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

$\left.\right\}$ 1st Semester (chapters 1–5)

$\left.\right\}$ 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

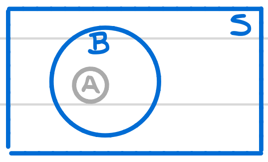
Definition 1.1.2 An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

& Event Operations

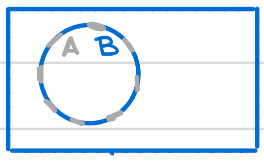
A = Event, subset of S.

Event A occurs if outcome of experiment is in set A.

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

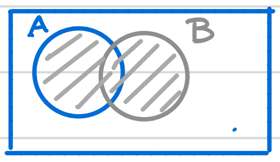


- Equality $A = B \iff A \subset B$ and $B \subset A$

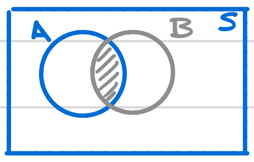


Operations

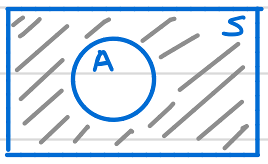
- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$



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



- Complementation: $A^c = \{\underbrace{x}_{\text{all elements of } x} \underbrace{:}_{\text{such that}} \underbrace{x \notin A}_{x \text{ not in } A}\}$



Properties of Events

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

a. Commutativity $A \cup B = B \cup A$,
$A \cap B = B \cap A$;

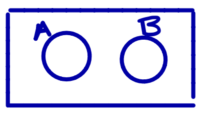
b. Associativity $A \cup (B \cup C) = (A \cup B) \cup C$,
$A \cap (B \cap C) = (A \cap B) \cap C$;

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

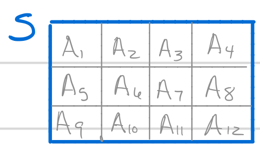
Disjoint

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)

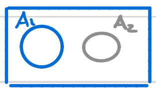
Definition 1.2.4 Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function P with domain $\mathcal{B}$ that satisfies

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

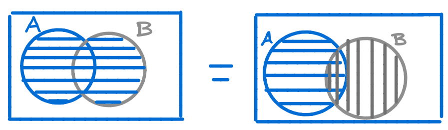
Theorem 1.2.9 *If P is a probability function and A and B are any sets in $\mathcal{B}$, then*

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

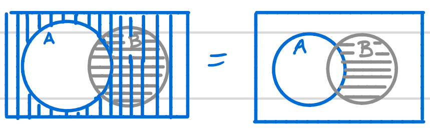
b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

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

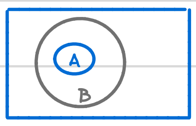
b)



a)



c)



§ 1.2 cont.

Probability functions

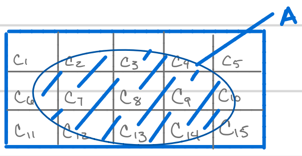
A, Partition $C_1, C_2, \ldots$

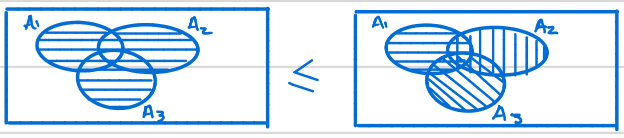
Union

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. $P(\cup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} P(A_i)$ *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	$\binom{n}{r}$	$\binom{n+r-1}{r}$

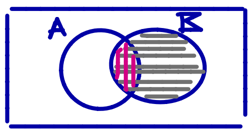
most important for us. → Unordered

§1.3 Conditional Prob + Independence

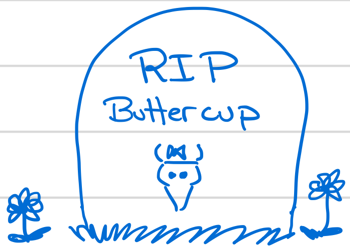
Conditional Probability

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

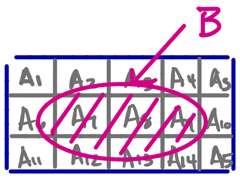
(1.3.1) 
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$



