Review of Probability

Tonglin Zhang
Discrete Distributions

Let $X$ be an integer-valued random variable. Then, its CDF (cumulative distribution function) $F(x)$ is

$$F(x) = P(X \leq x) = \sum_{a \leq x} P(X = a) = \sum_{a \leq x} p(a),$$

where $p(a)$ is the PMF (probability mass function) of $X$. Then,

- $F(x)$ is non-decreasing.
- $F(x)$ is a step function with jumps at $a$ if $p(a) > 0$.
- $p(a) = F(a) - F(a^-)$, where $F(a^-) = \lim_{x \to a^-} F(a)$.
- $F(x)$ is right-continuous at any $x \in \mathbb{R}$.
- $F(-\infty) = 0$ and $F(\infty) = 1$. 
Discrete Distributions

- **Binomial distribution**: identically and independently repeat 0–1 experiment $n$ times. The total number $X$ follows the binomial distribution, denoted by $Bin(n, p)$, where $p = P(1)$ in a single experiment. The PMF is

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x},$$

for $x = 0, 1, \cdots, n$.

- The PMF of $Poisson(\lambda)$ is

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda},$$

for $x = 0, 1, \cdots$. 
Continuous Distributions

Let $X$ be a real-valued random variable with CDF $F$. It is a continuous random variable if there exists a function $f$ such that

$$F(x) = \int_{-\infty}^{x} f(t) dt.$$ 

Then,

- $F(x)$ is continuous in $x$.
- For any $x$, $P(X = x) = 0$.
- $f(x) \geq 0$.
- $\int_{-\infty}^{\infty} f(x) dx = 1$. 
Expected Values

Let $X$ be a random variable. Then, its expected value is

$$E(X) = \int_{-\infty}^{\infty} x F(dx).$$

If $X$ is discrete, then

$$E(X) = \sum_x x p(x).$$

If $X$ is continuous, then

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

For a function $h$, there is

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)F(dx).$$

If $X$ is discrete, then

$$E[h(X)] = \sum_{x} h(x)p(x).$$

If $X$ is continuous, then

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx.$$
To define expected value, we need

\[ E(X) = \int_{-\infty}^{\infty} x |F(dx)| < \infty. \]

Otherwise, the expected value does not exist. It is used similarly for \( E[h(X)] \).

The expected value is a constant (not random).

The variance is defined by expected value, i.e.,

\[ V(X) = E[(X - EX)^2]. \]

The concept has been extended to the conditional expected value, which is random.
It is extended to multivariate random variables, which is a expected vector. Let \( \mathbf{x} = (X_1, \cdots, X_n)^\top \) be a random vector, then the \( i \)th component of \( \mathbf{E}\mathbf{x} \) is \( \mathbf{E}(X_i) \). We define the mean vector of \( \mathbf{X} \) as

\[
\mathbf{E}(\mathbf{x}) = \begin{pmatrix}
\mathbf{E}X_1 \\
\mathbf{E}X_2 \\
\vdots \\
\mathbf{E}X_n
\end{pmatrix}.
\]
It is used to define covariance and correlation. Let $X_1$ and $X_2$ be random variables. Then, the covariance between $X_1$ and $X_2$ is

$$\text{Cov}(X_1, X_2) = E[(X_1 - E X_1)(X_2 - E X_2)]$$

and the correlation between $X_1$ and $X_2$ is

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}}.$$ 

Clearly, there is $\text{Cov}(X, X) = V(X)$. The covariance has been extended to a matrix form.
Expected Values

Let \( \mathbf{x} = (X_1, \cdots, X_m) \) and \( \mathbf{y} = (Y_1, \cdots, Y_n) \) be random vectors. Then,

\[
\text{Cov}(\mathbf{x}, \mathbf{y}) =
\begin{pmatrix}
\text{Cov}(X_1, Y_1) & \text{Cov}(X_1, Y_2) & \cdots & \text{Cov}(X_1, Y_n) \\
\text{Cov}(X_2, Y_1) & \text{Cov}(X_2, Y_2) & \cdots & \text{Cov}(X_2, Y_n) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(X_m, Y_1) & \text{Cov}(X_m, Y_2) & \cdots & \text{Cov}(X_m, Y_n)
\end{pmatrix}
\]
By $\text{Cov}(X_i, Y_j) = \mathbb{E}(X_i Y_j) - \mathbb{E}(X_i)\mathbb{E}(Y_j)$. We further have

