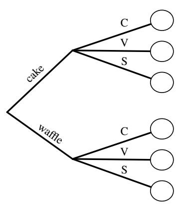
Introduction

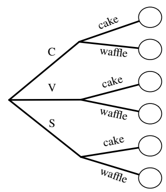
This is a cheatsheet compiled for the First-Year Econometrics sequence at Florida State University as taught by Dr. Zuehlke. While a lot of the material presented here was compiled and aggregated by myself (but not invented!), I could not have done this without the original template as graciously provided by William Chen. His website is `http://wzchen.com`, and he provides the original "Probability Cheatsheet" there. His version is far more expansive (and arguably better) than mine. Additional thanks are also due to Dr. Zuehlke whose lecture notes compose a large extent of this cheatsheet. The goal of this cheatsheet is to provide a quick refresher or desktop reference for the forgetful econometrician. Any errors are my own and please don't hesitate to contact me with suggestions or corrections at `oam18@my.fsu.edu` .

Probability Concepts

Counting

Addition & Multiplication Rules





Let's say we have a compound experiment (an experiment with multiple components). If the 1st component has n_1 possible outcomes, the 2nd component has n_2 possible outcomes, . . . , and the rth component has n_r possible outcomes. If only one component can be chosen, then we have $n_1 + n_2 + \cdots + n_r$ ways to select an outcome. If you can select one outcome per component, then overall there are $n_1 n_2 \dots n_r$ possibilities for the whole experiment.

Sampling Table

The sampling table gives the number of possible samples of size k out of a population of size n, under various assumptions about how the sample is collected.

	Order Matters	Not Matter
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Some Distributions for Discrete Draws

	Replace	No Replace
Fixed # trials (n)	Binomial (Bern if $n = 1$)	HGeom
Draw until r success	NBin (Geom if $r = 1$)	NHGeom

Naive Definition of Probability

If all outcomes are equally likely, the probability of an event A happening is:

$$P_{\text{naive}}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of outcomes}}$$

Thinking Conditionally

Independence

Independent Events A and B are independent if knowing whether A occurred gives no information about whether B occurred. More formally, A and B (which have nonzero probability) are independent if and only if one of the following equivalent statements holds:

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

Conditional Independence A and B are conditionally independent given C if $P(A \cap B|C) = P(A|C)P(B|C)$. Conditional independence does not imply independence, and independence does not imply conditional independence.

Unions, Intersections, and Complements

De Morgan's Laws A useful identity that can make calculating probabilities of unions easier by relating them to intersections, and vice versa. Analogous results hold with more than two sets.

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

Joint, Marginal, and Conditional

Joint Probability $P(A \cap B)$ or $P(A, B) \leftarrow$ Probability of A and B.

Marginal (Unconditional) Probability $P(A) \leftarrow$ Probability of A.

Conditional Probability $P(A|B) = P(A, B)/P(B) \leftarrow$ Probability of A, given that B occurred.

Conditional Probability *is* Probability $P(A|B)$ is a probability function for any fixed B. Any theorem that holds for probability also holds for conditional probability.

Probability of an Intersection or Union

Intersections via Conditioning

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

Unions via Inclusion-Exclusion

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Law of Total Probability (LOTP)

Let $B_1, B_2, B_3, \dots B_n$ be a *partition* of the sample space (i.e., they are disjoint and their union is the entire sample space).

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

Special case of LOTP with B and B^c as partition:

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

Bayes' Rule

If $A_1, A_2, \dots, A_n$ form a partition of the Sample Space, S, and B is any event, then:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j)P(B|A_j)}{P(B)} = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

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

We can also write

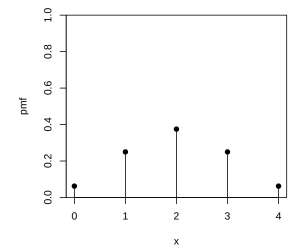
$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(B \cap C|A)P(A)}{P(B \cap C)}$$

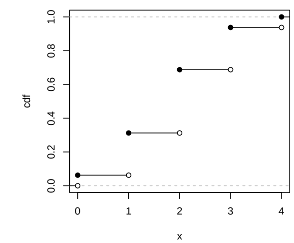
Random Variables and their Distributions

PMF, CDF, and Independence

The Probability Mass Function (pmf) Gives the probability that a *discrete* random variable takes on the value x.

$$f(x) \equiv P(X = x)$$





The PMF satisfies $f(x) \geq 0$ and $\sum_x f(x) = 1$.

