Course Overview

C&B Chapters:

1: Probability Theory
2: Transformations / Expectations
3: Common Families of Distributions
4: Multiple Random Variables
5: Properties of a Random Sample
6: Principles of Data Reduction
7: Point Estimation
8: Hypothesis Testing
9: Interval Estimation
10: Asymptotic Evaluations

1st semester (chapters 1–5)
2nd semester (chapters 6–10)

1st: Semester: Learning Tools
(info we'd like at our finger tips)

2nd semester: Applying Tools
(why we do what we do...)

Chapter - 1: Probability Theory

§1.1 Set Theory

Sample Space

Definition 1.1.1 The set, S, of all possible outcomes of a particular experiment is called the *sample space* for the experiment.

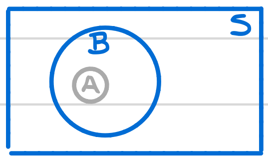
Events

Definition 1.1.2 An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

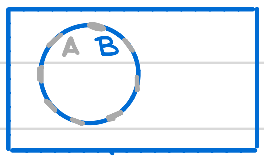
& Event Operations

A = Event, subset of S.
Event A occurs if outcome of experiment is in set A.

- Containment $A \subset B \Leftrightarrow x \in A \Rightarrow x \in B$

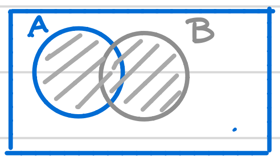


- Equality $A = B \Leftrightarrow A \subset B$ and $B \subset A$

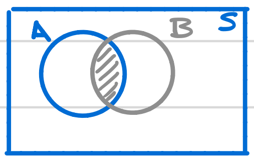


Operations

- Union: $A \cup B = \{x : x \in A$ or $x \in B\}$

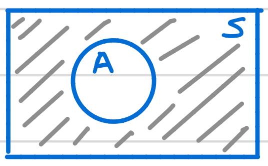


- Intersection: $A \cap B = AB = \{x : x \in A$ and $x \in B\}$



- Complementation: $A^c = \{x : x \notin A\}$

all elements of x — such that — x not in A



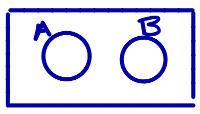
Properties of Events

Theorem 1.1.4 *For any three events, A, B, and C, defined on a sample space S,*

a. Commutativity	$A \cup B = B \cup A,$ $A \cap B = B \cap A;$
b. Associativity	$A \cup (B \cup C) = (A \cup B) \cup C,$ $A \cap (B \cap C) = (A \cap B) \cap C;$
c. Distributive Laws	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$
d. DeMorgan's Laws	$(A \cup B)^c = A^c \cap B^c,$ $(A \cap B)^c = A^c \cup B^c.$

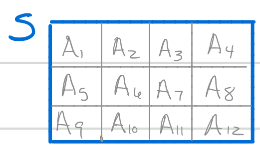
Disjoint

Definition 1.1.5 Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \emptyset$. The events $A_1, A_2, \ldots$ are *pairwise disjoint* (or *mutually exclusive*) if $A_i \cap A_j = \emptyset$ for all $i \neq j$.



Partition

Definition 1.1.6 If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$, then the collection $A_1, A_2, \ldots$ forms a *partition* of S.



§1.2 Basics of Probability

$\mathcal{B}$ Borel field or σ-algebra

Definition 1.2.1 A collection of subsets of S is called a *sigma algebra* (or *Borel field*), denoted by $\mathcal{B}$, if it satisfies the following three properties:

a. $\emptyset \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$).

b. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation).

c. If $A_1, A_2, \ldots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

For MS Theory, our problems will be well-behaved....
PhD Theory...

Probability function (Axioms of probability or Kolmogorov Axioms)

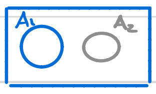
Definition 1.2.4 Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function P with domain $\mathcal{B}$ that satisfies

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.

2. $P(S) = 1$.

3. If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

(Simplified Axiom-3)
A_1 & A_2 are disjoint



$\Pr(A_1 \cup A_2) = P(A_1) + P(A_2)$

Probability $A \in \mathcal{B}$
sum over probs of subsets of S in A.
S finite

Theorem 1.2.6 *Let $S = \{s_1, \ldots, s_n\}$ be a finite set. Let $\mathcal{B}$ be any sigma algebra of subsets of S. Let $p_1, \ldots, p_n$ be nonnegative numbers that sum to 1. For any $A \in \mathcal{B}$, define $P(A)$ by*

$$P(A) = \sum_{\{i: s_i \in A\}} p_i.$$

(The sum over an empty set is defined to be 0.) Then P is a probability function on $\mathcal{B}$. This remains true if $S = \{s_1, s_2, \ldots\}$ is a countable set.

Probabilities $A \in \mathcal{B}$
$\emptyset$, A, A^c

Theorem 1.2.8 *If P is a probability function and A is any set in $\mathcal{B}$, then*

a. $P(\emptyset) = 0$, *where $\emptyset$ is the empty set;*

b. $P(A) \leq 1$;

c. $P(A^c) = 1 - P(A)$.

Probabilities
$B \cap A^c$, $A \cup B$, $A \subset B$

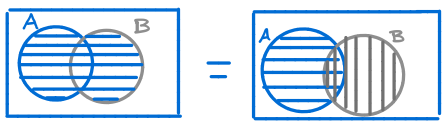
Theorem 1.2.9 *If P is a probability function and A and B are any sets in $\mathcal{B}$, then*

a. $P(B \cap A^c) = P(B) - P(A \cap B)$;

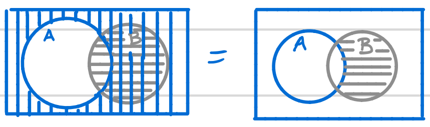
b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

c. *If $A \subset B$, then $P(A) \leq P(B)$.*

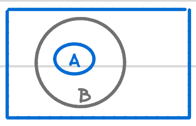
b)



a)



c)



Ms Theory II Review /4

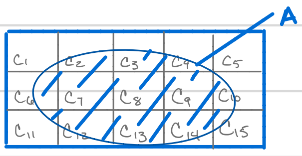
§ 1.2 cont.

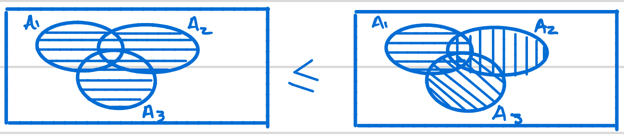
Probability functions
A, Partition $C_1, C_2, \ldots$
Union

Theorem 1.2.11 *If P is a probability function, then*

a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ *for any partition $C_1, C_2, \ldots$;*

b. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ *for any sets $A_1, A_2, \ldots$.* *(Boole's Inequality)*

a) 

b) 

Ordering tasks + objects

Sesame Street
Count von Count
(1972 – ...)
wikipedia accessed 8/7/18

Theorem 1.2.14 *If a job consists of k separate tasks, the ith of which can be done in n_i ways, $i = 1, \ldots, k$, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.*

Table 1.2.1. *Number of possible arrangements of size r from n objects*

	Without replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

most important for us. → Unordered

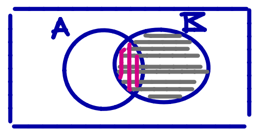
---

§1.3 Conditional Prob + Independence

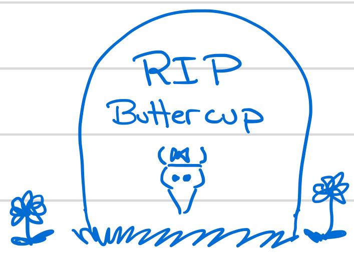
Conditional Probability

Definition 1.3.2 If A and B are events in S, and $P(B) > 0$, then the *conditional probability of A given B*, written $P(A|B)$, is

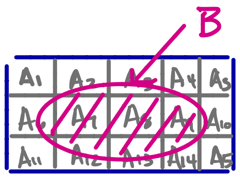
(1.3.1) 
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$



$P(B) > 0$



Theorem 1.3.5 (Bayes' Rule) *Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \ldots$,*



$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}. = \frac{P(B|A_i)P(A_i)}{P(B)}$$

T. Bayes.

Statistical Independence

$A \perp B$

Definition 1.3.7 Two events, A and B, are *statistically independent* if

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

Theorem 1.3.9 *If A and B are independent events, then the following pairs are also independent:*

a. *A and B^c,*

b. *A^c and B,*

c. *A^c and B^c.*

§1.3 cont.
Mutual Independence

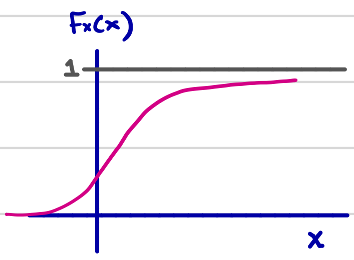
Definition 1.3.12 A collection of events $A_1, \ldots, A_n$ are *mutually independent* if for any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P\left(A_{i_j}\right).$$

§1.4 Random Variables

Definition 1.4.1 A *random variable* is a function from a sample space S into the real numbers.

§1.5 Dist'n ft'ns



CDF's.

Definition 1.5.1 The *cumulative distribution function* or *cdf* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \le x), \quad \text{for all } x. \leftarrow -\infty < x < \infty$$

Theorem 1.5.3 *The function $F(x)$ is a cdf if and only if the following three conditions hold:*

a. $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$.

b. *$F(x)$ is a nondecreasing function of x.*

c. *$F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x\downarrow x_0} F(x) = F(x_0)$.*

Continuous or Discrete

Definition 1.5.7 A random variable X is *continuous* if $F_X(x)$ is a continuous function of x. A random variable X is *discrete* if $F_X(x)$ is a step function of x.

Identically distributed

$F_X(x) = F_Y(x)$ $\forall x$

Definition 1.5.8 The random variables X and Y are *identically distributed* if, for every set $A \in \mathcal{B}^1$, $P(X \in A) = P(Y \in A)$.

well behaved σ-algebra → Not 'pathological cases'

Theorem 1.5.10 *The following two statements are equivalent:*

a. *The random variables X and Y are identically distributed.*

b. *$F_X(x) = F_Y(x)$ for every x.*

§1.6 Density & Mass functions

Discrete

Definition 1.6.1 The *probability mass function* (*pmf*) of a discrete random variable X is given by

$$f_X(x) = P(X = x) \quad \text{for all } x.$$

discrete cdf

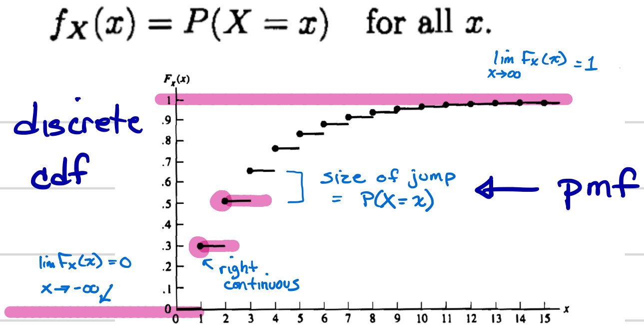


Figure 1.5.2. *Geometric cdf, $p = .3$*

Note cdf defined for $-\infty < x < \infty$

§1.6 cont.

