Chs 3 & 4: Univariate pr. dsn (1 at a time)

(2 or more at a time)

MULTIVARIATE PROBABILITY DISTRIBUTIONS (Chapter 5)

We'll first look at the *discrete* case; the *continuous* case will be discussed later. (3) mixed case

The discrete case

Example 1 A die is rolled. Let X = no. of 6's and Y = no. of even numbers. Find the joint probability distribution of X and Y.

Number on die	1	2	3	4	5	6
Value of X	0	0	0	0	0	1
Value of Y	0	1	0	1	0	1

same

$P(X = 1, Y = 1) = P(6) = 1/6$
$P(X = 0, Y = 1) = P(2 \text{ or } 4) = 2/6 = 1/3$
$P(X = 0, Y = 0) = P(1 \text{ or } 3 \text{ or } 5) = 3/6 = 1/2.$

We say that X and Y have a *joint probability distribution*.

The *joint pdf* (pmf) of X and Y is

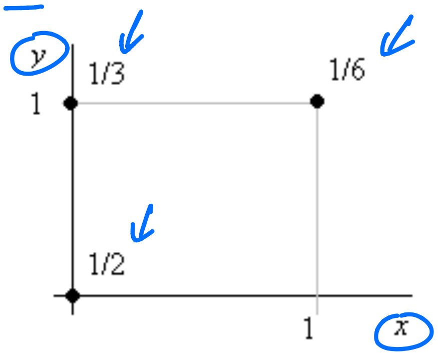
$$p(x,y) = P(X = x, Y = y) = \begin{cases} 1/2, & x = y = 0 \\ 1/3, & x = 0, y = 1 \\ 1/6, & x = y = 1 \\ 0, & \text{otherwise} \end{cases}$$

The joint probability distribution of X and Y can also be presented in other ways.

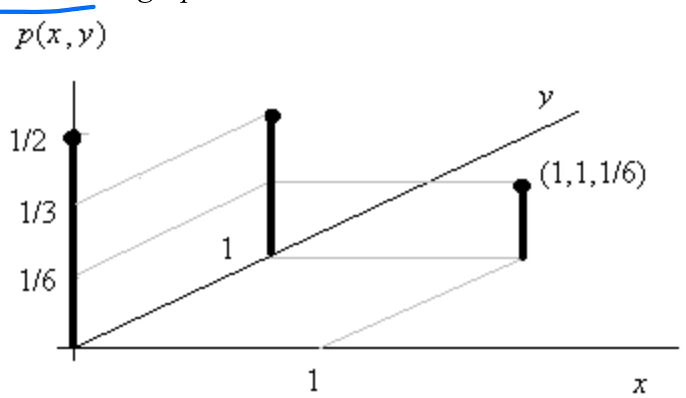
Table of $p(x,y)$:

		y = 0	y = 1
x	0	1/2	1/3
	1		1/6

Graph (two-dimensional top view):



Three-dimensional graph:



Two properties of discrete joint pdfs:

1. $0 \le p(x,y) \le 1$ for all x and y

2. $\sum_{x,y} p(x,y) = 1$. (Or, equivalently, $\sum_x \sum_y p(x,y) = 1$.)

In Example 1: 1/2, 1/3 and 1/6 are all in the interval $[0,1]$; and 1/2 + 1/3 + 1/6 = 1.

The *joint cdf* of X and Y is

$$F(x,y) = P(X \le x, Y \le y).$$

In Example 1 observe that:

