Econometrics Summary

Mitch Downey

January 22, 2024

Contents

Part I
Math References

1 Common Distributions

Distribution	pdf/pmf	CDF	Mean	Variance	$M_x(t)$	$\Phi_x(t)$
Uniform: $U(\theta_1, \theta_2)$	$\frac{1}{\theta_2-\theta_1}\ \forall x \in [\theta_1, \theta_2]$	$\frac{x-\theta_1}{\theta_2-\theta_1}$	$\frac{\theta_1+\theta_2}{2}$	$\frac{\theta_2-\theta_1)^2}{12}$	$\frac{e^{t\theta_2}-e^{t\theta_1}}{t(\theta_2-\theta_1)}$	–
Normal: $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	No closed form: $\Phi(\frac{x-\mu}{\sigma})$	μ	σ^2	$\exp(\mu t + \frac{t^2\sigma^2}{2})$	–
Exponential: $\exp(\lambda)$	$\frac{1}{\lambda}e^{-x/\lambda}\ \forall x > 0$	$1 - e^{-x/\lambda}$	λ	λ^2	$\frac{1}{1-\lambda t}$	–
Gamma: $\Gamma(\alpha, \beta)$	$\big(\Gamma(\alpha)\beta^\alpha\big)^{-1}x^{\alpha-1}e^{-x/\beta}\ \forall x > 0$	$\frac{\gamma(\alpha, \frac{x}{\beta})}{\Gamma(\alpha)}$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$	–
Chi-squared: χ^2_v	$\frac{x^{v/2-1}e^{-x/2}}{2^{v/2}\Gamma(\frac{v}{2})}\ \forall x > 0$	$\frac{\gamma(\frac{v}{2}, \frac{x}{2})}{\Gamma(\frac{v}{2})}$	v	$2v$	$(1-2t)^{-v/2}$	–
Beta: $\beta(\alpha, \beta)$	$\big[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\big]x^{\alpha-1}(1-x)^{\beta-1}\ \forall x \in (0,1)$	–	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	No closed form	–
Binomial: $bin(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}\ \forall x \geq 0$	$\sum_{i=0}^{x}\binom{n}{i}p^i(1-p)^{n-i}$	np	$np(1-p)$	$\big[pe^t + (1-p)\big]^n$	–
Geometric	$p(1-p)^{x-1}\ \forall x \geq 0$	$1-(1-p)^x$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$	–
Poisson: $Pois(\lambda)$	$\frac{\lambda^x e^{-\lambda}}{x!}\ \forall x \geq 0$	$\sum_{i=0}^{x}\frac{\lambda^i}{i!}$	λ	λ	$\exp\big(\lambda(e^t-1)\big)$	–
Negative Binomial: $nb(r,p)$	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$	–	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$	–
Cauchy; location: x_0, scale: γ	$\frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right]$	$\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}$	–	–	–	$\exp(x_0 it - \gamma\lvert t\rvert)$

Part II
220A

2 Measure Theory

Definition 1. A σ**-field**, called $\mathcal{F}$ is a collection of subsets of Ω that satisfies

1. $\emptyset \in \mathcal{F}$

2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

3. $A_1, A_2, ... \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Definition 2. A π**-system** is a collection of subsets of Ω that is closed under finite intersections. Note, all σ-fields are π-systems but not all π-systems are σ-fields because 1) they might not be closed under complements and 2) they might not be closed under countably infinite intersections.

Theorem 1. (Special Case of More General π-λ **Theorem**). Suppose that 1) P_1 and P_2 are probability measures on $(\Omega, \mathcal{F})$ and 2) $\mathcal{P}$ is a π-system such that $\sigma(\mathcal{P}) = \mathcal{F}$. Then $P_1 A = P_2 A \ \ \forall A \in \mathcal{P} \Rightarrow P_1 A = P_2 A \ \ \forall A \in \mathcal{F}$.

Theorem 2. (**Bayes' Theorem**).

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

Definition 3. A sequence of σ-fields, $\mathcal{A}_1, \mathcal{A}_2, ... \subset \mathcal{F}$ are **mutually independent** if any sequence of of sets $A_n \in \mathcal{A}_n, n \in \mathbb{N}$ is mutually independent (meaning any finite collection of those sets is independent: $P(A \cap B) = P(A)P(B)$).

