Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

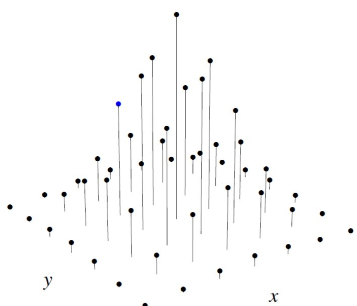
Previous Lecture

- Moment Generating Functions (MGFs)
- Moments via Derivatives of MGFs
- MGF of Location-Scale Transformation

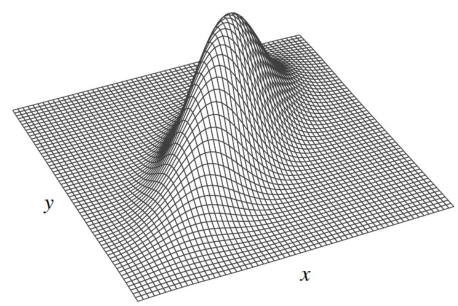
7- Multivariable Distributions

Recall: The distr of X provides complete info about the prob of X falling into any subset of $\mathbb{R}$.

Analogously, the **joint distr** of X and Y gives complete info about the prob of the vector (X, Y) falling in any subset of $\mathbb{R}^2$.



Discrete



Continuous

We will also discuss:

The **marginal distr** of X is the distr of X, ignoring the value of Y.

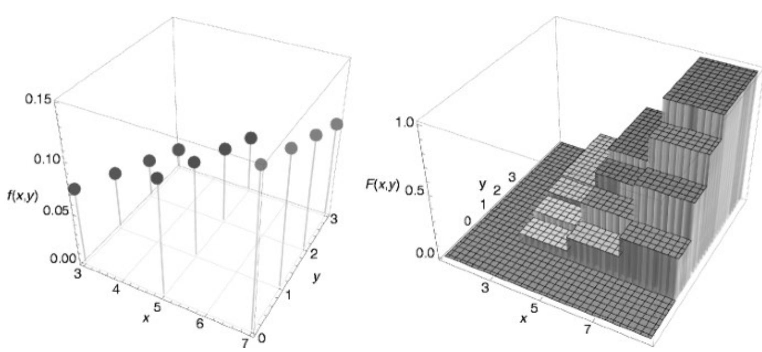
The **conditional distr** of X given $Y = y$ is the updated distr for X after observing $Y = y$.

§7.1 - Joint, Marginal, and Conditional Distr

Joint Distr

For the Joint CDF, the discrete and cont versions are the same:

Def (**Discrete & Cont Joint CDF**): The joint CDF of X and Y is $F_{X,Y}(x,y) := P(X \leq x, Y \leq y)$.



