# MATH493 Probability

Albert Peng

December 18, 2022

# Contents

# 1 Basic Probability Foundations

We define the **sample space** $\Omega$ as all possible events. Events $E$ are thus just subsets of $\Omega$ ($E \subseteq \Omega$). It is also the case that $\emptyset = \Omega^c$.

Mutually disjoint sets $A_1, A_2, ...$ are sets where $A_i \cap A_j = \emptyset$, $\forall i \neq j$.

**Theorem.** De Morgan's Laws state that

$$(A \cup B)^c = A^c \cap B^c \qquad (A \cap B)^c = A^c \cup B^c$$

The **axiomatic definition of probability** states that the probability must satisfy

1. $P(E) \geq 0$
2. $P(\Omega) = 1$
3. For mutually exclusive events $E_1, ..., E_i$,

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

# 2 Conditional Probability and Independence

**Definition.** For events $A$ and $B$ with $P(B) > 0$:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

$$P(B) = P(B \cap A) \cup P(B \cap A^C) = P(A)P(B|A) + P(A^c)P(B|A^c)$$

## 2.1 Bayes Theorem

The **Bayes Theorem** is important in allowing us to compute unknown probability given known info.

$$P(E_1 \,|E_2) = \frac{P(E_1)P(E_2 \mid E_1)}{P(E_2)} = \frac{P(E_1)P(E_2 \mid E_1)}{P(E_1)P(E_2 \mid E_1) + P(E_1^c)P(E_2 \mid E_1^c)} \tag{2}$$

**Definition.** $A$ being independent of $B$ means that knowing $B$ has occurred does not change $P(A)$. This implies that $P(A) = P(A \mid B)$, which means

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

**Corollary.** $A, B$ independent $\iff A, B^c$ independent $\iff A^c, B$ independent $\iff A^c, B^c$ independent.

Note that this is very different from disjoint events. Disjoint sets are never independent unless one or both has probability 0.

# 3 Random Variable

Mathematically, a *random variable* is a function from $\Omega$ to real numbers.

We denote the set of possible values (or state space) of $X$ by $\mathcal{X}$, where

$$\mathcal{X} = \{X(\omega) : \mathcal{X} \in \Omega\}$$

The function $p_X(x) = P(X = x), x \in \mathbb{R}$. is called the **probability mass function (PMF)**. The PMF of a discrete random variables satisfy the properties that

$$p_X(x) \geq 0 \text{ for all } x \qquad \text{and} \qquad \sum_{x \in \mathcal{X}} p_X(x) = 1$$

We can also describe probability as its **cumulative distribution function (CDF)**, where

$$F_X(x) = F(x) = P(X \leq x), \ -\infty < x < \infty$$

## 3.1 Probabilities of Limiting Sets

Supposes that events $B_1, ...,$ are increasing such that

$$B_1 \subset B_2 \subset \ldots, \qquad B = B_1 \cup B_2 \cup \ldots$$

Then we can write $B_n \uparrow B$ and

$$P(B) = P(\cup_{n=1}^{\infty} B_n) = \lim_{n \to \infty} P(B_n)$$

## 3.2 Expected Values

Suppose $X$ is discrete with PMF $p_X(x)$ for $x \in \mathcal{X}$, then

$$E(X) = \sum_{x \in \mathcal{X}} x p_X(x), \qquad \text{if } \sum_{x \in \mathcal{X}} |x| \, p_X(x) < \infty$$

**Pull-back trick** is where we assume that $\Omega$ is a countable space and expectation exists. Then

$$\mu = E(X) = \sum_{x \in \mathcal{X}} x p_X(x) = \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

Now, let $\mathcal{X} = \{x_1, x_2, ...\}$. Then $E_i = \{\omega : X(\omega) = x_i, \}, i = 1, 2, ...$ partitions $\omega$. In other words, $E_i \cap E_j = \emptyset$, $E_1 \cup E_2 \cup ...$

$$E(X_1+X_2) = \sum_{\omega \in \Omega} (X_1+X_2)(\omega) P(\omega) = \sum_{w \in \Omega} \{X_1 \omega + X_2 \omega P\} P(\omega) = \sum_{\omega \in \Omega} X_1(\omega) P(\omega) + \sum_{\omega \in \Omega} X_2(\omega) P(\omega)$$

Also let $g$ be a function and $g(X)$ is a discrete random variable. Then

$$E(g(x)) = \sum_{x \in \mathcal{X}} g(x) p_X(x), \text{ if } \sum_{x \in \mathcal{X}} |g(x)| \, p_X(x) < \infty$$

Common properties include

1. $E(aX) = a(E(X)), \exists\, a \in \mathbb{R}$
2. $E(c) = c, \exists\, c \in \mathbb{R}$

### 3.3 Variance

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

Common properties include: (proofs trivial)

1. $Var(X) \geq 0 \quad (E(X^2) \geq \mu^2)$
2. $Var(X) = 0 \implies P(X = \mu) = 1$
3. $Var(aX + b) = a^2 Var(X),\ \exists\, a, b \in \mathbb{R}$
4. $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$ does not apply in most situations.

