Statistic

John Joy

April 2024

Contents

1 Probability Fundamentals

1.1 Set

1.1.1 Definitions of Set

Definition 1 (set)**.** ***Set*** *is a collection of some elements.*

In this note, we will consider sets as events. Sets have some properties as follow:

1. Order of elements doesn't matter.
2. Repeated elements will be regarded as one single element.

For example, $\{1, 2, 3\} = \{2, 3, 1\}$, and $\{1, 1, 2, 4\} = \{1, 2, 4\}$. We always denote a set with a pair of curly brackets($\{\}$) which contains some elements. Specially, if a set contains no elements, we also call it a set. And we denote it as $\emptyset$.

1.1.2 Operations on sets

Definition 2 (Cartesian product)**.** *Two sets A, B, we define their **Cartesian product** as*

$$A \times B := \{(\omega_1, \omega_2) : \textit{ for } \omega_1 \in A \textit{ and } \omega_2 \in B\}$$

Theorem 1 (Product rule)**.** *If $X_1, X_2, \ldots, X_k$ are finite sets, then*

$$|X_1 \times X_2 \times \cdots \times X_k| = \Pi_{i=1}^{k} |X_i|$$

Example 1. *For example, if $A = \{1, 2, 3\}$ and $B = \{4, 5, 6\}$, then*

$$A \times B = \{(1, 4), (1, 5), (1, 6), (2, 4), (2, 5), (2, 6), (3, 4), (3, 5), (3, 6)\}$$

Definition 3 (Disjoint or mutually exclusive)**.** *If two sets A, B, they have no common element, which means $A \cap B = \emptyset$.*

Definition 4 (A measure of set)**.** *We denote the $|A|$ or $\#A$ is the element number in the set.*

Example 2.

- *If $A = \{1, 2, 3\}$, then $\#A = 3$.*
- *If $A = \{1, 2, 2, 4\}$, then $\#A = 3$.*
- *If $A = \emptyset$, then $\#A = 0$.*
- *If A is an infinite set, then $\#A = \infty$*

1.2 Probability

Definition 5 (experiment)**.** *Experiment is the trial that can produce uncertain outcomes.*

Remark 1. *If a trial have a certain outcome, it doesn't belong to experiment. Such as the sun will rise tomorrow.*

Definition 6 (Sample space)**.** *All outcomes of an experiment compose a set. The set is called **Sample space**. Always denoted as* Ω *or* S*.*

Example 3. *Tossing a coin is a simple experiment. Here we can take the sample space* $\Omega = \{H, T\}$*, with the outcome* H *representing 'heads' and the outcome* T *representing 'tails'.*

Definition 7 (Event)**.** *From the perspective of set, an event is a **subset** of the sample space. From the perspective of experiment, an event is some **outcomes** of the experiment.*

Example 4. *An experiment is to toss a coin three times. The outcome can be denoted as* $(\omega_1, \omega_2, \omega_3)$*, where* $\omega_i \in \{H, T\}$ *is the i-th coin toss. The sample space* Ω *is*

$$\begin{aligned} \Omega &= \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{H, T\} \; for \; all \; i = 1, 2, 3\} \\ &= \{H, T\} \times \{H, T\} \times \{H, T\} = \{H, T\}^3 \end{aligned}$$

some examples of event are following

- $A_1 = \{(T, T, T), (H, H, H)\}$ *is the event that* `'all three tosses land on the same side'`*.*
- $A_2 = \{(T, T, H), (T, H, T), (H, T, T), (H, H, H)\}$ *the event* `' an odd number of heads appear'`*.*

In many cases, we want to know how probably an event happen. We need to measure the probability of an event.

Definition 8 (Probability)**.** *The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.*

Definition 9 (Probability distribution)**.** *A probability distribution or probability measure on a sample space* Ω *is a function* $\mathbb{P} : A \to \mathbb{R}$*, where* $A \subseteq \Omega$*, so that:*

1. $\mathbb{P}(A) \in [0, 1]$ *for every event* $A \subseteq \Omega$*;*
2. $\mathbb{P}(\emptyset) = 0$ *and* $\mathbb{P}(\Omega) = 1$*;*
3. (a) *for every sequence of pairwise disjoint events* $A_1, A_2, \ldots, A_k \subseteq \Omega$ *we have*

$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathbb{P}(A_i)$$

(b) *Further more, for countable infinite sequence of pairwise disjoint events* $A_1, A_2, \ldots, A_k \subseteq \Omega$ *we have*

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

Lemma 1. *Suppose that* Ω *is a discrete sample space.*

(a) *If* $\mathbb{P}$ *is a probability distribution on* Ω *then*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}), \quad \text{for all events } A \subseteq \Omega.$$

(b) *Let* $p_\omega \geq 0$ for all $\omega \in \Omega$ *with* $\sum_{\omega \in \Omega} p_\omega = 1$. *Then the function* $\mathbb{P}$ *defined by*

$$\mathbb{P}(A) := \sum_{\omega \in A} p_\omega, \quad \text{for all events } A \subseteq \Omega,$$

is a probability distribution on Ω.