Definition 4. A function $X : \Omega \to \mathbb{R}^k$ is $\mathcal{F}$**-measurable** if $\forall \ B \in \mathcal{B}(\mathbb{R}^k)$, $X^{-1}B \in \mathcal{F}$. A **random variable** is an $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$. A **random vector** is an $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}^k$. Note that this implies your σ-field must be at least as "precise" as your random variable. Also note that $\emptyset$ is always an element of any σ-field, so it's possible to take some $B \in \mathcal{B}(\mathbb{R}^k)$ that nothing maps to, and that's ok.

Definition 5. Given a random vector X, the σ**-field generated by** X, denoted $\sigma(X)$, is the collection of sets $\{X^{-1}B : B \in \mathcal{B}(\mathbb{R}^k)\}$. That is, $\sigma(X)$ is the collection of inverse images of Borell sets. $\sigma(X)$ is a σ-field. Note that if X is $\mathcal{F}$-measurable, then $\sigma(X) \in \mathcal{F}$. It is a refined version of $\mathcal{F}$, where the elements that aren't useful (because they don't map to $\mathbb{R}^k$) are dropped.

Theorem 3. Let $\mathcal{C} \subset \mathcal{B}(\mathbb{R}^k)$ be a collection of sets such that $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^k)$. A function $X : \Omega \to \mathbb{R}^k$ is a random vector ($\mathcal{F}$-measurable) iff $X^{-1}C \in \mathcal{F} \ \forall \ C \in \mathcal{C}$. In other words, if you have a set that can generate $\mathcal{B}(\mathbb{R}^k)$, and for any element C of that set, $X^{-1}C \in \mathcal{F}$, then X is $\mathcal{F}$-measurable. That is, you don't have to check for any element in $\mathcal{B}(\mathbb{R}^k)$, but only elements of some set than can generate $\mathcal{B}(\mathbb{R}^k)$. This has very much the same application and feel of the π-λ theorem.

Example 1. Suppose we flip a coin twice and record the outcome. We may define $\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$ and let $\mathcal{F}$ be the set of all subsets of Ω including $\emptyset$ and Ω. Let $X : \Omega \to \mathbb{R}$ be the number of heads tossed. That is,

$$X(H,H) = 2, \quad X(H,T) = 1, \quad X(T,H) = 1, \quad X(T,T) = 0.$$

The σ-field $\sigma(X)$ is given by

$$\sigma(X) = \Big\{\emptyset, \Omega, \{(H,H)\}, \{(H,T),(T,H)\}, \{(T,T)\}, \{(H,H),(T,T)\}$$
$$\{(H,H),(H,T),(T,H)\}, \{(T,H),(H,T),(T,T)\}\Big\}$$

This σ-field represents the information learned about Ω by observing X. Note that it does not allow us to distinguish between (H,T) and (T,H) since they both correspond to $X=1$.

Definition 6. A random vector $X:\Omega \to \mathbb{R}^k$ **induces a probability measure** P_X on $\big(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)\big)$ according to

$$P_X A = PX^{-1}A = P\{\omega \in \Omega : X(\omega) \in A\}, A \in \mathcal{B}(\mathbb{R}^k)$$

Note that $\mathcal{F}$-measurability of X is essential for P_X to be properly defined. Intuitively, $\mathcal{P}$ maps elements of $\mathcal{F}$ to $[0,1]$. Since all elements of $\mathcal{F}$ are collections of elements of Ω, $\mathcal{P}$ might as well map Ω subsets. Given some X, any of those Ω subsets (the ones in $\mathcal{F}$) can be rewritten as the pre-image of sets in $\mathbb{R}^k$ (since X is $\mathcal{F}$-measurable). So for any $A \in \mathcal{B}(\mathbb{R}^k)$, you can think about the pre-image of that A ($X^{-1}A$). Then P_X simply maps A to the same place that $\mathcal{P}$ mapped $X^{-1}A$ to. So it's basically a shortcut, where we map sets of $\mathcal{B}(\mathbb{R}^k)$ directly into $[0,1]$ instead of going to their pre-image (an element of $\mathcal{F}$) and mapping it into $[0,1]$ using $\mathcal{P}$.

