Joint Distributions and Independence (Ch. 5.4)

Dason Kurkiewicz

Iowa State University

June 13, 2013

Outline

Example: bearings

- Consider multiple random variables at the same time.
- Suppose you're manufacturing ring bearings (nominal inner diameter 1.00 in) on rods (nominal diameter 0.99 in). Let:
  - $X =$ the inside diameter of the next ring bearing
  - $Y =$ rod diameter where the ring is located
- We might want to know probabilities like

$$P(X < Y)$$

since if $X < Y$, the assembly cannot be made.

Example: bearings

- A **joint probability function** for discrete random variables X and Y is a nonnegative function $f(x, y)$ such that:

$$f(x, y) = P(X = x \text{ and } Y = y)$$

as a distribution, $f \geq 0$ and:

$$\sum_{x,y} f(x, y) = 1$$

- For the discrete case, it is useful to give $f(x, y)$ in a table.
- Example: suppose:
  - X = torque required to loosen bolt #3 in the next apparatus.
  - Y = torque for bolt #4.

where all torques are rounded to the nearest integer.

Example: torque (blank entries are 0)

$f(x, y)$ for the Bolt Torque Problem

$y \setminus x$	11	12	13	14	15	16	17	18	19	20
20								2/34	2/34	1/34
19							2/34			
18			1/34	1/34			1/34	1/34	1/34	
17					2/34	1/34	1/34	2/34		
16				1/34	2/34	2/34			2/34	
15	1/34	1/34			3/34					
14					1/34			2/34		
13					1/34					

- $P(X = 18 \text{ and } Y = 17) = \frac{2}{34}$
- $P(X = 14 \text{ and } Y = 19) = 0$

Your turn: torque

$f(x, y)$ for the Bolt Torque Problem

$y \setminus x$	11	12	13	14	15	16	17	18	19	20
20								2/34	2/34	1/34
19							2/34			
18			1/34	1/34			1/34	1/34	1/34	
17					2/34	1/34	1/34	2/34		
16				1/34	2/34	2/34			2/34	
15	1/34	1/34			3/34					
14					1/34			2/34		
13					1/34					

Calculate:

1. $P(X \geq Y)$
2. $P(|X - Y| \leq 1)$

Answers: torque

y \ x	11	12	13	14	15	16	17	18	19	20
20										*
19									*	*
18								*	*	*
17							*	*	*	*
16						*	*	*	*	*
15					*	*	*	*	*	*
14				*	*	*	*	*	*	*
13			*	*	*	*	*	*	*	*

Combinations of bolt 3 and bolt 4 torques with $x \geq y$

Answers: torque

$$P(X \geq Y) = \sum_{x \geq y} f(x,y)$$
$$= f(20,20) + f(20,19) + f(20,18) + \cdots + f(13,13)$$

Dropping all the $f(x,y)$ values that equal 0:

$$= f(15,13) + f(15,14) + f(15,15) + f(16,16)$$
$$+ f(17,17) + f(18,14) + f(18,17) + f(18,18)$$
$$+ f(19,16) + f(19,18) + f(20,20)$$
$$\frac{1}{34} + \frac{1}{34} + \frac{3}{34} + \frac{2}{34} + \cdots + \frac{1}{34} = \frac{17}{34}$$

Answers: torque

y \ x	11	12	13	14	15	16	17	18	19	20
20									*	*
19								*	*	*
18							*	*	*	
17						*	*	*		
16					*	*	*			
15				*	*	*				
14			*	*	*					
13		*	*	*						

Combinations of bolt 3 and bolt 4 torques with $|x - y| \leq 1$

Answers: torque

$$
\begin{aligned}
P(X \geq Y) &= \sum_{x \geq y} f(x,y) \\
&= f(13,13) + f(14,13) + f(14,14) + \cdots + f(20,20)
\end{aligned}
$$

Dropping all the $f(x,y)$ values that equal 0:

$$
\begin{aligned}
&= f(15,14) + f(15,15) + f(15,16) + f(16,16) \\
&+ f(16,17) + f(17,17) + f(17,18) + f(18,17) \\
&+ f(18,18) + f(19,18) + f(19,20) + f(20,20) \\
&= \frac{18}{34}
\end{aligned}
$$

Marginal distributions

- The **marginal distributions** of X and Y, which have joint pmf $f(x, y)$, are:

$$f_X(x) = \sum_y f(x, y)$$
$$f_Y(y) = \sum_x f(x, y)$$

- $f_X(x)$ is just the ordinary, univariate pmf of X.

