Conditional Distributions and Joint Distributions

Rob McCulloch

1. Conditional, Joint and Marginal Distributions
2. Bayes Theorem
3. Several Variables
4. Independence
5. IID

1. Conditional, Joint, and Marginal Distributions

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

For example we may need want to:

- describe our uncertainty about several quantities together
  (the joint distribution)
- understand how learning values about some variables affects our beliefs about others
  (the conditional distribution).

Suppose you are thinking about sales next quarter.
In order to think about sales, it may be helpful to think about sales *and* what will happen to the 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 (what kind of random variable is this?)

Let's assume $p(E = 1) = 0.7$

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

s	$p(S = s \mid E = 1)$	s	$p(S = s \mid E = 0)$
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*

s	$p(S = s \mid E = 1)$	s	$p(S = s \mid E = 0)$
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*.

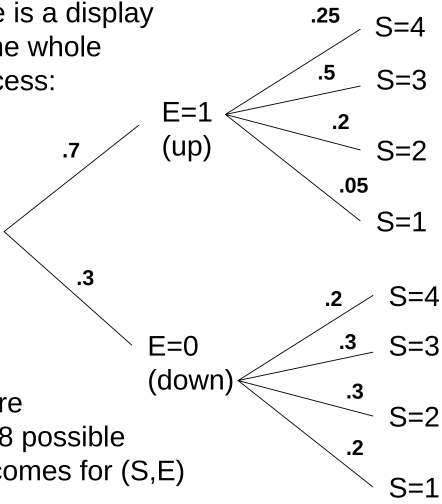
The way S is related to E is captured by the difference between the conditional distributions !!!

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?

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

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

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)

We can specify the distribution of the pair of random variables (S, E) by listing all possible pairs and the corresponding probability.

(s, e)	$p(S = s, E = e)$
(1,1)	.035
(2,1)	.14
(3,1)	.35
(4,1)	.175
(1,0)	.06
(2,0)	.09
(3,0)	.09
(4,0)	.06

Question: What is $Pr(S = 1)$?

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

In general the notation is...

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

Important relationships

Relationship between the joint and conditional...

$$
\begin{aligned}
Pr(Y = y, X = x) &= Pr(X = x) \times Pr(Y = y | X = x) \\
&= Pr(Y = y) \times Pr(X = x | Y = y)
\end{aligned}
$$

Relationship between joint and marginal...

$$
\begin{aligned}
Pr(X = x) &= \sum_y Pr(X = x, Y = y) \\
Pr(Y = y) &= \sum_x Pr(X = x, Y = y)
\end{aligned}
$$

A Note on Notation

We have used the notations

$$P(Y = y), \;\; P(Y = y, X = x), \;\; P(Y = y \mid X = x)$$

You will see all kinds of similar, but not exactly the same notations for these fundamental concepts.

For example, we will sometimes use $p(x, y)$ in place of $P(X = x, Y = y)$ when it is clear from the context what we mean.

For example in the (S, E) example I could write $P(S = s, E = e)$ or just $p(s, e)$.

The Two-Way Table Display of the Joint Distribution

This is a nice way to display a joint distribution.
Same information as when we just listed the (s, e) pairs and their probabilities but this way we can see some things more easily.

S

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

For example, you can see why the marginals are called "the marginals".

Conditionals from Joints

We derived the joint distribution of (E, S) from the marginal for E and the conditional $S \mid E$.

You can also calculate the conditional from the joint by doing it the other way

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

$\Rightarrow$

$$Pr(Y = y \mid X = x) = \frac{Pr(Y = y, X = x)}{Pr(X = x)}$$

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

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

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

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

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

2. Bayes Theorem

So, in general, you can compute the joint from marginals and conditionals and the other way around.

How you think about stuff depends on what's easiest or what you know, or what you care about.

Suppose you toss two coins: X is the first, Y is the second.
In each case 1 means a head and 0 a tail.
What is $P(X = 1, Y = 1)$ = P(two heads) ?