When expanded, we have

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) - 2E[(X_1 - \mu_1)(X_2 - \mu_2)]$$

The last term is defined to be the **covariance**, where

$$Cov(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$$

### 3.4 Common Sense Random Variables

For Bernoulli situations, we have $E(x) = p$ and $Var(X) = p(1-p)$

Binomial random variables is when $n$ Bernoulli random variables are repeated, with

$$Y_n \sim Bin\,(n, p) \qquad P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$E(Y) = np \qquad Var(Y) = np(1-p)$$

### 3.5 Non-trivial things

$$E(I_1) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = \frac{1}{p}$$

$$\frac{d}{dx} \sum_{n=0}^{\infty} x^n = \frac{d}{dx} \frac{1}{1-x} = \frac{1}{(1-x)^2} = \sum_{n=0}^{\infty} n \cdot x^{n-1}$$

$$\sum_{n=0}^{\infty} n \cdot x^{n-1}(1-x) = \frac{1}{1-x}$$

Let $x = 1 - p$, then

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

To compute $E(I^2)$, we differentiate twice and get

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

With $n-1=k$,

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

$$\vdots$$

### 3.6 Middle School Level distributions: geometric

Geometric distribution has "lack of memory" property, where

$$P(I > k+n \,|\, I > n) = P(I > k)$$

*Proof. Trivial* To prove again:

$$P(I_1 = k_1 \text{ and } I_2 = k_2)$$

$$\vdots$$

$$P(I_2 = k_2) = \sum_{k_1=1}^{\infty} P(I_1 = k_1 \text{ and } I_2 = k_2) = \sum_{k_1=1}^{\infty}(1-p)^{k_1-1}\cdot p\cdot(1-p)^{k_2-1}\cdot p = ... = (1-p)^{k_2-1}\cdot p\cdot 1.$$

#### 3.6.1 Waiting Times

The pmf of $W_r$ is

$$P(w_r = n) = P(r\text{th success in } n \text{ toss})$$

Thus,

$$\begin{aligned} P(W_r = n) = P(Y_{n-1} = r-1, X_n = 1) &= \binom{n-1}{r-1}p^{r-1}(1-p)^{n-1-(r-1)}\cdot p \\ &= \binom{n-1}{r-1}p^r(1-p)^{n-r}, \qquad n = r, r+1... \end{aligned}$$

$W_r$ is known as **negative binomial** random variable with $r$ and $p$ such that $W_r \sim NB(r,p)$. It follows that

$$\sum_{n=r}^{\infty}\binom{n-1}{r-1}p^r(1-p)^{n-r} = 1$$

For expected values we know that $W_r = \sum_{i=1}^{r} I_r$ so

$$E(W_r) = \sum_{i=1}^{r} I_i = \frac{1}{p} + ... + \frac{1}{p} = \frac{r}{p}$$

$$Var(W_r) > Var(I_1) + ... + Var(I_2) = \frac{r(1-p)}{p^2}$$

An important property is that

$$[W_r > n] \equiv [Y_n < r] \implies \sum_{k=n+1}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} = \sum_{k=0}^{r-1} \binom{n}{k} p^k (1-p)^{n-k}$$

We can reach maximum peaks by comparing successive negative binomial probabilities:

$$R(y) = \frac{p_w(y)}{p_w(y-1)} = \left(\frac{r-1}{\phantom{x}}\right)$$

### 3.6.2 Hypergeometric Distribution

Given that there are $n$ successes among $N$ trials, what is the probabillity that there are $x$ successes among the first $m$ trials? Here $x$ is nonnegative integer and $m < N$.

$$P(Y_m = x \mid Y_N = n) = \frac{\binom{m}{n} \cdot \binom{N-m}{n-x}}{\binom{N}{n}}, \qquad x = \max(0, n+m-N), ..., \min(n, m)$$

$$E(H) = \frac{mn}{N} \qquad Var(H) = \frac{mn(N-m)}{N^2} \cdot \frac{N-n}{N-1}$$

Look at slides for a bit more detail

If $N \to \infty$, $\frac{n}{M} \to p$ and $H$ becomes $\text{Bin}(m, p)$

### 3.6.3 Poisson Distribution

$$\mathbb{P}(Y = y) = e^{-\lambda} \frac{\lambda^y}{y!}, \quad \mu = \mathbb{E}(X) = \lambda, \, Var(Y) = \lambda$$

Proving $E(\lambda)$ simply requires difference of infinite geometric sum.

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{y=0}^{\infty} y \, \mathbb{P}(y) \\ &= \sum_{y=0}^{\infty} \frac{y^{e^{-\lambda} \lambda^y}}{y!} = e^{-\lambda} \sum_{\lambda=1}^{\infty} \frac{\lambda^y}{(y-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \end{aligned}$$

**Law of Small Numbers:** With large $n$, small $p$ and $\lambda = np$, then

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

*Proof.* Begin with the normal formula for binomial distribution. Substituting $p = \frac{\lambda}{n}$ and considering the convergence of $n$ and $p$,

$$\frac{n!}{y!(n-y)!} \cdot \frac{\lambda^y}{n^y} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-y} = \frac{1}{y!} \cdot \lambda^y \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y} = \frac{\lambda^y e^{-\lambda}}{y!}$$

□

La Cam's theory...

