Computational Statistics II

Bhashkar Paudyal

June 5, 2025

Abstract

This course was taught by **Prof. Dr Dhiraj Giri** . This is a continuation of the course **Computational Statistics I**.

Contents

Chapter 1

Random Variable

1.1 Introduction to Random Variable

Definition 1.1.1 (Random Experiment). A random experiment in probability refers to any well-defined procedure that produces uncertain outcomes when repeated under identical conditions.

Definition 1.1.2 (Sample Space). Sample Space denoted by Ω is the set of all possible event denoted by ξ.

Definition 1.1.3 (Random Variable). Random Variable denoted by X is a function that maps the elements in Ω i.e the **Domain of Random Variable** denoted by:
$D_X = \{\xi \in \Omega | X(\xi) = x_k, x_k \in R_X\}$ to a number in $\mathbb{R}$. It can assume many values and these values that a Random Variable can assume are in the **Range of the Random Variable** denoted by:
$R_X = \{x_1, x_2, x_3, \ldots | x_k = X(\xi_k)\}$. So, the inverse function $X^{-1}(x_k) = \xi_k$.

A random variable may be continuous or discrete.

Note. # means no. of something.

1.2 Cumulative Distribution Function

Definition 1.2.1 (Cumulative Distribution Function (CDF)). CDF denoted by F_X is given by:

$$\forall x \in \mathbb{R},\ F_X(x) = P[X \leq x] \tag{1.1}$$

It is an increasing function i.e, $x < y \Leftrightarrow F_X(x) < F_X(y)$.

1.3 Discrete Random Variable

For a Discrete Random Variable X, R_X is a finite or countable infinite set.

Definition 1.3.1 (Probability Mass Function (PMF)). PMF denoted by P_X is given by:

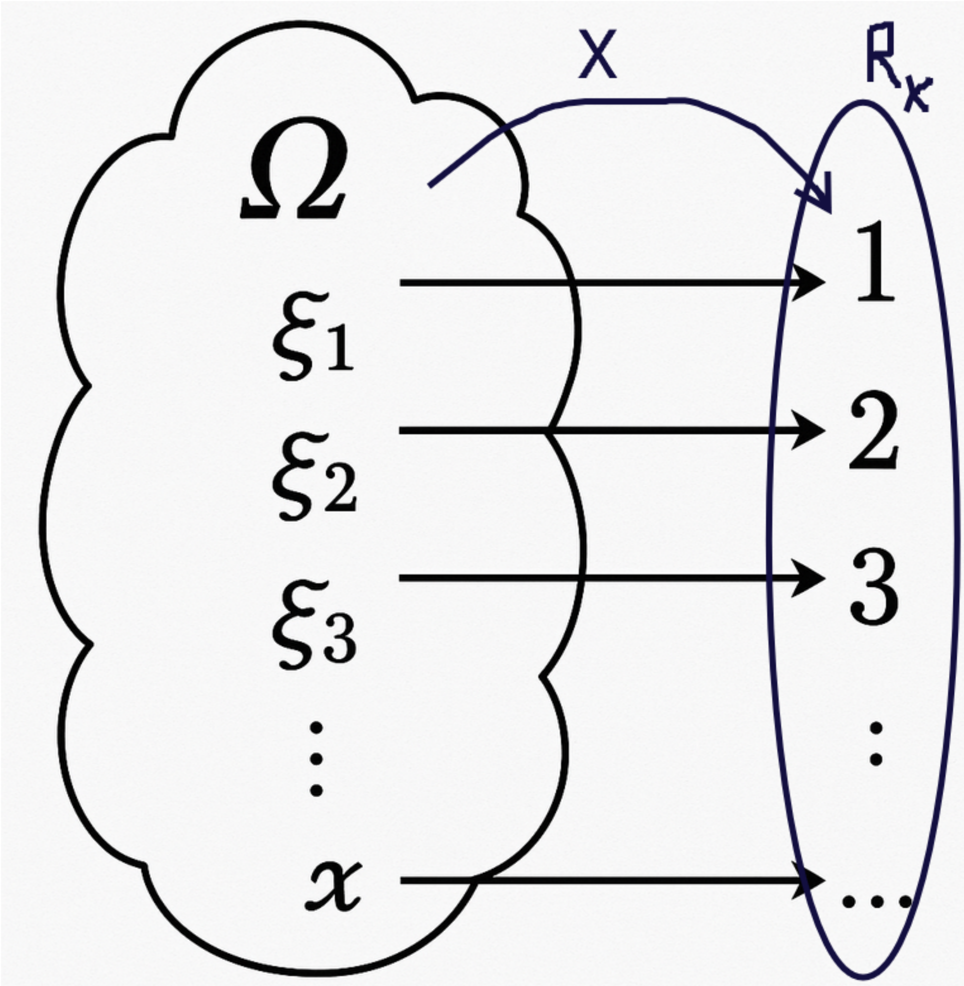

Figure 1.1: Random Variable Sample Space Mapping

$$P_X(x) = P[X = x] \tag{1.2}$$

It gives the probability that X assumes the value x.

