Random variables

Random variables assign a number to each possible outcome of a random process.

Random variables have a distribution.

A probability function, $p(x) = \Pr(X = x)$.

or a density function, $f(x)$

(for which areas under the curve correspond to probabilities)

Mean (or *expected value*) of R.V., $E(X)$

The typical value.

Standard deviation , $SD(X)$

A measure of the spread of the distribution.
The typical deviation from the mean.

Y = a + b X

If X is a random variable, and a and b are numbers,

$Y = a + b X$ *is another random variable.*

$E(Y) = a + b\, E(X)$

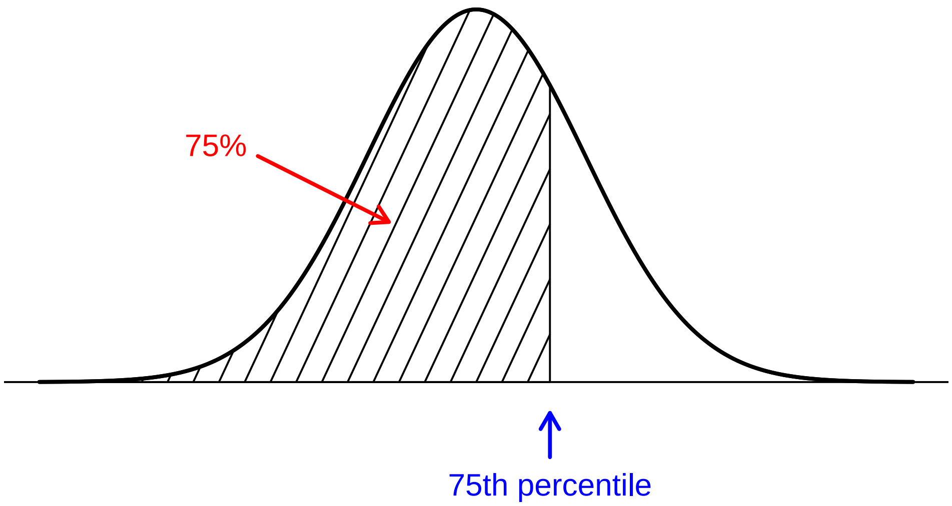
$SD(Y) = |b|\, SD(X)$

Important example:

If $Z = [\, X - E(X)\,] / SD(X)$

then $E(Z) = 0$ and $SD(Z) = 1$.

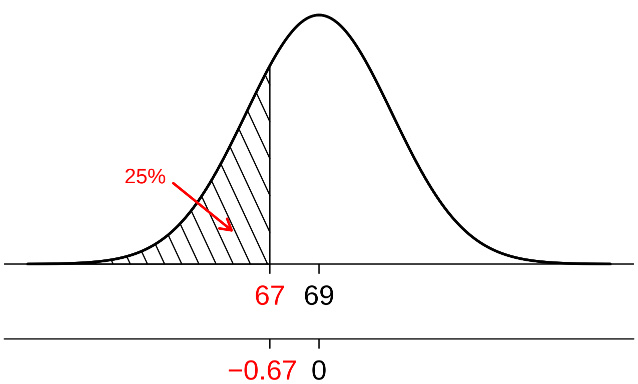
Percentiles



Example

Suppose the heights of adult males in the U.S. are approximately normal distributed, with mean = 69 in and SD = 3 in.

What is the 25th percentile?



$X \sim \mathrm{N}(\mu{=}69,\ \sigma{=}3)$

$Z = (X - 69)/3 \sim \mathrm{N}(0,1)$

$X = Z \times 3 + 69$

$\Pr(Z \leq -0.67) \approx 0.25$

$\Pr(X \leq -0.67 \times 3 + 69) \approx 0.25$

$\Pr(X \leq 67) \approx 0.25$

Multiple Random Variables

We essentially always consider multiple RV's at once.