# 4 Continuous Random Variables

**Definition.** A **continuous random variable** $X$ takes a continuous range of values as opposed to a discrete random variable for which the set of possible values is at most countable. Typically, a continuous random variable variable $X$ is described by a probability density function (PDF) $f_X(x)$, so that a probability within the interval $[a, b]$ is given by

$$p_X(x) = \int_a^b f_X(x)\,dx$$

We also have $F_X(x) = P(X \leq x)$, where $X$ is continuous iff $F_X(x)$ is continuous. Assume that $F_X(x)$ is differentiable and $f_X(x) = F'_X(x)$.

A density $f_X$ also satisfies

$$f_X(x) \geq 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

Note that $f_X(x)$ is not a probability, so we can have $f_X(x) > 1$.

For any continuous random variables $X$, then for any number $c$, $P(X = c) = 0$.

## 4.1 Expectation and Variance

Suppose $X$ is a continuous random variable with PDF $f(x)$. Then

$$\mu = \mathbb{E}(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx, \qquad \text{provided} \int_{-\infty}^{\infty} |g(x)|f(x)\,dx < \infty \tag{3}$$

$$\sigma^2 = Var(X) = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = E(X^2) - \mu^2 \tag{4}$$

## 4.2 Uniform Distribution

In uniform distribution, $X$ takes values in an interval $[a, b]$ :

$$f(x) = \frac{1}{(b-a)},\ a \leq x \leq b$$

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

## 4.3 Exponential Random Variables

Exponential Random Variables is useful to describe time for an event to ccur or life-time of a component, where

$$X \sim exp(\lambda), f_X(x) = \lambda e^{\lambda x},\ x \geq 0$$
$$E(X) = \frac{1}{\lambda},\ Var(X) = \frac{1}{\lambda^2}$$

This also has the memoryless property.

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

## 4.4 Poisson Process

Let $T_1, ..., T_n$ be times when an event occurs. We use $N(t)$ to dente the number of events up to and including time $t$. Also let $N(J)$ be the number of events in the time interval $J$.

Then, $N(t)$ where $t \geq 0$ is a **Poisson process** with rate $\lambda$ iff

1. Outcomes in disjoint intervals are independent. Formally, $J_1$ and $J_2$ are independent.
2. $N(t) \sim Pois(\lambda t)$ for all $t \geq 0$. In other words, $N(J) \sim Pois(\lambda |J|)$

We can conclude that

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t} \implies T_1 \sim e^{\lambda}$$

Now, let $I_1 = T_1, I_2 = T_2 - T_1, I_3 = T_3 - T_2, ...I_k = T_k - T_{k-1}$ and $\{X_n^{(m)}\}$ be outcomes of coin tossing experiments where success probability $= p_m$ and $n = \frac{1}{m}, \frac{2}{m}, ....$ Thus, Number of successes up to time $t$ is essentially number of successes in $mt$ tosses. As $m$ approaches infinity, we have $mp_m$ approach $\lambda$.

The probability of no heads up to time $t$ is

$$(1 - pm)^{mt} = (1 - \frac{\lambda_m}{m})^{mt} \to e^{-\lambda t}$$

Now, we have

$$P(T_k > t) \equiv P(N(t) < k) \equiv P(Pois(\lambda t) < k) = \sum_{i=1}^{k-1} \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

Therefore if $F_k$ is CDF of $T_k$,

$$T_k > t \equiv N(t) < k, N(t) \sim Pois(\lambda t) \implies 1 - F_k(t) = P(T_k > t) = P(Pois(\lambda t) < k) = \sum_{i=1}^{k-1} \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

Differentiating both sides and taking derivative, we have the PDF where

$$f_k(t) = \lambda \frac{e^{-\lambda t}(\lambda t)^{k-1}}{(k-1)!}$$

Here, $f_k$ is known as gamma density with parameters $k$ and $\lambda$, where

$$T_k \sim gamma(k, \lambda)$$

### 4.4.1 Gamma Variables

Continued from previous as continuation of exponential random variables. We define

$$\Gamma(r) = \int_0^\infty e^{-x} x^{r-1}\, dx,\ r > 0$$

We have $X \sim gamma(r, \lambda)$ where PDF is

$$f_X(x) = \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda}$$

since $\Gamma(r) = (r-1)!$ for integer values of $r$. This is due to the fact that $\Gamma(r) = (r-1)\Gamma(r-1)$. The CDF $F_X(x)$ is not known analytically unless $r$ is a positive integer.

$$E(X) = \frac{\Gamma(r+1)}{\Gamma(r)\lambda} = \frac{r}{\lambda}, Var(X) = \frac{r}{\lambda^2}$$

$$\begin{aligned}
E(X) = \int_0^\infty x f_X(x)\, dx &= \int_0^\infty \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x} \\
&= \frac{1}{\Gamma(r)} \int_0^\infty \lambda^r x^r e^{-\lambda x} \\
&= \frac{\Gamma(r+1)}{\lambda\Gamma(r)} \int_0^\infty \frac{\lambda^{r+1}}{\Gamma(r+1)} x^r e^{-\lambda x} \\
&= \frac{\Gamma(r+1)}{\lambda\Gamma(r)} = \frac{r\Gamma(r)}{\lambda\Gamma(r)} = \frac{r}{\lambda}
\end{aligned}$$