Continuous pdf

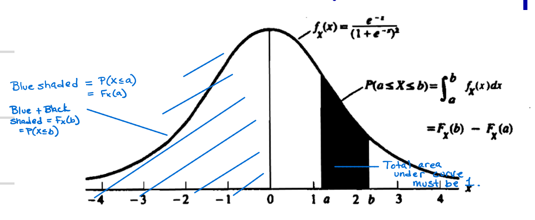


Figure 1.6.1. *Area under logistic curve*

Definition 1.6.3 The *probability density function* or *pdf*, $f_X(x)$, of a continuous random variable X is the function that satisfies

(1.6.3) $$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \text{for all } x.$$

$f_X(x) \ge 0 \ \forall x$

pmf: $\sum f_X(x) = 1$ pdf: $\int f_X(x)dx = 1$

Theorem 1.6.5 *A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if*

a. $f_X(x) \ge 0$ *for all* x.

b. $\sum_x f_X(x) = 1$ *(pmf)* or $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ *(pdf)*.

Chapter 2: Transformations + Expectations

§2.1 Dist'ns of ft'ns of R.V.

Continuous X cdf $F_Y(y)$ where $Y = g(X)$

cdf $F_Y(y)$

$X \sim F_X(x)$

$Y = g(X)$

Theorem 2.1.3 *Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as in (2.1.7).* ← $\mathcal{X} = \{x : f_X(x) > 0\}$ + $\mathcal{Y} = \{y : y = g(x)$ for some $x \in \mathcal{X}\}$.

a. *If g is an increasing function on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.*

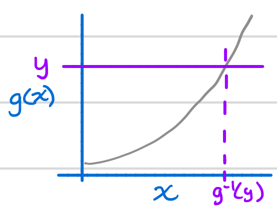
b. *If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.*

increasing g

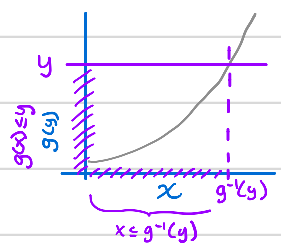
$F_Y(y) = F_X(g^{-1}(y))$

$y \in \mathcal{Y}$ ↑ Remember sample space!

$g(x)$ is increasing — Find $F_Y = P(Y \le y)$



$\{x \in \mathcal{X} : g(x) \le y\}$
$= \{x \in \mathcal{X} : g^{-1}(g(x)) \le g^{-1}(y)\}$
$= \{x \in \mathcal{X} : x \le g^{-1}(y)\}$



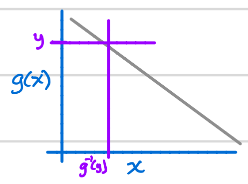
$$F_Y(y) = \int_{\{x \in \mathcal{X} : x \le g^{-1}(y)\}} f_X(x)\,dx = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx = F_X(g^{-1}(y)).$$

decreasing g

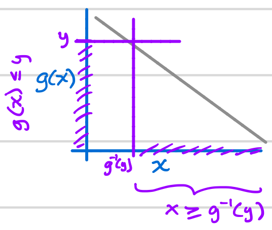
$F_Y(y) = 1 - F_X(g^{-1}(y))$

$y \in \mathcal{Y}$

$g(x)$ is decreasing — Find $F_Y = P(Y \le y)$



$\{x \in \mathcal{X} : g(x) \le y\}$
$= \{x \in \mathcal{X} : g^{-1}(g(x)) \ge g^{-1}(y)\}$
$= \{x \in \mathcal{X} : x \ge g^{-1}(y)\}$



$$F_Y(y) = \int_{g^{-1}(y)}^{\infty} f_X(x)\,dx = 1 - F_X(g^{-1}(y))$$

pdf $f_Y(y)$

$X \sim f_X(x)$

$Y = g(X)$

g monotone

Theorem 2.1.5 *Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Let $\mathcal{X}$ and $\mathcal{Y}$ be defined by (2.1.7). Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by*

(2.1.10) $$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \underbrace{\left| \frac{d}{dy} g^{-1}(y) \right|}_{\text{don't forget}} & y \in \mathcal{Y} \leftarrow \text{don't forget!} \\ 0 & \text{otherwise.} \end{cases}$$

2.1.7: $\mathcal{X} = \{x : f_X(x) > 0\}$ $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$.

§2.1 cont.

pdf $f_Y(y)$

$X \sim f_X(x)$
$Y = g(X)$

g Not monotone

Theorem 2.1.8 *Let X have pdf $f_X(x)$, let $Y = g(X)$, and define the sample space $\mathcal{X}$ as in (2.1.7). Suppose there exists a partition, $A_0, A_1, \ldots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_1(x), \ldots, g_k(x)$, defined on $A_1, \ldots, A_k$, respectively, satisfying*

i. *$g(x) = g_i(x)$, for $x \in A_i$,*
ii. *$g_i(x)$ is monotone on A_i,*
iii. *the set $\mathcal{Y} = \{y\colon\ y = g_i(x)$ for some $x \in A_i\}$ is the same for each $i = 1, \ldots, k$, and*
iv. *$g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$, for each $i = 1, \ldots, k$.*

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X\left(g_i^{-1}(y)\right) \left|\frac{d}{dy} g_i^{-1}(y)\right| & y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

X continuous w/ cdf $F_X(x)$
$Y = F_X(x)$
$Y \sim U(0,1)$

$P(Y \le y) = y,\ 0 < y < 1$

Theorem 2.1.10 (Probability integral transformation) *Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0, 1)$, that is, $P(Y \le y) = y, 0 < y < 1$.*

Why so important? If you can generate a $U(0,1)$, you can generate random variables from any continuous dist'n.

§2.2 Expected Values

Definition 2.2.1 The *expected value* or *mean* of a random variable $g(X)$, denoted by $\mathrm{E}\,g(X)$, is

$$\mathrm{E}\,g(X) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\,dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ is discrete,} \end{cases}$$

provided that the integral or sum exists. If $\mathrm{E}|g(X)| = \infty$, we say that $\mathrm{E}\,g(X)$ does not exist. (Ross 1988 refers to this as the "law of the unconscious statistician." We do not find this amusing.)

LOTUS - see appendix. lecture-5 6631.

Theorem 2.2.5 *Let X be a random variable and let a, b, and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,*

Proof in C&B pg. 57

a. $\mathrm{E}(ag_1(X) + bg_2(X) + c) = a\mathrm{E}\,g_1(X) + b\mathrm{E}\,g_2(X) + c$.
b. *If $g_1(x) \ge 0$ for all x, then $\mathrm{E}\,g_1(X) \ge 0$.*
c. *If $g_1(x) \ge g_2(x)$ for all x, then $\mathrm{E}\,g_1(X) \ge \mathrm{E}\,g_2(X)$.*
d. *If $a \le g_1(x) \le b$ for all x, then $a \le \mathrm{E}\,g_1(X) \le b$.*

§2.3 Moments & mgfs

Definition 2.3.1 For each integer n, the nth *moment of* X (or $F_X(x)$), μ'_n, is

$$\mu'_n = \mathrm{E}\,X^n.$$

The nth *central moment of* X, μ_n, is

$$\mu_n = \mathrm{E}(X - \mu)^n,$$

where $\mu = \mu'_1 = \mathrm{E}\,X$.

Moments & Central Moments

Definition 2.3.2 The *variance* of a random variable X is its second central moment, $\mathrm{Var}\,X = \mathrm{E}(X - \mathrm{E}\,X)^2$. The positive square root of $\mathrm{Var}\,X$ is the *standard deviation* of X.

§2.3 cont.

$Var(aX+b) = a^2 Var(X)$

Theorem 2.3.4 *If X is a random variable with finite variance, then for any constants a and b,*

$$\operatorname{Var}(aX+b) = a^2 \operatorname{Var} X.$$

Mgfs

Definition 2.3.6 Let X be a random variable with cdf F_X. The *moment generating function (mgf) of X* (or F_X), denoted by $M_X(t)$, is

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

provided that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $\mathrm{E}e^{tX}$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

More explicitly, we can write the mgf of X as

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx \qquad \text{if } X \text{ is continuous,}$$

or

$$M_X(t) = \sum_x e^{tx} P(X = x) \qquad \text{if } X \text{ is discrete.}$$

It is very easy to see how the mgf generates moments. We summarize the result in the following theorem.

Mgf $\longrightarrow E[X^n]$ moments

Theorem 2.3.7 *If X has mgf $M_X(t)$, then*

$$\mathrm{E}\, X^n = M_X^{(n)}(0),$$

where we define

$$M_X^{(n)}(0) = \left.\frac{d^n}{dt^n} M_X(t)\right|_{t=0}.$$

That is, the nth moment is equal to the nth derivative of $M_X(t)$ evaluated at $t = 0$.

mgf's $\rightarrow F_X(x)$

$F_X(u) \overset{?}{=} F_Y(u) \quad \forall u$

a) $E[X^r] = E[Y^r] \quad r = 0, 1, \ldots$
+ bounded support

b) $M_X(t) = M_Y(t)$
$\forall t \quad -h < t < h \quad h > 0$

Theorem 2.3.11 *Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist.*

a. *If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if $\mathrm{E}\, X^r = \mathrm{E}\, Y^r$ for all integers $r = 0, 1, 2, \ldots$.*

b. *If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.*

$M_X(t)$ exists $\forall t$
$-h < t < h \quad h > 0$ ← many dist'ns

F_X (or f_X) $\longrightarrow M_X(t) \longrightarrow \mu_r' = E[X^r]$
$r = 0, 1, 2, \ldots$

bounded support ← fewer dist'ns

MS Theory II

Review /9

§ 2.3 cont.

mgfs converge $\overset{\text{imply}}{\Longrightarrow}$ cdfs converge

Theorem 2.3.12 (Convergence of mgfs) *Suppose $\{X_i,\ i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that*

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t), \quad \text{for all } t \text{ in a neighborhood of } 0,$$

and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x).$$

That is, convergence, for $|t| < h$, of mgfs to an mgf implies convergence of cdfs.

$\lim_{i\to\infty} M_{X_i}(t) = M_X(t)$ $\forall |t| < h$

$\Longrightarrow$

$\lim_{i\to\infty} F_{X_i}(x) = F_X(x)$

example sequence of R.V.'s

$\bar{X}_1 = X_1/1$

$\bar{X}_2 = (X_1 + X_2)/2$

$\bar{X}_3 = (X_1 + X_2 + X_3)/3$

$\vdots$

$\bar{X}_n = \sum_{i=1}^{n} X_i / n$

each $\bar{X}_i$ has own mgf $M_{\bar{X}_i}(t)$ and cdf $F_{\bar{X}_i}(t)$

Lemma 2.3.14 *Let $a_1, a_2, \ldots$ be a sequence of numbers converging to a, that is, $\lim_{n\to\infty} a_n = a$. Then*

$$\lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n = e^a.$$

We'll use this lemma a few times this year. I will remind you when we need it

proof of CLT.

$M_{ax+b} = e^{bt} M_x(at)$

Theorem 2.3.15 *For any constants a and b, the mgf of the random variable $aX + b$ is given by*

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

Interchange $\int$ + $\frac{d}{d\theta}$.

$\frac{d}{d\theta} \int f(x,\theta)\, dx \overset{?}{=} \int \frac{\partial}{\partial\theta} f(x,\theta)\, dx$

- yes if bounds of integration not ft'n of θ.
- if bounds $-\infty$ +/or ∞, can if exponential family.