Example: Sample a random student from the class.

$$\text{Let } X = \begin{cases} 0 & \text{if female} \\ 1 & \text{if male} \end{cases}$$

$$Y = \begin{cases} 0 & \text{if Windows} \\ 1 & \text{if Mac} \\ 2 & \text{if Both} \end{cases}$$

Joint distribution

X = gender (0 if female, 1 if male)

Y = OS (0 if Windows, 1 if Mac, 2 if both)

	y = 0	y = 1	y = 2
x = 0	0.55	0.14	0.01
x = 1	0.22	0.06	0.01

$p_{XY}(x,y) = \Pr(X=x \text{ and } Y=y)$

Marginal distributions

X = gender (0 if female, 1 if male)

Y = OS (0 if Windows, 1 if Mac, 2 if both)

		y = 0	y = 1	y = 2	
x	0	0.55	0.14	0.01	0.70
	1	0.22	0.06	0.01	0.30
		0.77	0.20	0.02	

Pr(X=x) = $\sum_y$ Pr(X=x and Y=y)   Pr(Y=y) = $\sum_x$ Pr(X=x and Y=y)

Conditional distribution

X = gender (0 if female, 1 if male)

Y = OS (0 if Windows, 1 if Mac, 2 if both)

		y = 0	y = 1	y = 2	
x	0	0.55	0.14	0.01	0.70
	1	0.22	0.06	0.01	0.30
		0.77	0.20	0.02	

Pr(Y=0 | X=1) = Pr(X=1 and Y=0) / Pr(X=1) = 0.22/0.30 $\approx$ 0.733

Pr(Y=1 | X=1) = Pr(X=1 and Y=1) / Pr(X=1) = 0.06/0.30 = 0.200

Pr(Y=2 | X=1) = Pr(X=1 and Y=2) / Pr(X=1) = 0.01/0.30 $\approx$ 0.033

Independence

Random variables X and Y are independent if:

$p_{XY}(x,y) = p_X(x)\, p_Y(y)$ for every pair x,y

In other words/symbols:

$\Pr(X = x \text{ and } Y = y) = \Pr(X = x)\, \Pr(Y = y)$ for every pair x,y

Equivalently,

$\Pr(X = x \mid Y = y) = \Pr(X = x)$ for all x,y

$\Pr(Y = y \mid X = x) = \Pr(Y = y)$ for all x,y

Another example

Sample a couple who are both carriers of some disease gene

X = no. children they have

Y = no. affected children they have

		x						
	$p_{XY}(x,y)$	0	1	2	3	4	5	$p_Y(y)$
	0	0.160	0.248	0.124	0.063	0.025	0.014	0.634
	1	0	0.082	0.082	0.063	0.034	0.024	0.285
y	2	0	0	0.014	0.021	0.017	0.016	0.068
	3	0	0	0	0.003	0.004	0.005	0.012
	4	0	0	0	0	0.000	0.001	0.001
	5	0	0	0	0	0	0.000	0.000
	$p_X(x)$	0.160	0.330	0.220	0.150	0.080	0.060	

$\Pr(Y = y \mid X = 2)$

$p_{XY}(x,y)$		x: 0	1	2	3	4	5	$p_Y(y)$
y	0	0.160	0.248	0.124	0.063	0.025	0.014	0.634
	1	0	0.082	0.082	0.063	0.034	0.024	0.285
	2	0	0	0.014	0.021	0.017	0.016	0.068
	3	0	0	0	0.003	0.004	0.005	0.012
	4	0	0	0	0	0.000	0.001	0.001
	5	0	0	0	0	0	0.000	0.000
$p_X(x)$		0.160	0.330	0.220	0.150	0.080	0.060	

y	0	1	2	3	4	5
Pr(Y=y | X=2)	0.564	0.373	0.064	0.000	0.000	0.000

$\Pr(X = x \mid Y = 1)$

$p_{XY}(x,y)$		x: 0	1	2	3	4	5	$p_Y(y)$
y	0	0.160	0.248	0.124	0.063	0.025	0.014	0.634
	1	0	0.082	0.082	0.063	0.034	0.024	0.285
	2	0	0	0.014	0.021	0.017	0.016	0.068
	3	0	0	0	0.003	0.004	0.005	0.012
	4	0	0	0	0	0.000	0.001	0.001
	5	0	0	0	0	0	0.000	0.000
$p_X(x)$		0.160	0.330	0.220	0.150	0.080	0.060	

x	0	1	2	3	4	5
Pr(X=x | Y=1)	0.000	0.288	0.288	0.221	0.119	0.084

Example

Sample a random rat from Baltimore.

