Topic 7: Random Variables and Distribution Functions*

September 22 and 27, 2011

1 Introduction

From the universe of possible information, we ask a question. To address this question, we might collect quantitative data and organize it, for example, using the empirical cumulative distribution function. With this information, we are able to compute sample means, standard deviations, medians and so on.

Similarly, even a fairly simple probability model can have an enormous number of outcomes. For example, flip a coin 332 times. Then the number of outcomes is more than a google (10^{100}) – a number 100 quintillion times the number of elementary particles in the known universe. We may not be interested in an analysis that considers separately every possible outcome but rather some simpler concept like the number of heads or the longest run of tails. To focus our attention on the issues of interest, we take a given outcome and compute a number. This function is called a **random variable**.

statistics	**probability**
universe of information	sample space - Ω and probability - P
$\Downarrow$	$\Downarrow$
ask a question and collect data	define a random variable X
$\Downarrow$	$\Downarrow$
organize into the empirical cumulative distribution function	organize into the cumulative distribution function
$\Downarrow$	$\Downarrow$
compute sample means and variances	compute distributional means and variances

Table I: Corresponding notions between statistics and probability. Examining probabilities models and random variables will lead to strategies for the collection of data and inference from these data.

Definition 1. *A* **random variable** *is a real valued function from the sample space.*

$$X : \Omega \to \mathbb{R}.$$

Generally speaking, we shall use capital letters near the end of the alphabet, e.g., X, Y, Z for random variables. The range of a random variable is sometimes called the **state space**.

Exercise 2. *Roll a die twice and consider the sample space* $\Omega = \{(i, j); i, j = 1, 2, 3, 4, 5, 6\}$ *and give some random variables on* Ω.

Exercise 3. *Flip a coin 10 times and consider the sample space* Ω*, the set of 10-tuples of heads and tails, and give some random variables on* Ω.

We often create new random variables via composition of functions:

$$\omega \to X(\omega) \to f(X(\omega))$$

Thus, if X is a random variable, then so are

$$X^2, \quad \exp \alpha X, \quad \sqrt{X^2+1}, \quad \tan^2 X \quad \lfloor X \rfloor$$

and so on. The last of these, rounding down X to the nearest integer, is called the **floor function**.

Exercise 4. *How would we use the floor function to round down a number x to n decimal places.*

2 Distribution Functions

Having define a random variable of interest, X, the question typically becomes, "What are the chances that X lands in some subset of values A?" For example,

$$A = \{\text{odd numbers}\}, \quad A = \{\text{greater than } 1\}, \quad \text{or} \quad A = \{\text{between 2 and 7}\}.$$

We write

$$\{\omega \in \Omega; X(\omega) \in A\} \tag{1}$$

to indicate those outcomes ω which have $X(\omega)$, the value of the random variable, in the subset A. We shall often abbreviate (1) to the shorter statement $\{X \in A\}$. Thus, for the example above, we may write the events

$$\{X \text{ is an odd number}\}, \quad \{X \text{ is greater than } 1\} = \{X > 1\}, \quad \{X \text{ is between 2 and 7}\} = \{2 < X < 7\}$$

to correspond to the three choices above for the subset A.

Many of the properties of random variables are not concerned with the specific random variable X given above, but rather depends on the way X *distributes its values*. This leads to a definition in the context of random variables that we saw previously with quantitive data..

Definition 5. *A* **(cumulative) distribution function** *of a random variable X is defined by*

$$F_X(x) = P\{\omega \in \Omega; X(\omega) \le x\}.$$

Recall that with a data set, we called the analogous notion the *empirical cumulative distribution function*. Using the abbreviated notation above, we shall typically write the less explicit expression

$$F_X(x) = P\{X \le x\}$$

for the distribution function.

Exercise 6. *Show that*

1. $\{X \in B\}^c = \{X \in B^c\}$
2. *For sets $B_1, B_2, \ldots,$*

$$\bigcup_i \{X \in B_i\} = \{X \in \bigcup_i B\}.$$

For the complement of $\{X \le x\}$, we have the **survival function**

$$\bar{F}_X(x) = P\{X > x\} = 1 - P\{X \le x\} = 1 - F_X(x).$$

Choose $a < b$, then the event $\{X \le a\} \subset \{X \le b\}$. Their set theoretic difference

$$\{X \le b\} \setminus \{X \le a\} = \{a < X \le b\}.$$

In words, the event that X is less than or equal to b but not less than or equal to a is the event that X is greater than a and less than or equal to b. Consequently, by the difference rule for probabilities,

$$P\{a < X \le b\} = P(\{X \le b\} \setminus \{X \le a\}) = P\{X \le b\} - P\{X \le a\} = F_X(b) - F_X(a).$$

Thus, we can compute the probability that a random variable takes values in an interval by subtracting the distribution function evaluated at the endpoints of the intervals. Care is needed on the issue of the inclusion or exclusion of the endpoints of the interval.

Example 7. *To give the cumulative distribution function for X, the sum of the values for two rolls of a die, we start with the table*

x	2	3	4	5	6	7	8	9	10	11	12
$P\{X = x\}$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

and create the graph.