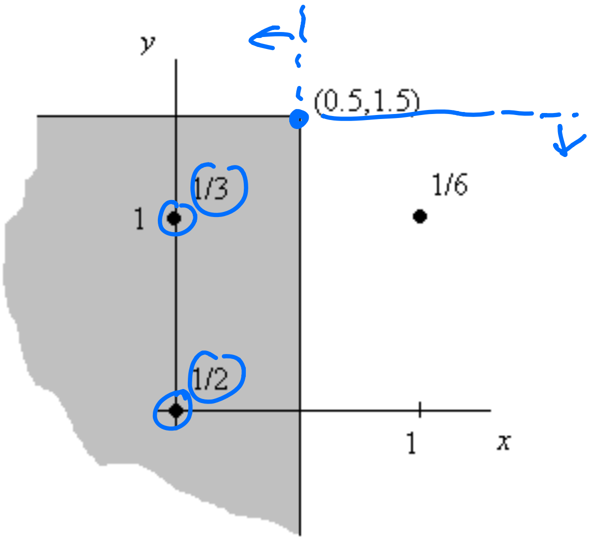
$F(0,0) = P(X \le 0, Y \le 0) = p(0,0) = 1/2$

$F(0,0.5) = P(X \le 0, Y \le 0.5) = p(0,0) = 1/2$

$F(0,1) = P(X \le 0, Y \le 1) = p(0,0) + p(0,1) = 1/2 + 1/3 = 5/6$

$F(0.5,1.5) = P(X \le 0.5, Y \le 1.5) = p(0,0) + p(0,1) = 1/2 + 1/3 = 5/6$ etc.

The following figure illustrates the working for $F(0.5,1.5)$. The region to the left of and below (0.5,1.5) is shaded, and we see that this region contains (0,0) and (0,1). So we sum the joint pdf $f(x,y)$ over those points to get the joint cdf $F(0.5,1.5)$. The region includes its boundary lines. If there were any points (x,y) with positive $f(x,y)$ on those boundaries, the values of $f(x,y)$ at those points would also contribute to $F(0.5,1.5)$.

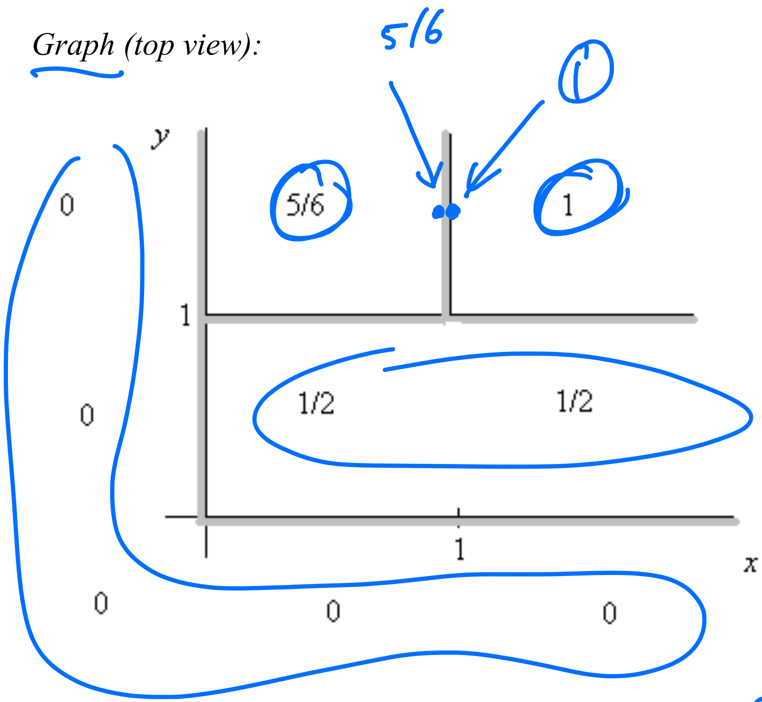


(eg $<$ $\not\equiv$ $\not<$)

We find that X and Y have joint cdf
$$F(x,y)=\begin{cases}0, & x<0 \text{ or } y<0 \text{ (or both)}\\ 1/2, & x\ge 0, 0\le y<1\\ 5/6, & 0\le x<1, y\ge 1\\ 1, & x,y\ge 1\end{cases}$$

