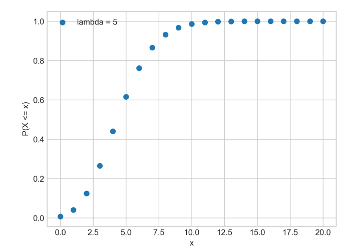


Figure 4.2: Cumulative distribution function.

4.2 Exponential Distribution

If X is the waiting time until the first event occurs in a Poisson process, then X is an exponential random variable

Definition 9: Exponential Distribution
The exponential distribution is a probability distribution that models the time until the first event occurs in a Poisson process.
The exponential distribution is characterized by a single parameter:

- λ - the rate of events occurring in the Poisson process.

The probability density function of an exponential random variable is given by [4]:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The expected value and variance of an exponential random variable are given by:

$$E[T] = \int_0^{+\infty} x f(x) dx = \frac{1}{\lambda} \qquad Var[T] = \frac{1}{\lambda^2}$$

The exponential distribution is memoryless, meaning that the probability of an event occurring in the next interval (waiting time) is independent of the time that has already passed.

4.3 Gamma Distribution

Definition 10: Gamma Function
The gamma function is defined as:

$$\Gamma(z) = \int_0^{+\infty} x^{z-1} e^{-x} dx$$

It is defined as long as the exponent z is positive.

Properties of the gamma function:

- $\Gamma(z+1) = z\Gamma(z)$
- $\Gamma(1) = 1$
- $\Gamma(n+1) = n!$

The most important property of the gamma function is its recursive definition[5]:

4

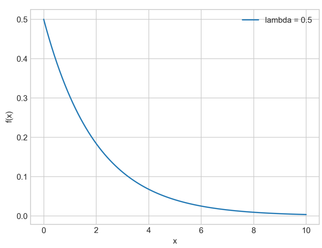


Figure 4.3: probability density function of an exponential random variable with $\lambda = 0.5$.

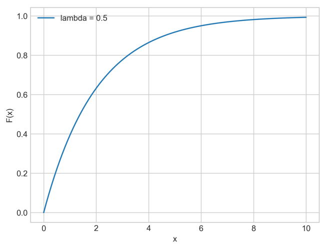


Figure 4.4: Cumulative distribution function.

5 Integration by part:

$$\int u dv = uv - \int v du$$

Proposition 1.

$$\Gamma(z+1) = z\Gamma(z)$$

Proof.

$$\begin{aligned}\Gamma(z+1) &= \int_0^{+\infty} x^z e^{-x} dx \\ &= [-x^z e^{-x}]_0^{+\infty} + z\int_0^{+\infty} x^{z-1} e^{-x} dx = z\Gamma(z)\end{aligned} \tag{4.1}$$

From this property, we can see that the gamma function is a generalization of the factorial function. For any positive integer n, we have $\Gamma(n+1) = n!$.

Starting from the gamma function, we can define the gamma distribution[6].

Definition 11: Gamma Distribution
The gamma distribution is a probability distribution that models the sum of n exponential random variables.
The gamma distribution is characterized by two parameters:

- n - the number of exponential random variables.
- λ - the rate of events occurring in the Poisson process.

$$X \sim \text{Ga}(\alpha, \lambda)$$

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Proposition 2.
The expected value of $X \sim \text{Ga}(\alpha, \lambda)$ is given by:

$$E[X] = \frac{\alpha}{\lambda}$$

Proof.

$$\begin{aligned}E[X] &= \int_0^{+\infty} x f(x) dx = \int_0^{+\infty} x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x}\end{aligned} \tag{4.2}$$

We can see that this is similar to the integral of the gamma function with $\alpha = \alpha + 1$, to make it equal, we need to multiply by $\frac{\lambda}{\lambda}$.

$$\begin{aligned}E[X] &= \frac{1}{\lambda} \frac{\overbrace{\Gamma(\alpha+1)}^{=\alpha\Gamma(\alpha)}}{\Gamma(\alpha)} \underbrace{\int_0^{+\infty} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx}_{=1} \\ &= \frac{\alpha}{\lambda}\end{aligned} \tag{4.3}$$

[6]

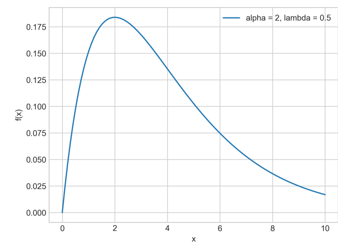


Figure 4.5: probability density function of a gamma random variable with $\alpha = 2$ and $\lambda = 0.5$.

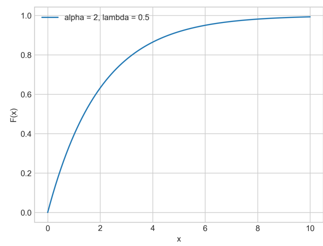


Figure 4.6: Cumulative distribution function.