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 $i = 1, 2, \ldots$,*



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

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

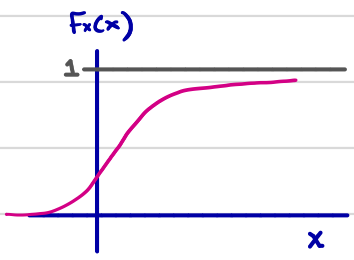
Definition 1.3.12 A collection of events $A_1, \ldots, A_n$ are *mutually independent* if for any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have

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

§ 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

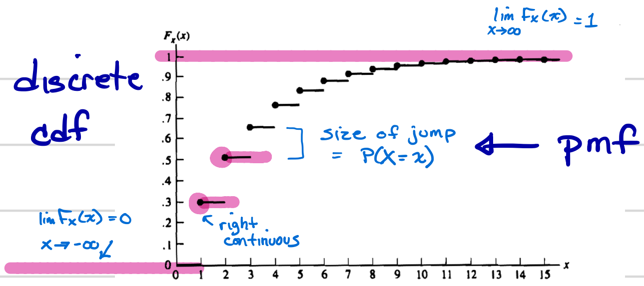


Figure 1.5.2. *Geometric cdf, $p = .3$*

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

§1.6 cont.

Continuous pdf

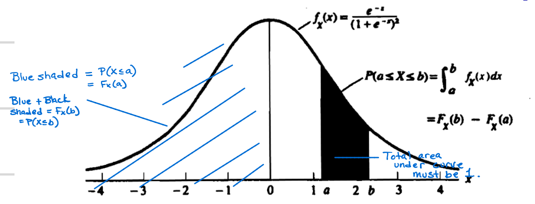


Figure 1.6.1. *Area under logistic curve*

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

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

pmf: $\Sigma f_X(x) = 1$ 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. $f_X(x) \geq 0$ *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 $y \in \mathcal{Y}$.*

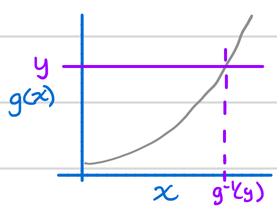
b. *If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.*

increasing g

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

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

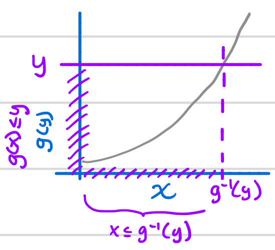
$g(x)$ is increasing Find $F_Y = P(Y \le y)$



$\{x \in \mathcal{X} : g(x) \le y\}$

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

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



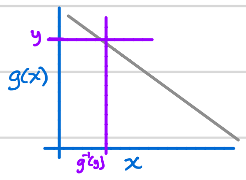
$$F_Y(y) = \int_{\{x \in \mathcal{X}:\, x \le g^{-1}(y)\}} f_X(x)\,dx = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,dx = F_X(g^{-1}(y)).$$

decreasing g

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

$y \in \mathcal{Y}$

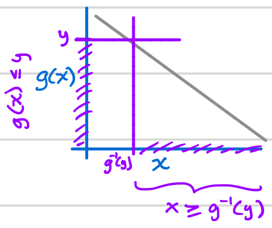
$g(x)$ is decreasing Find $F_Y = P(Y \le y)$



$\{x \in \mathcal{X} : g(x) \le y\}$

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

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



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

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

Moments & Central Moments

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

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

The *n*th *central moment of* X, μ_n, is

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

where $\mu = \mu_1' = \mathrm{E}\, X$.

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 $\rightarrow F_X(x)$

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

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

§ 2.3 cont.

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

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

$\Rightarrow$

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

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.