The implication of this is that the CDF for Discrete Random Variable is given by:

$$F_X(x) = P[X \leq x_n] = \sum_{k=1}^{n} P_X(x_k) = \sum_{k=1}^{n} P[X = x_k] \tag{1.3}$$

Then,

Corollary 1.3.1 ($P(a < X \leq b) = F_X(b) - F_X(a)$). |
Here,
Take the equation from probability addition theorem,

$$P(X \leq a) + P(a < X \leq b) = P(X \leq b) \tag{1.4}$$

$$\text{or, } P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

$$\text{or, } P(a < X \leq b) = F_X(b) - F_X(a)$$

From Above, one can prove the following:

1. $P(a \leq X \leq b) = F_X(b) - F_X(a) + P[X = a]$
2. $P(a \leq X < b) = F_X(b) - F_X(a) + P[X = a] - P[X = b]$
3. $P(a < X < b) = F_X(b) - F_X(a) - P[X = b]$

1.4 Continuous Random Variable

For a Continuous Random Variable X, R_X is an uncountable infinite set. The probability at a point is 0 i.e:
$P(X = x) = 0$.
For continuous random variable, the probability can only be calculated at a given interval. The probability:

$$P[a < x < b] = \int_a^b f_X(x)dx \tag{1.5}$$

where,
$f_X(x)$ is the probability density function such that:

$$f_X(x) = F'_X(x) \tag{1.6}$$

1.5 Expectation of a Random Variable

The Expectation of a Random Variable denoted by $E(X)$ is the expected value of X that is most likely to occur on average. So, it is also the mean of a random variable denoted by μ and is given by:

$$E(X) = \mu = \sum xP_X(x) = \sum xP(X = x).$$

The expectation of constant is the constant itself.

$$E(a) = a$$

1.5.1 Linearity of Expectation

Expectation if Linear i.e for random variables X, A and B.

1. $E(A + B) = E(A) + E(B)$
2. $E(aA) = aE(A)$

1.6 Variance of a Random Variable

The Variance of a random variable, denoted by $Var(X)$ is the Expectation of the square of the deviation of values from the mean of the random variable in question and is given by:

$$Var(X) = E((X - \mu)^2) = E(X^2) - E(x)^2.$$

Chapter 2

Binomial Distribution

Lecture 2: Second Lecture

2.1 Binomial Expansion

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

2.1.1 Properties of Binomial Expansion

1. There are $n+1$ elements.
2. $\binom{n}{r} = \binom{n}{n-r}$
3. $(n-r)+r = n$
4. $t_{r+1} = \binom{n}{r} q^{n-r} p^r$

2.2 Head Tail Experiment(HTE)

2.2.1 Outcomes of HTE

A HTE has 2 possible outcomes. So it is Bernoulli Process.

1. Head (H)
2. Tail (T)

We set Head to be success (1) and Tail to be failure (0).

Then, we have the probability of success as:

$$p = \frac{1}{2}$$

And of failure as:

$$q = \frac{1}{2}$$

Experiment:
Let, n be the #trials i.e we toss a coin n times.

The outcome of this experiment is that we get r # success and $(n-r)$ failure.

Note. The Probability that we get exactly r # success is same as:

$$\binom{n}{r} q^{n-r} p^r$$

because as stated before, there are r success. So the probability:

$$P(H \cap H \cap H \ldots \text{r times}) = \prod^{r} p = p^r$$

and similarly, q^{n-r} i.e.

$$P_X(r) = P(X = r) = \binom{n}{r} q^{n-r} p^r \tag{2.1}$$

Note. The Probability that we get at most r # success is same as $F_X(r)$:

$$F_X(r) = P(X \leq r) = \sum_{n=0}^{r} P_X(r) \tag{2.2}$$

2.3 Exclusive and Independent Events

Definition 2.3.1 (Mutually Exclusive Events). Two events A and B are mutually exclusive (disjoint) if they cannot occur simultaneously. Mathematically:

$$P(A \cap B) = 0$$

Definition 2.3.2 (Independent Events). Two events A and B are independent if the occurrence of one does not affect the probability of the other. Mathematically:

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

or,

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

2.4 Bernoulli Process

In this whole experiment, we have tossed a coin many times. First we fixed how many times we toss. (We fix n). We have the two possible events, one head and the other tail. These events are mutually exclusive and independent. We also know the probability of success and this probability is constant (fixed) throughout the experiment.
What we have defined here is a process called **Bernoulli Process**.

Remark. A Bernoulli process consists of many independent *Bernoulli Trials*, each trial having exactly two possible outcomes (often called success and failure).

Definition 2.4.1 (Bernoulli Process). A *Bernoulli process* is a sequence of independent and

identically distributed (i.i.d.) *Bernoulli trials*, where each trial results in one of two possible outcomes: "success" (with probability p) or "failure" (with probability $1-p$). The number of trials n is fixed in advance, and the probability of success p remains constant throughout the experiment.
Formally, the process consists of random variables $X_1, X_2, \ldots, X_n$, where each

$$R_{X_i} = \begin{cases} 1, & \text{with probability } p \quad \text{(success)} \\ 0, & \text{with probability } 1-p \quad \text{(failure)} \end{cases}$$

represents range of random variable for i^{th} trial among the n trials and X_i are independent.

Here,

1. n is fixed
2. Events are mutually exclusive
3. Trials are independent
4. p is known and fixed throughout the experiment.

2.5 Probability Distribution for Bernoulli Trial

The distribution resulted by this probability mass function for $n = 1$.

$$P_X(r) = p^r$$

Here, $r \in \{0, 1\}$ since there cannot be more than 1 success in one trial. It is a special case of sec:binomial-distribution

1. $\mu = p$
2. $\sigma^2 = p(1-p) = pq$

2.6 Binomial Distribution

The distribution resulted by this probability mass function is a binomial distribution.