4.4 Uniform Distribution

If all outcomes in an interval are equally likely, we have a uniform distribution. Its probability density function is given by:

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

The cumulative distribution function is:

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

The expected value and variance of a uniform random variable are given by:

$$E[X] = \frac{a+b}{2} \qquad Var[X] = \frac{(b-a)^2}{12}$$

4.5 Weibull Distribution

The Weibull distribution is a continuous probability distribution that is often used to model the lifetime of objects[7]. It is characterized by two parameters:

The scale parameter λ determines the location of the distribution, while the shape parameter α determines its shape.

$$X \sim \text{Weib}(\alpha, \lambda)$$

Its cumulative distribution function is given by:

$$F_X(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-\lambda t^\alpha} & t \geq 0 \end{cases}$$

Its probability density function is given by:

$$f(x) = \lambda \alpha x^{\alpha - 1} e^{-\lambda x^\alpha} \mathbb{1}_(0, +\infty)(x)$$

Remark 2.
The Weibull distribution is a generalization of the exponential distribution. When $\alpha = 1$, the Weibull distribution reduces to the exponential distribution.

[7]

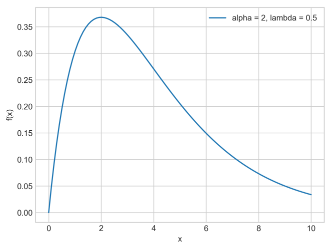


Figure 4.7: probability density function of a Weibull random variable with $\lambda = 0.5$ and $\alpha = 2$.

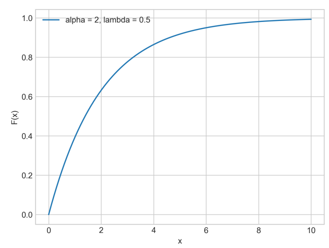


Figure 4.8: Cumulative distribution function.

4.6 Beta Distribution

X ≈ Beta(a, b), with a, b > 0

Moments

- $f_X(x)$:

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

with

$$c = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

- $E(X)$:

$$E(X) = \int_0^1 x\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}dx = E(X) = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a}(1-x)^{b-1}dx$$

As $X \sim \text{Beta}(\alpha+1, b)$:

$$f_X(x) = \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)}x^a(1-x)^{b-1} \quad \text{and} \quad F_X = \int f_X = 1:$$

Divide and multiply by $\Gamma(a+b+1)$ and $\Gamma(a+1)$:

$$\frac{\Gamma(a+1)}{\Gamma(a)} \cdot \frac{\Gamma(a+b)}{\Gamma(a+b+1)} \cdot \int_0^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)}x^a(1-x)^{b-1}dx = \frac{a}{a+b}$$

> **Remark 3.**
> Note that Uniform(0,1) is $X \sim \text{Beta}(1,1)$.

Properties

$X \sim \text{Beta}(a, 1)$, so:

$$f_X(x) = ax^{a-1}, \quad 0 < x < 1$$

Let $Y = -\ln(x)$, where $x = h(y) = e^{-y}$:

$$f_Y(y) = f_X(e^{-y})\,|-e^{-y}| = a(e^{-y})^{a-1}e^{-y} = ae^{-ay}, \quad 0 < y < +\infty$$

Thus, $Y \sim \text{Exp}(a)$.

4.7 Normal Distribution

$X \sim \text{Normal}(\mu, \sigma), \quad \mu, \sigma > 0$

Moments

- $f_X(x)$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- $E(X)$:

$$E(X) = \mu$$

- $Var(X)$:

$$\text{Var}(X) = \sigma^2$$

Properties

Given $X \sim$ Normal(μ, σ), if $Y = aX + b$, then:

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

This is the only continuous random variable with this property.

4.8 Standard Normal Distribution

X ≈ Standard Normal(0, 1), with a, b > 0

Assumptions

The Standard Normal distribution is the result of the standardization of a normal distribution through the transformation:

$$z = \frac{x - \mu}{\sigma}$$

Moments

- $f_Z(z)$:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

- $F_Z(z)$:

$$F_Z(z) = \Phi(z)$$

Properties

Given $Z \sim N(0,1)$, if $Y = Z^2$ (note: the square of $N(0,1)$ is a chi-square distribution with 1 degree of freedom, and the chi-square distribution is a particular case of the Gamma distribution with parameters $\frac{1}{2}$, $\frac{1}{2}$):

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ *, & y > 0 \end{cases}$$

$$F(y \leq t) = P(y \leq t) = P(z^2 = t) = P(-\sqrt{t} \leq z \leq \sqrt{t}) = \Phi(\sqrt{t}) - \Phi(-\sqrt{t})$$

Since $\Phi(-\sqrt{t}) = 1 - \Phi(\sqrt{t})$, we get:

$$2\Phi(\sqrt{t}) - 1$$

The probability density function of Y is:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(2\Phi(\sqrt{y}) - 1)$$

Using the chain rule:

$$f_Y(y) = 2 \cdot \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{y}{2}\right) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} \cdot \exp(-\frac{1}{2}y)$$

This is a Gamma distribution with parameters $\frac{1}{2}$, $\frac{1}{2}$.

5 Expected Values and Variances

Chapter abstract:

Outline

5.1 Expected Value of a Linear Transformation

$$E(aX+b) = aE(X) + b$$

This property shows that the expectation of a linear transformation of X follows this simple rule, where a is a constant multiplier and b is a constant shift.