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 $M_{X_i}(t)$ and cdf $F_{X_i}(t)$

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:

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

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	$P(X = x\|N) = \frac{1}{N}$; $x = 1, 2, \ldots, N$; $N = 1, 2, \ldots$
mean and variance	$\mathrm{E}X = \frac{N+1}{2}$, $\mathrm{Var}\,X = \frac{(N+1)(N-1)}{12}$
mgf	$M_X(t) = \frac{1}{N}\sum_{i=1}^{N} e^{it}$

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

Bernoulli (p)

Bernoulli(p)

pmf	$P(X = x\|p) = p^x(1-p)^{1-x}$; $x = 0, 1$; $0 \le p \le 1$
mean and variance	$\mathrm{E}X = p$, $\mathrm{Var}\,X = p(1-p)$
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	$P(X = x\|n, p) = \binom{n}{x} p^x (1-p)^{n-x}$; $x = 0, 1, 2, \ldots, n$; $0 \le p \le 1$
mean and variance	$\mathrm{E}X = np$, $\mathrm{Var}\,X = np(1-p)$
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.

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

Binomial Theorem

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

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

If $x = y = 1$

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

Poisson (λ)

Poisson(λ)

pmf	$P(X = x\|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, \ldots$; $0 \le \lambda < \infty$
mean and variance	$\mathrm{E}X = \lambda$, $\mathrm{Var}\,X = \lambda$
mgf	$M_X(t) = e^{\lambda(e^t - 1)}$

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

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

Negative Binomial (r, p)

(# trials to r^{th} success)

Negative binomial(r, p)

pmf	$P(X = x\|r, p) = \binom{r+x-1}{x} p^r (1-p)^x$; $x = 0, 1, \ldots$; $0 \le p \le 1$
mean and variance	$\mathrm{E}X = \frac{r(1-p)}{p}$, $\mathrm{Var}\,X = \frac{r(1-p)}{p^2}$
mgf	$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$, $t < -\log(1-p)$
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.)

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

MS Theory I

Review/11

§ 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 $\quad P(X=x|p)=p(1-p)^{x-1};\quad x=1,2,\dots;\quad 0\le p\le 1$

mean and variance $\quad EX=\frac{1}{p},\quad \text{Var}\,X=\frac{1-p}{p^2}$

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

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

notes $\quad Y=X-1$ is negative binomial$(1,p)$. The distribution is *memoryless*: $P(X>s|X>t)=P(X>s-t)$.

§ 3.3 Continuous Dist'ns

Uniform (a, b)

Uniform(a,b)

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

pdf $\quad f(x|a,b)=\frac{1}{b-a},\quad a\le x\le b$

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

mean and variance $\quad EX=\frac{b+a}{2},\quad \text{Var}\,X=\frac{(b-a)^2}{12}$

mgf $\quad M_X(t)=\frac{e^{bt}-e^{at}}{(b-a)t}$

notes $\quad$ If $a=0$ and $b=1$, this is a special case of the beta $(\alpha=\beta=1)$.

Gamma (α, β)

Gamma Dist'n

Gamma(α,β)

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

pdf $\quad f(x|\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta},\quad 0\le x<\infty,\quad \alpha,\beta>0$

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

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

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

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

mean and variance $\quad EX=\alpha\beta,\quad \text{Var}\,X=\alpha\beta^2$

mgf $\quad M_X(t)=\left(\frac{1}{1-\beta t}\right)^\alpha,\quad t<\frac{1}{\beta}$

notes $\quad$ 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).

Normal (μ,σ^2)

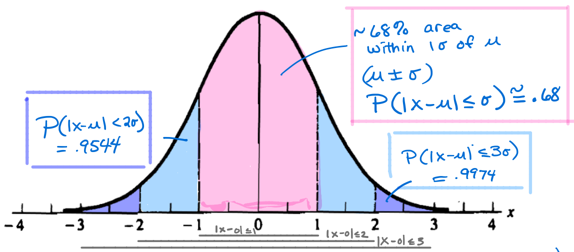


Figure 3.3.1. *Standard normal density* $(\mu=0,\sigma=1)$

Normal(μ,σ^2)

pdf $\quad f(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)},\quad -\infty<x<\infty,\quad -\infty<\mu<\infty,\quad \sigma>0$

mean and variance $\quad EX=\mu,\quad \text{Var}\,X=\sigma^2$

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

mgf $\quad M_X(t)=e^{\mu t+\sigma^2t^2/2}$

notes $\quad$ Sometimes called the *Gaussian* distribution.

