9 RANDOM VARIABLES

Defn: A random variable is a rule which assigns a numerical value to each possible outcome of an experiment

Example: Toss a coin: S = H, T
Call a head 1 and a tail 0
S = { 1, 0 }

Random variables are **DISCRETE** or **CONTINUOUS**

- **Discrete Random Variable**:
  A random variable is discrete if its values can assume isolated points on the number line.
  Examples:

  number of telephone calls in an hour

  number of jobs arriving for service

- **Continuous Random Variable:**
  A random variable is continuous if its values can assume all points in a particular interval.

  Examples:

  Heights, Weights;

  Lifetime of a component;

  cpu time;

  response time

Probability Distributions Functions

1. Toss a fair coin:

   Let $X = 1$ if head occurs, and $X = 0$ if tail occurs. Since the coin is fair, $P(X = 1) = P(X = 0) = 1/2$. The pdf is

Outcome (x)	0	1
Probability ($P(X = x)$)	.5	.5

2. Tossing two fair coins:

   Let X be the number of heads. Then the values of X are 0, 1 or 2. The sample space consists of the following equally likely outcomes. $\{(T, T), (H, T), (T, H)$ and $(H, H)\}$. So $P(X = 0) = 1/4$, $P(X = 1) = 2/4$, and $P(X = 2) = 1/4$. The pdf is

Outcome (x)	0	1	2
Probability ($P(X = x)$)	.25	.5	.25

3. Draw a card from a deck:

   Let X equal to 1 if a spade is drawn and 0 otherwise. $P(X = 1) = 13/52$ and $P(X = 0) = 39/52$. The pdf is

Outcome (x)	0	1
Probability ($P(X = x)$)	.75	.25

4. Roll a fair die:

   Let X equal the face number, then $X = 1, 2, 3, 4, 5, 6$. $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 6) = 1/6$.

Outcome (x)	1	2	3	4	5	6
Probability ($P(X = x)$)	1/6	1/6	1/6	1/6	1/6	1/6

   Notice that in all cases the probabilities sum to one.

Probability Density Function (PDF) satisfies

- $P(X = x) \geq 0$
- $\sum_x P(X = x) = 1$

Example: Hardware Failures

In order to obtain a model for hardware failures in a computer system, the number of crashes occurring each week was observed over a period of one year. It was found that

0 failures occurred in each of 9 weeks,

1 failure occurred in each of 14 weeks,

2 failures occurred in each of 13 weeks,

3 failures occurred in each of 9 weeks,

4 failures occurred in each of 4 weeks,

5 failures occurred in each of 2 weeks, and

6 failures occurred in 1 week.

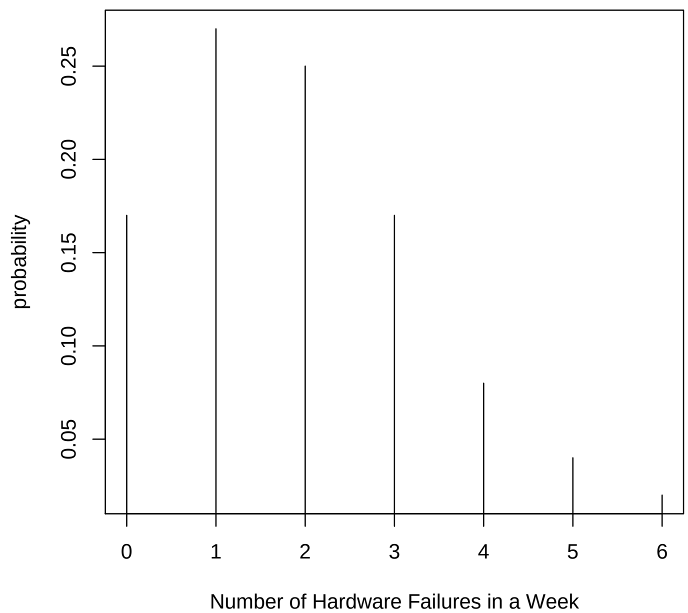
Use R to obtain relative frequencies

```
weeks <- c(9, 14, 13, 9, 4, 2, 1)
sum(weeks)
[1] 52
probability <- weeks/52
round(probability, 2) #rounding the  probabilities to 2 decimal places
[1] 0.17 0.27 0.25 0.17 0.08 0.04 0.02
```