For the variance,

$$\begin{aligned}
E(X^2) = \int_0^\infty x^2 f_X(x)\, dx &= \int_0^\infty \frac{1}{\Gamma(r)} \lambda^r x^{r+1} e^{-\lambda x} \\
&= \frac{1}{\Gamma(r)} \int_0^\infty \lambda^r x^{r+1} e^{-\lambda x} \\
&= \frac{\Gamma(r+1)}{\lambda\Gamma(r)} \int_0^\infty \frac{\lambda^{r+1}}{\Gamma(r+1)} x^r e^{-\lambda x} \\
&= \frac{\Gamma(r+2)}{\lambda^2\Gamma(r)} = \frac{(r+1)\Gamma(r+1)}{\lambda^2\Gamma(r)} = \frac{r(r+1)}{\lambda^2}
\end{aligned}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}$$

## 4.5 Normal Distribution

Normal distribution $X$ has the density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2},\ -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0 \tag{5}$$

The *standard normal density* with $\mu = 0, \sigma = 1$ is written as $\Phi(X)$ where

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

Note the **68-95-99.7**% rule.

Normal distribution is often used to approximate a lot of distributions such as binomial, poisson, and gamma.

For $X \sim \text{Bin}(n, p)$, $n$ large and $p$ not close to 0 or 1. then

$$X \approx N(np, np(1-p))$$

For $X \sim \text{pois}(\lambda)$ and $\lambda$ large, then

$$X \approx N(\lambda, \lambda)$$

For $X \sim \text{gamma}(r, \lambda)$ and $r$ large, then

$$N \approx N\left(\frac{r}{\lambda}, \frac{r}{\lambda^2}\right)$$

When we are approximating binomial distribution $X$ with normal distribution $Y$, we should use

$$P(X \leq x) \approx P(Y \leq x + 0.5)$$

because of the discontinuity of binomial distribution.

## 4.6 Transformation of Continuous Random Variables

Suppose we have distribution $X$ and $Y$ such that $Y = aX + b$, where $a > 0$.

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(aX + b \leq y) \\ &= P(aX \leq y - b) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

$$f_Y(y) = F_Y'(y) = f_X\left(\frac{y-b}{a}\right) \frac{d}{dy}\frac{y-b}{a} = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}$$

This could be generalized as $Y = \varphi(X)$ where $\varphi$ is a monotone differentiable function. Thus,

$$F_Y(y) = P(\varphi(X) \leq y) = P(X < \varphi^{-1}(y)) = 1 - F_X(\varphi^{-1}(y)), \qquad f_Y(y) = f_X(\varphi^{-1}(y)) \left|\frac{d}{dy}\varphi^{-1}(y)\right|$$

**Example.** Suppose $Y = e^X = \varphi(X)$ and $X \sim N(\mu, \sigma^2)$.

$$X = \ln Y = \varphi^{-1}(Y) \implies \frac{d}{dy}\varphi^{-1}(Y) = \frac{1}{y} \implies f_Y(y) = f_X(\varphi^{-1}(Y))\frac{d}{dy}\varphi^{-1}(Y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\log y - \mu)^2}$$

$Y$ is known here as the log normal random variable.

With uniform random variable $Y \sim unif(0, 1)$,

$$P(Y \leq y) = P(F_X(x) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(x)) = y, 0 < y < 1$$

**Example.** $X \sim N(0, 1)$. Find PDF of $Y = X^2$.

$f_X(x) = \phi(X) \implies F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$

PDF of $Y$ is

$$f_Y(y) = \frac{d}{dy}(2\Phi(X) - 1) = 2\Phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}, y > 0 = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} y^{-\frac{1}{2}}$$

This is $Y \sim gamma(\frac{1}{2}, \frac{1}{2})$, and is also known as $\chi_1^2$ distribution (chi-square random variable with degrees of freedom = 1)

# 5 Jointly Distributed Random Variables

$X, Y$ are jointly continuous with density $f_{X,Y}$ if for any $a \leq b$ and $c \leq d$,

$$P(a \leq X \leq b,\, c \leq Y \leq d) = \int\int_{[a,b]\times[c,d]} f_{X,Y}(x, y)\, dx\, dy$$

Conceptually, the double integral is the volume under the surface given by $\{(x, y, z) : z = f_{X,Y}(x, y)\}$ over a region covered by $[a, b] \times [c, d]$.

We can get **marginal densities** from

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \qquad f_Y(x, y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

*Exercise: Prove this (not fully trivial); see slides*

Look at example from slides. (example 3 expansion below)

$$\mathbb{P}(...) = e^{-\lambda x} \cdot \frac{\lambda\delta_x e^{-\lambda\delta_x}}{1} \cdot e^{-\lambda(y-x-\delta_x)} \cdot \frac{\lambda\delta_y e^{-\lambda\delta_y}}{1} = \lambda^2 e^{-\lambda y}\delta_x\delta_y e^{-\lambda\delta_y} \approx \lambda^2 e^{-\lambda y}\delta_x\delta_y + \text{ smaller terms}$$

Calculating marginal densiities for example 3,

$$f_y(y) = \int_0^y \lambda^2 e^{-\lambda y}\, dx = \Big[\lambda^2 e^{-\lambda y} x\Big]_0^y = \lambda^2 y e^{-\lambda y}, y > 0 \implies \frac{1}{\Gamma(2)}\lambda^2 y^2 e^{-\lambda y}, y > 0$$

$$f_X(x) = \int_x^{\infty} \lambda^2 e^{-\lambda y}\, dx = \lambda e^{-\lambda x}, x > 0$$

## 5.1 Joint CDFs

We define the joint CDF of the continuous bivariate random variable as

$$F_{X,Y}(x, y) = P\left(X \leq x, Y \leq y\right) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, du\, dv$$

Maybe add in parts from slides.