Beta (α, β)

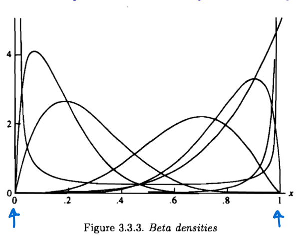


Figure 3.3.3. Beta densities

strictly increasing $\alpha>1,\beta=1$; U shaped $(\alpha<1,\beta<1)$

strictly decreasing $\alpha=1,\beta>1$; Unimodal $(\alpha>1,\beta>1)$

symmetric $(\alpha=\beta)$

Beta(α,β)

bounded support — used to model proportions

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

pdf $\quad f(x|\alpha,\beta)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad 0\le x\le 1,\quad \alpha>0,\quad \beta>0$

mean and variance $\quad EX=\frac{\alpha}{\alpha+\beta},\quad \text{Var}\,X=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

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

mgf $\quad M_X(t)=1+\sum_{k=1}^\infty\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$

notes $\quad$ 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.

Cauchy (θ, σ)

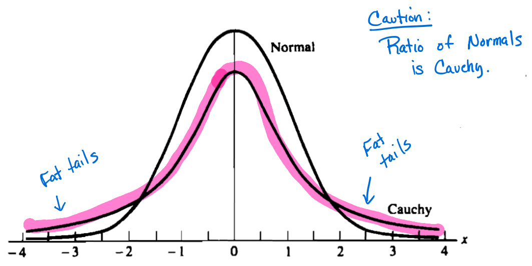


Figure 3.3.5. *Standard normal density and Cauchy density*

Cauchy(θ, σ)

pdf	$f(x\mid\theta,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}, \quad -\infty < x < \infty; \quad -\infty < \theta < \infty, \quad \sigma > 0$
mean and variance	do not exist ← $\quad I_{(-\infty,\infty)}^{(x)}$
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.

Lognormal (μ, σ^2)

Lognormal(μ, σ^2) $\quad I_{[0,\infty)}^{(x)}$

pdf	$f(x\mid\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log x-\mu)^2/(2\sigma^2)}}{x}, \quad 0 \le x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$
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) $\quad I_{(-\infty,\infty)}^{(x)}$

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

§ 3.4

♡

Exponential Families

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

- $\underline{\theta}$ represents parameter
- $h(x) \ge 0$ ← 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.
- $c(\underline{\theta}) \ge 0$
- $w_1(\theta), \ldots, w_k(\theta)$ ft'ns of θ - don't depend on x

Take $f(x\mid\underline{\theta})$ & 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*

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

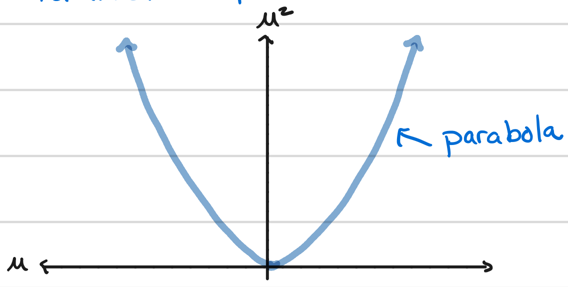
$$\mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j^2} \log c(\boldsymbol{\theta}) - \mathrm{E}\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X)\right). \tag{3.4.5}$$

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 $N(\mu, \mu^2)$

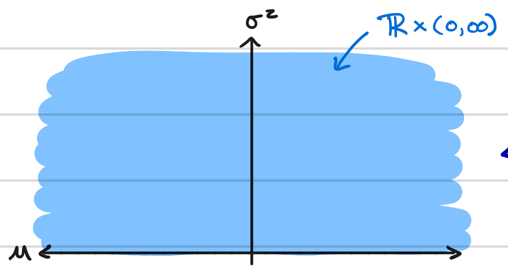


Curved Exponential

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

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

§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

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

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.

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

standard pdf ($\mu=0$)

location family ($\mu=2$) (shaped unchanged).