$$
\text{Cov}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix}
\mathbb{E}(X_1 Y_1) & \mathbb{E}(X_1 Y_2) & \cdots & \mathbb{E}(X_1 Y_n) \\
\mathbb{E}(X_2 Y_1) & \mathbb{E}(X_2 Y_2) & \cdots & \mathbb{E}(X_2 Y_n) \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}(X_m Y_1) & \mathbb{E}(X_m Y_2) & \cdots & \mathbb{E}(X_m Y_n)
\end{pmatrix}
- \begin{pmatrix}
\mathbb{E}(X_1)\mathbb{E}(Y_1) & \mathbb{E}(X_1)\mathbb{E}(Y_2) & \cdots & \mathbb{E}(X_1)\mathbb{E}(Y_n) \\
\mathbb{E}(X_2)\mathbb{E}(Y_1) & \mathbb{E}(X_2)\mathbb{E}(Y_2) & \cdots & \mathbb{E}(X_2)\mathbb{E}(Y_n) \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}(X_m)\mathbb{E}(Y_1) & \mathbb{E}(X_m)\mathbb{E}(Y_2) & \cdots & \mathbb{E}(X_m)\mathbb{E}(Y_n)
\end{pmatrix},
$$

$$
= \mathbb{E}(\mathbf{xy}^\top) - \mathbb{E}(\mathbf{x})\mathbb{E}(\mathbf{y})^\top,
$$

which is an $m \times n$ matrix.
Note that $\text{Cov}(y, x)$ is an $n \times m$ matrix. We have $\text{Cov}(y, x) \neq \text{Cov}(x, y)$. Actually, there is

$$\text{Cov}(y, x) = [\text{Cov}(x, y)]^T.$$

The variance-covariance matrix of $x$ is

$$V(x) = \text{Cov}(x) = \text{Cov}(x, x).$$
Mean and variance formulae

Let $u = a^\top x + b = \sum_{i=1}^{n} a_i X_i + b$ with $\mu = E x$ and $V x = \Sigma$. Let $a = (a_1, \cdots, a_m)$ be an $n$-dimensional vector and $b$ be a constant. Then,

$$E(u) = \sum_{i=1}^{m} a_i E(X_i) + b = a^\top \mu + b.$$
Let
\[ u = \sum_{i=1}^{m} a_i X_i + b = \mathbf{a}^\top \mathbf{x} + b \]
and
\[ v = \sum_{i=1}^{n} c_i Y_i + d = \mathbf{c}^\top \mathbf{y} + d. \]
Then,
\[ \text{Cov}(u, v) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i c_j \text{Cov}(X_i, Y_j) = \mathbf{a}^\top \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{c}. \]
Moreover, we have

\[ \text{Cov}(Ax, By) = ACov(x, y)B^\top \]

and

\[ V(Ax) = \text{Cov}(Ax, Ax) = A\Sigma A^\top. \]
Expected Values

Important Summary:

(a) For any fixed matrix $A$ and fixed vector $b$, there is

$$E(Ax + b) = AEx + b;$$

(b) The definition of covariance matrix is

$$\text{Cov}(x, y) = E(xy^\top) - ExE^\top y;$$

(c) The linear transformation formula of covariance matrix is

$$\text{Cov}(Ax, By) = ACov(x, y)B^\top;$$

(d) The linear transformation of the variance-covariance matrix is

$$V(Ax) = AV(x)A^\top.$$
Example 1: Suppose \( \mathbf{x} = (X_1, X_2, X_3, X_4)^\top \) is a four-dimensional normally distributed random vector with expected vector or mean vector
\[
\mathbf{\mu} = (1.5, -1.3, 0.5, 2.0)^\top
\]
and variance-covariance matrix
\[
\Sigma = \begin{pmatrix}
3.7 & 1.5 & 0.5 & 0 \\
1.5 & 2.5 & -0.7 & 0 \\
0.5 & -0.7 & 2.0 & 0 \\
0 & 0 & 0 & 1.8
\end{pmatrix}.
\]