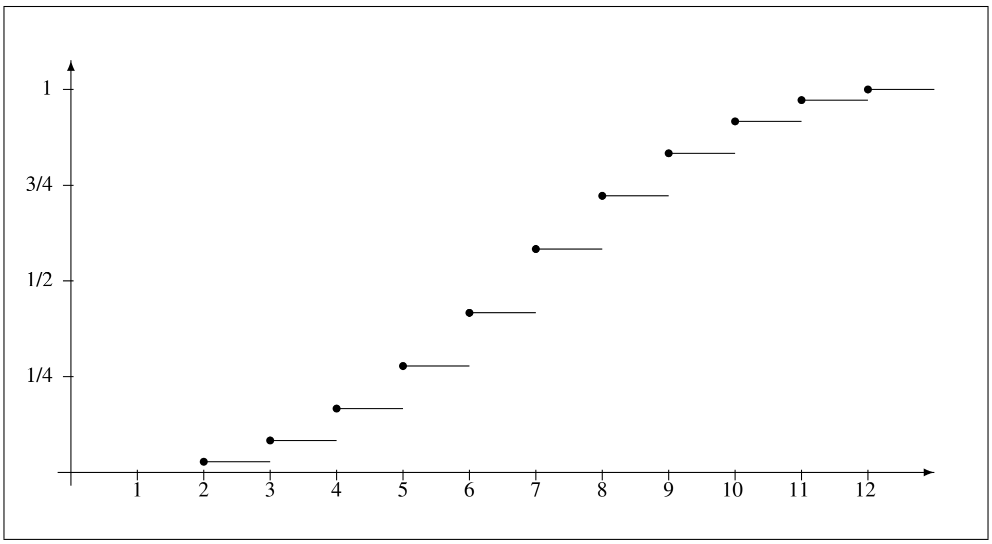


Figure 1: Graph of F_X, the cumulative distributtion function for the sum of the values for two rolls of a die.

If we look at the graph of this cumulative distribution function, we see that it is constant in between the possible values for X and that the jump size at x is equal to $P\{X = x\}$. In this example, $P\{X = 5\} = 4/36$, the size of the jump at $x = 5$. In addition,

$$F_X(5) - F_X(2) = P\{2 < X \le 5\} = P\{X = 3\} + P\{X = 4\} + P\{X = 5\} = \sum_{2<x\le 5} P\{X = x\}$$
$$= \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{9}{36}.$$

We shall call a random variable **discrete** if it has a finite or countably infinite state space. Thus, we have in general that:

$$P\{a < X \le b\} = \sum_{a<x\le b} P\{X = x\}.$$

Exercise 8. *Let X be the number of heads on three independent flips of a biased coin that turns ups heads with probability p. Give the cumulative distribution function F_X for X. Use* R *to give a plot of F_X.*

Exercise 9. *Let X be the number of spades in a collection of three cards. Give the cumulative distribution function for X. Use* R *to plot this function.*

Exercise 10. *Find the cumulative distribution function of $Y = X^3$ in terms of F_X, the distribution function for X.*

3 Properties of the Distribution Function

A distribution function F_X has the following properties:

1. F_X is nondecreasing.
   Let $x_1 < x_2$, then $\{X \le x_1\} \subset \{X \le x_2\}$ and by the monotonicity rule for probabilities

   $$P\{X \le x_1\} \le P\{X \le x_2\}, \quad \text{or written in terms of the distribution function,} \quad F_X(x_1) \le F_X(x_2)$$

2. $\lim_{x\to\infty} F_X(x) = 1$.
   Let $x_n \to \infty$ be an increasing sequence. Then $x_1 < x_2 < \cdots$

   $$\{X \le x_1\} \subset \{X \le x_2\} \subset \cdots .$$

   Thus,

   $$P\{X \le x_1\} \le P\{X \le x_2\} \le \cdots .$$

   For each outcome ω, eventually, for some n, $X(\omega) \le x_n$, and

   $$\bigcup_{n=1}^{\infty} \{X \le x_n\} = \Omega.$$

   Now, use the first continuity property of probabilities.

3. $\lim_{x\to-\infty} F_X(x) = 0$.
   Let $x_n \to -\infty$ be a decreasing sequence. Then $x_1 > x_2 > \cdots$

   $$\{X \le x_1\} \supset \{X \le x_2\} \supset \cdots$$

   Thus,

   $$P\{X \le x_1\} \ge P\{X \le x_2\} \ge \cdots$$

   For each outcome ω, eventually, for some n, $X(\omega) \le x_n$, and

   $$\bigcap_{n=1}^{\infty} \{X \le x_n\} = \emptyset.$$

   Now, use the second continuity property of probabilities.

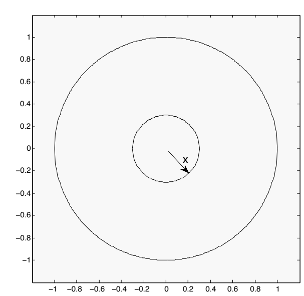


Figure 2: Dartboard.

The cumulative distribution function F_X of a discrete random variable X is constant except for jumps. At the jump, F_X is **right continuous**,

$$\lim_{x\to x_0+} F_X(x) = F_X(x_0).$$

Exercise 11. *Prove the statement concerning the right continuity of the distribution function from the continuity property of a probability.*

Definition 12. *A* **continuous random variable** *has a cumulative distribution function F_X that is differentiable.*

So, distribution functions for continuous random random variables increase smoothly. To show how this can occur, we will develop an example of a continuous random variable.

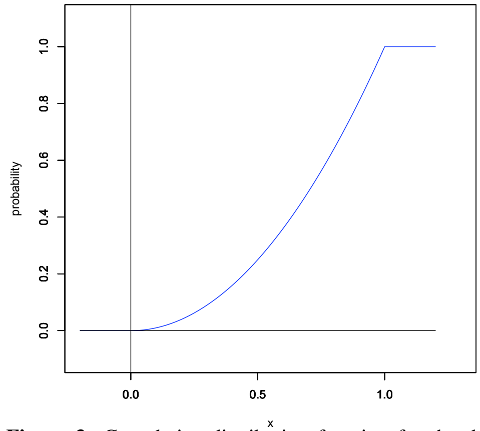


Figure 3: Cumulative distribution function for the dartboard random variable.