(1) Directly figure out the joint distribution.
There are 4 possible outcomes for the two coins and each is equally likely so it is $1/4$.

(2) Figure out some marginals and conditionals.
$P(X = 1, Y = 1) = P(X = 1) * P(Y = 1 \mid X = 1) =$
$(1/2) * (1/2) = 1/4$.

Bayes Theorem refers to the situation where we build a model for (Y, X) by thinking about

$$Pr(Y = y \mid X = x), \;\; Pr(X = x).$$

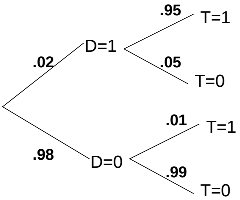
and then, having observed $Y = y$ compute

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

Example: Disease Testing

Disease testing example...

Let $D = 1$ indicate you have a disease.
Let $T = 1$ indicate that you test positive for it
We know the marginal of D and the conditional of T given D.



$\Longrightarrow$

		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?

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

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

The computation of $p(x|y)$ from $p(x)$ and $p(y|x)$ is called Bayes theorem...

$$p(x|y) = \frac{p(y,x)}{p(y)} = \frac{p(y,x)}{\sum_x p(y,x)} = \frac{p(x)p(y|x)}{\sum_x p(x)p(y|x)}$$

In the disease testing example:

$$p(D=1|T=1) = \frac{p(T=1|D=1)p(D=1)}{p(T=1|D=1)p(D=1)+p(T=1|D=0)p(D=0)}$$

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

3. Several Variables

Of course, we often want to think about more than two variables at a time.

We can extend the ideas we used with two variables to many variables.

Suppose we have the three random variables,

$$(Y_1, Y_2, Y_3)$$

Then,

$$Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) =$$

$$Pr(Y_1 = y_1)\, Pr(Y_2 = y_2 \mid Y_1 = y_1)\, Pr(Y_3 = y_3 \mid Y_1 = y_1, Y_2 = y_2)$$

Or, using more succinct notation:

$$p(y_1, y_2, y_3) = p(y_1)\, p(y_2 \mid y_1)\, p(y_3 \mid y_1, y_2)$$

You can keep going for any number of variables !!

Example

Suppose we have 10 voters.
4 are republican and 6 are democratic.

We randomly choose 3.

Let Y_i be 1 if the i^{th} voter is a democrat and 0 otherwise, $i = 1, 2, 3$.

What is

$$Pr(Y_1 = 1, Y_2 = 1, Y_3 = 1)$$

What is the probability of getting three democrats in a row ??

$Pr(Y_1 = 1, Y_2 = 1, Y_3 = 1) =$

$Pr(Y_1 = 1)\, p(Y_2 = 1 \mid Y_1 = 1)\, Pr(Y_3 = 1 \mid Y_1 = 1, Y_2 = 1)$

$= (6/10)(5/9)(4/8)$

$= (1/6) = .167.$

When we randomly pick a person from a population of people, and then randomly pick a second from the ones left, and so on, we call it *sampling without replacement*.

If we put the person back each time and randomly choose from the whole group each time, then we call it *sampling with replacement*.

Suppose we sample from our 10 voters with replacement.

Now what is

$$Pr(Y_1 = 1, Y_2 = 1, Y_3 = 1)$$

$$Pr(Y_1 = 1, Y_2 = 1, Y_3 = 1) =$$

$$Pr(Y_1 = 1)\, p(Y_2 = 1 \mid Y_1 = 1)\, Pr(Y_3 = 1 \mid Y_1 = 1, Y_2 = 1)$$

$$= (6/10)(6/10)(6/10)$$

$$= .6^3 = .216$$

Notice that when we sample with replacement

$$p(Y_2 = 1 \mid Y_1 = y_1) \text{ and } Pr(Y_3 = 1 \mid Y_1 = y_1, Y_2 = y_2)$$

do not depend on y_1 and y_2.