The **Cumulative Distribution Function (cdf)** gives the probability that a random variable is less than or equal to x.

$$F_X(x) = P(X \leq x)$$

The CDF is an increasing, right-continuous function with

$$F_X(x) \to 0 \text{ as } x \to -\infty \text{ and } F_X(x) \to 1 \text{ as } x \to \infty$$

Independence Intuitively, two random variables are independent if knowing the value of one gives no information about the other. Discrete r.v.s X and Y are independent if for *all* values of x and y

$$P(X = x \cap Y = y) = P(X = x)P(Y = y)$$

Expected Value

Expected Value (a.k.a. *mean*, *expectation*, or *average*) is a weighted average of the possible outcomes of our random variable. Mathematically, if $x_1, x_2, x_3, \dots$ are all of the distinct possible values that X can take, the expected value of X is

$$E(X) = \sum_i x_i P(X = x_i) = \mu$$

Linearity For any r.v.s X and Y, and constants a, b, c,

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

Same distribution implies same mean If X and Y have the same distribution, then $E(X) = E(Y)$ and, more generally, $E(g(X)) = E(g(Y))$

Conditional Expected Value is defined like expectation, only conditioned on any event A.

$$E(X|A) = \sum_x x P(X = x|A)$$

Variance and Standard Deviation

$$\text{Var}(X) = E\left(X - E(X)\right)^2 = E(X^2) - (E(X))^2 = \sigma^2$$

$$\text{SD}(X) = \sqrt{\text{Var}(X)} = \sigma$$

Continuous RVs, LOTUS, UoU

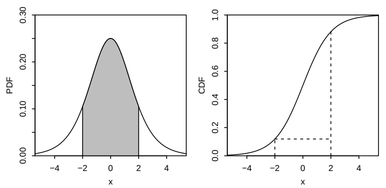
Continuous Random Variables (CRVs)

What is the Probability Density Function (PDF)? The PDF f is the derivative of the CDF F.

$$F'(x) = f(x)$$

A PDF is nonnegative and integrates to 1. By the fundamental theorem of calculus, to get from PDF back to CDF we can integrate:

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



To find the probability that a CRV takes on a value in an interval, integrate the PDF over that interval.

$$F(b) - F(a) = \int_a^b f(x)dx$$

How do I find the expected value of a CRV? Analogous to the discrete case, where you sum x times the PMF, for CRVs you integrate x times the PDF.

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

LOTUS

Expected value of a function of an r.v. The expected value of X is defined this way:

$$E(X) = \sum_x xP(X = x) \text{ (for discrete } X\text{)}$$

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

The **Law of the Unconscious Statistician (LOTUS)** states that you can find the expected value of a *function of a random variable*, $h(X)$, in a similar way, by replacing the x in front of the PMF/PDF by $h(x)$ but still working with the PMF/PDF of X:

$$E(h(X)) = \sum_x h(x)P(X = x) \text{ (for discrete } X\text{)}$$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx \text{ (for continuous } X\text{)}$$

Moments and MGFs

Moments

Let X have mean μ and standard deviation σ, then:
The k^{th} **moment** of a RV, X, is $E[X^k] = \sum\!\!\!\!\!\!\int x^k f(x)$.
The k^{th} **central moment** of a RV, X, is
$E[(X-\mu)^k] = \sum\!\!\!\!\!\!\int (x-\mu)^k f(x)$.
The mean and variance are important summaries of the shape of a distribution.

Mean $E(X) = \mu_1$

Variance $\text{Var}(X) = \mu_2 - \mu_1^2$

Moment Generating Functions

MGF For any random variable X, the function

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

is the **moment generating function (MGF)** of X, if it exists for all t in some open interval containing 0.

