1 Random Variables and Distributions

Random variables are the main link between probability and statistics. In statistics we observe numbers (or data) as the result of an experiment, and a random variable links the numbers to the probability structure of the experiment.

Definition of random variable

Let the sample space of an experiment be $\mathcal{S}$. A *random variable* is a mapping, or function, from $\mathcal{S}$ to the real number line.

We will use capital letters late in the alphabet, such as X, Y and Z, to denote random variables. If X is a random variable, then X associates with each $s \in \mathcal{S}$ a real number $X(s)$.

$X(s)$ is a function of the outcome.

The notation X and $X(s)$ parallels the notation we often see in math classes, where

- f is used to denote a function, and
- $f(x)$ is the value of f at argument x.

So,

$$X \Longleftrightarrow f \quad \text{and} \quad X(s) \Longleftrightarrow f(x).$$

A value of X will often be denoted x, i.e., $X(s) = x$.

Example 12 Suppose that a coin is tossed four times. Then there are $2^4 = 16$ possible sequences of tosses (such as $HTHT$). Let X be the number of heads until the first tail. For our example sequence, the mapping is $X(HTHT) = 1$. We can form a table of the mapping for all possible sequences:

$$X(TTTT) = 0 \quad X(TTTH) = 0 \quad X(TTHT) = 0 \quad X(TTHH) = 0$$

$$X(THTT) = 0 \quad X(THTH) = 0 \quad X(THHT) = 0 \quad X(THHH) = 0$$

$$X(HTTT) = 1 \quad X(HTTH) = 1 \quad X(HTHT) = 1 \quad X(HTHH) = 1$$

$$X(HHTT) = 2 \quad X(HHTH) = 2 \quad X(HHHT) = 3 \quad X(HHHH) = 4$$

NOTE: Just because $|\{X=0\}|$ is $8/16$ $\not\Rightarrow$ $P(X=0) = 1/2$. That would only be the case if each outcome was equally likely.

$\rightarrow A \subset \mathbb{R}$

Let X be a random variable defined on a sample space $\mathcal{S}$, and let A be some subset of the real numbers. We then define $P(X \in A)$ by

$$P(X \in A) = P\left(\{s \in \mathcal{S} : X(s) \in A\}\right).$$

The probabilities so defined by all relevant subsets A is called the *probability distribution* of X.

We use $P(X = x)$ as a shorthand for $P(X \in \{x\})$.

Example 13 In the experiment of Example 12, we assume that the coin is a fair coin. Then each outcome has probability $\frac{1}{16}$.

For example, we have

$$P(X = 1) = P(\{HTTT, HTTH, HTHT, HTHH\}) = \frac{4}{16} = 0.25.$$

Similar reasoning yields:

x	0	1	2	3	4
$P(X = x)$	0.5	0.25	0.125	0.0625	0.0625

Any other probability of interest concerning the random variable X may be determined from these probabilities.

There are two main types of random variables: *discrete* and *continuous*.

Remember, X is a mapping from $\mathcal{S}$ to some subset of the real numbers.

The domain of the mapping is $\mathcal{S}$, and we'll call the range R_X.

If R_X is countable, then X is a *discrete* random variable. If R_X is not countable, then X is a *continuous* random variable. When X is continuous, R_X is usually an interval or a union of disjoint intervals.

When $\mathcal{S}$ is countable, then X *must* be discrete, while if $\mathcal{S}$ is uncountable, then X can be either discrete or continuous.

2 Discrete Random Variables

The *probability function (or probability mass function)* of a discrete random variable is a function p_X defined by

$$p_X(x) = P(X = x) \quad \text{for every real number } x.$$

Write the range of X as $R_X = \{x_1, x_2, \ldots\}$. Then

$$\sum_{i=1}^{\infty} p_X(x_i) = 1.$$

For any subset A of real numbers, we may express $P(X \in A)$ as

$$P(X \in A) = \sum_{x \in A \cap R_X} p_X(x).$$

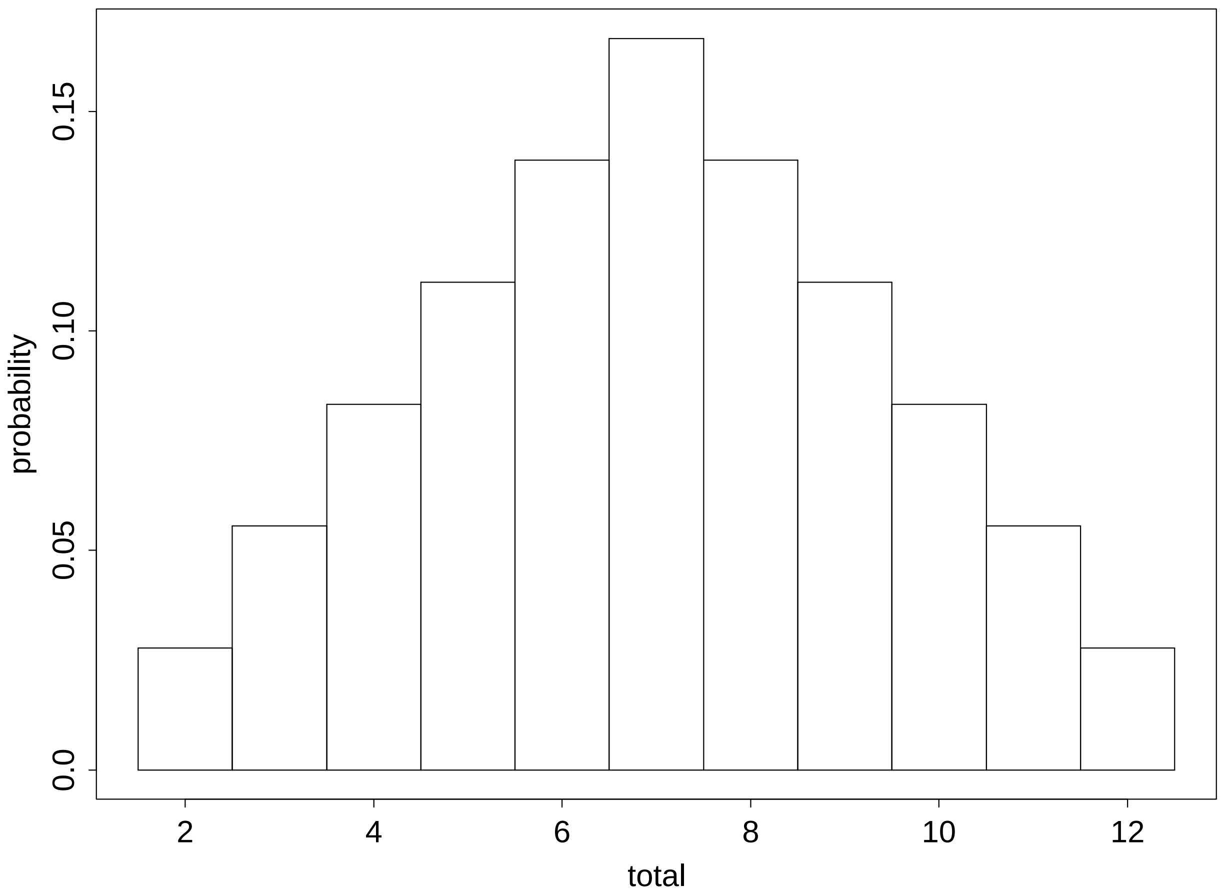
Example 4 revisited Consider again our dice experiment. If the dice are balanced, then the probability of each of the 36 different outcomes is the same. In this case, for each (i, j)

$$P(\{(i,j)\}) = \frac{1}{36}. \qquad \text{Why?}$$

We define the random variable X to be total on the two dice. We can compute the probability mass function for X from the probabilities on the original sample space. For example,

$$p_X(6) = P(X = 6) = P(\{(1,5),(2,4),(3,3),(4,2),(5,1)\}) = \frac{5}{36}.$$

We can likewise find the probability of any other possible total. A graph of the probability mass function is given below.


0.0
0.05
0.10
0.15
probability
2
4
6
8
10
12
total

Example 4 continued

Suppose now we let $A = \{x : x \leq 4\}$. We are interested in $P(X \in A)$. We note that

$P(X \le 4) = 1/6$

$$R_X = \{2, 3, 4, \ldots, 11, 12\} \text{ and } A \cap R_X = \{2, 3, 4\}.$$

We have two ways of finding $P(X \in A)$.

1. We can use the original probability space:

$$P(X \in A) = P(\{(1,1), (1,2), (2,1), (1,3), (2,2), (3,1)\}) = \frac{6}{36} = \frac{1}{6}$$

2. We can use the probability mass function of X:

$$P(X \in A) = \sum_{x=2}^{4} p_X(x) = p_X(2) + p_X(3) + p_X(4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}$$

A function that is often useful in defining and proving properties of a random variable is an *indicator function*.

The *indicator function* of an event A is defined as

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

Example: **Exercise 2.1.6 (modified a little)** Let $\mathcal{S} = \{1, 2, 3, 4\}$, $X = I_{\{1,2\}}$, $Y = I_{\{2,3\}}$, $Z = I_{\{3,4\}}$, and $W = X - 2Y + Z$.

$$W(1) = X(1) - 2Y(1) + Z(1) = 1 - 2(0) + 0 = 1$$

$$W(2) = X(2) - 2Y(2) + Z(2) = 1 - 2(1) + 0 = -1$$

$$W(3) = X(3) - 2Y(3) + Z(3) = 0 - 2(1) + 1 = -1$$

$$W(4) = X(4) - 2Y(4) + Z(4) = 0 - 2(0) + 1 = 1$$

We'll discuss several probability mass functions (pmfs) that are important in statistical applications. These are the pmfs for the **Bernoulli distribution**, **discrete uniform distribution**, the **binomial distribution**, the **negative binomial distribution**, the **Poisson distribution**, and the **hypergeometric distribution**.

2.1 Bernoulli distribution

The simplest nontrivial discrete random variable X takes on only two values, 0 and 1. Suppose that $0 < \theta < 1$. The probability mass function (pmf) of X is

$$\begin{aligned} p_X(1) &= \theta \\ p_X(0) &= 1 - \theta \\ p_X(x) &= 0, \qquad \text{otherwise.} \end{aligned}$$

We can also write the pmf as

$$p_X(x) = \begin{cases} \theta^x (1-\theta)^{1-x} & \text{if } x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise.} \end{cases}$$

2.2 Discrete uniform distribution

useful in nonparametric statistics.

The discrete uniform probability mass function p_X is defined for a positive integer k by

$$p_X(x) = \begin{cases} 1/k, & x = 1, 2, \ldots, k \\ 0, & \text{otherwise.} \end{cases}$$

This probability mass function is the distribution of a random variable X that has range $\{1, \ldots, k\}$ and is equally likely to take on any value in this range.

The uniform distribution arises in the analysis of *ranks* in statistics. Suppose that k numbers are drawn randomly from an *infinite* population (to be defined later).

Let R be the rank of the first number drawn among all k numbers. So, $R = 1$ if the first number drawn is the smallest one, $R = 2$ if the first number drawn is the next to the smallest, and so on.

It turns out that R has a discrete uniform distribution in this case.

2.3 Binomial distribution

Binomial experiment

1. Observe a sequence of n trials, where n is fixed in advance.
2. Each trial results in one of two possible outcomes; call them "success" and "failure" (S and F).
3. The trials are independent of each other.
4. The probability of S on any one trial is θ where $0 < \theta < 1$. (Note: θ remains the same from trial to trial.)

Define X to be the number of successes among the n trials of a binomial experiment. Then the probability mass function of X has the following form:

$$P(X=x) = p_X(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & x = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise.} \end{cases}$$

This pmf is called the *binomial pmf*.

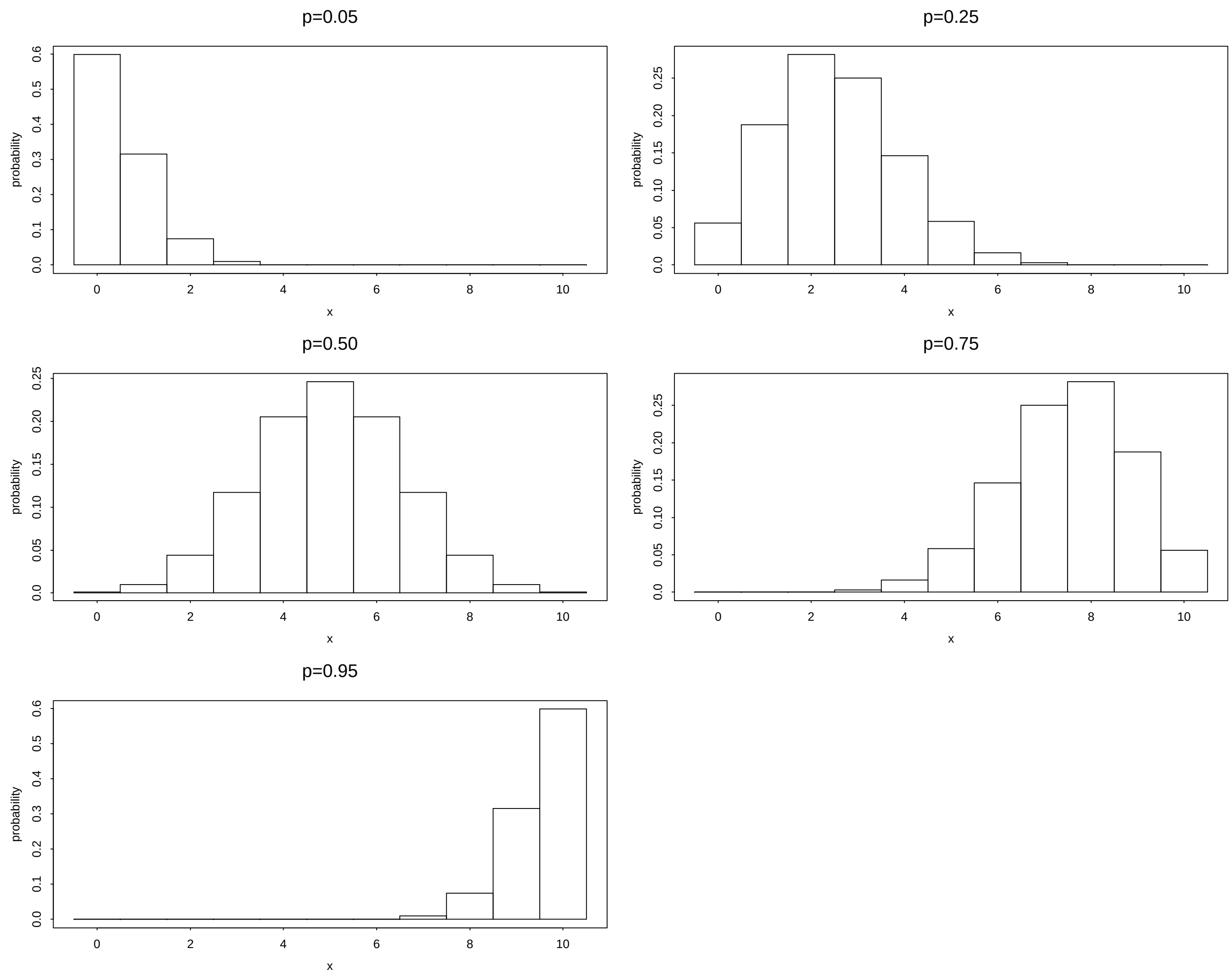
We can use ideas from Chapter 1 to prove that p_X is the pmf of the number of successes X in a binomial experiment.

There is a statistical application for the binomial distribution in sampling from a finite population. Suppose a population consists of N items, M of which are defective and $N - M$ nondefective.