Theorem 2.4.1 (Leibnitz's Rule) *If $f(x,\theta)$, $a(\theta)$, and $b(\theta)$ are differentiable with respect to θ, then*

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x,\theta)\, dx = f(b(\theta),\theta) \frac{d}{d\theta} b(\theta) - f(a(\theta),\theta) \frac{d}{d\theta} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\, dx.$$

Notice that if $a(\theta)$ and $b(\theta)$ are constant, we have a special case of Leibnitz's Rule:

$$\underbrace{\frac{d}{d\theta} \int_a^b f(x,\theta)\, dx}_{\text{ft'n of } \theta} = \underbrace{\int_a^b \frac{\partial}{\partial\theta} f(x,\theta)\, dx}_{\text{ft'n } x,\theta}.$$

partial derivative

Similarly for 'exponential families' we can assume that we can interchange differentiation and summation.

$\frac{d}{d\theta}$ and $\sum_{i=1}^{\infty}$ or $\sum_{i=1}^{n}$.

Chapter 3: Common Dist'ns

§ 3.2 Discrete Dist'ns

Countable Sample Space

Discrete

Discrete Uniform (N)

Discrete uniform

get cdf by summing

pmf $P(X = x|N) = \frac{1}{N}; \quad x = 1, 2, \ldots, N; \quad N = 1, 2, \ldots$

mean and variance $\mathrm{E}X = \frac{N+1}{2}, \quad \mathrm{Var}\, X = \frac{(N+1)(N-1)}{12}$

$I_{[1,\ldots N]}^{(x)} = \begin{cases} 1 & x \in [1,\ldots N] \\ 0 & \text{else} \end{cases}$

mgf $M_X(t) = \frac{1}{N}\sum_{i=1}^{N} e^{it}$

Bernoulli (p)

Bernoulli(p)

pmf $P(X = x|p) = p^x(1-p)^{1-x}; \quad x = 0, 1; \quad 0 \le p \le 1$

mean and variance $\mathrm{E}X = p, \quad \mathrm{Var}\, X = p(1-p)$

mgf $M_X(t) = (1-p) + pe^t$

Binomial with n=1

$I_{[0,1]}^{(x)}$

Binomial (n, p)

We will assume n is known, unless stated otherwise.

Binomial(n, p)

usually parameter n assumed known

pmf $P(X = x|n,p) = \binom{n}{x}p^x(1-p)^{n-x}; \quad x = 0, 1, 2, \ldots, n; \quad 0 \le p \le 1$

mean and variance $\mathrm{E}X = np, \quad \mathrm{Var}\, X = np(1-p)$

$I_{[0,1,\ldots,n]}^{(x)}$

mgf $M_X(t) = [pe^t + (1-p)]^n$

notes Related to Binomial Theorem (Theorem 3.2.2). The *multinomial* distribution (Definition 4.6.2) is a multivariate version of the binomial distribution.

Binomial Theorem

Theorem 3.2.2 (Binomial Theorem) *For any real numbers x and y and integer $n \ge 0$,*

(3.2.4) $$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

If $x=y=1$

$(1+1)^n = 2^n = \sum_{i=0}^{n} \binom{n}{i}$

Poisson (λ)

Poisson(λ)

pmf $P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, \ldots; \quad 0 \le \lambda < \infty$

mean and variance $\mathrm{E}X = \lambda, \quad \mathrm{Var}\, X = \lambda$

$I_{[0,1,2,\ldots)}^{(x)}$

mgf $M_X(t) = e^{\lambda(e^t - 1)}$

$\boxed{\text{Tool: } e^y = \sum_{i=0}^{\infty} y^i/i!}$

Negative Binomial (r, p)

(# trials to r^{th} success)

Negative binomial(r, p)

pmf $P(X = x|r,p) = \binom{r+x-1}{x}p^r(1-p)^x; \quad x = 0, 1, \ldots; \quad 0 \le p \le 1$

mean and variance $\mathrm{E}X = \frac{r(1-p)}{p}, \quad \mathrm{Var}\, X = \frac{r(1-p)}{p^2}$

$I_{[0,1,\ldots)}^{(x)}$

mgf $M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r, \quad t < -\log(1-p)$

notes An alternate form of the pmf is given by $P(Y = y|r,p) = \binom{y-1}{r-1}p^r(1-p)^{y-r}$, $y = r, r+1, \ldots$. The random variable $Y = X + r$. The negative binomial can be derived as a gamma mixture of Poissons. (See Exercise 4.34.)

§ 3.2 Cont.

Geometric (p)

\# trials to 1^{st} success

note: can be reparameterized to # trials before 1^{st} success $X = [0, 1, \dots)$

Geometric(p)

pmf	$P(X = x\|p) = p(1-p)^{x-1}; \quad x = 1, 2, \dots; \quad 0 \le p \le 1$
mean and variance	$\mathrm{E}X = \frac{1}{p}, \quad \mathrm{Var}\, X = \frac{1-p}{p^2}$
mgf	$M_X(t) = \frac{pe^t}{1-(1-p)e^t}, \quad t < -\log(1-p)$
notes	$Y = X - 1$ is negative binomial$(1, p)$. The distribution is *memoryless*: $P(X > s\|X > t) = P(X > s - t)$.

$I_{[1, 2, \dots)}(x)$

§ 3.3 Continuous Dist'ns

Uniform (a, b)

Uniform(a, b)

Sample space is f'tn of parameters $a \le x \le b$

pdf	$f(x\|a, b) = \frac{1}{b-a}, \quad a \le x \le b$
mean and variance	$\mathrm{E}X = \frac{b+a}{2}, \quad \mathrm{Var}\, X = \frac{(b-a)^2}{12}$
mgf	$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$
notes	If $a = 0$ and $b = 1$, this is a special case of the beta $(\alpha = \beta = 1)$.

$I_{[a, b]}(x)$

Gamma (α, β)

Gamma Dist'n

Gamma(α, β)

pdf	$f(x\|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty, \quad \alpha, \beta > 0$
mean and variance	$\mathrm{E}X = \alpha\beta, \quad \mathrm{Var}\, X = \alpha\beta^2$
mgf	$M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha, \quad t < \frac{1}{\beta}$
notes	Some special cases are exponential $(\alpha = 1)$ and chi squared $(\alpha = p/2, \beta = 2)$. If $\alpha = \frac{3}{2}$, $Y = \sqrt{X/\beta}$ is *Maxwell*. $Y = 1/X$ has the *inverted gamma distribution*. Can also be related to the Poisson (Example 3.2.1).

$I_{[0, \infty)}(x)$

$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$

$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

$\Gamma(n) = (n-1)!$ $n > 0$, integer

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Normal (μ, σ^2)

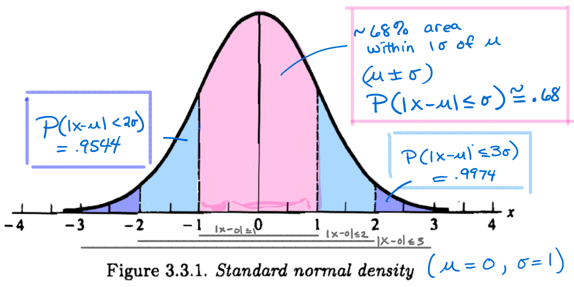


Figure 3.3.1. *Standard normal density* $(\mu = 0, \sigma = 1)$

Normal(μ, σ^2)

pdf	$f(x\|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$
mean and variance	$\mathrm{E}X = \mu, \quad \mathrm{Var}\, X = \sigma^2$
mgf	$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$
notes	Sometimes called the *Gaussian* distribution.

$I_{(-\infty, \infty)}(x)$

Beta (α, β)

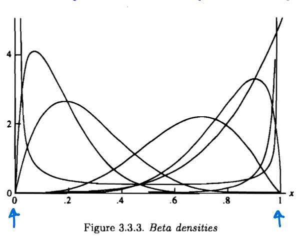


Figure 3.3.3. Beta densities

strictly increasing $\alpha > 1, \beta = 1$; strictly decreasing $\alpha = 1, \beta > 1$;

U shaped $(\alpha < 1, \beta < 1)$; unimodal $(\alpha > 1, \beta > 1)$; symmetric $(\alpha = \beta)$

Beta(α, β)

bounded support; used to model proportions

pdf	$f(x\|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \le x \le 1, \quad \alpha > 0, \quad \beta > 0$
mean and variance	$\mathrm{E}X = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}\, X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
mgf	$M_X(t) = 1 + \sum_{k=1}^\infty \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r}\right) \frac{t^k}{k!}$
notes	The constant in the beta pdf can be defined in terms of gamma functions, $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Equation (3.2.18) gives a general expression for the moments.

$I_{[0, 1]}(x)$

$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$

MS Theory II — Review/12

Cauchy (θ, σ)

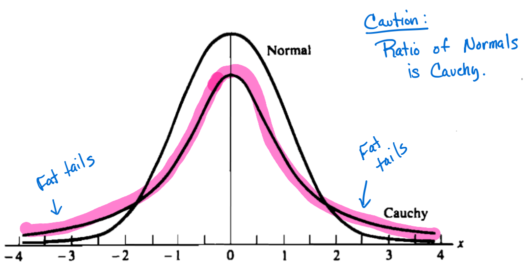


Figure 3.3.5. *Standard normal density and Cauchy density*

Cauchy(θ, σ)

pdf	$f(x\|\theta,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}, \quad -\infty < x < \infty; \quad -\infty < \theta < \infty, \quad \sigma > 0$
mean and variance	do not exist ←
mgf	does not exist ←
notes	Special case of Student's t, when degrees of freedom $= 1$. Also, if X and Y are independent $\mathrm{n}(0,1)$, X/Y is Cauchy.

$I^{(x)}_{(-\infty,\infty)}$

Lognormal (μ, σ^2)

Lognormal(μ, σ^2)

$I^{(x)}_{[0,\infty)}$

pdf	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log x-\mu)^2/(2\sigma^2)}}{x}, \quad 0 \le x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$
mean and variance	$\mathrm{E}X = e^{\mu+(\sigma^2/2)}, \quad \mathrm{Var}\, X = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$
moments (mgf does not exist)	$\mathrm{E}X^n = e^{n\mu+n^2\sigma^2/2}$
notes	Example 2.3.5 gives another distribution with the same moments.

Double exponential (μ, σ)

Double exponential(μ, σ)

$I^{(x)}_{(-\infty,\infty)}$

pdf	$f(x\|\mu,\sigma) = \frac{1}{2\sigma}e^{-\|x-\mu\|/\sigma}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$
mean and variance	$\mathrm{E}X = \mu, \quad \mathrm{Var}\, X = 2\sigma^2$
mgf	$M_X(t) = \frac{e^{\mu t}}{1-(\sigma t)^2}, \quad \|t\| < \frac{1}{\sigma}$
notes	Also known as the *Laplace* distribution.

§ 3.4

Exponential Families

$$f(x|\underline{\theta}) = h(x)\,c(\underline{\theta})\exp\left(\sum_{i=1}^{k} w_i(\underline{\theta})\,t_i(x)\right)$$

- $\underline{\theta}$ represents parameter
- $h(x) \geq 0$ ← can include indicator of support/sample space
- $t_1(x), t_2(x) \ldots t_k(x)$ ft'ns of x - don't depend on θ.
- $c(\underline{\theta}) \geq 0$
- $w_1(\underline{\theta}), \ldots, w_k(\underline{\theta})$ ft'ns of θ - don't depend on x

Take $f(x|\underline{\theta})$ + identify $h(x), c(\underline{\theta}), w_i(\underline{\theta})$ + $t_i(x)$
- ideally k as small as possible!

→ Note: If sample space is ft'n of θ, then it won't fit w/ $h(x)$ or $c(\theta)$.