2.6.1 Symmetricity of Binomial Distribution

Remark. If $p = 0.5$, distribution is symmetric

Proof. If probability of success is same as the failure, there is no skewness in the distribution.

⊛

If $p < 0.5$, distribution is skewed right(+ve).

Proof. If p is less than q, there will be low chances of success so distribution will be skewed to left.

⊛

If $p > 0.5$, distribution is skewed left(-ve).

Proof. If p is more than q, there will be high chances of success so distribution will be skewed to right.

⊛

When n is sufficiently large, binomial distribution is symmetric.

2.6.2 Expectation of Binomial Distribution

Given a binomial random variable $X \sim \text{Binomial}(n, p)$ with probability mass function:

$$P_X(r) = P(X = r) = \binom{n}{r} q^{n-r} p^r, \quad r = 0, 1, 2, \ldots, n,$$

where $q = 1 - p$.

Definition 2.6.1. The expectation $E[X]$ is defined as:

$$E[X] = \sum_{r=0}^{n} r\, P_X(r) = \sum_{r=0}^{n} r \binom{n}{r} q^{n-r} p^r.$$

Note. The term for $r = 0$ is zero since it is multiplied by r, so we can start the sum from $r = 1$:

$$E[X] = \sum_{r=1}^{n} r \binom{n}{r} q^{n-r} p^r.$$

Rewrite $r\binom{n}{r}$ using the identity:

$$r\binom{n}{r} = n\binom{n-1}{r-1}.$$

Therefore,

$$E[X] = \sum_{r=1}^{n} n\binom{n-1}{r-1} q^{n-r} p^r = n \sum_{r=1}^{n} \binom{n-1}{r-1} q^{n-r} p^r.$$

Change the index of summation by letting $k = r - 1$. When $r = 1$, $k = 0$, and when $r = n$, $k = n - 1$. So,

$$E[X] = n \sum_{k=0}^{n-1} \binom{n-1}{k} q^{n-(k+1)} p^{k+1} = np \sum_{k=0}^{n-1} \binom{n-1}{k} q^{(n-1)-k} p^k.$$

Notice that the sum is the binomial expansion of $(p + q)^{n-1}$:

$$\sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{(n-1)-k} = (p + q)^{n-1} = 1^{n-1} = 1.$$

Hence,

$$E[X] = np \cdot 1 = np.$$

Remark. The expectation of a binomial random variable $X \sim \text{Binomial}(n, p)$ is

$$\boxed{E[X] = np.} \tag{2.3}$$

2.6.3 Variance of Binomial Distribution

Recall that the variance of a random variable X is defined as:

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

We have already derived that, [1]

$$E[X] = np.$$

Next, we compute $E[X^2]$:

$$E[X^2] = \sum_{r=0}^{n} r^2 P_X(r) = \sum_{r=0}^{n} r^2 \binom{n}{r} q^{n-r} p^r,$$

where $q = 1 - p$.

Rewrite r^2 as $r(r-1) + r$, so that

$$\begin{aligned} E[X^2] &= \sum_{r=0}^{n} [r(r-1) + r] \binom{n}{r} q^{n-r} p^r \\ &= \sum_{r=0}^{n} r(r-1) \binom{n}{r} q^{n-r} p^r + \sum_{r=0}^{n} r \binom{n}{r} q^{n-r} p^r. \end{aligned}$$

We already know the second sum is $E[X] = np$. Focus on the first sum:

$$S = \sum_{r=0}^{n} r(r-1) \binom{n}{r} q^{n-r} p^r.$$

Using the combinatorial identity:

$$r(r-1) \binom{n}{r} = n(n-1) \binom{n-2}{r-2},$$

we get

$$S = n(n-1) \sum_{r=2}^{n} \binom{n-2}{r-2} q^{n-r} p^r.$$

Change the index by letting $k = r - 2$. When $r = 2$, $k = 0$, and when $r = n$, $k = n - 2$:

$$\begin{aligned} S &= n(n-1) \sum_{k=0}^{n-2} \binom{n-2}{k} q^{n-(k+2)} p^{k+2} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} q^{(n-2)-k} p^k. \end{aligned}$$

[1] See **??**

Recognize the sum as the binomial expansion of $(p+q)^{n-2} = 1^{n-2} = 1$:

$$S = n(n-1)p^2.$$

Putting it all together:

$$\begin{aligned} E[X^2] &= S + E[X] \\ &= n(n-1)p^2 + np. \end{aligned}$$

Finally, compute the variance:

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= n(n-1)p^2 + np - n^2p^2. \end{aligned}$$

Simplify:

$$\begin{aligned} \mathrm{Var}(X) &= np + n(n-1)p^2 - n^2p^2 \\ &= np + n^2p^2 - np^2 - n^2p^2 \\ &= np - np^2 \\ &= np(1-p) \\ &= npq. \end{aligned}$$

Remark. The variance of a binomial random variable $X \sim \mathrm{Binomial}(n, p)$ is

$$\boxed{\mathrm{Var}(X) = npq,}$$

where $q = 1 - p$.