## 5.2 Independence

Two random variables $X, Y$ are independent iff for every $A$ and $B$ the events $[X \in A]$ and $[Y \in B]$ are independent, then $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$. Equivalently, two events are independent iff $\forall x, y$,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \qquad f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

## 5.3 Conditional PMF and PDF

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(y)}{f_Y(y)} \qquad f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

## 5.4 Transformation of Bivariate Random Variables

Assume we have random variables $(X, Y)$ with joint PDF $f_{X,Y}(x, y)$. Let $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Let $g(g_1, g_2)$ be injective and surjective so that we can solve $X = \psi_1(U, V), \qquad Y = \psi_2(U, V)$. Then,

$$f_{U,V}(u, v) = f_{X,Y}\left(\psi_1(U, V), \psi_2(U, V)\right) \times |\text{Jacobian Determinant}|, \text{ where } \mathbb{J} = \begin{vmatrix} \frac{\partial \psi_1}{\partial U} & \frac{\partial \psi_1}{\partial V} \\ \frac{\partial \psi_2}{\partial U} & \frac{\partial \psi_2}{\partial V} \end{vmatrix}$$

**Example.** $X \sim N(0, 1)$ and $Y \sim N(0, 1)$ and $X$ and $Y$ are independent. Now, we have $U = X + Y$ and $V = X - Y$.

First, we get $X = \frac{U+V}{2}$ and $Y = \frac{U-V}{2}$. The Jacobian determinant $= -\frac{1}{2}$. We also know that $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$. Thus,

$$f_{U,V}(u, v) = f_{X,Y}\left(\frac{U+V}{2}, \frac{U-V}{2}\right) \cdot \frac{1}{2} = \frac{1}{4\pi} e^{-\frac{u^2+v^2}{4}} = \frac{1}{\sqrt{4\pi}} e^{-\frac{u^2}{4}} \cdot \frac{1}{\sqrt{4\pi}} e^{-\frac{v^2}{4}} \implies U, V \sim N(0, 2)$$

**Example.** $X, Y \sim exp(1)$ and independent. $U = X + y, V = X - Y$.

$$f_{U,V}(u, v) = f_{X,Y}\left(\frac{U+V}{2}, \frac{U-V}{2}\right) \cdot \frac{1}{2}$$

$$\begin{cases} 0 < x < \infty \\ 0 < y < \infty \end{cases} \implies \begin{cases} 0 < \frac{U+V}{2} < \infty \\ 0 < \frac{U-V}{2} < \infty \end{cases} \implies \begin{cases} -v < u < \infty \\ v < u < \infty \end{cases} \implies 0 < |v| < u < \infty$$

$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right)\frac{1}{2} = e^{-\frac{U+V}{2}} \cdot e^{-\frac{U-V}{2}} \cdot \frac{1}{2} = \frac{1}{2}e^{-u}, 0 < |v| < u < \infty$$

$$f_U(u) = \int_{-u}^{u} \frac{1}{2}e^{-u}\, dv = ue^{-u}, 0 < u < \infty \implies u \sim gamma(2,1)$$

$$f_V(v) = \int_{|V|}^{\infty} \frac{1}{2}e^{-u}\, du = \frac{1}{2}e^{-|V|}, -\infty < v < \infty$$

Here, $V$ is known as a double exponential random variable.

**Example.** $X \sim gamma(r,\lambda), y \sim gamma(s,\lambda)$. $X, Y$ independent. $U = X+Y, V = \frac{X}{X+Y}$.

Since $T1 \,|\, T_2 = t \sim unif(0,1) \implies \frac{T_1}{T_2} \,|\, T_2 = t \sim unif(0,1)$.

$$X = UV, Y = U - X = U(1-V), 0 < u < \infty, 0 < v < 1 \implies \mathbb{J} = \begin{vmatrix} V & U \\ 1-V & -U \end{vmatrix} = -U$$

$$f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y) = \frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x}\cdot \frac{1}{\Gamma(s)}\lambda^s y^{s-1}e^{-\lambda y} = \frac{1}{\Gamma(r)\Gamma(s)}\lambda^{r+s}x^{r-1}y^{s-1}e^{-\lambda(x+y)}$$

$$f_{U,V}(u,v) = f_{X,Y}\left(uv, u(1-v)\right)\cdot u = \frac{1}{\Gamma(r)\Gamma(s)}\lambda^{r+s}u^{r+s-1}e^{-\lambda u}v^{r-1}(1-v)^{s-1}$$

$$= \frac{1}{\Gamma(r+s)}u^{r+s-1}\lambda^{r+s}e^{-\lambda u} \cdot \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \cdot v^{r-1}(1-r)^{s-1}$$

Thus, $U, V$ independent with $U \sim gamma(r+s,\lambda)$ and $V \sim beta(r,s)$.