Let
\[
\mathbf{y} = \begin{pmatrix}
0.7X_1 - 0.5X_2 - 1.2X_3 + 1.1X_4 + 0.5 \\
2.7X_1 - 1.9X_2 + 0.7X_3 - 0.1X_4 + 0.4
\end{pmatrix}.
\]

Show that \( X_4 \) and \((X_1, X_2, X_3)^\top\) are independent. Compute the mean vector and variance-covariance matrix of \( \mathbf{y} \). Compute \( \text{Cov}(\mathbf{y}, X_4) \) and \( \text{Cov}(X_4, \mathbf{y}) \).
Solution: Note that two normal random vectors are independent if and only if their covariance vector is \( \mathbf{0} \). By

\[
\text{Cov}(X_4, \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}) = (\text{Cov}(X_4, X_1), \text{Cov}(X_4, X_2), \text{Cov}(X_4, X_3)) = (0, 0, 0),
\]

we conclude that \( X_4 \) and \((X_1, X_2, X_3)^\top\) are independent.
We want to use matrix to express $\mathbf{y}$, which is

$$\mathbf{y} = \begin{pmatrix} 0.7 & -0.5 & -1.2 & 1.1 \\ 2.7 & -1.9 & 0.7 & -0.1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} + \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix}$$

$$= \begin{pmatrix} 0.7 & -0.5 & -1.2 & 1.1 \\ 2.7 & -1.9 & 0.7 & -0.1 \end{pmatrix} \mathbf{x} + \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix}.$$
Then,

\[
E(y) = \begin{pmatrix} 0.7 & -0.5 & -1.2 & 1.1 \\ 2.7 & -1.9 & 0.7 & -0.1 \end{pmatrix} E(x) + \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix}
\]

\[
= \begin{pmatrix} 0.7 & -0.5 & -1.2 & 1.1 \\ 2.7 & -1.9 & 0.7 & -0.1 \end{pmatrix} \begin{pmatrix} 1.5 \\ -1.3 \\ 0.5 \\ 2.0 \end{pmatrix} + \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix}
\]

\[
= \begin{pmatrix} 5.00 \\ 6.37 \end{pmatrix}.
\]
Further,

\[
\text{Cov}(\mathbf{y}) = \begin{pmatrix}
0.7 & -0.5 & -1.2 & 1.1 \\
2.7 & -1.9 & 0.7 & -0.1
\end{pmatrix}
\text{Cov}(\mathbf{x})
\begin{pmatrix}
0.7 & 2.7 \\
-0.5 & -1.9 \\
-1.2 & 0.7 \\
1.1 & -0.1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
4.766 & 0.744 \\
0.744 & 25.358
\end{pmatrix}.
\]

Thus,

\[
\mathbf{y} \sim \mathcal{N}\left[\begin{pmatrix} 5.00 \\ 6.37 \end{pmatrix}, \begin{pmatrix} 4.766 & 0.744 \\ 0.744 & 25.358 \end{pmatrix}\right].
\]
Note that $X_4 = (0, 0, 0, 1)x$. We have

\[
\text{Cov}(y, X_4) = \text{Cov}\left\{ \left( \begin{array}{cccc}
0.7 & -0.5 & -1.2 & 1.1 \\
2.7 & -1.9 & 0.7 & -0.1 \\
\end{array} \right) x, (0, 0, 0, 1)x \right\}
\]

\[
= \left( \begin{array}{c}
1.98 \\
-0.18 \\
\end{array} \right).
\]

and

\[
\text{Cov}(X_4, y) = \left( \begin{array}{c}
1.98 \\
-0.18 \\
\end{array} \right).
\]
Independence

Let $X_1, \cdots, X_n$ be random variables with the joint CDF $F(x_1, \cdots, x_n)$ and the marginal CDF $F_i(x_i)$, respectively. They are independent if and only if

$$F(x_1, \cdots, x_n) = \prod_{i=1}^{n} F_i(x_i).$$
Independence

Equivalently, for discrete random variables, the condition becomes

\[ p(x_1, \cdots, x_n) = \prod_{i=1}^{n} p_i(x_i), \]

and for continuous random variables, the condition becomes

\[ f(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_i(x_i), \]

where are the joint and marginal PMFs, and the joint and marginal PDFs, respectively. The joint PMF or PDF is also called the likelihood function.
If $X_1, \cdots, X_n$ are independent, then