Proposition 1. *Let* $\mathbb{P}$ *be a probability distribution on sample space* Ω. *Then the following hold:*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ for all $A \subseteq \Omega$.
2. $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ for all $A, B \subseteq \Omega$.
3. $\mathbb{P}(A) \leq \mathbb{P}(B)$ for all $A, B \subseteq \Omega$ *with* $A \subseteq B$.
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *for all* $A, B \subseteq \Omega$
5. $\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} \mathbb{P}(A_i)$ *for all* $A_1, \ldots, A_k \subseteq \Omega$.

Theorem 2 (Inclusion-exclusion principle)**.** *Suppose* Ω *is a sample space, and* $A_1, A_2, \ldots, A_n \subseteq \Omega$, *we have*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\emptyset \neq J \subseteq \{1,\ldots,n\}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{j \in J} A_j\right)$$

1.2.1 Conditional probability

Definition 10 (Conditional probability)**.** *Let* $\mathbb{P}$ *be a probability distribution on a sample space* Ω *and let* $A, B \subseteq \Omega$ *be events where* $\mathbb{P}(B) > 0$. *The* ***conditional probability*** *of* A ***given*** B *is defined by*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

For $\mathbb{P}(B) = 0$ *then we set* $\mathbb{P}(A|B) := 0$.

Remark 2. *Show that if $\mathbb{P}(B) > 0$ then the map $A \to \mathbb{P}(A|B)$ is a probability distribution. This is often called the **conditional distribution** of $\mathbb{P}$ **given** B*

1. *Since $\mathbb{P}(A \cap B) \geq 0$ and $\mathbb{P}(B) > 0$, then $\mathbb{P}(A|B) \geq 0$. And $A \cap B \subseteq B$, thus by Probability properties 1.3 we have $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, therefore*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1$$

2. $\mathbb{P}(\emptyset|B) = \dfrac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = 0$

   $\mathbb{P}(\Omega|B) = \dfrac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$

3. *For countable infinite sequence of pairwise disjoint events $A_1, A_2, \ldots, A_k, \ldots \subseteq \Omega$, we have*

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) &= \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \cap B\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbb{P}(B)} \\
&= \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{\infty} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)
\end{aligned}$$

1.2.2 Independent Event

Definition 11 (Independent)**.** *If we know one event happened or not have no effect on the other event, the two events should be called **independent**. In math, suppose two event A, B, and $\mathbb{P}(B) \neq 0$, if A and B are **independent**, then*

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

*and more generally, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$. Generally independent can be given as **any two events** A, B*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

1.2.3 Expectation

Definition 12 (random variable)**.** *A random process or variable with a numerical outcome.*

Definition 13 (Expected Value of A Discrete Random Variable)**.** *If X takes outcomes $x_1, \ldots, x_k$ with probabilities $\mathbb{P}(X = x_1), \cdots, \mathbb{P}(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:*

$$\mathbb{E}(X) = x_1 \cdot \mathbb{P}(X = x_1) + \cdots + x_k \cdot \mathbb{P}(X = x_k) = \sum_{i=1}^{k} x_i \mathbb{P}(X = x_i)$$

The Greek letter μ may be used in place of the notation $\mathbb{E}(X)$. And if series $\sum_{i=1}^{k} |x_i| \mathbb{P}(X = x_i)$ is divergent, then the $\mathbb{E}(X)$ doesn't exist.

Definition 14 (Expected Value of A Continuous Random Variable)**.** *If X is a continuous random variable, and its probability density function is $f_X(x)$, then the expected value of X is :*

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$$

If integral $\int_{-\infty}^{+\infty} |x| f_X(x) dx$ is divergent, then the $\mathbb{E}(X)$ doesn't exist.

In physics, the expectation holds the same meaning as the center of gravity.

Theorem 3 (Linear Combination)**.** *If X and Y are random variables, then a linear combination of the random variables is given by*

$$aX + bY$$

where $a, b \in \mathbb{R}$. Then the expectation of the combination is

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

Theorem 4. *If random varianles X and Y are independent, then*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Proposition 2. *The expectation of a function g applied on a random variable X is*

$$\mu = \mathbb{E}(g(X)) = \begin{cases} \sum_{k \in S_X} g(k) \cdot \mathbb{P}(X = k), & \text{if } X \text{ is discrete;} \\ \int_{-\infty}^{\infty} g(t) f_X(t) dt, & \text{if } X \text{ is continuous with p.d.f } f_X \end{cases}$$

1.2.4 Variance

Definition 15 (Variance)**.** *If X takes outcomes $x_1, \ldots, x_k$ with probabilities $\mathbb{P}(X = x_1), \ldots, \mathbb{P}(X = x_k)$ and expected value $\mu = \mathbb{E}(X)$, then the* ***variance*** *of X, denoted by $\mathbb{V}\mathrm{ar}(X)$ or the symbol σ^2, is*

$$\sigma^2 = (x_1 - \mu)^2 \mathbb{P}(X = x_1) + \cdots + (x_k - \mu)^2 \mathbb{P}(X = x_k) = \sum_{j=1}^{k} (x_j - \mu)^2 \mathbb{P}(X = x_j)$$

The standard deviation of X, labeled σ, is the square root of the variance.