However, for non-linear transformations, the expectation does not necessarily follow through the transformation:

$$E(g(X)) \neq g(E(X))$$

This is incorrect in general. The only special case where this holds is for a constant function, or where $g(x)$ is a linear transformation. For instance, $E(1/X)$ is not $1/E(X)$.

5.2 Expected Value of a Function

You can find the expected value of a function $g(X)$ as:

$$E(Y) = E(g(X))$$

This can be computed without needing the explicit distribution of Y. The method differs for discrete and continuous random variables:

- For discrete variables:

$$E(g(X)) = \sum_i g(x_i)P(x_i)$$

- For continuous variables:

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

Example 12.
Expected value of a function with $P(X=k)=\frac{1}{n}, \quad 1 \leq k \leq n$: Given X number of trials and $P(X=k)=\frac{1}{n}$:

$$E(x)=1\cdot\frac{1}{n}\cdots+n\frac{1}{n}=\frac{1}{n}\cdot(1+2+3)=\frac{n+1}{2}$$

Example 13.
Example: Expected Value of X^2 For example, given X, the number of trials, and $P(X=k)=\frac{1}{n}, \quad 1 \leq k \leq n$, the expected value can be computed as:

$$E(X^2)=\sum_i x_i^2 P(x_i) \quad \text{or} \quad E(X^2)=\int_{-\infty}^{\infty} x^2 f_X(x)dx$$

5.3 Variance of a Random Variable

The variance of a random variable X is:

$$\text{Var}(X)=E[(X-E(E))^2]=E[(X-\mu)^2]=E(X^2-2\mu X+\mu^2)$$

$$E(X^2)-2\mu E(X)+\mu^2=E(X^2)-\mu^2=E(X^2)-[E(X)]^2$$

Chebishev Inequality: For any R.V. P()

$$P(|X-E(X)|\geq\eta)\leq\frac{\text{Var}(X)}{\eta^2}, then for \eta=k*\sigma \quad P(|X-E(X)|\geq k\sigma)\leq\frac{1}{k^2}$$

This shows that variance is the difference between the expected value of X^2 and the square of the expected value of X.

5.4 Variance of a Linear Transformation

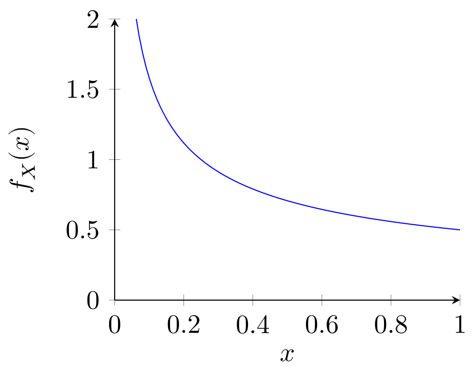
For a linear transformation $aX+b$, the variance follows the rule:

$$\text{Var}(aX+b)=a^2\text{Var}(X)$$

The constant b shifts the expected value but does not affect the variance.

5.5 Example 1

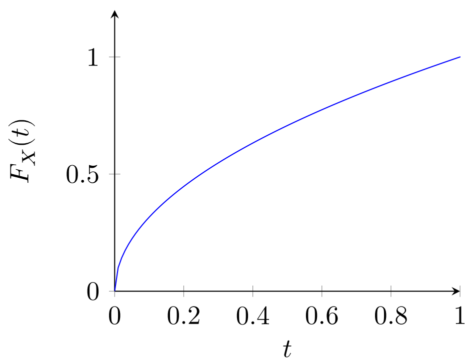
$$f_X(x)=\begin{cases}\frac{1}{2\sqrt{x}} & 0\leq x\leq 1\\ 0 & \text{otherwise}\end{cases}$$



This is a valid density function because π is greater than 0 and the integral from 0 to 1 is 1.

The cumulative distribution function is:

$$F_X(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \sqrt{t} & \text{if } 0 < t < 1 \\ 1 & \text{if } t \geq 1 \end{cases}$$



The expected value is:

$$E(x) = \int_0^1 x \cdot f_X(x)dx = \int_0^1 x \cdot \frac{1}{2\sqrt{x}}dx = \int_0^1 \frac{1}{2} \cdot \sqrt{x}dx = [\frac{1}{2} \cdot \frac{2}{3} \cdot x^{3/2}]_0^1 = \frac{1}{3}$$

The expected value of X^2 is:

$$E(X^2) = \int_0^1 x^2 \cdot f_X(x)dx = \int_0^1 x^2 \cdot \frac{1}{2\sqrt{x}}dx = \int_0^1 \frac{1}{2} \cdot x^{3/2}dx = [\frac{1}{2} \cdot \frac{2}{5} \cdot x^{5/2}]_0^1 = \frac{1}{5}$$