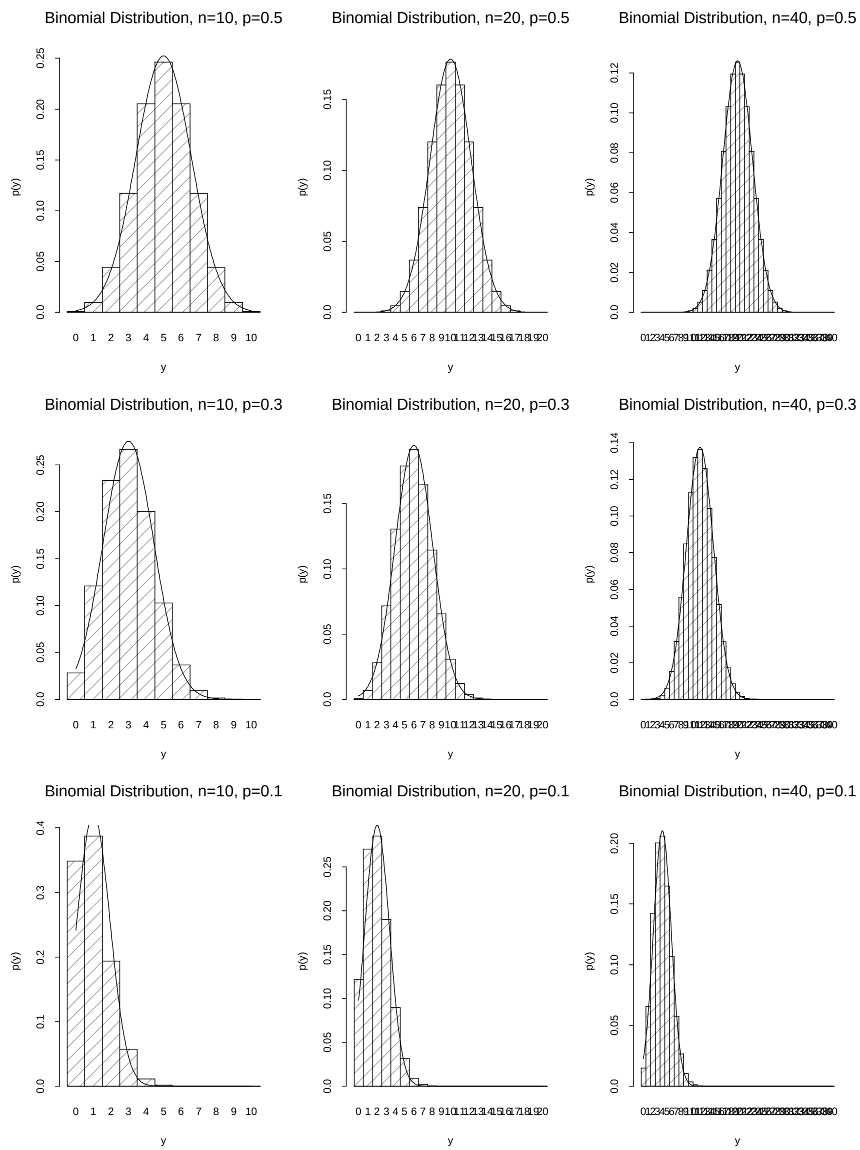
Suppose we randomly select n items from the population with replacement. Let X be the number of defective items among the n selected. Then X has the binomial distribution with $\theta = M/N$.

To argue that the distribution is binomial, just verify the conditions of the binomial experiment in this setting.

Various Binomial Distributions for $n = 10$



Some Binomial Distributions for $n = 10, 20$ and 40



2.4 Negative Binomial Probability Distribution

Consider the following simple experiment:

We flip a coin until we have tossed 5 heads.

Note:

1. The experiment consists of a sequence of independent trials.
2. Each trial is identical and can result in one of two outcomes (S or F).
3. The probability of S equals θ $(0 < \theta < 1)$ for each trial.
4. We continue the experiment until r successes have been observed.

Any experiment meeting all of the above conditions is called a

Negative Binomial Experiment.

If Y is our random variable representing *the number of failures obtained before obtaining r successes*, then Y has a negative binomial distribution:

$$Y \sim \text{Negative Binomial}(r, \theta)$$

r = number of S

θ = probability of S

The probability mass function of a negative binomial rv is given by

$$p_Y(y) = P(Y = y) = \binom{r-1+y}{r-1} \theta^r (1-\theta)^y, \qquad y = 0, 1, 2, \ldots$$

- If $r = 1$ and $X =$ the number of failures until the first success, we have a *geometric* distribution with pmf

$$P(X = x) = p_X(x) = \theta(1-\theta)^x, \qquad x = 0, 1, 2, 3, \ldots$$

- Some books define the negative binomial distribution as *the number of trials it takes to obtain the r^{th} success*. This changes the formulas a little.

2.5 Hypergeometric Distribution

The hypergeometric distribution applies to "sampling without replacement" from a finite population containing two types of items.

We have a population of size N containing M defective items and $N - M$ nondefective items. We randomly select n items ($n \leq N$) *without replacement*.

Define X to be the number of defectives in the sample. Then X has the probability mass function

$$p_X(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = M_1, \ldots, M_2, \\ 0 & \text{otherwise,} \end{cases}$$

where $M_1 = \max(0, n - (N - M))$, and $M_2 = \min(n, M)$.

This is the pmf of the *hypergeometric distribution*.

2.6 Poisson Distribution

Consider these random variables:

- Number of phone calls received per hour by AAA emergency service.
- Number of customers logging onto Amazon Prime in a 5 minute interval.
- Number of trees in an area of forest.
- Number of bacteria in a culture.

A random variable X, the number of successes occurring during a given time interval or in a specified region, is called a *Poisson* random variable. The corresponding distribution of

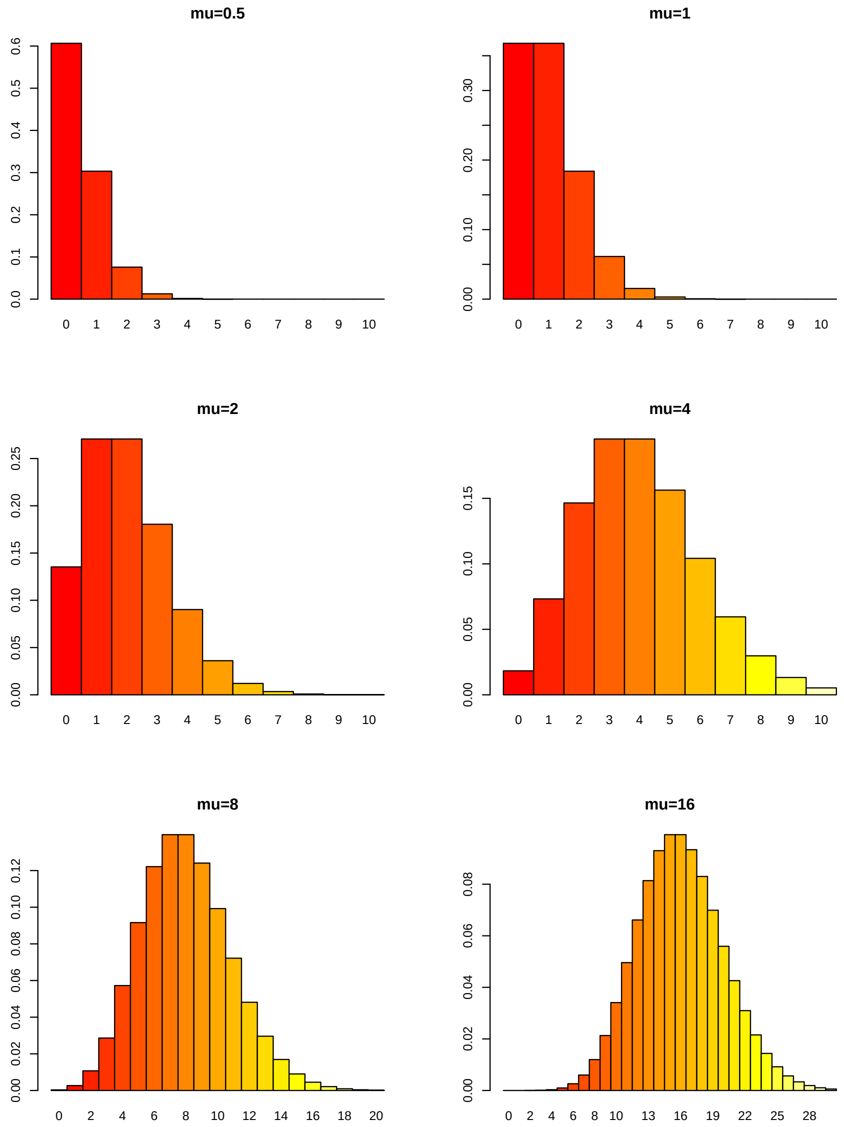
$$Y \sim \text{Poisson}(\lambda)$$

Look of power series expression for exponential function.

where λ is the rate for the given time or area, has pmf

$$p_Y(y) = P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots, \ \lambda > 0.$$

Some Poisson Distributions



START Class Notes 9/13/21

2.7 Some Relationships among the Distributions

- If we sample without replacement and n is small relative to N and M, we can approximate the hypergeometric distribution by using the binomial distribution with $\theta = \frac{M}{N}$:

$$X \sim \text{Hypergeometric}(N, M, n) \;\rightarrow\; X \sim \text{Binomial}(n, \theta = \frac{M}{N})$$

- Let X be a binomial random variable with probability distribution $X \sim \text{Binomial}(n, \theta)$. When $n \rightarrow \infty$ and $\theta \rightarrow 0$ and $\lambda = n\theta$ remains fixed at $\lambda > 0$, then

$$X \sim \text{Binomial}(n, \theta) \rightarrow X \sim \text{Poisson}(\lambda = n\theta)$$

As a rule of thumb, this approximation can be safely applied if:

$$n \geq 100 \qquad \theta \leq 0.01 \qquad n\theta \leq 20$$

3 Continuous Random Variables

The probability distribution of a continuous random variable is defined in a fundamentally different way.

Associated with any *absolutely continuous rv* X (which we will call a continuous rv) is a function f (sometimes written f_X) called a *probability density function (pdf)* or *density function*. This pdf f is a function such that, for any interval (a, b),

$$\underbrace{P(a \leq X \leq b) = \int_a^b f(x)\, dx.}$$

Any function f with the two properties

(i) $f(x) \geq 0$ for all real numbers x, and

(ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

is a pdf.

Note that $f(x)$ is not itself a probability. Instead, the units for $f(x)$ are "probability per unit x." The quantity $f(x)dx$ represents the probability of an infinitesimally small interval centered at x.

A sort of paradox for continuous random variables is that $P(X = x) = 0$ for every real number x. Of course, this doesn't mean that every value of X is impossible. It's simply a consequence of the fact that

$$P(X=b) = \lim_{\epsilon \to 0} P(b \le X \le b + \epsilon) = \lim_{\epsilon \to 0} \int_b^{b+\epsilon} f(x)\, dx = 0.$$

The definition of the p.d.f. also implies that

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

for a continuous random variable X.

NOTE: For discrete case, we must worry about the inequalities.

3.1 Uniform distribution

For $L < R$, define the pdf f by

$$f(x) = \begin{cases} 1/(R-L), & L \leq x \leq R \\ 0, & \text{otherwise.} \end{cases}$$

This is the pdf of the *uniform distribution* on the interval $[L, R]$. Why is this function a pdf?

The distribution is called uniform because, on $[L, R]$, the probability is spread uniformly over the interval. Note that, for $L \leq x < x + \delta \leq R$,

$$P(x \leq X \leq x+\delta) = \int_x^{x+\delta} f(t)\,dt = (R-L)^{-1} \cdot t \Big|_x^{x+\delta}$$

$$= \frac{\delta}{R-L},$$

the only thing that matters the diff btwn the two points i.e. the prob is independent of location.

which doesn't depend on x.

Remark: Any nonnegative function g such that

$$\int_{-\infty}^{\infty} g(x)\,dx < \infty$$

may be transformed into a pdf simply by multiplying it by a constant. Why?

Example 14 Consider the function

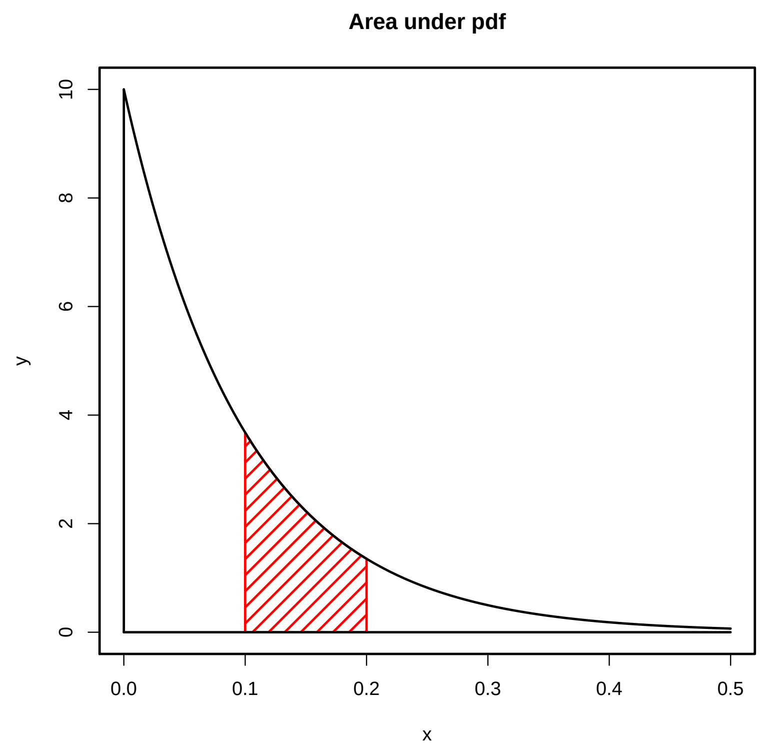
$$g(x) = \begin{cases} e^{-10x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

We have

$$\begin{aligned} \int_{-\infty}^{\infty} g(x)\,dx &= \int_{0}^{\infty} e^{-10x}\,dx \\ &= (-1/10)\cdot e^{-10x}\Big|_0^{\infty} = 1/10 \end{aligned}$$

So $10\int_{-\infty}^{\infty} g(x)\,dx = 1$. Hence, if we define $f(x)$ to equal $10g(x)$ for all x, then f is a pdf.

Suppose X has pdf f. What is the probability that X is between 1/10 and 1/5?



$$P(1/10 \leq X \leq 1/5) = \int_{1/10}^{1/5} 10e^{-10x}\,dx = e^{-1} - e^{-2} \approx 0.233.$$

3.2 The Normal Distribution

affects horizontal scale of plot

Quadratic function of x makes it normal. The log is easy to work w/ b/c we just get a quadratic function of X.

Define

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

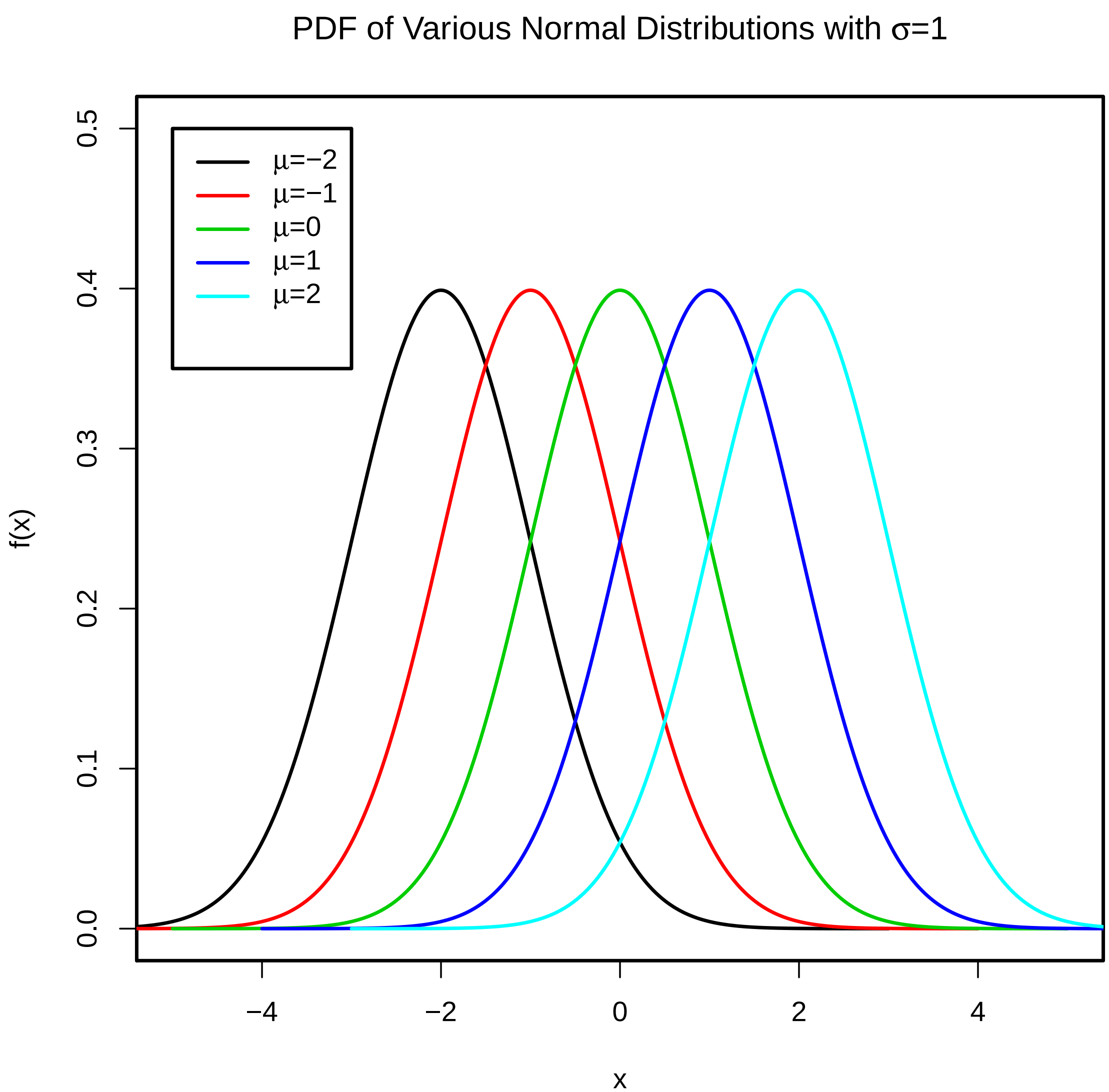
where μ is any real number and σ any positive real number. This function is the probability density function of the *normal distribution*.

