Random Variables

A <u>random variable</u> X (← always use capital letter) is a variable which represents the outcome of a random event.

Examples

1/ X = height of a randomly chosen person in this class.

2/ X = Wait time for a subway train.

3/ X = Lifetime of a light bulb.

X a <u>discrete random variable</u> (D.R.V.) ⇒ X can only take discrete values.
E.g. $X = 1, 2, 3, 4 \ldots$

X a <u>continuous random variable</u> (C.R.V.) ⇒ X can take any value on some interval.
E.g. X in $[0, \infty)$.

Examples

1/ X = Age to <u>nearest year</u> of randomly chosen person is <u>discrete</u>

2/ X = Exact age of a randomly chosen person is <u>continuous</u>.

Pr(Outcome) = **Probability** of outcome = Number in $[0,1]$ which represents the likelihood of outcome occuring.

Pr(Outcome) = 0 ⇒ Outcome will never occur.

Pr(Outcome) = 1 ⇒ Outcome will always occur.

Pr(Outcome) = A ⇒ If event is repeated many times the outcome occurs (100A) % of the time.

E.g. $Pr(X = n)$ = Probability random variable takes value n.

$Pr(a \le X \le b)$ = Probability random variable takes value between a and b.

Fact : If X is a D.R.V. with possible values $a_1, a_2, a_3, \dots, a_n$

$$\Rightarrow \quad Pr(X = a_1) + Pr(X = a_2) + \dots + Pr(X = a_n) = 1$$

If X is a D.R.V. with possible values $a_1, a_2, a_3, \dots$ ← infinite sequence

$$\Rightarrow \quad Pr(X = a_1) + Pr(X = a_2) + \dots = 1$$

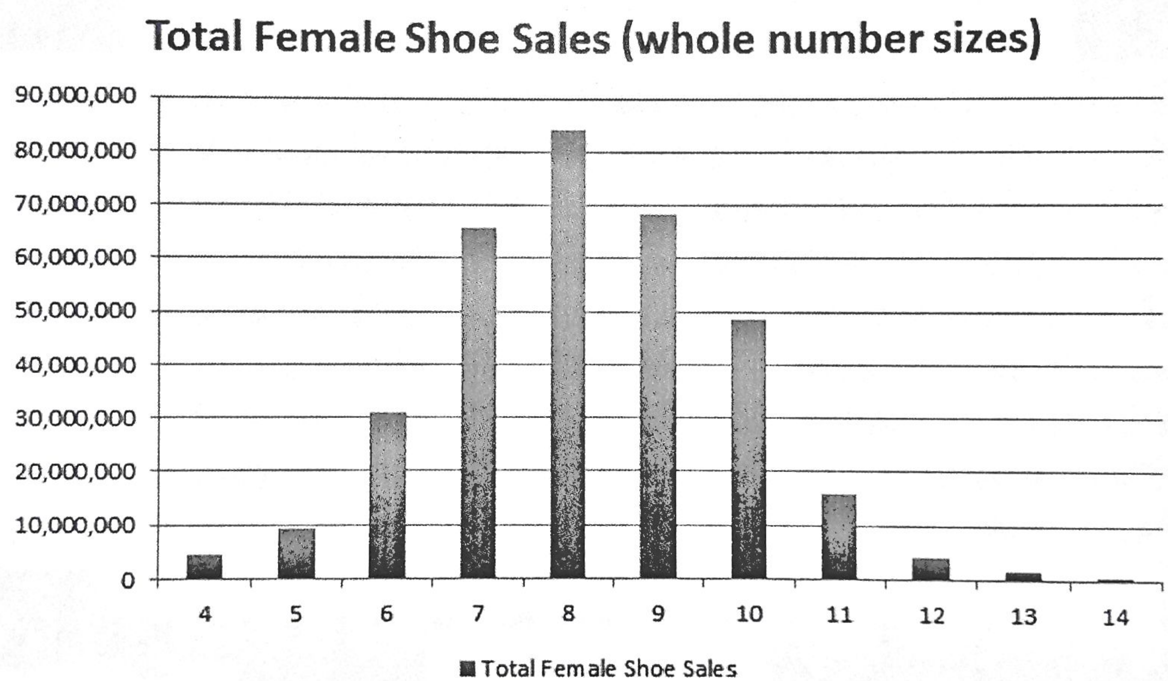
↗ infinite series

Can we come up with a good way to visualise probabilities?

<u>Example</u> X = Shoe size of randomly selected woman.

If we take whole number sizes this is a <u>D.R.V.</u>

We can represent the entire sample set in a histogram / bar graph of total shoe sales :



What is $Pr(X = 8) = ?$

$$Pr(X=8) = \frac{\text{Number of women with shoe size 8}}{\text{Total number of women}} = \frac{\text{Area of bar at 8}}{\text{Total area of all bars}}$$

Clever Idea : Scale histogram so total area of all bars is 1.

$\Rightarrow$ $Pr(X = 8)$ = Area of bar at 8.