Example 13. *Consider a dartboard having unit radius. Assume that the dart lands randomly uniformly on the dartboard.*

Let X be the distance from the center. For $x \in [0,1]$,

$$F_X(x) = P\{X \leq x\} = \frac{\text{area inside circle of radius } x}{\text{area of circle}} = \frac{\pi x^2}{\pi 1^2} = x^2.$$

Thus, we have the distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x^2 & \text{if } 0 < x \leq 1, \\ 1 & \text{if } x > 1. \end{cases}$$

The first line states that X cannot be negative. The third states that X must be below 1, and the middle lines describes how X distributes is values between 0 and 1. For example,

$$F_X\left(\frac{1}{2}\right) = \frac{1}{4}$$

indicates that with probability 1/4, the dart will land within 1/2 unit of the center of the dartboard.

Exercise 14. *An* **exponential random variable** X *has cumulative distribution function*

$$F_X(x) = P\{X \leq x\} = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 - \exp(-\lambda x) & \text{if } x > 0 \end{cases}$$

for some $\lambda > 0$. Show that F_X has the properties of a distribution function.

Its value at x can be computed in R using the command `pexp(x,0.1)` for $\lambda = 1/10$ and drawn using

```
> curve(pexp(x,0.1),0,80)
```

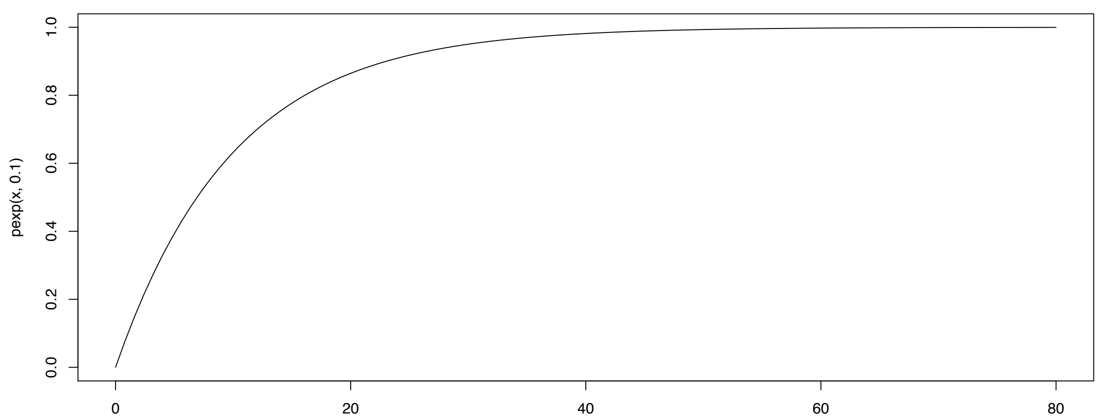


Figure 4: Cumulative distribution function for an exponential random variable with $\lambda = 1/10$.

Exercise 15. *The time until the next bus arrives is an exponential random variable with $\lambda = 1/10$ minutes. A person waits for a bus at the bus stop until the bus arrives, giving up if when the wait reaches 20 minutes. Give the cumulative distribution function for T the time that the person remains at the bus station and sketch a graph.*

Even though the cumulative distribution function is defined for every random variable, we will often use other characterizations, namely, the mass function for discrete random variable and the density function for continuous random variables. Indeed, we typically will introduce a random variable via one of these two functions. In the next two sections we introduce these two concepts and develop some of their properties.

4 Mass Functions

Definition 16. *The* **(probability) mass function** *of a discrete random variable X is*

$$f_X(x) = P\{X = x\}.$$

The mass function has two basic properties:

- $f_X(x) \ge 0$ for all x in the state space.
- $\sum_x f_X(x) = 1$.

Example 17. *Let's make tosses of a biased coin whose outcomes are independent . We shall continue tossing until we obtain a toss of heads. Let X denote the random variable that gives the number of tails before the first head. Let p denote the probability of heads in any given toss. Then*

$$\begin{aligned} f_X(0) &= P\{X = 0\} = P\{H\} = p \\ f_X(1) &= P\{X = 1\} = P\{TH\} = (1-p)p \\ f_X(2) &= P\{X = 2\} = P\{TTH\} = (1-p)^2 p \\ &\vdots \qquad\quad \vdots \qquad\quad \vdots \\ f_X(x) &= P\{X = x\} = P\{T\cdots TH\} = (1-p)^x p \end{aligned}$$

So, the probability mass function $f_X(x) = (1-p)^x p$. Because the terms in this mass function form a geometric sequence, X is called a **geometric random variable***. Recall that a geometric sequence $c, cr, cr^2, \ldots, cr^n$ has sum*

$$s_n = c + cr + cr^2 + \cdots + cr^n = \frac{c(1-r^{n+1})}{1-r}$$

for $r \neq 1$. If $|r| < 1$, then the infinite sum

$$c + cr + cr^2 + \cdots + cr^n + \cdots = \lim_{n\to\infty} s_n = \frac{c}{1-r}.$$

In this situation the ratio $r = 1-p$. Consequently, for positive integers a and b,

$$\begin{aligned} P\{a < X \le b\} &= \sum_{x=a+1}^{b} (1-p)^x p = (1-p)^{a+1} p + \cdots + (1-p)^b p \\ &= \frac{(1-p)^{a+1}p - (1-p)^{b+1}p}{1-(1-p)} = (1-p)^{a+1} - (1-p)^{b+1} \end{aligned}$$

