Session 7: Jointly distributed Random Variables

Susan Thomas

http://www.igidr.ac.in/~susant

susant@mayin.org

IGIDR

Bombay

Goals

- Joint probability distributions
- Marginal probabilities
- Conditional probability distributions
- Conditional density functions
- Independance
- Covariance
- Conditional mean and variance

Discrete RVs

One discrete variable

X	$\Pr(x)$
0	0.2
1	0.2
2	0.6
	1

- Being a pdf, the probabilities are all ≥ 0 and add up to one.
- Knowing the pdf is the most you can know about X.

The joint distribution of X and Y

- Suppose an experiment consists of producing 2 random outcomes X and Y.
- X assumes values 0, 1, 2. Y assumes values 0 and 1.
- The joint distribution $\Pr(X = x, Y = y)$ shows all probabilities of the events that can come about.
- These correspond to statements $\Pr((X = x)\text{and}(Y = y))$.

Joint distribution: $\Pr(X = x, Y = y)$

	Y		
X	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

- All the cells contain joint probabilities.
- They add up to 1
- This joint pdf is the most you can know about the joint variation of X and Y.

Recovering $\Pr(X)$

	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

- How to reduce from $\Pr(X = x, Y = y)$ to $\Pr(X)$?

Recovering $\Pr(X)$

	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

- How to reduce from $\Pr(X = x, Y = y)$ to $\Pr(X)$?
- Addup all the ways in which you can get $X = 2$

Recovering $\Pr(X)$

	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

- How to reduce from $\Pr(X = x, Y = y)$ to $\Pr(X)$?
- Addup all the ways in which you can get $X = 2$

Recovering $\Pr(X)$

	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	0.6
			1

- How to reduce from $\Pr(X = x, Y = y)$ to $\Pr(X)$?
- Addup all the ways in which you can get $X = 2$

Recovering $\Pr(X)$

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
			1

- How to reduce from $\Pr(X = x, Y = y)$ to $\Pr(X)$?
- Addup all the ways in which you can get $X = 2$
- Addup along the rows of the joint to get $\Pr(X)$
- Takes us back to the pdf of X = (0.2, 0.2, 0.6).

Same idea for recovering $\Pr(Y)$

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
			1

- Suppose we're interested in $\Pr(Y = 1)$.

Same idea for recovering $\Pr(Y)$

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
			1

- Suppose we're interested in $\Pr(Y = 1)$.
- Y can be 1 in 3 different ways. Adding up, we get 0.7.

Same idea for recovering $\Pr(Y)$

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
		0.7	1

- Suppose we're interested in $\Pr(Y = 1)$.
- Y can be 1 in 3 different ways. Adding up, we get 0.7.

Same idea for recovering $\Pr(Y)$

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- Suppose we're interested in $\Pr(Y = 1)$.
- Y can be 1 in 3 different ways. Adding up, we get 0.7.
- Similarly, we get $\Pr(Y = 0)$.
- Now we know the full distribution of Y.

“Joint” versus “Marginal” distribution

	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

- The joint distribution contains all knowable facts. From the joint, we got the 2 univariate distributions.

"Joint" versus "Marginal" distribution

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- The joint distribution contains all knowable facts. From the joint, we got the 2 univariate distributions.
- Written in the margins of the table, so the name "marginal" distributions.
- The joint is the fundamental underlying information; the marginals flow from that.

Changing your mind as information unfolds

What do you know about Y?

- Suppose X and Y have this joint distribution.
- Suppose I challenge you to make a statement about Y.
- What is the best that you can say?
- You would say:

0	1
0.3	0.7

- This is your best knowledge about Y.

Suppose the outcome of X was unfolded

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- You believe that the outlook for Y is (0.3, 0.7).
- *Suppose I told you* X. Would it change your views about the outlook for Y?

Suppose the outcome of X was unfolded

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- You believe that the outlook for Y is (0.3, 0.7).
- *Suppose I told you* X. Would it change your views about the outlook for Y?
- Example: Suppose I told you $X = 0$
- Now you know that $Y = 0$ just can't happen!

How your views change when information unfolds

- What is going on here is something remarkable!
- You believed that Y was the pdf (0.3, 0.7)
- I told you that X had come out to 0.
- Now your beliefs about Y change; now your "conditional pdf" is (0, 1), i.e. you now believe that Y will come out to 1 with certainty.