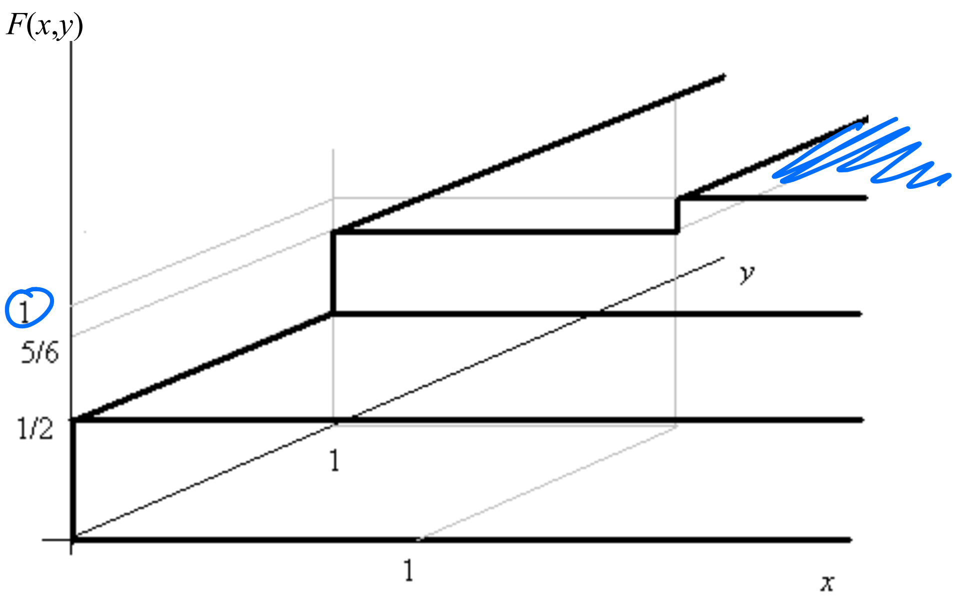
(?)

Graph (top view):



The shading here indicates that, for example, $F(1,1.5) = 1$, $F(0.999,1.5) = 5/6$.

3-d graph (non-assessable):



> Some properties of all joint cdf's:
>
> 1. $F(x,y) \to 0$ as $x \to -\infty$ or $y \to -\infty$ (or both).
> 2. $F(x,y) \to 1$ as $x \to \infty$ and $y \to \infty$.
> 3. $F(x,y)$ is nondecreasing in both x and y directions.
> 4. $F(x,y)$ is right-continuous in both x and y directions.

Note that with these properties in mind, the joint cdf of X and Y in our example could be written more simply as

$$F(x,y) = \begin{cases} 1/2, & x > 0, 0 < y < 1 \\ 5/6, & 0 < x < 1, y > 1 \end{cases}$$

But, for clarity, it is best to write *joint* cdf's in full detail.

Now for some more definitions.

ch2

by LTP

> The *marginal pdf* of X is
>
> $$P(X=x) = p(x) = \sum_y p(x,y) = \sum_y P(X=x, Y=y)$$
>
> This pdf defines the *marginal probability distribution* of X.
>
> We may also write $p(x)$ as $p_X(x)$.

In our example:

$$f_X(0) = \sum_y p(0,y) = p(0,0) + p(0,1) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

$$f_X(1) = \sum_y p(1,y) = p(1,1) = \frac{1}{6}.$$

Thus $\boxed{f(x) = \begin{cases} 5/6, & x = 0 \\ 1/6, & x = 1 \end{cases}}$

In words we may say that X's marginal probability distribution is Bernoulli with parameter 1/6. That is, $X \sim \text{Bern}(1/6)$.

(This makes sense: X is the number of 6's on one roll of a die.)

Similarly, we find that $Y \sim \text{Bern}(1/2)$.