Exercise 18. *Establish the formula above for s_n.*

The mass function and the cumulative distribution function for the geometric random variable with parameter $p = 2/3$ can be found in R by writing

```
> x<-c(0:10)
> f<-dgeom(x,1/3)
> F<-pgeom(x,1/3)
```

The initial `d` indicates **d**ensity and `p` indicates the **p**robability from the distribution function.

```
> data.frame(x,f,F)
    x           f         F
1   0 0.333333333 0.3333333
2   1 0.222222222 0.5555556
3   2 0.148148148 0.7037037
4   3 0.098765432 0.8024691
5   4 0.065843621 0.8683128
6   5 0.043895748 0.9122085
7   6 0.029263832 0.9414723
8   7 0.019509221 0.9609816
9   8 0.013006147 0.9739877
10  9 0.008670765 0.9826585
11 10 0.005780510 0.9884390
```

Exercise 19. *Check that the jumps in the cumulative distribution function for the geometric random variable above is equal to the values of the mass function.*

We can simulate 100 geometric random variables with parameter $p = 1/3$ using `rgeom(100,1/3)`.

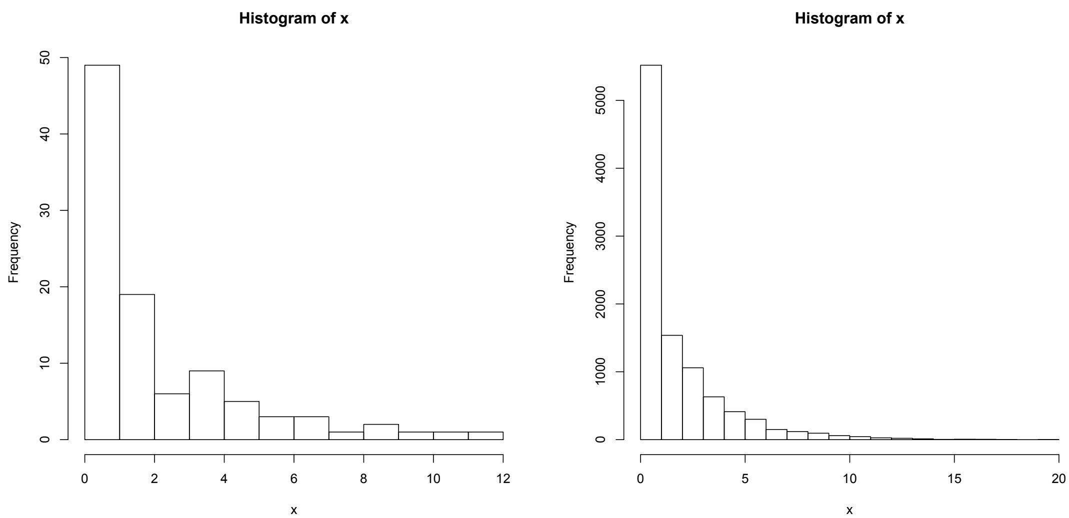


Figure 5: Histogram of 100 and 10,000 simulated geometric random variables with $p = 1/3$. Note that the histogram looks much more like a geometric series for 10,000 simulations. We shall see later how this relates to the law of large numbers.

5 Density Functions

Definition 20. *For X a random variable whose distribution function F_X has a derivative. The function f_X satisfying*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

is called the **probability density function** *and X is called a* **continuous random variable***.*

By the fundamental theorem of calculus, the density function is the derivative of the distribution function.

$$f_X(x) = \lim_{\Delta x \to 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = F'_X(x).$$

In other words,

$$F_X(x + \Delta x) - F_X(x) \approx f_X(x)\Delta x.$$

We can compute probabilities by evaluating definite integrals

$$P\{a < X \leq b\} = F_X(b) - F_X(a) = \int_a^b f_X(t)\, dt.$$

The density function has two basic properties that mirror the properties of the mass function:

- $f_X(x) \geq 0$ for all x in the state space.
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1.$

Return to the dart board example, letting X be the distance from the center of a dartboard having unit radius. Then,

$$P\{x < X \leq x+\Delta x\} = F_X(x+\Delta x) - F_X(x) \approx f_X(x)\Delta x = 2x\Delta x$$

and X has density

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 2x & \text{if } 0 \leq x \leq 1, \\ 0 & \text{if } x > 1. \end{cases}$$

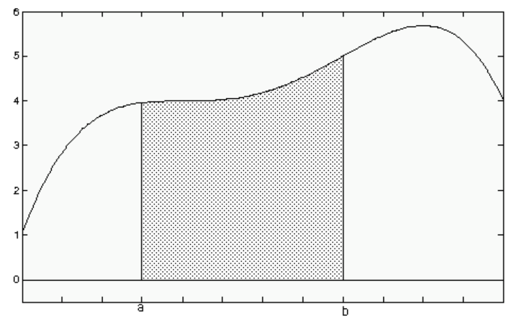


Figure 6: The probability $P\{a < X \leq b\}$ is the area under the density function, above the x axis between $y = a$ and $y = b$.

Exercise 21. *Let f_X be the density for a random variable X and pick a number x_0. Explain why $P\{X = x_0\} = 0$.*

Example 22. *Density functions do not need to be bounded, for example, if we take*

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{c}{\sqrt{x}} & \text{if } 0 < x < 1, \\ 0 & \text{if } 1 \leq x. \end{cases}$$

Then, to find the value of the constant c, we compute the integral

$$1 = \int_0^1 \frac{c}{\sqrt{t}}\, dt = 2c\sqrt{t}\Big|_0^1 = 2c.$$

So $c = 1/2$.