How to change our mind

- *It's more important to be correct than to be consistent.*
  – John Kenneth Galbraith
- *When the facts change, I change my mind. And what do you do, Sir?*
  – John Maynard Keynes.
- Probability gives us the scientific way to change our minds when new information unfolds.
- Research in psychology tells us that a common flaw in human reasoning is to inadequately learn from new data. Formal reasoning through probability theory will help produce better decisions.

The conditional probability

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- We start with knowing the marginal pdf.

The conditional probability

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- We start with knowing the marginal pdf.
- Now I tell you $X = 2$.

The pdf of Y once you know $X = 2$

	0	1	
0			
1			
2	0.2	0.4	0.6

- Now nothing else in the pdf matters but one row.
- You know that 0.2 and 0.4 convey the relative weights of 0 versus 1.
- But they don't add up to 1 - so they do not make a pdf.
- So we divide each by 0.6 to make the probabilities

$\Pr(Y|X = 2)$

Y	0	1
$\Pr(Y\|X = 2)$	0.333	0.666

- This is the conditional probability of Y given $X = 2$.

Problem:

For the given joint of X, Y, compute the conditional distribution $\Pr(X|Y = 1)$.

Problem:

For the given joint of X, Y, compute the conditional distribution $\Pr(X|Y = 1)$.

	0	1	2
X	0.286	0.143	0.571

It is indeed different from (.2, .2, .6) so unfolding $Y = 1$ did change my mind.

Generalising,

$$\text{Conditional} = \frac{\text{joint}}{\text{marginal}}$$

i.e.

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

Jargon

- We write $\Pr(Y|X=2)$ for the "conditional" probability.
- To emphasise the contrast, we call $\Pr(Y)$ the "unconditional" probability.
- The terms "unconditional" and "marginal" mean the same thing.
- So you must be perfect in juggling the words:
  joint,
  marginal / unconditional,
  conditional.

Marvellous interpretation

- Your views start out as the marginal / unconditional of Y.
- X unfolds.
- Now your views shift to the conditional distribution of Y.
- The shift from the unconditional to the conditional reflects your learning from the data for X.

Note a consequence

$$\text{Conditional} = \frac{\text{joint}}{\text{marginal}}$$

so

$$\text{Conditional} \cdot \text{marginal} = \text{joint}$$

i.e.

$$\Pr(Y|X = x) \cdot \Pr(X = x) = \Pr(Y = y, X = x)$$

Independence

Independence

- Suppose I tell you the outcome for X.
- Suppose *it does not change your views about* Y.
- Then X and Y are "independent".
- Example: If I tell you about tourist arrivals in Montenegro, it doesn't change your views about the pdf of Nifty returns next year.

Example

	0	1	
0	0.06	0.14	0.2
1	0.06	0.14	0.2
2	0.18	0.42	0.6
	0.3	0.7	1

- Work out the conditional $\Pr(Y|X=2)$.
- It is the same as the unconditional $\Pr(Y)$.
- Telling you that $X=2$ changed nothing.
- If, for all cases, the conditional is the same as the unconditional, then you have independence.

Independence

X and Y are independent
if and only if

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

Independence: Joint is product of marginals

	0	1	
0			0.2
1			0.2
2			0.6
	0.3	0.7	1

Independence: Joint is product of marginals

	0	1	
0	0.06	0.14	0.2
1	0.06	0.14	0.2
2	0.18	0.42	0.6
	0.3	0.7	1

Independence: Joint is product of marginals

	0	1	
0	0.06	0.14	0.2
1	0.06	0.14	0.2
2	0.18	0.42	0.6
	0.3	0.7	1

Compare against the example we used earlier –

0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

Evaluating whether there is independence

Operationally: For every cell in the joint, test whether the joint is the product of the marginals.
If for even one cell, the test fails, then independence is absent.

Flipping between joint and marginals

- If you know the joint, you can always compute the marginals.
- If you only know the marginals, in general, you can't recover the joint.
- Only in one special case – independence – can this reverse step happen.
- If I tell you the two marginals, and if I tell you that there is independence, then (and only then) you can obtain the full joint.

Summary

What have we learned?

1. The joint distribution contains all knowable information.
2. Given a joint, we can always make the marginal pdf.
3. Given a joint, we can always make the conditional pdf.
4. Our belief shifts from marginal to conditional when information unfolds.
5. Independence: When information unfolds and our beliefs don't change.
6. The marginals are enough to reconstruct the joint if and only if there is independence.
7. Without independence, if you only know the marginals, there is plenty that you don't know.