And variance can also be defined by expectation:

$$\mathbb{V}\mathrm{ar}(X) := \mathbb{E}(X - \mathbb{E}(X))^2$$

Proposition 3. *If X is a random variable, and it has expectation $\mathbb{E}(X)$ and variance $\mathbb{V}\text{ar}(X)$, then*

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Proof.

$$\begin{aligned}\mathbb{V}\text{ar}(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 \\ &= \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2\right) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + (\mathbb{E}(X))^2 \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2\end{aligned}$$

□

Theorem 5. *Let $X_1, \dots, X_n$ be discrete or continuous random variables with well-defined expectations and variances. If $X_1, \dots, X_n$ are* ***independent****, then*

$$\mathbb{V}\text{ar}(X_1 + \cdots + X_n) = \sum_{k=1}^{n} \mathbb{V}\text{ar}(X_k)$$

1.2.5 Median

Definition 16 (Median)**.** *Let X be a discrete or continuous random variable. A value $m \in \mathbb{R}$ is called a median of X if $\mathbb{P}(X \geq m) \geq 1/2$ and $\mathbb{P}(X \leq m) \leq 1/2$.*

1.3 Discrete variables

Definition 17 (Discrete random variable)**.** *Let Ω be a discrete sample space. A function $X : \Omega \to \mathbb{R}$ is called a* ***discrete random variable****. The image of X is denoted by S_x.*

Let $\mathbb{P}$ be a probability distribution on a discrete sample space Ω and let $X : \Omega \to \mathbb{R}$ be a discrete random variable with image S_X. Then the function $p_X : S_X \to [0, 1]$ defined by $p_X(x) := \mathbb{P}(X = x)$ is a discrete probability distribution on S_X.

Definition 18 (Probability mass function and cumulative distribution function)**.** *Let $X : \Omega \to \mathbb{R}$ be a discrete random variable. The* ***(probability) mass function (p.d.f)*** *of X is the map $p_X : S_X \to [0, 1]$ given by*

$$p_X(x) := \mathbb{P}(X = x).$$

The ***(cumulative) distribution function (c.d.f)*** *of X is the map $F_X : \mathbb{R} \to [0, 1]$ given for all $t \in \mathbb{R}$ by*

$$F_X(t) := \mathbb{P}(X \leq t) = \sum_{x \in S_X, x \leq t} \mathbb{P}(X = x).$$

Proposition 4. *Let X be a discrete random variable. For its distribution function F_X, we have*

1. *F_X is monotonically increasing.*
2. *$\lim_{t \to -\infty} F_X(t) = 0$*
3. *$\lim_{t \to \infty} F_X(t) = 1$.*

1.3.1 The Binomial Distribution

Definition 19 (binomial distribution)**.** *The **binomial distribution** with parameters n and p, where $n \in \mathbb{N}$ and $p \in [0,1]$, is the probability distribution on $\{0, 1, \ldots, n\}$ given by*

$$\mathrm{bin}_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k \in \{0, \ldots, n\}.$$

A random variable X is said to follow the binomial distribution with parameters n and p if $\mathbb{P}(X = k) = \mathrm{bin}_{n,p}(k)$ for $k \in S_X = \{0, 1, 2, \ldots, n\}$. We denote this by writing $X \sim \mathrm{bin}_{n,p}$.

Proposition 5. *A random variable $X \sim \mathrm{bin}_{n,p}$, then*

$$\mathbb{E}(X) = np$$

$$\mathbb{V}\mathrm{ar}(X) = np(1-p)$$

Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k \cdot \mathrm{bin}_{n,p}(k) \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} p^k (1-p)^{n-k-1} \\
&= np(p + (1-p))^{n-1} \\
&= np
\end{aligned}$$

By formular
$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$
we just need to calculate $\mathbb{E}(X^2)$.

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{n} k^2 \cdot \text{bin}_{n,p}(k) \\
&= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} \\
&= np \left(\sum_{k=1}^{n} (k-1) \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} \right. \\
&\qquad \left. + \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} \right) \\
&= np \left(\sum_{k=1}^{n} (k-1) \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} + 1 \right) \\
&= np \left(\sum_{k=0}^{n-1} k \frac{(n-1)!}{k!(n-k)!} p^{k}(1-p)^{n-k-1} + 1 \right) \\
&= np \left(\mathbb{E}(\text{bin}_{n-1,p}) \right) \\
&= np((n-1)p + 1) \\
&= (np)^2 + np(1-p)
\end{aligned}$$

Therefore the variance of X is
$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = (np)^2 + np(1-p) - (np)^2 = np(1-p)$$