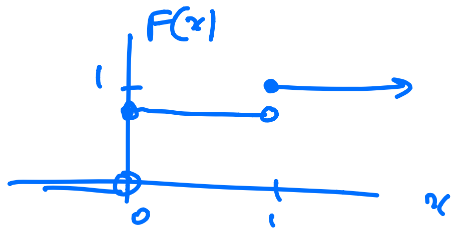
Note that what we have done is equivalent to computing column and row totals:

$f(x,y)$:

		y: 0	y: 1	$f(x)$
x	0	1/2	1/3	5/6
	1	0	1/6	1/6
$f(y)$	→	1/2	1/2	

> The *marginal cdf* of X is
> $$F(x) = P(X \le x) = \sum_{t \le x} f_X(t) \qquad \text{(Ch3)}$$
> This is just the ordinary cdf of X, and can be computed in the usual way.

For example, $F(x) = \begin{cases} 0, & x < 0 \\ 5/6, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$



> The "*conditional pdf* of X given that $Y = y$" is
> $$f(x \mid y) = \frac{f(x,y)}{f(y)}. \qquad \left(\text{Or equivalently, } p(x \mid y) = \frac{P(X = x, Y = y)}{P(Y = y)}.\right)$$
> $P(X=x \mid Y=y)$
>
> by Ch2
>
> This density defines the *conditional probability distribution* of X given that $Y = y$.

In our example, what's the conditional probability distribution of X given that $Y = 1$?

$\boxed{p(x\,|\,1) = \dfrac{p(x,1)}{p_Y(1)}}$ for $x = 0,1$.

Explicitly: $p_{X|Y}(0\,|\,1) = \dfrac{p_{X,Y}(0,1)}{p_Y(1)} = \dfrac{1/3}{1/2} = \dfrac{2}{3}$

$p_{X|Y}(1\,|\,1) = \dfrac{p_{X,Y}(1,1)}{p_Y(1)} = \dfrac{1/6}{1/2} = \dfrac{1}{3}$.

So $p(x\,|\,1) = \begin{cases} 2/3, & x = 0 \\ 1/3, & x = 1 \end{cases}$

Thus $(X|Y = 1) \sim$ Bern(1/3). ✓

2, 4 or 6

(This makes sense: If an even number comes up (2, 4 or 6) then there is obviously a one-in-three chance of that number being 6. So $P(X = 1|\ Y = 1) = 1/3$, etc.)

What is the dsn of X given that $Y = 0$?

$Y = 0$ implies that a 1, 3 or 5 comes up, meaning that a 6 definitely does *not* come up.
So $P(X = 1|Y = 0) = 0$ and $P(X = 0|Y = 0) = 1$.
(If $Y = 0$ then $X = 0$ with probability one.) w.p. 1

Thus $(X|Y = 0) \sim Bern(0)$, and $p(x\,|\,0) = I(x = 0) = \begin{cases} 1, & x = 0 \\ 0, & x = 1 \end{cases}$

This is an example of a *degenerate* dsn (a discrete dsn with only one possible value).

The *conditional cdf* of $(X|Y = y)$ is
$F(x\,|\,y) = P(X \le x\,|\,Y = y)$. ✓ not $P(X \le x|Y \le y)$ (x)

This function can be computed in the same way as the marginal cdf of X, but using $p(x|y)$ instead of $p(x)$.

For example, $F(x\,|\,1) = \begin{cases} 0, & x < 0 \\ 2/3, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$

Note: For clarity, we could also write this as $\boxed{F_{X|Y}(x\,|\,1)}$. $= P(X \le x|Y=1)$

$F(1|1) = ?$ confusing
$= F_{X|Y}(1,1)$ or $F_{Y|X}(1,1)$

Independence of random variables

Recall that two events A and B are independent if $P(AB) = P(A)P(B)$. Similarly...

Two random variables X and Y are *independent* if

$p(x,y) = p(x)p(y)$ for all x and y. (*)

f(x,y) = f(x) f(y)

We then write $X \perp Y$.

If (*) is false for some x and y, then X and Y are *dependent*, and we write $X \not\perp Y$.

In our example, are X and Y independent?

Recall that $p_{X,Y}(1,1) = 1/6$, $p_X(1) = 1/6$, $p_Y(1) = 1/2$.

