Topic 7

Random Variables and Distribution Functions

Joint Distributions and Simulations

Outline

Joint Distributions

Experimental procedures use data based on multiple observations. 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, X_1 and X_2, this means looking at the probability of events,

$$P\{X_1 \in B_1, X_2 \in B_2\}.$$

For discrete random variables, take $B_1 = \{x_1\}$ and $B_2 = \{x_2\}$ 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 $B_1 = (x_1, x_1 + \Delta x_1]$ and $B_2 = (x_2, x_2 + \Delta x_2]$ and ask that for some function f_{X_1,X_2}, the joint probability density function satisfies

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

Independent Random Variables

For independent discrete random variables, we have that

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

In this case, we say that the joint probability mass function is the product of the marginal mass functions. A similar identity for independent continuous random variables.

Exercise. Roll two dice and consider equally likely outcomes. Let X_1 be the value on the first die and X_2 be the value on the second. Show that X_1 and X_2 are independent.

Simulating Discrete Random Variables

The `sample` command has been used to generate simple random sample, a procedure called sampling without replacement. The `sample` command can also be used to simulate a discrete random variable. For example, for the mass function

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

```
> x<-c(1:4);f<-c(0.4,0.3,0.2,0.1) #create a vector of values and mass function
> sum(f)                          #check that the sum is one
[1] 1
> data<-sample(x,80,replace=TRUE,prob=f) #simulate 80 independent observations
> table(data)                     #make a table of the simulated values
data
 1  2  3  4
33 20 16 11
```

Simulating Discrete Random Variables

Exercise.

1. Repeat the simulation and compare the values using the `table` command.
2. For equally likely outcomes, repeat the simulation, remove `prob=f` and make a table. Describe the differences in the tabulated values.
3. Simulate 80 observations of a fair die and make a table.
4. Simulate 80 observations of an unfair die whose mass function is given below.

x	1	2	3	4	5	6
$f_X(x)$	1/12	1/12	1/12	1/4	1/4	1/4

5. Compute the mean and standard deviation of the observations.

Describe the differences in the tabulated values for the fair and unfair die.

Simulating Continuous Random Variables

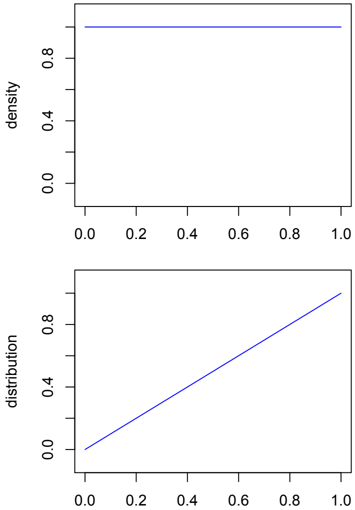
The `runif(n)` command is used to simulate n independent random variables U that are uniformly distributed on the interval $[0,1]$.

The density function is

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

The distribution function is

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



Probability Transform

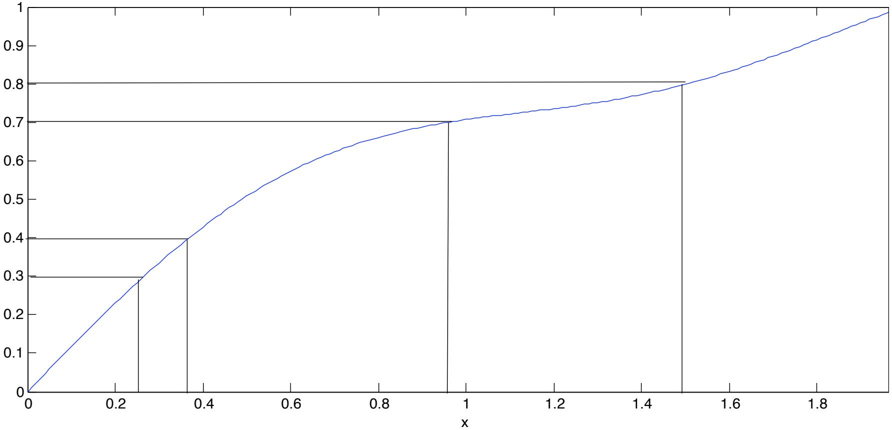


Figure: Cumulative Distribution Function and the Inverse Function

Probability Transform

If X a continuous random variable with a density f_X that is positive everywhere in its domain, then

- the distribution function $F_X(x) = P\{X \leq x\}$ is strictly increasing.
- F_X has a inverse function F_X^{-1}, known as the quantile function.
  - For example, $F_X^{-1}(1/2)$ is the median, $F_X^{-1}(3/4)$ is the third quartile
- For values of u between 0 and 1, note that

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

  The distribution function for the random variable U.
- If we can simulate U, we can simulate a random variable with distribution F_X via the quantile function

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

Probability Transform

For the dart board, for x between 0 and 1, the distribution function

$$u = F_X(x) = x^2 \quad \text{and thus the quantile function} \quad x = F_X^{-1}(u) = \sqrt{u}.$$

We can simulate independent observations of the distance from the center of the dart board

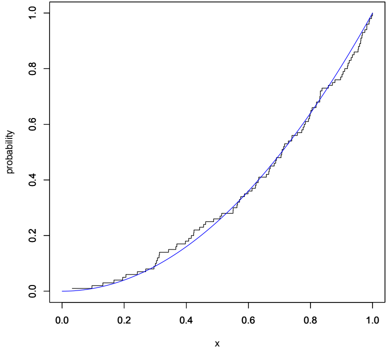
$$X_1, X_2, \ldots, X_n$$

of the dart board by simulating independent uniform random variables $U_1, U_2, \ldots U_n$ and taking the quantile function

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

Probability Transform

```
> u<-runif(100)              #simulate 100 uniform random variables
> x<-sqrt(u)                 #find the quantile functions
> xd<-seq(0,1,0.01)          #set a sequence to graph F(x)=x^2
> 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="blue")
```



Exercise. Perform the simulation of dart throws. Give the 0.25, 0,50, and 0.75 quantiles for both the distribution function F_X and for the simulated values.