Example 2. Recall the random variable X from the previous example. P_X can be written as

$$P_X A = \frac{1}{4}1_{\{0\in A\}} + \frac{1}{2}1_{\{1\in A\}} + \frac{1}{4}1_{\{2\in A\}}, A \in \mathcal{B}(\mathbb{R}^k)$$

Theorem 4. The distribution function F_X satisfies the following properties:

1. F_X is right-continuous. That is, $\lim_{\delta\downarrow 0} F_X(x+\delta) = F_X(x),\ \forall x \in \mathbb{R}$.
2. F_X is nondecreasing.
3. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$

Theorem 5. If X and Y are random vectors and $F_X(x) = F_Y(x)\ \forall x \in \mathbb{R}^k$ then $P_X B = P_Y B\ \forall B \in \mathcal{B}(\mathbb{R}^k)$. In other words, a cdf uniquely defines a random variable. Two variables with the same cdf's will have the same induced probability measures (and are thus identical).

Definition 7. A sequence of random vectors $X_1, X_2, ...$ are **mutually independent random vectors** if $\sigma(X_1), \sigma(X_2), ...$ are mutually independent.

Theorem 6. Suppose $X:\Omega\to\mathbb{R}^k$ and $Y:\Omega\to\mathbb{R}^l$ are random vectors. Define $Z:\Omega\to\mathbb{R}^{k+l}$ by $Z(\omega) = \big(X(\omega), Y(\omega)\big)$. Then X and Y are independent iff $F_Z(x,y) = F_X(x)F_Y(y)$. This is known as the factorization of distribution functions.

2.1 Lebesgue Integration

Definition 8. Let $(M, \mathcal{M})$ be a measurable space. A function $\phi: M \to \mathbb{R}$ is a **simple function** if there exists $A_1, ..., A_n \in \mathcal{M}$ and $c_1, ..., c_n \in \mathbb{R}$ such that

$$\phi(x) = \sum_{i=1}^{n} c_i 1_{\{x\in A_i\}} \quad \forall x \in M$$

Think of the c_i's as weights. Thus, suppose you can create a finite set of subsets of $\mathcal{M}$ that corresponds to elements of M (a σ-field made up of elements of $\mathcal{M}$). These A_i's don't have to be disjoint. Suppose to each A_i you assign some "weight" which is just a value from $\mathbb{R}$. Choose any $x \in M$, take all the A_i's that x belongs to (could be one or multiple), and add up the c_i's for those A_i's. If that (finite) sum is the value of ϕ evaluated at x, then ϕ is a simple function.

Definition 9. Let $(M, \mathcal{M}, \mu)$ be a measure space. The **Lebesgue Integral** of a nonnegative simple function ϕ with respect to μ is

$$\int \phi d\mu = \int \phi(x) d\mu(x) = \sum_{i=1}^{n} c_i \mu(A_i)$$

This is a generalization of the Rieman Integral. In cases when the Rieman integral is well-defined, the two are idential. Imagine Rieman sums. The A_i's are the width of the rectangles. The c_i's are the height of the rectangles. This generalizes that. You don't have to deal with rectangles, you can deal with any arbitrary sets (A_i) of the domain. They don't even have to be evenly sized.

The following steps generalize the Lesbesgue Integral to all $\mathcal{M}$-measurable functions.

1. If f is nonnegative but not simple, define $\int f d\mu = sup\{\int \phi d\mu : 0 \leq \phi \leq f, \phi \text{ simple}\}$.
2. If f is not nonnegative, define f^+ as the positive portions and f^- as the negative portions. Then $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$.

Theorem 7. (**Monotone Convergence Theorem**). Suppose $f_1, f_2, ...$ is a sequence of non-negative real valued functions on M, with $f_n \leq f_{n+1}$ for all n and $\lim_{n\to\infty} f_n(x) = f(x)$. Let f_n and f be $\mathcal{M}$-measurable. Then $\lim_{n\to\infty} \int f_n(x) d\mu = \int f(x) d\mu$.

Theorem 8. (**Dominated Convergence Theorem**). Suppose $f_1, f_2, ...$ is a sequence of nonnegative real valued functions on M, with $\lim_{n\to\infty} f_n(x) = f(x)$. Let f_n and f be $\mathcal{M}$-measurable. If there exists an integrable nonnegative function $g : M \to \mathbb{R}$ such that $\mid f_n \mid \leq g$ for all n, then $\lim_{n\to\infty} \int f_n(x) d\mu = \int f(x) d\mu$.

