Review of Probability [1]

Jasmine(Yu) Hao

Faculty of Business and Economics
Hong Kong University

June 8, 2021

[1] This section is based on Stock et al. (2020), Chapter 2.

Probabilities and Outcomes

- An **outcomes** is a specific result:
  - Coin toss: either H or T.
  - Roll of dice: $1, 2 \ldots, 6$.
- The **probability** of an **outcome** is the proportion of the time that the outcome occurs in the long run.
  - Fair coin toss: 50 % chance of heads.
- The **sample space** is the set of all possible outcomes.
  - In a coin flip the sample space is $S = \{H, T\}$.
  - If two coins are flipped the sample space is $S = \{HH, HT, TH, TT\}$.
- An **event** is a subset of the sample space.
  - Roll a die $A = \{1, 2\}$.

Probability Function

Definition 1 (Probability Function)

A function $\mathbb{P}$ which assigns a numerical value to events is called a probability function if it satisfies the following Axioms of Probability:

1. $\mathbb{P}(A) \geq 0$.
2. $\mathbb{P}(S) = 1$.
3. If $A_1, A_2, \ldots$ are disjoint then $\mathbb{P}| \cup_{j=1}^{N} A_j| = \sum_{j=1}^{N} A_j$.

Events

Theorem 2 (Properties of probability functions)

For two events, A and B,

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(A) \leq 1$.
4. **Monotone Probability Inequality**: *If* $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.
5. **Inclusion-Exclusion Principle**: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
6. **Boole's Inequality**: $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.
7. **Bonferroni's Inequality**: $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$.

Exercise: show **Bonferroni's Inequality**.(Hint: use the above theorems.)

Conditional Probability

Definition 3 (Conditional Probability)

If $\mathbb{P}(B) > 0$, then the **conditional probability** of A given B is given by

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

$\mathbb{P}(B)$ is the **marginal probability** of event B.

$$\mathbb{P}(B) = \mathbb{P}(A \cup B) + \mathbb{P}(A^c \cup B).$$

Joint Events

Take two events H and C.

- let be H the event that an individual's monthly wage exceeds RMB 8000,
- let M be the event that the individual has a master's degree.

Table: Joint Distribution

	Master degree	Non-master degree	Any education
High wage	0.19	0.12	0.31
Low wage	0.17	0.52	0.69
Any wage	0.36	0.64	1.00

Conditional Probability - Example

The probability of earning a high wage conditional on high education is

$$
\begin{aligned}
&\mathbb{P}(\text{High wage}|\text{Master degree}) \\
=&\frac{\mathbb{P}(\text{High wage} \cup \text{Master degree})}{\mathbb{P}(\text{High wage} \cup \text{Master degree}) + \mathbb{P}(\text{Low wage} \cup \text{Master degree})} \\
=&\frac{0.16}{0.36} = 0.53.
\end{aligned}
$$

Similarly, the probability of earning a high wage conditional on non-master degree is
$\mathbb{P}(\text{High wage}|\text{Non-master degree}) = \frac{0.12}{0.64} = 0.19.$

Independence

Definition 4 (Independence)

The events A and B are independent if $\mathbb{P}(A \cup B) = \mathbb{P}(A)\mathbb{P}(B)$

Theorem 5 (Independence)

If A and B are independent with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, then

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

Some facts:

- When events are independent then joint probabilities can be calculated by multiplying individual probabilities.
- If A and B are disjoint then they cannot be independent.

Bayes Rule

Theorem 6 (Bayes Rule)

If $\mathbb{P}(A) > 0$ *and* $\mathbb{P}(B) > 0$ *then*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}.$$

Proof.

Use the Properties of probability functions. □

Random Variables

Definition 7 (Random variable)

A random variable is a real-valued outcome; a function from the sample space S to the real line $\mathbb{R}$.

For example, X is a mapping from the coin flip sample space to the real line, with T mapped to 0 and H mapped to 1.

$$X = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T.} \end{cases}$$

Properties of random variables.