MGF of linear functions If we have $Y = aX + b$, then

$$M_Y(t) = E(e^{t(aX+b)}) = e^{bt}E(e^{(at)X}) = e^{bt}M_X(at)$$

Summing Independent RVs by Multiplying MGFs. If X and Y are independent, then

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX})E(e^{tY}) = M_X(t) \cdot M_Y(t)$$

The MGF of the sum of two random variables is the product of the MGFs of those two random variables.

Joint PDFs and CDFs

Joint Distributions

The **joint CDF** of X and Y is

$$F(x, y) = P(X \leq x, Y \leq y)$$

In the discrete case, X and Y have a **joint PMF**

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

In the continuous case, they have a **joint PDF**

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

The joint PMF/PDF must be nonnegative and sum/integrate to 1.

Conditional Distributions

Conditioning and Bayes' rule for discrete r.v.s

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

Conditioning and Bayes' rule for continuous r.v.s

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

Marginal Distributions

To find the distribution of one (or more) random variables from a joint PMF/PDF, sum/integrate over the unwanted random variables.

Marginal PMF from joint PMF

$$P(X = x) = \sum_y P(X = x, Y = y)$$

Marginal PDF from joint PDF

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

Limits of marginals and expected values: Note that the only time you are integrating over *all* the possible values of Y for the continuous case is when you are trying to find its expected value.
For example, given iid RV's X, Y, w/ pdf $f(x, y) = 3x,\ 0 < y < x < 1$
$\Rightarrow E[X] = \int_{-\infty}^{\infty} xf_X(x)dx = \int_{-\infty}^{\infty} x[\int_{-\infty}^{\infty} f(x, y)dy]dx =$
$\int_{-\infty}^{\infty} x[\int_0^x 3xdy]dx = \int_{-\infty}^{\infty} x[3x^2,\ 0 < x < 1]dx = \int_0^1 x3x^2dx = \frac{3}{4}$
However note that $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f(x, y)\ dy\ dx$, which using the previous joint pdf: $= \int_0^1\int_0^x xy\ 3x\ dy\ dx$

Independence of Random Variables

Random variables X and Y are independent if and only if any of the following conditions holds:

- Joint CDF is the product of the marginal CDFs
- Joint PMF/PDF is the product of the marginal PMFs/PDFs
- Conditional distribution of Y given X is the marginal distribution of Y
- ie no funny limits, and pdf/cdf/pmf's *can* be factored out

Multivariate LOTUS

LOTUS in more than one dimension is analogous to the 1D LOTUS.
For discrete random variables:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)P(X = x, Y = y)$$

For continuous random variables:

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

Covariance and Transformations

Covariance and Correlation

Covariance is the analog of variance for two random variables.

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

Note that

$$\text{Cov}(X, X) = E(X^2) - (E(X))^2 = \text{Var}(X)$$

Correlation is a standardized version of covariance that is always between -1 and 1.

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Covariance and Independence If two random variables are independent, then they are uncorrelated. The converse is not necessarily true (e.g., consider $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$).

$$X \perp\!\!\!\perp Y \longrightarrow \text{Cov}(X, Y) = 0 \longrightarrow E(XY) = E(X)E(Y)$$

Covariance and Variance The variance of a sum can be found by

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

If X and Y are independent then they have covariance 0, so

$$X \perp\!\!\!\perp Y \Longrightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Covariance Properties For random variables W, X, Y, Z and constants a, b:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$
$$\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$$
$$\text{Cov}(aX, bY) = a * b * \text{Cov}(X, Y)$$
$$\text{Cov}(W + X, Y + Z) = \text{Cov}(W, Y) + \text{Cov}(W, Z) + \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

Correlation is location-invariant and scale-invariant For any constants a, b, c, d with a and c nonzero,

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

Transformations

One Variable Transformations Let's say that we have a random variable X with PDF $f_X(x)$, but we are also interested in some function of X. We call this function $Y = g(X)$. Also let $y = g(x)$. If g is differentiable and strictly increasing (or strictly decreasing), then the PDF of Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

The derivative of the inverse transformation is called the **Jacobian**.