The variance is:

$$Var(X) = \frac{1}{5} - (\frac{1}{3})^2 = \frac{4}{45}$$

5.6 Deriving the Marginal Distribution Function of Y

Given X and $Y = g(X)$, if g is differentiable and invertible (meaning $y = g(x) \leftrightarrow x = h(y)$), then

$$f_Y(y) = f_X(h(y)) \cdot |h'(y)|$$

5.7 Example 1

Given the probability density function of X as:

$$f_X(x) = \begin{cases} \frac{x-1}{2} & \text{if } 1 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Then using the indicator function:

$$f_X(x) = \frac{x-1}{2} \cdot \mathbb{1}_{[1,3]}(x)$$

Let's consider the transformation $Y = \ln(X)$. It is invertible and differentiable with $x = h(y) = e^y$. Applying the aforementioned formula, we get:

$$f_Y(y) = f_X(h(y)) \cdot |h'(y)| = \frac{e^y - 1}{2} \cdot \mathrm{I}_{[\ln(1), \ln(3)]}(y) \cdot e^y \cdot |e^y| = \frac{e^y(e^y-1)}{2} \cdot \mathrm{I}_{[\ln(1), \ln(3)]}(y)$$

Let's consider the transformation $Y = X^2$. Generally not invertible, but it is on the support 1-3 such that $x = h(y) = \sqrt{y}$.

$$f_Y(y) = f_X(\sqrt{y}) \cdot \left| \frac{1}{2\sqrt{y}} \right| = \frac{\sqrt{y}-1}{2} \cdot \mathrm{I}_{[1,9]}(y) \cdot \sqrt{y} \cdot \frac{1}{2\sqrt{y}} = \frac{\sqrt{y}-1}{4\sqrt{y}} \cdot \mathrm{I}_{[1,9]}(y)$$

5.8 Deriving the Cumulative Distribution Function of Y

$$F_y(t) = P(Y \le t) = P(g(X) \le t) - P(g(X) \le t) = P(X \in At)$$

5.9 Example 2

$$F_x(t) = \begin{cases} 0 & \text{if } t \le 1 \\ \frac{(t-1)^2}{4} & \text{if } 1 < t < 3 \\ 1 & \text{if } t \ge 3 \end{cases}$$

If Y = X^2, then the CDF of Y is:

$$F_Y(t) = P(Y \le t) = P(X^2 \le t) = P(X \le \sqrt{t})$$

Two cases:

$$\begin{cases} 0 & \text{if } t \le 0 \\ P(-\sqrt{t} \le x \le \sqrt{t}) & \text{if } t > 0 \end{cases}$$

$$F_X(\sqrt{t}) - F_X(-\sqrt{t}) = \begin{cases} 0, & \text{if } \sqrt{t} < 1 \\ \left(\frac{\sqrt{t}-1}{4}\right)^2, & \text{if } 1 \le \sqrt{t} < 3 \\ 1, & \text{if } \sqrt{t} \ge 93 \end{cases} = \begin{cases} 0, & \text{if } t < 1 \\ \left(\frac{\sqrt{t}-1}{4}\right)^2, & \text{if } 1 < t < 9 \\ 1, & \text{if } t \ge 9 \end{cases}$$

5.10 Example 3

$$f_X = \frac{x^2}{18} \cdot \mathbb{1}_{[-3,3]}(x)$$

$y = x^2$: Not invertible, use formula for F_X.

$$F_X(t) = \begin{cases} 0 & \text{if } t < -3 \\ \frac{t^2+27}{54} & \text{if } -3 \leq t \leq 3 \\ 0 & \text{if } t > 3 \end{cases} \rightarrow F_Y(t) = \begin{cases} 0 & \text{if } t < 0 \\ * & \text{if } 0 \leq t \leq 9 \\ 1 & \text{if } t > 9 \end{cases}$$

Clearly, being:

$$F_Y(t) = P(y \leq t) = P(x^2 \leq t) = P(-\sqrt{t} \leq x \leq \sqrt{t}) =$$

$$F_X(\sqrt{t}) - F_X(-\sqrt{t}) = \frac{\sqrt{t}^2 + 27}{54} - \frac{-\sqrt{t}^2 + 27}{54} = \frac{t}{27}$$

Then:

$$* = f_Y = \frac{dF_Y}{dt} = \frac{1}{27} \cdot I_{[0,9]}(y)$$

6 Conditional Probability and Sequences of R.V.s

Chapter abstract: In day 6, we discuss the concept of conditional probability distributions in the context of discrete and continuous random variables. We then introduce the concept of sample mean and variance of a set of random variables. Finally, we discuss the concept sequences of random variables and their convergence.

Outline

6.1 Exercise

$$X \sim \text{Gamma}(\alpha, \lambda) \qquad Y \sim \text{Gamma}(\beta, \lambda)$$

$$\begin{cases} V = \frac{x}{y} \\ W = X + Y \end{cases}$$

$$\begin{cases} V = \frac{x}{y} \\ W = Y(1+V) \end{cases}$$

$$\begin{cases} X = \frac{VW}{1+V} \\ Y = \frac{W}{1+V} \end{cases}$$

$$J = det \begin{vmatrix} \frac{W(1+V)-vw}{(1+V)^2} & \frac{v}{1+v} \\ -\frac{w}{(1+v)^2} & \frac{1}{1+v} \end{vmatrix} = \frac{w}{(1+v)^2}$$