□

Proposition 6. *Let X, Y be independent random variables. If $X \sim \text{bin}_{n,p}$ and $Y \sim \text{bin}_{m,p}$ for $n, m \geq 1$ and $0 \leq p \leq 1$, then $X + Y \sim \text{bin}_{n+m,p}$.*

Proof.
$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{i=0}^{m} \mathbb{P}(Y=i)\mathbb{P}(X=k-i) \\
&= \sum_{i=0}^{m} \binom{m}{i} p^i (1-p)^{m-i} \binom{n}{k-i} p^{k-i}(1-p)^{n-k+i} \\
&= p^k (1-p)^{m+n-k} \sum_{i=0}^{m} \binom{m}{i}\binom{n}{k-i}
\end{aligned}$$

Since hypergeometric distribution is a probability distribution, we have

$$\sum_{i=0}^{m} \frac{\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}} = 1$$

Thus

$$\sum_{i=0}^{m} \binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}$$

Therefore

$$\mathbb{P}(X+Y=k) = p^k(1-p)^{m+n-k}\sum_{i=0}^{m}\binom{m}{i}\binom{n}{k-i} = p^k(1-p)^{m+n-k}\binom{m+n}{k}$$

which is equivalent to $X+Y \sim \mathrm{bin}_{m+n,p}$. □

1.3.2 The Hypergeometric Distribution

Definition 20 (Hypergeometric distribution)**.** *The **hypergeometeic distribution** with parameters $n, r, t \geq 1$ with $n, r \leq t$ is the probability distribution on $\{0, 1, \ldots, n\}$ given by*

$$\mathrm{hyp}_{n,r,t}(k) = \frac{\binom{r}{k}\binom{t-r}{n-k}}{\binom{t}{n}}, \text{ for } k \in \{0, 1, \ldots, n\}$$

A random variable X is said to follow the hypergeometric distribution with parameters n, r, t if $\mathbb{P}(X=k) = \mathrm{hyp}_{n,r,t}(k)$ for $k \in S_X = \{0, 1, \ldots, n\}$. In this case we will write $X \sim \mathrm{hyp}_{n,r,t}$

Proposition 7. *If random variable $X \sim \mathrm{hyp}_{n,r,t}$, then*

$$\mathbb{E}(X) = np$$

where $p = \frac{r}{t}$

$$\mathbb{V}\mathrm{ar}(X) = np(1-p)\frac{t-n}{t-1}$$

1.3.3 The Geometric Distribution

The geometric distribution is the probability which represents the time taken until the first success.

Definition 21 (Geometric distribution)**.** *The geometric distribution with parameter p, with $p \in (0,1)$, is the probability distribution on $\{1, 2, \ldots\}$ given by*

$$\mathrm{geo}_p(k) = p(1-p)^{k-1}, \text{ for } k \in \{1, 2, \ldots\}$$

A random variable X follows this distribution if $S_X = \{1, 2, \ldots\}$ and $\mathbb{P}(X = k) = \mathrm{geo}_p(k)$ for $k \in S_X = \{1, 2, \ldots\}$. We then write $X \sim \mathrm{geo}_p$

Proposition 8. *The **p.m.f** of Geometric distribution* geo_p *is*

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

*The **cumulative function** of geometric distribution is*

$$F_X(k) = \mathbb{P}(X \le k) = \sum_{i=1}^{k} p(1-p)^{i-1} = 1 - (1-p)^k$$

Proposition 9. *If random variable* $X \sim \mathrm{geo}_p$*, then*

$$\mathbb{E}(X) = \frac{1}{p}$$

$$\mathbb{V}\mathrm{ar}(X) = \frac{1-p}{p^2}$$

Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{\infty} k \cdot \mathrm{geo}_p(k) \\
&= \sum_{k=1}^{\infty} kp(1-p)^{k-1} \\
&= p \sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= p \cdot \frac{1}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=1}^{\infty} k^2 \cdot \mathrm{geo}_p(k) \\
&= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} \\
&= p \sum_{k=1}^{\infty} k^2(1-p)^{k-1} \\
&= p\left(\frac{2}{p^3} - \frac{1}{p^2}\right) \\
&= \frac{2}{p^2} - \frac{1}{p}
\end{aligned}$$

Therefore

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

□

Remark 3. *The above proof used some conclusions of the series* $\sum_{n=0}^{\infty} x^n$.

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

$$\sum_{k=1}^{\infty} k^2 x^{k-1} = -\frac{1}{(1-x)^2} + \frac{2}{(1-x)^3}$$

for all $x \in (-1, 1)$.