Useful Facts

Convolutions of Random Variables

A convolution of n random variables is simply their sum. For the following results, let X and Y be *independent*.

1. $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2) \longrightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
2. $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p) \longrightarrow X + Y \sim \text{Bin}(n_1 + n_2, p)$. $\text{Bin}(n, p)$ can be thought of as a sum of i.i.d. $\text{Bern}(p)$ r.v.s.
3. $X \sim \text{Gamma}(a_1, \lambda)$, $Y \sim \text{Gamma}(a_2, \lambda)$ $\longrightarrow X + Y \sim \text{Gamma}(a_1 + a_2, \lambda)$. $\text{Gamma}(n, \lambda)$ with n an integer can be thought of as a sum of i.i.d. $\text{Expo}(\lambda)$ r.v.s.

4. $X \sim \text{NBin}(r_1, p)$, $Y \sim \text{NBin}(r_2, p)$
   $\longrightarrow X + Y \sim \text{NBin}(r_1 + r_2, p)$. NBin$(r, p)$ can be thought of as a sum of i.i.d. Geom(p) r.v.s.
5. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$
   $\longrightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Special Cases of Distributions

1. $\text{Bin}(1, p) \sim \text{Bern}(p)$
2. $\text{Beta}(1, 1) \sim \text{Unif}(0, 1)$
3. $\text{Gamma}(1, \lambda) \sim \text{Expo}(\lambda)$
4. $\chi_n^2 \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$
5. $\text{NBin}(1, p) \sim \text{Geom}(p)$

Inequalities

1. **Cauchy-Schwarz** $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$
2. **Markov** $P(X \geq a) \leq \frac{E|X|}{a}$ for $a > 0$
3. **Chebyshev** $P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$ for $E(X) = \mu, \text{Var}(X) = \sigma^2$
4. **Jensen** $E(g(X)) \geq g(E(X))$ for g convex; reverse if g is concave
5. **Chebychev's Inequality**: gives rough bounds on certain probabilities.
   $E[X] = \mu, V(X) = \sigma^2 \Rightarrow \forall \varepsilon > 0,\ P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

Conditional Expectation

Conditioning on an Event We can find $E(Y|A)$, the expected value of Y given that event A occurred. A very important case is when A is the event $X = x$. Note that $E(Y|A)$ is a *number*. For example:

- The expected value of a fair die roll, given that it is prime, is $\frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 5 = \frac{10}{3}$.
- Let Y be the number of successes in 10 independent Bernoulli trials with probability p of success. Let A be the event that the first 3 trials are all successes. Then
  $$E(Y|A) = 3 + 7p$$
  since the number of successes among the last 7 trials is Bin$(7, p)$.
- Let $T \sim \text{Expo}(1/10)$ be how long you have to wait until the shuttle comes. Given that you have already waited t minutes, the expected additional waiting time is 10 more minutes, by the memoryless property. That is, $E(T|T > t) = t + 10$.

Discrete Y	Continuous Y
$E(Y) = \sum_y yP(Y = y)$	$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$
$E(Y\|A) = \sum_y yP(Y = y\|A)$	$E(Y\|A) = \int_{-\infty}^{\infty} y f(y\|A) dy$

Conditioning on a Random Variable We can also find $E(Y|X)$, the expected value of Y given the random variable X. This is *a function of the random variable* X. It is *not* a number except in certain special cases such as if $X \perp\!\!\!\perp Y$. To find $E(Y|X)$, find $E(Y|X = x)$ and then plug in X for x. For example:

- If $E(Y|X = x) = x^3 + 5x$, then $E(Y|X) = X^3 + 5X$.
- Let Y be the number of successes in 10 independent Bernoulli trials with probability p of success and X be the number of successes among the first 3 trials. Then $E(Y|X) = X + 7p$.
- Let $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$. Then $E(Y|X = x) = x^2$ since if we know $X = x$ then we know $Y = x^2$. And $E(X|Y = y) = 0$ since if we know $Y = y$ then we know $X = \pm\sqrt{y}$, with equal probabilities (by symmetry). So $E(Y|X) = X^2, E(X|Y) = 0$.