Your turn: torque

- Calculate the marginal pmfs of X and Y

$f(x, y)$ for the Bolt Torque Problem

$y \setminus x$	11	12	13	14	15	16	17	18	19	20
20								2/34	2/34	1/34
19							2/34			
18			1/34	1/34			1/34	1/34	1/34	
17					2/34	1/34	1/34	2/34		
16				1/34	2/34	2/34			2/34	
15	1/34	1/34			3/34					
14					1/34			2/34		
13					1/34					

Answers: torque

- Take the column sums to calculate f_X at each x.
- Take the row sums to calculate f_Y at each y.

x	$f_X(x)$
11	1/34
12	1/34
13	1/34
14	2/34
15	9/34
16	3/34
17	4/34
18	7/34
19	5/34
20	1/34

y	$f_Y(y)$
13	5/34
14	2/34
15	5/34
16	6/34
17	7/34
18	7/34
19	3/34
20	1/34

Answers: torque

- It is customary to write the marginal pmfs in the margins of the table of the joint pmf.

Joint and Marginal Probabilities for X and Y

$y \setminus x$	11	12	13	14	15	16	17	18	19	20	$f_Y(y)$
20								2/34	2/34	1/34	5/34
19							2/34				2/34
18			1/34	1/34			1/34	1/34	1/34		5/34
17					2/34	1/34	1/34	2/34			6/34
16				1/34	2/34	2/34			2/34		7/34
15	1/34	1/34			3/34						5/34
14					1/34			2/34			3/34
13					1/34						1/34
$f_X(x)$	1/34	1/34	1/34	2/34	9/34	3/34	4/34	7/34	5/34	1/34	

Conditional distributions

- The **conditional distribution** of Y given $X = x$ is a function, $f_{Y|X=x}$, given by:

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)}$$

- To make sense of conditional distributions, return to the torque example...

Example: torque

Joint and Marginal Probabilities for X and Y

y \ x	11	12	13	14	15	16	17	18	19	20	$f_Y(y)$
20								2/34	2/34	1/34	5/34
19							2/34	0			2/34
18			1/34	1/34			1/34	1/34	1/34		5/34
17					2/34	1/34	1/34	2/34			6/34
16				1/34	2/34	2/34		0	2/34		7/34
15	1/34	1/34			3/34			0			5/34
14					1/34			2/34			3/34
13					1/34			0			1/34
$f_X(x)$	1/34	1/34	1/34	2/34	9/34	3/34	4/34	7/34	5/34	1/34	

- For example, $f_{Y|X=18}(20) = \frac{2/34}{7/34} = 2/7$. That makes sense because:
  - Since $f_X(18) = 7/34$, we expect roughly 7 out of every 34 cases to have $X = 18$.
  - Since $f_{X,Y}(18, 20) = 2/34$, we expect roughly 2 of those 7 cases to also have $Y = 20$.

Example: torque

y	13	14	15	16	17	18	19	20
$f_{X,Y}(18,y)$	$2/34$	0	$1/34$	$2/34$	0	0	$2/34$	0
$f_{Y\mid X=18}(y)$	$2/7$	0	$1/7$	$2/7$	0	0	$2/7$	0

- $\sum_{y=13}^{20} f_{X,Y}(18,y) = f_X(18) = 7/34$
- $\sum_{y=13}^{20} f_{Y|X=18}(y) = 1$
- The conditional distribution, $f_{Y|X=18}$ is the renormalized column of the joint distribution corresponding to $X = 18$.

Your turn: torque

Joint and Marginal Probabilities for X and Y

$y \setminus x$	11	12	13	14	15	16	17	18	19	20	$f_Y(y)$
20								2/34	2/34	1/34	5/34
19							2/34				2/34
18			1/34	1/34			1/34	1/34	1/34		5/34
17					2/34	1/34	1/34	2/34			6/34
16				1/34	2/34	2/34			2/34		7/34
15	1/34	1/34			3/34						5/34
14					1/34			2/34			3/34
13					1/34						1/34
$f_X(x)$	1/34	1/34	1/34	2/34	9/34	3/34	4/34	7/34	5/34	1/34	