The density of X and Y is given by:

$$f_{X,Y}(x,y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \frac{\lambda^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y}$$

X and Y are two independent random variables, therefore the joint distribution of V and W is given by:

$$f_{V,W}(v,w) = f_{X,Y}(x,y)\,|J| = f_X\left(\frac{vw}{1+v}\right) f_Y\left(\frac{w}{1+v}\right) \frac{w}{(1+v)^2}$$

$$f_{V,W}(v,w) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} (\frac{vw}{1+v})^{\alpha-1} e^{-\lambda \frac{vw}{1+v}} (\frac{w}{1+v})^{\beta-1} e^{-\lambda \frac{w}{1+v}} \frac{w}{(1+v)^2}$$

$$f_{V,W}(v,w) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \frac{v^{\alpha-1}}{(1+v)^{\alpha+\beta}} w^{\alpha+\beta-1} e^{-\lambda w} \mathbb{1}_{(0,+\infty)}(v) \mathbb{1}_{(0,+\infty)}(w)$$

If the joint density is the product of two functions, then the two random variables are independent.

The joint density is therefore

$$f_{V,W}(v,w) = \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \frac{v^{\alpha-1}}{(1+v)^{\alpha+\beta}} \frac{\lambda^{\alpha+\beta}}{(1+v)^(\alpha+\beta)} e^{-\lambda w} \mathbb{1}_{(0,+\infty)}(v) \mathbb{1}_{(0,+\infty)}(w)$$

6.2 Conditional Distributions

Discrete Case

Consider two random variables X and Y with the following joint distribution:

	1	2	3
0	0.1	0.2	0.1
1	0.2	0.1	0.3

We can calculate the conditional distribution[8] of Y given X:

[8] **Conditional Distribution:** the probability distribution of a random variable, calculated according to the rules of conditional probability after observing the realization of another random variable.

$$P(Y=1|X=0) = \frac{P(Y=1, X=0)}{P(X=0)} = \frac{0.1}{0.4}$$

More in general:

$$P_{Y|X}(y_i|x_i) = \frac{P_{XY}(y_i, x_i)}{P_X(x_i)}$$

Expected values can be evaluated in the same way:

$$E(Y|X=0) = 1 \cdot 0.25 + 2 \cdot 0.5 + 3 \cdot 0.25 = 2$$

The same can be done for the variance:

$$Var(Y|X=0) = E(Y^2|X=0) - E(Y|X=0)^2$$

$$E(Y^2|X=0) = 1^2 \cdot 0.25 + 2^2 \cdot 0.5 + 3^2 \cdot 0.25 = 4.5$$

$$Var(Y|X=0) = 4.5 - 2^2 = 0.5$$

Continuous Case

$$f_{XY}(x,y) = \begin{cases} \frac{15}{8} xy^2 & (x,y) \in T \\ 0 & \text{otherwise} \end{cases} = \frac{15}{8} \mathbb{1}_{(0,1)}(x) \mathbb{1}_{(0,2x)}(y)$$

Can we construct the conditional distribution of X and Y? In this case, we cannot use the formula $P_{Y|X}(y_i|x_i) = \frac{P_{XY}(y_i,x_i)}{P_X(x_i)}$ because the probability of X is zero. We can, however, use the formula for the continuous case:

$$f_{Y|X}(y|x) = \frac{\frac{15}{8} xy^2 \mathbb{1}_{(0,1)}(x) \mathbb{1}_{(0,2x)}(y)}{5x^4 \mathbb{1}_{(0,1)}(x)} = \frac{3}{8} \frac{y^2}{x^3} \mathbb{1}_{(0,2x)}(y)$$

We can now calculate the expected value of Y given X, as the integral of y times the conditional density of Y given X:

$$E(Y|X=x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy = \int_{-\infty}^{+\infty} y \frac{3}{8} \frac{y^2}{x^3} \mathbb{1}_{(0,2x)}(y) dy$$

$$E(Y|X=x)=\int_0^{2x}\frac{3}{8}y^3x^{-3}dy=\frac{3}{8}x^{-3}\frac{y^4}{4}\Big|_0^{2x}=\frac{3}{8}x^{-3}\frac{16x^4}{4}=\frac{3}{2}x$$

The variance can be calculated in the same way:

$$Var(Y|X=x)=E(Y^2|X=x)-E(Y|X=x)^2$$

$$E(Y^2|X=x)=\int_{-\infty}^{+\infty}y^2f_{Y|X}(y|x)dy=\int_0^{2x}\frac{3}{8}y^4x^{-3}dy$$

$$=\frac{3}{8}x^{-3}\frac{y^5}{5}\Big|_0^{2x}=\frac{12}{5}x^2$$

So the variance is:

$$Var(Y|X=x)=\frac{12}{5}x^2-\left(\frac{3}{2}x\right)^2=\frac{3}{20}x^2$$

Remark 4.
If, in the discrete or continuous case, you construct the conditional distribution of Y given X, the expected value and variance of Y given X are functions of X. This holds true unless X and Y are independent.