What happens for Y_1 does not affect what we think will happen for Y_2 and what happens for Y_1 *and* Y_2 does not affect what will happen for Y_3.

In this case we say the random variables are *independent*.

4. Independence

Given a bunch of random variables, we say they are independent of each other if the conditional distribution of any one of them does not depend on anything you might observe for any of the others.

Example

Suppose I am about to toss 100 coins.

Let Y_i be 1 if the i^{th} coin is a head and 0 otherwise.

What is $Pr(Y_3 = 1)$?

What is $Pr(Y_3 = 1 \mid Y_1 = 1, Y_2 = 0)$?
What is $Pr(Y_3 = 1 \mid Y_1 = 0, Y_2 = 1)$?
What is $Pr(Y_3 = 1 \mid Y_1 = 0, Y_2 = 0)$?
What is $Pr(Y_3 = 1 \mid Y_1 = 1, Y_2 = 1)$?

What is

$Pr(Y_{100} = 1 \mid Y_1 = 1, Y_2 = 1, \ldots, Y_{99} = 1)$ first 99 are heads?

What is

$Pr(Y_1 = 1, Y_2 = 1, Y_3 = 1, \ldots, Y_{100} = 1)$ 100 heads in a row!!?

Independence, Conditional Equals Marginal

If random variables are independent then the conditional is the marginal.

For two random variables X and Y if X and Y are independent then,

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

does not depend on x.

We also have:

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

What you believe about Y knowing $X = x$, is the same as what you believe if you know nothing about X.

Example

X is 1 if first coin is head.
Y is 1 if second coin is head.

What is

$P(Y = 1 \mid X = 1)$?

What is

$P(Y = 1 \mid X = 0)$?

What is

$P(Y = 1)$?

Independence, Joints, and Marginals

If X and Y are independent then the joint is the product of the marginals:

$$\begin{aligned} p(x, y) &= p(x)\, p(y \mid x) \\ &= p(x)\, p(y) \end{aligned}$$

This also works "the other way", that is, if the joint is the product of the marginals then they are independent.

Example

You are about to manufacture two parts.

$X = 1$ if part one fails, 0 else.
$Y = 1$ if part two fails, 0 else.

The table below gives the joint distribution of X and Y.

		X = 0	X = 1
Y	0	.72	.08
	1	.18	.02

		X	
		0	1
Y	0	.72	.08
	1	.18	.02

$Pr(Y = 1 \mid X = 0) = .18/.9 = .2$
$Pr(Y = 1 \mid X = 1) = .02/.1 = .2$
$Pr(Y = 1) = 18 + .02 = .2$

$Pr(Y = 1, X = 1) = .02$
$Pr(Y = 1)Pr(X = 1) = .2 * .1 = .02.$

X and Y are independent.

For random variables Y_i, if they are independent we have

$p(y_1, y_2, \ldots, y_n) =$

$p(y_1)\, p(y_2 \mid y_1)\, p(y_3 \mid y_1, y_2)\, \ldots\, p(y_n \mid y_1, y_2, \ldots, y_{n-1})$

$= p(y_1)\, p(y_2)\, p(y_3)\, \ldots\, p(y_n)$

Example

If Y_i is 1 if the i^{th} coin is a head, 0 else, $i = 1, 2, \ldots, 10$, what is

$$p(1, 1, \ldots, 1)$$

10 heads in a row?

$p(1, 1, \ldots, 1) =$

$= p(1)\, p(1)\, p(1)\, \ldots\, p(1)$

$= .5^{10} = 0.0009765625.$

5. IID

Suppose we are about to toss 100 coins.

Let Y_i be 1 if heads, 0 else,

We usually think the Y's are independent.

In addition, we usually think they are *identically distributed*, that is, each one has *the same marginal distribution*.

What is $Pr(Y_{20} = 1)$?
What is $Pr(Y_{98} = 1)$?

When random variables are independent and identically distributed we say they are **IID**.

I: independent

ID: identically distributed.

In our coins example, each $Y_i \sim \text{Bernoulli}(.5)$.