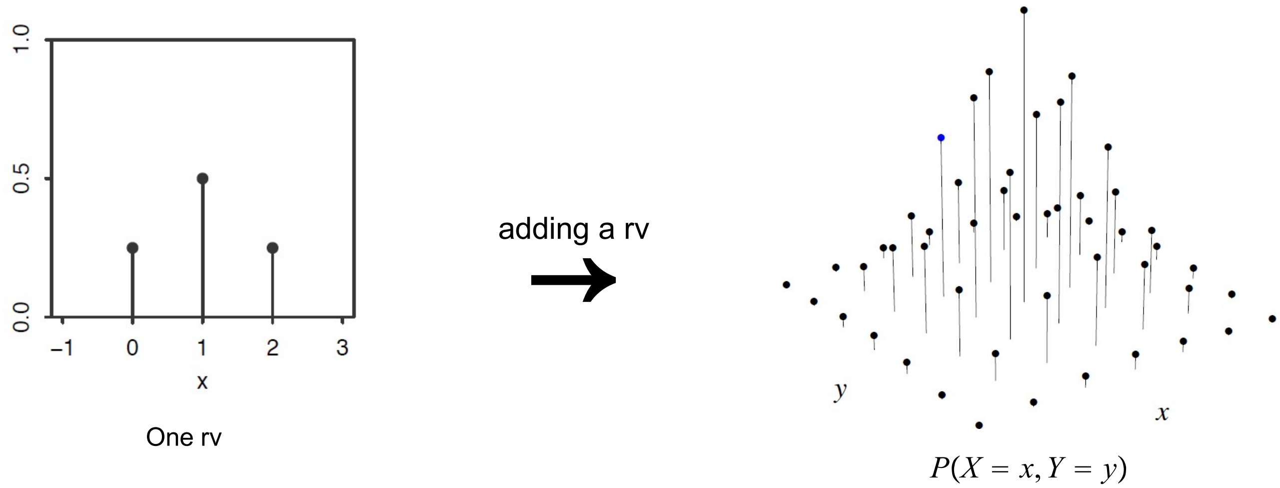
Discrete PF and CDF

But things diverge from here, so let's first focus on discrete rvs to start with.

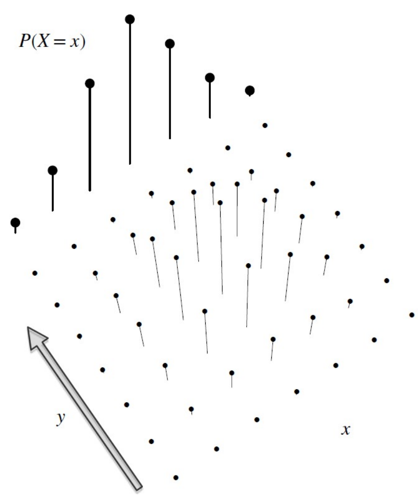
❗ In particular, the above discrete CDF is difficult to visualize, so instead, we focus on the joint PF.

Def (**Discrete Joint PF**): The joint PF of discrete X and Y is $p_{X,Y}(x,y) := P(X = x, Y = y)$.

Still must be positive and sum to one ($\Sigma_x\Sigma_y P(X = x, Y = y) = 1$).



Def (**Discrete Marginal PF**): For discrete X and Y, the marginal (or unconditional) PF of X is $P(X = x) = \sum_y P(X = x, Y = y)$. ("marginalizing out Y")



sum up all the probs along y for any

Ex (**Joint Discrete PF**): Let X and Y have PF: $f(x,y) = \begin{cases} \frac{1}{6}|x+y| & \text{for } x = -1, 0 \text{ and } y = -2, 0 \\ 0 & \text{otherwise.} \end{cases}$

a. Find $P(X = -1, Y = -2)$.

$$P(X = -1, Y = -2) = f(-1,-2) = \tfrac{1}{6}|-1-2| = \tfrac{1}{2}.$$

b. Find $P(X \geq -1, Y = 0)$.

$$P(X \geq -1, Y = 0) = f(-1,0) + f(0,0) = \tfrac{1}{6}|-1+0| + \tfrac{1}{6}|0+0| = \tfrac{1}{6}.$$

c. Find the marginal PF of X.

$$\sum_y \frac{1}{6}|x+y| = \frac{1}{6}(|x-2| + |x+0|) = \frac{|x-2|+|x|}{6}$$

$$= \frac{|x-2|-x}{6} = \frac{(-x+2)-x}{6} = \frac{1-x}{3}.$$

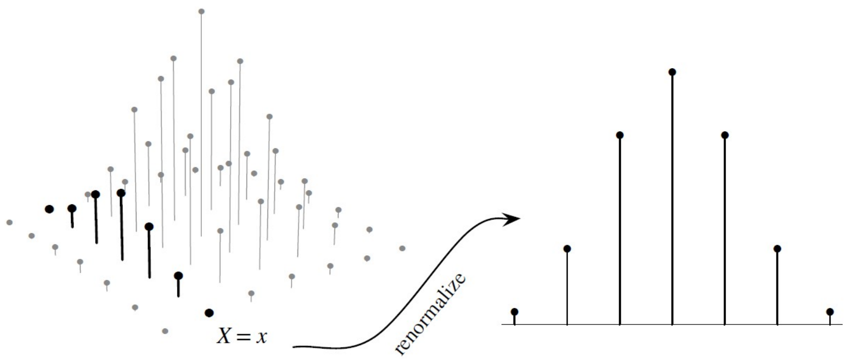
$$f_X(x) = \begin{cases} \frac{1-x}{3} & \text{for } x = -1, 0 \\ 0 & \text{otherwise.} \end{cases}$$

Conditional PF

Suppose we observe the value of X and want to update our distr of Y to reflect this knowledge?