Thus $p_{X,Y}(1,1) \neq p_X(1)p_Y(1)$.

Therefore X and Y are not independent.

NB: If $p(x|y) = p(x)$ or $p(y|x) = p(y)$ then $X \perp Y$.

If $p(x|y) \neq p(x)$ or $p(y|x) \neq p(y)$ then $X \not\perp Y$.

In our example, $p_{X|Y}(1|1) = 1/3$ and $p_X(1) = 1/6$.

These are not the same, and therefore $X \not\perp Y$.

Ch 3: $Eg(X) = \sum_x g(x) f(x)$

Multivariate expectation

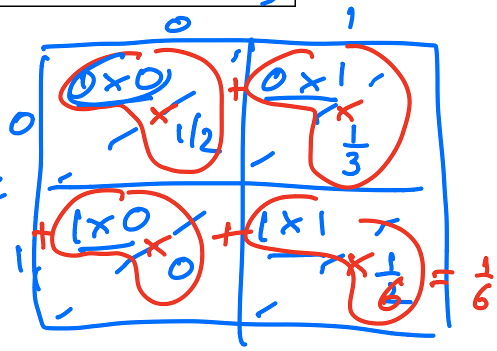
$$Eg(X,Y) = \sum_{x,y} g(x,y)p(x,y) \quad = \sum_x \sum_y g(x,y) f(x,y)$$

In our example, what is the expected value of XY?

$1 \times 0 \; f(1,0)$

$$E(XY) = \sum_{x,y} xyp(x,y) = 0(0)p(0,0) + 0(1)p(0,1) + 1(1)p(1,1)$$

$$= 0 + 0 + p(1,1) = 1/6.$$



Also, what is the expected value of $(X+1)^Y$?

$$E\{(X+1)^Y\} = \sum_{x,y}(x+1)^y p(x,y) = (0+1)^0 p(0,0) + (0+1)^1 p(0,1) + (1+1)^1 p(1,1)$$

$$= (0+1)^0(1/2) + (0+1)^1(1/3) + (1+1)^1(1/6) = 7/6.$$

(For each example here, it may help to draw a matrix showing a cell for each value of the pair (x,y). In each cell write the value of the pdf and the value of the function.)

Covariance and correlation

The *covariance* between X and Y is

$$Cov(X,Y) = E\{(X - EX)(Y - EY)\}.$$

What's the covariance between X and Y in our example?

Recall that $X \sim$ Bern(1/6) and $Y \sim$ Bern(1/2).
Therefore EX = 1/6 and EY = 1/2.

It follows that

$$(x-EX)\;(y-EY)$$

$$Cov(X,Y) = \sum_{x,y}\left(x-\frac{1}{6}\right)\left(y-\frac{1}{2}\right)p(x,y)$$

$$= \left(0-\frac{1}{6}\right)\left(0-\frac{1}{2}\right)p(0,0) + \left(0-\frac{1}{6}\right)\left(1-\frac{1}{2}\right)p(0,1) + \left(1-\frac{1}{6}\right)\left(1-\frac{1}{2}\right)p(1,1)$$

$$= \left(\frac{1}{12}\right)\frac{1}{2} + \left(-\frac{1}{12}\right)\frac{1}{3} + \left(\frac{5}{12}\right)\frac{1}{6} = \frac{1}{12}.$$

A useful result: $Cov(X,Y) = E(XY) - (EX)EY$.

Proof: LHS = $E\{(X-\mu_X)(Y-\mu_Y)\} = E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)$

$$= E(XY) - \mu_Y EX - \mu_X EY + \mu_X\mu_Y$$

$$= E(XY) - \mu_Y\mu_X - \cancel{\mu_X\mu_Y} + \cancel{\mu_X\mu_Y} = \text{RHS}.$$ QED

Let's illustrate this result by using it to check $Cov(X,Y)$ in our example.

Recall that $E(XY)$ = 1/6.