- The **expected value** is the long-run average of the random variable.
- The **standard deviation** measures the spread of a probability distribution.

Discrete Random Variables

The set $\mathcal{X}$ is discrete if it has a finite or countably infinite number of elements.

Definition 8 (Discrete random variable)

If there is a discrete set $\mathcal{X}$ such that $\mathbb{P}(X \in \mathcal{X}) = 1$ then X is a discrete random variable.

The smallest set $\mathcal{X}$ with this property is the support of X.

Definition 9 (Probability mass function)

The probability mass function of a random variable is $\pi(x) = \mathbb{P}(X = x)$, the probability that X equals the value x.

- The **probability distribution** of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur.

Expectation

For a dsicrete variable X with the support of $\mathcal{X}$, the expctation is computed as

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} \pi(x)x.$$

- $X = 1$ with the probability of p and $X = 0$ with probability $1 - p$. The expected value is $\mathbb{E}(X) = 1 * p + 0 * (1 - p) = p$.

The expectation of the function of X, $g(X)$ is computed as

$$\mathbb{E}(g(X)) = \sum_{x \in \mathcal{X}} \pi(x)g(x).$$

St. Petersburg Paradox

Note that the expectation of a distribution does not neccesarily exist.

- A single player in which a fair coin is tossed at each stage.
- The initial stake begins at \$ 2 and is doubled every time heads appears.
- The first time a tail appears, the game ends and the player wins whatever is in the pot.
- How much would a rational agent pay to get in the bet.

X has the support of $\mathcal{X} = \{2^k : k = 1, \ldots\}$. The probability distribution is defined by $\pi(2^k) = 2^{-k}$. Then the expectation of X is

$$\mathbb{E}(x) = \sum_{k=1}^{\infty} 2^k \pi(2^k) = 1 + 1 \ldots = \infty.$$

Cumulative distribution function

The **cumulative probability distribution(**CDF**)** is the probability that the random variable is less than or equal to a particular value,

$$F(x) = \mathbb{P}(X \leq x),$$

where the probability event is $X \leq x$.

Theorem 10 (Properties of a CDF)

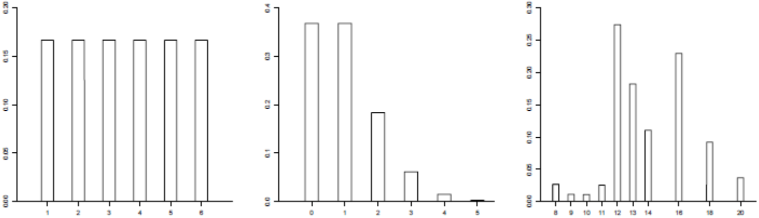
If $F(x)$ is a distribution function, then

1. *$F(x)$ is non-decreasing.*
2. $\lim_{x \to -\infty} F(x) = 0$.
3. $\lim_{x \to \infty} F(x) = 1$.
4. *$F(x)$ is right-continuous,* $\lim_{x \downarrow x_0} F(x) = F(x_0)$.

Example-Probability mass function

Some examples for discrete variables.

- For a fair dice toss, the support is $\mathcal{X} = \{1, 2 \ldots, 6\}$ with the probability mass function is $\pi(x) = \frac{1}{6}$ for $x \in \mathcal{X}$.
- An example of infinite countable random variable is the Poisson distribution, the probability mass function is $\pi(x) = \frac{e^{-1}}{x!}, x = 0, 1, \ldots$.

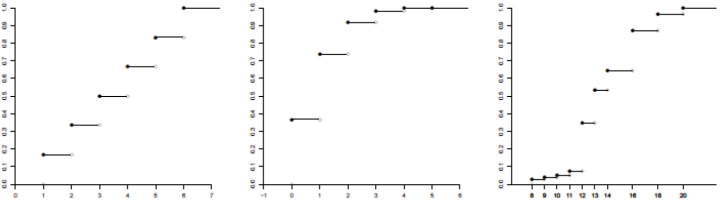


(a) Die Toss (b) Poisson (c) Education

Example-Probability function

Some examples for discrete variables.