→ We can interchange $\int_{-\infty}^{\infty}$ + $\frac{d}{d\theta}$ for exponential families ☺
- Needed for a few important Theorem proofs!

§ 3.4 cont.

C+B introduce indicator ft'n + use when sample space is ft'n of θ (not exponential family)

Definition 3.4.5 The *indicator function* of a set A, most often denoted by $I_A(x)$, is the function

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

An alternative notation is $I(x \in A)$.

Find expectations by taking derivatives! ☺ without mgf

exponential family

Theorem 3.4.2 *If X is a random variable with pdf or pmf of the form (3.4.1), then*

$$\text{(3.4.4)} \quad \mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial}{\partial \theta_j} \log c(\boldsymbol{\theta});$$

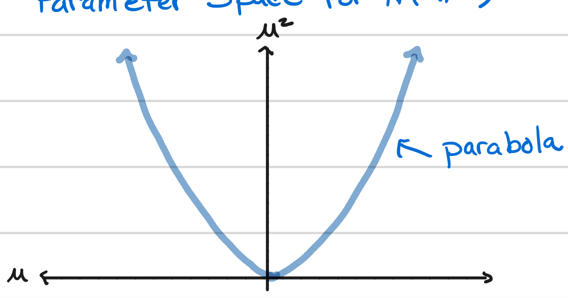
$$\text{(3.4.5)} \quad \mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j^2} \log c(\boldsymbol{\theta}) - \mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X)\right).$$

Curved exponential - $\dim(\theta) < k$

Definition 3.4.7 A *curved exponential family* is a family of densities of the form (3.4.1) for which the dimension of the vector $\boldsymbol{\theta}$ is equal to $d < k$. If $d = k$, the family is a *full exponential family*. (See also Miscellanea 3.8.3.)

$$f(x|\underline{\theta}) = h(x)\, c(\underline{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\underline{\theta})\, t_i(x)\right)$$

Parameter Space for $N(\mu, \mu^2)$

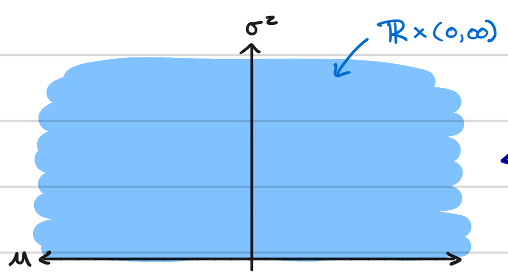


Curved Exponential

Full exponential - $\dim(\theta) = k$

statistical nirvana → A state of perfect happiness; an ideal or idyllic place ☺

Parameter Space for $N(\mu, \sigma^2)$



Full exponential

← open rectangle ☺ in $\mathbb{R}^k = \mathbb{R}^2$

Natural Parameter Exponential Family

An exponential family is sometimes reparameterized as

$$\text{(3.4.7)} \qquad f(x|\eta) = h(x) c^*(\eta) \exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right).$$

§3.4 Cont.

Natural parameter Exponential

since pdf $c^*(\eta) = \left[\int_{-\infty}^{\infty} h(x)\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right)\right]^{-1}$

$\eta_i = w_i(\theta)$, $i=1,\dots,k$

$$f(x|\eta) = h(x)c^*(\eta)\exp\left(\sum_{i=1}^{k}\eta_i t_i(x)\right).$$

$h(x)$, $t(x)$ same as original parameterization

$\mathcal{H}$ = natural parameter space

$$\mathcal{H} = \left\{\eta=(\eta_1,\dots,\eta_k) : \underbrace{\left(\int_{-\infty}^{\infty} h(x)\exp\sum_{i=1}^{k}\eta_i t_i(x)\,dx\right)}_{\frac{1}{c^*(\eta)}} < \infty\right\}$$

integral replaced by sum for discrete

§ 3.5 Location & Scale Families

Theorem 3.5.1 *Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants. Then the function*

$$g(x|\mu,\sigma) = \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

μ = location param

σ = scale param

$f(x|\mu=0, \sigma=1) = f(x)$

Location family

$f(x-\mu)$

Definition 3.5.2 Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$, is called the *location family with standard pdf* $f(x)$ and μ is called the *location parameter* for the family.

$f(x|\mu=0) = f(x-\mu|\mu)$

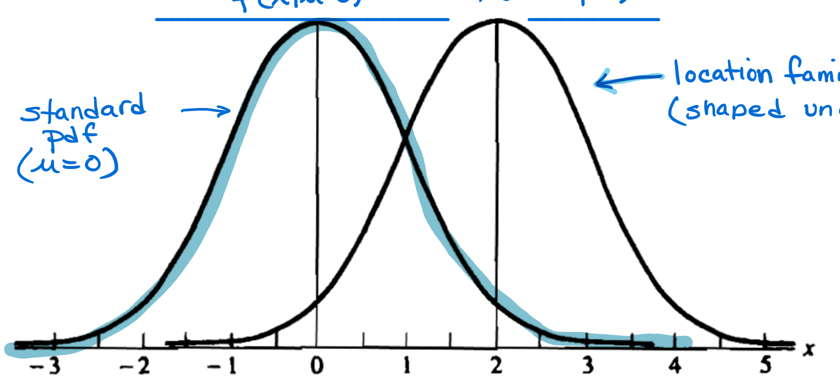


Figure 3.5.1. *Two members of the same location family: means at 0 and 2*

(same shape – shifted right by $\mu=2$)

$f(x-\mu) = f(x)$

Scale Family

$\frac{1}{\sigma}f(x/\sigma)$

Definition 3.5.4 Let $f(x)$ be any pdf. Then for any $\sigma > 0$, the family of pdfs $(1/\sigma)f(x/\sigma)$, indexed by the parameter σ, is called the *scale family with standard pdf* $f(x)$ and σ is called the *scale parameter* of the family.

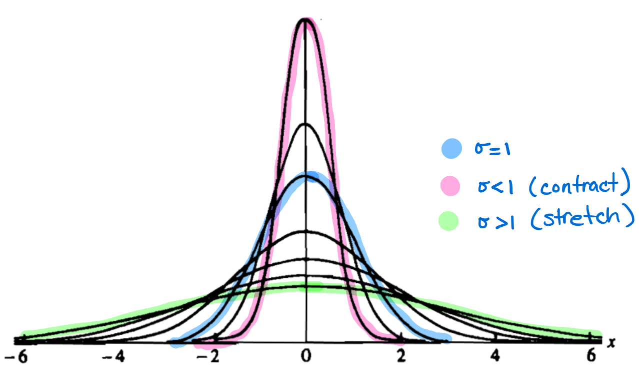


Figure 3.5.3. *Members of the same scale family*

§3.5 cont.

Location - Scale family

$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

Definition 3.5.5 Let $f(x)$ be any pdf. Then for any μ, $-\infty < \mu < \infty$, and any $\sigma > 0$, the family of pdfs $(1/\sigma)f((x-\mu)/\sigma)$, indexed by the parameter (μ, σ), is called the *location–scale family with standard pdf* $f(x)$; μ is called the *location parameter* and σ is called the *scale parameter.*

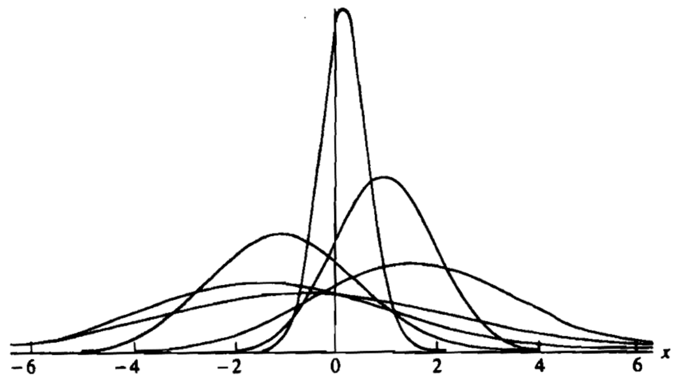


Figure 3.5.4. *Members of the same location–scale family*

$f_x(x) = {}^1/_\sigma f\left(\frac{x-\mu}{\sigma}\right)$

$\Rightarrow$ iff $\Leftarrow$

$f(z)$ and $X = \sigma Z + \mu$

Theorem 3.5.6 *Let $f(\cdot)$ be any pdf. Let μ be any real number, and let σ be any positive real number. Then X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.*

$f(z)$ w/ $E[Z]=0$ $Var[Z]=1$

$X = \sigma Z + \mu$

$f_x(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

$E[X] = \mu$ $Var[X] = \sigma^2$

Theorem 3.5.7 *Let Z be a random variable with pdf $f(z)$. Suppose* EZ *and* Var Z *exist. If X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$, then*

$$\mathrm{E}X = \sigma \mathrm{E}Z + \mu \quad \text{and} \quad \mathrm{Var}\, X = \sigma^2 \mathrm{Var}\, Z.$$

In particular, if $\mathrm{E}Z = 0$ *and* $\mathrm{Var}\, Z = 1$, *then* $\mathrm{E}X = \mu$ *and* $\mathrm{Var}\, X = \sigma^2$.

§3.6 Inequalities & ~~Identities~~

Cheby's

Theorem 3.6.1 (Chebychev's Inequality) *Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,*

$$P(g(X) \geq r) \leq \frac{\mathrm{E}g(X)}{r}.$$

X	univariate
(X, Y)	bivariate
$(X_1, \ldots, X_n)$	multivariate

Chapter 4: Multiple Random Variables

Random Variable

Recall **Definition 1.4.1** A *random variable* is a function from a sample space S into the real numbers.

Random Vector

Definition 4.1.1 An *n-dimensional random vector* is a function from a sample space S into $\Re^n$, n-dimensional Euclidean space.

§4.2 Conditional Dist'ns and Independence

Discrete

joint pmf

$f(x,y) = P(X=x, Y=y)$

Definition 4.1.3 Let (X, Y) be a discrete bivariate random vector. Then the function $f(x, y)$ from $\Re^2$ into $\Re$ defined by $f(x, y) = P(X = x, Y = y)$ is called the *joint probability mass function* or *joint pmf* of (X, Y). If it is necessary to stress the fact that f is the joint pmf of the vector (X, Y) rather than some other vector, the notation $f_{X,Y}(x, y)$ will be used.

marginal pmf's
$f_X(x)$, $f_Y(y)$

Theorem 4.1.6 *Let (X, Y) be a discrete bivariate random vector with joint pmf $f_{X,Y}(x, y)$. Then the marginal pmfs of X and Y, $f_X(x) = P(X = x)$ and $f_Y(y) = P(Y = y)$, are given by*

$$f_X(x) = \sum_{y \in \Re} f_{X,Y}(x, y) \quad \text{and} \quad f_Y(y) = \sum_{x \in \Re} f_{X,Y}(x, y).$$

$$\text{Sample Space} \longrightarrow f_{xy}(x,y) \longrightarrow f_x(x), f_y(y)$$

$$\cancel{\longleftarrow} \qquad\qquad \cancel{\longleftarrow}$$

different joint pmfs may have same marginals.