$$E(X_i X_j) = E(X_i)E(X_j);$$

$$\text{Cov}(X_i, X_j) = 0;$$

$$\text{Corr}(X_i, X_j) = 0;$$

and

$$V(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b) = a_1^2 V(X_1) + a_2^2 V(X_2) + \cdots + a_n^2 V(X_n).$$
**Example:** Suppose \(X_1, X_2, X_3\) are independent random variables with \(E(X_1) = 1.6, E(X_2) = 0.8, E(X_3) = -0.9, V(X_1) = 2.5, V(X_2) = 1.7,\) and \(V(X_3) = 2.1.\) Let

\[ Y_1 = 0.5X_1 - 0.7X_2 - 0.3X_3 + 0.4 \]

and

\[ Y_2 = -X_1 - 1.1X_2 + 0.7X_3 + 0.2. \]

Compute \(E(Y_1), E(Y_2), V(Y_1), V(Y_2),\) and \(\text{Cov}(Y_1, Y_2).\)
Solution:

\[ E(Y_1) = E(0.5X_1 - 0.7X_2 - 0.3X_3 + 0.4) \]
\[ = 0.5E(X_1) - 0.7E(X_2) - 0.3E(X_3) + 0.4 \]
\[ = 0.51 \]

and

\[ E(Y_2) = E(-X_1 - 1.1X_2 + 0.7X_3 + 0.2) \]
\[ = -E(X_1) - 1.1E(X_2) + 0.7E(X_3) + 0.2 \]
\[ = -3.11. \]
Note that $\text{Cov}(X_i, X_j) = 0$ for all $i \neq j$. We have,

$$V(Y_1) = (0.5)^2 V(X_1) + (-0.7)^2 V(X_2) + (-0.3)^2 V(X_3)$$
$$= 1.647$$

$$V(Y_2) = (-1)^2 V(X_1) + (-1.1)^2 V(X_2) + (0.7)^2 V(X_3)$$
$$= 5.586$$

and

$$\text{Cov}(Y_1, Y_2)$$
$$= \text{Cov}(0.5X_1 - 0.7X_2 - 0.3X_3, -X_1 - 1.1X_2 + 0.7X_3)$$
$$= (0.5)(-1.0)V(X_1) + (-0.7)(-1.1)V(X_2) + (-0.3)(0.7)V(X_3)$$
$$= -0.382.$$
Example: Let $X_1, \cdots, X_n$ be iid with common mean $\mu$ and commone variance $\sigma^2$. Let

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

Compute the mean and variance of $\bar{X}$.

Solution:

$$E(\bar{X}) = E \left\{ \frac{1}{n} X_1 + \frac{1}{n} X_2 + \cdots + \frac{1}{n} X_n \right\}$$

$$= \frac{1}{n} E(X_1) + \frac{1}{n} E(X_2) + \cdots + \frac{1}{n} E(X_n)$$

$$= \frac{1}{n} \mu + \frac{1}{n} \mu + \cdots + \frac{1}{n} \mu$$

$$= \mu$$
The two formulae are used very often.
Binomial distribution

A discrete random variable $X$ is binomial if its PMF is

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

for some $p \in (0, 1)$, denoted by $X \sim \text{Bin}(n, p)$. A binomial random variable is a Bernoulli random variable if $n = 1$. If $X \sim \text{Bin}(n, p)$, then

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

and

$$\text{V}(X) = np(1 - p).$$
Poisson distribution

A discrete random variable \( X \) is Poisson if its PMF is

\[
p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \ldots,
\]

for some \( \lambda > 0 \), denoted by \( X \sim \text{Poisson}(\lambda) \). If \( X \sim \text{Poisson}(\lambda) \), then

\[
E(X) = V(X) = \lambda.
\]
Negative binomial distribution

A discrete random variable is negative binomial if its PMF is

\[ p(x) = \frac{\Gamma(x + k)}{\Gamma(k)\Gamma(x + 1)} \left( \frac{k}{\mu + k} \right)^k \left( 1 - \frac{k}{\mu + k} \right)^x, \quad x = 0, 1, \ldots, \]

for some \( \mu, k > 0 \), denoted by \( X \sim NB(\mu, k) \). If \( X \sim NB(\mu, k) \), then

\[ E(X) = \mu \]