- For a fair dice toss, the support is $\mathcal{X} = \{1, 2 \ldots, 6\}$ with the probability mass function is $\pi(x) = \frac{1}{6}$ for $x \in \mathcal{X}$.
- An example of infinite countable random variable is the Poisson distribution, the probability mass function is $\pi(x) = \frac{e^{-1}}{x!}, x = 0, 1, \ldots.$



(a) Die Toss (b) Poisson (c) Education

Continuous Random Variables

- The **probability density function(**$p.d.f$**)** area under the probability density function between any two points is the probability that the random variable falls between those two points.
  - the probability for a continous variable to take any value is 0.
  - definition is different from discrete random variables.
- When $F(x)$ is differentiable, the density function is $f(x) = \frac{dF(x)}{dx}$.

Theorem 11 (Properties of density function)

A function $f(x)$ is a density function if and only if

- $f(x) \geq 0 \forall x.$
- $\int_0^\infty f(x)dx = 1.$

Example - Continuous Variables

- Uniform distribution. The CDF is
  $$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \leq x \leq 1 \\ 1 & \text{if} \quad x > 1 \end{cases}.$$ The PDF is
  $$f(x) = \begin{cases} 1 & \text{if} \qquad 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}.$$
- Exponential distribution. The CDF is
  $$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 - \exp(-x) & \text{if} x \geq 0 \end{cases}.$$ The PDF is
  $$f(x) = \exp(-x), x \geq 0.$$

Expectation

If X is a continuous random variable with the density function $f(x)$, its expectation is defined as

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

when the integral is convergent.
The expectation of the function of X, $g(X)$ is computed as

$$\mathbb{E}(g(X)) = \sum_{x \in \mathcal{X}} \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Some examples:

- $f(x) = 1$ if $0 \leq x \leq 1$, $\mathbb{E}(X) = \int_0^1 xf(x) = 0.5$.
- $f(x) = \exp(-x)$ if $x \geq 0$,
  $\mathbb{E}(X) = \int_0^{\infty} x \exp(-x)dx = 1$(integration by part).

Mean, variance and Higher Moment

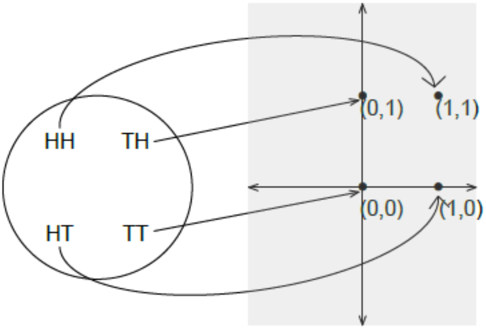
Suppose X is a random variable(either discrete or continous).

- The **mean** of X is $\mu = \mathbb{E}(X)$.
- The **variance** of X is $\sigma^2 = var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$.
  - The **standard deviation** of X is the positive root of the variance, $\sigma = \sqrt{\sigma^2}$.
- The $m-th$ **moment** of X is $\mu'_m = \mathbb{E}(X^m)$ and the $m-th$ **central moment** of X is $\mu_m = \mathbb{E}((X - \mathbb{E}(X))^m)$.
  - The **skewness** of X is defined as $skewness = \frac{\mathbb{E}((X-\mathbb{E}(X))^3)}{\sigma^3}$. If the distribution is symmetric, the skewness is 0.
  - The **kurtosis** of X is defined as $skewness = \frac{\mathbb{E}((X-\mathbb{E}(X))^4)}{\sigma^4}$.