No.of Failures	0	1	2	3	4	5	6
Probability	.17	.27	.25	.17	.08	.04	.02

Graphical Display in *R*

```
Failures<- 0:6
plot(Failures, probability,
 xlab="Number of Hardware Failures in a Week",
 type="h")
```



Example: Stack Size

A particular Java assembler interface was used 2000 times, and the operand stack size was observed. It was found that

a zero stack size occurred 100 times,

a stack size of 1 occurred 200 times

a stack size of 2 occurred 500 times

a stack size of 3 occurred 500 times

a stack size of 4 occurred 400 times

a stack size of 5 occurred 200 times

a stack size of 6 occurred 80 times

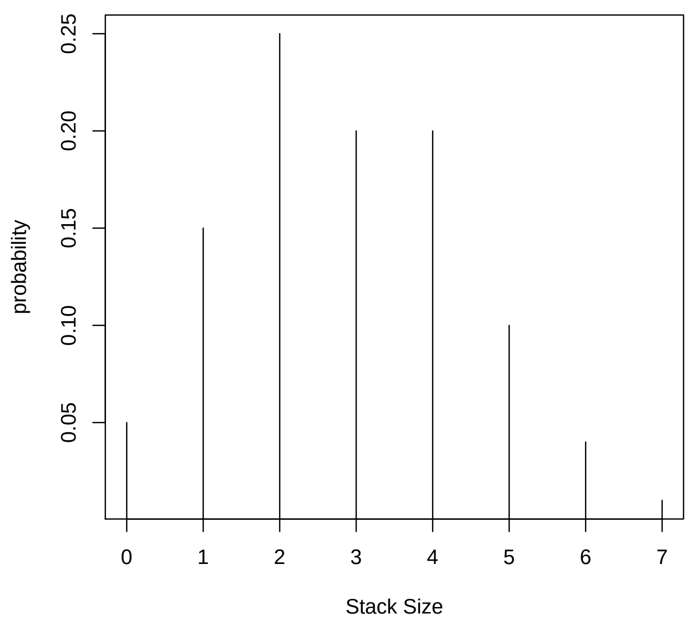
a stack size of 7 occurred 20 times

PDF: Stack Size

Stack Size (x)	0	1	2	3	4	5	6	7
Probability ($P(X = x)$)	.05	.10	.25	.25	.20	.10	.04	.01

Graphical Display in *R*

```
stacksize <- 0:7
probability <- c(.05, .10, .25, .20, .20, .10, .04, .01)
plot(stacksize, probability, xlab = "Stack Size", type = "h")
```



Cumulative Distribution Function (CDF):

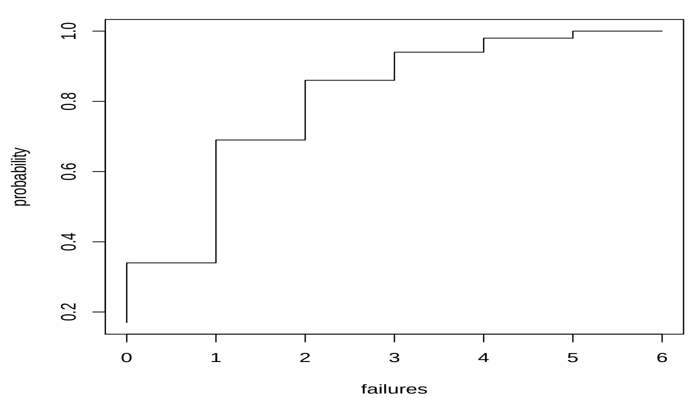
$$F(x) = P(X \leq x)$$

No.of Failures	0	1	2	3	4	5	6
CDF P$(X \leq x)$	.17	.44	.69	.86	.94	.98	1.00

In R

```
cumprob <- c(.17, .44, .68, .86, .94, 1)
plot(Failures, cumprob, xlab = "failures", ylab = "probability'',
            type = "S")
```

wnich gives



Summarising Random Variabless

The Mean of a Sample

Example 1: Averages

Salaries of 6 recent computer science graduates (000s euro).