We can succinctly describe coin tossing by

$$Y_i \sim \text{Bernoulli}(.5), \quad IID$$

Example

Suppose we have 10 voters. 4 are republican and 6 are democratic.

We randomly choose 3, sampling *with replacement*.

Let Y_i be 1 if the i^{th} voter is a democrat and 0 otherwise, $i = 1, 2, 3$.

How can we describe the joint distribution of (Y_1, Y_2, Y_3), are they IID?

Example

Suppose we have 10 voters. 4 are republican and 6 are democratic.

We randomly choose 3, sampling *without replacement*.

Let Y_i be 1 if the i th voter is a democrat and 0 otherwise, $i = 1, 2, 3$.

How can we describe the joint distribution of (Y_1, Y_2, Y_3), are they IID?

Example

Suppose we have 1,000,000 voters. 400,000 are republican and 600,000 are democratic.

We randomly choose 3, sampling *without replacement*.

Let Y_i be 1 if the i^{th} voter is a democrat and 0 otherwise, $i = 1, 2, 3$.

How can we describe the joint distribution of (Y_1, Y_2, Y_3), are they IID?

Example

Suppose I am about to toss a die 100 times.

Let D_i be the outcome for the i^{th} toss
(a number in $\{1, 2, 3, 4, 5, 6\}$).

Are the D_i IID?

Example

Suppose an experienced NBA player is about to take repeated free-throws.

Let Y_i be 1 if he makes the i^{th} attempt and 0 otherwise.

Are these Y_i iid Bernoulli?

This is known at the “hot hand” question.

Most people who play sports believe that the can get get “hot” so that if they made the last few, they are more likely to make the next one.

However, if you look at the data, it looks IID Bernoulli!!

How do you look at the data to see if it looks IID Bernoulli.
That is covered in the notes “Modeling with IID Bernoulli Draws”.

Example

Suppose the first penalty in an NHL game is on team A.

For subsequent penalties $P_i = 1$ if the penalty is on a different team than the previous one and 0 otherwise.

Are the P's independent?

Are they IID?

These are not IID.

If the last two (or three!) penalties were on the same team, it becomes quite a bit more likely that the next penalty will be on the other team.

See
Reversal of fortune: a statistical analysis of penalty calls in the national hockey league", (2014), Journal of Quantitative Analysis in Sports 10 (2), 207-224 (Jason Abrevaya and Robert McCulloch)

Example

Suppose you are monitoring a stock and for every 10 minute interval, you record whether the price went up or down.

Let U_i be 1 if it goes up in the i^{th} interval, 0 otherwise.

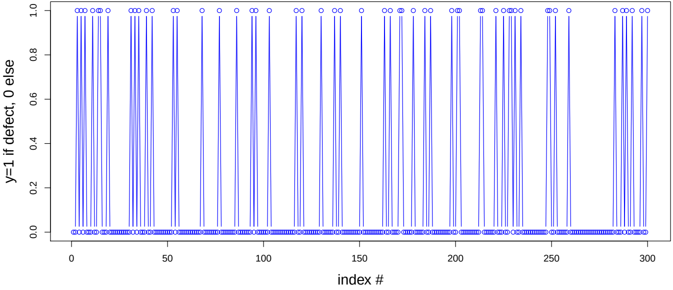
Are the U_i IID?

Of course, this is a much studied question.

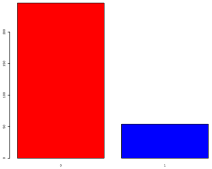
We leave this to your finance courses but just note that it is very interesting how little dependence there is!!!!

Example: Modeling Defects:

You are in charge of a manufacturing process.
For the last 300 parts made, we have 1 if the part is defective and 0 otherwise.



54/300=.18 of the parts are defective.

