Chapter 04 - Random Variables-Distributions

October 24, 2018

1 Random Variables - Distributions

2 Probability Concepts

2.0.1 Sample Space

For a random experiment E, the set of all possible outcomes of E is called the **sample space** and is denoted by the letter S. For a coin-toss experiment, S would be the results "Head" and "Tail", which we may represent by $S = [H; T]$. Formally, the performance of a random experiment is the unpredictable selection of an outcome in S .

2.0.2 Events

An event A is merely a collection of outcomes, or in other words, a subset of the sample space. After the performance of a random experiment E we say that the event A occurred if the experiment's outcome belongs to A. We say that a bunch of events $A_1, A_2, A_3, \ldots$ are mutually exclusive or disjoint if $A_i \cap A_j = \emptyset$. for any distinct pair $A_i \neq A_j$. For instance, in the coin-toss experiment the events $A = [Head]$ and $B = [Tails]$ would be mutually exclusive.

2.0.3 Probability - Relative Frequency Approach

This approach states that the way to determine $P(Heads)$ is to flip the coin repeatedly, in exactly the same way each time. Keep a tally of the number of flips and the number of Heads observed. Then a good approximation to P(Heads) will be

$$P(Heads) = \frac{\text{number of observed Heads}}{\text{total number of flips}}$$

The mathematical underpinning of this approach is the celebrated Law of Large Numbers which may be loosely described as follows. Let E be a random experiment in which the event A either does or does not occur. Perform the experiment repeatedly, in an identical manner, in such a way that the successive experiments do not influence each other. After each experiment, keep a running tally of whether or not the event A occurred. Let S_n count the number of times that A occurred in the n experiments. Then the law of large numbers says that

$$\frac{S_n}{n} \to P(A) \text{ as } n \to \infty$$

2.0.4 Types of questions we want to answer with probabilities

Find the probability that the mean is . . .

English statement	Mathematical notation
"equal to fifty"	$P(\mu = 50)$
"not equal to fifty"	$P(\mu \neq 50)$
"not fifty"	$P(\mu \neq 50)$
"greater than fifty"	$P(\mu > 50)$
"at least fifty"	$P(\mu \geq 50)$
"less than fifty"	$P(\mu < 50)$
"no more than fifty"	$P(\mu \leq 50)$
"between 40 and 50 (incl)	$P(40 \leq \mu \leq 50)$
"between 40 and 50"	$P(40 \leq \mu \leq 50)$

3 Random Variable

Statistics and data mining are concerned with data. How do we link sample spaces Ω and events A to data? The link is provided by the concept of a random variable.

We already know about experiments, sample spaces, and events. In this section, we are interested in a number that is associated with the experiment. We conduct a random experiment E and after learning the outcome ω in S we calculate a number X. That is, to each outcome ω in the sample space we associate a number $X(\omega) = x$.

Definition 1 *A Random variable X is a function*

$$X : S \rightarrow R$$

that assigns a real number $X(\omega) = x$ to each outcome ω of a procedure associated with the sample space S.

Example 1 *Examples of random variables:*

- *Height*
- *Weight*
- *Number of males in a class*

Example 2 *Let E be the experiment of flipping a coin twice. The sample space is $S = [HH, HT, TH, TT]$. Now define the random variable X = the number of heads. That is, for example, $X(HH) = 2$, while $X(HT) = 1$. We may make a table of the possibilities:*

$\omega \in S$	HH	HT	TH	TT
$X(\omega) = x$	2	1	1	0

Taking a look at the second row of the table, we see that the support of X – the set of all numbers that X assumes – would be $S_X = [0, 1, 2]$.