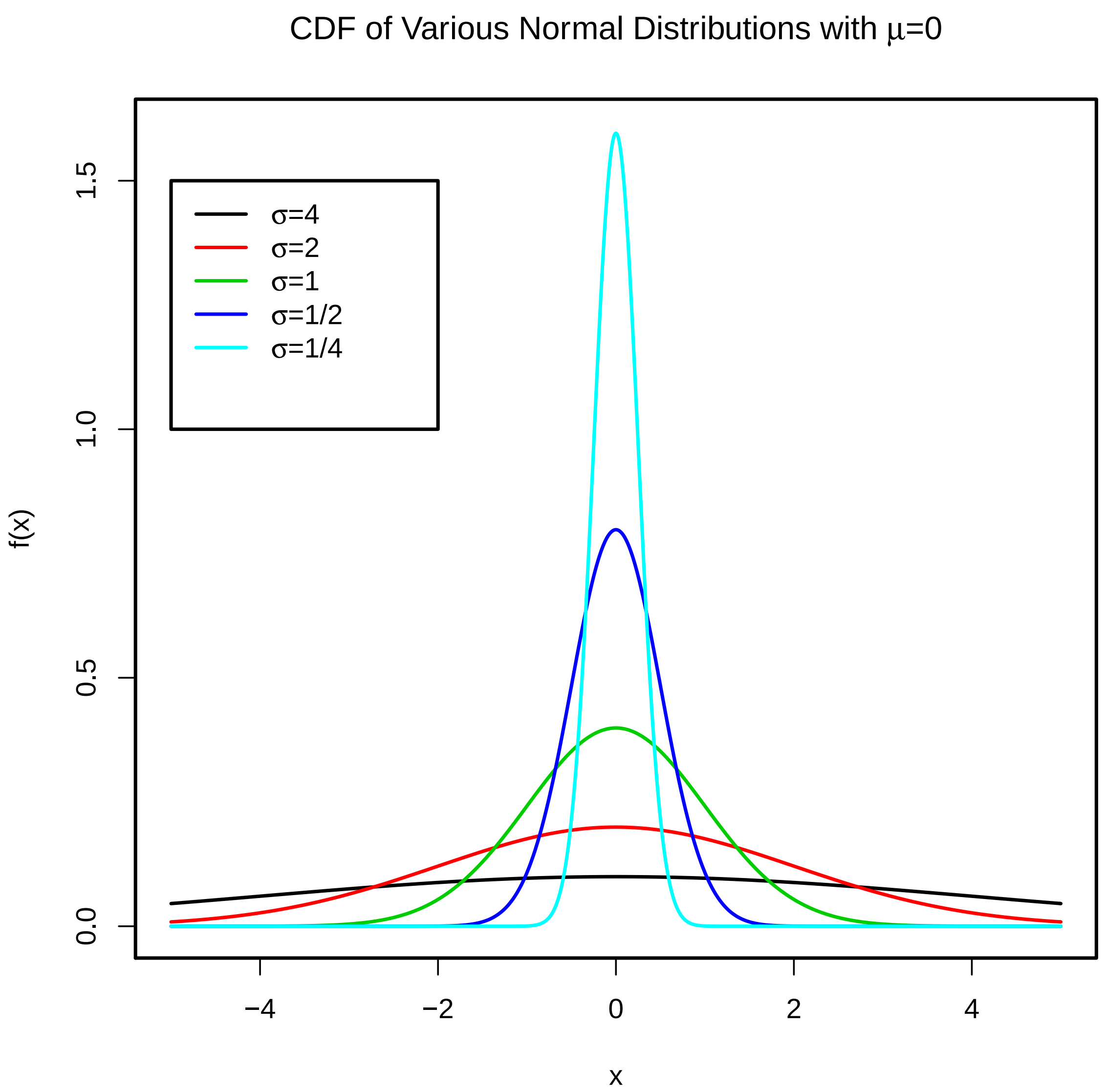
We indicate that the rv X has the above density $f(\cdot)$ by the notation

$$X \sim N(\mu, \sigma^2).$$

- The parameters of $f(\cdot)$ are the mean μ, and the variance, σ^2 (or the standard deviation, σ), respectively.
- Suppose that $X \sim N(\mu, \sigma^2)$ where $\mu = 0$ and $\sigma^2 = 1$. Then X is said to have the *standard normal distribution*. We will often use the notation Z for a standard normal r.v., ϕ for the p.d.f. of a standard normal distribution, and Φ for the c.d.f. of a standard normal distribution.

Some Normal Distributions




CDF of Various Normal Distributions with μ=0
σ=4
σ=2
σ=1
σ=1/2
σ=1/4
f(x)
0.0
0.5
1.0
1.5
−4
−2
0
2
4
x

3.3 Gamma Distribution

The gamma distribution is related to the gamma function, a very important function in applied mathematics. The gamma function is

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \quad \alpha > 0.$$

Some properties of the gamma function are:

1. For $\alpha > 1$, $\quad \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
2. If n is positive integer: $\Gamma(n) = (n-1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

A rv X with the pdf

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, I_{(0,\infty)}(x).$$

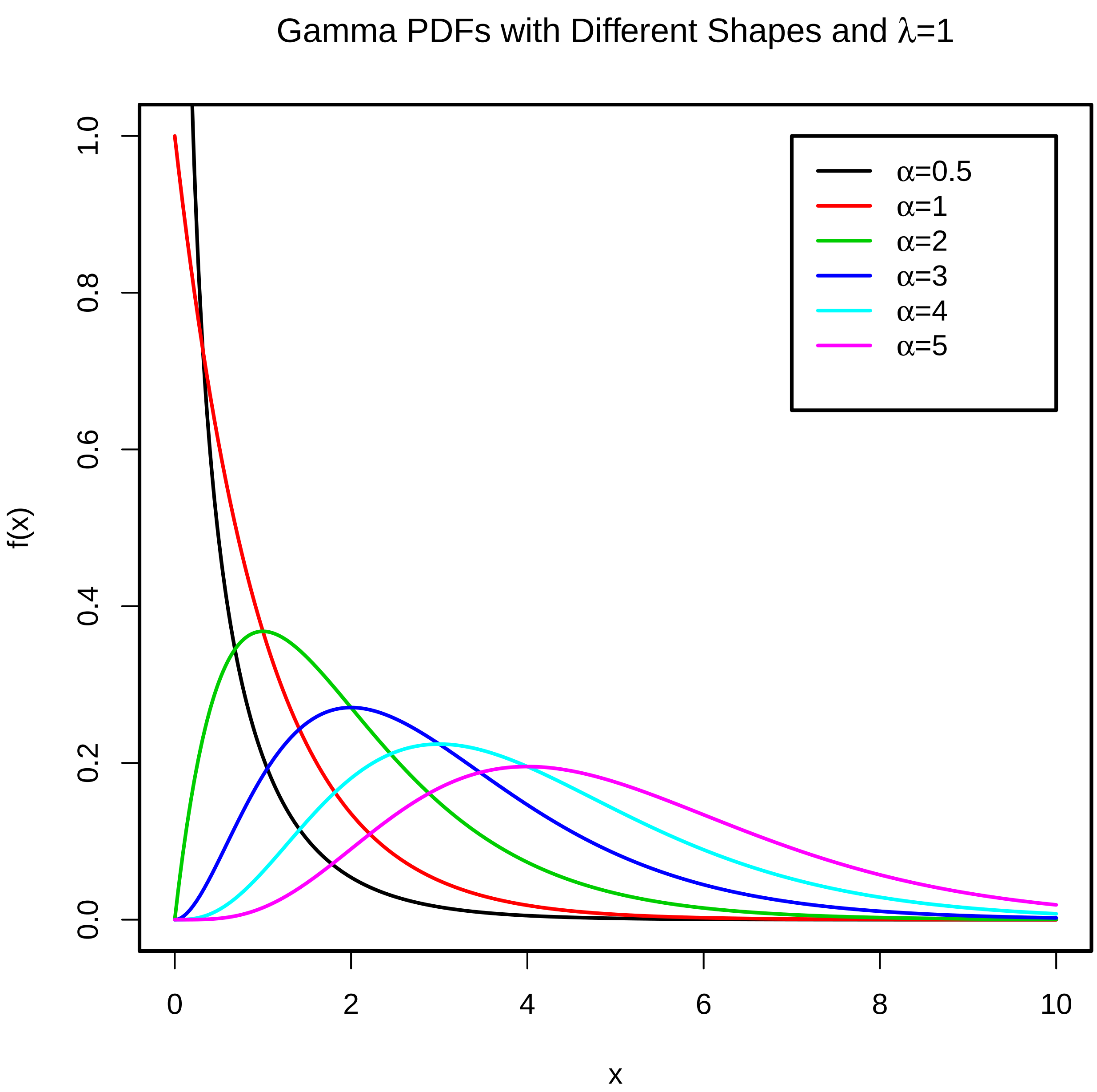
is said to have the gamma(α, λ) distribution. The two parameters α and λ can be any **positive** numbers.

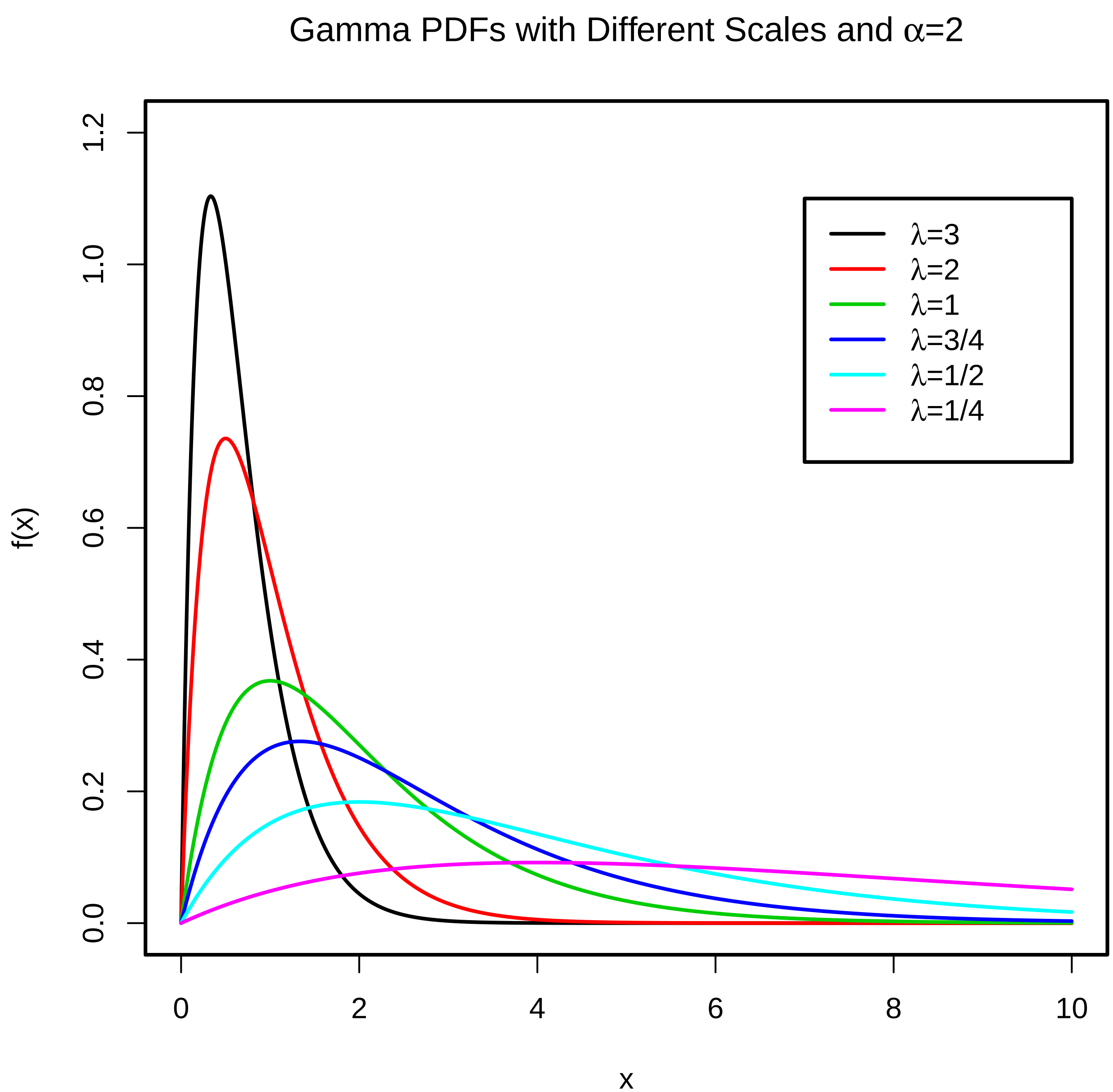
Let's prove that $f(\,\cdot\,|\alpha,\lambda)$ is a density. First of all it's nonnegative. Now show that it integrates to 1.

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\,dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-y}\, d\left(\frac{y}{\lambda}\right) \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y}\,dy \\
&= 1.
\end{aligned}$$

- α is called a shape parameter, and λ is called a scale parameter. On the next two slides we see the gamma densities corresponding to various values of these parameters.

- The gamma distribution is a popular model for variables that are inherently positive. In particular, it is often used to model distributions of *lifetimes*, such as those of people, electrical components, etc.


Gamma PDFs with Different Shapes and λ=1
α=0.5
α=1
α=2
α=3
α=4
α=5
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10


Gamma PDFs with Different Scales and α=2
λ=3
λ=2
λ=1
λ=3/4
λ=1/2
λ=1/4
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
2
4
6
8
10

The exponential distribution (encountered in Example 23) is a special case of the gamma distribution in which $\alpha = 1$. The exponential density is

$$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x).$$

The exponential distribution has an interesting *memoryless* property. Suppose X has an exponential distribution and let $x > 0$ and $\delta > 0$.

$$\begin{aligned} P(X > x + \delta | X > x) &= \frac{P(X > x + \delta)}{P(X > x)} = \frac{e^{-\lambda(x+\delta)}}{e^{-\lambda x}} \\ &= e^{-\lambda\delta} = P(X > \delta). \end{aligned}$$

So, in terms of lifetimes, this says that if, say, a person lives to at least age x, the probability that he/she will live at least another δ years **is the same as** the unconditional probability (i.e., at age 0) that the person will live more than δ years. In this sense, the person doesn't "remember" how long he/she has lived. Unfortunately, lifetimes of people aren't memoryless, and hence not exponential. The exponential model does work, though, for certain other lifetime distributions.