Conditional pmf

$f(y|x) =$

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

$= f(x,y)/f_X(x)$

Definition 4.2.1 Let (X, Y) be a discrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) > 0$, the *conditional pmf of* Y *given that* $X = x$ is the function of y denoted by $f(y|x)$ and defined by

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

For any y such that $P(Y = y) = f_Y(y) > 0$, the *conditional pmf of* X *given that* $Y = y$ is the function of x denoted by $f(x|y)$ and defined by

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

Continuous

$P((X,Y) \in A) = \int_A \int f(x,y)\, dx\, dx$

$\mathbb{R}^2 \to \mathbb{R}$

$A \subset \mathbb{R}^2$

Definition 4.1.10 A function $f(x, y)$ from $\Re^2$ into $\Re$ is called a *joint probability density function* or *joint pdf of the continuous bivariate random vector* (X, Y) if, for every $A \subset \Re^2$,

$$P((X, Y) \in A) = \underbrace{\int_A \int}_{} f(x, y)\, dx\, dy.$$

double integrals, integrate over all $(x,y) \in A$.

$$\boxed{E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy}$$

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

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

1)

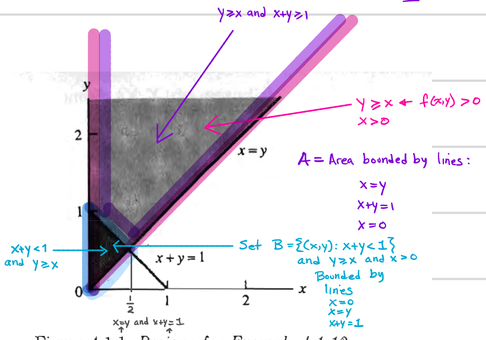


Figure 4.1.1. *Regions for Example 4.1.12*

Graphs almost always needed!

Continuous cont.
Conditional pdf

$f(y|x) = \frac{f(x,y)}{f(x)}$

Definition 4.2.3 Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the *conditional pdf of* Y *given that* $X = x$ is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = \frac{f(x,y)}{f_X(x)}.$$

For any y such that $f_Y(y) > 0$, the *conditional pdf of* X *given that* $Y = y$ is the function of x denoted by $f(x|y)$ and defined by

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Independence : $X \perp Y$

$f(x,y) = f_x(x) f_y(y)$

$\rightarrow X \perp Y$

Definition 4.2.5 Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called *independent random variables* if, for every $x \in \Re$ and $y \in \Re$,

$$f(x,y) = f_X(x) f_Y(y). \tag{4.2.1}$$

$X \perp Y \underset{\text{if}}{\overset{\text{only if}}{\rightleftarrows}} f(x,y) = g(x) h(y)$

Lemma 4.2.7 *Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$. Then X and Y are independent random variables if and only if there exist functions $g(x)$ and $h(y)$ such that, for every $x \in \Re$ and $y \in \Re$,*

$$f(x,y) = g(x)h(y).$$

$X \perp Y$:

$P(X \in A, Y \in B)$
$= P(X \in A) P(Y \in B)$

$E[g(X) h(Y)]$
$= E(g(X)) E[g(Y)]$

Theorem 4.2.10 *Let X and Y be independent random variables.*

a. *For any $A \subset \Re$ and $B \subset \Re$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$; that is, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent events.*

b. *Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then*

$$\mathrm{E}\left(g(X)h(Y)\right) = \left(\mathrm{E}g(X)\right)\left(\mathrm{E}h(Y)\right).$$

$X \perp Y$

$M_{X+Y}(t) = M_X(t) M_Y(t)$

Theorem 4.2.12 *Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by*

$$M_Z(t) = M_X(t)M_Y(t).$$

- $X \sim N(\mu, \sigma^2)$ + $Y \sim N(\gamma, \tau^2)$
- $X \perp Y$
- $X+Y \sim N(\mu+\gamma, \sigma^2+\tau^2)$

Theorem 4.2.14 *Let $X \sim \mathrm{n}(\mu, \sigma^2)$ and $Y \sim \mathrm{n}(\gamma, \tau^2)$ be independent normal random variables. Then the random variable $Z = X + Y$ has a $\mathrm{n}(\mu + \gamma, \sigma^2 + \tau^2)$ distribution.*

§4.3 Bivariate Transforms

Know $f(x,y)$

want $f(u,v)$

where $u = g_1(x,y)$
$v = g_2(x,y)$

§4.3 Bivariate Transforms

For a bivariate transform

Assume: X, Y bivariate random vector w/ known $f(x,y)$.

Consider: $\left.\begin{array}{l} U = g_1(X,Y) \\ V = g_2(X,Y) \end{array}\right\}$ where $g_1(x,y)$ & $g_2(x,y)$ are given.

If $B \subset \mathbb{R}^2$ then $(U,V) \in B \overset{\text{iff}}{\Longleftrightarrow} (X,Y) \in A$;
where $A = \{(x,y): (g_1(x,y), g_2(x,y)) \in B\}$

Thus: $P((U,V) \in B) = P((X,Y) \in A)$ and prob. dist'n of (U,V) is determined by prob. dist'n of (X,Y).

$(U,V) \in B$ must be consistent w/ A and $g_1(x,y), g_2(x,y)$

Discrete Example

Example Poisson $X \perp Y$

$$f_{XY}(x,y) = \frac{\theta^x e^{-\theta}}{x!} \frac{\lambda^y e^{-\lambda}}{y!} I_{[0,1,2\ldots)}(x) \, I_{[0,1,\ldots)}(y)$$

define: $\begin{array}{l} U = X+Y \\ V = Y \end{array}$ $\quad \begin{array}{l} g_1(x,y) = x+y \\ g_2(x,y) = y \end{array}$

$\mathcal{A} = \{(x,y): x = 0,1,2\ldots \text{ and } y = 0,1,2,\ldots\}$

$\mathcal{B} = \{(u,v): v = 0,1,2,\ldots \text{ and } u = v, v+1, \ldots\}$ ← or $\begin{array}{l} u = 0,1,2,\ldots \\ v = 0,\ldots u \end{array}$

for any (u,v) $\left.\begin{array}{l} x = u-v \\ y = v \end{array}\right\}$ satisfies $x+y = u$ and $y = v$

$A_{uv} = (u-v, v)$

$$f_{U,V}(u,v) = \frac{\theta^{u-v} e^{-\theta}}{(u-v)!} \frac{\lambda^v e^{-\lambda}}{v!} \underline{I_{[0,1,2,\ldots)}(v) \, I_{[v,v+1,v+2,\ldots)}(u(v))}$$

$0 \leq v \leq u < \infty$

$= I_{[0,1,\ldots)}(v) \, I_{[v,v+1,\ldots)}(u(v))$

$= I_{[0,1,\ldots)}(u) \, I_{[0,1,\ldots u]}(v(u))$

Find marginals

Find marginals

know $y = v \Rightarrow f_V(v) \sim \text{Pois}(\lambda)$

$$f_U(u) = \sum_{v=0}^{u} \frac{\theta^{u-v} e^{-\theta} \lambda^v e^{-\lambda}}{(u-v)! \, v!} = e^{-(\theta+\lambda)} \sum_{v=0}^{u} \frac{\theta^{u-v} \lambda^v}{(u-v)! \, v!} I_{[0,1,2\ldots)}(u)$$

$$= \frac{e^{-(\theta+\lambda)}}{u!} \sum_{v=0}^{u} \binom{u}{v} \theta^{u-v} \lambda^v$$

u	v (Y)	u−v (X)
0	0	0
1	0	1
1	1	0
2	0	2
2	1	1
2	2	0
⋮		

Theorem 3.2.2 (Binomial Theorem) *For any real numbers x and y and integer $n \geq 0$,*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}. \tag{3.2.4}$$

$$f_U(u) = \frac{e^{-(\theta+\lambda)}}{u!} (\theta+\lambda)^u \, I_{[0,1,2,\ldots)}(u)$$

Sum of independent Poisson, is Poisson
└ sum parameters.

Sum of Poissons ~ Poisson

Theorem 4.3.2 *If $X \sim$ Poisson(θ) and $Y \sim$ Poisson(λ) and X and Y are independent, then $X+Y \sim$ Poisson$(\theta+\lambda)$.*

$X \perp Y \rightarrow g(X) \perp h(Y)$

Theorem 4.3.5 *Let X and Y be independent random variables. Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent.*

Bivariate Transformation
Continuous

$$f_{U,V}(u,v) = f_{X,Y}\left(h_1(u,v), h_2(u,v)\right) |J|$$

where $u = g_1(x,y)$, $v = g_2(x,y)$ $\xrightarrow{\text{one-to-one}}$ $x = h_1(u,v)$, $y = h_2(u,v)$

$\mathcal{A} = \{(x,y): f_{X,Y}(x,y) > 0\}$

$\mathcal{B} = \{(u,v): u = g_1(x,y) \text{ and } v = g_2(x,y) \text{ for some } (x,y) \in \mathcal{A}\}$

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}$$ and $|J|$ = absolute value (J)

§4.4 Hierarchical Models & Mixture Dist'n

Conditional Expectation Identity

$E[X] = E\left[E[X|Y]\right]$

Theorem 4.4.3 *If X and Y are any two random variables, then*

(4.4.1) $$\mathrm{E}X = \mathrm{E}\left(\mathrm{E}(X|Y)\right),$$

provided that the expectations exist.

Mixture Dist'n

Definition 4.4.4 A random variable X is said to have a *mixture distribution* if the distribution of X depends on a quantity that also has a distribution.

Conditional Variance Identity

$\mathrm{Var}(X) = E\left[\mathrm{Var}(X|Y)\right] + \mathrm{Var}\left[E[X|Y]\right]$

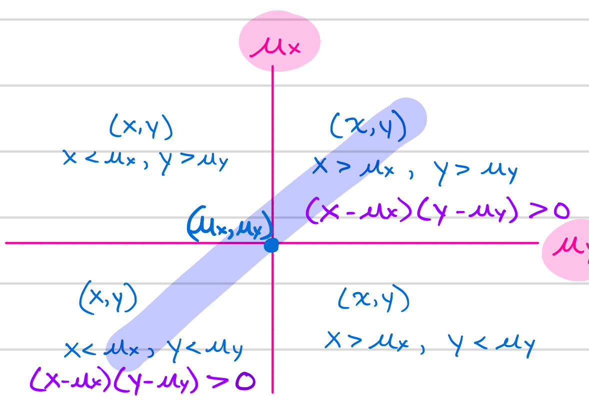
Theorem 4.4.7 (Conditional variance identity) *For any two random variables X and Y,*

(4.4.4) $$\mathrm{Var}\,X = \mathrm{E}\left(\mathrm{Var}(X|Y)\right) + \mathrm{Var}\left(\mathrm{E}(X|Y)\right),$$

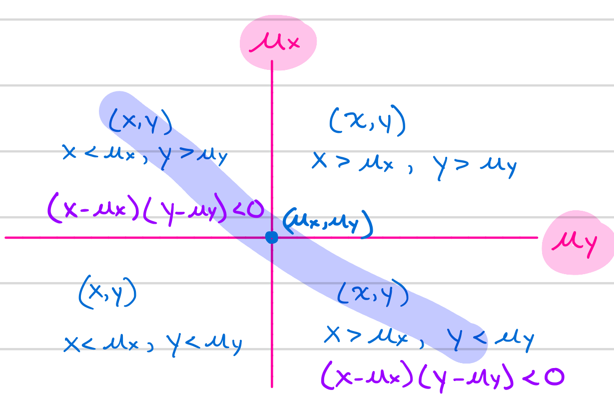
provided that the expectations exist.

§4.5 Covariance and Correlation

Positive & Negative Covariance / Correlation



Positive Covariance/Correlation



Negative Covariance/Correlation

MS Theory II

Review | 20

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

Definition 4.5.1 The *covariance of* X *and* Y is the number defined by

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left((X-\mu_X)(Y-\mu_Y)\right).$$

$\rho_{xy} = \dfrac{Cov(X,Y)}{\sigma_x \sigma_y}$

Definition 4.5.2 The *correlation of* X *and* Y is the number defined by

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

The value ρ_{XY} is also called the *correlation coefficient*.