and

\[ V(X) = \mu + \frac{\mu^2}{k}. \]
Well Known Distributions

Uniform distribution

A continuous random variable $X$ is uniform if its PDF is

$$f(x) = \frac{1}{\theta}, 0 \leq x \leq \theta,$$

for some $\theta > 0$, denoted by $X \sim \text{Uniform}(\theta)$. If $X \sim \text{Uniform}(\theta)$, then

$$E(X) = \frac{\theta}{2}$$

and

$$V(X) = \frac{\theta^2}{12}.$$
Exponential distribution

A continuous random variable $X$ is exponential if its PDF is

$$f(x) = \lambda e^{-\lambda x}, \ x > 0,$$

for some $\lambda > 0$, denoted by $X \sim \text{Exp}(\lambda)$. If $X \sim \text{Exp}(\lambda)$, then

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

and

$$V(X) = \frac{1}{\lambda^2}.$$
Well Known Distributions

Gamma distribution

A continuous random variable $X$ is gamma if its PDF is

$$f(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}, \quad x > 0,$$

for some $\alpha, \beta > 0$, denoted by $X \sim \text{Gamma}(\alpha, \beta)$. If $X \sim \Gamma(\alpha, \beta)$, then

$$E(X) = \frac{\alpha}{\beta}$$

and

$$V(X) = \frac{\alpha}{\beta^2}.$$
Beta distribution

A continuous random variable $X$ is beta if its PDF is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0, 1),$$

for some $\alpha, \beta > 0$, denoted by $X \sim Beta(\alpha, \beta)$. If $X \sim Beta(\alpha, \beta)$, then

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

and

$$V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$
Normal distribution

A continuous random variable $X$ is normal (or Gaussian) if its PDF is

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

for some $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$, denoted by $X \sim N(\mu, \sigma^2)$. If $X \sim N(\mu, \sigma^2)$, then

$$E(X) = \mu$$

and

$$V(X) = \sigma^2.$$
\( \chi^2 \) distribution

A continuous random variable \( X \) is \( \chi^2 \) if its PDF is

\[
f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, x > 0,
\]

for some \( \nu > 0 \), denoted by \( X \sim \chi^2_\nu \), where \( \nu \) is called the degrees of freedom. This is

\[
\chi^2_\nu = \Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right).
\]

If \( X \sim \chi^2_\nu \), then \( \mathbb{E}(X) = \nu \) and \( \mathbb{V}(X) = 2\nu \). There is the way to construct \( \chi^2_n \) distribution for a positive integer \( n \).
Important property

Let $Z_1, \cdots, Z_n \sim iid \ N(0, 1)$, then $X = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$. This is useful in statistics.

$t$-distribution. If $Y \sim N(0, 1)$ and $Z \sim \chi_\nu^2$ independently, then

$$X = \frac{Y}{\sqrt{Z/\nu}}$$

follows $t_\nu$-distribution, where $\nu$ is also called the degrees of freedom.
**F-distribution.** If $Y \sim \chi^2_{\nu_1}$ and $Z \sim \chi^2_{\nu_2}$ independently, then

$$X = \frac{Y/\nu_1}{Z/\nu_2} \sim F_{\nu_1, \nu_2},$$

where $\nu_1$ is called the first and $\nu_2$ is called the second degrees of freedom.
**Multivariate normal distribution.** Let $\mathbf{x}$ be an $n$-dimensional random vector. It is a multivariate normal distribution with $\mu = \mathbb{E}(\mathbf{x})$ and $\Sigma = \text{Cov}(\mathbf{x})$ if its PDF is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\Sigma)|} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mu)^\top \Sigma^{-1} (\mathbf{y} - \mu) \right\}, \mathbf{x} \in \mathbb{R}^n.$$ 

It is denoted by $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$. 
If $\mu = 0$ is a vector of zeros and $\Sigma = I$ is the identity matrix. Then,

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2}.$$ 

If $\mu = (\mu, \cdots, \mu)^\top$ and $\Sigma = \sigma^2 I$, then

$$f(x) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}.$$ 

The above two cases are often used.
Let $\mathbf{A}$ be a matrix and $\mathbf{a}$ be a vector. Then, $\mathbf{Ax} + \mathbf{a}$ is also multivariate normal. By the mean and variance formulae, we have

$$\mathbf{Ax} + \mathbf{a} \sim N(\mathbf{A} \mu + \mathbf{a}, \mathbf{A} \Sigma \mathbf{A}^\top).$$

A multivariate normal distribution is independent if and only if its variance-covariance matrix is diagonal.