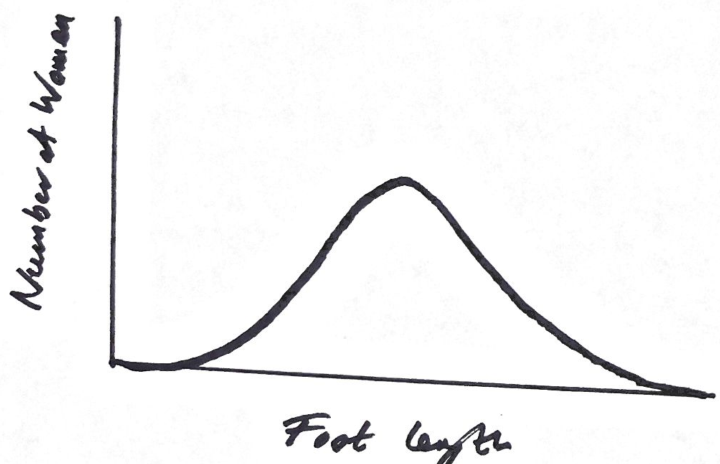
or $Pr(7 \leq X \leq 11)$ = Sum of areas of bars 7, 8, 9, 10 and 11

So we can interpret probabilities as areas.

Now let X = Exact foot length of a randomly chosen women.

Because it's exact foot length X is a C.R.V.

We could still represent the entire sample set with a graph:



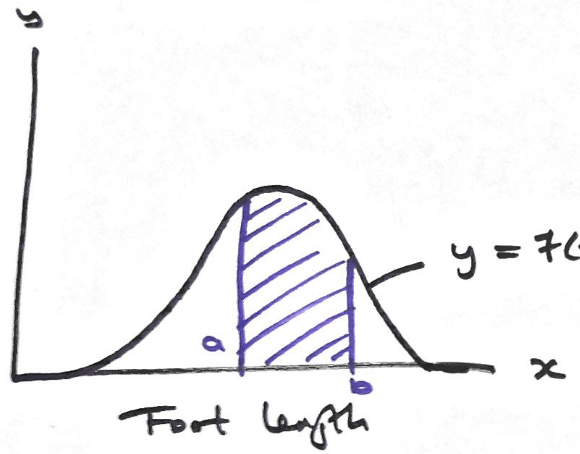
What is $Pr(a \leq X \leq b)$?

$$Pr(a \leq X \leq b) = \frac{\text{Number of women with foot length between a and b}}{\text{Total number of women}} = \frac{\text{Area under curve between a and b}}{\text{Total area under curve}}$$

Clever Idea : Scale so <u>total</u> area under curve is 1

$\Rightarrow$ $Pr(a \leq X \leq b)$ = Area under curve between a and b.

Scaled Picture



$Pr(a \leq X \leq b)$ = Area (shaded)

$= \int_a^b f(x)\,dx$

$y = f(x)$ called the Probability Density Function (P.D.F) of the C.R.V. X.

Conclusions

X <u>D.R.V</u> with possible values $a_1, a_2, a_3, \ldots$ $\Rightarrow$ there are numbers $Pr(X = a_n)$ for $n \geq 1$ with

1/ $Pr(X = a_n) \geq 0$ for $n \geq 1$

2/ $\sum_{n=1}^{\infty} Pr(X = a_n) = 1$

X C.R.V. with possible values in interval $[A, B]$ $\Rightarrow$ there is a function f on $[A, B]$ called the probability density function of X such that

1/ $f(x) \geq 0$ on $[A, B]$

2/ $\int_A^B f(x)\,dx = 1$

3/ $Pr(a \leq X \leq b) = \int_a^b f(x)\,dx$

<u>Remark</u> Any f on $[A, B]$ which satisfies 1/ and 2/ is called a P.D.F.

If we replace $[A, B]$ with $[A, \infty)$, $(-\infty, B]$ or $(-\infty, \infty)$ we replace 2/ with $\int_A^\infty f(x)\,dx = 1$, $\int_{-\infty}^B f(x)\,dx = 1$, $\int_{-\infty}^\infty f(x)\,dx = 1$

Study of C.R.V.s = Study of P.D.F.s

<u>Example</u> Find k such that $f(x) = kx^2$ is a P.D.F. on $[0, 4]$

1/ $f(x) \geq 0$ on $[0, 4]$ $\Rightarrow$ $k \geq 0$

2/ $\int_0^4 f(x)\,dx = \int_0^4 kx^2\,dx = \frac{k}{3}x^3 \Big|_0^4 = \frac{64k}{3}$

$\int_0^4 f(x)\,dx = 1 \Rightarrow \frac{64}{3}k = 1 \Rightarrow k = \frac{3}{64}$.

Example Find B such that $f(x) = \sin(x)$ is a P.D.F. on $[0, B]$.