Chapter 3

Poisson Distribution

Lecture 3: Third Lecture

Example (Simpler Example: Customers at a Cafe). Imagine a small cafe where, on average, 2 customers arrive every minute during the lunch rush. We want to find the probability that exactly 5 customers will arrive in a specific one-minute interval during this period. This scenario can be modeled using the Poisson distribution.
Here, the factors of the Poisson distribution are:

- λ (lambda): The average rate of arrivals. In this case, $\lambda = 2$ customers per minute. This average is assumed to be constant during the lunch rush period.
- x: The specific number of arrivals we are interested in. Here, $x = 5$ customers.
- e: Euler's number, approximately 2.71828.

Explanation of "specific one-minute interval": The term "specific one-minute interval" refers to any chosen single minute within the timeframe where the average arrival rate ($\lambda = 2$ customers/minute) is valid – in this case, during the lunch rush. It doesn't mean a pre-selected or unique minute, but rather *any* individual minute you decide to observe. For example:

- The minute from 12:05 PM to 12:06 PM.
- The minute from 1:10 PM to 1:11 PM.
- Any other one-minute slot during the lunch rush.

The Poisson distribution calculates the probability of observing x events (5 customers) in such a defined interval, assuming the rate λ is constant across all such intervals and events occur independently. So, for any given minute you pick during the lunch rush, the probability of 5 customers arriving in *that particular minute* is what we are calculating.
The probability mass function (PMF) for the Poisson distribution is:

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

We want to calculate $P(X = 5)$ when $\lambda = 2$.

Plugging the values into the formula:

$$P(X = 5) = \frac{e^{-2} \cdot 2^5}{5!}$$

Let's break down the calculation:

- $e^{-2} \approx 0.135335$
- $2^5 = 32$
- $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

So,

$$P(X = 5) = \frac{0.135335 \times 32}{120}$$
$$P(X = 5) = \frac{4.33072}{120}$$
$$P(X = 5) \approx 0.036089$$

Therefore, the probability of exactly 5 customers arriving in any specific one-minute interval during the lunch rush is approximately 0.0361, or about 3.61

3.1 Introduction

It is used to model the probability of a given number of events occurring within a fixed interval of time or space, assuming these events happen with a known constant mean rate and independently of the time since the last event.

Definition 3.1.1 (Poisson Distribution). A discrete random variable X is said to follow a Poisson distribution if its probability mass function (PMF) is given by:

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

for $x = 0, 1, 2, \ldots$. Here:

- x is the number of occurrences of an event (a non-negative integer).
- λ (lambda) is the average number of events occurring in the given interval. It is the sole parameter of the distribution.
- e is Euler's number (approximately 2.71828).
- $x!$ is the factorial of x.

The notation for a random variable X following a Poisson distribution with mean λ is $X \sim \mathcal{P}(\lambda)$ or $X \sim \text{Poisson}(\lambda)$. Conditions for using Poisson distribution include:

- Events occur at random and independently. The occurrence of one event does not affect the probability of another event.
- The mean number of events (λ) occurring within a given interval is known and constant.

3.2 Meaning of $P(X = x)$

The expression $P(X = x)$ represents the probability that exactly x events will occur in a specified interval of time or space. For instance, if X represents the number of emails received per hour and $X \sim \mathcal{P}(10)$, meaning on average 10 emails are received per hour, then $P(X = 5)$ would be the probability of receiving exactly 5 emails in a particular hour. This is calculated using the Poisson formula:

$$P(X = 5) = \frac{e^{-10}10^5}{5!}$$

The value of x can be any non-negative integer (0, 1, 2, ...), reflecting that there's theoretically no upper limit to the number of events that can occur.

3.3 Derivation of Expectation and Variance

For a Poisson-distributed random variable $X \sim \mathcal{P}(\lambda)$, the expectation (mean) and variance are both equal to λ.

3.3.1 Expectation $E[X]$

The expectation of a discrete random variable X is given by $E[X] = \sum_{x=0}^{\infty} xP(X = x)$.

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} \quad \text{(The term for } x = 0 \text{ is 0)} \\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \\
\text{Let } k = x - 1, &\text{ so } x = k + 1. \text{ When } x = 1, k = 0. \\
E[X] &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} \\
&= e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
\text{Since } \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} &= e^{\lambda} \text{ (Taylor series expansion of } e^{\lambda}\text{)}, \\
E[X] &= e^{-\lambda} \lambda e^{\lambda} \\
&= \lambda
\end{aligned}$$

Thus, the expectation of a Poisson distribution is λ.

3.3.2 Variance Var(X)

The variance of a random variable X is given by $\text{Var}(X) = E[X^2] - (E[X])^2$. First, we calculate $E[X(X-1)]$:

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda}\sum_{x=2}^{\infty} x(x-1)\frac{\lambda^x}{x!} \quad \text{(Terms for } x = 0, 1 \text{ are 0)} \\
&= e^{-\lambda}\sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!}
\end{aligned}$$

Let $j = x - 2$, so $x = j + 2$. When $x = 2, j = 0$.

$$\begin{aligned}
E[X(X-1)] &= e^{-\lambda}\sum_{j=0}^{\infty} \frac{\lambda^{j+2}}{j!} \\
&= e^{-\lambda}\lambda^2\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}
\end{aligned}$$

Since $\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}$,

$$\begin{aligned}
E[X(X-1)] &= e^{-\lambda}\lambda^2 e^{\lambda} \\
&= \lambda^2
\end{aligned}$$

We know that $E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X]$. So, $E[X^2] = E[X(X-1)] + E[X]$. Using $E[X] = \lambda$ and $E[X(X-1)] = \lambda^2$:

$$E[X^2] = \lambda^2 + \lambda$$

Now, we can find the variance:

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

Thus, the variance of a Poisson distribution is also λ.