Example 3 *Flip a coin ten times and define the outcome ω. Let $X(\omega)$ be the number of heads in the sequence ω. For example, if $\omega = HHTHHTHHTT$, then $X(\omega) = 6$.*

Example 4 *Let E be the experiment of flipping a coin repeatedly until observing a Head. The sample space would be $S = [H, TH, TTH, TTTH, \ldots]$. Now define the random variable $Y =$* ***the number of Tails before the first head****. Then the support of Y would be $S_Y = [0, 1, 2, \ldots]$.*

Example 5 *Let $S = \{(x, y) : x^2 + y^2 \leq 1\}$ be the unit disk. Consider drawing a point at random from S. A typical outcome is of the form $\omega = (x, y)$. Some examples of random variables are $X(\omega) = x$, $Y(\omega) = y$, $Z(\omega) = x + y$, and $W(\omega) = \sqrt{x^2 + y^2}$.*

Example 6 *When throwing a dice, the possible outcomes are 1, 2, 3, 4, 5, or 6 eyes. The sample space consists of all these six outcomes.*

Examples of events are:

1. *The dice shows an even number of eyes, i.e., 2, 4, or 6.*
2. *The dice shows an odd number of eyes, i.e., 1, 3, or 5.*
3. *The dice shows at most 3 eyes.*
4. *The dice shows exactly 3 eyes.*

As the last example shows, an event can consist of just one outcome. The events in example 1 and 2 are complementary events. This means that

- *The two events are mutually exclusive, i.e., the two events have no outcome in common.*
- *The two events together contain all outcomes of the sample space.*

Remark 1 *The sum of the probabilities of two complementary events is 1.*

Notice that X denotes the random variable and x denotes a particular value of X.

Example 7 *Flip a coin twice and let X be the number of heads. Then, $P(X = 0) = P(TT) = 1/4$, $P(X = 1) = P(HT, TH) = 1/2$ and $P(X = 2) = P(HH) = 1/4$. The random variable and its distribution can be summarized as follows:*

ω	$P(\omega)$	$X(\omega)$
TT	*1/4*	*0*
TH	*1/2*	*1*
HT	*1/2*	*1*
HH	*1/4*	*2*

x	$P(X = x)$
0	*1/4*
1	*1/2*
2	*1/4*

3.0.1 Review

Remark 2 ***A random variable (*) is a mathematical function on the sample space.***

The mathematical function will often be the identity! For instance, in example 6, the sample space consists of the outcomes 1, 2, 3, 4, 5, or 6 (eyes on a dice). The random variable is simply the number itself, i.e., the number of eyes shown; no mathematical operation is done! This number will vary randomly, hence the term random variable.

4 Types of random variables

- **discrete** are countable. Have an associated **probability distribution**. (ex.Number of males)
- **continuous** have infinitely many values. Have an associated **probability density function**. (ex. Height)

5 Discrete random variables

5.1 Probability Mass Function

Consider a discrete random variable X that takes values $S_X = \{x_i\}_{i=1}^{k}$

Definition 2 ***Probability Mass Function*** *Every discrete random variable X has associated with it a probability mass function (PMF)* $f_X : S_X \to [0,1]$ *defined by*

$$f_X(x) = P(X = x)$$

in the space S_X*.*

Two key properties of probability mass functions

$$\sum_{i=1}^{k} P(X = x_i) = 1$$

with k possible values of X.

$$0 \leq P(X = x_i) \leq 1 \text{ for } i = 1, 2, \ldots, k$$

Example 8 *Toss a coin 3 times. The sample space would be* $S = [HHH, HTH, THH, TTH, HHT, HTT, THT, TTT]$*. Now let X be the number of Heads observed. Then X has support* $S_X = [0, 1, 2, 3]$*. Assuming that the coin is fair and was tossed in exactly the same way each time, it is not unreasonable to suppose that the outcomes in the sample space are all equally likely.*

```
In [28]: # In this case the PMF is
         OmegaX = c(0,1,2,3)
         Px <- c(1/8,3/8,3/8,1/8)
         table <- cbind(OmegaX,Px)
         table
```