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

Theorem 4.5.3 *For any random variables* X *and* Y,

$$\mathrm{Cov}(X,Y) = \mathrm{E}XY - \mu_X\mu_Y.$$

$X \perp Y \rightarrow Cov(X,Y) = 0$

$X \perp Y \rightarrow \rho_{xy} = 0$

Theorem 4.5.5 *If* X *and* Y *are independent random variables, then* $\mathrm{Cov}(X,Y) = 0$ *and* $\rho_{XY} = 0$.

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

Theorem 4.5.6 *If* X *and* Y *are any two random variables and* a *and* b *are any two constants, then*

$$\mathrm{Var}(aX+bY) = a^2\mathrm{Var}\,X + b^2\mathrm{Var}\,Y + 2ab\mathrm{Cov}(X,Y).$$

If X *and* Y *are independent random variables, then*

$$\mathrm{Var}(aX+bY) = a^2\mathrm{Var}\,X + b^2\mathrm{Var}\,Y.$$

$-1 \le \rho_{xy} \le 1$

$|\rho_{xy}| = 1 \overset{iff}{\Leftrightarrow} Y = aX + b$

Theorem 4.5.7 *For any random variables* X *and* Y,

a. $-1 \le \rho_{XY} \le 1$.

b. $|\rho_{XY}| = 1$ *if and only if there exist numbers* $a \ne 0$ *and* b *such that* $P(Y = aX + b) = 1$. *If* $\rho_{XY} = 1$, *then* $a > 0$, *and if* $\rho_{XY} = -1$, *then* $a < 0$.

Bivariate Normal

$\begin{pmatrix} X \\ Y \end{pmatrix} \sim MVN\left(\begin{matrix} \mu_x \\ \mu_y \end{matrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \right)$

Cov(x,y)

$\rho = Corr(X,Y)$

MVN = multivariate normal

Definition 4.5.10 Let $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$, and $-1 < \rho < 1$ be five real numbers. The *bivariate normal pdf with means* μ_X *and* μ_Y, *variances* σ_X^2 *and* σ_Y^2, *and correlation* ρ is the bivariate pdf given by

$$f(x,y) = \left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}\right)^{-1} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$.

Parameters $\underbrace{\mu_x, \mu_y \quad \sigma_x, \sigma_y}$ ρ

Properties

a. $f_X(x) \sim N(\mu_x, \sigma_x^2)$

b. $f_Y(y) \sim N(\mu_y, \sigma_y^2)$

c. correlation between X & Y $= \rho_{xy} = \rho$.

d. dist'n of $aX + bY \sim N(a\mu_x + b\mu_y,\ a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)$

proof homework.

§4.6 Multivariate Dist'ns

X, Y bivariate

$X_1, X_2, \ldots, X_n$ multivariate

bivariate (X, Y)

multivariate $(X_1, X_2, \ldots, X_n)$ $\leftarrow$ Generalize from bivariate to multivariate

Notation

C & B: $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ ← bold = vector

$\mathbf{x} = (x_1, x_2, \ldots, x_n)$

Sam: $\underset{\sim}{X} = (X_1, \ldots X_n)$ ← vector

$\underset{\sim}{x} = (x_1, \ldots x_n)$ ← vector

$\underset{\sim}{X}$ sample space is subset of $\mathbb{R}^n$.

Multinomial Dist'n (generalize binomial)

Discrete

Definition 4.6.2 Let n and m be positive integers and let $p_1, \ldots, p_n$ be numbers satisfying $0 \leq p_i \leq 1$, $i = 1, \ldots, n$, and $\sum_{i=1}^n p_i = 1$. Then the random vector $(X_1, \ldots, X_n)$ has a *multinomial distribution with m trials and cell probabilities* $p_1, \ldots, p_n$ if the joint pmf of $(X_1, \ldots, X_n)$ is

$$f(x_1, \ldots, x_n) = \frac{m!}{x_1! \cdots \cdot x_n!} p_1^{x_1} \cdot \cdots \cdot p_n^{x_n} = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$$

on the set of $(x_1, \ldots, x_n)$ such that each x_i is a nonnegative integer and $\sum_{i=1}^n x_i = m$.

Generalized Binomial Thm.

Theorem 4.6.4 (Multinomial Theorem) *Let m and n be positive integers. Let $\mathcal{A}$ be the set of vectors $\mathbf{x} = (x_1, \ldots, x_n)$ such that each x_i is a nonnegative integer and $\sum_{i=1}^n x_i = m$. Then, for any real numbers $p_1, \ldots, p_n$,*

$$\underbrace{(p_1 + \cdots + p_n)^m}_{1^m} = \sum_{\mathbf{x} \in \mathcal{A}} \frac{m!}{x_1! \cdots \cdot x_n!} p_1^{x_1} \cdot \cdots \cdot p_n^{x_n}$$

If $\sum p_i = 1$

If $\sum_{i=1}^n p_i = 1 \Rightarrow$ multinomial sums to 1.

Independence

$f(\underset{\sim}{x}) = \prod_{i=1}^n f_{X_i}(x_i)$

if $X_i \perp X_j \; \forall i \neq j$ & mutually independent

Definition 4.6.5 Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be random vectors with joint pdf or pmf $f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Let $f_{\mathbf{X}_i}(\mathbf{x}_i)$ denote the marginal pdf or pmf of $\mathbf{X}_i$. Then $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are called *mutually independent random vectors* if, for every $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$,

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = f_{\mathbf{X}_1}(\mathbf{x}_1) \cdot \cdots \cdot f_{\mathbf{X}_n}(\mathbf{x}_n) = \prod_{i=1}^n f_{\mathbf{X}_i}(\mathbf{x}_i).$$

If the X_is are all one-dimensional, then $X_1, \ldots, X_n$ are called *mutually independent random variables.*

mutual independence $\Rightarrow \underbrace{X_j \perp X_i \;\; \forall i,j}_{\text{pairwise } \perp}$

$\not\Leftarrow$

$E[g_1(X_1) \cdot \ldots \cdot g_n(X_n)] = E[g(X_1)] \cdots E[g(X_n)]$

Theorem 4.6.6 (Generalization of Theorem 4.2.10) *Let $X_1, \ldots, X_n$ be mutually independent random variables. Let $g_1, \ldots, g_n$ be real-valued functions such that $g_i(x_i)$ is a function only of x_i, $i = 1, \ldots, n$. Then*

$$\mathrm{E}(g_1(X_1) \cdot \cdots \cdot g_n(X_n)) = (\mathrm{E}g_1(X_1)) \cdot \cdots \cdot (\mathrm{E}g_n(X_n)).$$

$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \ldots \cdot M_{X_n}(t)$

Theorem 4.6.7 (Generalization of Theorem 4.2.12) *Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $Z = X_1 + \cdots + X_n$. Then the mgf of Z is*

$$M_Z(t) = M_{X_1}(t) \cdot \cdots \cdot M_{X_n}(t).$$

$X_1, \ldots X_n$ iid

$M_{X_1+X_2+\cdots+X_n}(t) = (M_X(t))^n$

In particular, if $X_1, \ldots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = (M_X(t))^n.$$

Independence Cont.

$Z = \sum_{i=1} (a_i X_i + b_i)$

$M_Z(t) = e^{t\sum b_i} \prod_{i=1}^{n} M_{X_i}(a_i t)$

Corollary 4.6.9 *Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Let $Z = (a_1X_1 + b_1) + \cdots + (a_nX_n + b_n)$. Then the mgf of Z is*

$$M_Z(t) = (e^{t(\Sigma b_i)})M_{X_1}(a_1 t) \cdot \cdots \cdot M_{X_n}(a_n t).$$

$X_1, \ldots, X_n$ mutually independent $N(\mu_i, \sigma_i^2)$

$\sum X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$

Corollary 4.6.10 *Let $X_1, \ldots, X_n$ be mutually independent random variables with $X_i \sim \mathrm{n}(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Then*

$$Z = \sum_{i=1}^{n} (a_i X_i + b_i) \sim \mathrm{n}\left(\sum_{i=1}^{n} (a_i \mu_i + b_i), \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

$\underset{\sim}{X_1}, \ldots, \underset{\sim}{X_n}$ mutually $\perp$ R. Vectors

$\Longleftrightarrow$

$f(\underset{\sim}{x_1}, \ldots \underset{\sim}{x_n}) = g_1(\underset{\sim}{x_1}) \ldots g_n(\underset{\sim}{x_n})$

Theorem 4.6.11 (Generalization of Lemma 4.2.7) *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be random vectors. Then $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are mutually independent random vectors if and only if there exist functions $g_i(\mathbf{x}_i), i = 1, \ldots, n$, such that the joint pdf or pmf of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ can be written as*

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = g_1(\mathbf{x}_1) \cdot \cdots \cdot g_n(\mathbf{x}_n).$$

$\underset{\sim}{X_1}, \ldots, \underset{\sim}{X_n}$ $\perp$ Random Vectors

$U_i = g_i(X_i)$ $i = 1, \ldots n$ mutually $\perp$

Theorem 4.6.12 (Generalization of Theorem 4.3.5) *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random vectors. Let $g_i(\mathbf{x}_i)$ be a function only of $\mathbf{x}_i, i = 1, \ldots, n$. Then the random variables $U_i = g_i(\mathbf{X}_i), i = 1, \ldots, n$, are mutually independent.*

Multivariate Transforms

$f_U(u_1, \ldots, u_n) =$

$\sum_{i=1}^{k} f_X(h_1(\underset{\sim}{u}), \ldots h_n(\underset{\sim}{u})) \, |J_i|$

partition transformation one-to-one on A_i

Transform $(X_1, \ldots X_n)$ w/ pdf $f_{\underset{\sim}{X}}(x_1, \ldots x_n)$ $\mathcal{A} = \{\underset{\sim}{x} : f_{\underset{\sim}{X}}(\underset{\sim}{x}) > 0\}$

to $(U_1, \ldots, U_n)$

where $U_1 = g_1(X_1, \ldots, X_n), \ldots, U_n = g_n(X_1, \ldots, X_n)$ $\mathcal{A} = \{(x,y) : f_{X,Y}(x,y) > 0\}$

and transformation is one-to-one from A_i to $\mathcal{B}$ $i = 1, \ldots, k$

$\overbrace{A_1, \ldots A_k}^{\text{partition}} + A_0$ $P(\underset{\sim}{X} \in A_0) = 0$

- inverses: $x_1 = h_{1i}(u_1, \ldots u_n)$ $\vdots$ $x_n = h_{ni}(u_1, \ldots u_n)$ } 'well behaved' (details C&B) $\rightarrow$ $\mathcal{B} = \{(\underset{\sim}{u}) : u_1 = g_{1i}(\underset{\sim}{x}), \ldots u_n = g_{ni}(\underset{\sim}{x})$ for some $(\underset{\sim}{x}) \in \mathcal{A}\}$

Don't forget the Jacobian!

$$J_i = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \cdots & \frac{\partial x_1}{\partial u_n} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \cdots & \frac{\partial x_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \cdots & \frac{\partial x_n}{\partial u_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_{1i}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{1i}(\mathbf{u})}{\partial u_2} & \cdots & \frac{\partial h_{1i}(\mathbf{u})}{\partial u_n} \\ \frac{\partial h_{2i}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{2i}(\mathbf{u})}{\partial u_2} & \cdots & \frac{\partial h_{2i}(\mathbf{u})}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_{ni}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{ni}(\mathbf{u})}{\partial u_2} & \cdots & \frac{\partial h_{ni}(\mathbf{u})}{\partial u_n} \end{vmatrix},$$

Then

$$f_{\mathbf{U}}(u_1, \ldots, u_n) = \sum_{i=1}^{k} f_{\mathbf{X}}(h_{1i}(u_1, \ldots, u_n), \ldots, h_{ni}(u_1, \ldots, u_n))|J_i|. \tag{4.6.7}$$

Cauchy-Schwartz Inequality

(use in a few proofs).