3.4 Beta Distribution

The beta distribution is often used to model random variables that are restricted to the interval $(0, 1)$. A random variable X with the pdf

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} I_{(0,1)}(x)$$

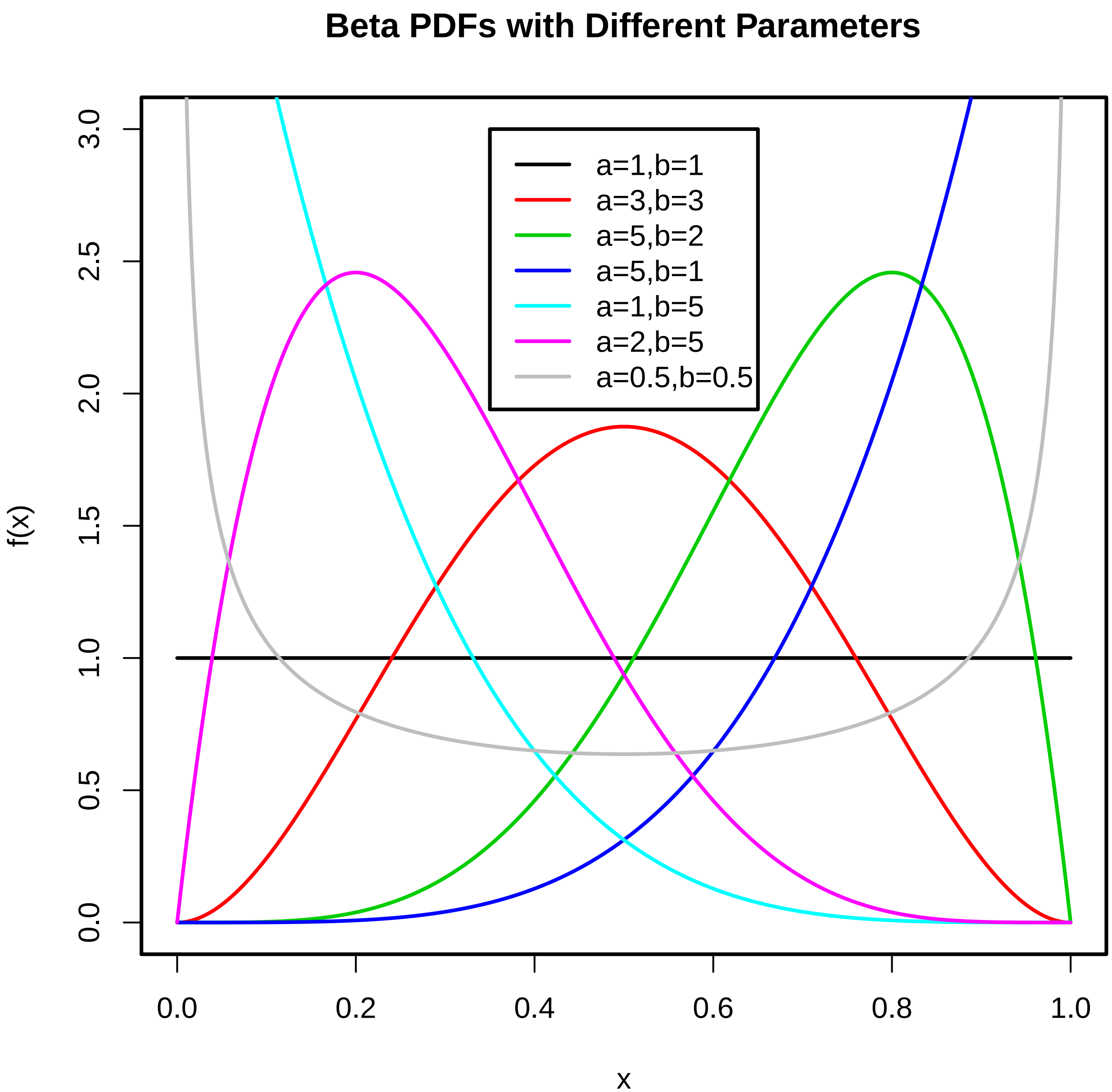
is said to have the beta(a, b) distribution. The parameters a and b can be any positive numbers. We will use the beta distribution in Bayesian statistics.

The *beta function* is defined as

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

We could write the beta pdf using the beta function as

$$f(x) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1} I_{(0,1)}(x).$$


Beta PDFs with Different Parameters
a=1,b=1
a=3,b=3
a=5,b=2
a=5,b=1
a=1,b=5
a=2,b=5
a=0.5,b=0.5
f(x)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
x

4 Cumulative Distribution Functions

Any random variable, whether discrete or continuous, may be characterized by its *cumulative distribution function* (cdf), which is also called its *distribution function*.

The cdf F_X of a rv X is defined by

$$F_X(x) = P(X \leq x), \quad \text{for all } x.$$

Note: We sometimes write $F(x)$ for $F_X(x)$.

For a discrete rv X with p.m.f. p_X, the cdf may be expressed

$$F_X(x) = \sum_{\{k:k\leq x\}} p_X(k), \quad \text{for all } x.$$

If X is continuous with pdf f_X, then

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

Any cdf F_X satisfies the following properties:

1. F_X is a nondecreasing function, meaning that whenever $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$.
2. F_X is right continuous.
3. $\lim_{x \to -\infty} F_X(x) = 0$
4. $\lim_{x \to \infty} F_X(x) = 1$

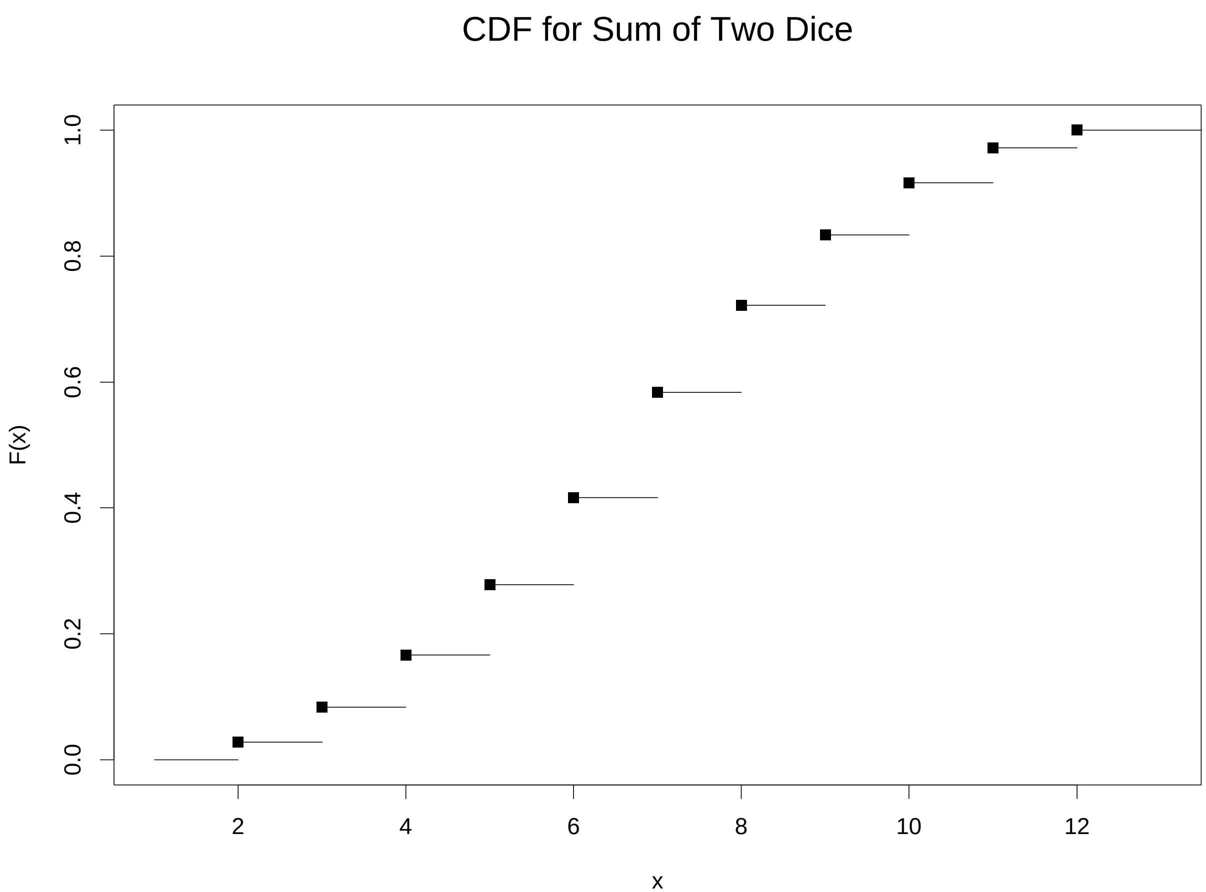
Right continuity of F_X means that, for every x,

$$\lim_{\epsilon \downarrow 0} F_X(x + \epsilon) = F_X(x).$$

We also note that the distribution of a random variable X with cdf F_X is completely determined by the cdf F_X.

4.1 Cumulative distribution functions of discrete rvs

Example: Consider again the rv $X =$ the sum of two dice. The c.d.f. of X is given in the following figure:



The previous plot illustrates a couple of things about the cdfs of discrete random variables.

- If F_X is the cdf of a *discrete* rv X, then F_X is continuous at every x such that $P(X = x) = 0$.
- If x is such that $P(X = x) > 0$, then F_X is not left continuous at x. In fact, the height of the jump is

$$F_X(x) - F_X(x-) = P(X = x)$$

- The cdf of a discrete rv is a *step function*. This means it's constant for awhile then jumps up, is constant for awhile again and then jumps up, etc.

- Consider $F_X(x)$ at points near $x = 2$ for the distribution on the previous page. We have $F_X(2) = P(X \leq 2) = P(X = 2) = \frac{1}{36}$.

  Now suppose $2 < x < 3$. Then

$$F_X(x) = P(X \leq x) = P(X = 2) = \frac{1}{36},$$

  and so

$$\lim_{\epsilon \downarrow 0} F_X(2 + \epsilon) = F(2).$$

- If $x < 2$, then $F_X(x) = P(X \leq x) = 0$, and hence

$$\lim_{\epsilon \uparrow 0} F_X(2 - \epsilon) = 0 \neq F_X(2).$$

  So, we've shown that F_X is right, but not left, continuous at 0. The same thing is true at any point where the cdf of a discrete rv jumps up.

4.2 Cumulative distribution functions of continuous rvs

The properties of the cdf of a continous rv differ from those of a cdf of a discrete rv.

- If F_X is the cdf of a *continuous* rv, then F_X is continuous at all x.

- We can obtain the pdf of a continuous rv X from the cdf F_X:

  Suppose a continuous random variable X has cdf F_X. Then the pdf f_X of X is

  $$f_X(x) = \frac{dF_X(x)}{dx}$$

  at each x for which F_X is differentiable. At points x where F_X is not differentiable, we may define $f_X(x)$ however we like without affecting the fact that

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

Example 16 Let X have the pdf

$$f_X(x) = \begin{cases} 10e^{-10x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

Find the cdf of X. Clearly, $F(x) = 0$ for $x < 0$. Now let $x \ge 0$.

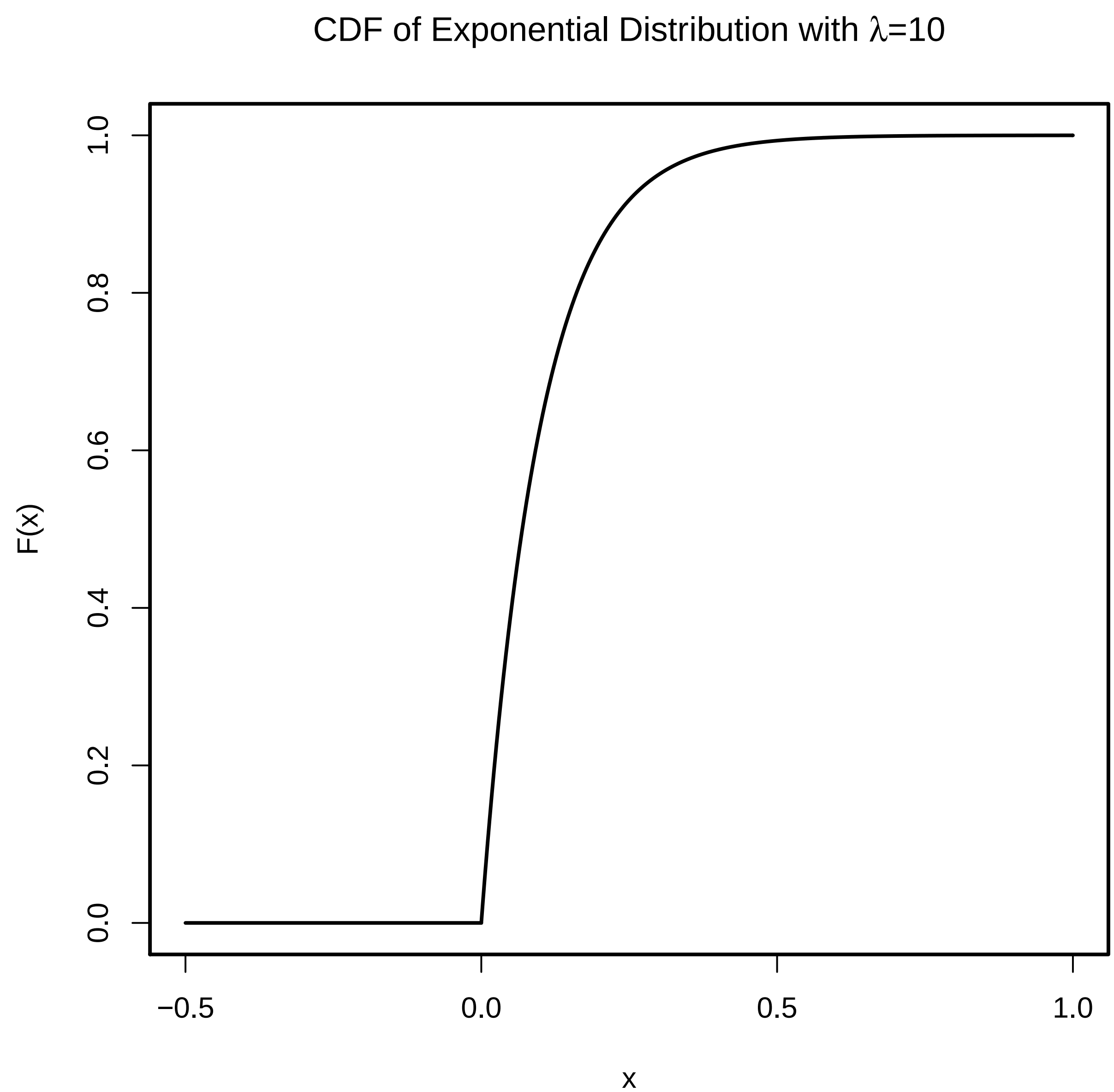
$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f(u)\,du = \int_0^x 10e^{-10u}\,du \\ &= -e^{-10u}\Big|_0^x = 1 - e^{-10x}. \end{aligned}$$