Properties of Conditional Expectation

1. $E(Y|X) = E(Y)$ if $X \perp\!\!\!\perp Y$
2. $E(h(X)W|X) = h(X)E(W|X)$ (**taking out what's known**) In particular, $E(h(X)|X) = h(X)$.
3. $E(E(Y|X)) = E(Y)$ (**Law of Iterated Expectation (LIE)**, a.k.a. Law of Total Expectation, a.k.a. Adam's Law)

Law of Iterated Expectation (LIE) can also be written in a way that looks analogous to **LOTP**. For any events $A_1, A_2, \ldots, A_n$ that partition the sample space,
$$E(Y) = E(Y|A_1)P(A_1) + \cdots + E(Y|A_n)P(A_n)$$
For the special case where the partition is A, A^c, this says
$$E(Y) = E(Y|A)P(A) + E(Y|A^c)P(A^c)$$

Eve's Law (a.k.a. Law of Total Variance)
$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

Method of Moments (MoM) Estimators

MoM for

- $\mu = E[X_i] = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
- $E[X_i^2] \frac{\sum_{i=1}^n X_i}{n}$
- $Var(X_i) = E[X_i^2] - (E[X_i])^2$

χ^2 (Chi-Square) Distribution

Let us say that X is distributed χ_n^2. We know the following:

Story A Chi-Square(n) is the sum of the squares of n independent standard Normal r.v.s.

Properties and Representations

$$Z_1 + Z_2 + \cdots + Z_k \text{ for i.i.d. } Z_i \sim \mathcal{N}(0, 1) \Rightarrow Y = \sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

t-distribution

suppose $Z \sim N(0, 1), Y \sim \chi^2(k), Z \perp\!\!\!\perp Y \Rightarrow T \equiv \frac{Z}{\sqrt{\frac{Y}{k}}}$

F-Distribution

- Suppose $X \sim \chi^2(n),\ Y \sim \chi^2(m),\ X \perp\!\!\!\perp Y \Rightarrow F \equiv \frac{X/n}{Y/m} \sim F(n, m)$. That is, the F-distribution is the ratio of two Chi-Squared Distributed variables divided by their respective degrees of freedom.
- $E[F] = \frac{m}{m-2},\ m > 2,\ Var(F) = \frac{(2n+m-2)\sqrt{8(m-4)}}{(m-6)\sqrt{n(n+m-2)}}$
- $F \sim F(n, m) \Rightarrow$ we denote the $(1 - \alpha)$ quantile by$F_{\alpha,n,m}$, *ie* $P(F > F_{\alpha,n,m}) = \alpha$
- $F_{1-\alpha,m,n} = \frac{1}{F_{\alpha,n,m}}$
- Used for ratio of variances

LLN & CLT

Law of Large Numbers (LLN)

Let $X_1, X_2, X_3 \ldots$ be i.i.d. with mean μ. The **sample mean** is
$$\bar{X}_n = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$
The **Law of Large Numbers** states that as $n \to \infty$, $\bar{X}_n \to \mu$ with probability 1. For example, in flips of a coin with probability p of Heads, let X_j be the indicator of the jth flip being Heads. Then LLN says the proportion of Heads converges to p (with probability 1).

Central Limit Theorem (CLT)

Asymptotic Distributions using CLT

We use $\xrightarrow{D}$ to denote *converges in distribution to* as $n \to \infty$. The CLT says that if we standardize the sum $X_1 + \cdots + X_n$ then the distribution of the sum converges to $\mathcal{N}(0, 1)$ as $n \to \infty$:
$$\frac{1}{\sigma\sqrt{n}}(X_1 + \cdots + X_n - n\mu_X) \xrightarrow{D} \mathcal{N}(0, 1)$$
In other words, the CDF of the left-hand side goes to the standard Normal CDF, Φ. In terms of the sample mean, the CLT says
$$\frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\sigma_X} \xrightarrow{D} \mathcal{N}(0, 1)$$