- Calculate:
  1. $f_{Y|X=15}(y)$
  2. $f_{Y|X=20}(y)$
  3. $f_{X|Y=18}(x)$

Answers: torque

1.

y	13	14	15	16	17	18	19	20
$f_{Y\|X=15}(y)$	1/9	1/9	3/9	2/9	2/9	0	0	0

2.

y	13	14	15	16	17	18	19	20
$f_{Y\|X=20}(y)$	0	0	0	0	0	0	0	1

3.

x	11	12	13	14	15	16	17	18	19	20
$f_{X\|Y=18}(x)$	0	0	1/5	1/5	0	0	1/5	1/5	1/5	0

Given a set of marginal distributions, there are many possible joint distributions.

- What do you notice about each of the following joint distributions?

Distribution 1

$y \backslash x$	1	2	3	
3	.4	0	0	.4
2	0	.4	0	.4
1	0	0	.2	.2
	.4	.4	.2	

Distribution 2

$y \backslash x$	1	2	3	
3	.16	.16	.08	.4
2	.16	.16	.08	.4
1	.08	.08	.04	.2
	.4	.4	.2	

Given a set of marginal distributions, there are many possible joint distributions.

- What do you notice about each of the following joint distributions?

Distribution 1

y \ x	1	2	3	
3	.4	0	0	.4
2	0	.4	0	.4
1	0	0	.2	.2
	.4	.4	.2	

Distribution 2

y \ x	1	2	3	
3	.16	.16	.08	.4
2	.16	.16	.08	.4
1	.08	.08	.04	.2
	.4	.4	.2	

1. Given $X = x$, you know what Y has to be (and vice versa).
2. Each $P(X = x, Y = y)$ is just $P(X = x) \cdot P(Y = y)$; i.e., X and Y have no influence on each other.

A look at distribution 2

y \ x	1	2	3	
3	.16	.16	.08	.4
2	.16	.16	.08	.4
1	.08	.08	.04	.2
	.4	.4	.2	

- Among just the cases when $X = 1$:
  - $Y = 3$ every 16 out of $(16 + 16 + 8) = 40$ times: i.e., with probability $\frac{16}{40} = 0.4$
  - Same with $Y = 2$
  - $Y = 1$ every 8 out of $(16 + 16 + 8) = 40$ times: i.e., with probability 0.2
- So pmf of Y given $X = 1$ is the same as the marginal pmf of Y.

Independence

- Discrete random variables X and Y are independent (written $X \perp\!\!\!\perp Y$) if for all x and y,

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

  where $\mid$ means "given".

- If $X \perp\!\!\!\perp Y$, then:

$$P(Y = y \text{ and } X = x) = P(X = x) \cdot P(Y = y)$$
$$f(x, y) = f_X(x) \cdot f_Y(y)$$

- If X and Y are not only independent but also have the same marginal distribution, then they are **independent and identically distributed**, abbreviated **iid**.

Outline

Continuous joint distributions

- A **joint probability density function** (pdf) for two continuous random variables X and Y is a nonnegative function with:

$$\int \int f(x,y)dxdy = 1$$

$$P((X,Y) \in R) = \int \int_R f(x,y)dxdy$$

where R is some region of $\mathbb{R}^2$.

Example: sales desk

- $S =$ true excess time (over a 7.5 s threshold) required to complete the next sale
- $R =$ excess time measured with a stopwatch

$$f(s,r) = \begin{cases} \frac{1}{16.5} e^{-s/16.5} \frac{1}{\sqrt{2\pi(0.25)}} e^{-(r-s)^2/2(0.25)} & s > 0 \\ 0 & \text{otherwise} \end{cases}$$

$f(s, r)$ is valid.

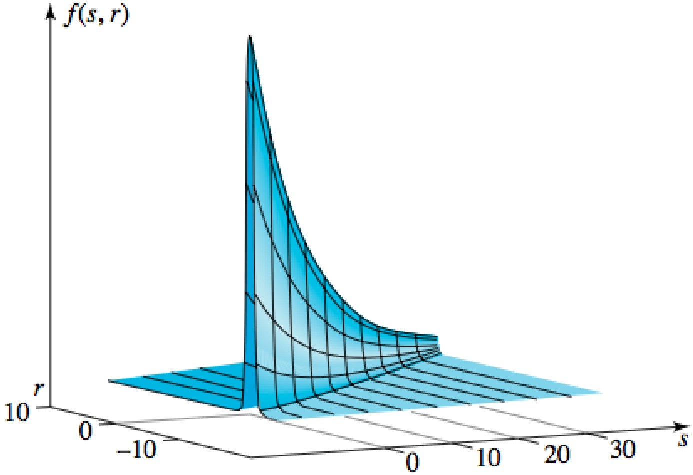
$$\int\int f(s,r)ds\ dr = \int_0^\infty \int_{-\infty}^\infty \frac{1}{16.5\sqrt{2\pi(0.25)}} e^{-(s/16.5)-((r-s)^2/2(0.25))} dr\ ds$$