For $0 \leq a < b \leq 1$,

$$P\{a < X \leq b\} = \int_a^b \frac{1}{2\sqrt{t}}\, dt = \sqrt{t}\Big|_a^b = \sqrt{b} - \sqrt{a}.$$

Exercise 23. *Give the cumulative distribution function for the random variable in the previous example.*

Exercise 24. *Let X be a continuous random variable with density f_X, then the random variable $Y = aX + b$ has denisty*

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

(Hint: Begin with the definition of the cumulative distribution function F_Y for Y. Consider the cases $a > 0$ and $a < 0$ separately.)

6 Joint Distributions

Because we will collect data on several observations, we must, as well, consider more than one random variable at a time in order to model our experimental procedures. Consequently, we will expand on the concepts above to the case of multiple random variables and their joint distribution. For the case of two random variables, this means looking at the probability of

$$P\{X_1 \in A_1, X_2 \in A_2\}.$$

For discrete random variables take $A_1 = \{x_1\}$ and $A_2 = \{x_2\}$ and define the **joint probability mass function**

$$f_{X_1,X_2}(x_1, x_2) = P\{X_1 = x_1, X_2 = x_2\}.$$

For continuous random variables, we consider $A_1 = (x_1, x_1 + \Delta x_1]$ and $A_2 = (x_2, x_2 + \Delta x_2]$ and ask that for some function f_{X_1,X_2}, the **joint probability density function** to satisfy

$$P\{x_1 < X_1 \le x_1 + \Delta x_1, x_2 < X_2 \le x_2 + \Delta x_2\} \approx f_{X_1,X_2}(x_1, x_2)\Delta x_1 \Delta x_2.$$

Example 25. *Generalize the notion of mass and density functions to more than two random variables.*

6.1 Independent Random Variables

Many of our experimental protocols will be designed so that observations are independent. More precisely, we will say that two random variables X_1 and X_2 are **independent** if any two events associated to them are independent, i.e.,

$$P\{X_1 \in A_1, X_2 \in A_2\} = P\{X_1 \in A_1\}P\{X_2 \in A_2\}.$$

For discrete random variables,

$$f_{X_1,X_2}(x_1, x_2) = P\{X_1 = x_1, X_2 = x_2\} = P\{X_1 = x_1\}P\{X_2 = x_2\} = f_{X_1}(x_1)f_{X_2}(x_2).$$

The joint probability mass function is the product of the **marginal mass functions**. For continuous random variables,

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2)\Delta x_1 \Delta x_2 &\approx P\{x_1 < X_1 \le x_1 + \Delta x_1, x_2 < X_2 \le x_2 + \Delta x_2\} \\ &= P\{x_1 < X_1 \le x_1 + \Delta x_1\}P\{x_2 < X_2 \le x_2 + \Delta x_2\} \approx f_{X_1}(x_1)\Delta x_1 f_{X_2}(x_2)\Delta x_2 \\ &= f_{X_1}(x_1)f_{X_2}(x_2)\Delta x_1 \Delta x_2. \end{aligned}$$

The joint probability density function

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

is the product of the **marginal density functions**.

Exercise 26. *Generalize the notion of independent mass and density functions to more than two random variables.*

Soon, we will be looking at n independent observations $x_1, x_2, \ldots, x_n$ arising from an unknown density or mass function f. Thus, the joint density is

$$f(x_1)f(x_2)\cdots f(x_n).$$

Generally speaking, the density function f will depend on the choice of a parameter value θ. (For example, the unknown parameter in the density function for an exponential random variable that describes the waiting time for a bus.) Given the data arising from the n observations, the **likelihood function** arises by consider this joint density as a function of the variable θ. We shall learn how the study of the likelihood plays a major role in parameter estimation and in the testing of hypotheses.

Often we will explore the properties of the data through simulation. Thus, we present methods for simulating first discrete and then continuous random variables.

7 Simulating Discrete Random Variables in R

One goal for this course is to provide the tools needed to design inferential procedures based on sound principles of statistical science. Thus, one of the very important uses of statistical software is the ability to generate pseudo-data to simulate the actual data. This provides the opportunity to test and refine methods of analysis in advance of the need to use these methods on genuine data.

The `sample` command is used to create simple and stratified random samples. This is using the default R command of **sampling without replacement**. We can use this command to simulate discrete random variables. To do this, we need to give the state space in a vector `x` and a mass function `f`. Then to give a sample of n independent random variables we use `sample(x,n,replace=TRUE,prob=f)`

Example 27. *Let X be described by the mass function*

x	1	2	3	4
$f_X(x)$	0.1	0.2	0.3	0.4

Then to simulate 50 independent observations from this mass function:

```
> x<-c(1,2,3,4)
> f<-c(0.1,0.2,0.3,0.4)
> sum(f)
[1] 1
> data<-sample(x,50,replace=TRUE,prob=f)
> data
 [1] 2 4 3 3 1 3 2 4 2 4 1 3 3 3 4 3 4 1 2 3 4 4 4 4 3 4 4 4 4 1 2 4 4 4 3 3 4 4 1 2 4
[42] 3 4 4 3 4 2 3 4 3
```

Notice that 1 is the least represented value and 4 is the most represented. If the command $\hat{p}$rob=f is omitted, then `sample` *will choose uniformly from the values in the vector* `x`..