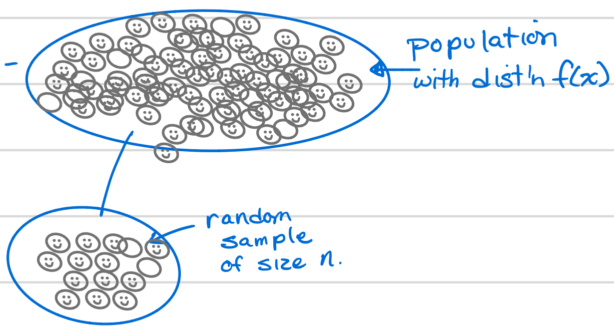
§4.7 Inequalities (Cauchy-Schwartz Inequality).

Theorem 4.7.3 (Cauchy–Schwarz Inequality) *For any two random variables X and Y,*

$$|\mathrm{E}XY| \le \mathrm{E}|XY| \le (\mathrm{E}|X|^2)^{1/2}(\mathrm{E}|Y|^2)^{1/2}. \tag{4.7.4}$$

§ 5.1 Basic Concepts of a Random Sample

- Random Sampling



Definition 5.1.1 The random variables $X_1, \ldots, X_n$ are called a *random sample of size n from the population* $f(x)$ if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \ldots, X_n$ are called *independent and identically distributed random variables with pdf or pmf* $f(x)$. This is commonly abbreviated to iid random variables.

$X_1, \ldots, X_n$ random sample and are independent and identically distributed (iid Random Variables)

Since iid

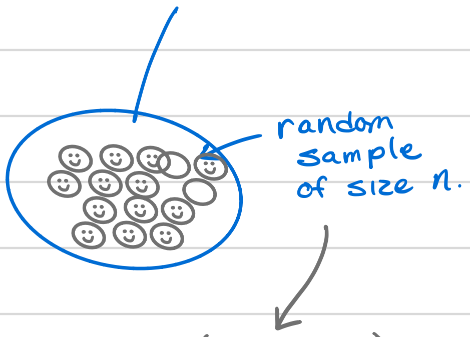
$$f(x_1, \ldots, x_n) = f(x_1) f(x_2) \cdot \ldots \cdot f(x_n) = \prod_{i=1}^{n} f(x_i)$$

If pdf is ft'n of parameter (θ)

$$f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)$$

Later (Chapter 7) we'll estimate θ from $X_1, \ldots, X_n$.

§ 5.2 Sums of RVs from a random sample



$Y = T(x_1, \ldots, x_n)$

Definition 5.2.1 Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \ldots, X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a *statistic*. The probability distribution of a statistic Y is called the *sampling distribution of* Y.

Statistic: function of data not parameter, θ.

$\bar{X} = \Sigma X_i / n$

Definition 5.2.2 The *sample mean* is the arithmetic average of the values in a random sample. It is usually denoted by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

$S^2 = \dfrac{\Sigma (X_i - \bar{X})^2}{(n-1)}$

Definition 5.2.3 The *sample variance* is the statistic defined by

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

The *sample standard deviation* is the statistic defined by $S = \sqrt{S^2}$.

least squares

$(n-1) s^2 = \Sigma x_i^2 - n \bar{x}^2$

Theorem 5.2.4 *Let $x_1, \ldots, x_n$ be any numbers and $\bar{x} = (x_1 + \cdots + x_n)/n$. Then*

a. $\min_a \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$, ← least squares

b. $(n-1) s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. ← simplifies calculation

$X_1, \dots, X_n$ iid

$E[\Sigma g(X_i)] = n E[g(X_1)]$

$Var[\Sigma g(X_i)] = n Var[g(X_1)]$

Lemma 5.2.5 *Let $X_1, \dots, X_n$ be a random sample from a population and let $g(x)$ be a function such that $\mathrm{E}g(X_1)$ and $\mathrm{Var}\, g(X_1)$ exist. Then*

iid

$$\mathrm{E}\left(\sum_{i=1}^{n} g(X_i)\right) = n\left(\mathrm{E}g(X_1)\right) \tag{5.2.1}$$

← true for identically distribute does not require independe

and

$$\mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\left(\mathrm{Var}\, g(X_1)\right). \tag{5.2.2}$$

← true for iid

$X_1, \dots X_n$ iid

$E[X] = \mu$ $Var[X] = \sigma^2$

- $E[\bar{X}] = \mu$
- $Var[\bar{X}] = \sigma^2/n$
- $E[S^2] = \sigma^2$

Theorem 5.2.6 *Let $X_1, \dots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then*

a. $\mathrm{E}\bar{X} = \mu$, ← unbiased estimator of μ.

b. $\mathrm{Var}\, \bar{X} = \dfrac{\sigma^2}{n}$,

c. $\mathrm{E}S^2 = \sigma^2$. ← unbiased estimator of σ^2.

---

mgf of $\bar{X}$

$M_{\bar{X}}(t) = [M_X(t/n)]^n$

Theorem 5.2.7 *Let $X_1, \dots, X_n$ be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is*

$$M_{\bar{X}}(t) = [M_X(t/n)]^n.$$

↑ useful if we recognize this mgf.

---

dist'n sums

- useful if mgf doesn't exist.

Theorem 5.2.9 *If X and Y are independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w)\, dw. \tag{5.2.3}$$

---

exponential family

Statistic

$\sum_{j=1}^{n} t_1(X_j), \dots, \sum_{j=1}^{n} t_k(X_j)$

is exponential family

Sampling from Exponential Family

Theorem 5.2.11 *Suppose $X_1, \dots, X_n$ is a random sample from a pdf or pmf $f(x|\theta)$, where*

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$$

is a member of an exponential family. Define statistics $T_1, \dots, T_k$ by (or $U_1, \dots U_n$)

$$U_i = T_i(X_1, \dots, X_n) = \sum_{j=1}^{n} t_i(X_j), \quad i = 1, \dots, k.$$

If the set $\{(w_1(\theta), w_2(\theta), \dots, w_k(\theta)), \theta \in \Theta\}$ contains an open subset of $\Re^k$, then the distribution of $(T_1, \dots, T_k)$ is an exponential family of the form

$$f_T(u_1, \dots, u_k|\theta) = H(u_1, \dots, u_k)[c(\theta)]^n \exp\left(\sum_{i=1}^{k} w_i(\theta)u_i\right). \tag{5.2.6}$$

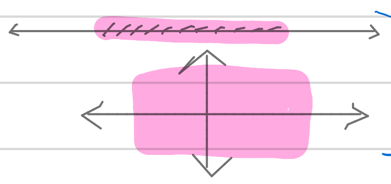
curved exponential

not open rectangle

$w_1(\theta) \dots w_k(\theta)$ $\theta \in \Theta$ > open subset on $\mathbb{R}^k$

$\mathbb{R}^1$ $w_1(\theta)$ subset (not point) of $\mathbb{R}^1$

$\mathbb{R}^2$ $w_1(\theta)$ subset (not line) of $\mathbb{R}^2$



exclude curved exponential

$k > \dim(\Theta)$

§5.3 Sampling from Normal Dist'n

$X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$

a) $\bar{X} \perp S^2$

b) $\bar{X} \sim N(\mu, \sigma^2/n)$

c) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

Theorem 5.3.1 *Let $X_1, \dots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ distribution, and let $\bar{X} = (1/n)\sum_{i=1}^n X_i$ and $S^2 = [1/(n-1)]\sum_{i=1}^n (X_i - \bar{X})^2$. Then*

a. *$\bar{X}$ and S^2 are independent random variables.*

b. *$\bar{X}$ has a $\mathrm{n}(\mu, \sigma^2/n)$ distribution,* proof example 5.2.8 (Notes pg 5).

c. *$(n-1)S^2/\sigma^2$ has a chi squared distribution with $n-1$ degrees of freedom.*

$Z \sim N(0,1)$

a) $Z^2 \sim \chi^2_1$

$X_i \sim \chi^2_{p_i}$

b) $\sum X_i \sim \chi^2_{\sum p_i}$

Lemma 5.3.2 (Facts about chi squared random variables) *We use the notation χ^2_p to denote a chi squared random variable with p degrees of freedom.*

a. *If Z is a $\mathrm{n}(0,1)$ random variable, then $Z^2 \sim \chi^2_1$; that is, the square of a standard normal random variable is a chi squared random variable.*

b. *If $X_1, \dots, X_n$ are independent and $X_i \sim \chi^2_{p_i}$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1 + \cdots + p_n}$; that is, independent chi squared variables add to a chi squared variable, and the degrees of freedom also add.*

$X_j \sim N(\mu_j, \sigma_j^2)$

$U_i = \sum a_{ij} X_j \quad i = 1 \dots k$

$V_r = \sum b_{rj} X_j \quad r = 1 \dots m$

$U_i \perp V_r \overset{\text{iff}}{\Longleftrightarrow} \mathrm{Cov}(U_i, V_r) = 0$

$(U_1, \dots, U_k) \perp (V_1 \dots V_m)$

$\overset{\text{iff}}{\Longleftrightarrow}$

$U_i \perp V_r \quad \forall i, r$

Lemma 5.3.3 *Let $X_j \sim \mathrm{n}(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$, independent. For constants a_{ij} and b_{rj} ($j = 1, \dots, n$; $i = 1, \dots, k$; $r = 1, \dots, m$), where $k + m \le n$, define*

$$U_i = \sum_{j=1}^n a_{ij} X_j, \quad i = 1, \dots, k,$$

$$V_r = \sum_{j=1}^n b_{rj} X_j, \quad r = 1, \dots, m.$$

a. *The random variables U_i and V_r are independent if and only if $\mathrm{Cov}(U_i, V_r) = 0$. Furthermore, $\mathrm{Cov}(U_i, V_r) = \sum_{j=1}^n a_{ij} b_{rj} \sigma_j^2$.*

b. *The random vectors $(U_1, \dots, U_k)$ and $(V_1, \dots, V_m)$ are independent if and only if U_i is independent of V_r for all pairs i, r ($i = 1, \dots, k$; $r = 1, \dots, m$).*

$X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$

$\sqrt{n}\dfrac{(\bar{X} - \mu)}{S} \sim t_{n-1}$

Definition 5.3.4 Let $X_1, \dots, X_n$ be a random sample from a $\mathrm{n}(\mu, \sigma^2)$ distribution. The quantity $(\bar{X} - \mu)/(S/\sqrt{n})$ has *Student's t distribution with $n-1$ degrees of freedom*. Equivalently, a random variable T has Student's t distribution with p degrees of freedom, and we write $T \sim t_p$ if it has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}, \quad -\infty < t < \infty. \tag{5.3.6}$$