## 5.5 Beta Distribution

$U \sim beta(\alpha,\beta)$ if

$$f_U(u) = \frac{1}{B(\alpha,\beta)}u^{\alpha-1}(1-u)^{\beta-1}, \text{ where } B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \text{ and } 0 < u < 1$$

$$E(U) = \frac{\alpha}{\alpha+\beta}, Var(U) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$E(v) = \int_0^1 v\frac{1}{B(\alpha,\beta)}v^{\alpha-1}(1-\alpha)^{\beta-1} = \frac{1}{B(\alpha,\beta)}\int_0^1 v^{\beta+1-1}(1-v)^{\alpha-1}\, du = \frac{B(r+1,s)}{B(r,s)} = \frac{r}{r+s}$$

Note that $beta(1,1) \equiv unif(0,1)$

Sum of independent $\chi_1^2$ random variables

Suppose $Z_1, \ldots$ are i.i.d. $N(0,1)$ random variables. The distribution of $Z_1^2 + \ldots + Z_n^2$. Recall that $Z_1^2 \sim \chi_1^2 \equiv gamma(\frac{1}{2},\frac{1}{2}) \implies Z_1^2 + Z_2^2 \sim gamma(\frac{1}{2}+\frac{1}{2},\frac{1}{2}) \implies Z_1^2 + \ldots + Z_n^2 \sim gamma(\frac{n}{2},\frac{1}{2}) \equiv \chi_n^2$

### 5.6 Bivariate Normal Distribution

$(X,Y) \sim f_{X,Y}(x,y)$. Suppose $f_X(x) \equiv N(\mu_1, \sigma_1^2), f_Y(y) \equiv N(\mu_2, \sigma_2^2)$, we have joint probability

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right) - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right)\right]$$

With this bivariate normal distributions, we have properties

$$X \sim N(\mu_1, \sigma_1^2),\ X \sim Y(\mu_2, \sigma_2^2)$$

$$f_{X|Y}(x|y) \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)\right), \quad f_{Y|X}(y|x) \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y-\mu_1), \sigma_2^2(1-\rho^2)\right),$$

$$aX + bY \sim N\left(a\mu_1 + b\mu_2, a^2\sigma_1^2 + 2ab\rho\sigma_1\sigma_2 + b^2\sigma_2^2\right)$$

If $\rho = 0$, then $X$ and $Y$ are independent.

### 5.7 $T$ Distribution

Suppose that $Z \sim N(0,1)$, $W$ follows $\chi_n^2$ and $Z$ and $W$ are independent. Define

$$T = \frac{Z}{\sqrt{\frac{W}{n}}}$$

Then,$T$ follows $t$ dis tribution with $n$ degrees of freedom with density

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

*Proof.* Let $T = Z/\sqrt{\frac{W}{n}}$, $V = W \implies Z = T\sqrt{\frac{V}{n}}$. The jacobian matrix is

$$\det\begin{bmatrix} \frac{V}{n} & * \\ 0 & 1 \end{bmatrix} \implies \sqrt{\frac{V}{n}}$$

$$f(z,w) = f_Z(z)f_W(w) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\frac{1}{\Gamma(\frac{n}{2})}\frac{1}{2}^{\frac{n}{2}} w^{\frac{n}{2}-1}e^{-\frac{w}{2}}$$

$$f_{T,V}(t,v) = f_{Z,W}\left(t\sqrt{\frac{V}{n}}, V\right)\sqrt{\frac{V}{n}} \implies f(t,v) = \frac{1}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\left(\frac{1}{2}\right)^{\frac{n+1}{2}} V^{\frac{n+1}{2}+1}e^{-\frac{V}{2}(1+\frac{t^2}{n})}$$

Thus, find $f_T(t)$ after transformation yields

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

□

### 5.8 $F$ Distribution

Suppose that $W$ follows $\chi^2_m$, $V \sim \chi^2_n$ independent. Define

$$F = (W/m)/(V/n)$$

# 6 Properties of Expectation

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$

$X_1, ..., X_n$ are jointly continuous with density $f$ and $Y = g(X_1, ..., X_n)$. Then

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(x_1, ..., x_n) f(x_1, ..., x_n) dx_1 ... dx_n$$

Define **covariance** between $X$ and $Y$ as

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - E(X)E(Y)$$

Since $[(X-\mu_x)(Y-\mu_Y)] \leq \frac{1}{2}[(X-\mu_X)^2(Y-\mu_Y)^2] \implies 2Cov(X,Y) \leq Var(X) + Var(Y)$.

Properties:

- $Cov(X,Y) = Var(X)$
- $Cov(aX,Y) = aCov(X,Y)$
- $CoV(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$

In general,

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, X_j)$$

We define **correlation** between $X$ and $Y$ as

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}, \qquad -1 \leq \rho_{X,Y} \leq 1$$

Suppose $X, Y \sim f_{X,Y}(x,y)$ and $E(X) = \mu_1, E(Y) = \mu_2, Var(X) = \sigma_1^2, Var(Y) = \sigma_2^2$. If $U = \frac{X-\mu_1}{\sigma_1}, V = \frac{Y-\mu_2}{\sigma_2}$, then $E(U) = 0 = E(V), Var(U) = 1 = Var(V)$, $Cov(U,V) = E(UV) = Corr((X,Y)$ since

$$E(UV) = E\left(\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right) = \frac{E((X-\mu_1)(Y-\mu_2))}{\sigma_1\sigma_2} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Consider $Var(U+V) = Var(U) + Var(V) + 2Cov(U,V) = 1 + 1 + 2\rho_{U,V} \implies 2 + 2\rho_{U,V} \geq 0 \implies \rho_{U,V} \geq -1$ Similarly, $Var(U-V) = 2 - 2\rho_{U,V} \geq 0 \implies \rho_{U,V} \leq 1$.