3 Probability

3.1 Densities

Definition 10. Let μ and ν be two measures on $(M, \mathcal{M})$. We say ν is **absolutely continuous** with respect to μ, written $\nu << \mu$, if $\mu(A) = 0 \Rightarrow \nu(A) = 0 \ \forall A \in \mathcal{M}$. Note that this does not say, $\nu(A) \leq \mu(A)$. There are two ways to show ν is <u>not</u> absolutely continuous with respect to μ:

1. $\exists A s.t. \mu(A) = 0 \wedge \nu(A) \neq 0$
2. (If ν is a probability measure), the CDF of ν is not continuous

Theorem 9. (**Radon-Nikodym Theorem**). Let μ and ν be measures on $(M, \mathcal{M})$. If $\nu << \mu$, then there exists a $\mathcal{M}$-measurable function $f : M \to \mathbb{R}$ such that

$$\nu(A) = \int_A f d\mu, \ \forall A \in \mathcal{M}$$

In this case, ν has **density** f with respect to μ. The function f is the **Radon-Nikodym derivative**. f is not necessarily unique, but any f, f^* have to be equal a.e. (μ-a.e.). Intuitively, think of ν as a probability measure induced by some continous random variable and μ as the Lebesgue measure. Note that $\nu << \mu$ implies that sets with no Lebesgue measure (e.g., singletons and collections of singletons) have no probability under ν. This is consistent with the intuitive definition of a continous random variable, where the probability of any single value is zero. Then this establishes that all continous random variables have a density, and defines that "density" simply as some function that integrates to the CDF.

Note that if $\nu = \int f d\mu$, then whenever $\mu(A) = 0$ then $\nu(A) = \int_A f d\mu = 0$ obviously. So $\nu = \int f d\mu \Rightarrow \nu << \mu$.

All this says, essentially, that P_X has a pdf (continous) iff it is absolutely continuous with respect to the Lebesgue measure. Equivalently, P_X has a pmf (discrete) iff it is absolutely continous with respect to the counting measure.

Example 3. Consider $X \sim N(0,1),\ Y \equiv max\{0, X\}$. Like all random variables, Y has a CDF. But half of its mass is at a single point ($Y = 0$), so it can't have a pdf (not continous wrt Lebesgue measure), so it's not continous. But hte other half of its mass is spread out over an uncountable infinite set ($Y > 0$), so it can't have a pmf (not continous wrt counting measure), so it's not discrete. It's neither, but is still a random variable.

Example 4. Consider $U \sim U(0,1),\ X \equiv cos(2U),\ Y \equiv sin(2U),\ Z \equiv (X,Y)$. Then Z is just the line that makes a circle (radius 1), but not the area of the circle. So the Lebesgue measure of Z is 0. Thus, P_Z is not absolutely continous with respect to the Lebesgue measure, so Z doesn't have a pdf. X and Y have marginal pdf's, but no joint pdf.

Definition 11. A distribution function F_X is called absolutely continous if P_X is absolutely continous with respect to the Lebesgue measure. Note: Absolute continuity of F_X implies continuity of F_X. (Generally, continuity of F_X also implies absolute continuity of P_X, but there are some esoteric counter-examples.)

3.2 Inequalities

Theorem 10. (**Markov's Inequality**). For any r.v. X, $\varepsilon > 0$, and $p \geq 0$,

$$P(|X| \geq \varepsilon) \leq \frac{E\big(|X|^p\big)}{\varepsilon^p}$$

When $p = 2$, this is called **Chebyshev's Inequality**.

Theorem 11. (**Holder's Inequality**). For any r.v.'s X and Y and $p, q > 1$ where $\frac{1}{p} + \frac{1}{q} = 1$,

$$E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$$

When $p = q = 2$, this is called the **Cauchy-Schwarz Inequality**. $(E|X|^p)^{1/p}$ is called the L-p norm and can be written $||X||_p$.

Theorem 12. (**Liapounov's Inequality**). For any r.v. X and $1 \leq r \leq s < \infty$,

$$(E|X|^r)^{1/r} \leq (E|X|^s)^{1/s}$$

Thus, L-p norms are nondecreasing in p.

Theorem 13. (**Minkowski's Inequality**). For any r.v.'s X and Y and $1 \geq p < \infty$,

$$(E|X+Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$$

This establishes the triangle inequality for L-p norms.

