Chapter 1
Probability Theory
Mass Functions and Density Functions

Outline

Mass Functions

The (probability) mass function of a discrete random variable X is

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

The mass function has two basic properties:

- $f_X(x) \geq 0$ for all $x \in S$, the state space.
  - Probabilities are non-negative.
- $\sum_x f_X(x) = 1$.
  - The collection

    $$C_x = \{\omega; X(\omega) = x\}$$

    for all $x \in S$, forms a partition of the probability space, Ω.

Mass Functions

Let's make tosses of a biased coin whose outcomes are independent. Let X denote the random variable that gives the number of tails before the first head and p denote the probability of heads in any given toss. Then

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

Because the terms in this mass function form a geometric sequence, X is called a geometric random variable

Mass Functions

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 $\lim_{n\to\infty} r^n = 0$ and thus s_n has a limit as $n \to \infty$. In this case, the infinite sum is the limit

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

Thus, $\sum_{x=0}^{\infty} f_X(x) = \sum_{x=0}^{\infty} (1-p)^x p = \frac{p}{1-(1-p)} = 1.$

$$\bar{F}(b) = P\{X > b\} = \sum_{x=b+1}^{\infty} f_X(x) = \sum_{x=b+1}^{\infty} (1-p)^x p = \frac{(1-p)^{b+1} p}{1 - (1-p)} = (1-p)^{b+1}$$

and $F_X(b) = 1 - (1-p)^{b+1}$ for $b = 0, 1, 2, \ldots$.

Mass Functions

Exercise. We use R to investigate a geometric random variable with $p = 1/3$. Enter the commands

```
> x<-0:10          #creates a sequence from 0 to 10
> f<-dgeom(x,1/3) #gives the mass function for these values
> F<-pgeom(x,1/3) #gives the distribution function for these values
> data.frame(x,f,F)
```

- Check that the jumps in the cumulative distribution function $F_X(x) - F_X(x-1)$ is equal to the values of the mass function.
- Find $P\{X \leq 4\}$, $P\{2 < X \leq 5\}$, and $P\{X \geq 5\}$.

Answer. 0.8683128, 0.2085048, 0.1316872

Mass Functions

Example An urn contains balls numbers 1 through n. From these r are removed at random. Let X be the *maximum* value of the removed balls.

If they are removed with replacement, then the distribution function

$$F_X(x) = P\{X \le x\} = \left(\frac{x}{n}\right)^r.$$

The mass function

$$f_X(x) = P\{x-1 < X \le x\} = F_X(x) - F_X(x-1) = \left(\frac{x}{n}\right)^r - \left(\frac{x-1}{n}\right)^r.$$

Mass Functions

If they are removed without replacement, then the distribution function

$$F_X(x) = P\{X \le x\} = \frac{\binom{x}{r}}{\binom{n}{r}} = \frac{(x)_r}{(n)_r}.$$

The mass function

$$\begin{aligned} f_X(x) &= P\{x-1 < X \le x\} = F_X(x) - F_X(x-1) \\ &= \frac{\binom{x}{r}}{\binom{n}{r}} - \frac{\binom{x-1}{r}}{\binom{n}{r}} = \frac{\binom{x-1}{r-1}}{\binom{n}{r}} \end{aligned}$$

by the Pascal triangle identity.

Mass Functions

Example. (mixtures) Let X_1 and X_2 be discrete random variables with respective mass functions f_1 and f_2. Flip a biased coin that lands heads with probability p. With heads, observe X_1 and with tails observe X_2. By the law of total probability, the mass function of the observation X,

$$\begin{aligned} f_X(x) &= P\{X = x\} \\ &= P\{X = x|\text{heads}\}P\{\text{heads}\} + P\{X = x|\text{tails}\}P\{\text{tails}\} \\ &= f_1(x)p + f_2(x)(1-p) \end{aligned}$$

Exercise. Generalize the notion of mixture to more than two random variables.