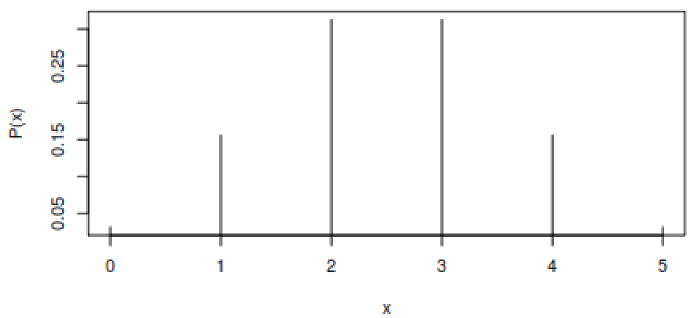
OmegaX	Px
0	0.125
1	0.375
2	0.375
3	0.125

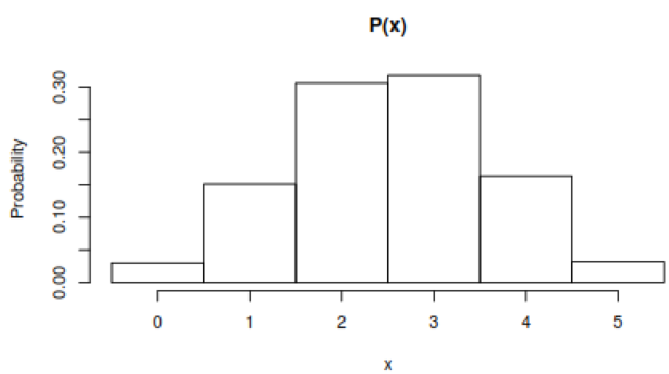
Ways we can represent $P(X = x_i)$
Three basic ways we can represent $P(X = x_i)$

1. **Table**

	P(x)
0	0.03125
1	0.15625
2	0.31250
3	0.31250
4	0.15625
5	0.31125

2. **Probability distribution plot**





3. **Probability histogram**

5.2 MEAN AND STANDARD DEVIATION OF $P(X = x_i)$

Given $P(X = x_i)$, we can find the mean and standard deviation:

Definition 3 *Mean and standard deviation of discrete distribution*

Expected value *of a random variable is the mean value μ. The expected value E in terms of the probability distribution is:*

$$E = \mu = \sum_i^k x_i P(x_i)$$

Standard deviation *of a probability distribution is given by:*

$$\sigma = \sqrt{\sum_i^k (x_i - \mu)^2 P(x_i)}$$

```
In [22]: Px <- c(0.03125, 0.15625, 0.31250, 0.31250, 0.15625, 0.03125)
         x <- c(0:5)
         xPx <- x*Px
         E = sum(xPx)
         E
```

2.5

```
In [23]: t1 <-cbind(x,Px,xPx)
         t1
```

x	Px	xPx
0	0.03125	0.00000
1	0.15625	0.15625
2	0.31250	0.62500
3	0.31250	0.93750
4	0.15625	0.62500
5	0.03125	0.15625

```
In [24]: si=(x-2.5)^2*Px
         t2 <- cbind(x, Px, si)
         t2
```

x	Px	si
0	0.03125	0.1953125
1	0.15625	0.3515625
2	0.31250	0.0781250
3	0.31250	0.0781250
4	0.15625	0.3515625
5	0.03125	0.1953125

```
In [25]: s2 = sum(si)
         s2
         s2root = sqrt(s2)
         s2root
```

1.25
1.11803398874989

5.3 Cumulative distribution function (CDF)

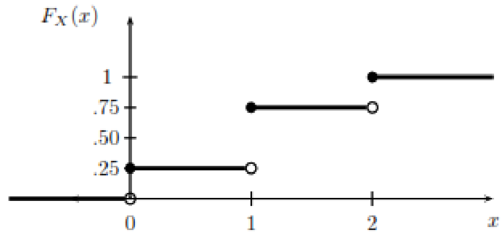
Related to the probability mass function $f_X(x) = P(X = x)$ is another important function called the cumulative distribution function (CDF), F_X. It is defined by the formula

$$F_X(t) = P(X \leq t)\, , -\infty < t < \infty$$

.