Proposition 10 (memoryless property)**.** *Let* $X \sim \text{geo}_p$ *with* $p \in (0,1)$. *Then*

$$\mathbb{P}(X > n+m | X > n) = \mathbb{P}(X > m) \text{ for all } n \geq 1, m \geq 0.$$

Proof. Because $m \geq 0$, we have

$$\{X > n+m\} \subseteq \{X > n\}$$

Thus

$$\{X > n+m\} \cap \{X > n\} = \{X > n+m\}$$

Therefore

$$\begin{aligned}
\mathbb{P}(X > n+m | X > n) &= \frac{\mathbb{P}(\{X > n+m\} \cap \{X > n\})}{\mathbb{P}(X > n)} \\
&= \frac{\mathbb{P}(X > n+m)}{\mathbb{P}(X > n)} \\
&= \frac{(1-p)^{n+m}}{(1-p)^n} \\
&= (1-p)^m \\
&= \mathbb{P}(X > m)
\end{aligned}$$

□

1.3.4 The Poisson Distribution

The poisson distribution can be seen as an approximation of binomial distribution with parameters n and p if $p \approx \lambda/n$ for some $\lambda > 0$ and n is large.

Theorem 6 (Law of small numbers)**.** *Let* $p_n, n \in \mathbb{N}$ *be a sequence of real numbers with* $0 \leq p_n \leq 1$ *for all* $n \in \mathbb{N}$ *and* $n \cdot p_n \to \lambda > 0$ *as* $n \to \infty$*, Then , for all* $k \in \{0, 1, \ldots\}$

$$\lim_{n \to \infty} \text{bin}_{n, p_n}(k) = e^{-k} \frac{\lambda^k}{k!}$$

Proof. □

Definition 22 (Poisson distribution)**.** *Let $\lambda > 0$. The* ***Poisson distribution*** *with parameter $\lambda > 0$ is the probability distribution on $\{0, 1, 2, \ldots\}$. given by*

$$\mathrm{Poi}_\lambda(k) = e^{-k} \frac{\lambda^k}{k!}$$

A random variable X is said to follow the Poisson distribution with parameter λ if $\mathbb{P}(X = k) = \mathrm{Poi}_\lambda(k)$ for $k \in S_X = \{0, 1, 2, \ldots\}$. In this case we will write $X \sim \mathrm{Poi}_\lambda$ or $X \sim \mathrm{Pois}(\lambda)$.

Remark 4. ***The value λ then corresponds to the average value of the quantity under investigation.***

Proposition 11. *Let X, Y be independent discrete random variables with $X \sim Pois(\lambda)$ and $Y \sim Pois(\mu)$ where $\lambda, \mu > 0$, then*

$$X + Y \sim Pois(\lambda + \mu)$$

Proof. Since $X \sim Pois(\lambda), Y \sim Pois(\mu)$, so we have

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

and

$$\mathbb{P}(Y = k) = \frac{\mu^k}{k!} \mathrm{e}^{-\mu}$$

we note

$$\{X + Y = k\} = \bigcup_{i=0}^{k} \{X = i \wedge Y = k - i\} = \bigcup_{i=0}^{k} \{X = i\} \cap \{Y = k - i\}$$

Because X and Y are independent, so

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \mathbb{P}\left(\bigcup_{i=0}^{k}\{X=i\}\cap\{Y=k-i\}\right) \\
&= \sum_{i=0}^{k}\mathbb{P}\left(\{X=i\}\cap\{Y=k-i\}\right) \\
&= \sum_{i=0}^{k}\mathbb{P}\left(\{X=i\}\right)\mathbb{P}\left(\{Y=k-i\}\right) \\
&= \sum_{i=0}^{k}\frac{\lambda^i}{i!}\mathrm{e}^{-\lambda}\cdot\frac{\mu^{k-i}}{(k-i)!}\mathrm{e}^{-\mu} \\
&= \mathrm{e}^{-(\lambda+\mu)}\sum_{i=0}^{k}\frac{\lambda^i\mu^{k-i}}{i!(k-i)!} \\
&= \frac{\mathrm{e}^{-(\lambda+\mu)}}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda^i\mu^{k-i} \\
&= \frac{\mathrm{e}^{-(\lambda+\mu)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda^i\mu^{k-i} \\
&= \frac{\mathrm{e}^{-(\lambda+\mu)}}{k!}(\lambda+\mu)^k
\end{aligned}$$

Therefore $X+Y$ meet the distribution $Pois(\lambda+\mu)$ ☐

Proposition 12. *If a random variable* $X \sim \mathrm{Poi}_\lambda$, *then*

$$\mathbb{E}(X) = \lambda$$

$$\mathbb{V}\mathrm{ar}(X) = \lambda$$

Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k\cdot\mathbb{P}(X=k) \\
&= \sum_{k=0}^{\infty} k\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \\
&= \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\mathrm{e}^{-\lambda} \\
&= \lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \\
&= \lambda
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot \mathbb{P}(X = k) \\
&= \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} \mathrm{e}^{-\lambda} \\
&= \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} \mathrm{e}^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \mathrm{e}^{-\lambda} \\
&= \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} + \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \\
&= \lambda^2 + \lambda
\end{aligned}
$$

Therefore

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

□

1.3.5 Independent random variables

Definition 23 (Independence of random variables)**.** *Let $X_1, \dots, X_n$ be discrete random variables with $X_i : \Omega \to \mathbb{R}$ for $i \in \{1, \dots, n\}$. We say that $X_1, \dots, X_n$ are independent if for any elements $x_1 \in S_{X_1}, x_2 \in S_{X_2}, \dots, x_n \in S_{X_n}$ we have*

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

Proposition 13. *Let $X_1, \dots, X_n$ be independent random variables. Then given $A_i \subseteq S_X$, for $i = 1, \dots, n$, we have*

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$$

Proposition 14. *Let $X_1, \dots, X_n$ be independent random variables and*

$$f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$$

be arbitrary functions. Then the random variables $f_1 \circ X_1, \dots, f_n \circ X_n$ are independent.