Linear Algebra Review

Quadratic Forms

A matrix $\underset{q\times q}{A}$ is:

- **Positive Definite (p.d.)** if for any conformable vector $\underset{q\times 1}{X}$, $X'AX > 0$ for $X \neq 0$.
- **Positive Semi-Definite (p.s.d.)** if for any conformable vector $\underset{q\times 1}{X}$, $X'AX \geq 0$ for $X \neq 0$.
- **Negative Definite (n.d.)** if for any conformable vector $\underset{q\times 1}{X}$, $X'AX < 0$ for $X \neq 0$.
- **Negative Semi-Definite (n.s.d.)** if for any conformable vector $\underset{q\times 1}{X}$, $X'AX \leq 0$ for $X \neq 0$.
- **Indefinite** if $X'AX < 0$ for some X and $X'AX > 0$ for some other X.
- $\underset{(1\times q)}{X'}\underset{(q\times q)}{A}\underset{(q\times 1)}{X} = \sum_{i=1}^q \sum_{j=1}^q a_{ij}x_ix_j = a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + \ldots + a_{1q}x_1x_q + a_{21}x_1x_2 + a_{22}x_2^2 + \ldots + a_{qq}x_q^2$
- Any matrix of the from $X'X$ is **positive semi-definite (p.s.d.)**.
- If $\underset{n\times k}{X}$ where $\rho(X) = k$, that is, X has *full column rank*, then $X'X$ is p.d.
- If $\underset{n\times k}{X}$, $\rho(X) = k$, $\underset{n\times n}{D}$, $d_{ij} > 0 \forall i$, $\Rightarrow X'DX$ is p.d. for diagonal matrix D.
- If $\underset{n\times k}{X}$, $\rho(X) = k$, $\underset{n\times n}{D}$, $d_{ij} > 0 \forall i$, $\Rightarrow -X'DX$ is n.d. for diagonal matrix D.

Derivatives

Let $f(X)$ denote a *scalar*-valued function where $\underset{n\times 1}{X} \Rightarrow f : \mathbb{R}^n \to \mathbb{R}^1$.

Then we will define $\frac{\partial \underset{1\times 1}{f(X)}}{\partial \underset{n\times 1}{X}} = \begin{bmatrix} \frac{\partial f(X)}{\partial X_1} \\ \frac{\partial f(X)}{\partial X_1} \\ \vdots \\ \frac{\partial f(X)}{\partial X_n} \end{bmatrix} = \nabla f(X)$.

Likewise $\frac{\partial f(X)}{\partial X'} = \begin{bmatrix} \frac{\partial f(X)}{\partial X_1} & \frac{\partial f(X)}{\partial X_1} & \cdots & \frac{\partial f(X)}{\partial X_n} \end{bmatrix}$.

