

Discrete vs Continuous RV.

A discrete rv. can take only a finite # of values, that are counted using positive integers

- # children / household
- # trips / week

Dummy variables are artificial variables that measure the existence of some qualitative characteristic.

$$D = \begin{cases} 1 & \text{if event occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Example $D_g = \begin{cases} 1 & \text{if female} \\ 0 & \text{if male.} \end{cases}$

A continuous RV. can take any real value in at least 1 interval on the real # line.

There are an uncountably infinite # of possibilities.

$GNP \in [0, \infty)$

weights

interest rates.

Prices.

Probability Distributions

We can make statements about the probability of certain values of a R.V. occurring by specifying a probability distribution.

If there is an event A that is the outcome of some experiment then the probability of A , $P(A)$, is defined as the relative frequency with which A occurs in many (as in infinite) experimental trials. For any event

$$0 \leq P(A) \leq 1$$

and the total prob of observing all events = 1.

Discrete R.V.

The probability density function of a discrete R.V. lists each possible outcome and the prob with which it occurs.

Example: Die. (fair)

$x = \# \text{ dots}$	$f(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

$$P(X=2) = \frac{1}{6}$$

Continuous

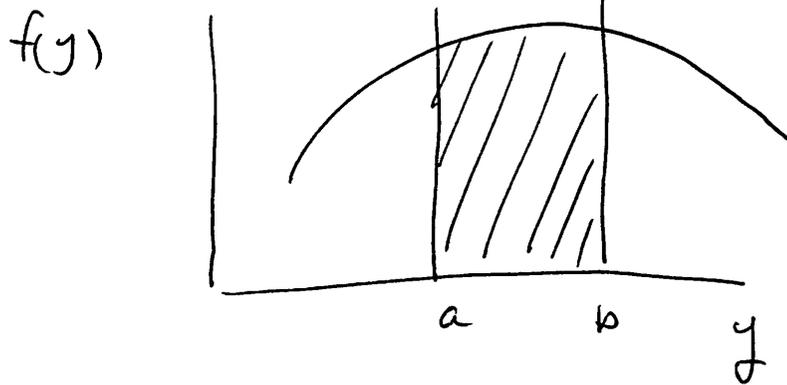
For a continuous r.v. the pdf is represented by an equation which can be described graphically as a line or curve.

To get probability of an event (an interval in this case) you compute an area below the curve.

$$f(y) \quad y \in \mathcal{S} \text{ (sample space)}$$

$$P(a \leq Y \leq b)$$

17



To be a P.D.F. the function $f(y)$ must be such that

i) $f(y) \geq 0$ for all y .

ii) $\int_{-\infty}^{\infty} f(y) dy = 1$

iii) $\int_a^b P(a \leq Y \leq b) = \int_a^b f(y) dy.$

Measure of P.D.F.

Expectations of R.V.

Mathematical Expectations reveal important characteristics of R.V. For instance

$E(X)$ is the mean of the statistical population of the R.V. X .

It is the average value of X in an infinite # of statistical trials

If X is a discrete r.v. that can take on values x_1, x_2, \dots, x_n with probs $f(x_1), f(x_2), \dots, f(x_n)$, respectively then the expected value of X is

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \dots + x_n f(x_n)$$

Die

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

It is a weighted average where the weights are the probabilities of various values of X occurring.

Continuous

$$E(X) = \int_S x f(x) dx$$

Expectations of functions of R.V.s

Suppose I have some function of X $g(x)$ that I want the mean of.

$$E[g(X)] = g(x_1) \cdot f(x_1) + g(x_2) \cdot f(x_2) + \dots + g(x_n) \cdot f(x_n)$$

$$= \sum_n g(x_n) f(x_n) \quad \text{discrete}$$

$$\approx \int_S g(x) f(x) dx \quad \text{continuous.}$$

$$\begin{aligned}
 E(X^2) &= 1^2 \frac{1}{6} + 2^2 \frac{1}{6} + 3^2 \frac{1}{6} + 4^2 \frac{1}{6} + 5^2 \frac{1}{6} + 6^2 \frac{1}{6} = \\
 &= \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} = \frac{91}{6}
 \end{aligned}$$

The variance is an example

$$\text{Var}(X) = E[(X - E(X))^2] = \sigma_x^2 \geq 0$$

Var is the average squared distance between X and its mean. Measures dispersion.

Properties of Expectations

Let a, b be constants and X be a r.v. with mean μ, σ^2 $X \sim (\mu, \sigma^2)$

$$\begin{aligned}
 E[a + b(X)] &= E(a) + b E(X) \\
 &= a + b\mu
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(a + bX) &= E[(a + bX - E(a + bX))^2] \\
 &= E[b^2(X - E(X))^2] = b^2 \text{Var}(X)
 \end{aligned}$$

- 1) Adding a constant to a r.v. shifts the mean but has no effect on variance.
- 2) Multiply a r.v. by a constant and its variance Δ 's by the square of the constant.