$$= \int_0^\infty \frac{1}{1.65} e^{-s/16.5} \left\{ \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi(0.25)}} e^{-(r-s)^2/2(0.25)} dr \right\} ds$$

$$= \int_0^\infty \frac{1}{16.5} e^{-s/16.5} ds$$

$$= 1$$

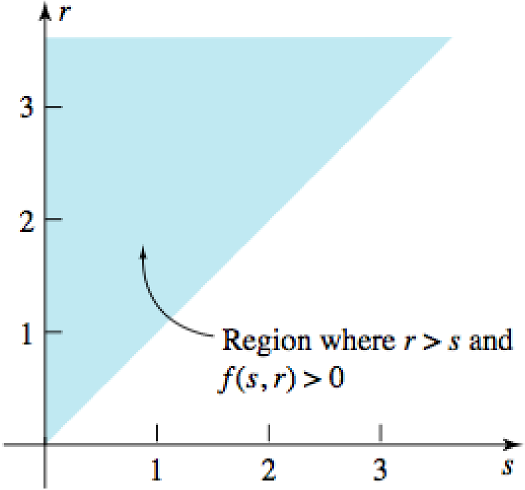
A look at $f(s, r)$



Checking for measurement bias: P(measured excess time > actual excess time)

$$
\begin{aligned}
P(R > S) &= \int\int_{r>s} f(s,r) ds\ dr \\
&= \int_0^\infty \int_s^\infty f(s,r) dr\ ds \\
&= \int_0^\infty \frac{1}{16.5} e^{-s/16.5} \left\{ \int_s^\infty \frac{1}{\sqrt{2\pi(0.25}} e^{-(r-s)^2/2(0.25)} dr \right\} ds \\
&= \int_0^\infty \frac{1}{16.5} e^{-s/16.5} \left\{ \frac{1}{2} \right\} ds \\
&= \frac{1}{2}
\end{aligned}
$$

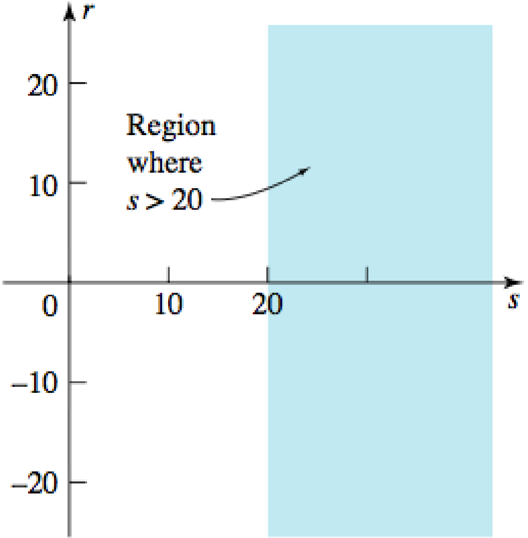
Checking for measurement bias: region of integration



Probability of taking too long

$$
\begin{aligned}
P(S > 20) &= \int\int_{s>20} f(s,r)dr\ ds \\
&= \int_{20}^{\infty}\int_{-\infty}^{\infty} f(s,r)dr\ ds \\
&= \int_{20}^{\infty} \frac{1}{16.5} e^{-s/16.5} \left\{ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(0.25)}} e^{-(r-s)^2/s(0.25)} \right\} ds \\
&= \int_{20}^{\infty} e^{-s/16.5} ds \\
&= e^{-20/16.5} \\
&\approx 0.30
\end{aligned}
$$

Probability of taking too long: region of integration



Continuous marginal and conditional distributions

- For continuous random variables X and Y, the **marginal distribution** of X is:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

- The **conditional distribution** of Y given $X = x$ is:

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}$$