3.4 Applications of Poisson Distribution

The Poisson distribution is a versatile tool used in various fields to model the number of occurrences of an event over a specific interval. Some key applications include:

- **Telecommunications:** Predicting the number of calls arriving at a call center per hour, or the number of network failures per week. This helps in staffing and resource allocation.
- **Quality Control:** Modeling the number of defects or errors in a manufactured product (e.g., defects per square meter of material or per batch of items).
- **Finance and Insurance:** Estimating the number of insurance claims per month or bankruptcies filed per month.

- **Biology and Medicine:** Counting the number of bacteria in a culture, radioactive decay particles per minute, or instances of a rare disease in a population. Ecology example from StudyPug could involve modeling the number of trees affected by a disease per acre.
- **Astronomy:** Modeling the number of meteorites of a certain size striking the Earth in a year.
- **Traffic Engineering:** Predicting the number of cars arriving at an intersection or a toll booth in a specific time interval.
- **Retail Management:** Forecasting customer arrivals at a store or checkout queues, helping in optimizing staff schedules and queue management.

The Poisson distribution is particularly useful for modeling rare events or counting processes where events occur independently and at a constant average rate.

Chapter 4

Geometric Distribution

Lecture 4: Fourth Lecture

4.0.1 Introduction

Example. In HTE, each trial is a Bernoulli Trial. Let's say we toss the coin k times before we get our first head (success). Then, there are k failures and 1 success. So, the probability of tossing the coin k times before our first head is:

$$P[X = k] = q^k p$$

Here, X is a Random Variable that models the # trials before our first success. (i.e., number of failures)

4.1 Definition

Definition 4.1.1. A **Geometric Distribution** models the number of successive, independent Bernoulli trials required to achieve the first success, or alternatively, the number of failures before the first success. Let p be the probability of success on any given trial, and $q = 1 - p$ be the probability of failure.
There are two common parameterizations:

1. The random variable X represents the number of failures *before* the first success. The support of X is $\{0, 1, 2, \dots\}$.
   $X \sim \textbf{Geom}(p)$
2. The random variable Y represents the number of trials *until* (and including) the first success. The support of Y is $\{1, 2, 3, \dots\}$. $Y \sim \textbf{Geom}(p)$

It is clear that $Y = X + 1$.

4.2 PMF of Geometric Distribution

For the random variable X that represents the number of failures before we get the first success (support $x \in \{0, 1, 2, \dots\}$):

$$\boxed{P[X = x] = (1 - p)^x p = q^x p} \tag{4.1}$$

For the random variable Y that represents the number of trials to get the first success (support

$y \in \{1, 2, 3, \dots\}$):

$$\boxed{P[Y = y] = (1-p)^{y-1}p = q^{y-1}p} \tag{4.2}$$

4.3 CDF of Geometric Distribution

For X (number of failures before first success, $x \geq 0$):
The CDF is $F_X(x) = P[X \leq x]$.

$$\begin{aligned}
P[X \leq x] &= \sum_{k=0}^{x} P[X = k] \\
&= \sum_{k=0}^{x} q^k p \\
&= p \sum_{k=0}^{x} q^k \\
&= p \left(\frac{1 - q^{x+1}}{1 - q} \right) \\
&= p \left(\frac{1 - q^{x+1}}{p} \right) \\
&= 1 - q^{x+1}
\end{aligned}$$

So,

$$\boxed{F_X(x) = P[X \leq x] = 1 - q^{x+1}, \quad \text{for } x = 0, 1, 2, \dots} \tag{4.3}$$

For Y (number of trials until first success, $y \geq 1$): The CDF is $F_Y(y) = P[Y \leq y]$.

$$\begin{aligned}
P[Y \leq y] &= \sum_{k=1}^{y} P[Y = k] \\
&= \sum_{k=1}^{y} q^{k-1} p \\
&= p \sum_{k=1}^{y} q^{k-1} \\
&\text{Let } j = k - 1, \text{ then the sum becomes } \sum_{j=0}^{y-1} q^j \\
&= p \left(\frac{1 - q^{(y-1)+1}}{1 - q} \right) \\
&= p \left(\frac{1 - q^y}{p} \right) \\
&= 1 - q^y
\end{aligned}$$

Alternatively, using $Y = X+1$: $P[Y \leq y] = P[X+1 \leq y] = P[X \leq y-1] = 1-q^{(y-1)+1} = 1-q^y$.
So,

$$\boxed{F_Y(y) = P[Y \leq y] = 1 - q^y, \quad \text{for } y = 1, 2, 3, \dots} \tag{4.4}$$