Bivariate random variables

A pair of **bivariate random variables** is a pair of numerical outcomes; a function from the sample space to $\mathbb{R}^2$.
A pair of bivariate random variables are typically represented by a pair of uppercase Latin characters such as (X, Y). Specific values will be written by a pair of lower case characters, e.g. (x, y).

Figure: Tossing two coins



Joint distribution functions

The **joint distribution function(Joint CDF)** of (X, Y) is defined as $F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\})$.

- A pair of random variables is **discrete** if there is a discrete set $(P) \in \mathbb{R}^2$ such that $\mathbb{P}((X, Y) \in \mathscr{P}) = 1$.
  - The set $\mathcal{P}$ is the support of (X, Y) and consists of a set of points in $\mathbb{R}^2$.
  - The **joint probability mass function** is defined as $p(x, y) = \mathbb{P}(X = x, Y = y)$.
- The pair (X, Y) has a continuous distribution if the joint distribution function $F(x, y)$ is **continuous** in (x, y).
  - When $F(x, y)$ is continuous and differentiable its **joint density (joint PDF)** $f(x, y)$ equals $f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$.

Bivariate expectation

The **expected value** of real-valued $g(X, Y)$ is

$$\mathbb{E}(g(X,Y)) = \sum_{(x,y)\in\mathbb{R}^2, \pi(x,y)>0} g(x,y)\pi(x,y),$$

for discrete variables and

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy.$$

Marginal distribution

The **marginal distribution(marginal CDF)** of X is

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq \infty) = \lim_{y \to \infty} F(x, y).$$

- In the continuous case,

$$F_X(x) = \lim_{y \to \infty} \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v) du dv = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f(u, v) du dv.$$

- The **marginal densities(marginal PDF)** of X is the derivative of the marginal CDF of X,

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^{\infty} \int_{-\infty}^{x} f(u, v) du dv = \int_{-\infty}^{\infty} f(x, y) dy.$$

- Similarly, the marginal PDF of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Conditional distributions I

The conditional cumulative distributions:

- The **conditional distribution function** of Y given $X = x$ is

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y | X = x)$$

  for any x such that $\mathbb{P}(X = x) > 0$, If X has a discrete distribution.

- For continous X, Y, the **conditional distribution** of Y given $X = x$ is

$$F_{Y|X}(y|x) = \lim_{\epsilon \downarrow 0} \mathbb{P}(Y \leq y | x - \epsilon \leq X \leq x + \epsilon).$$

  If $F(x, y)$ is differentiable w.r.t x and $f_X(x) > 0$,

$$F_{Y|X}(y|x) = \frac{\frac{\partial}{\partial x} F(x, y)}{f_X(x)}.$$

The conditional density:

- For continous variable (X, Y), the conditional density function (conditional PDF) is defined by $f_{Y|X}(y|x) = \frac{d}{dy} f_{Y|X}(y|x)$.

Independence I

- Recall that two events A and B are independent if the probability that they both occur equals the product of their probabilities, thus $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- Consider the events $A = \{X \leq x\}$ and $B = \{Y \leq y\}$.
- The probability that they both occur is $\mathbb{P}(A \cap B) = \mathbb{P}(X \leq x, Y \leq y) = F(x, y)$.
- If $F(x, y) = F_X(x)F_Y(y)$ then $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- The random variables X and Y are **statistically independent** if for all x, y, $F(x, y) = F_X(x)F_Y(y)$.
- The discrete random variables X and Y are **statistically independent** if for all x, y, $\pi(x, y) = \pi_X(x)\pi_Y(y)$.
- If X, Y have differentiable density function, X, Y are statistically independent if $f(x, y) = f_X(x)f_Y(y)$.