Example 14.
Let's go back to the previous example:

	1	2	3
0	0.1	0.2	0.1
1	0.2	0.1	0.3

We know that $E(Y|X=0)=2$, what is $E(Y|X=1)$?

$$E(Y|X=1)=1\cdot\frac{0.2}{0.6}+2\cdot\frac{0.1}{0.6}+3\cdot\frac{0.3}{0.6}=\frac{13}{6}$$

As in the continuous case, the expected value of Y given X is a function of X.

$$E(X|Y=x)=h(x)=\begin{cases}2 & x=0\\ \frac{13}{6} & x=1\end{cases}$$

General case

$$E(Y|X)=h(X)\leftarrow \text{random variable}$$

In this case, there is no $X=x$ in the conditional expectation, so we need to calculate the expected value of Y given X as a function of X, not of x.

So, in the example above:

$$E(Y|X)=\frac{3}{2}X$$

Definition 12: Conditional Expectation
The conditional expectation of Y given X is a random variable $h(X)$. (a function of X)

$$E(Y|X) = h(X)$$

Properties:
- $E(E(Y|X)) = E(Y)$ (Tower Property)
- $E(Yg(X)|X) = g(X)E(Y|X)$
- $Var(Y|X)$ is a r.v.
- $Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$

Example 15.
Suppose that Y is a random variable "duration of battery", and X is the r.v. "percentage of an element"

$$X \sim \text{Uniform}(1,3) \qquad (Y|X=x) \sim \text{Exp}(\lambda = x)$$

What is the average duration of the battery? i.e. $E(Y)$

$$E(Y|X=x) = \frac{1}{x} \qquad E(Y|X) = \frac{1}{X}$$

Therefore, we can use the Tower Property:

$$E(Y) = E(E(Y|X)) = E\left[\frac{1}{X}\right] = \int_1^3 \frac{1}{x}\frac{1}{2}dx = \frac{1}{2}\int_1^3 \frac{1}{x}dx = \frac{1}{2}\ln(3)$$

Example 16.
The duration of a call is: $T_1 \sim \text{Exp}(\lambda = \frac{1}{2})$
The number of calls is: $N \sim \text{Poisson}(\lambda = 60)$
The total time spent on calls is therefore:

$$Y = \sum_{i=1}^{N} T_i$$

What is the expected value and variance of the total time spent on calls?

$$(Y|N=n) = \sum_{i=1}^{n} T_i \sim \text{Gamma}(n, \frac{1}{2})$$

$$E(Y|N=n) = E(Ga(n, \frac{1}{2})) = \frac{n}{\lambda} = 2n$$

$$E(Y) = E(E(Y|N)) = E(2N) = 2E(N) = 2 \cdot 60 = 120$$

Calculating the variance:

$$Var(Y|N=n) = n \cdot \frac{1}{\lambda^2} = 4n$$

$$Var(Y) = Var(E(Y|N)) + E(Var(Y|N))$$

$$Var(Y) = Var(2N) + E(4N) = 4Var(N) + 4E(N) = 4 \cdot 60 + 4 \cdot 60 = 480$$

Example 17.
Suppose $X_1, \dots, X_n$ are independent and identically distributed random variables with $X_n \sim$ Bern(p).
Let $Y = \sum_{i=1}^{n} X_i$.
What is $E(X_1|Y)$?

The distribution of Y is a binomial: $Y \sim$ Bin(n,p).

$$E(X_1|Y=k) = 0 \cdot P(X_1 = 0|Y=k) + 1 \cdot P(X_1 = 1|Y=k)$$

$$E(X_1|Y=k) = P(X_1 = 1|Y=k) = \frac{P(X_1 = 1, Y = k)}{P(Y=k)}$$

The possible values of k are $0, 1, \dots, n$, so:

$$E(X_1|Y=k) = \begin{cases} 0 & k = 0 \\ ? & k = 1, \dots, n \end{cases}$$

We can rewrite the conditional expectation as:

$$E\left(X_1 | \sum_{i=1}^{n} X_i = k\right) = \frac{P(X_1 = 1, \sum_{i=1}^{n} X_i = k)}{P(\sum_{i=1}^{n} X_i = k)}$$

The two events in the numerator are not independent, so we have to rewrite it to solve the problem. The probability of $X_1 = 1$ and $\sum_{i=1}^{n} X_i = k$ is the same as the probability of $X_1 = 1$ and $X_2 + \dots + X_n = k-1$:

$$P(X_1 = 1, \sum_{i=1}^{n} X_i = k) = \frac{P(X_1 = 1, \sum_{i=2}^{n} X_i = k-1)}{P(\sum_{i=1}^{n} X_i = k)}$$

$$= \frac{P(X_1 = 1)P(\sum_{i=2}^{n} X_i = k-1)}{P(\sum_{i=1}^{n} X_i = k)} = \frac{p \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k}}{\binom{n}{k} p^k q^{n-k}}$$

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

The final result is therefore:

$$E(X_1|Y=k) = \begin{cases} 0 & k = 0 \\ \frac{k}{n} & k = 1, \dots, n \end{cases} = \frac{k}{n}$$

Therefore:

$$E(X_1|Y) = \frac{Y}{n}$$

6.3 Sample Mean and Variance

Take $X_1, \dots, X_n$ i.i.d. We can define the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Defining $m = E(X_1)$, we can calculate the expected value of the sample mean:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\cdot n\cdot m = m$$

With the variance of X_1 defined as v, we can calculate the variance of the sample mean:

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\cdot n\cdot v = \frac{v}{n}$$

This holds true for any sampling distribution.