It follows that $Cov(X,Y) = \frac{1}{6} - \frac{1}{6}\left(\frac{1}{2}\right) = \frac{1}{12}$, as before.

The *correlation* between X and Y is

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}.$$

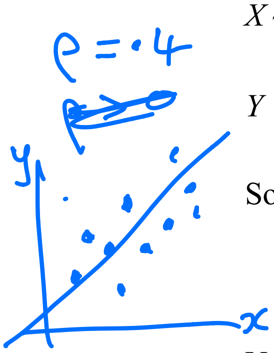
rho

What's the correlation between X and Y in our example?

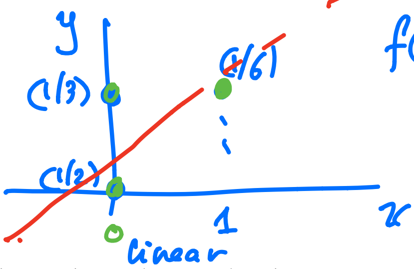
$$X \sim \text{Bern}(1/6) \Rightarrow VarX = \frac{1}{6}\left(1-\frac{1}{6}\right) = \frac{5}{36} \Rightarrow SD(X) = \frac{\sqrt{5}}{6}.$$

$$Y \sim \text{Bern}(1/2) \Rightarrow VarY = \frac{1}{2}\left(1-\frac{1}{2}\right) = \frac{1}{4} \Rightarrow SD(Y) = \frac{1}{2}.$$

So $\rho = \dfrac{1/12}{\dfrac{\sqrt{5}}{6}\times\dfrac{1}{2}} = 0.4472.$

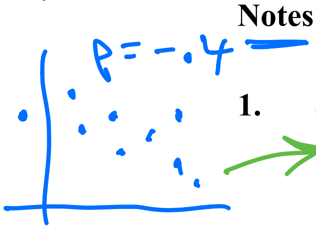


positive linear relationship

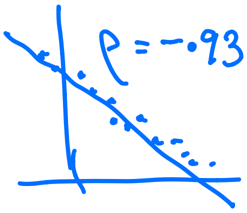


f(x,y):

Notes

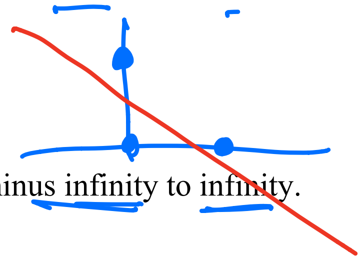


1. ρ provides information about the *relationship* between X and Y.
   If $\rho > 0$ then *high* values of X are associated with *high* values of Y (eg $\rho = 0.4472$ above).
   If $\rho < 0$ then *high* values of X are associated with *low* values of Y.

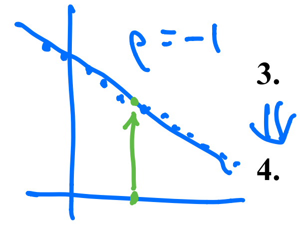


2. $-1 \le \rho \le 1.$

Corr(X,Y)
ρ =



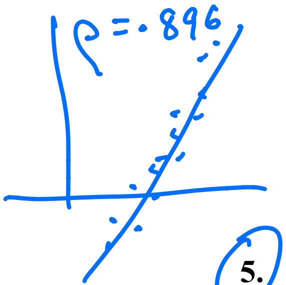
(By contrast, $Cov(X,Y)$ can be anything from minus infinity to infinity. So ρ is easier to interpret.)



3. $X \perp Y \Rightarrow \rho = 0$. (Prove this as an exercise.)

4. $\rho \ne 0 \Rightarrow X \not\perp Y$.

(In logical parlance, this follows from Note 3 by the *principle of contraposition*. The *contrapositive* of P => Q is notQ => notP. For example, since it is true that all dogs are animals, it follows by contraposition that if something is not an animal, it is also not a dog.)



5. $\rho = 0 \not\Rightarrow X \perp Y$.

(A proof of this fact is provided by Example 5.24 in the text.)