*Other distributions.* We have many other distributions. Examples include, the noncentral $\chi^2$, noncentral $t$, noncentral $F$, and etc.
Example: Suppose \( X_1, \cdots, X_n \) are independent \( \mathcal{N}(0, \sigma^2) \). Then 
\( X_i/\sigma \) for all \( i \in \{1, \cdots, n\} \) are independent \( \mathcal{N}(0, 1) \). Thus

\[
\sum_{i=1}^{n} \left( \frac{X_i}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i^2
\]

is \( \chi^2_n \), indicating that

\[
\sum_{i=1}^{n} X_i^2 \sim \sigma^2 \chi^2_n.
\]
Exponential Family

Theoretically, an exponential family distribution has a PDF or a PMF by

$$f(y_i) = \exp \left[ \frac{y\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]$$

for $i \in \{1, \cdots, n\}$. We have $E(y_i) = b'(\theta_i)$ and $V(y_i) = a(\phi)b''(\theta_i)$. Exponential family distributions include

- Normal
- Binomial
- Poisson
- Many others
Convergence

Let $X_1, \cdots, X_n$ be random variables. Let $Y_n$ be a function of $X_1, \cdots, X_n$. Let $Z$ be either a random variable or a constant. Both do not depend on $X_1, \cdots, X_n$. Then, we have the following convergence:

- Convergence almost surely (with probability one):

  $$P\left( \lim_{n \to \infty} Y_n = Z \right) = 1$$

  and denoted as $X_n \overset{a.s.}{\to} Z$. This is usually too strong.
Convergence

- Convergence in probability:

\[ \lim_{{n \to \infty}} P(\left| Y_n - Z \right| \geq \epsilon) = 0 \]

for any \( \epsilon > 0 \), and denoted by \( X_n \xrightarrow{P} Z \). This is often used.

- Convergence in distributions:

\[ P(Y_n \leq x) = P(Z \leq x) \]

if \( x \) is a continuous point of \( P(Z \leq x) \), the CDF of \( Z \), denoted by \( Y_n \xrightarrow{D} Z \).
There is
\[ Y_n \xrightarrow{a.s.} Z \Rightarrow Y_n \xrightarrow{P} Z \Rightarrow Y_n \xrightarrow{D} Z. \]

The method to show convergence in probability is Chebyshev’s inequality. Let \( \mu = \mathbb{E}(X) \) and \( \sigma^2 = \mathbb{V}(X) \). Then,
\[
P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.
\]

The method to show convergence in distribution is the characteristic function. Since it is too complicated, it will not be taught.
The CLT (central limit theorem) states that the distribution of a random variable or a random vector is approximately normal. Theoretically, one needs to prove whether the CLT holds. In practice, one can simply use the conclusion without the proof. This is usually correct. Many well-known conclusions have been obtained already. To apply the CLT, we only need the expected value and variance. Then, we can approximately compute the probabilities.
Central Limit Theorem

- If $X \sim Bin(n, p)$, then
  \[ X \overset{\text{approx}}{\sim} N(np, np(1 - p)) \]
  when $n$ is large (e.g. $n > 20$).

- If $X \sim \text{Poisson}(\lambda)$, then
  \[ X \overset{\text{approx}}{\sim} N(\lambda, \lambda) \]
  when $\lambda$ is large (e.g. $\lambda > 5$).
Example: Suppose \( X \sim Bin(10^4, 0.45) \). Compute \( P(4430 \leq X \leq 4570) \) by the CLT.

Solution: By \( E(X) = np = 10^4(0.45) = 4500 \) and \( V(X) = 10^4(0.45)(1 - 0.45) = 2475 \). We have

\[
P(4430 \leq X \leq 4570) = P(4430 \leq Bin(10^4, 0.45) \leq 4570) \\
\approx P(4430 \leq N(4500, 2475) \leq 4570) \\
= \Phi\left(\frac{4570 - 4500}{\sqrt{2475}}\right) - \Phi\left(\frac{4430 - 4500}{\sqrt{2475}}\right) \\
= \Phi(1.41) - \Phi(-1.41) \\
= 0.8414.
\]
**Example:** Suppose \( X \sim \text{Poisson}(150) \). Compute \( P(130 \leq X \leq 170) \) by the CLT.

