Conditional, Joint and Marginal Distributions

In general we want to use probability to address problems involving more than one variable at the time

Think back to our first question on the returns of my portfolio... if we know that the economy will be growing next year, does that change the assessment about the behavior of my returns?

We need to be able to describe what we think will happen to one variable relative to another...

Conditional, Joint and Marginal Distributions

How are my sales impacted by the overall economy?

Let E denote the performance of the economy next quarter... for simplicity, say $E = 1$ if the economy is expanding and $E = 0$ if the economy is contracting. Let's assume $pr(E = 1) = 0.7$

Conditional, Joint and Marginal Distributions

Let S denote my sales next quarter... and let's suppose the following probability statements:

S	$pr(S\|E=1)$	S	$pr(S\|E=0)$
1	0.05	1	0.20
2	0.20	2	0.30
3	0.50	3	0.30
4	0.25	4	0.20

These are called *Conditional Distributions*

Conditional, Joint and Marginal Distributions

S	$pr(S\|E=1)$	S	$pr(S\|E=0)$
1	0.05	1	0.20
2	0.20	2	0.30
3	0.50	3	0.30
4	0.25	4	0.20

- In blue is the conditional distribution of S given $E = 1$
- In red is the conditional distribution of S given $E = 0$
- We read: *the probability of Sales of 4 ($S = 4$)* **given (or conditional on)** *the economy is growing ($E = 1$) is 0.25*

Conditional, Joint and Marginal Distributions

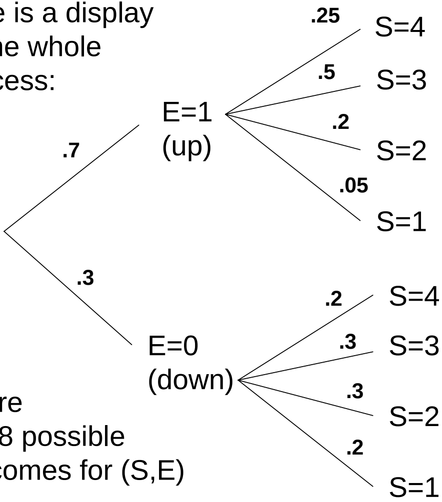
The conditional distributions tell us about about what can happen to S for a given value of E... but what about S and E jointly?

$$\begin{aligned} pr(S = 4 \quad \text{and} \quad E = 1) &= pr(E = 1) \times pr(S = 4 | E = 1) \\ &= 0.70 \times 0.25 = 0.175 \end{aligned}$$

In english, 70% of the times the economy grows and 1/4 of those times sales equals 4... 25% of 70% is 17.5%

Conditional, Joint and Marginal Distributions

here is a display of the whole process:



p(S=4 and E=1) = .7*.25 = .175

p(S=3 and E=1) = .7*.5 = .35

p(S=2 and E=1) = .7*.2 = .14

p(S=1 and E=1) = .7*.05 = .035

p(S=4 and E=0) = .3*.2 = .06

p(S=3 and E=0) = .3*.3 = .09

p(S=2 and E=0) = .3*.3 = .09

p(S=1 and E=0) = .3*.2 = .06

There are 8 possible outcomes for (S,E)

Conditional, Joint and Marginal Distributions

We call the probabilities of E and S together the joint distribution of E and S.

In general the notation is...

- $pr(Y = y, X = x)$ is the joint probability of the random variable Y equal y **AND** the random variable X equal x.
- $pr(Y = y|X = x)$ is the conditional probability of the random variable Y takes the value y **GIVEN** that X equals x.
- $pr(Y = y)$ and $pr(X = x)$ are the marginal probabilities of $Y = y$ and $X = x$

Conditional, Joint and Marginal Distributions

Why we call marginals marginals... the table represents the joint and at the margins, we get the marginals.

E \ S	1	2	3	4	
0	.06	.09	.09	.06	.3
1	.035	.14	.35	.175	.7
	.095	.23	.44	.235	1

Conditional, Joint and Marginal Distributions

Example... Given $E = 1$ what is the probability of $S = 4$?

S

E		1	2	3	4	
	0	.06	.09	.09	.06	.3
	1	.035	.14	.35	.175	.7
		.095	.23	.44	.235	1

$$pr(S = 4|E = 1) = \frac{pr(S = 4, E = 1)}{pr(E = 1)} = \frac{0.175}{0.7} = 0.25$$

Conditional, Joint and Marginal Distributions

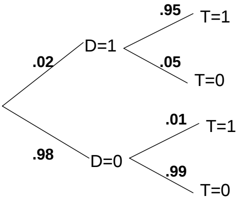
Example... Given $S = 4$ what is the probability of $E = 1$?

		S				
		1	2	3	4	
E	0	.06	.09	.09	.06	.3
	1	.035	.14	.35	.175	.7
		.095	.23	.44	.235	1

$$pr(E = 1|S = 4) = \frac{pr(S = 4, E = 1)}{pr(S = 4)} = \frac{0.175}{0.235} = 0.745$$

Disease Testing Example

Let $D = 1$ indicate you have a disease
Let $T = 1$ indicate that you test positive for it



$\Longrightarrow$

		D	
		0	1
T	0	.9702=.98*.99	.001
	1	.0098	.019

If you take the test and the result is positive, you are really interested in the question: Given that you tested positive, what is the chance you have the disease?

Disease Testing Example

		D = 0	D = 1
T	0	.9702	.001
	1	.0098	.019

$$pr(D = 1 | T = 1) = \frac{0.019}{(0.019 + 0.0098)} = 0.66$$

Disease Testing Example

- Try to think about this intuitively... imagine you are about to test 100,000 people.
- we assume that about 2,000 of those have the disease.
- we also expect 1% of the disease-free people to test positive, ie, 980, and 95% of the sick people to test positive, ie 1,900. So, we expect a total of 2,880 positive tests.
- Choose one of the 2,880 people at random... what is the probability that he/she has the disease?

$$p(D = 1|T = 1) = 1,900/2,880 = 0.66$$

- isn't that the same?!

Prosecutor's Fallacy

- Assume you are in a jury for a murder trial in a city of roughly 10 million people.
- In his opening remarks, the prosecutor tells you that the defendant has been arrested on the strength of a single, overwhelming piece of evidence: that his DNA matched a sample of DNA taken from the scene of the crime.
- To convince you of the strength of this evidence, the prosecutor calls a forensic scientist to the stand, who testifies that the probability that an innocent person's DNA would match the sample found at the crime scene is only one in a million
- Would you vote to convict this man?

Prosecutor's Fallacy

- Denote G for guilty ... what is $P(G)$=?
- Denote DNA for a positive DNA match... what is $P(DNA|not\ G)$?
- What is the relevant probability to inform your decision? $P(DNA|not\ G)$ or $P(not\ G|DNA)$? Calculate and compare both.