Now, let $V = \rho U + \sqrt{1-\rho^2}Z$ where $Z \sim N(0,1)$. Then

$$\begin{aligned} E(UV) &= E(\rho U^2 + \sqrt{1-\rho^2}UZ) \\ &= E(\rho U^2) + E(\sqrt{1-\rho^2}UV) \\ &= \rho E(U^2) + \sqrt{1-\rho^2}E(UZ) \\ &= \rho + \sqrt{1-\rho^2} \cdot 0 = \rho \end{aligned}$$

## 6.1 Conditional Mean and Variance

The **Conditional mean** and **conditional variance** is defined as

$$\mu_{Y|X} = E(Y|X) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|X)\, dy$$

$$Var(Y|X) = \int_{-\infty}^{\infty} (y - \mu_{Y|X})^2 f_{Y|X}(y|X)\, dy$$

- $E(E(Y|X)) = E(Y)$
- $Var(Y) = E\left[Var(Y|X)\right] + Var\left[E(Y|X)\right]$
- $E(Y|X)$ minimizes $E(Y - f(X))^2$ for all functions $f$.

**Example.** Suppose events happpen with poisson process of rate $\lambda$. Let $X$ be the time of the first event and $Y$ be the time of the second event. The joint PDF is defined as $f_{X,Y}(x,y) = \lambda^2 e^{-\lambda y}$. Find the conditional mean variance of $X|Y$ and $Y|X$.

$$f_X(x) = \lambda e^{-\lambda x}, x > 0, \qquad f_Y(y) = \lambda^2 y e^{-\lambda y}$$

$$\implies f_{Y|X}(y|x) = \frac{\lambda^2 e^{\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)}, f_{X|Y}(x|y) = \frac{1}{y} \implies X|Y \sim Unif(0,y), Y|X \sim x + exp(\lambda)$$

Thus, $E(X|Y) = \frac{y}{2}, Var(X|Y) = \frac{y^2}{12}, E(Y|X) = X + \frac{1}{\lambda}, Var(Y|X) = \frac{1}{\lambda^2}$.

$$E(E(X|Y)) = E(\frac{Y}{2}) = \frac{2}{\lambda}\cdot\frac{1}{2} = \frac{1}{\lambda}, E(Var(X|Y)) = E\left(\frac{Y^2}{1^2}\right) = \frac{Var(Y) + (E(Y))^2}{12} = \frac{2}{\lambda^2} + \left(\frac{2}{\lambda}\right)^2 = \frac{1}{2\lambda^2}$$

$$E(E(Y|X)) = E\left(X = \frac{1}{\lambda}\right) = \frac{2}{\lambda},\ E(Var(Y|X)) = \frac{1}{\lambda^2}$$

$$Var(E(X|Y)) = Var(\frac{Y}{2}) = \frac{1}{2\lambda^2}, Var(E(Y|X)) = Var\left(X + \frac{1}{\lambda}\right) = 0$$

Look at proofs for the formula above.

For second:

$$\mu(X) = E(Y|X). RHS = E(Var(Y|X)) + Var(E(Y|X)) = E\left[E(Y^2|X) - (E(Y|X))^2\right] + Var(\mu(X))$$

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

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

For third:, $\mu(X) = E(Y|X)$

$$E(Y - f(X))^2 = E(Y - \mu(X) + \mu(X) - f(x))^2$$

For the $2ab$ term,

$$E(\underbrace{(Y - \mu(X)(\mu(X) - f(X)))}_{W}) = E(W) = E(E(W|X))$$

$$E(W|X) = E((Y - \mu(X)(\mu(X) - f(X))|X)) = (\mu(X) - f(X))E(Y - \mu(X)|X)$$

$$E(Y - \mu(X)|X) = E(Y - E(Y|X)|X) = E(Y|X) - E(Y|X) = 0 \implies E(W) = 0$$

$$\implies E(Y - f(X))^2 = E[(Y - \mu(X))^2] + \underbrace{E(\mu - f(x))^2}_{=0} = E(Y|X)$$

**Example.** In the month of October, the mean and the variance of the amount of rainfall on a rainy day are respectively 1cm and 1cm2. However, the number of rainy days in October follows $Bin(30, 0.2)$. implying that average number of rainy days in October is 6 and variance of the number of rainy days is 24 and 5. Calculate the mean and variance of total amount of rainfall in the month of October.

Let $Y_i$ be defined as the amount of rainfall on the $i$th rainy day, with $i = [1, N]$ and $N =$ total number of rainy days. Define $T$ as the total amount of rainfal, where $T = Y_1 + ... + Y_N$. $N \sim Bin(30, 0.2) \implies E(X) = 6, Var(X) = 4.8$. Given $E(Y_i) = 1, Var(Y_i) = 1$.