Sample Variance

The sample variance is defined differently, depending on wether m is known or not.

If m is known, the sample variance is:

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2m\sum_{i=1}^{n} X_i + nm^2$$

The expected value of the sample variance is $E(S_0^2) = v$.

If m is unknown, the sample variance is:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

This can be rewritten as:

$$S_n^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \underbrace{2\bar{X}_n\sum_{i=1}^{n} X_i}_{=2\bar{X}_n^2} + n\bar{X}_n^{\,2}\right]$$

$$S_n^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^{\,2}\right]$$

Case of a Normal Distribution

Take $X_1, \dots, X_n$ i.i.d. with $X_i \sim \mathrm{N}(\mu, \sigma^2)$. Then the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is normally distributed with: $\bar{X}_n \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$.

The sample variance in the case where m is known is distributed as:

$$\frac{nS_0^2}{\sigma^2} \sim \chi^2(n)$$

In the case where m is unknown, the sample variance is distributed as:

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

6.4 Sequence of R.V.s

Definition 13: Limit of a Sequence
In a sequence of real numbers $a_1, a_2, \dots, a_n$, the limit is defined as:

$$\lim_{n\to\infty} a_n = L \in \mathbb{R} \quad \forall \epsilon > 0 \quad \exists n_\epsilon \in \mathbb{N} \quad | \quad \forall n \geq n_\epsilon \implies |a_n - L| < \epsilon$$

Intuitively, the limit exists if there exists an n large enough so that after that n all the terms are within ϵ distance of the limit. (dist$(a_n, L) < \epsilon$)

Limits can be defined in spaces other than $\mathbb{R}$, as long as there is a way to define the distance between two elements.

Sequence of Random Variables

Take a sequence of random variables $X_1, X_2, \dots, X_n, \dots$. Suppose that all the random variables are defined on the same probability space $(\Omega, \mathcal{A}, P)$.

Definition 14: Sure Convergence
We say that $X_n \to Y$ surely if:

$$\forall \omega \in \Omega \quad X_n(\omega) \to Y(\omega)$$

In other words, for every sequence of outcomes $X_n(\omega)$, the limit of the sequence is $X(\omega)$.

This definition is very strong, and is not very useful in practice.

Definition 15: Almost Sure Convergence
We say that $X_n \to Y$ almost surely if:

$$P(\{\omega \in \Omega | X_n(\omega) \to Y(\omega)\}) = 1$$

In other words, the set of outcomes for which the sequence of random variables converges to the limit has probability 1.

Property:

$$g: \mathbb{R} \to \mathbb{R} \quad \text{continuous} \quad \implies \quad g(X_n) \to g(Y) \quad \text{almost surely}$$

Definition 16: Convergence in Probability
We say that $X_n \to Y$ in probability if:

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

In other words, the probability that the distance between X_n and Y is less than ϵ converges to 1 as n goes to infinity.

Definition 17: Convergence in Mean of Order k
We say that $X_n \to Y$ in mean of order $k \geq 1$ if:

$$E(|X_n - Y|^k) \to 0$$

Definition 18: Convergence in Quadratic Mean
We say that $X_n \to Y$ in quadratic mean if:

$$E(|X_n - Y|^2) \to 0$$

From a sequence of random variables, we can define the sequence of sample means:

$$\bar{X}_1 = X_1 \quad \bar{X}_2 = \frac{X_1 + X_2}{2} \quad \bar{X}_3 = \frac{X_1 + X_2 + X_3}{3} \quad \dots$$

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

We can demonstrate that the sample mean converges to the expected value of the random variable:

$$Var(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n \cdot v = \frac{v}{n}$$

$$E(|\bar{X}_n - m|^2) = Var(\bar{X}_n) = \frac{v}{n} \to 0$$

Definition 19: Strong Law of Large Numbers
If $X_1, X_2, \dots, X_n$ are i.i.d. with $E(X_i) = m$ and $Var(X_i) = v$, then:

$$\bar{X}_n \to m \quad \text{almost surely}$$

If $X_n \to Y$ almost surely, then $X_n \to Y$ in probability. The converse is not true.

$$X_n \to Y \text{ almost surely} \quad \underset{\not\Leftarrow}{\Rightarrow} \quad X_n \to Y \text{ in probability}$$

Moreover, if $X_n \to Y$ in order k, then $X_n \to Y$ in probability. The converse is again not true.

$$X_n \to Y \text{ in order } k \quad \underset{\not\Leftarrow}{\Rightarrow} \quad X_n \to Y \text{ in probability}$$

Convergence in Distribution

Definition 20: Convergence in Distribution
We say that $X_n \to Y$ in distribution if:

$$F_{X_n}(t) \to F_Y(t) \quad \forall t \quad \text{where} \quad F_Y \quad \text{is continuous}$$