Joint Density functions

Frequently, we make probability statements about two or more r.v.
To do this, we have to know their joint p.d.f.

Example:

$$G = \begin{cases} 1 & \text{male} \\ 0 & \text{female} \end{cases} \quad P = \begin{cases} 0 & \text{Democrat} \\ 1 & \text{Rep.} \\ 2 & \text{other} \end{cases}$$

		Gender		h(p)
		Fem 0	male 1	
Party	0	.20	.27	.47
	1	.30	.10	.4
	2	.06	.07	.13
P(S)		.56	.44	

So, The Prob($P=1$ and $G=0$) = ~~.27~~ .30
 And so on. The probabilities must add to 1 which indicates that the entire sample space is covered by the J.P.D.F.

$f(1,0) = .3$

in the continuous case there is a multidimensional surface and the total volume under that surface must = 1.

$$X, Y \sim f(x, y)$$

Marginal densities

Given the joint pdf, we can make probability statements about individual elements using the marginal pdf.

Let $f(x, y)$ is joint pdf of X, Y

$$f(x, \cdot) = \sum_y f(x, y)$$

For the continuous case

$$f(x) = \int f(x, y) dy$$

Example:

Females
Add across
party

$$P(G=0 \text{ (Female)}) = f(0) = \sum_{P,G} f(1,0) + f(2,0) + f(3,0) \\ = .2 + .3 + .06 = .56$$

Republicans.
Add across
gender

$$P(P=1) = h(1) = \sum_{P,G} f(1,0) + f(1,1) \\ = .30 + .10 = .40$$

Conditional pdf.

Often the probability of an event occurring depends on whether other events have occurred.

$$P(\text{wet street} \mid \text{rain})$$

These conditional probabilities can be determined from knowledge of the joint and marginal p.d.f's.

$$P[X=x \mid Y=y] = f(x|y) = \frac{f(x,y)}{f(y)}$$

What is the probability of ~~Matrix~~ ~~Rep~~ a person being Republican given that she the person selected is Female

$$P[P=\text{Rep} \mid G=\text{Female}] \begin{matrix} \swarrow \text{Rep \& Female} \\ \searrow \text{Female} \end{matrix}$$

$$f(P=1 \mid G=1) = \frac{f(1,1)}{f(1)} = \frac{.30}{.56} = .535$$

$$\text{Dem} \mid \text{Female} \quad P(P=D \mid G=F) = .20 / .56 = .357$$

$$\text{Other} \mid \text{Female} \quad P(P=\text{Other} \mid G=\text{Female}) = .06 / .56 = .107$$

Independent R.V.s

If the prob of observing one event is not affected by the outcome of other R.V.s then they are statistically independent.

Recall that in def of conditional prob.

Def In general
$$f(x|y) = \frac{f(x,y)}{f(y)}$$

$$\therefore f(x,y) = f(x|y) \cdot f(y)$$

But, if $f(x|y)$ is the same as $f(x)$ which means, knowledge of y has not affect on the prob of observing x then

$f(x,y) = f(y) \cdot f(x)$ and x & y are said to be statistically independent.

If $f(y) \cdot f(x) \neq f(x,y)$ for at least 1 combo. then they are NOT stat indep.

$$P(P=1) \cdot P(G=0) = (.4)(.56) = .224 \neq .30$$

\therefore They can't be independent.

Expected values of several R.V.S.

One question that is often asked, "Do variables tend to move together?"

A measure of linear association is covariance

$$\text{Cov}(X, Y) = E \left[(X - E(X))(Y - E(Y)) \right] = \sum_x \sum_y [(x - E(X))(y - E(Y)) f(x, y)]$$

Voting example.