Proof. Let $y_1, y_2, \dots, y_n \in \mathbb{R}$. Set $A_i = \{x \in \mathbb{R} : f(x) = y_i\}$. Then

$$
\begin{aligned}
\mathbb{P}(f_1 \circ X_1 = y_1, \cdots, f_n \circ X_n = y_n) &= \mathbb{P}(X_1 \in A_1, \cdots, X_n \in A_n) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \\
&= \prod_{i=1}^{n} \mathbb{P}(f_i \circ X_i = y_i)
\end{aligned}
$$

□

Proposition 15. *Let X, Y be independent discrete random variables and $Z = X + Y$. Then , $S_z = \{x + y : x \in S_X, y \in S_Y\}$. For all $z \in S_Z$, we have*

$$\mathbb{P}(X + Y = z) = \sum_{x \in S_X} \mathbb{P}(X = x)\mathbb{P}(Y = z - x)$$

1.4 Continuous Random Variables

Definition 24 (Densities). *A function f: $\mathbb{R} \to \mathbb{R}$ is called a (probability) density function if*

1. *$f(x) \geq 0$ for all $x \in \mathbb{R}$, and*
2. *$\int_{-\infty}^{\infty} f(x)dx = 1$.*

If X is a random variable, and a function f_X is a probability of X, the $f_X(x)$ is called a PDF(p.d.f.) of X.

Definition 25 (Continuous random variables). *A function $X : \Omega \to \mathbb{R}$ is called a **continuous random variable** if there is a* p.d.f f_X *with*

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx, \ -\infty \leq a < b \leq \infty$$

Definition 26 (Cumulative Distribution function). *The cumulative distribution function (c.d.f) of a continuous random variable X with p.d.f f_X is the map $F_X : \mathbb{R} \to [0, 1]$ defined for all $t \in \mathbb{R}$ by*

$$F_X(t) := \mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(x)dx$$

The $F_X : \mathbb{R} \to [0, 1]$ is called a **c.d.f** *of X.*

Proposition 16. *The* **c.d.f** *of continuous variable has the following properties:*

1. *F_X is non-decreasing.*
2. $\lim_{t \to +\infty} F_X(t) = 1$
3. $\lim_{t \to -\infty} F_X(t) = 0$

Theorem 7. *If X is a continuous random variable, then $\mathbb{P}(X = x) = 0$ for any real number x.*

1.4.1 Uniform Distribution

Definition 27 (Uniform Distribution). *X is uniformly distributed on an interval (a, b) if it has the density function*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & if\ a < x < b \\ 0 & otherwise \end{cases}$$

*the notation of **Uniform Distribution** is $X \sim U(a, b)$ or $X \sim \mathrm{unif}_{[a,b]}$. In this case, the distribution function F_X is given by*

$$F_X(t) = \begin{cases} 0 & \text{if } t < a, \\ \frac{t-a}{b-1} & \text{if } a \leq t \leq b, \\ 1 & \text{if } t > b. \end{cases}$$

1.4.2 Normal Distribution

Definition 28 (Error function)**.** *The (Gauss) error function Φ is the map $\Phi : \mathbb{R} \to [0, 1]$ given by*

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{x^2/2} dx, \quad \textit{for } t \in \mathbb{R}.$$

Proposition 17. *The error function $\Phi(t)$meets*

- $\Phi(-t) = 1 - \Phi(t), \quad$ *for* $t \in \mathbb{R}$
- $\Phi(0) = 0.5$

Definition 29 (Normal Distribution)**.** *The* **p.d.f** *of* ***normal distribution*** *with parameters μ and σ is*

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

where $\mu \in \mathbb{R}$ and $\sigma > 0$. We write $X \sim N(\mu, \sigma^2)$. The ***distribution function*** *F_X is then*

$$F_X(t) = \Phi(\frac{t-\mu}{\sigma})$$

When $\mu = 0$ and $\sigma = 1$, this distribution is called **the standard normal distribution**.

Proposition 18. *Let $\mathcal{N} \sim N(0, 1)$ and let $\mu \in \mathbb{R}$ and $\sigma > 0$. Then*

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

Proposition 19.