Theorem 14. (**Jensen's Inequality**). For any r.v. X and convex function $g : \mathbb{R} \to \mathbb{R}$,

$$g(EX) \leq Eg(X)$$

For any concave function $g : \mathbb{R} \to \mathbb{R}$,

$$g(EX) \geq Eg(X)$$

3.3 Multiple Variables

Definition 12. Let X be a random vector on $(\Omega, \mathcal{F}, P)$. Given $B \in \mathcal{F}$, the **conditional probability** of B given X, denoted $P(B|X)$ is the $\sigma(X)$-measurable random variable

$$\int_A P(B|X)(\omega)dP(\omega) = P(A \cap B), \quad \forall\ A \in \sigma(X)$$

When we say that $P(B|X)$ is $\sigma(X)$-measurable, it means that if I tell you some $A \in \sigma(X)$, then $P(B|X)$ is just a number. But if I don't give you that A, then $P(B|X)$ draws its randomness from which A we will get, so it's a random variable itself. When you integrate over all possible values of A, you get the joint probability.

Theorem 15. $P(B|X)$ always exists and is unique up to P-a.e. equivalence. So if $P(B|X) = P(B|Y)\ \forall\ B$ then $X = Y$ (on a set of probability 1).

Definition 13. Let X and Y be random vectors on $(\Omega, \mathcal{F}, P)$ with $E|Y| < \infty$. The **conditional expectation** of Y given X, denoted $E(Y|X)$ is the $\sigma(X)$-measurable random variable

$$\int_A E(Y|X)(\omega)dP(\omega) = \int_A Y(\omega)dP(\omega)$$

Theorem 16. $E(Y|X)$ always exists and is unique up to P-a.e. equivalence.

Definition 14. Let (X, Y) be a continuous pair of r.v.'s with joint pdf $f : \mathbb{R}^2 \to \mathbb{R}$. The **conditional pdf** of Y is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{\int f(x,z)dz}$$

The denominator $\int f(x,z)dz$ is called the **marginal pdf** of X. For any value of x, it will tell you the density of X at that value, accounting for all the various possible values of Y (each with their respective probabilities). The intuition behind the definition is that $P(Y|X) = \frac{P(X \cap Y)}{P(X)}$ (which is <u>wrong</u> technical notation). So the probability of Y, given X, is the probability of X and Y, <u>divided</u> by (inverse weighted by) the probability of X, since we already know that X happened.

Theorem 17.

$$E(Y|X) = \int y f_{Y|X}(y|x)dy$$

Definition 15. Let Y be a random vector on $(\Omega, \mathcal{F}, P)$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-field. The **conditional expectation** of Y given $\mathcal{A}$, denoted $E(Y|\mathcal{A})$ is the $\mathcal{A}$-measurable random variable

$$\int_A E(Y|\mathcal{A})(\omega)dP(\omega) = \int_A Y(\omega)dP(\omega), \quad \forall\ A \in \mathcal{A}$$

Theorem 18. If Y is an $\mathcal{A}$-measurable random variable with $E|Y| < \infty$ then $E(Y|\mathcal{A}) = Y$ a.s.

Theorem 19. (**Law of Iterated Expectations**). Let Y be a random vector on $(\Omega, \mathcal{F}, P)$ with $E|Y| < \infty$. Suppose that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ are all σ-fields. Then

$$E\big(E(Y|\mathcal{F}_2)|\mathcal{F}_1\big) = E\big(E(Y|\mathcal{F}_1)|\mathcal{F}_2\big) = E(Y|\mathcal{F}_1)$$

In other words, the conditional expectation is always that based on the least informative σ-field (the smallest σ-field). This is intuitive. If you take the conditional expectation of $E(Y|\mathcal{F}_2)$ based on $\mathcal{F}_1$, then you might as well not have had $\mathcal{F}_2$ at all.

Note if $\mathcal{F}_1$ is the trivial sigma field $(\{\emptyset, \mathcal{F}\})$ then $E\left(E(Y|\mathcal{F}_2)\right) = E(Y)$.

Theorem 20. The following are sufficient to determine the independence of two random variables (probably the easiest way to check independnece):

1. $F_{XY} = F_X F_Y$
2. $f_{XY} = f_X f_Y$
3. $F_{XY} = g(X)h(Y)$, where g is not a function of Y and h is not a function of X
4. $f_{XY} = g(X)h(Y)$, where g and h are as above

Theorem 21. If X and Y are independent random variables, then $Cov(X, Y) = 0$. The converse is not true. In general, $Cov(X, Y) = 0 \nRightarrow X, Y$ are independent. However, if X and Y are both multivariate normally distributed, then the implication does hold ($Cov(X, Y) = 0 \Rightarrow X, Y$).