8 Probability Transform

For X a continuous random variable with a density f_X that is positive everywhere in its domain, the distribution function F_X is strictly increasing. In this case F_X has a inverse function F_X^{-1}, called the **quantile function**.

Exercise 28. *$F_X(x) \le u$ if and only if $x \le F_X^{-1}(u)$.*

The **probability transform** follows from an analysis of the random variable

$$U = F_X(X)$$

Note that F_X has range from 0 to 1. It cannot take values below 0 or above 1. Thus, the cumulative distribution function

$$F_U(u) = 0 \text{ for } u < 0 \quad \text{and} \quad F_U(u) = 1 \text{ for } u \ge 1.$$

For values of u between 0 and 1, note that

$$P\{F_X(X) \le u\} = P\{X \le F_X^{-1}(u)\} = F_X(F_X^{-1}(u)) = u.$$

Thus, the distribution function for the random variable U,

$$F_U(u) = \begin{cases} 0 & u < 0, \\ u & 0 \le u < 1 \\ 1 & 1 \le u \end{cases}$$

Thus, if we can simulate U, we can simulate a random variable with distribution F_X via the quantile function

$$X = F_X^{-1}(U). \tag{2}$$

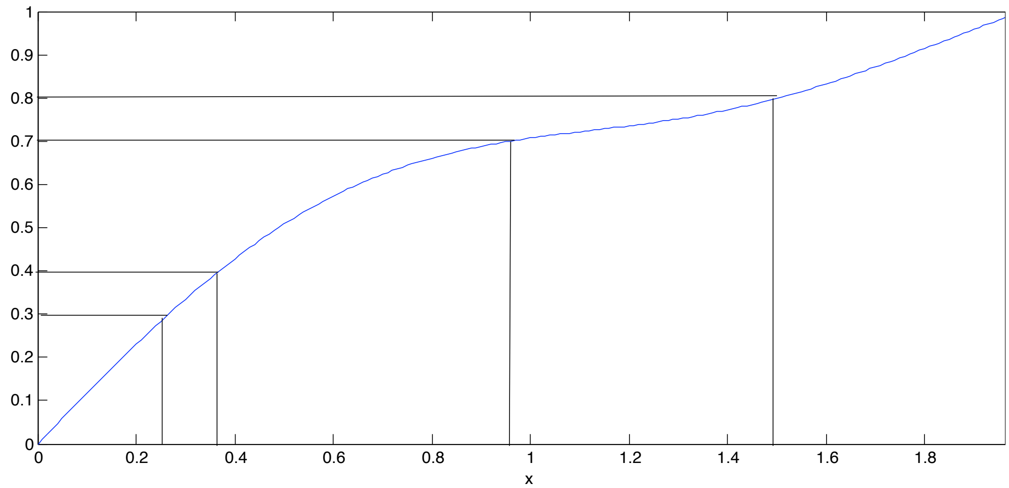


Figure 7: Illustrating the Probability Transsform. First simulate uniform random variables $u_1, u_2, \dots, u_n$ on the interval $[0,1]$. About 10% of the random numbers should be in the interval $[0.3, 0.4]$. This corresponds to the 10% of the simulations on the interval $[0.28, 0.38]$ for a random variable with distribution function F_X shown. Similarly, about 10% of the random numbers should be in the interval $[0.7, 0.8]$ which corresponds to the 10% of the simulations on the interval $[0.96, 1.51]$ for a random variable with distribution function F_X, These values on the x-axis can be obtained from taking the inverse function of F_X, i.e., $x_i = F_X^{-1}(u_i)$.

Take a derivative to see that the density

$$f_U(u) = \begin{cases} 0 & u < 0, \\ 1 & 0 \le u < 1 \\ 0 & 1 \le u \end{cases}$$

Because the random variable U has a constant density over the interval of its possible values, it is called **uniform** on the interval $[0,1]$ and the identity (2) is called the **probability transform**. This transform is illustrated in Figure 7. It accomplished in R via the `runif` command. We can see how this works in the following example.

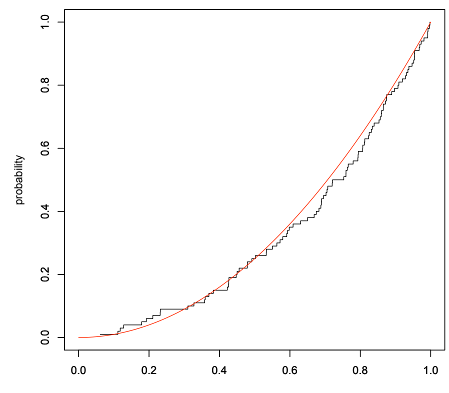


Figure 8: The distribution function (red) and the empirical cumulative distribution function (black) based on 100 simulations of the dart board distribution. R commands given below.

Example 29. *For the dart board,*

$$u = F_X(x) = x^2 \quad \textit{and thus} \quad x = F_X^{-1}(u) = \sqrt{u}.$$

We can simulate independent observations of the distance from the center $X_1, X_2, \dots, X_n$ of the dart by simulating independent uniform random variables $U_1, U_2, \dots U_n$ and taking the transform

$$X_i = \sqrt{U_i}.$$

```
> u<-runif(100)
> x<-sqrt(u)
> xd<-seq(0,1,0.01)
> plot(sort(x),1:length(x)/length(x),type="s",xlim=c(0,1),ylim=c(0,1),
+ xlab="x",ylab="probability")
> par(new=TRUE)
> plot(xd,xd^2,type="l",xlim=c(0,1),ylim=c(0,1),xlab="",ylab="",col="red")
```

Exercise 30. *If U is uniform on $[0,1]$, then so is $V = 1 - U$.*

Sometimes, it is easier to simulate X using $F_X^{-1}(V)$.