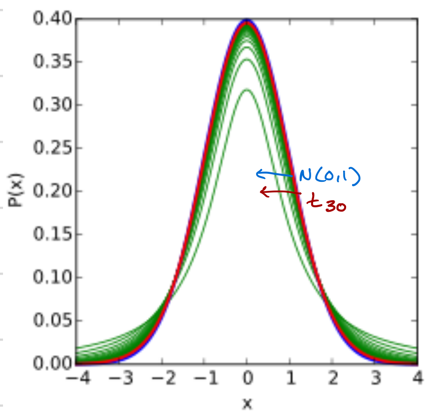
$X_1, \dots X_n$ iid $N(\mu_x, \sigma_x^2)$

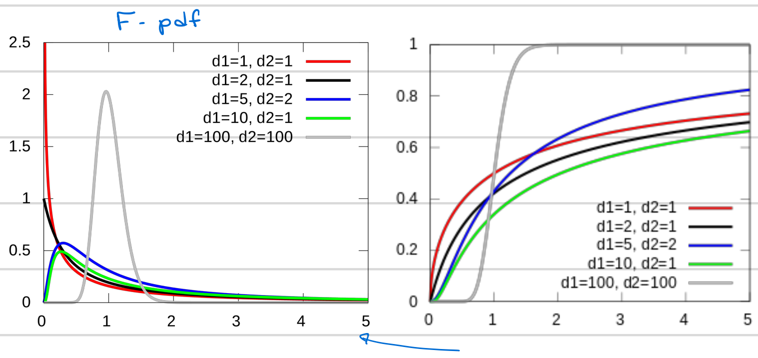
$Y_1, \dots, Y_m$ iid $N(\mu_y, \sigma_y^2)$

$\dfrac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F_{n-1, m-1}$

Definition 5.3.6 Let $X_1, \dots, X_n$ be a random sample from a $\mathrm{n}(\mu_X, \sigma_X^2)$ population, and let $Y_1, \dots, Y_m$ be a random sample from an independent $\mathrm{n}(\mu_Y, \sigma_Y^2)$ population. The random variable $F = (S_X^2/\sigma_X^2)/(S_Y^2/\sigma_Y^2)$ has *Snedecor's F distribution with $n-1$ and $m-1$ degrees of freedom.* Equivalently, the random variable F has the F distribution with p and q degrees of freedom if it has pdf

$$f_F(x) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1 + (p/q)x]^{(p+q)/2}}, \quad 0 < x < \infty. \tag{5.3.9}$$





Order Stats dist'ns

cdf relationship to

bin (n, F_X)

§ 5.4 Order Statistics   Things that make you say "hmmm".

- Assume you take a random sample of $n=4$ from a $U(0,1)$ dist'n.
  X_1, X_2, X_3, X_4, iid $U(0,1)$

$f_X(x) = 1\, I_{[0,1]}(x)$ and $F_X(x) = \int_0^x dt = x \quad 0 \le x \le 1$

- Find $P(\text{all values} \ge \frac{1}{2})$

  Method: Figure out $Pr(\text{success})$ for each X_i
  - Determine how many successes are needed
  - $P(\text{success}) = \frac{1}{2} = 1 - F_X(\frac{1}{2})$
  - Need all X_i ($n=4$) success : $\left(\frac{1}{2}\right)^4$

  $F_X(\frac{1}{2}) = \frac{1}{2}$
  $1 - F_X(\frac{1}{2}) = \frac{1}{2}$
  $= Pr(\text{success})$

- Find $P(\text{minimum (smallest } X) \ge \frac{1}{2})$

  Same as all values $\ge \frac{1}{2}$

- Find $P(\text{smallest value} < \frac{1}{4})$

  $P(\text{success}) = \frac{1}{4}$
  1 or more successes ($< \frac{1}{4}$)
  $\sum_{i=1}^{4} \binom{4}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{4-i}$

  $F_X(\frac{1}{4}) = \frac{1}{4}$
  $= Pr(\text{success})$

- Find $P(\text{largest value} \ge \frac{3}{4})$

  $P(\text{success}) = \frac{1}{4}$
  1 or more successes ($\ge \frac{1}{4}$)
  $\sum_{i=1}^{4} \binom{4}{i} \left(\frac{1}{4}\right)^i \left(\frac{3}{4}\right)^{4-i}$

  $F_X(\frac{3}{4}) = \frac{3}{4}$
  $1 - F_X(\frac{3}{4}) = \frac{1}{4}$
  $= Pr(\text{success})$

- Find $P(\text{at least half of values} \ge \frac{1}{2}) = Pr(\text{at least 2 values} \ge \frac{1}{2} \mid n=4)$

  $P(\text{success}) = \frac{1}{2}$
  Need 2 or more successes
  $\sum_{i=2}^{4} \binom{4}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{4-i}$

  $F_X(\frac{1}{2}) = \frac{1}{2}$
  $1 - F_X(\frac{1}{2}) = \frac{1}{2}$
  $= Pr(\text{success})$

- Find $Prob(3^{rd} \text{ smallest} \ge \frac{1}{2})$

  $X_{(1)}$ = smallest X_i
  $X_{(2)}$ = second smallest X_i
  $X_{(3)}$ = 3^{rd} smallest X_i = 2^{nd} largest
  $X_{(4)}$ = Largest X_i

  success: ≥ 2 values $\ge \frac{1}{2} \mid n=4$ $(2,3,4 \ge \frac{1}{2})$
  or failures: ≤ 2 values $< \frac{1}{2} \mid n=4$ $(2,1,0 < \frac{1}{2})$

  $\sum_{i=2}^{4} \binom{4}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{4-i}$ or $\sum_{i=0}^{2} \binom{4}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{4-i}$

Order Stats

Definition 5.4.1 The *order statistics* of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, \ldots, X_{(n)}$.

The order statistics are random variables that satisfy $X_{(1)} \leq \cdots \leq X_{(n)}$. In particular,

$$X_{(1)} = \min_{1 \leq i \leq n} X_i,$$
$$X_{(2)} = \text{second smallest } X_i,$$
$$\vdots$$
$$X_{(n)} = \max_{1 \leq i \leq n} X_i.$$

percentiles

Definition 5.4.2 The notation $\{b\}$, when appearing in a subscript, is defined to be *the number b rounded to the nearest integer in the usual way.* More precisely, if i is an integer and $i - .5 \leq b < i + .5$, then $\{b\} = i$.

Discrete pmf

Theorem 5.4.3 *Let $X_1, \ldots, X_n$ be a random sample from a discrete distribution with pmf $f_X(x_i) = p_i$, where $x_1 < x_2 < \cdots$ are the possible values of X in ascending order. Define*

$$P_0 = 0$$
$$P_1 = p_1$$
$$P_2 = p_1 + p_2$$
$$\vdots$$
$$P_i = p_1 + p_2 + \cdots + p_i$$
$$\vdots$$

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics from the sample. Then

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \tag{5.4.2}$$

$\left\{ \begin{array}{l} j \text{ or more successes} \\ P(\text{success}) = P_i = P(X_\ell \leq x_i) \quad \ell = 1, \ldots, n \text{ (iid)} \end{array} \right.$

and

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}\right]. \tag{5.4.3}$$

continuous pdf

Theorem 5.4.4 *Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}. \tag{5.4.4}$$

continuous

$f_{X_{(i)}, X_{(j)}}(u, v)$

Theorem 5.4.6 *Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \leq i < j \leq n$, is*

$$f_{X_{(i)}, X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} \times [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j} \tag{5.4.7}$$

for $-\infty < u < v < \infty$.

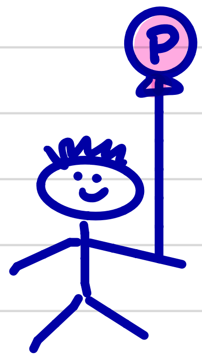
$f_{X_{(1)} \ldots X_{(n)}}(x_1, \ldots, x_n)$

$$f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! f_X(x_1) \cdot \cdots \cdot f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty \\ 0 & \text{otherwise.} \end{cases}$$

$\hookrightarrow$ $n!$ permutations of $\underline{x}$ give same order stats

Where to next?

§ 5.5 Convergence Concepts



$\xrightarrow{P}$

Definition 5.5.1 A sequence of random variables, $X_1, X_2, \ldots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \underbrace{P(|X_n - X| \geq \epsilon)}_{\text{probability}} = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

start here

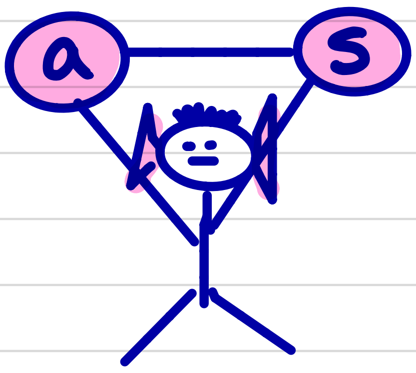
WLLN

$\bar{X}_n \xrightarrow{P} E[X]$

Theorem 5.5.2 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.



$\xrightarrow{a.s.}$

Definition 5.5.6 A sequence of random variables, $X_1, X_2, \ldots$, *converges almost surely* to a random variable X if, for every $\epsilon > 0$,

$$P(\underbrace{\lim_{n\to\infty} |X_n - X| < \epsilon}_{\text{pointwise convergence}}) = 1.$$

start here

SLLN

$\bar{X}_n \xrightarrow{a.s.} E[X]$

Theorem 5.5.9 (Strong Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$, and define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges almost surely to μ.

$\xrightarrow{\mathcal{L}}$

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if

convergence in Law

$$\lim_{n\to\infty} F_{X_n}(x) = \underline{F_X(x)}$$

at all points x where $F_X(x)$ is continuous.

start here (or mgf).

CLT

$f_{\bar{X}} \xrightarrow{\mathcal{L}} N(\mu, \sigma^2/n)$

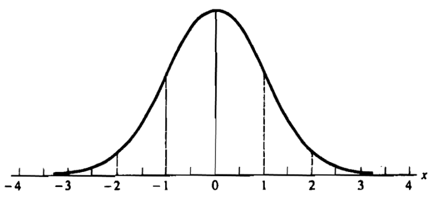


Figure 3.3.1. *Standard normal density*

Theorem 5.5.14 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

$$\therefore \quad \frac{\sqrt{n}\,(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

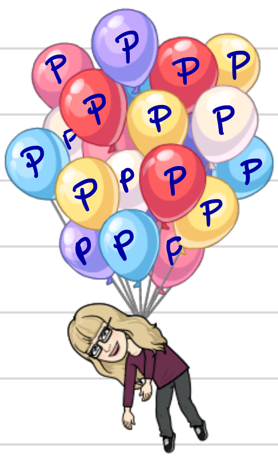
proof: homework-2

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathrm{E}X_i = \mu$ and $0 < \mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

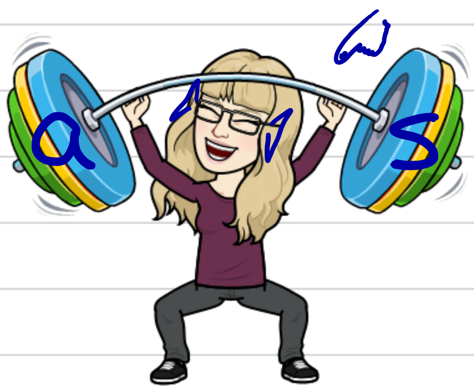
$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

proof uses characteristic ft'n $E[e^{itX}]$ $(i^2 = -1)$



$\xrightarrow{P}$



$\xrightarrow{a.s.}$

$\xrightarrow{\mathcal{L}}$