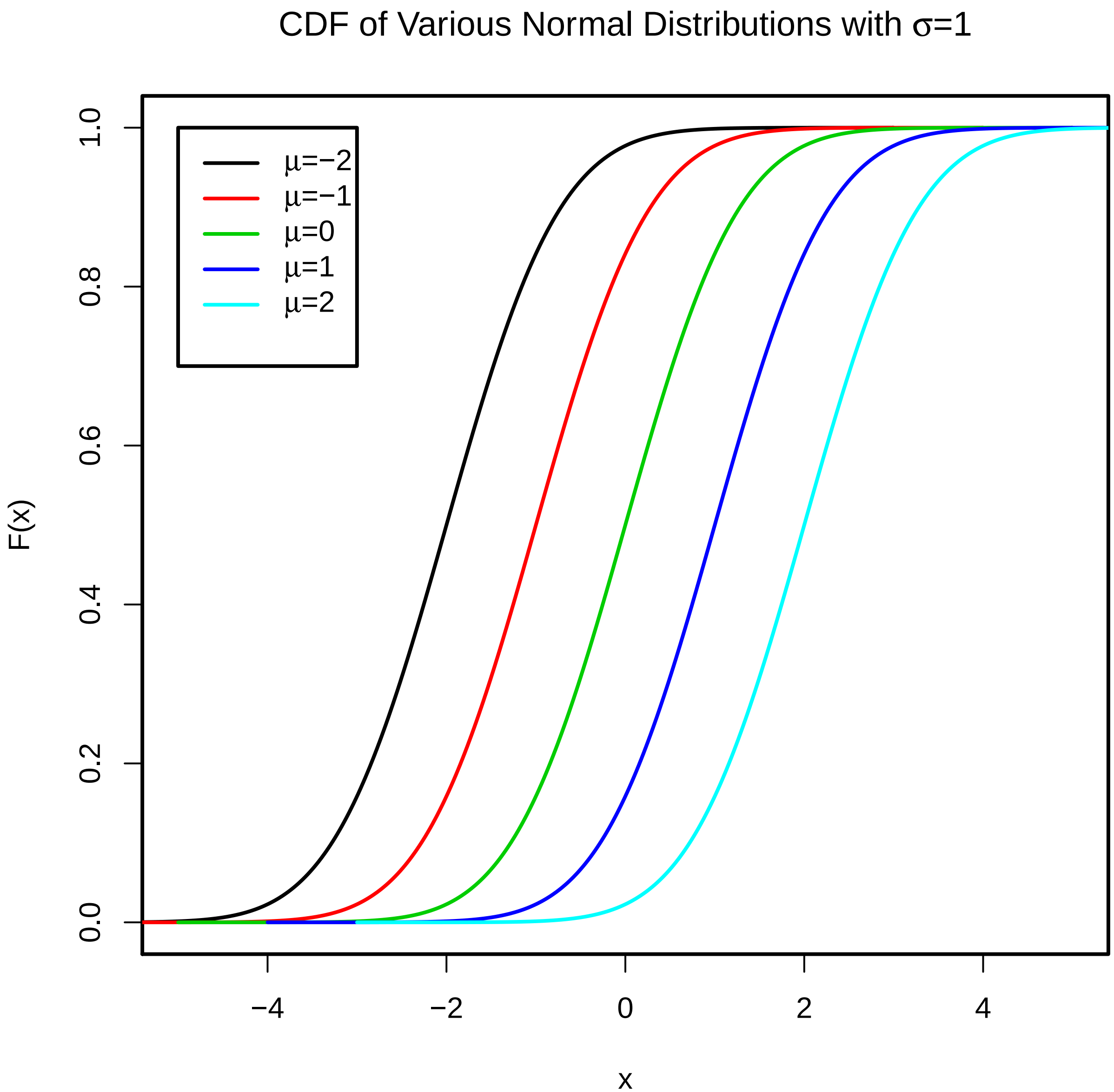
The cdf of X is thus

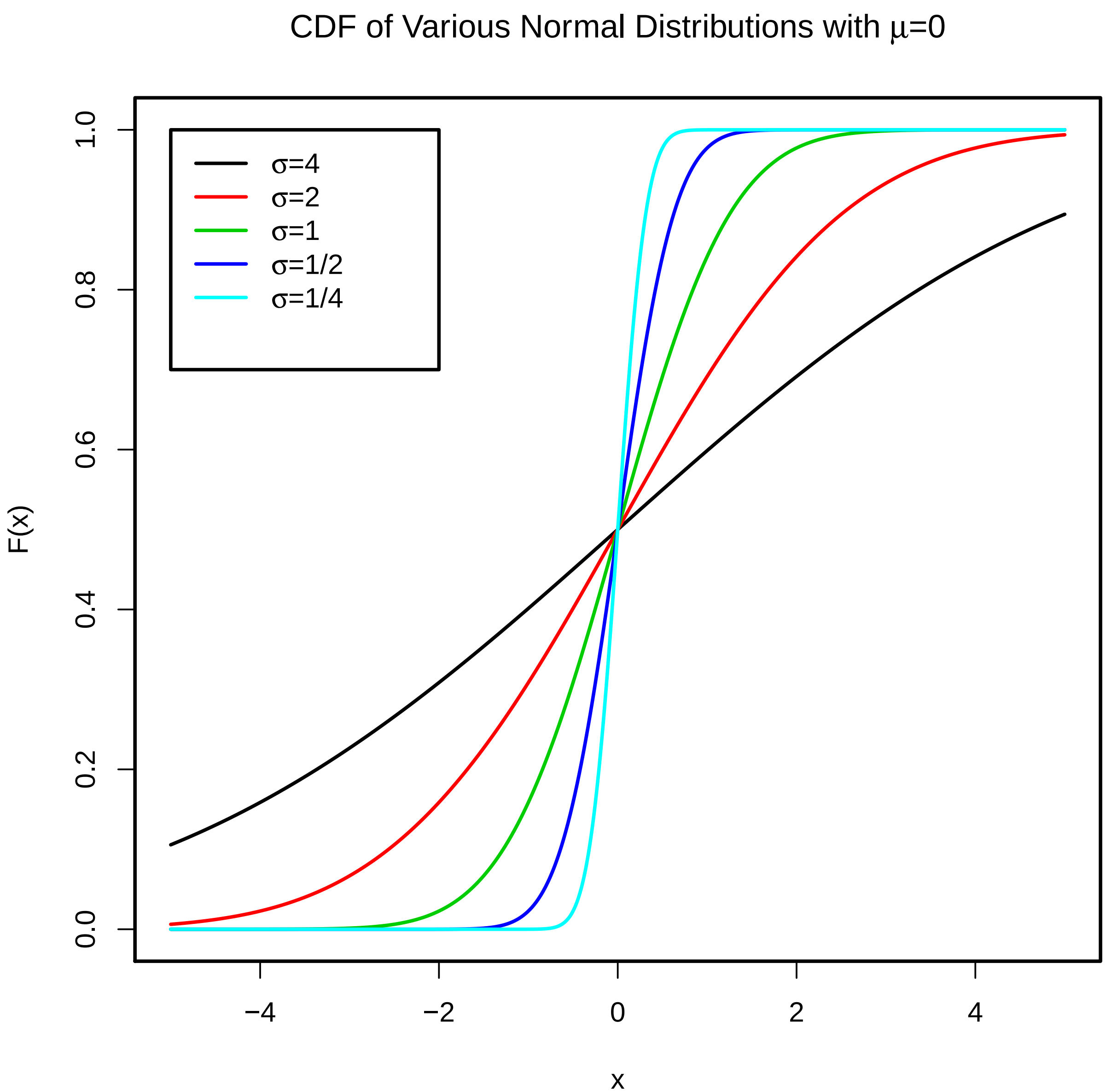
$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-10x}, & x \ge 0. \end{cases}$$

Note that F_X is continuous everywhere, including $x = 0$. Why?

A plot of the cdf is given on the next page.


CDF of Exponential Distribution with λ=10
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
−0.5
0.0
0.5
1.0
x


CDF of Various Normal Distributions with σ=1
μ=−2
μ=−1
μ=0
μ=1
μ=2
F(x)
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
x


CDF of Various Normal Distributions with μ=0
σ=4
σ=2
σ=1
σ=1/2
σ=1/4
F(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4

5 Quantile functions

The quantile function of X is essentially the inverse of the cdf. The quantile function provides a value such that the cdf equals a given probability from 0 to 1. In fact, whenever F is a 1-1 function, the quantile function *is* the inverse of F. (Note: F can only be 1-1 when X is a continuous rv.)

Suppose F is 1-1. Then we define its quantile function Q (or the p^{th} quantile x_p) by

$$Q(p) = x_p = F^{-1}(p), \quad 0 < p < 1,$$

i.e., $Q(p) = x_p$ is the number x_p such that $F(x_p) = p$.

We call $Q(p)$ the "pth quantile of X," or the "$100p$th percentile of X." The interpretation of, for example, $Q(.75) = x_{.75}$ is that 75% of the distribution is less than or equal to $Q(.75)$.

Example 16 Suppose that X has the cdf

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ 1 - e^{-10x}, & x \geq 0. \end{cases}$$

We will now find the quantile function.

Set

$$p = F(x) = 1 - e^{-10x}$$

Then

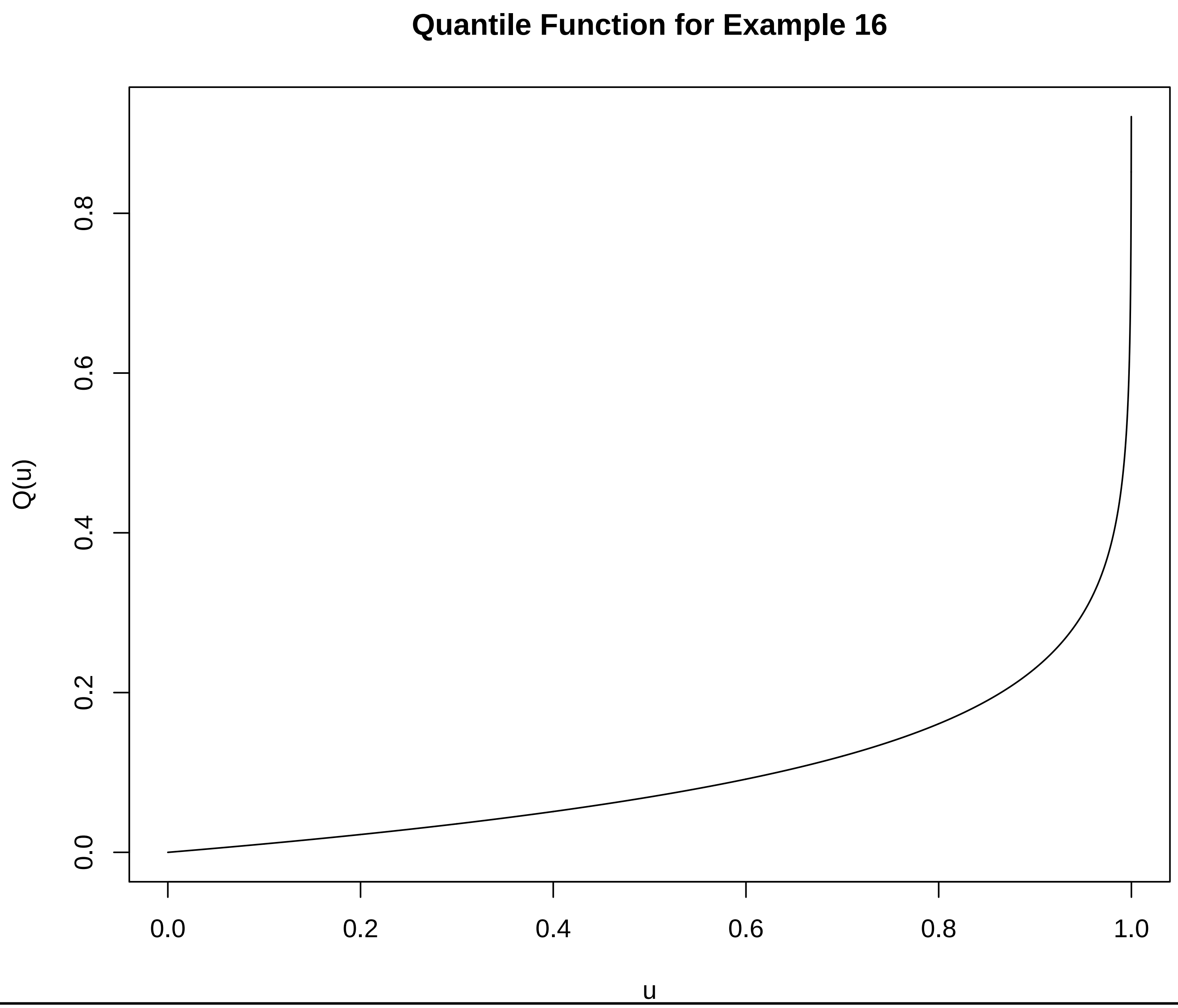
$$1 - p = e^{-10x}$$

and

$$\log(1 - p) = -10x$$

where $\log$ donotes the natural logarithm (i.e., $\log_e$).

Thus, $Q(p) = -\log(1 - p)/10$. A plot of the quantile function appears on the next slide.


Quantile Function for Example 16
Q(u)
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
u

6 Functions of a Random Variable

Suppose we are interested in some function $Y = h(X)$ of a random variable X. Then Y is also a random variable with some probability distribution. For example, we might want to know the probability distribution of $Y = X^2$, or that of $Y = \cos(X)$.

6.1 Functions of a Discrete RV

Consider the case where X is discrete and has pmf p_X, and suppose we want to know the pmf of $Y = h(X)$. The pmf p_Y of Y is

$$
\begin{aligned}
p_Y(y) &= P(Y = y) \\
&= P(h(X) = y) \\
&= \sum_{\{x : h(x) = y\}} p_X(x).
\end{aligned}
$$

Example: Suppose that $X \sim$ binomial$(3, \theta)$. Let $Y = (X-1)^2$. Find the pmf of Y.

We first observe that the range of the rv X is $R_X = \{0, 1, 2, 3\}$. Then the range of the rv Y is $R_Y = \{y : y = h(x)$ for some $x \in R_X\} = \{0, 1, 4\}$.

We obtain the pmf of Y as

$$\begin{aligned}
p_Y(0) &= P[Y=0] = P[\{x : (x-1)^2 = 0\}] = p_X(1) = 3\theta(1-\theta)^2 \\
p_Y(1) &= P[Y=1] = P[\{x : (x-1)^2 = 1\}] = p_X(0) + p_X(2) \\
&= (1-\theta)^3 + 3\theta^2\,(1-\theta) \\
p_Y(4) &= P[Y=4] = P[\{x : (x-1)^2 = 4\}] = p_X(3) = \theta^3 \\
p_Y(y) &= 0, \qquad \text{for all other } y
\end{aligned}$$

6.2 Functions of a Continuous RV

Suppose X is continuous and consider the function $Y = h(X)$. The cdf of Y is

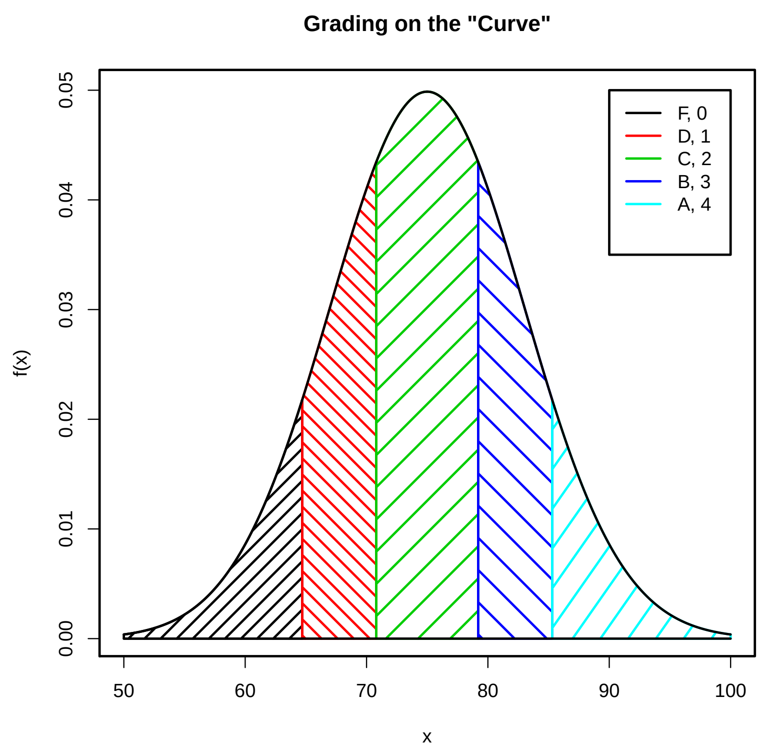
$$\begin{aligned} F_Y(y) &= P(h(X) \leq y) \\ &= \int_{\{x:h(x)\leq y\}} f_X(x)\,dx, \end{aligned}$$

where f_X is the pdf of X.

Now, Y can be continuous or discrete in the case where X is continuous.

A fairly common practice is to *categorize* continuous data.

Example: Suppose that a college professor *grades on a curve.* The scores (X) on a test are assumed to be normally distributed with $\mu = 75$ and $\sigma = 8$. He wants cut-offs so that the top 10% receive As, the next 20% Bs, the middle 40% Cs, the next 20% Ds, and the rest Fs. We could carry this out using a plot of the normal distribution $\mu = 75$ and $\sigma = 8$:



Suppose that $h(X)$ denotes the grade corresponding to a score of X on a four-point scale. To accomplish the desired distribution of grades, the professor uses the function

$$h(x) = \begin{cases} 0, & x \leq 64.7 \\ 1, & 64.7 < x \leq 70.8 \\ 2, & 70.8 < x \leq 79.2 \\ 3, & 79.2 < x \leq 85.3 \\ 4, & x > 85.3. \end{cases}$$

Verify that $Y = h(X)$ has pmf

y	0	1	2	3	4
$P(Y = y)$	0.10	0.20	0.40	0.20	0.10

If Y is continuous, then we may find its pdf as follows:

- Determine the cdf F_Y of Y as described on Slide 54.
- Differentiate F_Y to get the pdf, i.e., $f_Y(y) = \frac{d}{dy}F_Y(y) = F'_Y(y)$.

Remark: It is always important to determine the range of the rv Y, defined by $\{y : f_Y(y) > 0\}$. This range can be determined from the cdf of Y by seeing for which values of y the cdf is increasing ($f_Y(y) > 0$) and for which values of y the cdf is constant ($f_Y(y) = 0$).

PDF of a Linear Transformation:

Suppose that X has pdf f_X and $Y = aX + b$ where $a > 0$, the pdf of Y is

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

The cdf of Y is

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) \\ &= P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

Thus,

$$f_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Example 20 Suppose that $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ where $a > 0$, then $Y \sim N(a\mu + b, a^2\sigma^2)$.

We now let f_X be the $N(\mu, \sigma^2)$ density. Then we obtain

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-b-a\mu}{a\sigma}\right)^2\right], \quad -\infty < y < \infty.$$

We recognize this as the pdf of a $N(a\mu + b, a^2\sigma^2)$ distribution.

Special Cases:

- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.
- If $Z \sim N(0, 1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

Example 21 Suppose that U is uniform on the interval $(0,1)$. Let $X = -\log(U)$. Find the pdf of X.

The pdf of U is $f_U(u) = 1$ for $0 < u < 1$. The c.d.f. is $F_U(u) = u$ for $0 < u < 1$.

The c.d.f. of X is given by

$F_X(x) = P(X \leq x) = P(-\log(U) \leq x) = P(U \geq e^{-x}) = 1 - e^{-x}$ for $0 < x < \infty$ and $F_X(x) = 0$ for $x \leq 0$.

Thus, the pdf of X is

$$f_X(x) = F_X'(x) = \begin{cases} e^{-x}, \ x > 0 \\ 0, \ x \leq 0 \end{cases}.$$

Example 22 Suppose that X has the standard normal distribution with pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \ -\infty < x < \infty.$$

Let $Y = X^2$. Find the pdf of Y.

For $y > 0$,

$$\begin{aligned} P[Y \le y] &= P[X^2 \le y] = P[-\sqrt{y} \le X \le \sqrt{y}] \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx. \end{aligned}$$

Then

$$\begin{aligned} f_Y(y) &= F'_Y(y) = \frac{d}{dy} 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 2\frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \frac{1}{2} \frac{1}{\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y/2} = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \text{ for } y > 0. \end{aligned}$$

Remark: When using the cdf of X to find the cdf of Y, the range R_Y of the rv Y is determined by its cdf. In Example 22, the range of $Y = X^2$ is $R_Y = (0, \infty)$ since $F_Y(y) = 0$ for $-\infty < y \le 0$. If you use the formula on the next slide to obtain the pdf of Y, you need to take care in obtaining the range of Y.