Example 31. *For an exponential random variable, set*

$$u = F_X(x) = 1 - \exp(-\lambda x), \quad \textit{and thus} \quad x = -\frac{1}{\lambda}\ln(1-u)$$

Consequently, we can simulate independent exponential random variables $X_1, X_2, \ldots, X_n$ by simulating independent uniform random variables $V_1, V_2, \ldots V_n$ and taking the transform

$$X_i = -\frac{1}{\lambda}\ln V_i.$$

Example 32.

9 Answers to Selected Exercises

2. The sum, the maximum, the minimum, the difference, the value on the first die, the product.

3. The roll with the first H, the number of T, the longest run of H, the number of Ts after the first H.

4. $\lfloor 10^n x \rfloor / 10^n$

6. A common way to show that two events A_1 and A_2 are equal is to pick an element $\omega \in A_1$ and show that it is in A_2. This proves $A_1 \subset A_2$. Then pick an element $\omega \in A_2$ and show that it is in A_1, proving that $A_2 \subset A_1$. Taken together, we have that the events are equal, $A_1 = A_2$. Sometimes the logic needed in showing $A_1 \subset A_2$ consist now just of implications, but rather of equivalent statements. (We can indicate this with the symbol $\Longleftrightarrow$.) In this case we can combine the two parts of the argument. For this exercise, as the lines below show, this is a successful strategy.

We follow an arbitrary outcome $\omega \in \Omega$

1. $\omega \in \{X \in B\}^c \Longleftrightarrow \omega \notin \{X \in B\} \Longleftrightarrow X(\omega) \notin B \Longleftrightarrow X(\omega) \in B^c \Longleftrightarrow \omega \in \{X \in B^c\}$. Thus, $\{X \in B\}^c = \{X \in B^c\}$
2. $\omega \in \bigcup_i \{X \in B_i\} \Longleftrightarrow \omega \in \{X \in B_i\}$ for some $i \Longleftrightarrow X(\omega) \in B_i$ for some $i \Longleftrightarrow X(\omega) \in \bigcup_i B_i \Longleftrightarrow \omega \in \{X \in \bigcup_i B\}$. Thus, $\bigcup_i \{X \in B_i\} = \{X \in \bigcup_i B\}$.

8. For three tosses of a biased coin, we have

x	0	1	2	3
$P\{X = x\}$	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Thus, the cumulative distribution function,

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0, \\ (1-p)^3 & \text{for } 0 \le x < 1, \\ (1-p)^3 + 3p(1-p)^2 = (1-p)^2(1+2p) & \text{for } 1 \le x < 2, \\ (1-p)^2(1+2p) + 3p^2(1-p) = 1 - p^3 & \text{for } 2 \le x < 3, \\ 1 & \text{for } 3 \le x \end{cases}$$

9. From the example in the section *Basics of Probability*, we know that

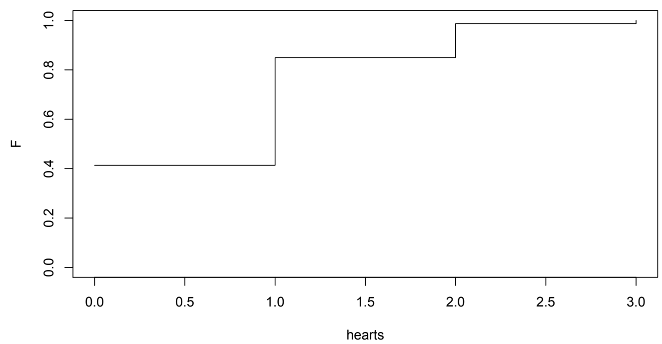
x	0	1	2	3
$P\{X = x\}$	0.41353	0.43588	0.13765	0.01294

To plot the distribution function, we use,

```
> f<-choose(13,hearts)*choose(39,3-hearts)/choose(52,3)
> F<-cumsum(f)
> plot(hearts,F,ylim=c(0,1),type="s")
```

Thus, the cumulative distribution function,

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0, \\ 0.41353 & \text{for } 0 \le x < 1, \\ 0.84941 & \text{for } 1 \le x < 2, \\ 0.98706 & \text{for } 2 \le x < 3, \\ 1 & \text{for } 3 \le x \end{cases}$$



10. The cumulative distribution function for Y,

$F_Y(y) = P\{Y \le y\} = P\{X^3 \le y\} = P\{X \le \sqrt[3]{y}\} = F_X(\sqrt[3]{y})$.

11. Let $x_n \to x_0$ be a decreasing sequence. Then $x_1 > x_2 > \cdots$

$$\{X \le x_1\} \supset \{X \le x_2\} \supset \cdots, \quad \bigcap_{n=1}^{\infty} \{X \le x_n\} = \{X \le x_0\}.$$

(Check this last equality.) Then $P\{X \le x_1\} \ge P\{X \le x_2\} \ge \cdots$. Now, use the second continuity property of probabilities to obtain $\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P\{X \le x_n\} = P\{X \le x_0\} = F_X(x_0)$.

14. We use the fact that the exponential function is increasing, and that $\lim_{u\to\infty} \exp(-u) = 0$. Using the numbering of the properties above

1. For $x < 0$, F_X is constant, $F_X(0) = 0$ and $\exp(-\lambda x)$ is decreasing. Thus, $1 - \exp(-\lambda x)$ is increasing for $x \ge 0$.
2. $\lim_{x\to\infty} \exp(-\lambda x) = 0$. Thus, $\lim_{x\to\infty} 1 - \exp(-\lambda x) = 1$.
3. Because $F_X(x) = 0$ for all $x < 0$, $\lim_{x\to-\infty} F_X(x) = 0$.

