Multiple Random Variables

Multiple random variables

We essentially always consider multiple random variables at once.

$\longrightarrow$ The key concepts: Joint, conditional and marginal distributions, and independence of RVs.

Let X and Y be discrete random variables.

$\longrightarrow$ Joint distribution:

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

$\longrightarrow$ Marginal distributions:

$$p_X(x) = \Pr(X = x) = \sum_y p_{XY}(x,y)$$
$$p_Y(y) = \Pr(Y = y) = \sum_x p_{XY}(x,y)$$

$\longrightarrow$ Conditional distributions:

$$p_{X|Y=y}(x) = \Pr(X = x \mid Y = y) = p_{XY}(x,y) \,/\, p_Y(y)$$

Example

Sample a couple who are both carriers of some disease gene.

X = number of children they have
Y = number of affected children they have

$p_{XY}(x,y)$	y \ x	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 | X = 2)

$p_{XY}(x,y)$	y \ x	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	

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 | 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

Independence

Random variables X and Y are independent if

$\longrightarrow$ $p_{XY}(x,y) = p_X(x)\, p_Y(y)$

for every pair x,y.

In other words/symbols:

$\longrightarrow$ $\Pr(X = x \text{ and } Y = y) = \Pr(X = x)\Pr(Y = y)$

for every pair x,y.

Equivalently,

$\longrightarrow$ $\Pr(X = x \mid Y = y) = \Pr(X = x)$

for all x,y.

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		
$p_{XY}(x,y)$		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, $f_{XY}(x,y)$.

$\longrightarrow$ The marginal densities are obtained by integration:

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

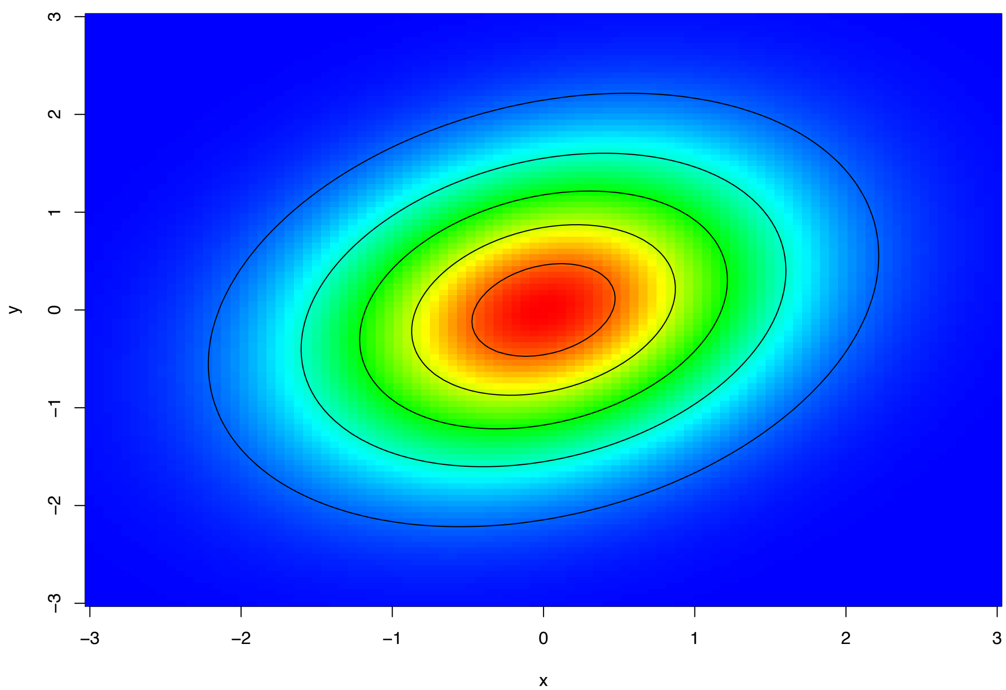
$\longrightarrow$ Conditional density:

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

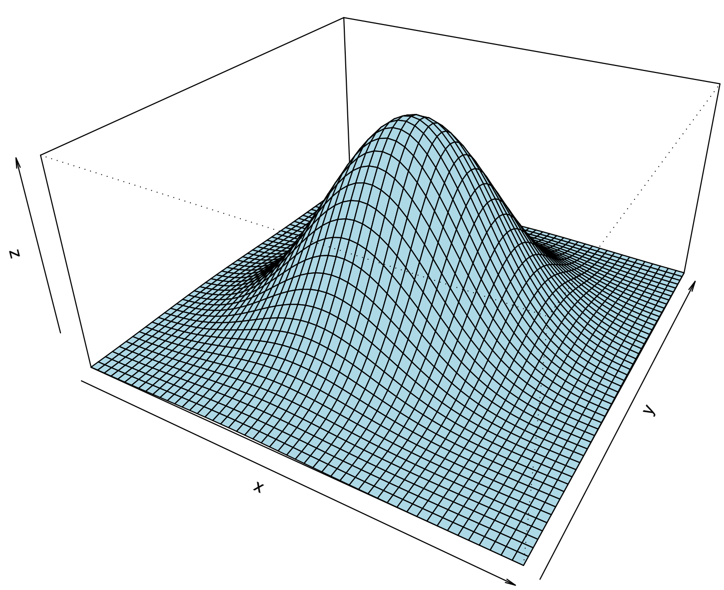
$\longrightarrow$ *X* and *Y* are independent if:

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

The bivariate normal distribution



The bivariate normal distribution



iid

More jargon:

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

- $\longrightarrow$ they are independent,
- $\longrightarrow$ they all have the same distribution.

Usually such RVs are generated by

- $\longrightarrow$ repeated independent measurements, or
- $\longrightarrow$ random sampling from a large population.

Means and SDs

$\longrightarrow$ Mean and SD of sums of random variables:

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

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

$\longrightarrow$ Mean and SD of means of random variables:

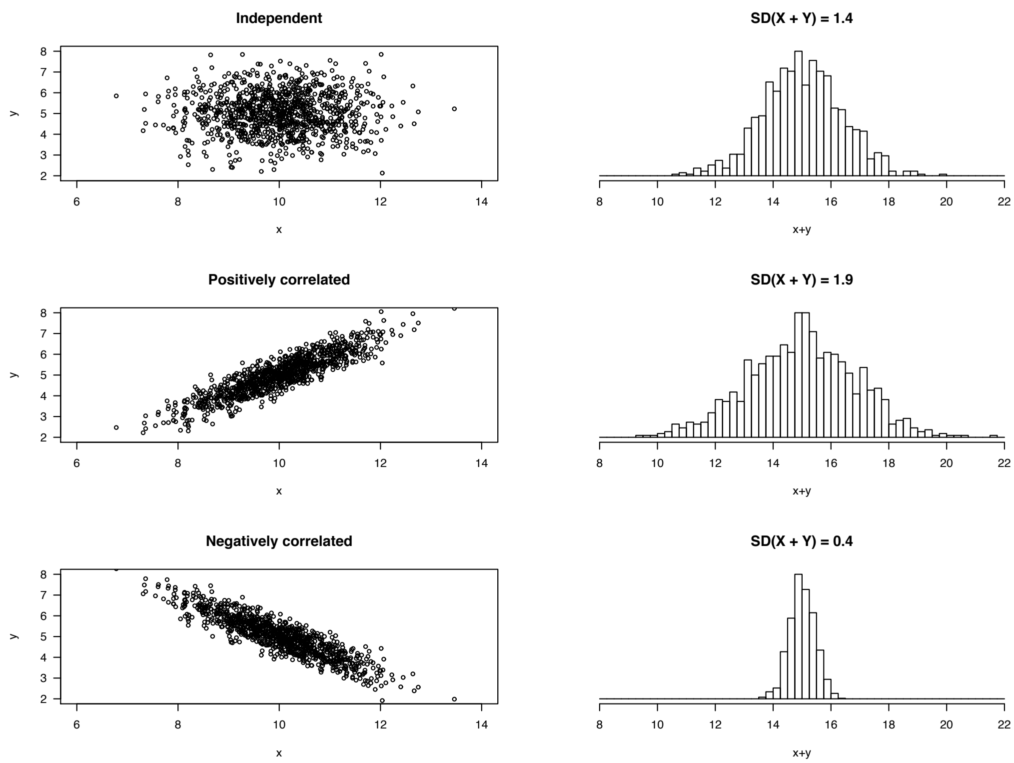
$\mathrm{E}(\sum_i X_i \,/\, n) = \sum_i \mathrm{E}(X_i)/n$ — no matter what

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

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

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

Example


Independent
SD(X + Y) = 1.4
Positively correlated
SD(X + Y) = 1.9
Negatively correlated
SD(X + Y) = 0.4
x
y
x+y