4.4 Expectation of Geometric Distribution

For X (number of failures before first success):
We use the identity $\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$ for $|p| < 1$.
Differentiating with respect to x:

$$\sum_{k=1}^{\infty} kp^{k-1} = \frac{1}{(1-p)^2}$$

Multiplying by p:

$$\sum_{k=1}^{\infty} kp^{k} = \frac{p}{(1-p)^2}$$

Since the $k = 0$ term is $0 \cdot p^0 = 0$, this is also

$$\sum_{k=0}^{\infty} kp^{k} = \frac{p}{(1-p)^2}$$

Now,

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} xP[X = x] \\
&= \sum_{x=0}^{\infty} xq^x p \\
&= p\sum_{x=0}^{\infty} xq^x \\
&= p\left(\frac{q}{(1-q)^2}\right) \\
&= p\left(\frac{q}{p^2}\right) \\
&= \frac{q}{p}
\end{aligned}$$

$$\boxed{\therefore E[X] = \frac{q}{p} = \frac{1-p}{p}} \tag{4.5}$$

For Y (number of trials until first success): Derivation (substitute $Y = X + 1$):

$$\begin{aligned}
E[Y] &= E[X+1] \\
&= E[X] + E[1] \\
&= \frac{q}{p} + 1 \\
&= \frac{q+p}{p} \\
&= \frac{1}{p}
\end{aligned}$$

Alternatively, using PMF of Y:

$$\begin{aligned}
E[Y] &= \sum_{y=1}^{\infty} yP[Y = y] \\
&= \sum_{y=1}^{\infty} yq^{y-1}p \\
&= p\sum_{y=1}^{\infty} yq^{y-1} \\
&= p\left(\frac{1}{(1-q)^2}\right) \quad \text{(using the derivative identity from above)} \\
&= p\left(\frac{1}{p^2}\right) \\
&= \frac{1}{p}
\end{aligned}$$

$$\boxed{\therefore E[Y] = \frac{1}{p}} \tag{4.6}$$

4.5 Variance of Geometric Distribution

For X (number of failures before first success):
We know,

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

We have,

$$\sum_{k=0}^{\infty} pp^k = \frac{p}{(1-p)^2}$$

Differentiating with respect to p:

$$\begin{aligned}
\sum_{k=1}^{\infty} k^2 p^{k-1} &= \frac{d}{dp}\left(\frac{p}{(1-p)^2}\right) \\
&= \frac{1+p}{q^3} \\
\text{or, } \sum_{k=0}^{\infty} k^2 p^k &= \frac{p(1+p)}{q^3} \quad \text{(Multiplying by } p\text{)}
\end{aligned}$$

$\because$ the $k = 0$ term is 0.

Now, for $E[X^2]$:

$$\begin{aligned}
E[X^2] &= \sum_{x=0}^{\infty} x^2 P[X=x] \\
&= \sum_{x=0}^{\infty} x^2 q^x p \\
&= p \sum_{x=0}^{\infty} x^2 q^x \\
&= p \left(\frac{q(1+q)}{(1-q)^3} \right) \\
&= p \left(\frac{q(1+q)}{p^3} \right) \\
&= \frac{q(1+q)}{p^2}
\end{aligned}$$

Now, $Var(X)$:

$$\begin{aligned}
Var(X) &= E[X^2] - (E[X])^2 \\
&= \frac{q(1+q)}{p^2} - \left(\frac{q}{p} \right)^2 \\
&= \frac{q+q^2}{p^2} - \frac{q^2}{p^2} \\
&= \frac{q}{p^2}
\end{aligned}$$

$$\boxed{\therefore Var(X) = \frac{q}{p^2} = \frac{1-p}{p^2}} \tag{4.7}$$

For Y (number of trials until first success): Derivation (substitute $Y = X+1$): We use the property $Var(aZ+b) = a^2 Var(Z)$. Here, $Y = X+1$, so $a = 1, b = 1$.

$$\begin{aligned}
Var(Y) &= Var(X+1) \\
&= Var(X) \quad \text{(since adding a constant does not change variance)} \\
&= \frac{q}{p^2}
\end{aligned}$$

$$\boxed{\therefore Var(Y) = \frac{q}{p^2} = \frac{1-p}{p^2}} \tag{4.8}$$

Chapter 5

Negative Binomial Distribution

Lecture 5: Fifth Lecture

Example (Coin Tossing for Multiple Heads). Suppose we are flipping a fair coin repeatedly. Each flip is an independent Bernoulli trial. Let p be the probability of getting a Head (success), and $q = 1 - p$ be the probability of getting a Tail (failure). We are interested in the number of Tails, X, observed before we obtain exactly r Heads.
For instance, let $r = 3$ (we want 3 Heads) and assume the probability of a Head is $p = 0.5$. We continue flipping the coin until we have observed 3 Heads. The number of Tails(failures) X that occurred before the 3rd Head is a random variable following a Negative Binomial distribution. Possible outcomes for X could be:

- HHH: 0 Tails ($k = 0$). The sequence ends with the 3rd Head.
- THHH: 1 Tail ($k = 1$).
- HTHH: 1 Tail ($k = 1$).
- HHTH: 1 Tail ($k = 1$).
- TTHHH: 2 Tails ($k = 2$).

This scenario perfectly illustrates the conditions for a Negative Binomial distribution:

- The experiment consists of a sequence of independent trials (each coin flip).
- Each trial has only two possible outcomes: success (Head) or failure (Tail).
- The probability of success (p) is constant for each trial.
- The experiment continues until a fixed number of successes (r) is achieved.

The random variable of interest, X, is the number of failures (Tails) that occur before the r-th success (Head).

5.1 Definition

Definition 5.1.1 (Negative Binomial Distribution). The Negative Binomial distribution, sometimes referred to as the Pascal distribution, is a discrete probability distribution. It models the random variable X that represents the # failures that occur before a specified (constant or fixed) number of successes, denoted by r, is achieved.