1/ $B \leq \pi$ as $\sin(x) \leq 0$ on $[\pi, 2\pi]$

2/ $$\int_0^B \sin(x)\,dx = -\cos(x)\Big|_0^B = -\cos(B) - (-\cos(0))$$
$$= 1 - \cos(B)$$

$$\int_0^B \sin(x)\,dx = 1 \iff 1 - \cos(B) = 1 \iff \cos(B) = 0 \Rightarrow B = \frac{\pi}{2}$$

Example Consider a cell population which is growing. When a cell is T days old it divides into two new cells.

Experimental Fact: If cell population is big, the proportion of cells at various ages is constant.

X = age of randomly chosen cell.

X is a C.R.V. with possible values in $[0, T]$

<u>Cool Fact</u> : X has P.D.F. $f(x) = \frac{2\ln(2)}{T} e^{-\frac{\ln(2)}{T}x}$

Verify this is a P.D.F. on $[0,T]$

1/ $\frac{2\ln(2)}{T} e^{-\frac{\ln(2)}{T}x} \geqslant 0$ on $[0,T]$

2/ $\int_0^T \frac{2\ln(2)}{T} e^{-\frac{\ln(2)}{T}x} = \frac{\left(\frac{2\ln(2)}{T}\right)}{\left(-\frac{\ln(2)}{T}\right)} e^{-\frac{\ln(2)}{T}x} \Big|_0^T$

$= -2\left(e^{-\ln(2)} - 1\right) = -2 \cdot \left(\frac{1}{2} - 1\right) = 1$

What is the probability a randomly chosen cell is age less than $\frac{T}{2}$?

$Pr\left(0 \leq X \leq \frac{T}{2}\right) = \int_0^{T/2} \frac{2\ln(2)}{T} e^{-\frac{\ln(2)}{T}x} dx = -2e^{-\frac{\ln(2)}{T}x} \Big|_0^{T/2}$

$$= -2\left(e^{-\frac{\ln(2)}{2}} - 1\right) = -2\left(\frac{1}{\sqrt{2}} - 1\right) = 2 - \sqrt{2}$$

<u>Example</u> The parent company for a franchised chain of restaurants claims the proportion of their new restaurants making a profit has P.D.F. $f(x) = 12x(1-x)^2$.

What is the probability that more than or equal to 50% of new restaurants will make a profit?

X = Proportion of new restaurants making a profit

X C.R.V. on $[0,1]$ with P.D.F. $f(x) = 12x(1-x)^2$

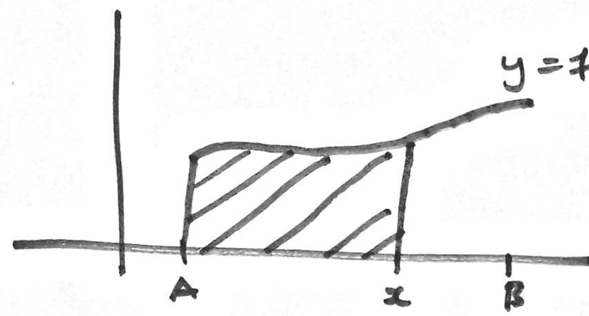
$$Pr(0.5 \le X \le 1) = \int_{0.5}^{1} 12x(1-x)^2\, dx = \int_{0.5}^{1} 12x - 24x^2 + 12x^3\, dx$$

$$= 6x^2 + 8x^3 + 3x^4 \Big|_{0.5}^{1} = 0.3125$$

Let X be a C.R.V. on $[A, B]$ with P.D.F. $f(x)$

Define the new function $F(x) = Pr(X \le x)$ ← "Cumulative Distribution Function" of X

$$F(x) = \int_A^x f(t)\,dt \quad (x \leftarrow \text{variable}) \quad \Rightarrow$$



$F(x) = \text{Area}(▨)$

Facts

1/ $F(A) = 0$ and $F(B) = 1$

2/ $F(x)$ is increasing on $[A, B]$

} Follow from properties of P.D.F.

3/ $F(x) = \int_A^x f(t)\,dt \Rightarrow F'(x) = f(x)$ (Fundamental Theorem)

($F(x)$ an antiderivative of $f(x)$)

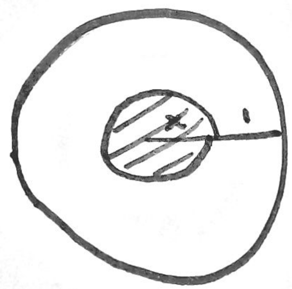
$$\Rightarrow Pr(a \le X \le b) = \int_a^b f(x)\,dx = F(b) - F(a)$$

Example: Select a random point within a circle of radius 1

X = distance from center of this random point.

What is the P.D.F. of X?

$Pr(X \le x) = ?$



$\Rightarrow Pr(X \le x) = \dfrac{\text{Area of small circle}}{\text{Area of whole circle}}$

$\Rightarrow Pr(X \le x) = \dfrac{\pi x^2}{\pi \cdot 1^2} = x^2 \quad (= F(x))$

$\Rightarrow f(x) = 2x$