$E(T) = E(E(T|N))$, with $E(T|N = E(Y_1 + ... + Y_N|N) = E(Y_1) + ... + E(Y_N) = 1 + ... + 1 = N = 6 \implies E(T) = 6$

$Var(T) = \mathbb{E}(Var(T|N)) = Var(\mathbb{E}(T|N))$ where $Var(T|N) = Var(Y_1 + ... + Y_N|N) = Var(Y_1 + ... + Y_N) = N$

$$Var(T) = \mathbb{E}(N) + Var(N) = 6 + 4.8 = 10.8$$

## 6.2 Moment Generating Functions

The **moment generating functions** of a random variable $X$ denoted by $M_X(t)$ is

$$M_X(t) = E[e^{tX}]$$

This can be used to find the moments of distribution, where

$$e^{tX} = \sum_{i=0}^{\infty} \frac{(tX)^i}{i!} \implies M_X(t) = E[e^{tX}] = 1 + tE[X] + \frac{2tE[X^2]}{2!} + ... = 1 + tm_1 + \frac{2tm_2}{2!} + ...$$

In particular, $M_X(0) = 1, M'_X(0) = E[X], M''_X(0) = E[X^2], ...,$.

Look at Binomial, Poisson, Uniform, Normal, Chi-squared moment generating functions.

**Example.** Let $X \sim Bin(n, p)$.

$$\mathbb{E}(e^{tX}) = \sum_{X=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \sum_{X=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} = (e^t p + (1-p))^n$$

**Example.** Let $X \sim Pois(\lambda)$.

$$\mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda \lambda^x}}{x!}$$
$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} e^{e^t \lambda} = e^{e^t \lambda - \lambda}$$

**Example.** $X \sim exp(\lambda)$

$$\mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x}\, dx = \lambda \int_0^{\infty} e^{(t-\lambda)x}\, dx = \left[\frac{\lambda(e^{t-\lambda} x)}{t - \lambda}\right]_0^{\infty} = \frac{\lambda}{t-\lambda}, t < \lambda$$

$$\implies M(t) = \begin{cases} \mathbb{E}[e^{tx}], & t < \lambda \\ \infty, & t \geq \lambda \end{cases} M(t)$$

# 7 Limit Theorems

For notations, we say $c_n \to 0$ if, for every $\epsilon > 0$, we can find $N(\epsilon)$ such that $|c_n - c| < \epsilon$ for all $n \geq N(\epsilon)$.

Let $\{Z_n\}$ be a sequence of random variables that converges.

For $L^2$ convergence, we say $\{Z_n\} \to c$ if $\mathbb{E}(Z_n - c)^2 \to 0$.

$$\mathbb{E}(Z_n - c)^2 = E(Z_n - \mu_n + \mu_n - c)^2$$
$$= \mathbb{E}(Z_n - \mu_n)^2 + \mathbb{E}(\mu - c)^2 - 2\,\mathbb{E}[\underbrace{(Z_n - \mu_n)(\mu_n - c)}_{=0}]$$
$$= Var(Z_n) + Bias^2(Z_n) \to 0$$

if $Var(Z_n) = 0$ and $\mu_n \to c$.

In probability, we say $\{Z_n\} \to c$ if for all $\epsilon > 0$,

$$\lim_{n \to \infty} P\left(|Z_n - c| > \epsilon\right) = 0$$

**Example.** Suppose $X_1, X_2, ...$ are i.i.d. random variables with $\bar{X}_n = (X_1 + ... + X_n)/n$. Suppose $\forall i$, $\mathbb{E}(X_i) = \mu, Var(X_i) = \sigma^2$. Then $E(\bar{X}_n) = \mu, Var(\bar{X}_n) = \sigma^2/n$. $E(\bar{X}_n - \mu)^2 = Var(\bar{X}_n) = \sigma^2/n \to 0$ as $n \to \infty \implies \bar{X}_n \xrightarrow{L^2} \mu$. also

**Markov's inequality** states that for non-negative random variable $X$, then for any $a > 0$,

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

**Chebyshev's inequality** states that if $X$ is a random variable with mean $\mu$ and variance $\sigma^2$, then for any $k > 0$,

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

For epsilon $\epsilon$, we can conclude

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \ni n \to \infty \implies \bar{X}_n \xrightarrow{\mathcal{P}} \mu$$

**The Weak Law of Large Number** states that for i.i.d. random variables $X_1, ..., X_n$ with $\mathbb{E}[X_i] = \mu$, then for any $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\frac{X_1 + ... + X_n}{n} - \mu\right| \geq \varepsilon\right] \to 0 \text{ as } n \to \infty$$

We say sequence $\{X_n\}$ converges to $X$ weakly and write $X_n \to_d X$ if $F_{X_n}(x) \to F_X(x)$ for all $x$. Thus, we can approximate $P(a \leq X_n \leq b) \approx P(a \leq X \leq b)$.

**Central Limit Theorem** states that for i.i.d. random variables $X_1, ..., X_n$ with mean $\mu$ and $\sigma^2$, then

$$\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} \approx N(0, 1) \text{ as } n \to \infty$$

**Strong Law of Large Number** states that for i.i.d. random variables $X_1, ..., X_n$ with $E[X_i] = \mu$, then

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{X_1 + ... + X_n}{n} = \mu\right] = 1$$