Independence II

Theorem 12

If X, Y are independent and continuously distributed, then

$$f_{Y|X}(y|x) = f(y),$$
$$f_{X|Y}(x|y) = f(x).$$

Theorem 13

(Bayes Theorem for Densities)

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty} f(x,y)dy}.$$

Independence III

Theorem 14

If X and Y are independent then for any functions, $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}|g(X)| < \infty$ and $\mathbb{E}|h(Y)| < \infty$, then

$$\mathbb{E}(g(X)h(G)) = \mathbb{E}_X(g(X))\mathbb{E}_Y(h(Y)).$$

Covariance and correlation I

- If X and Y have finite variances, the **covariance** between X and Y is

$$cov(X,Y) = \mathbb{E}(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

- The **correlation** between X and Y is

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}.$$

Covariance and correlation II

- If X and Y are independent with finite variances, then X and Y are uncorrelated.
  - The reverse is not true. For example, suppose that $X \sim U[-1, 1]$. Since it is symmetrically distributed about 0 we see that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^3] = 0$. Set $Y = X^2$. Then $cov(X, Y) = \mathbb{E}[X^3] - \mathbb{E}[X^2]\mathbb{E}[X] = 0$. Thus X and Y are uncorrelated yet are fully dependent!
- If X and Y have finite variances, $\text{var}(X, Y) = \text{var}(X) + \text{var}(Y) + 2cov(X, Y)$.
- If X and Y are independent, $\text{var}(X, Y) = \text{var}(X) + \text{var}(Y)$.

Conditional Expectation

Just as the expectation is the central tendency of a distribution, the conditional expectation is the central tendency of a conditional distribution.

The **conditional expectation(conditional mean)** of Y given $X = x$ is the expected value of the conditional distribution $F_{Y|X}(y|x)$ and is written as $\mathbb{E}(Y|X = x)$.

- For discrete random variables, it is defined as

$$\mathbb{E}(Y|X = x) = \frac{\sum_y y\pi(x, y)}{\pi_X(x)}.$$

- For continuous random variables, it is defined as

$$\mathbb{E}[Y|X = x] = \frac{\int_y yf(x, y)}{f_X(x)}.$$

Law of Iterated Expectations I

- The function $\mathbb{E}[Y|X = x]$ is not random, it is a feature of the distribution function. But it is useful to treat the conditional expectation as a random variable.
- Consider $m(X) = \mathbb{E}[Y|X]$ a transformation of X.
- We can take expectation with respect to $m(X)$

Theorem 15 (Law of Iterated Expectations(LIE))

If $\mathbb{E}[Y] < \infty$, *then* $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.

Law of Iterated Expectations II

Proof.

For discrete random variables X, Y,

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X]] &= \sum_x \pi_X(x)\mathbb{E}[Y|X=x] \\
&= \sum_x \pi_X(x) \frac{\sum_y y\pi(x,y)}{\pi_X(x)} \\
&= \sum_x \sum_y y \\
&= \mathbb{E}[Y].
\end{aligned}
$$

Law of Iterated Expectations III

Proof.(Cont.)

For continuous random variables X, Y,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y|X]] &= \int_{-\infty}^{\infty} f_X(x)\mathbb{E}[Y|X=x]dx \\ &= \int_{-\infty}^{\infty} f_X(x)\frac{\int_{-\infty}^{\infty} yf(x,y)fy}{\pi_X(x)}dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\ dy\ dx \\ &= \mathbb{E}[Y].\end{aligned}$$

$\square$

Conditional Variance

The **conditional variance** of Y given $X = x$ is the variance of the condition distribution $F_{Y|X}(y|x)$ and is written as $\text{var}(Y|X = x)$ or $\sigma^2_Y(x)$. It equals

$$\text{var}(Y|X = x) = \mathbb{E}[(Y - m(x))^2|X = x],$$

where $m(x) = \mathbb{E}[Y|X = x]$.
Note that $\text{var}(Y) = \mathbb{E}[\text{var}(Y|X)] + \text{var}(\mathbb{E}[Y|X])$, where

- $\mathbb{E}[\text{var}(Y|X)]$ is the **within group variance**.
- $\text{var}(\mathbb{E}[Y|X])$ is the **across group variance**.