$$E(X) =$$

$$\text{Cov}(P, G) = E \left[(P - E(P))(G - E(G)) \right]$$

$$E(P) = \sum p \cdot h(p) = 0 \cdot .47 + 1 \cdot .4 + 2 \cdot (.13) = .66$$

$$E(G) = \sum g \cdot f(g) = 0 \cdot (.56) + 1 \cdot (.44) = .44$$

In this example, there are 6 terms to add together. For instance

$$(0 - E(.66))($$

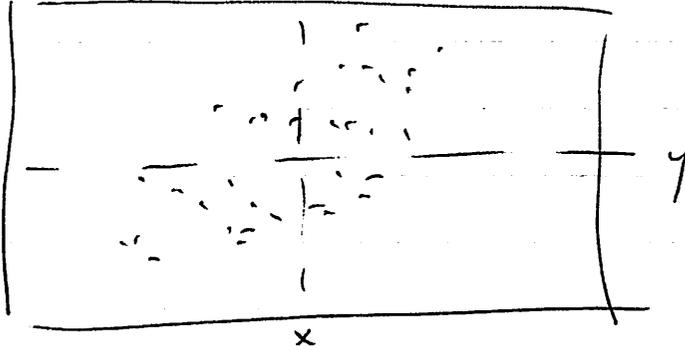
$$\begin{aligned} \text{Cov}(P, G) = & (0 - .66)(0 - .44) f(0, 0) + (0 - .66)(1 - .44) f(0, 1) \\ & + (1 - .66)(0 - .44) f(1, 0) + (1 - .66)(1 - .44) f(1, 1) \\ & + (2 - .66)(0 - .44) f(2, 0) + (2 - .66)(1 - .44) f(2, 1) \end{aligned}$$

=

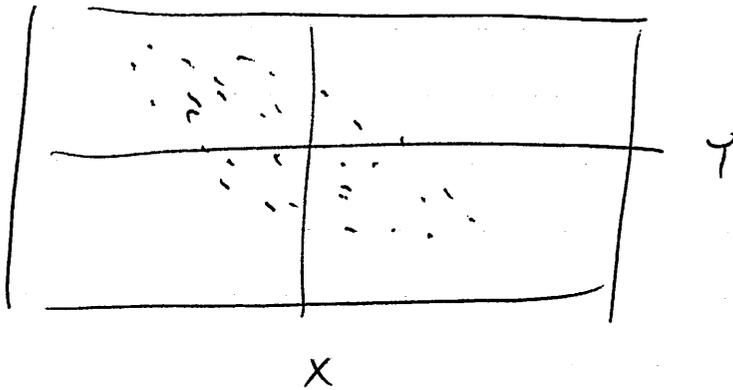
Positive covariance means that large values of X tend to be associated with large values of Y

Negative

$$\text{Cov}(X, Y) > 0$$

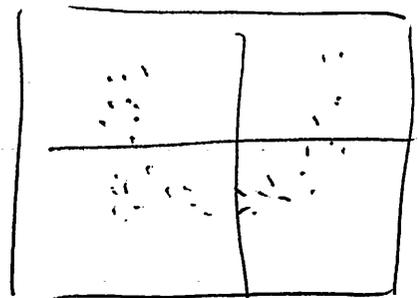
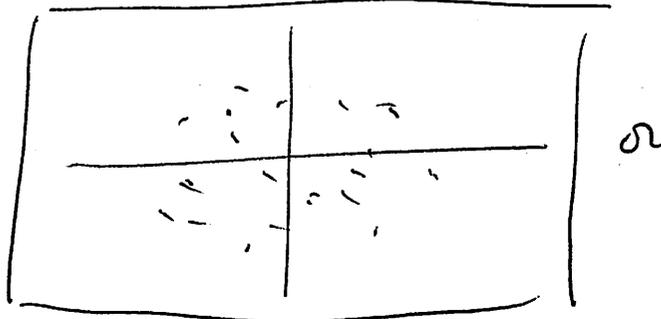


$$\text{Cov}(X, Y) < 0$$



$$\text{Cov}(X, Y) = 0$$

Linear Assoc



The magnitude of $\text{Cov}(X, Y)$ is impossible to interpret since it can be changed simply by changing the scale by which your variables are measured.

Pennies to dollars

So, correlation is often used instead.

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

$$-1 \leq \rho \leq 1$$

Values closer to 1 or -1 indicate stronger pos and neg relationships.

Independence $\Rightarrow \text{Cov}(X, Y) = 0$

But converse not true.

Rule:

c_1, c_2, \dots, c_n be constants
and X_1, X_2, \dots, X_n be R.V.

$$E[c_1 X_1 + c_2 X_2 + \dots + c_n X_n] =$$

$$c_1 E(X_1) + c_2 E(X_2) + \dots + c_n E(X_n)$$

$$\text{Var}(c_1 X_1 + c_2 X_2) = c_1^2 \text{Var}(X_1) + c_2^2 \text{Var}(X_2) \\ + 2c_1 c_2 \text{Cov}(X_1, X_2)$$

$$\text{Var}(c_1 X_1 + c_2 X_2 + \dots + c_n X_n) =$$

$$c_1^2 \text{Var}(X_1) + c_2^2 \text{Var}(X_2) + \dots + c_n^2 \text{Var}(X_n) \\ + \cancel{2c_1 c_2 \text{Cov}(X_1, X_2)} + \cancel{2c_1 c_3 \text{Cov}(X_1, X_3)} \\ + \sum_{i \neq j} c_i c_j \text{Cov}(X_i, X_j)$$

If X_i are indep then the cov terms
become zero.