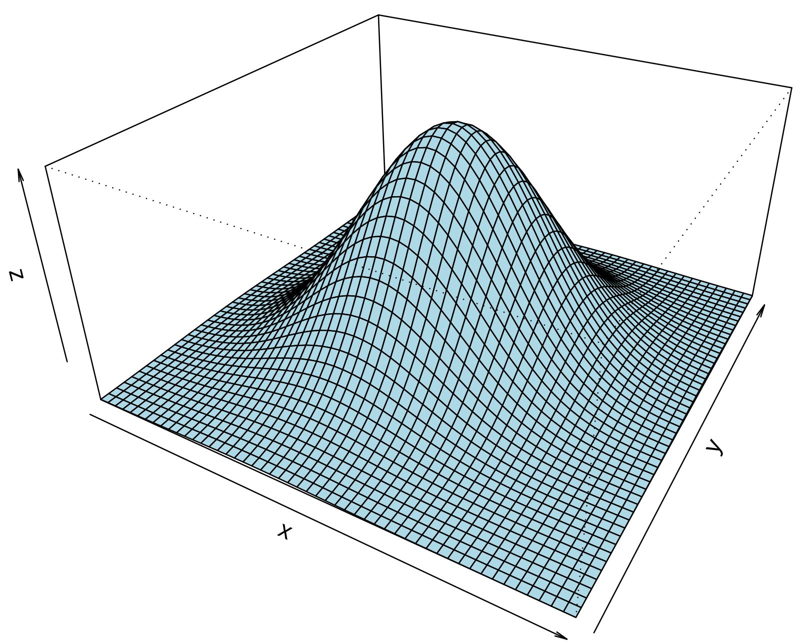
X = 1 if the rat is infected with virus A, and = 0 otherwise

Y = 1 if the rat is infected with virus B, and = 0 otherwise

		x = 0	x = 1	
$p_{XY}(x,y)$		0	1	$p_Y(y)$
y	0	0.72	0.18	0.90
	1	0.08	0.02	0.10
	$p_X(x)$	0.80	0.20	

Continuous random variables

Continuous random variables have joint densities, say $f_{XY}(x,y)$.



Continuous random variables

Continuous random variables have joint densities, say $f_{XY}(x,y)$.

The marginal densities are obtained by integration (areas of slices):

$$f_X(x) = \int f_{XY}(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int f_{XY}(x,y)\,dx$$

Conditional densities:

$$f_{X|Y=y}(x) = f_{XY}(x,y)/f_Y(y)$$

$$f_{Y|X=x}(y) = f_{XY}(x,y)/f_X(x)$$

X and *Y* are independent if

$f_{XY}(x,y) = f_X(x)\, f_Y(y)$ for all x,y

iid

More jargon:

Random variables $X_1, X_2, X_3, \ldots, X_n$ are said to be independent and identically distributed (iid) if:

(a) they are independent and

(b) they all have the same distribution

Usually:

- Repeated independent measurements
- Random sampling from a large population

Means and SDs

Mean and SD of sums of random variables:

$E(\sum_i X_i) = \sum_i E(X_i)$ — no matter what

$SD(\sum_i X_i) = \sqrt{\sum_i \{SD(X_i)\}^2}$ — if the X_i are independent

Mean and SD of means of random variables:

$E(\sum_i X_i / n) = \sum_i E(X_i)/n$ — no matter what

$SD(\sum_i X_i/n) = \sqrt{\sum_i \{SD(X_i)\}^2}/n$ — if the X_i are independent

If the X_i are iid with mean μ and SD σ:

$E(\sum_i X_i / n) = \mu$ and $SD(\sum_i X_i / n) = \sigma/\sqrt{n}$