Definition 16 (Standard Normal Dist.)

A random variable Z has the **standard normal distribution**, write $Z \sim N(0,1)$, if it has the density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}), x \in \mathbb{R}.$$

Note the standard normal density is typically written as $\phi(x)$. The CDF does not have closed form but is written as $\Phi(x)$.
If $X \sim N(\mu, \sigma^2)$ and $\sigma > 0$ then X has the density

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), x \in \mathbb{R}.$$

Multivariate Normal I

Let $Z_1, \ldots, Z_m$ be i.i.d $N(0,1)$. The joint density is the product of the marginal densities:

$$\begin{aligned} f(x_1, \ldots, x_m) &= f(x_1) \ldots f(x_m) \\ &= \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{m} x_i^2\right). \end{aligned}$$

Multivariate Normal II

Let $\mathbf{Z} = [Z_1, \ldots, Z_m]^\top$ be an m-component random vector following standard normal distribution $\mathbf{Z} \sim N(0, \mathbf{I}_m)$ and $\mathbf{X} = \mu + \mathbf{BZ}$ for $q \times m$ matrix $\mathbf{B}$, then $\mathbf{X}$ has the **multivariate normal distribution**, written as $\mathbf{X} \sim N(\mu, \Sigma)$, where $\Sigma = \mathbf{BB}^\top$.
The PDF of $\mathbf{X}$ is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{q/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{(x-\mu)^\top \Sigma^{-1}(x-\mu)}{2}\right).$$

Multivariate Normal III

Properties of multivariate normal distributions.

1. Any linear combination of $X_1, \ldots, X_m$ is normally distributed.
2. The marginal distribution of each random variable is normal.
3. If the covariance of X_1 and X_2 is 0, then X_1 and X_2 are independent. The reverse is true.
4. If X_1 and X_2 are normally distributed with the joint density of $f(x_1, x_2)$, then the marginal distribution of X_1 given X_2 is a linear function of X_2: $\mathbb{E}[X_1|X_2 = x_2] = a + bx_2$.

χ-squared, t and F distribution

The **chi-squared distribution** is the distribution of the sum of m squared independent standard normal distributed variables.

- Let $\mathbf{Z} \sim N(0, \mathbf{I}_m)$ be multivariate standard normal, then $\mathbf{Z}^\top \mathbf{Z} \sim \chi^2_m$.
- If $mbX \sim N(0, \Sigma)$ with Σ positive definite, then $\mathbf{X}^\top \Sigma^{-1} \mathbf{X}$.
- Let $Q_m \sim chi^2_m$ and $Q_r \sim chi^2_r$ be independent. Then $\frac{Q_m/m}{Q_r/r} \sim F_{m,r}$.
- Let $Z \sim N(0,1)$ and $Q_m \sim chi^2_m$ be independent, then $\frac{Z}{\sqrt{Q_r/r}}$ follows the t-distribution with m degree of freedom, t_m.

Sampling I

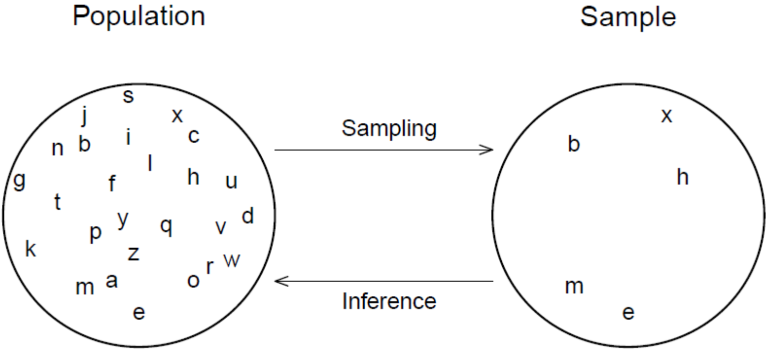
Definition 17 (independent and identically distributed(i.i.d))

The collection of random vectors $\{X_1, \ldots, X_n\}$ are **independent and identically distributed(i.i.d)** if they are mutually independent with identical marginal distributions.