A random variable X that follows a Negative Binomial distribution is commonly denoted as $X \sim \textbf{NB}(r, p)$
where,
r is the # success
p is the probability of success for one trial.

A random process is described by the Negative Binomial distribution if it adheres to the following conditions:

- The process is a sequence of independent trials.
- Each trial results in one of two outcomes, termed success or failure.
- The probability of success, p, is constant for all trials.
- The experiment is continued until a predetermined total of r successes are observed. Here, r must be a positive integer.

Remark. It's worth noting an alternative formulation where the distribution models the total number of trials $Y = X + r$ required to achieve r successes.

5.2 Probability Mass Function (PMF) of X

The **PMF** of $\text{NB}(r, p)$, $P(X = x)$, quantifies the probability of observing exactly x failures prior to achieving the r-th success in a series of independent Bernoulli trials with p the success probability.
That means, there are x failures and r success. So the total # trials is $k = x + r$. And first $x + r - 1$ trials represent $Y \sim \text{Bin}(n, p)$. Then,

$$P(X = x) = P(Y = r - 1) \cdot p$$

$$P(X = x) = \binom{x + r - 1}{r - 1} q^x p^{r-1} \cdot p$$
$$P(X = x) = \binom{x + r - 1}{r - 1} q^x p^r$$

$\therefore$ PMF for $X \sim NB(r, p)$ is:

$$\boxed{P(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x, \quad \text{for } x \in \mathbb{N}_0} \tag{5.1}$$

Here, $1 - p$ is often denoted as q, the probability of failure.

5.3 Expectation of X

The expectation (or mean) of a random variable $X \sim \text{NB}(r, p)$, denoted $E[X]$, can be derived by considering X as a sum of independent Geometric random variables.

Let X_i for $i = 1, \ldots, r$ be the number of failures between the $(i-1)$-th success and the i-th success.

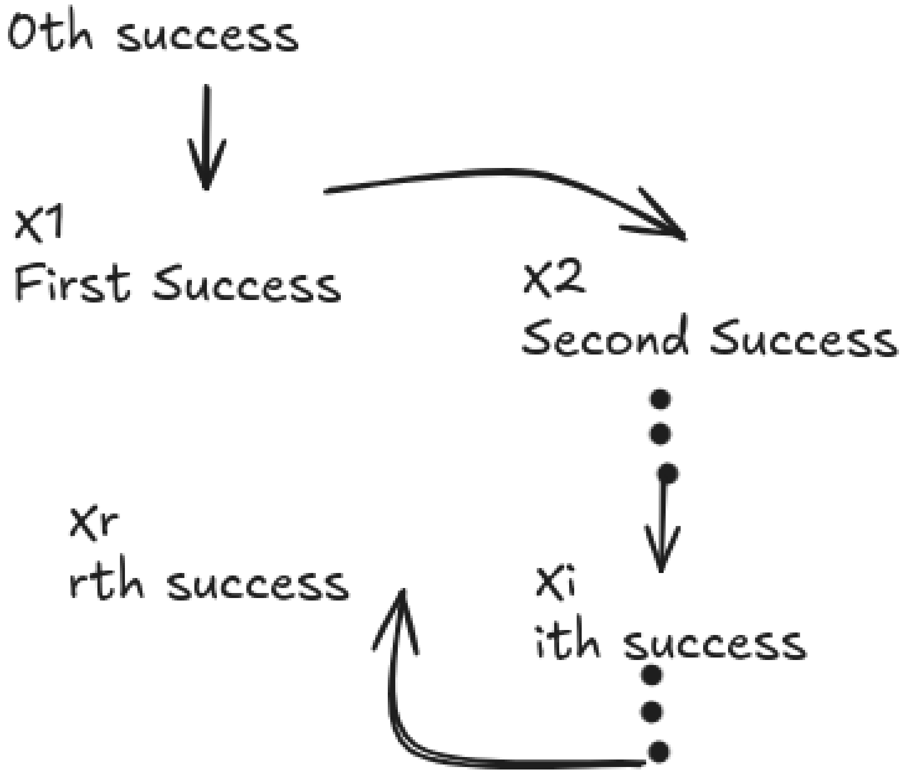

Figure 5.1: X_i follows Geometric distribution

Each X_i follows a Geometric distribution with success probability p, specifically $X_i \sim Geom(p)$. The PMF for such a variable (counting failures before the first success) is $P(X_i = k) = (1-p)^k p$ for $k = 0, 1, 2, \ldots$. The expectation of this form of Geometric distribution is $E[X_i] = \frac{q}{p}$.

The total number of failures X before r successes is the sum of these r independent Geometric variables:

$$X = X_1 + X_2 + \cdots + X_r$$

By the linearity of expectation:

$$\begin{aligned}
E[X] &= E[X_1 + X_2 + \cdots + X_r] \\
&= E[X_1] + E[X_2] + \cdots + E[X_r] \\
&= \sum_{i=1}^{r} E[X_i] \\
&= \sum_{i=1}^{r} \frac{q}{p} \\
&= r\frac{1-p}{p} \\
&= \frac{rq}{p}
\end{aligned}$$

Thus, the expectation of $X \sim NB(r, p)$ is:

$$E[X] = \frac{rq}{p} \tag{5.2}$$

$\mathbb{E}[X]$ is expected number of failures before r success

5.4 Variance of X

The variance of $X \sim NB(r, p)$, denoted $Var(X)$, can similarly be derived using the representation of X as the sum of r independent Geometric random variables $X_i \sim Geom(p)$, as defined in the expectation section. The variance of each X_i (number of failures before the first success, with success probability p) is $Var(X_i) = \frac{q}{p^2}$.