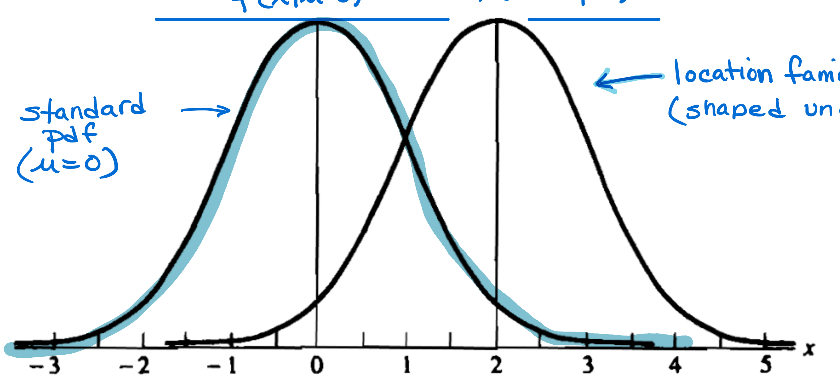


Figure 3.5.1. *Two members of the same location family: means at 0 and 2*

(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.

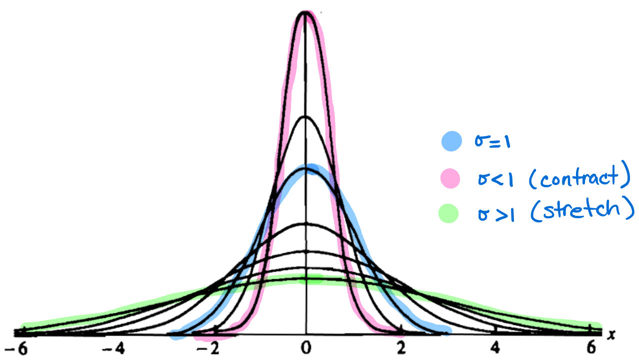


Figure 3.5.3. *Members of the same scale family*

§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.*

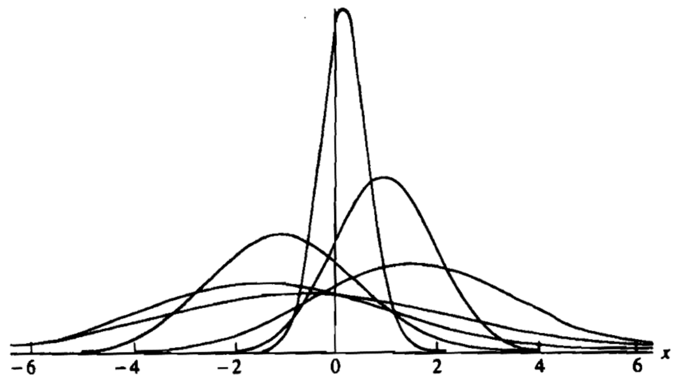


Figure 3.5.4. *Members of the same location–scale family*

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

$\Rightarrow$ 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*

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

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

§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 $r > 0$,*

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

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*

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

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

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

different joint pmfs may have same marginals.

Conditional pmf

$f(y|x) =$

$P(Y=y | X=x)$

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

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 $A \subset \Re^2$,

$$P((X,Y) \in A) = \underbrace{\int_A \int}_{\text{double integrals, integrate over all } (x,y) \in A} f(x,y)\, dx\, dy.$$

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

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

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

1)

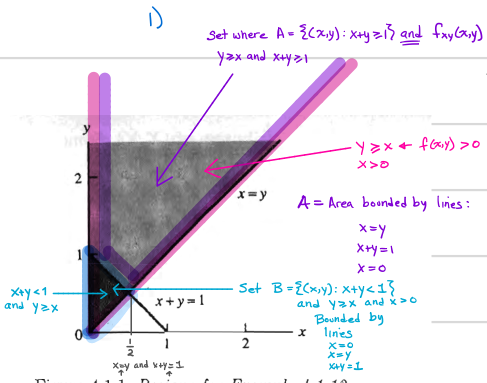


Figure 4.1.1. *Regions for Example 4.1.12*

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 : $X \perp Y$

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

$\rightarrow X \perp Y$

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

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 $y \in \Re$,

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

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.*