The range of X is defined as $R_X = \{x : f_X(x) > 0\}$. Then the range of $Y = h(X)$ is $R_Y = \{h(x) : x \in R_X\}$.

We next look at the case where the function h is monotone over the range of X. This occurs when the transformation is either strictly increasing or strictly decreasing over this range. In either case, the transformation h has an inverse, h^{-1}. Then for increasing functions h,

$$
\begin{aligned}
F_Y(y) &= P(h(X) \le y) \\
&= P(X \le h^{-1}(y)) \\
&= F_X(h^{-1}(y)),
\end{aligned}
$$

where F_X is the cdf of X.

Since

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

we can take its derivative to obtain the pdf. If X is continuous and h is strictly increasing and differentiable, then $f_X(x)\,dx = f_y(y)\,dy \Rightarrow f_y(y) = \frac{dx}{dy} f_X(x)$

$$f_Y(y) = F_Y'(y) = f_X(h^{-1}(y)) \cdot \frac{dh^{-1}(y)}{dy}.$$

In the same way it can be shown that if h is strictly *decreasing* and h is differentiable, then

$$f_Y(y) = -f_X(h^{-1}(y)) \cdot \frac{dh^{-1}(y)}{dy}.$$

Combining these two results, we may say that when h is differentiable and strictly increasing or decreasing, then

$$f_Y(y) = f_X(h^{-1}(y)) \cdot \left| \frac{dh^{-1}(y)}{dy} \right|.$$

Example 21 again Suppose that U is uniform on the interval $(0, 1)$. Let $Y = -\log(U)$. Find the pdf of Y.

The pdf of U is $f_U(u) = 1$ for $0 < u < 1$. Here $h(u) = -\log(u)$ is strictly decreasing for $0 < u < 1$. (Why?)

To obtain $h^{-1}(y)$, we set $y = h(u) = -\log(u)$ and solve for u:

We obtain $u = e^{-y} = h^{-1}(y)$. Thus, the pdf of Y is

$f_U(e^{-y})$

$$f_Y(y) = 1 \times \left| \frac{de^{-y}}{dy} \right| = |-e^{-y}| = e^{-y}, \; y > 0.$$

To verify this range for Y,

$$R_Y = \{-\log(u) : 0 < u < 1\} = (0, \infty).$$

7 Joint Distributions

In many experiments there are several numbers associated with each possible outcome. For the moment, suppose there are only two of interest. So, we have two functions, X and Y, each of which maps points in the sample space into points on the real number line. The joint behavior of two random variables, X and Y, depends on their joint probability distribution, which specifies $P[(X, Y) \in B]$ for any subset B of R^2.

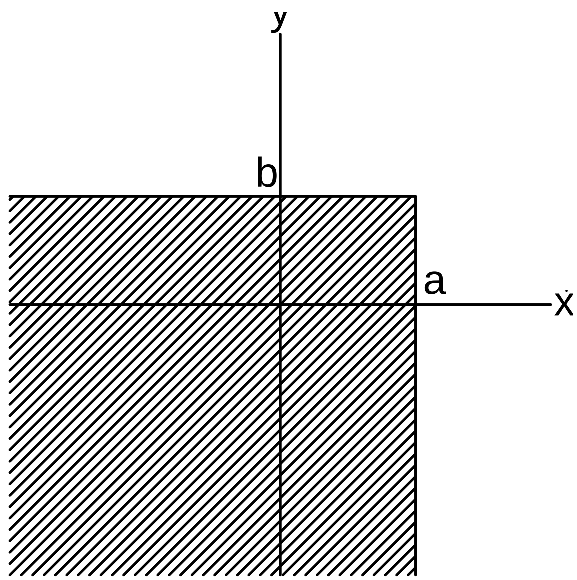
As for single random variables, we can specify a joint distribution using the cumulative distribution function:

comma ⇒ intersection

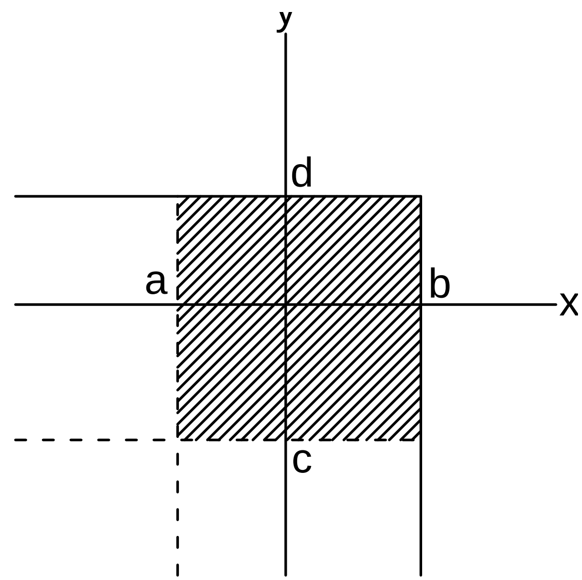
$$F(x, y) = P(X \leq x, Y \leq y).$$

We will investigate pmfs for bivariate discrete rvs and pdfs for bivariate continuous rvs.

F(a,b)=prob of shaded area



P(a<X<=b,c<Y<=d)=
prob of shaded area



Notice that we can obtain the probability that (X, Y) falls in a rectangle using the cdf:

$$P(a < X \leq b, c < Y \leq d) = F(b,d) - F(b,c) - F(a,d) + F(a,c).$$

7.1 Discrete Bivariate Distributions

We now consider the *joint distribution* of X and Y. In the case where both X and Y are discrete, the joint distribution is given by a frequency function (or probability mass function) p, defined by

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

for all ordered pairs (x, y). In this discrete case, there are at most countably many ordered pairs such that $p_{X,Y}(x, y) > 0$. The joint range of (X, Y) is $R_{X,Y} = \{(x, y) : p_{X,Y}(x, y) > 0\}$.

For any two-dimensional set A, we have

$$P((X, Y) \in A) = \sum_{(x,y) \in A \cap R_{X,Y}} p_{X,Y}(x, y)$$

Example 12 Again Suppose that a coin is tossed four times. Then there are $2^4 = 16$ possible sequences of tosses (such as $HTHT$). Let X be the number of heads until the first tail and Y be the number of heads in the four tosses. We can form a table of the mapping for all possible sequences:

Toss	(X,Y)	Toss	(X,Y)	Toss	(X,Y)	Toss	(X,Y)
$TTTT$	(0,0)	$TTTH$	(0,1)	$TTHT$	(0,1)	$TTHH$	(0,2)
$THTT$	(0,1)	$THTH$	(0,2)	$THHT$	(0,2)	$THHH$	(0,3)
$HTTT$	(1,1)	$HTTH$	(1,2)	$HTHT$	(1,2)	$HTHH$	(1,3)
$HHTT$	(2,2)	$HHTH$	(2,3)	$HHHT$	(3,3)	$HHHH$	(4,4)

Since each sequence is equally likely, we can find the joint pmf of (X,Y). For instance

$$P[(X,Y) = (0,1)] = P[\{(TTTH, TTHT, THTT)\}] = \frac{3}{16}.$$

Proceeding in this way we get the following joint distribution for X and Y as shown in this table of the joint pmf, $p_{X,Y}(x,y) = P[X = x, Y = y]$:

y

x		0	1	2	3	4
	0	1/16	3/16	3/16	1/16	0
	1	0	1/16	2/16	1/16	0
	2	0	0	1/16	1/16	0
	3	0	0	0	1/16	0
	4	0	0	0	0	1/16

We can use this table to find probabilities of events involving X and Y. For example,

$$
\begin{aligned}
P[X = Y] &= p_{X,Y}(0,0) + p_{X,Y}(1,1) + p_{X,Y}(2,2) + p_{X,Y}(3,3) + p_{X,Y}(4,4) \\
&= \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{5}{16}
\end{aligned}
$$

7.1.1 Multinomial Probability Distribution

Binomial case
$\theta_1 = p$
$\theta_2 = 1-p$
X successes
$n-X$ Failures

A **multinomial experiment** satisfies the following conditions:

1. Experiment consists of n trials.
2. Each trial can result in one of k mutually exclusive categories.
3. Trials are independent.
4. The probability of the i^{th} category is the same for each trial: θ_i, where $\theta_1 + \cdots + \theta_k = 1$.

The observable random variables are $(X_1, X_2, \ldots, X_k)$ where X_i = number of trials resulting in the i^{th} category, where $X_1 + \cdots + X_k = n$. The random variables $(X_1, \ldots, X_k)$ are said to have a multinomial distribution, Mult$(n, \theta_1, \ldots, \theta_k)$, with probability mass function

$$p(x_1, \ldots, x_k) = P[X_1 = x_1, \ldots, X_k = x_k] = \frac{n!}{x_1! \cdots x_k!}\theta_1^{x_1} \cdots \theta_k^{x_k}.$$

7.1.2 Marginal Distributions

Suppose X and Y are random variables having some joint distribution. When we obtain the probability distribution of, say, X from the joint distribution, the former distribution is referred to as the *marginal distribution* of X. (The marginal distribution of Y is defined similarly.)

For a pair of discrete random variables X and Y with joint probability mass function $p_{X,Y}$, the marginal pmf of X is

$$\begin{aligned} p_X(x) &= P(X = x) \\ &= \sum_{y} p_{X,Y}(x, y), \end{aligned}$$

after we sum across all values of y, the function no longer depends on y.

where the sum extends over all y in the range of Y. Similarly

$$\begin{aligned} p_Y(y) &= P(Y = y) \\ &= \sum_{x} p_{X,Y}(x, y). \end{aligned}$$

Example 12 again Find the marginal distributions corresponding to the joint distribution for X and Y.

NOTE: Knowing marginal doesn't give us joint. (not w/o assumptions)
Knowing joint does give us marginal

		y					
		0	1	2	3	4	$p_X(x)$
	0	1/16	3/16	3/16	1/16	0	8/16
	1	0	1/16	2/16	1/16	0	4/16
x	2	0	0	1/16	1/16	0	2/16
	3	0	0	0	1/16	0	1/16
	4	0	0	0	0	1/16	1/16
	$p_Y(y)$	1/16	4/16	6/16	4/16	1/16	

$$p_X(0) = 1/16+3/16+3/16+1/16 = 8/16, \quad p_X(1) = 1/16+2/16+1/16 = 4/16, \text{ etc.}$$

$$p_Y(0) = 1/16, \quad p_Y(1) = 3/16 + 1/16 = 4/16, \quad \text{etc.}$$

7.2 Continuous Joint Distributions

Suppose X and Y are both (absolutely) continuous random variables. The joint distribution of X and Y is given by a pdf f. For a subset A of the (x, y)-plane,

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\, dxdy.$$

So, the probability that (X, Y) lies in a given set is the volume under f that lies over the set.

Can derive marginal from joint

Cannot derive joint from marginal *

Just as in the univariate case, any nonnegative function f such that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dxdy = 1$$

is a pdf.

Example 18 Consider the joint pdf:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0 < x < 2, \quad 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

We see that $f_{X,Y}(x,y) \geq 0$ at all (x,y). Also,

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy &= \int_0^1\int_0^2 \frac{x(1+3y^2)}{4}dxdy \\ &= \int_0^1 (\frac{x^2}{8} + \frac{3x^2y^2}{8}\Big|_{x=0}^{x=2})dy \\ &= \int_0^1 (\frac{1}{2} + \frac{3y^2}{2})dy \\ &= \frac{y}{2} + \frac{y^3}{2}\Big|_0^1 \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

Example 19 Suppose X and Y have joint pdf

$$f_{X,Y}(x,y) = \begin{cases} 3(x+y), & 0 < x < 1, 0 < y < 1, \\ & 0 < x+y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Verify that this function is a pdf.

Now, find the following probabilities:

- $P(X < 1/2, Y < 1/2)$
- $P(X < Y)$
- $P(X < 1/2, Y < 3/4)$

In each case, one should first identify the region of interest. Drawing a picture is usually helpful.

$$P\left(X < \frac{1}{2}, Y < \frac{1}{2}\right) \;=\; \int_0^{1/2}\int_0^{1/2} 3(x+y)\,dxdy \;=\; 3/8.$$

$$\begin{aligned} P(X < Y) &= \int_0^{1/2}\int_x^{1-x} 3(x+y)\,dydx \\ &= 3\int_0^{1/2}(1/2 - 2x^2)\,dx \;=\; 1/2. \end{aligned}$$

$$P(X < 1/2, Y < 3/4) = \int_0^{3/4}\int_0^{1/4} f_{X,Y}(x,y)\,dxdy +$$

$$\int_{1/4}^{1/2}\int_0^{1-x} f_{X,Y}(x,y)\,dydx$$

$$= 36/128 + 41/128 \;=\; 77/128.$$

7.2.1 Bivariate Distribution Functions for Continuous RV

The joint *distribution function* of continuous random variables (X, Y) is the function F such that

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)\, du dv.$$

As in the univariate case, the pdf may be obtained from the df by differentiation. We have

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

at all (x, y) where the mixed partial derivative exists. Elsewhere, the pdf may be defined in any convenient way.

7.2.2 Marginal Distributions of Continuous RV

For continuous random variables X and Y with joint pdf $f_{X,Y}$, the marginal pdf f_X of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy, \quad \text{for all } x.$$

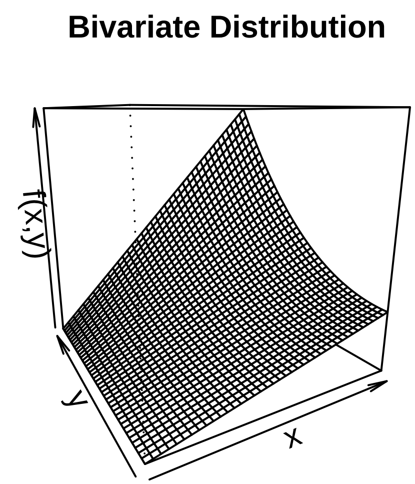
The marginal pdf f_Y of Y is

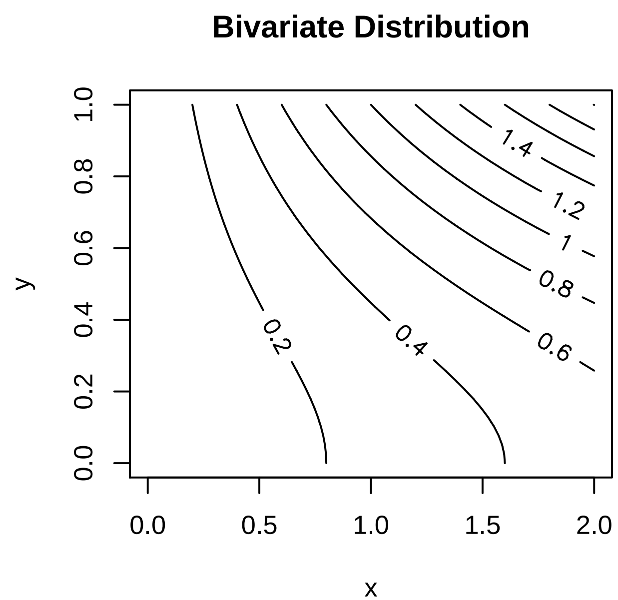
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx, \quad \text{for all } y.$$

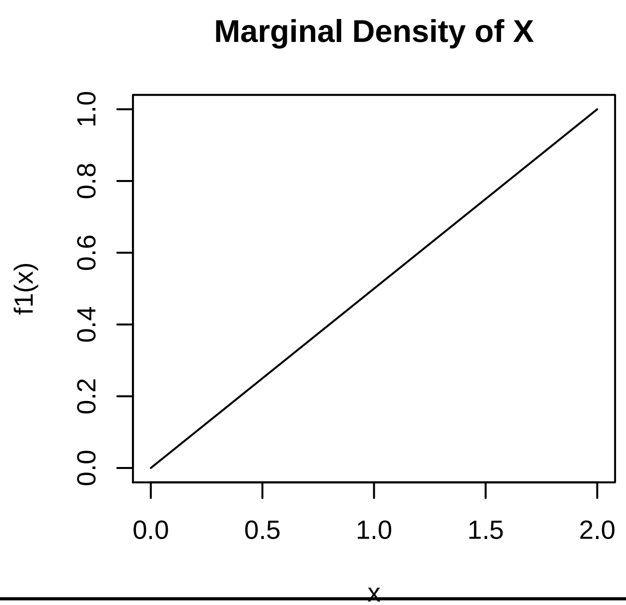
Example 18 again Suppose (X, Y) have the joint pdf of Example 18. Find the marginal pdfs of X and Y.