$$20.3, 14.9, 18.9, 21.7, 16.3, 17.7$$

The average of the sample is

$$\overline{x} = \frac{20.3 + 14.9 + 18.9 + 21.7 + 16.3 + 17.7}{6} = 18.3$$

Generally, if $x_1, x_2, \cdots x_n$ in a sample of size n,

Then the sample average is

$$\overline{x} = \frac{x_1 + x_2 + \cdots x_n}{n}$$

It may be written as

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example 2: Averages

One hundred applicants for a certain degree program had the follwing age distribution.

Age	No. of applicants
18	9
19	40
20	18
21	18
22	8
23	4
24	3

The mean is obtained as follows:

$$\overline{x} = \frac{(18*9)+(19*40)+(20*18)+(21*18)+(22*8)+(23*4)+(24*3)}{100}$$

$$\overline{x} = (18*.09)+(19*.40)+(20*.18)+(21*.18)+(22*.08)(23*.04)(24*.03) = 20$$

$$\sum xp(x)$$

Mean of a Random Variable

Definition:

The mean of a discrete random variable is defined as the weighted average of all possible values.

The weights are the probabilities of respective values of the random variable

$$E(X) = \sum_x xp(x)$$

The mean or expected value of a random variable X is often denoted by μ_x

Examples:

1. The number of hardware failures of a computer system in a week of operation has the following pdf:

No. of Failures (X)	0	1	2	3	4	5	6
Probability (P(X=x))	.18	.28	.25	.18	.06	.04	.01

   Calculate the expected number of failures in a week.

2. A quarter of the source programs submitted by a certain programmer compile successfully. Each day the programmer writes five programs. The compiling probabilities are:

No. that compiles	0	1	2	3	4	5
Probability	.237	.396	264	.088	.014	.001

   Calculate the average number of programs that compile in a day.

VARIANCES

Variance of Sample:

Spread of individual values from the mean.

Example 1: Variances

Salaries of 6 recent computer science graduates (000 euro)

$$20.3, 14.9, 18.9, 21.7, 16.3, 17.7$$

Recall $\overline{x} = 18.3$.

Calculation of s^2:

$$s^2 = \frac{[(20.3 - 18.3)^2 + (18.9 - 18.3)^2 + \ldots]}{5}$$

Generally, if $x_1, x_2, \cdots x_n$ is a sample of size n,

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$s^2 = \frac{[(20.3 - 18.3)^2 + (18.9 - 18.3)^2 + \ldots]}{5} = 6.368$$

Standard Deviation:

$$s = \sqrt{6.368} = 2.523$$

Example 2:   Variances

One hundred applicants for a certain degree program had the foll-wing age distribution.

Age	No. of applicants
18	9
19	40
20	18
21	18
22	8
23	4
24	3

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2 f_i}{\sum_i f_i}$$

The variance is obtained as follows:

$$\overline{x} = (18-20)^2 * .09 + (19-20)^2 * .40 + (20-20)^2 * .18 + (21-20)^2 * .18$$
$$+(22-20)^2 * .08 + (23-20)^2 * .04 + (24-20)^2 * .03$$

$$\sum (x-20)^2 p(x)$$

Check

$$s^2 = 2.1$$

The variance of a discrete random variable

Definition:

The variance is defined as the weighted average of the squared differences between each possible outcome and its mean ... the weights being the probability of the respective outcomes

$$V(X) = \sum_x (x - \mu_x)^2 p(x)$$

Or equivalently

$$V(X) = E\left(X - (E(X)\right)^2$$

The variance is often denoted by σ_x^2.

$$\sigma_x^2 \equiv E(X - \mu_x)^2$$

The standard deviation is denoted by σ_x.

$$\sigma_x \; = \sqrt{E(X - \mu_x)^2}$$

Worked Example

A quarter of the source programs submitted by a certain programmer compile successfully. Each day the programmer writes five programs. The compiling probabilities are:

No. that compiles	0	1	2	3	4	5
Probability	.237	.396	.264	.088	.014	.001

Calculate the expected number of programs that will compile per day

$E(X) = 0\times.237+1\times.396+2\times.264+3\times.088+4\times.014+5\times.001 = 1.25$

In R

```
x<- 0:5
prob <- c( .237,  .396,  .264,  .088,  .014,  .001)
expectation <-sum(x*prob)
expectation
[1] 1.249
```

Variance

X	0	1	2	3	4	5
P(X=x)	.237	.396	.264	.088	.014	.001
$(x-1.25)^2$	1.5625	.0625	.5625	3.0625	7.5625	14.0625
$(x-1.25)^2 p(x)$	.3703	.0247	.1485	.2695	.1059	.0140

$$V(X) = \sum_x (x-\mu_x)^2 p(x) = .933$$

In R

```
x<- 0:5
prob <- c( .237,  .396,  .264,  .088,  .014,  .001)
expectation <-sum(x*prob)
variance <- sum((x-expectation)^2 *prob)
```

Properties of Expectations

Example: Hardware Failures

The number of hardware failures X of a computer system in a week of operation, which has the probability density function:

Number of Failures (x)	0	1	2	3	4	5	6
Probability ($p(x)$)	.17	.27	.25	.17	.08	.04	.02

The monetary loss due to failures can be estimated in euros using the following linear function:

$$\text{Loss } = 10X + 200$$

Estimate

- the expected weekly loss;
- the variance of the expected weekly loss.

In this case we need to know

$$E(10X + 200)$$

$$V(10X + 200)$$

Recall:

$$E(X) = 1.904$$

and

$$V(X) = 2.0869$$

Properties

1. $E(X+c) = E(X) + c$:
   i.e. *adding a constant to each value has the effect of increasing the average by the constant amount*

$$\begin{aligned} E(X+c) &= \sum_x (x+c)p(x) \\ &= \sum_x xp(x) + c\sum_x p(x) \\ &= E(X) + c. \end{aligned}$$

   since $\sum_x p(x) = 1$ and $\sum_x xp(x) = E(X)$

2. $E(cX) = cE(X)$:
   i.e. *multplying each value by a constant amount increases the average by this amount*

$$\begin{aligned} E(cX) &= \sum_x cxp(x) \\ &= c\sum_x xp(x) \\ &= cE(X). \end{aligned}$$

   Hardware Failures:

$$E(10X + 200) = 10(1.904) + 200 = 219.04$$

3. $V(X+c) = V(X)$:
   i.e. *adding a constant to each value of the variable does not affect the variance*

$$
\begin{aligned}
V(X+c) &= E\left(X+c-E(X+c)\right)^2 \\
&= E\left(X-E(X)\right)^2 \\
&= V(X)
\end{aligned}
$$

4. $V(cX) = c^2 V(X)$:
   i.e. *multiplying each value of the variable by a constant c causes the variance to be multiplied by the square of that constant c^2*

$$
\begin{aligned}
V(cX) &= E\left(cX-E(cX)\right)^2 \\
&= E\left(cX-cE(X)\right)^2 \\
&= E\left(c(X-E(X))\right)^2 \\
&= \sum_x \left(c(x-E(X))\right)^2 p(x) \\
&= c^2 \sum_x \left(x-E(X)\right)^2 p(x) \\
&= c^2 E\left(X-E(X)\right)^2 \\
&= c^2 V(X)
\end{aligned}
$$

Hardware Failures:

$$V(10X+200) = 10^2 V(X) = 10^2(2.0869) = 208.69$$

Example:

The average salary of new employees in a computer firm is 27,500 euros with a variance of 400.

After negotiations with the trade union, it was agreed that employees would get a rise of 100 euros in addition to 10 percent increase on their basic salaries. What is the new average salary?

Solution

Let X = old salary; Y = new salary.

$$Y = 100 + 1.1X$$

Since $E(X) = 27,500$ euros, then

$E(Y) = 100+1.1E(X) = 100+1.1(27,500) = 30,350$ *euro*

Since V(X) = 400

$$V(Y) = 1.1^2 V(X) = 1.1^2(400) = 484$$