1. *If $X \sim N(\mu, \sigma)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$. Hence,*

$$(X - \mu)/\sigma \sim N(0, 1)$$

2. *If $X \sim N(\mu, \sigma^2)$, then $-X \sim N(-\mu, \sigma^2)$*
3. *If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ are independent, then $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$*
4. $\Phi(-z) = 1 - \Phi(z)$
5. $\varphi(-z) = \varphi(z)$

Example 5. *$X \sim N(5, 16)$ and $Y \sim N(-5, 9)$, X and Y are independent. What is the probability $\mathbb{P}(X - Y > 15)$?*

1.4.3 Exponential Distribution

Definition 30 (Exponential Distribution)**.** *If a function meet the format*

$$f_X(x) = \lambda e^{-\lambda x} \ (when\ x \geq 0)$$

*then X meet the **exponential distribution**, note as $X \sim \mathrm{Exp}(\lambda)$, or $X \sim \exp_\lambda$. The distribution function F_X is given by*

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0, & if\ t < 0 \\ 1 - e^{-\lambda t} & if\ t \geq 0 \end{cases}$$

Proposition 20 (Memory-less property)**.** *Let $X \sim \exp_\lambda$ with $\lambda > 0$. Then for $t, s \geq 0$, we have*

$$\mathbb{P}(X \geq t + s | X \geq t) = \mathbb{P}(X \geq s)$$

Proof. For given $t \geq 0$, we have $\mathbb{P}(X \geq t) = 1 - F_X(t) = \mathrm{e}^{-\lambda t}$. Thus

$$\mathbb{P}(X \geq t + s | X \geq t) = \frac{\mathbb{P}(X \geq t + s)}{\mathbb{P}(X \geq t)} = \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda s} = \mathbb{P}(X \geq s)$$

□

Proposition 21. *If $X \sim \exp_\lambda$, then*

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

$$\mathbb{V}\mathrm{ar}(X) = \frac{1}{\lambda^2}$$

1.4.4 Erlang Distribution

Suppose we want to model the time to complete k operation requires an exponential period of time to complete.

Definition 31 (Erlang Distribution)**.** *An Erlang distribution has the density function as*

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

*We call λ is the **rate parameter** and k is the **shape parameter**. We denote it as $X \sim \mathrm{Erlang}(\lambda, \mathrm{k})$.*

Theorem 8. *Let $X_1, X_2, \ldots, X_k$ are independent and $X_i \sim \exp_\lambda$, then the sum of them $X = \sum_{i=1}^k X_i \sim \mathrm{Erlang}(\lambda, k)$.*

Example 6. *Suppose you join a queue with three people ahead of you. One is being served and two are waiting. Their service times $S1, S2, S3$ are independent exponential random variables with common mean 2 minutes. What is the probability of that you wait more than 5 minutes in the queue?*

1.4.5 Gamma Distribution

Definition 32 (Gamma Distribution)**.** *The p.m.f. of Gamma distribution is*

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$$

Proposition 22. *If* $X \sim \text{Gamma}(\lambda, k)$*, then*

$$\mathbb{E}(X) = \frac{k}{\lambda}$$

$$\mathbb{V}\text{ar}(X) = \frac{k}{\lambda^2}$$

1.4.6 χ^2 Distribution

The χ^2 distribution is a special situation of Gamma distribution.

Definition 33 (χ^2 Distribution)**.** *The* $\text{Gamma}(1/2, 1/2)$ *distribution is called the* χ^2 *distribution with 1 degree of freedom.Its notation is* χ_1^2*.*

Definition 34. *Let* $X_1, X_2, \dots, X_n$ *are independent variables and* $X_i \sim \chi_1^2$ *for* $i = 1, 2, \dots, n$*, then*

$$\chi_n^2 := \sum_{i=1}^{n} X_i$$

Therefore $\text{Gamma}(1/2, n/2)$ *can be writted as* χ_n^2*.*

1.4.7 Beta Distribution

Definition 35 (Beta Function)**.** *Beta function is given by*

$$B(a, b) = \int_0^1 cx^{a-1}(1-x)^{b-1} dx$$

where $a, b > 0$ *and* $x \in (0, 1)$*.*

Remark 5. *Beta function and Gamma function have the relationship*

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

Definition 36 (Beta Distribution)**.** *A random variable has a beta distribution with parameters a and b if its p.d.f. is*

$$f(x) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1} \mathbb{I}_{\{0<x<1\}}$$

and we denote it $X \sim \text{Beta}(a, b)$*.*

Proposition 23. *If* $X \sim \text{Beta}(a, b)$*, Then*

- $\mathbb{E}(X) = \dfrac{a}{a+b}$
- $\mathbb{V}\text{ar}(X) = \dfrac{ab}{(a+b)^2(a+b+1)}$

1.5 The law of large numbers and the central limit theorem

1.5.1 Markov and Chebyshev's inequalities

Definition 37 (indicator variable)**.** *Let $A \subseteq \Omega$ be an event. The* ***indicator variable*** *of the event A is the random variable $\mathbb{I}_A : \Omega \to \mathbb{R}$ defined for all $\omega \in \Omega$ by*

$$\mathbb{I}_A(\omega) \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

Theorem 9 (Markov's inequality)**.** *Let $X : \Omega \to \mathbb{R}$ be a non-negtive random variable with well-defined expectation. Then, given any $t > 0$, we have*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof.