15. The distribution function has the graph shown in Figure 5.
The formula

$$F_T(x) = P\{X \le x\} = \begin{cases} 0 & \text{if } x < 0, \\ 1 - \exp(-x/10) & \text{if } 0 \le x < 20, \\ 1 & \text{if } 20 \le x. \end{cases}$$

18. For $r \ne 1$, write the expressions for s_n and rs_n and subtract.

$$\begin{aligned} s_n &= c + cr + cr^2 + \cdots + cr^n \\ rs_n &= \quad\;\; cr + cr^2 + \cdots + cr^n + cr^{n+1} \\ (1-r)s_n &= c \qquad\qquad\qquad\qquad\quad - cr^{n+1} = c(1 - r^{n+1}) \end{aligned}$$

Now divide by $1 - r$ to obtain the formula.

21. Let f_X be the density. Then

$$0 \le P\{X = x_0\} \le P\{x_0 - \Delta x < X \le x + \Delta x\} = \int_{x_0-\Delta x}^{x_0+\Delta x} f_X(x)\,dx.$$

Now the integral goes to 0 as $\Delta x \to 0$. So, we must have $P\{X + x_0\} = 0$.

23. Because the density is non-negative on the interval $[0,1]$, $F_X(x) = 0$ if $x < 0$ and $F_X(x) = 1$ if $x < 1$. For x between 0 and 1,

$$\int_0^x \frac{1}{2\sqrt{t}}\,dt = \sqrt{t}\Big|_0^x = \sqrt{x}.$$

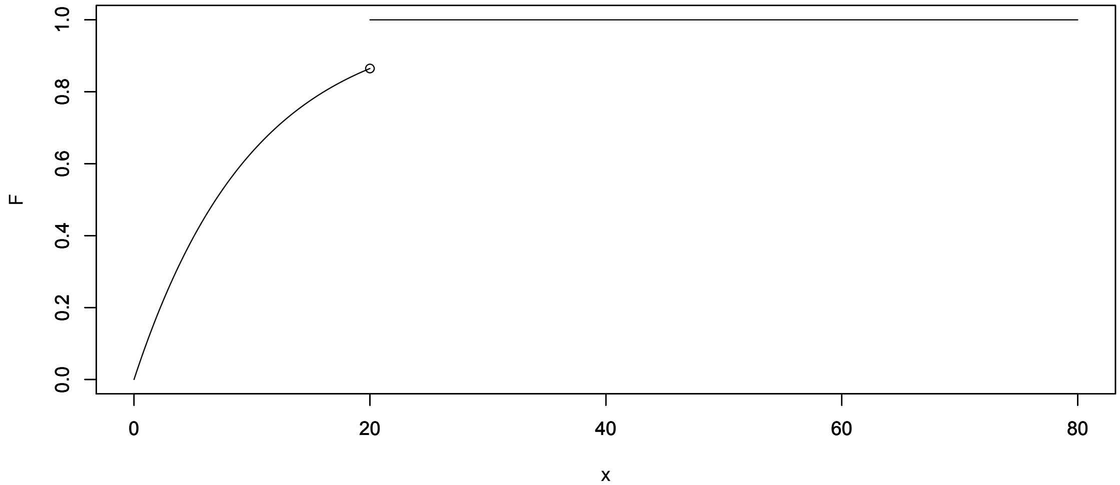


Figure 9: Cumulative distribution function for an exponential random variable with $\lambda = 1/10$ and a jump at $x = 20$.

Thus,

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \sqrt{x} & \text{if } 0 < x < 1, \\ 1 & \text{if } 1 \leq x. \end{cases}$$

24. The random variable Y has distribution function

$$F_Y(y) = P\{Y \leq y\} = P\{aX + b \leq y\} = P\{aX \leq y - b\}.$$

For $a > 0$

$$F_Y(y) = P\left\{X \leq \frac{y-b}{a}\right\} = F_X\left(\frac{y-b}{a}\right).$$

Now take a derivative and use the chain rule to find the density

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

For $a < 0$

$$F_Y(y) = P\left\{X \geq \frac{y-b}{a}\right\} = 1 - F_X\left(\frac{y-b}{a}\right).$$

Now the derivative

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

26. The joint density (mass function) for $X_1, X_2, \ldots, X_n$

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

is the product of the marginal densities (mass functions).

28. F_X is increasing and continuous, so the set $\{x; F_X(x) \leq u\}$ is the interval $(-\infty, F_X^{-1}(u)]$. In addition, x is in this inverval precisely when $x \leq F_X^{-1}(u)$.

30. Let's find F_V. If $v < 0$, then

$$0 \le P\{V \le v\} \le P\{V \le 0\} = P\{1 - U \le 0\} = P\{1 \le U\} = 0$$

because U is never greater than 1. Thus, $F_V(v) = 0$ Similarly, if $v \ge 1$,

$$1 \ge P\{V \le v\} \ge P\{V \le 1\} = P\{1 - U \le 1\} = P\{0 \le U\} = 1$$

because U is always greater than 0. Thus, $F_V(v) = 1$ For $0 \le v < 1$,

$$F_V(v) = P\{V \le v\} = P\{1 - U \le v\} = P\{1 - v \le U\} = 1 - P\{U < 1 - v\} = 1 - (1 - v) = v.$$

This matches the distribution function of a uniform random variable on $[0, 1]$.