Table of Distributions

	Distribution	PMF/PDF and Support	CDF	Expected Value	Variance	MGF
Discrete	Bernoulli $\text{Bern}(p)$	$P(X=1)=p$ $P(X=0)=q=1-p$	$F(x)=\begin{cases}0 & \text{if } x<0\\ q & \text{if } 0\leq x<1\\ 1 & \text{if } x\geq 1\end{cases}$	p	pq	$q+pe^t$
	Binomial $\text{Bin}(n,p)$	$P(X=k)=\binom{n}{k}p^k q^{n-k}$ $k\in\{0,1,2,\ldots n\}$	$F(x)=\sum\limits_{i=0}^{\lfloor k\rfloor}\binom{n}{i}p^i q^{n-i}$	np	npq	$(q+pe^t)^n$
	Geometric $\text{Geom}(p)$	$P(X=k)=q^k p$ $k\in\{0,\,1,\,2,\,\ldots\}$	$F(x)=1-(1-p)^{k+1}$	q/p	q/p^2	$\frac{p}{1-qe^t},\ qe^t<1$
	Negative Binomial $\text{NBin}(r,p)$	$P(X=n)=\binom{r+n-1}{r-1}p^r q^n$ $n\in\{0,\,1,\,2,\,\ldots\}$	Messy	rq/p	rq/p^2	$(\frac{p}{1-qe^t})^r,\ qe^t<1$
	Hypergeometric $\text{HGeom}(w,b,n)$	$P(X=k)=\binom{w}{k}\binom{b}{n-k}/\binom{w+b}{n}$ $k\in\{0,1,2,\ldots,n\}$	Messy	$\mu=\frac{nw}{b+w}$	$\left(\frac{w+b-n}{w+b-1}\right)n\frac{\mu}{n}(1-\frac{\mu}{n})$	Messy
	Poisson $\text{Pois}(\lambda)$	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$ $k\in\{0,\,1,\,2,\,\ldots\}$	Messy	λ	λ	$e^{\lambda(e^t-1)}$
Continuous	Uniform $\text{Unif}(a,b)$	$f(x)=\frac{1}{b-a}$ $x\in(a,b)$	$F(x)=\begin{cases}0 & \text{if } x<a\\ \frac{x-a}{b-a} & \text{if } a\leq x<b\\ 1 & \text{if } x\geq b\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
	Normal $\mathcal{N}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x\,-\,\mu)^2/(2\sigma^2)}$ $x\in(-\infty,\infty)$	Messy	μ	σ^2	$e^{t\mu+\frac{\sigma^2t^2}{2}}$
	Standard Normal $\Phi(0,1)$	$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-(x)^2/(2)}$ $x\in(-\infty,\infty)$	Messy	0	1	$e^{\frac{t^2}{2}}$
	Log-Normal $\mathcal{LN}(\mu,\sigma^2)$	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-\,(\log x\,-\,\mu)^2/(2\sigma^2)}$ $x\in(0,\infty)$	Messy	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2+2\mu})(e^{\sigma^2}-1)$	DNE
	Exponential $\text{Expo}(\lambda)$	$f(x)=\lambda e^{-\lambda x}$ $x\in(0,\infty)$	$F(x)=1-e^{\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t},\ t<\lambda$
	Erlang $\text{Erlang}_k(\lambda)$	$f(x)=\frac{\lambda^k e^{-\lambda x}x^{k-1}}{(k-1)!},\ x\geq 0$ $k=\{1,2,\ldots\},\ \lambda\in(0,\infty)$	$F(x)=1-\sum_{i=0}^{k-1}\frac{e^{-\lambda x}(\lambda x)^i}{i!}$	$\frac{k}{\lambda}$	$\frac{k}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^k$
	Gamma $\Gamma(\alpha,\beta)$	$f(x)=\left[\frac{1}{\Gamma(\alpha)\beta^\alpha}\right]x^{\alpha-1}e^{-\frac{x}{\beta}}$ $x\in(0,\infty)$	Messy	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha},\ t<\beta$
	Beta $\mathcal{B}(\alpha,\beta)$	$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ $x\in(0,1)$	Messy	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	Messy
	Chi-Square χ_n^2	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x\in(0,\infty)$	Messy	n	$2n$	$(1-2t)^{-n/2},\ t<1/2$
	Student-t t_n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$ $x\in(-\infty,\infty)$	Messy	0 if $n>1$	$\frac{n}{n-2}$ if $n>2$	DNE
	F $\mathcal{F}(d_1,d_2)$	$f\,(x;d_1,d_2)=\left(\sqrt{\frac{(d_1x)^{d_1}d_2^{d_2}}{(d_1x+d_2)^{d_1+d_2}}}\right)/x\mathcal{B}\left(\frac{d_1}{2},\frac{d_2}{2}\right)$ $x\in(0,+\infty)$ if $d_1=1$, o/w $x\in[0,+\infty)$	Messy	$\frac{d_2}{d_2-2}$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$	DNE
	Logistic sec^2	$\psi(x)=\frac{e^x}{1+e^x}\left[1-\frac{e^x}{1+e^x}\right]$ $x\in(-\infty,\infty)$	$\Psi(x)=\frac{e^x}{1+e^x}$	0	$\frac{\pi^2}{3}$	$e^t*\mathcal{B}(1-t,1+t)$ for $t\in(-1,1)$