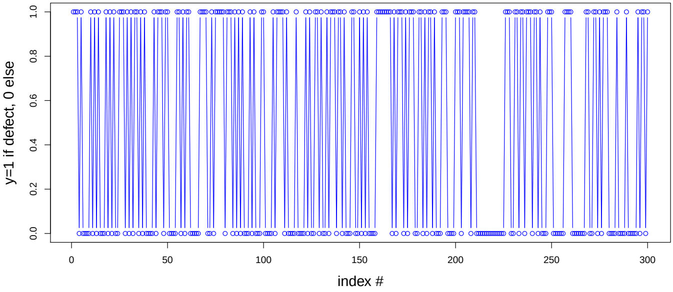


We are about to make the next part.

Let Y be 1 if it turns out to be defective and 0 if it is good.

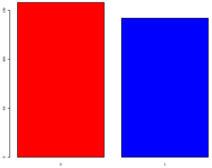
What is the distribution of Y?

Here are 300 coin tosses, 1 if heads, 0 if tails.

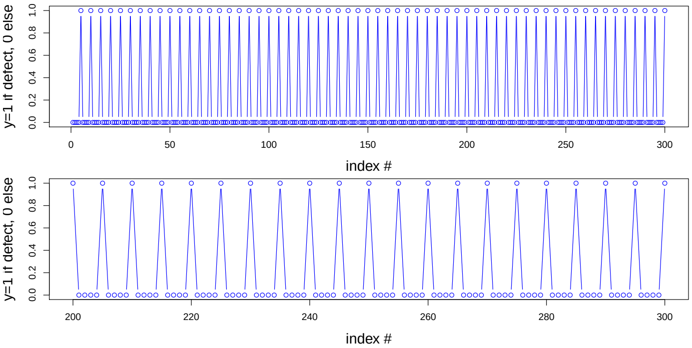


$142/300 = .47$ of the coins are heads.

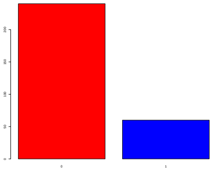
If $Y = 1$ means the next one turns out to be a head, what is the distribution of Y?



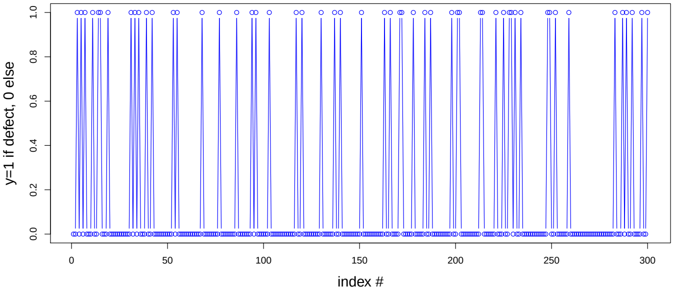
Here are 300 observations on another 0/1 process.
The second plot just shows the last 100.



60/300=.20 of the observations are 1.



If Y denotes the next one, what is the distribution of Y?



The defects look like the kind of thing you would get as IID draws from a Bernoulli, with $p = .18$.

We model:

$$Y_i \sim \text{Bernoulli}(.18).$$

So, a plausible choice for distribution of the next one is just $Y \sim \text{Bernoulli}(.18)$.

What is the probability the next part made is defective.?

What is the probability the next 10 parts are good?

Note:

For $Y_i \sim \text{Bernoulli}(p)$, the usual estimator of p is the observed fraction of 1's.

If we call this estimator $\hat{p}$, then,

- we have the *standard error*: $se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- and the associated 95% *confidence interval* for p: $\hat{p} \pm 2se(\hat{p})$.

Defect Example:

$\hat{p} = .18$, $se(\hat{p}) =$ sqrt(.18*(1-.18)/300) $= .022$

Approximate 95% confidence interval: .18 +/- .044 = (.14,.22).

Is it big?

Note:

For the the "sampling from a finite population problem" and the defects we used the IID Bernoulli model.

But,

the way we justify the use of the model
is completely different !!!!!

Note:

For the deterministic $0/1$ process we can summarize the numbers by saying we observed 20% ones. But, is that an estimate of anything?