**Solution:** By \( E(X) = V(X) - 150 \). We have

\[
P(130 \leq X \leq 170) = P(130 \leq \text{Poisson}(150) \leq 170) \\
\approx P(130 \leq N(150, 150) \leq 170) \\
= \Phi\left(\frac{170 - 150}{\sqrt{150}}\right) - \Phi\left(\frac{130 - 150}{\sqrt{150}}\right) \\
= \Phi(1.63) - \Phi(-1.63) \\
= 0.8969.
\]
Example: Flip a balanced die many times. Let \( T \) be the total number of the results. Let \( X_i \) be the result of the \( i \)th time. Then, \( T = \sum_{i=1}^{n} X_i \). Let \( \mu = \text{E}(X_i) \) and \( \sigma^2 = \text{V}(X_i) \). By the PMF of \( X_i \) given by \( P(X_i = 1) = \cdots = P(X_i) = 6 = 1/6 \). We obtain \( \mu = 3.5 \) and \( \sigma^2 = 2.9167 \). Note that the experiment is independent. We have

\[
\text{E}(T) = \text{E}\left(\sum_{i=1}^{n} X_i\right) = n\mu - 3.5n
\]

and

\[
\text{V}(T) = \text{V}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2 = 2.9167n.
\]

Thus,

\[
T \overset{\text{approx}}{\sim} N(3.5n, 2.9167n).
\]
Let $\bar{X} = T/n$. Then

$$E(\bar{X}) = E\left(\frac{T}{n}\right) = \frac{E(T)}{n} = 3.5$$

and

$$V(\bar{X}) = V\left(\frac{T}{n}\right) = \left(\frac{1}{n}\right)^2 V(T) = \frac{2.9167}{n}.$$ 

Thus,

$$\bar{X} \approx N(3.5, \frac{2.9167}{n}).$$
We can compute $P(|\bar{X} - 3.5| < 0.01)$ for a given $n$. For example, we have

$$P(|\bar{X} - 3.5| \leq 0.01) = P(3.5 - 0.01 \leq \bar{X} \leq 3.5 + 0.01)$$

$$= P(3.49 - 0.01 \leq \bar{X} \leq 3.51)$$

$$\approx P(3.49 \leq N(3.5, \frac{2.9167}{n}) \leq 3.51)$$

$$= \Phi(\frac{3.51 - 3.5}{\sqrt{2.9167/n}}) - \Phi(\frac{3.49 - 3.5}{\sqrt{2.9167/n}})$$

$$= \left\{ \begin{array}{ll}
0.1469, & n = 10^3, \\
0.4418, & n = 10^4, \\
0.9359, & n = 10^5.
\end{array} \right.$$

**Remark:** Exact computation by permutation or combination methods is impossible. One probably can only derive it for small $n$, such as $n = 2$ or 3.
Example: Let $X_1, \cdots, X_n$ be iid with common $\mu$ and $\sigma^2$. Let $\epsilon$ and $\alpha$ be positive. Use Chebyshev’s inequality. Find an $n$ such that

$$P(\left| \bar{X} - \mu \right| \leq \epsilon) \geq 1 - \alpha.$$

Solution: By

$$P(\left| \bar{X} - \mu \right| \geq \epsilon) \leq \frac{V(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2},$$

we have

$$P(\left| \bar{X} - \mu \right| \leq \epsilon) = 1 - P(\left| \bar{X} - \mu \right| > \epsilon) \geq 1 - P(\left| \bar{X} - \mu \right| \geq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}.$$
Central Limit Theorem

We only need to make

\[ 1 - \frac{\sigma^2}{n\epsilon^2} \geq 1 - \alpha, \]

which can be reached by

\[ \frac{\sigma^2}{n\epsilon^2} \leq \alpha \iff n \geq \frac{\sigma^2}{\alpha\epsilon^2}. \]

Clearly, \( P(|\bar{X} - \mu| \leq \epsilon) \) goes to 1 as \( n \to \infty \).
Summary

- Normal and related distributions include $\chi^2$, $t$-, and $F$-distributions. They are important.
- Constructions of these distributions are useful in statistics.