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} t \cdot f(t)dt = \int_{0}^{\infty} t \cdot f(t)dt \\ &= \int_{0}^{x} t \cdot f(t)dt + \int_{x}^{\infty} t \cdot f(t)dt \\ &\geq \int_{x}^{\infty} t \cdot f(t)dt \\ &\geq \int_{x}^{\infty} x \cdot f(t)dt \\ &= x \int_{x}^{\infty} f(t)dt \\ &= x \cdot \mathbb{P}(X \geq x) \end{aligned}$$

so

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}$$

□

Theorem 10 (Chebyshev's inequality)**.** *Let X be a random variable with well-defined expeactation and variance. Then, for all $\varepsilon > 0$, we have*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}\mathrm{ar}(X)}{\varepsilon^2}.$$

Proof. Consider the random variable $Y = (X - \mathbb{E}(X))^2$. Then $|X - \mathbb{E}(X)| \geq x$ if and only if $Y \geq x^2$ and moreover Y is non-negtive. Thus, according to Markov's inequality, we obtain

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq x) = \mathbb{P}(Y \geq x^2) &\leq \frac{\mathbb{E}(Y)}{x^2} \\ &= \frac{\mathbb{V}\mathrm{ar}(X)}{x^2} \end{aligned}$$

□

1.5.2 The law of large numbers

An infinite collection of random variables $X_1, X_2, \ldots$ are independent and identically distribution if:

- $X_1 \cdots, X_n$ are independent for all $n \in \mathbb{N}$, and
- all X_i follow the same distribution, that is $F_{X_i}(t) = F_{X_j}(t)$ for all $i, j \in \mathbb{N}$ and $t \in \mathbb{R}$.

Theorem 11 (The Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}(X_i) = \mu \in \mathbb{R}$ and $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2 > 0$ for $i = 1, 2, \cdots$. For $n \in \mathbb{N}$, let*

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

denote the average of the first n random variables. Then for all $\varepsilon > 0$, we have that

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) = 1.$$

Proof. By definition of $\overline{X}_n$, we have

$$\begin{aligned} \mathbb{E}(\overline{X}_n) &= \mathbb{E}\left(\frac{\sum_{k=1}^{n} X_k}{n}\right) \\ &= \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(X_k) \\ &= \mu \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}\mathrm{ar}(\overline{X}_n) &= \mathbb{V}\mathrm{ar}\left(\frac{\sum_{k=1}^{n} X_k}{n}\right) \\ &= \frac{1}{n^2} \mathbb{V}\mathrm{ar}\left(\sum_{k=1}^{n} X_k\right) \\ &= \frac{1}{n^2} \sum_{k=1}^{n} \mathbb{V}\mathrm{ar}(X_k) \\ &= \frac{\sigma}{n} \end{aligned}$$

According to Chebyshev's inequality, we have

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathbb{V}\mathrm{ar}(\overline{X}_n)}{\varepsilon^2} = \frac{1}{n} \cdot \frac{\sigma}{\varepsilon^2}$$

Thus

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) = \lim_{n\to\infty} \big(1 - \mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon)\big) \geq 1 - \lim_{n\to\infty} \frac{\sigma}{n\varepsilon^2} = 1$$

Because $\mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) \leq 1$, therefore

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) = 1$$

$\square$

Theorem 12 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}(X_i) = \mu \in \mathbb{R}$ and $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2 > 0$ for $i = 1, 2, \cdots$ For $n \in \mathbb{N}$, let $\bar{X}_n := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ denote the average of the first n random variables. Then, $\bar{X}_n$ is approximately $N(\mu, \frac{\sigma^2}{n})$ in the following sense:*

$$\lim_{n\to\infty} \mathbb{P}(\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \leq z) = \Phi(z) \quad \textit{for all } z \in \mathbb{R}$$

Remark 6. *For discrete random variables $X_1, \cdots, X_n, X = X_1 + \cdots + X_n$ with sample space $S_X \subseteq \mathbb{Z}$, we obtain the approximate formulas*

$$\begin{aligned}
\mathbb{P}(X = k) &\approx \Phi\left(\frac{k - n\mu + 0.5}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{k - n\mu - 0.5}{\sqrt{n\sigma^2}}\right) \\
\mathbb{P}(k \leq X \leq m) &\approx \Phi\left(\frac{m - n\mu + 0.5}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{k - n\mu - 0.5}{\sqrt{n\sigma^2}}\right) \\
\mathbb{P}(X \leq m) &\approx \Phi\left(\frac{m - n\mu + 0.5}{\sqrt{n\sigma^2}}\right) \\
\mathbb{P}(X \geq k) &\approx 1 - \Phi\left(\frac{k - n\mu - 0.5}{\sqrt{n\sigma^2}}\right)
\end{aligned}$$

Lemma 2. *If $X_1, X_2, \ldots$ are independent* ***normal*** *random variables with $\mathbb{E}(X_i) = \mu \in \mathbb{R}$ and $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2 > 0$ for $i = 1, 2, \ldots,$ then*

$$X := \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

exactly.