3.4 Moment Generating and Characteristics Functions

Definition 16. The **moment generating function** (mgf) is the function $M_X : \mathbb{R} \to [0, \infty]$

$$M_X(t) = E(e^{tX}) = \int e^{tX} dP = \int e^{tx} dP_X = \int e^{tx} f_X(x) dx$$

Theorem 22. The following are properties of the mgf:

1. $\forall a, b \in \mathbb{R}, \quad M_{aX+b}(t) = e^{bt} M_X(at)$
2. If X and Y are independent, then $M_{X+Y}(t) = M_X(t) M_Y(t)$
3. If M_X is finite in an open interval around zero, then it is infinitely differentiable on that interval and $E(X^n) = \frac{d^n}{dt^n} M_X(t)|_{t=0}, \ \forall n \in \mathbb{N}$
4. If X and Y are r.v.'s such that $M_X(t) = M_Y(t) < \infty$ for t in an open interval around zero, then $P_X = P_Y$ and $F_X = F_Y$. Note that this does not imply that the mgf is unique to a random variable because there are multiple random variables whose mgf's shoot to infinity immediately around zero.

Definition 17. The **characteristic function** (cf) of X is the function $\phi_X : \mathbb{R} \to [0, \infty]$

$$\phi_X(t) = E(e^{itX}) = \int e^{itX} dP = \int e^{itx} dP_X = \int e^{itx} f_X(x) dx$$

Theorem 23. The following are properties of the cf:

1. $\forall a, b \in \mathbb{R}, \quad \phi_{aX+b}(t) = e^{ibt} \phi_X(at)$
2. If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t)$
3. If $E(X^n) < \infty$ then ϕ_X is n-times differentiable and $E(X^n) = i^{-n} \frac{d^n}{dt^n} \phi_X(t)|_{t=0}$
4. If X and Y are r.v.'s such that $\phi_X(t) = \phi_Y(t)$, then $P_X = P_Y$ and $F_X = F_Y$. This means that cf's, unlike mgf's, uniquely identify a random variable.

Theorem 24. (**Levy's continuity theorem**). Suppose X is an r.v. and $X_1, X_2, ...$ a sequence of r.v.'s. Then $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all t, if and only if, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all x at which $F_X(x)$ is continuous. That is, if a sequence of characteristic functions converges to a characteristic function, then the sequence of distribution functions would also converge to the corresponding distribution function, and vice versa. This makes sense because both cf's and CDF's uniquely identify random variables, so they should have the same behavior (recall mgf's and pdf's don't uniquely identify random variables).

4 Statistics

Theorem 25. If $X_1, ..., X_n$ are iid random variables distributed $N(\mu, \sigma^2)$ then

1. $\bar{X}_n$ and S_n^2 are independent
2. $\bar{X}_n \sim N(\mu, \sigma^2/n)$
3. $(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2_{n-1}$

Definition 18. A random variable X has a **stable distribution** if $\bar{X}$ has the same distribution as the underlying $X_1, X_2, ...$ whenever they are independent.

Definition 19. X is **degenerate** if it maps a set of Ω with probability 1 into a single value $a \in \mathbb{R}$.

4.1 Asymptotics

4.1.1 Key Theorems

Theorem 26. (**Weak law of large numbers**). Let $X_1, X_2, ...$ be an iid sequence of random variables with $E|X_i| < \infty$. Then $\bar{X}_n \xrightarrow{p} E(X_i)$ as $n \to \infty$. So the sample mean always converges in probability to the expected value. The proof is pretty straightforward using Chebyshev's inequality. There are many other conditions under which we can prove the Law of Large Numbers. For more, see Theorem **??**.

Definition 20. A sequence of random variables $Z_1, Z_2, ...$ is said to **converge in distribution** to a random variable Z, written $Z_n \xrightarrow{d}$ as $n \to \infty$ if $lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$ for all x at which $F_Z(x)$ is continuous.This is weaker than pointwise convergence because it needn't converge at non-continuity points.