Since the X_i variables are independent, the variance of their sum is the sum of their variances:

$$\begin{aligned} Var(X) &= Var(X_1 + X_2 + \cdots + X_r) \\ &= Var(X_1) + Var(X_2) + \cdots + Var(X_r) \quad \text{(due to independence of } X_i\text{)} \\ &= \sum_{i=1}^{r} Var(X_i) \\ &= \sum_{i=1}^{r} \frac{q}{p^2} \\ &= \frac{rq}{p^2} \end{aligned}$$

Thus, the variance of the Negative Binomial distribution ($X \sim NB(r, p)$) is:

$$Var(X) = \frac{rq}{p^2} \tag{5.3}$$

An important property is that the variance of the Negative Binomial distribution can be larger than its mean, making it suitable for modeling overdispersed count data.

5.5 Alternate Formulation using $Y = X + r$

Here, Y is the random variable that represents the # trials to get r^{th} success.

5.5.1 PMF of Y

$$\boxed{P[Y = y] = \binom{y-1}{r-1} q^{y-r} \cdot p^r} \tag{5.4}$$

This measures the probability of getting r^{th} success in y trials.

5.5.2 Expectation of Y

$$\boxed{\mathbb{E}[Y] = \mathbb{E}[X + r] = \frac{r(1-p)}{p} + r = \frac{r}{p}} \tag{5.5}$$

This measures the expected number of trials to get r^{th} success.

5.5.3 Variance of Y

$$\boxed{\mathrm{Var}[Y] = \mathrm{Var}[X + r] = \mathrm{Var}[X] = \frac{rq}{p^2}} \tag{5.6}$$

Chapter 6

Hypergeometric Distribution

Lecture 6: Sixth Lecture

6.0.1 Introduction

The Hypergeometric Distribution is used to model the probability of x successes in n draws, *without replacement*, from a finite population of size N that contains exactly m successes. This contrasts with the Binomial distribution, where draws are made *with replacement* (or from a very large population), ensuring that the probability of success remains constant for each trial (i.e., trials are independent Bernoulli Trials).

Example. Consider an urn containing $N = 10$ balls. Of these, $m = 6$ are green (successes) and $N - m = 4$ are blue (failures). We draw a sample of $n = 3$ balls from the urn without replacement. We want to find the probability that exactly $x = 2$ of the drawn balls are green (and thus $n - x = 3 - 2 = 1$ ball is blue).
Let X be the Random Variable representing the number of green balls in our sample of 3.
The number of ways to choose 2 green balls from the 6 available green balls is $\binom{m}{x} = \binom{6}{2}$.
The number of ways to choose 1 blue ball from the 4 available blue balls is $\binom{N-m}{n-x} = \binom{4}{1}$.
The total number of ways to choose any 3 balls from the 10 available balls is $\binom{N}{n} = \binom{10}{3}$.
So, the probability of drawing exactly 2 green balls is:

$$P[X = 2] = \frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}}$$

Calculating the combinations:

- $\binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{6\times5}{2\times1} = 15$
- $\binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4}{1} = 4$
- $\binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10\times9\times8}{3\times2\times1} = 10 \times 3 \times 4/(3 \times 1) = 120$

Therefore,

$$P[X = 2] = \frac{15 \times 4}{120} = \frac{60}{120} = \frac{1}{2}$$

Here, X is a random variable that follows a Hypergeometric Distribution.

6.1 Definition

Definition 6.1.1. A **Hypergeometric Distribution** describes the probability of x successes in n draws, without replacement, from a finite population of size N where m objects are classified as successes and $N - m$ objects are classified as failures.
The parameters of the Hypergeometric distribution are:

- N: The total number of items in the population.
- m: The total number of items in the population that are classified as successes. $(0 \leq m \leq N)$
- n: The number of items drawn in a sample from the population (the sample size). $(0 \leq n \leq N)$

The random variable X represents the number of successes in the sample of size n. The value of x (a specific outcome for X) must satisfy:

$$\max(0, n - (N - m)) \leq x \leq \min(n, m)$$

This ensures that we don't try to choose more successes (or failures) than are available, or more items than are in the sample.

6.2 PMF of Hypergeometric Distribution

For the random variable X that represents the number of successes in a sample of size n drawn without replacement from a population of size N containing m successes. The probability mass function (PMF) is given by:

$$\boxed{P[X = x] = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}} \tag{6.1}$$

where:

- x is the number of successes in the sample.
- $\binom{a}{b}$ is the binomial coefficient, representing "a choose b", calculated as $\frac{a!}{b!(a-b)!}$.
- The support for x is $\max(0, n - (N - m)) \leq x \leq \min(n, m)$. For values of x outside this range, $P[X = x] = 0$.

Chapter 7

Continuos Random Variable

Lecture 7: Seventh Lecture

Definition 7.0.1 (Continuos Random Variable).

7.1 Probability Density Function

7.2 Cumulative Distribution Function

7.3 Expectation of Continuos Random Variable

7.4 Variance of Continuos Random Variable

Appendix