Next steps

- We built this in the context of discrete pdfs.
- This scales to continuous pdfs.

Continuous RVs

Continuous joint density functions

- The joint distribution of two continuous RVs X, Y is denoted as the joint probability density function:

$$f(X, Y)$$

- The marginal of X from $f(X, Y)$ is calculated as:

$$f_X(x) = \int_{-\infty}^{\infty} f(X, Y) dY$$

- The conditional density function of $X|Y = y$ is:

$$f_{X|Y=y}(x|Y = y) = f(X = x, y) / \int_{-\infty}^{\infty} f(X, Y) dX$$

Example: Conditional joint density functions

$$f(x,y) = \begin{cases} 6xy(2-x-y) & 0<x<1, 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

- What is the conditional density function of $(X|Y=y)$?

Example: Conditional joint density functions

$$f(x,y) = \begin{cases} 6xy(2-x-y) & 0<x<1, 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

- What is the conditional density function of $(X|Y=y)$?

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{\int_0^1 6xy(2-x-y)dx} \\ &= \frac{6x(2-x-y)}{(4-3y)} \end{aligned}$$

Summary statistics

Conditional expectation and variance

- $E(X|Y = y), \sigma^2_{(X|Y=y)}$
- In order to calculate the mean and the variance, we need to calculate the conditional density function.
- Once the conditional density function is calculated,

$$
\begin{aligned}
E(X|Y = y) &= \int x f(x, y) dX \\
Var(X|Y = y) &= \int x^2 f(x, y) dX
\end{aligned}
$$

Covariance

- Covariance between any pair of RVs (X, Y) is defined as:

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - \mu_X \mu_Y \end{aligned}$$

Covariance and correlation

- Covariance can be any number between $\infty, -\infty$.
- A standardised form of covariance is *correlation*. This is typically denoted as ρ_{xy} and is calculated as:

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- Correlations can be any number between -1 and 1.

Example: Discrete Covariance

	Y		
X	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

What is Cov(X,Y)?

Solution: Discrete Covariance

- Cov(X,Y) = E(XY) - $\mu_X \mu_Y$
- E(XY) :

XY	0	1	2
	0.5	0.1	0.4

1. E(XY) = 0*0.5 + 1*0.1 + 2*0.4 = 0.9
2. μ_X = 0*0.2 + 1*0.2 + 2*0.6 = 1.4
3. μ_Y = 0*0.3 + 1*0.7 = 0.7

Cov(X,Y) = 0.9 - 1.4*0.7 = -0.08

Mean and variance of $(X+Y)$

- E(X+Y) = E(X) + E(Y)
- Var(X+Y) = E$[((X+Y)-E(X+Y))^2]$

$$
\begin{aligned}
&= E[(X-E(X)+Y-E(Y))^2] \\
&= E[(X-E(X))^2]+E[(Y-E(Y))^2]+ \\
&\quad 2E[(X-E(X))(Y-E(Y))] \\
&= Var(X)+Var(Y)+2Cov(X,Y)
\end{aligned}
$$

Mean and variance of the returns of a 2-stock portfolio

The portfolio has funds V. Out of this, w% of funds invested in S_1, $(1-w)$% invested in S_2. S_1 has returns r_1 which is distributed as $N(\mu_1, \sigma_1^2)$. S_2 has returns $r_2 \sim N(\mu_1, \sigma_2^2)$.

- Portfolio returns: $r_p = wr_1 + (1-w)r_2$
- Expected returns: $\mathrm{E}(wr_1 + (1-w)r_2)$ = $w\mu_1 + (1-w)\mu_2$
- Variance: $\sigma_p^2 = \mathrm{Var}(wr_1 + (1-w)r_2)$

$$\begin{aligned}\sigma_p^2 &= w^2 var(r_1) + (1-w)^2 var(r_2) + w(1-w)cov(r_1, \\ &= w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + w(1-w)\rho(r_1, r_2)\sigma_1\sigma_2\end{aligned}$$

Problems

Problem 1: Calculating correlations

	Y		
X	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

What is $\rho_{(X,Y)}$?

Problem 2: Portfolio returns and variance

Two stocks (A and B), into which I put equal amounts of money, have the following mean and returns.

	Expected return (%)	σ (%)
A	15	15
B	12	9

The correlation between the two stocks returns is $\rho_{a,b} = 0.333$.

1. What is the expected rate of return on my portfolio?
2. What is it's standard deviation?