We know that all PMFs satisfy certain properties, and a similar statement may be made for CDFs. In particular, any CDF F_X satisfies

- F_X is nondecreasing



CDF_flipping_coin_2

- F_X is right-continuous
- $\lim_{t\to-\infty} F_X(t) = 0$ and $\lim_{t\to\infty} F_X(t) = 1$

Example 9 *CDF for flipping a coin twice*

Remark 3 *The CDF effectively contains all the information about the random variable.*

We say that X has the distribution F_X and we write $X \sim F_X$. In an abuse of notation we will also write $X \sim f_X$ and for the named distributions the PMF or CDF will be identified by the family name instead of the defining formula.

6 Continuous Distributions

Now we move to random variables whose support is a whole range of values, say, an interval (a, b). It is shown in later classes that it is impossible to write all of the numbers down in a list; there are simply too many of them.

6.1 Continuous Random Variables

6.1.1 Probability Density Functions

Continuous random variables have supports (range of values that take) that look like

$$S_X = [a, b] \ or \ (a, b)$$

;

or unions of intervals of the above form. Examples of random variables that are often taken to be continuous are: *the height or weight of an individual*, other physical measurements such as the length or size of an object, and *duration* of time (usually).

Every continuous random variable X has a *probability density function* (PDF) denoted f_X associated with it that satisfies three basic properties:

1) $f_X(x) > 0$ for $x \in S_X$,
2) $\int_{x\in S_X} f_X(x)dx = 1$, and
3) $P(X \in A) = \int_{x\in A} f_X(x)dx$, for an event $A \in S_X$.

Remark We can say the following about continuous random variables: Usually, the set A takes the form of an interval, for example, $A = [c, d]$, in which case

$$P(X \in A) = \int_c^d f_X(x)dx$$

- It follows that the probability that X falls in a given interval is simply the area under the curve of f_X over the interval.
- Since the area of a line $x = c$ in the plane is zero, $P(X = c) = 0$ for any value c.

In other words, the chance that X equals a particular value c is zero, and this is true for any number c. Moreover, when $a < b$ all of the following probabilities are the same:

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

- The PDF f_X can sometimes be greater than 1. This is in contrast to the discrete case; every nonzero value of a PMF is a probability which is restricted to lie in the interval $[0, 1]$.

We have defined the cumulative distribution function, F_X , for discrete random variables. Recall that it is defined by $F_X(t) = P(X \leq t)$, $-\infty < t < \infty$. While in the discrete case the CDF is unwieldy, in the continuous case the CDF has a relatively convenient form:

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(t)dx \;\; -\infty < t < \infty$$

There is a handy relationship between the CDF and PDF in the continuous case. Consider the derivative of

$$F'_X(t) = f_X(t)$$

The last equality being true by the Fundamental Theorem of Calculus.

Example 10 *Let X have a PDF $f(x) = 3x^2$ in the interval [0,1] find $P(0.14 \leq X \leq 0.7)$.*

```
In [26]: f <-function(x) 3*x^2
         integrate(f, lower = 0.14, upper = 0.71)
```

```
0.355167 with absolute error < 3.9e-15
```

6.1.2 Expectation and Variance of Continuous Random Variables

For a continuous random variable X the expected value of g(X) is

$$E_g(X) = \int_{x \in S} g(x) f_X(x)dx$$

.

One important example is the mean μ, also known as EX:

$$\mu = EX = \int_{x \in S} x f_X(x)dx$$

,

provided that $\int_S |x| f(x)dx$ is finite. Also there is the variance

$$\sigma^2 = E(X-\mu)^2 = \int_{x \in S} (x-\mu)^2 f_X(x) dx = EX^2 - (EX)^2$$

.

In addition, there is a standard deviation $\sigma = \sqrt{\sigma^2}$.

Example 11 *Let X have a PDF* $f(x) = 3/x^3$ *for* $x > 1$. *Find the mean (expected value)*

```
In [27]: g<- function(x) 3/x^3
         integrate(g, lower = 1, upper = Inf)
```

```
1.5 with absolute error < 1.7e-14
```