Density Functions

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, if Δx is small,

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

Density Functions

We can compute probabilities by evaluating definite integrals

$$
\begin{aligned}
P\{a < X \le b\} &= F_X(b) - F_X(a) \\
&= \int_a^b f_X(t)\, dt.
\end{aligned}
$$

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

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

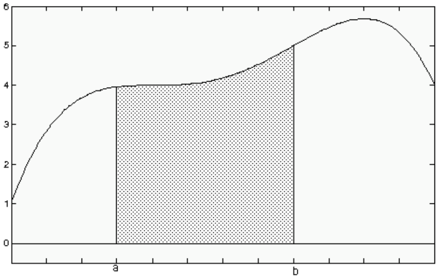

Figure: Integral of density function $f_X(x)$. Shaded region has area $P\{a < X \le b\}$.

Density Functions

Exercise. For X for a continuous random variable. For any x_0, explain why $P\{X = x_0\} = 0$.

Example. X is called a Pareto random variable if for some $\alpha > 0$ and $\beta > 0$, its density

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq \alpha, \\ \frac{\beta\alpha^\beta}{x^{\beta+1}} & \text{if } x > \alpha. \end{cases}$$

Note that $\int_{-\infty}^{\infty} f_X(x)\,dx = \int_\alpha^\infty \frac{\beta\alpha^\beta}{x^{\beta+1}}\,dx = -\alpha^\beta x^{-\beta}\Big|_\alpha^\infty = 0 - (-1) = 1$

Exercise. Show that X has distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq \alpha, \\ 1 - \left(\frac{\alpha}{x}\right)^\beta & \text{if } x > \alpha. \end{cases}$$

Answer. Check that F_X is a distribution function, and that $F_X'(x) = f_X(x)$

Density Functions

Exercise. Let X be a continuous random variable with density

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

1. Sketch a graph of f_X. Notice that f_X is not bounded.
2. Show that f_X is a density function and find its distribution function F_X.
   $F_X(x) = 0, x < 0, = \sqrt{x}, 0 \leq x < 1, = 1, 1 < x$
3. Use distribution function to find $P\{X \leq 1/4\}$ and $P\{1/4 \leq X \leq 1/2\}$.
   $1/2$, $\sqrt{1/2} - 1/2$

Exercise. Generalize the notion of mixtures to continuous random variables.

Hazard Functions

The hazard function $h(x)$ for a continuous nonnegative random variable is the rate of failure at time x. More precisely,

$$h(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} P\{X \le x + \Delta x | X \ge x\}$$

If X has cumulative distribution F_X and density f_X, then

$$\begin{aligned} \frac{1}{\Delta x} P\{X \le x + \Delta x | X \ge x\} &= \frac{P\{x < X \le x + \Delta x\}}{\Delta x \; P\{X \ge x\}} \\ &= \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x \; (1 - F_X(x))} \to \frac{f_X(x)}{(1 - F_X(x))} \end{aligned}$$

as $\Delta x \to 0$. Thus,

$$h(x) = \frac{f_X(x)}{(1 - F_X(x))} = -\frac{d}{dx} \ln(1 - F_X(x)).$$

Hazard Functions

The cumulative hazard

$$H(t) = \int_0^t h(x)\ dx = -\int_0^t \frac{d}{dx} \ln(1 - F_X(x))\ dx = -\ln(1 - F_X(t))$$

and consequently,

$$F_X(t) = 1 - \exp(-H(t))$$

Exercise. Find the hazard function for exponential distributions, for Pareto distributions.

Answer. For the exponential distriubtion,

$$F_X(t) = 1 - \exp(-\beta t), \quad \text{so} \quad h(x) = \beta.$$

For the Pareto distribution.

$$\frac{f_X(x)}{(1 - F_X(x))} = \frac{\beta\alpha^\beta x^{-\beta-1}}{\alpha^\beta x^{-\beta}} = \frac{\beta}{x}.$$

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

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

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<-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.

Exercise. Simulate the number of $\diamondsuit$'s from 5 cards drawn random from a standard deck. Compare your results with the mass function.

Simulating Discrete Random Variables

Exercise. Give the mass function for the random variable X, the maximum value on the roll of two independent die.

To simulate this,

```
> x<-replicate(36000,max(sample(6,2,replace=TRUE)))
> table(x)
x
    1     2     3     4     5     6
  977  3081  4949  7018  8982 10993
```

For the minimum,

```
> x<-replicate(36000,min(sample(6,2,replace=TRUE)))
> table(x)
x
    1     2     3     4     5     6
11036  9097  6878  5027  2943  1019
```