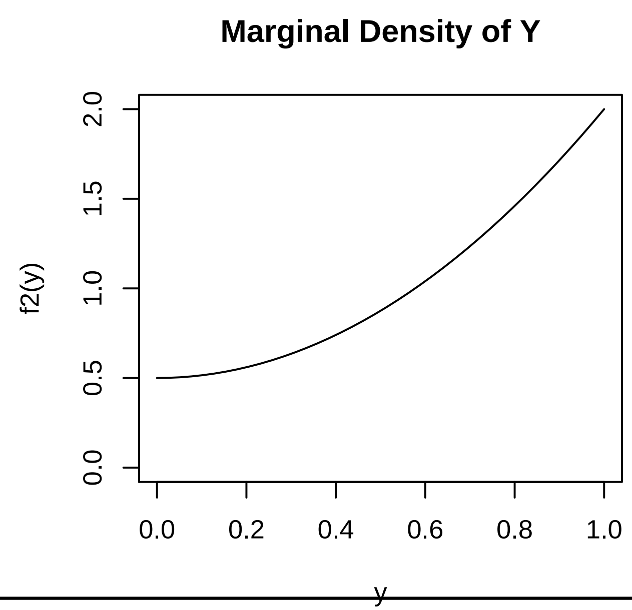
$$\begin{aligned} f_X(x) &= \int_0^1 \tfrac{x(1+3y^2)}{4} dy \\ &= \tfrac{xy}{4} + \tfrac{xy^3}{4} \Big|_{y=0}^{y=1} \\ &= \tfrac{x}{4} + \tfrac{x}{4} \\ &= \tfrac{2x}{4} \\ &= \tfrac{x}{2} \qquad 0 < x < 2 \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_0^2 \tfrac{x(1+3y^2)}{4} dx \\ &= \tfrac{x^2}{8} + \tfrac{3x^2y^2}{8} \Big|_{x=0}^{x=2} \\ &= \tfrac{4}{8} + \tfrac{12y^2}{8} \\ &= \tfrac{1}{2} + \tfrac{3y^2}{2} \\ &= \tfrac{1+3y^2}{2} \qquad 0 < y < 1 \end{aligned}$$


Bivariate Distribution
f(x,y)
y
x


Bivariate Distribution
0.2
0.4
0.6
0.8
1
1.2
1.4
0.0
0.5
1.0
1.5
2.0
x
0.0
0.2
0.4
0.6
0.8
1.0
y


Marginal Density of X
f1(x)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
x

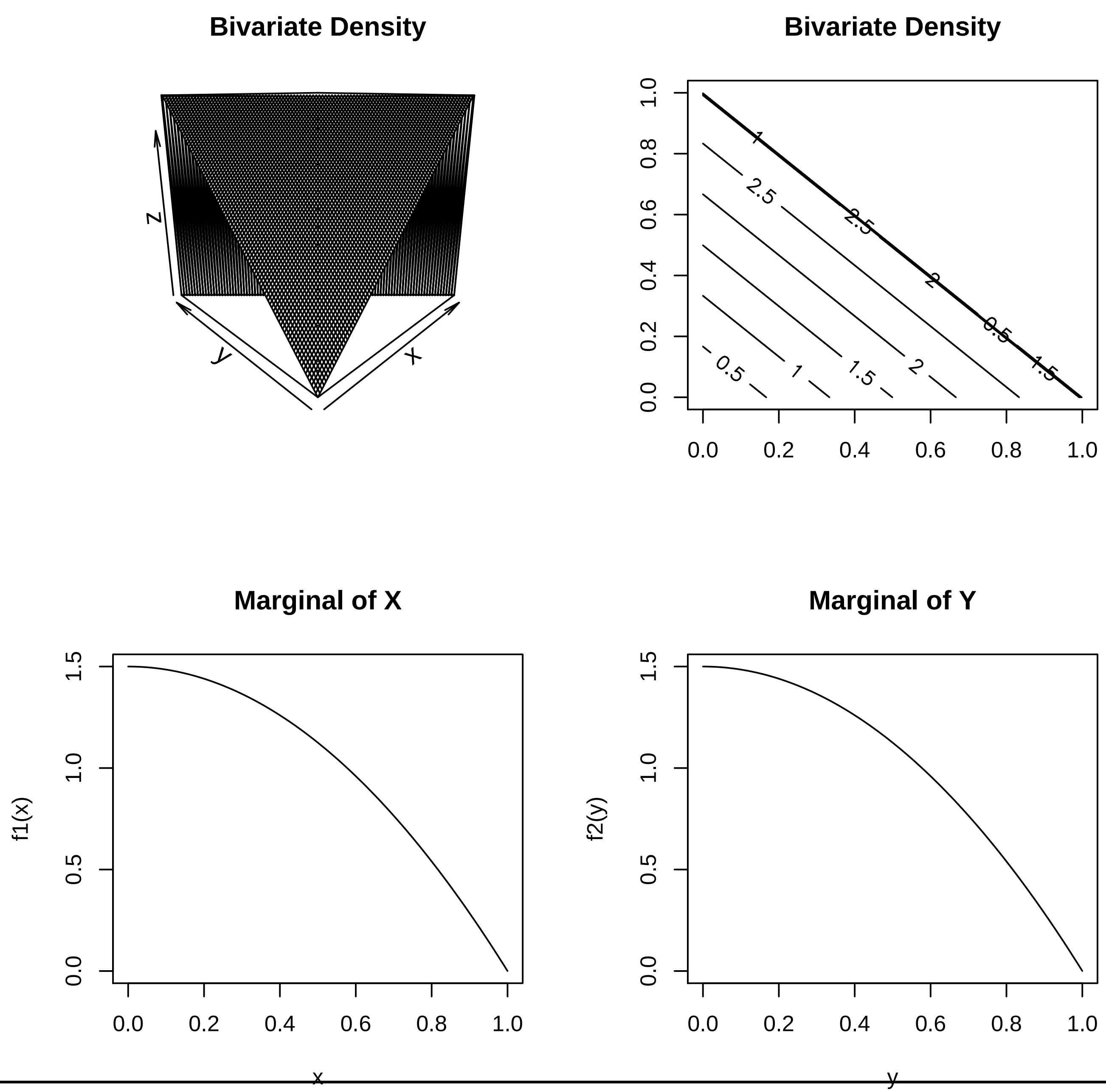

Marginal Density of Y
f2(y)
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
y

Example 19 again Suppose (X, Y) have the joint pdf of Example 19. Find the marginal pdfs of X and Y. If $x < 0$ or $x > 1$, $f_X(x) = 0$. Let $0 \le x \le 1$.

$$\begin{aligned} f_X(x) &= \int_0^{1-x} 3(x+y)\,dy \\ &= 3(xy + y^2/2)\Big|_0^{1-x} \\ &= 3[x(1-x) + (1-x)^2/2] \\ &= (3/2)(1-x^2). \end{aligned}$$

What about the marginal density of Y? Do we need to work to find it?

$f_Y(y) = 3/2\,(1-y^2)$


Bivariate Density
z
y
x
Bivariate Density
0.5
1
1.5
2
2.5
0.0 0.2 0.4 0.6 0.8 1.0
Marginal of X
f1(x)
x
0.0 0.5 1.0 1.5
0.0 0.2 0.4 0.6 0.8 1.0
Marginal of Y
f2(y)
y
0.0 0.5 1.0 1.5
0.0 0.2 0.4 0.6 0.8 1.0

7.2.3 The Bivariate Normal Distribution

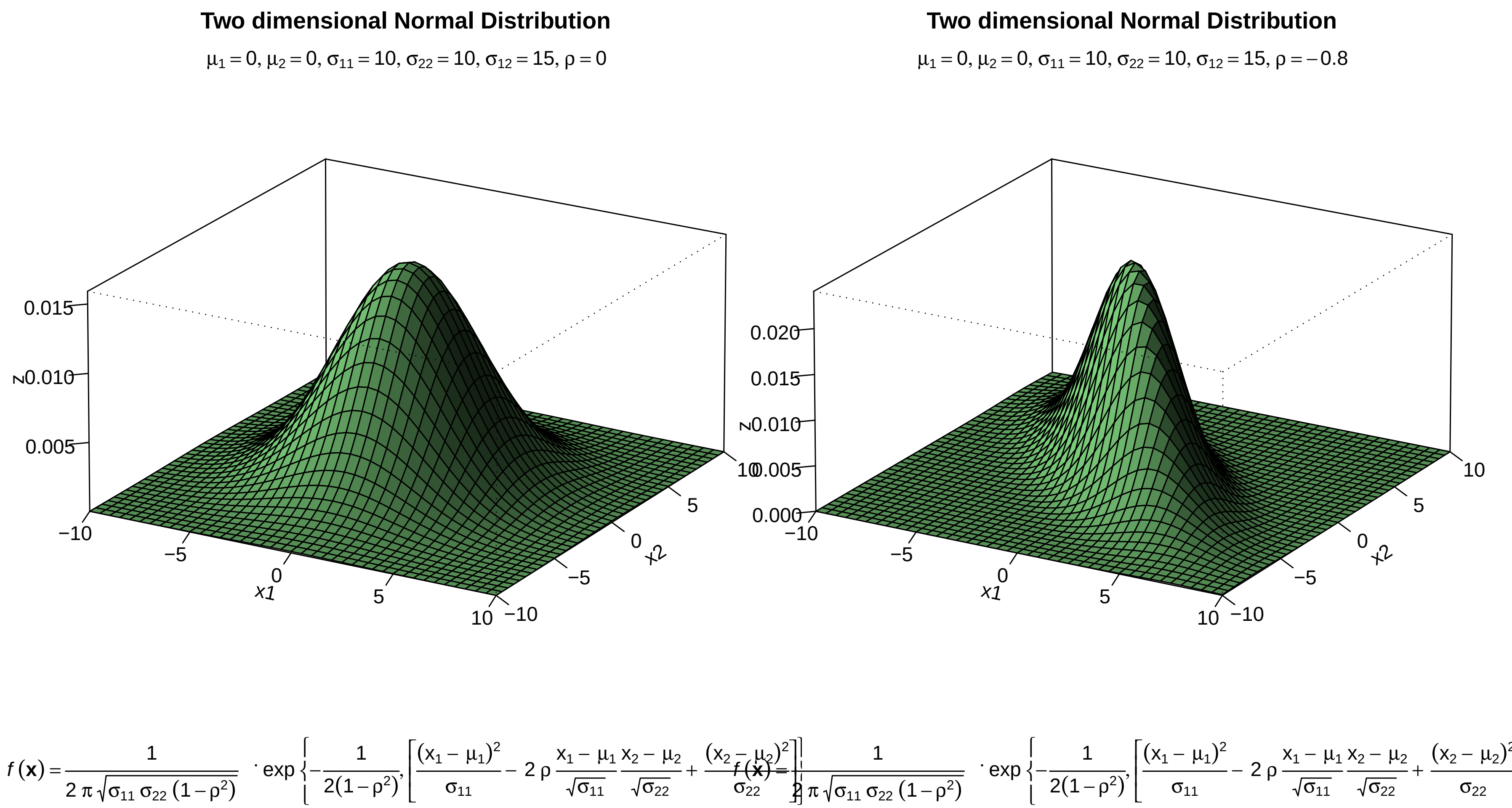
The random variables (X, Y) have a bivariate normal distribution if they have joint pdf

$$\begin{aligned} f_{X,Y}(x,y) \quad = \quad & \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}\right.\right. \\ & \left.\left. -\ 2\rho\frac{(x-\mu_X)}{\sigma_X}\frac{(y-\mu_Y)}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \end{aligned}$$

Properties:

The marginal distributions of X and Y are $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively.

If (X, Y) have a bivariate normal distribution, then $\rho = 0$ implies that X and Y are independent.



Two dimensional Normal Distribution
$\mu_1 = 0, \mu_2 = 0, \sigma_{11} = 10, \sigma_{22} = 10, \sigma_{12} = 15, \rho = 0$
Two dimensional Normal Distribution
$\mu_1 = 0, \mu_2 = 0, \sigma_{11} = 10, \sigma_{22} = 10, \sigma_{12} = 15, \rho = -0.8$
z
x1
x2
$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_{11}} - 2\rho\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}}\right]\right\}$

7.3 Independent Random Variables

Let X and Y be two random variables having some joint probability distribution. We say that X and Y are independent if and only if

$$P[(X \in A) \cap (Y \in B)] = P(X \in A)P(Y \in B)$$

for every two sets A and B of real numbers.

The notion of independence of two rvs is thus closely connected to the independence of events. X and Y are independent if and only if the events "$X \in A$" and "$Y \in B$" are independent for every A and B.

> *Intuition:* When X and Y are independent, neither random variable contains information about the other. So, knowing the value of X does not help one predict the value of Y, and vice versa.

Now let the events $A^{*}= [X \leq x]$ and $B^{*}= [Y \leq y]$. If X and Y are independent, then the joint cdf satisfies

$$\begin{aligned} F_{X,Y}(x,y) &= P[X \leq x, Y \leq y] \\ &= P(A \cap B) \\ &= P(A) \times P(B) \\ &= P[X \leq x] \times P[Y \leq y] = F_X(x)F_Y(y) \end{aligned}$$

for all values of x and y. It is also true that if the above equality holds for all (x, y), then X and Y are independent.

STOP 9/27/21

It turns out there are easier ways to check for independence.

Suppose X and Y are discrete with joint pmf $p_{X,Y}$ and marginal pmfs p_X and p_Y, respectively. Then X and Y are independent (as defined on the previous page) if and only if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{for all } (x,y).$$

Also, when X and Y are continuous with joint pdf f and respective marginal densities f_X and f_Y, then X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{for all } (x,y).$$

We will check whether X and Y are independent in Examples 17, 18, and 19.

7.4 Conditional Distributions

When we have two random variables X and Y, it is often of interest to know how one of the variables behaves when the other is known to take on some particular value.

For example, let X and Y be the age and weight, respectively, of a randomly selected child. In this case we might wish to know the distribution of weight among children having a given age, say nine years.

This leads us to a definition of the *conditional distribution* of Y given X. First we consider the case where both X and Y are discrete.

7.4.1 Conditional Distributions for Discrete RVs

Let X and Y have joint pmf p_{XY} and let X have marginal pmf p_X. If $p_X(x) > 0$, the conditional pmf of Y given that $X = x$ is

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= \frac{P[(X = x) \cap (Y = y)]}{P(X = x)} \\ &= \frac{p_{X,Y}(x, y)}{p_X(x)}. \end{aligned}$$

If p_Y is the marginal pmf of Y and $p_Y(y) > 0$, then the conditional pmf of X given that $Y = y$ is

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Consider $p_{Y|X}(y|x)$. It's important to understand that for each fixed x, $p_{Y|X}$ is a function of y. In general, if $x_1 \neq x_2$, then usually the function $p_{Y|X}(\cdot|x_1)$ will be different from $p_{Y|X}(\cdot|x_2)$.

We can rewrite the definition of conditional pmf as

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x).$$

We sum both sides of the first equality over all y to get

$$p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y).$$

Similary, we obtain

$$p_Y(y) = \sum_x p_{Y|X}(y|x)p_X(x).$$

Example 12 again Obtain the conditional pmf of X given $Y = y$.

We earlier found the marginal pmf of Y:

y	0	1	2	3	4
$p_Y(y)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

Then the conditional pmf of X given $Y = 0$ is

$$p_{X|Y}(0|0) = \frac{1/16}{1/16} = 1, \qquad x = 0.$$

The conditional pmf of X given $Y = 1$ is

$$p_{X|Y}(x|1) = \begin{cases} \frac{3/16}{4/16} = \frac{3}{4}, & x = 0 \\ \frac{1/16}{4/16} = \frac{1}{4}, & x = 1 \end{cases}$$

The conditional pmf of X given $Y = 2$ is

$$p_{X|Y}(x|2) = \begin{cases} \frac{3/16}{6/16} = \frac{1}{2}, & x = 0 \\ \frac{2/16}{6/16} = \frac{1}{3}, & x = 1 \\ \frac{1/16}{6/16} = \frac{1}{6} & x = 2. \end{cases}$$

We can use similar reasoning when $Y = 3, 4$.

Example 20 Let p be a constant in $(0, 1)$, and suppose X and Y have the following joint pmf:

$$p_{X,Y}(x, y) = \binom{y}{x} p^x(1-p)^{y-x} \cdot y^2/203, \quad x = 0, 1, \ldots, y;\ y = 2, \ldots, 8,$$

and 0 otherwise. Find the conditional pmf of X given $Y = y$, where y is a number in $\{2, 3, \ldots, 8\}$. (Why are other values of y not relevant?)

Let y be in $\{2, 3, \ldots, 8\}$. The marginal pmf of Y is

$$\begin{aligned} p_Y(y) &= \sum_{x=0}^{y} p(x, y) = (y^2/203) \sum_{x=0}^{y} \binom{y}{x} p^x(1-p)^{y-x} \\ &= (y^2/203)(p + 1 - p)^y = y^2/203. \end{aligned}$$

Therefore,

$$p_{X|Y}(x|y) = \binom{y}{x} p^x(1-p)^{y-x}, \quad x = 0, \ldots, y,$$

which is a binomial distribution with number of trials y and success probability p.