Convergence in distribution is a weaker form of convergence than other forms of convergence. Link between convergence in distribution and convergence in probability:

$$X_n \to Y \text{ in probability} \quad \Rightarrow \quad X_n \to Y \text{ in distribution}$$

The converse is generally not true. However, if Y is a constant, then the two forms of convergence are equivalent.

$$X_n \to y \in \mathbb{R} \text{ in distribution} \quad \Leftrightarrow \quad X_n \to y \in \mathbb{R} \text{ in probability}$$

Example 18.
Take a sequence of i.i.d. random variables $X_1, X_2, \dots, X_n, \dots$ with:

$$X_i \sim \text{Unif}(0,1)$$

We take the sequence $Y_n = \min(X_1, \dots, X_n)$. What is the limit of Y_n?

Recall that if $V_n = \min(X_1, \dots, X_n)$, then $F_{V_n}(t) = 1 - [1 - F_x(t)]^n$.

Since the distribution function of X is:

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

The distribution function of V_n is:

$$F_{V_n}(t) = \begin{cases} 0 & t < 0 \\ 1 - (1-t)^n & 0 \le t \le 1 \\ 1 & t > 1 \end{cases}$$

For the convergence in distribution, we need to calculate the limit of $F_{V_n}(t)$ as n goes to infinity:

$$\lim_{n \to \infty} F_{V_n}(t) = \begin{cases} 0 & t \le 0 \\ 1 & t > 0 \end{cases}$$

This is not the distribution function of any random variable, because it has a jump at 0, and in that jump the value is continuous from the left but not from the right.

This is the distribution of a discrete random variable, as it is not continuous but piecewise constant. From the definition oif the limit of a sequence of random variables, we can define the limit of $F_{V_n}(t)$ as:

$$F_Y(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$$

Which is the distribution function of a random variable Y that is equal to 0 with probability 1.

6.5 Central Limit Theorem

Definition 21: Central Limit Theorem
Let $X_1, \dots, X_n$ be i.i.d random variables with $m = E(X_i)$ and $v = Var(X_i)$.

$$(P)\left(\frac{\bar{X}_n - m}{\sqrt{\frac{v}{n}}} \leq t\right) \to \Phi(t) \quad \text{as} \quad n \to +\infty$$

or

$$\frac{\bar{X}_n - m}{\sqrt{\frac{v}{n}}} \to \mathrm{N}(0,1) \quad \text{in distribution}$$

For this to hold, we have to assume that the random variables have finite mean and variance. In other words, $E(X_i^2) < +\infty$.

In other words, if n is "large enough", then the distribution of the sample mean is approximately normal with $\bar{X}_n \approx N(m, \frac{v}{n})$.

If we multiply by n, we get:

$$\sum_{i=1}^{n} X_i \approx N(nm, nv)$$

Take $Y \sim \text{Bin}(n,p)$, then if n is large enough and p is not too close to 0 or 1, then $Y \approx N(np, npq)$.

With a binomial distribution $Y \sim \text{Bin}(n,p)$, if $n \to +\infty$ and $p \to 0$ such that $np \to \lambda$, then $Y \approx \text{Po}(\lambda)$. In other words, $X_n \to P(\lambda)$ in distribution.

Poisson random variables are also approximately normal if λ is large enough.[9]

In the case of Gamma distributions $Y \sim \text{Ga}(\alpha, \lambda)$, if α is large enough, then $Y \approx N(\alpha/\lambda, \alpha/\lambda^2)$.[10]

For a Chi-squared distribution with n degrees of freedom, if n is large enough, then $X \approx N(n, 2n)$.[11]

[9] The Poisson distribution with parameter λ is the sum of λ Poisson random variables with parameter 1.

[10] The Gamma distribution with parameters α and λ is the sum of α exponential random variables with parameter λ.

[11] The Chi-squared distribution with n degrees of freedom is the sum of n standard normal random variables squared (χ^2 with 1 d.f.).

Definition 22: Slutsky's Theorem
Let $X_n \to X$ in distribution. If $Y_n \to Y$ in probability, then:

$$X_n + Y_n \to X + Y \quad \text{in distribution}$$

$$X_n \cdot Y_n \to X \cdot Y \quad \text{in distribution}$$

$\forall n \quad P(Y_n = 0) = 0$ and $y \neq 0$.

Sampling from a Normal Distribution

Take $X_1, X_2, \ldots, X_n$ i.i.d. with $X_i \sim N(\mu, \sigma^2)$.

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}} = \underbrace{\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}}_{\sim N(0,1)} \underbrace{\sqrt{\frac{\sigma^2}{S_n^2}}}_{\to 1} \to Z \sim N(0,1) \text{in distribution}$$

Therefore, the distribution is approximately $N(0,1)$.

Sampling not from a Normal Distribution If $m = E(X_i)$ and $v = Var(X_i)$, then:

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}} = \underbrace{\frac{\bar{X}_n - \mu}{\sqrt{\frac{v}{n}}}}_{\sim Z} \underbrace{\sqrt{\frac{v}{S_n^2}}}_{\to 1} \to Z \sim N(0,1) \text{in distribution}$$