- A collection of random vectors $\{X_1, \ldots, X_n\}$ is a **random sample** from the population F if X_i are i.i.d with distribution F.
- The distribution F is called the **population distribution**. We refer to the distribution as the **data generating process(DGP)**.
- The **sample size** n is the number of individuals in the sample.

Sampling II

Figure: Sampling and Inference



Identification [2]

[2]For further reading, see Hansen (2021) Sec. 4.26.

Convergence

- A sequence of random variables $Z_n \in \mathbb{R}$ **converges in probability** to c as $n \to \infty$, denoted by $Z_n \to_p c$ or $plim_{n\to\infty} Z_n = c$, if for all $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|Z_n - c| \leq \delta) = 1.$$

- Let Z_n be a sequence of random variables or vectors with distribution $G_n(u) = \mathbb{P}(Z_n \leq u)$. We say that Z_n **converges in distribution** to Z as $n \to \infty$, denoted with $Z_n \to_d Z$, if for all u at which $G(u) = \mathbb{P}(Z \leq u)$ is continous, $G_n(u) \to G(u)$ as $n \to \infty$.

Law of Large Numbers I

Theorem 18 (Weak Law of Large Numbers(WLLN))

If X_i are independent and identically distributed and $\mathbb{E}(X) < \infty$, then as $n \to \infty$,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to_p \mathbb{E}(X).$$

An estimator $\hat{\theta}$ of a parameter θ is **consistent** if $\hat{\theta} \to_p \theta$ as $n \to \infty$.
Counter examples:

- $X_i = Z + U_i$, Z is common component and $\mathbb{E}[U] = 0$, $\bar{X}_i \to_p Z$ but not the sample mean.
- Suppose $\text{var}(X_i) = 1$ for $i \leq n/2$ and $\text{var}(X_i) = n$ for $i > n/2$. $\text{var}(\bar{X}_n) \to 1/2$, $\bar{X}_n$ does not converge in probability.

Continuous Mapping Theorem

Theorem 19 (Continuous Mapping Theorem)

If $Z_n \rightarrow_p c$ as $n \rightarrow \infty$ and $h(\cdot)$ is continuous at c then $h(Z_n) \rightarrow_p h(c)$ as $n \rightarrow \infty$.

Central Limit Theorem I

Theorem 20 (Central Limit Theorem(CLT))

If X_i are i.i.d. and $\mathbb{E}(X^2) < \infty$ then as $n \to \infty$

$$\sqrt{n}(\bar{X}_n \to \mu) \to_d N(0, \sigma^2),$$

where $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathbb{E}[(X - \mu)^2]$.

Theorem 21 (Slutsky's Theorem)

If $Z_n \to_d Z$ and $c_n \to_p c$ as $n \to \infty$, then

1. $Z_n + c_n \to_d Z + c$
2. $Z_n c_n \to_d Z_c$
3. $\frac{Z_n}{c_n} \to_d \frac{Z}{c}$ *if* $c \neq 0$.

Central Limit Theorem II

Theorem 22 (Delta Method)

If $\sqrt{n}(\hat{\theta} - \theta) \to_d \xi$ *and* $h(u)$ *is continuously differentiable in a neighborhood of* θ *then as* $n \to \infty$

$$\sqrt{n}(h(\hat{\theta}) - h(\theta)) \to_d \mathbf{H}'\xi$$

where $\mathbf{H}(u) = \frac{\partial}{\partial u} h(u)^\top$ *and* $\mathbf{H} = \mathbf{H}(\theta)$.

References I

Hansen, B. (2021). Introduction to econometrics. madison: University of wisconsin.

Stock, J. H., Watson, M. W., et al. (2020). *Introduction to econometrics*, volume 4. Pearson New York.