7.5 Conditional Densities for Continuous RVs

Now suppose X and Y are continuous rvs with joint pdf $f_{X,Y}$. If f_Y is the marginal pdf of Y and $f_Y(y) > 0$, then the conditional pdf of X given that $Y = y$ is

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

Likewise, when $f_X(x) > 0$, the conditional density of Y given $X = x$ is

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

We can reexpress these equations:

$$\begin{aligned} f_{X,Y}(x,y) &= f_{X|Y}(x|y)f_Y(y), \\ f_{X,Y}(x,y) &= f_{Y|X}(y|x)f_X(x). \end{aligned}$$

For continuous rvs, we have then

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

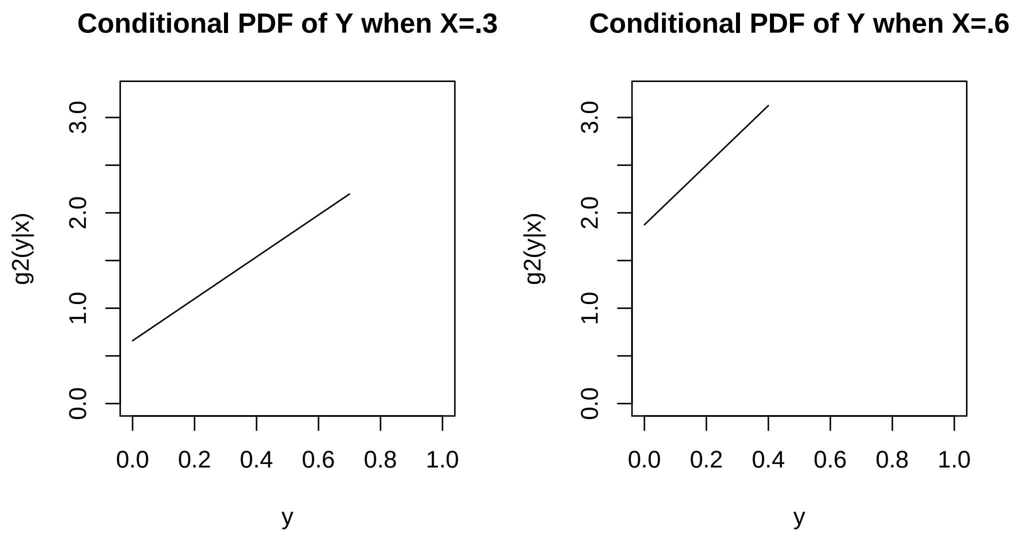
Example 19 again Let (X, Y) have the joint distribution given in Example 19. Find the conditional pdf of Y given $X = x$ for some $0 < x < 1$.

The marginal pdf of X is $f_X(x) = (3/2)(1 - x^2)$ for $0 < x < 1$.

The conditional pdf of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{3(x+y)}{(3/2)(1-x^2)} = \frac{2(x+y)}{1-x^2}$$

where $0 < y < 1 - x$.



7.6 Checking Independence via Conditional Distributions

Recall that X and Y are independent if and only if their joint pmf or pdf is the product of the two marginal pmfs or pdfs. We can also check for independence by seeing if one of the two conditional distributions equals its corresponding marginal distribution.

Discrete rvs X and Y are independent if and only if at each y such that $p_Y(y) > 0$,

$$p_{X|Y}(x|y) = p_X(x) \quad \text{for all } x.$$

Continuous rvs X and Y are independent if and only if at each y such that $f_Y(y) > 0$,

$$f_{X|Y}(x|y) = f_X(x) \quad \text{for all } x.$$

This parallels a similar result for events A and B. Suppose that $P(B) > 0$. Then A and B are independent if and only if

$$P(A|B) = P(A).$$

Example 18 again Suppose that (X, Y) have the joint pdf:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0 < x < 2, \quad 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The marginal pdf of X is $f_X(x) = \frac{x}{2}$, $0 < x < 2$. The conditional pdf of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{x(1+3y^2)}{4}}{\frac{x}{2}} = \frac{1+3y^2}{2} = f_Y(y).$$

Thus, X and Y are independent.

START MON 9/27/21

7.7 Extrema and Order Statistics

We are often use order statistics in statistical analysis. Let $X_1, \ldots, X_n$ be independent random variables from a continuous distribution with cdf F and pdf f. The order statistics are the values of $X_1, \ldots, X_n$ arranged in increasing order and are denoted by

$$\text{MIN} = X_{(1)} < X_{(2)} < \cdots < X_{(n)}. = \text{MAX}$$

We first find the cdf of the minimum $V = X_{(1)}$:

$$\begin{aligned} F_V(v) &= P(V \le v) = 1 - P(V > v) = 1 - P(X_1 > v, \ldots, X_n > v) \\ &= 1 - \prod_{i=1}^{n} P(X_i > v) = 1 - (1 - F_X(v))^n \end{aligned}$$

equality b/c the X_i have the same distribution

We differentiate to find the pdf:

$$f_V(v) = F_V'(v) = \frac{d}{dv}[1 - (1 - F(v))^n] = n(1 - F(v))^{n-1} f(v)$$

Similarly we first find the cdf of the maximum $U = X_{(n)}$:

$$\begin{aligned} F_U(u) &= P(U \le u) = P(X_1 \le u, X_2 \le u, \ldots, X_n \le u) \\ f_v(u) &= \prod_{i=1}^{n} P(X_i \le u) = [F_X(u)]^n \end{aligned}$$

by equality b/c X_i come from the same dist.

We differentiate to find the pdf:

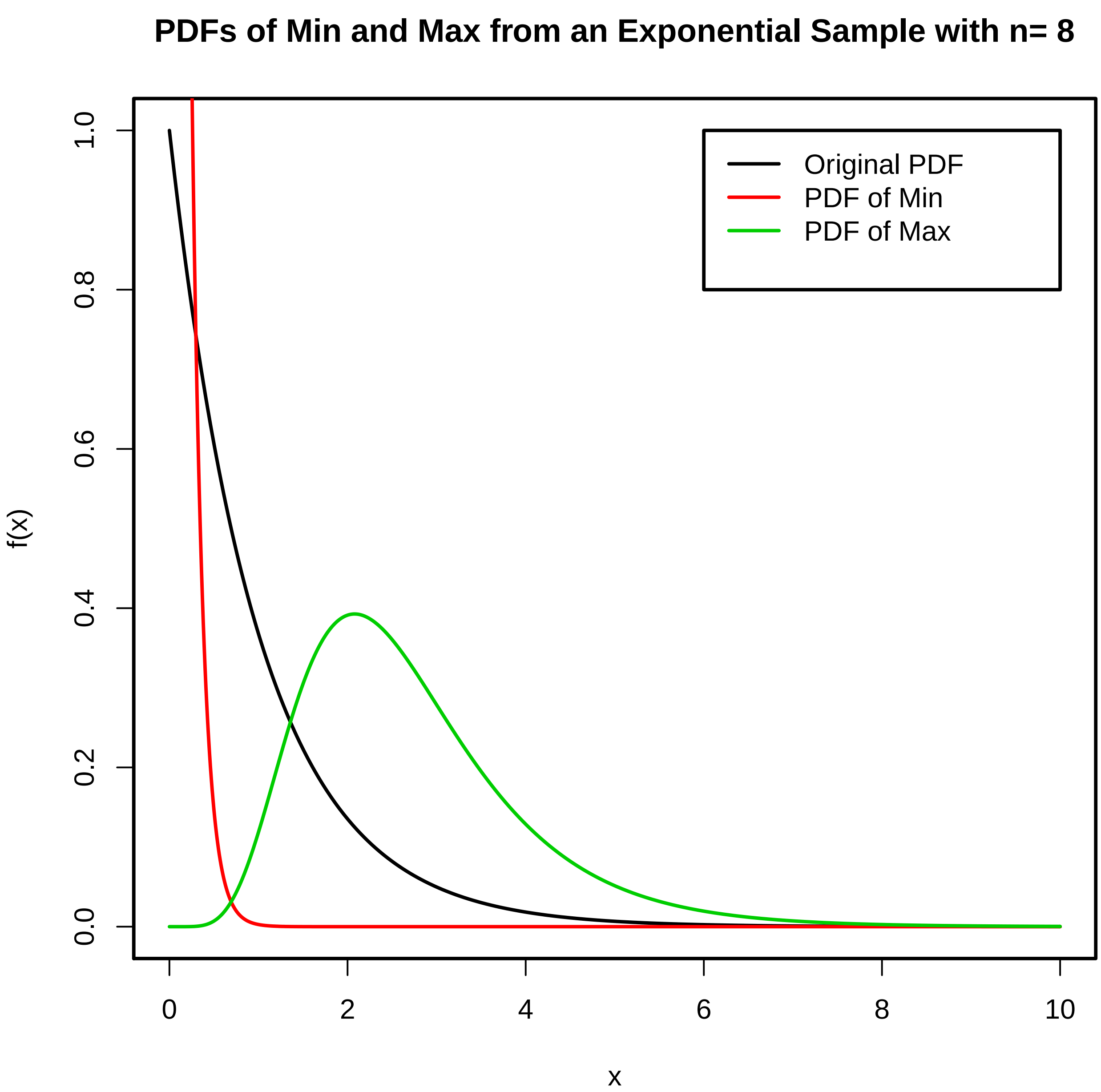
Exponential w/ coefficient of n constant & coefficient of 1.

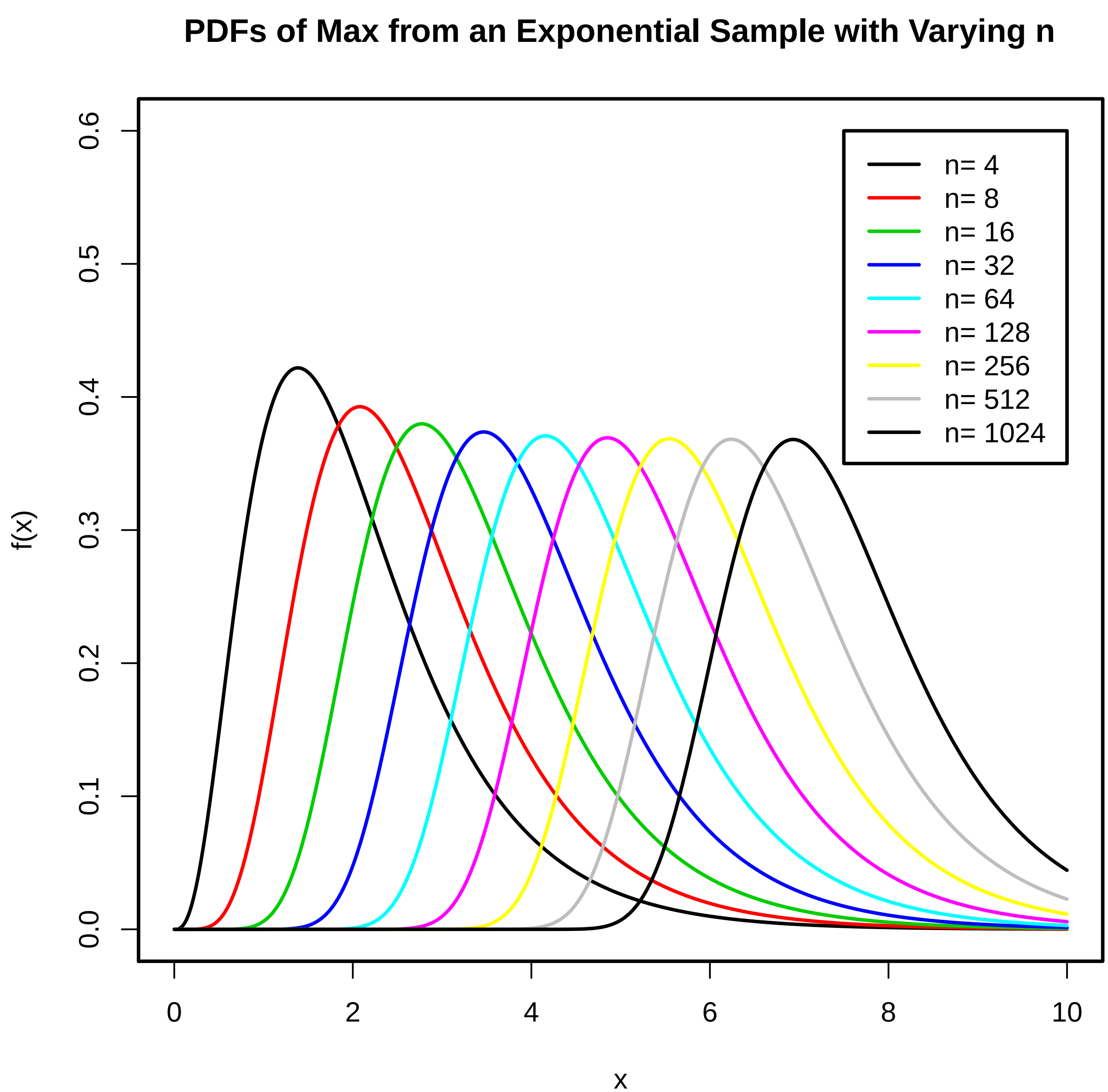
$$f_U(u) = F_U'(u) = \frac{d}{du}[F(u)]^n = n[F(u)]^{n-1} f(u)$$

Example: Let $X_1, \ldots, X_n$ be independent exponential(1) rvs. Find the pdfs of $V = \min\{X_1, \ldots, X_n\}$ and $U = \max\{X_1, \ldots, X_n\}$.

Here $f_X(x) = e^{-x}$ and $F(x) = 1 - e^{-x}$ for $x > 0$. Then

$$\begin{aligned} f_V(v) &= n[e^{-v}]^{n-1} e^{-v} = ne^{-nv}, \; v > 0 \\ f_U(u) &= n[1 - e^{-u}]^{n-1} e^{-u}, \; u > 0 \end{aligned}$$


PDFs of Min and Max from an Exponential Sample with n= 8
Original PDF
PDF of Min
PDF of Max
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10


PDFs of Max from an Exponential Sample with Varying n
f(x)
x
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
2
4
6
8
10
n= 4
n= 8
n= 16
n= 32
n= 64
n= 128
n= 256
n= 512
n= 1024

7.8 Convolutions

Sums of independent random variables play an important role in statistics. Suppose X and Y are independent random variables. What is the distribution of $Z = X + Y$?

7.8.1 Convolutions of Discrete RVs

Suppose that X and Y are independent discrete rvs with pmfs p_X and p_Y, respectively. Let $Z = X + Y$. Then

$$\begin{aligned} p_Z(z) &= P(Z = z) = P(X + Y = z) \\ &= \sum_{\{(x,y):x+y=z\}} p_X(x)p_Y(y) \\ &= \sum_y p_X(z-y)p_Y(y) = \sum_x p_X(x)\, p_Y(z-x) \end{aligned}$$

7.8.2 Convolutions of Continuous RVs

We will derive the distribution by finding the cdf and differentiating:

$$
\begin{aligned}
F_Z(z) &= P(X+Y \le z) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y)\, dx dy \\
&= \int_{-\infty}^{\infty} f_Y(y) \left[\int_{-\infty}^{z-y} f_X(x)\, dx \right] dy \\
&= \int_{-\infty}^{\infty} f_Y(y) F_X(z-y)\, dy.
\end{aligned}
$$

$x + y \le z$

$\Rightarrow x \le z - y$

F_Z is the cdf of $X + Y$. To find the density, differentiate F_Z:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \frac{d}{dz}\int_{-\infty}^{\infty} f_Y(y)F_X(z-y)\,dy$$
$$= \int_{-\infty}^{\infty} f_Y(y)\frac{dF_X(z-y)}{dz}\,dy$$
$$= \int_{-\infty}^{\infty} f_Y(y)f_X(z-y)\,dy.$$

if we didnt assume independence are iid replace this product w/ joint pdf

similar to * Σ on pg 101

The function f_Z defined by the very last integral is said to be the convolution of f_X and f_Y.

Remark: If we reverse the roles of X and Y in the above argument, we obtain the expression:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$$

Example 23 Suppose X and Y are independent and identically distributed, with each having density

$$f(x) = e^{-x} I_{(0,\infty)}(x),$$

which is an *exponential* density. Find the density, f_Z, of $Z = X + Y$. We know that $f_Z(z) = 0$ for $z \le 0$. Why? Let $z > 0$. $x > 0$, $z - x > 0$, $x < z$

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \\ &= \int_0^{\infty} e^{-x} e^{-(z-x)} I_{(0,\infty)}(z-x)\,dx \\ &= e^{-z} \int_0^z dx = ze^{-z}. \end{aligned}$$

So, the density of $X + Y$ is $f_Z(z) = ze^{-z} I_{(0,\infty)}(z)$, which is a special case of the *gamma density*.