Comment 1. Convergence in distribution does not imply convergence in mean. Mean comes from pdf and convergence in distribution is about CDF. CDF's converging doesn't mean pdf's converge.

Theorem 27. (**Lindberg-Levy Central Limit Theorem**). Let $X_1, X_2, ...$ be an iid sequence of random variables with expected value μ and finite variance σ^2. Then:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Essentially, z-scores are standard normally distributed in the limit no matter what the underlying distribution. Note, this is a special case of a broad class of theories called Central Limit Theorems. This one is easy to prove, so it's commonly used, but we don't actually need iid. For more conditions under which the CLT holds, see Theorem **??**.

4.1.2 Applications and Extensions

Definition 21. A sequence of statistics $\hat{\gamma}_1, \hat{\gamma}_2, ...$ is a **consistent estimator** of γ if $\hat{\gamma}_n \xrightarrow{p} \gamma$ as $n \to \infty$.

Definition 22. A sequence of random variables x_t is said to **converge in mean square** to μ is $\lim_{t\to\infty} E(x_t - \mu)^2 = 0$.

Theorem 28. Convergence in mean square implies asymptotic unbiasedness and consistency.

Theorem 29. Suppose $h : \mathbb{R}^k \to \mathbb{R}$ is continuous at $\mu \in \mathbb{R}^k$, and Borel measurable. Then $X_n \xrightarrow{p} \mu \;\Rightarrow\; h(X_n) \xrightarrow{p} h(\mu)$. So continuous functions maintain convergence in probability.

Theorem 30. If mgf of Z_n converges to mgf of Z then $Z_n \xrightarrow{d} Z$.

Theorem 31. (**Slutsky's Theorem**). If $X_n \xrightarrow{p} c$ and $Y_n \xrightarrow{d} Y$ then $X_n Y_n \xrightarrow{d} cY$ and $X_n + Y_n \xrightarrow{d} c + Y$.

Theorem 32. (**Continuous Mapping Theorem**). If $X_1, X_2, ...$ converges to X and $h : \mathbb{R} \to \mathbb{R}$ is continuous, then $h(X_n) \xrightarrow{d} h(X)$ as $n \to \infty$. This remains true if the set of points A at which h is discontinuous satisfies $P_X(A) = 0$. Essentially, continuous functions preserve convergence in distribution. This is similar to Theorem 29.

Theorem 33. (**Delta Method**). Suppose $Z_1, Z_2, ...$ satisfy $\sqrt{n}(Z_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ for some $\theta \in \mathbb{R}$. If 1) $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, 2) $g'(\theta)$ exists and is nonzero, then

$$\sqrt{n}\big(g(Z_n) - g(\theta)\big) \xrightarrow{d} N\big(0, \sigma^2 g'(\theta)^2\big)$$

Note there is also a 2^{nd}-order version of the Delta Method. Also, since the CLT establishes that $\bar{X}_n$ converges to the normal distribution, you can always use the Delta Method when you're dealing with the sample mean.

Definition 23. The following characterize **Big-O and little-o**:

1. $X_n \xrightarrow{p} 0 \;\Leftrightarrow\; X_n = o_p(1)$
2. $X_n \xrightarrow{p} c \;\Leftrightarrow\; X_n = c + o_p(1)$
3. If Y_n converges to zero, $X_n = o_p(Y_n)$ means X_n converges to zero faster than Y_n
4. If Y_n diverges to ∞, $X_n = o_p(Y_n)$ means X_n diverges to ∞ slower than Y_n
5. $X_n = o_p(Y_n) \Rightarrow \frac{X_n}{Y_n} \xrightarrow{p} 0$
6. If Y_n converges to zero, $X_n = O_p(Y_n)$ means X_n converges to zero at least as fast as Y_n
7. If Y_n diverges to ∞, $X_n = O_p(Y_n)$ means X_n diverges to ∞ at least as slow as Y_n
8. $X_n = O_p(Y_n)$ means $X_n = o_p(Z_n)$ for any sequence Z_n such that $Y_n = o_p(Z_n)$
9. $X_n = O_p(1)$ means X_n is not exploding off to ∞

Theorem 34. The following three facts hold:

1. $X_n = O_p(Y_n) \wedge X_n' = O_p(Y_n') \Rightarrow X_n X_n' = O_p(Y_n Y_n')$
2. Anything that converges in distribution is $O_p(1)$
3. $o(O_p(1)) = o_p(1)$