Def (**Discrete Conditional PF**). For discrete X and Y, the conditional PF of Y given $X = x$ is $P(Y = y|X = x) = \frac{P(Y=y,\, X=x)}{P(X=x)}$.

This is viewed as a function of y for fixed x.



Dividing by $P(X = x)$ ensures it still sums to 1 when summing over $Y = y$.

Discrete Bayes' Rule: $P(Y = y|X = x) = \frac{P(X=x|\,Y=y)P(Y=y)}{P(X=x)}$. (same as when they weren't in a joint distr)

Def (**Independence of Rvs**): X and Y are indep if for all x and y, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. (we saw this in §3.8, but this is new notation)

If X and Y are discrete, this is equivalent to the condition $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x, y.

It's also equivalent to the condition $P(Y = y|X = x) = P(Y = y)$ for all x, y as long as $P(X = x) > 0$.

❗ In general, marginal distr's ($P(X = x), P(X = x)$) don't determine the joint distr $P(X = x, Y = y)$.

This is why we need to study joint distrs in the first place!
But in the special case of indep, the marginal distr's are all we need to specify the joint distr.
We can get the joint PF by multiplying the marginal PFs!

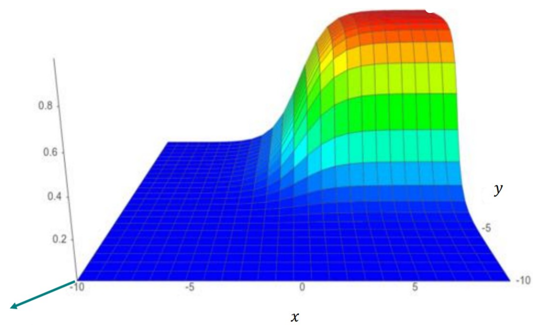
Activity 14: Problem 1

Continuous Joint, Marginal, and Conditional

For continous rvs, replace Σ w/ $\int$.

Let cont X and Y have joint PF: $f_{X,Y}(x,y)$.

Def (**Cont Joint CDF**): $F_{X,Y}(x,y) := P(X \le x, Y \le y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(s,t)dsdt.$



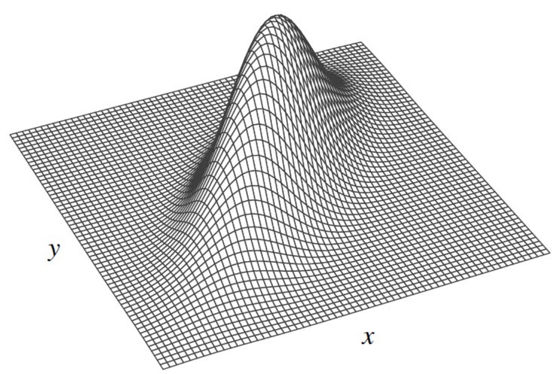
Cont Joint CDF $F_{X,Y}(x,y)$

Def (**Cont Joint PF**): If cont X and Y have joint CDF $F_{X,Y}$, their joint PF is the derivative of $F_{X,Y}$ w/respect to x and y:
$f_{X,Y}(x,y) := \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$

We require valid joint PFs to be nonnegative and integrate to 1:
$f_{X,Y}(x,y) \ge 0$; and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1.$

To get the prob of a 2D region, integrate the PF over that region.

For example, $P(X < 3,\ 1 < Y < 4) = \int_{1}^{4}\int_{-\infty}^{3} f_{X,Y}(x,y)dxdy.$



$f_{X,Y}(x,y)$

Def (**Marginal PF**): For cont X and Y w/joint PF $f_{X,Y}$, the marginal PF of X is $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$ ("marginalizing out Y ")

❗ Another way to find the marginal distr is by way of the joint CDF. Here we calculate the marginal CDF:

$$F_X(x) = P(X \le x) = \lim_{y\to\infty} P\big(X \le x,\ Y \le y\big) = \lim_{y\to\infty} F_{X,Y}(x,y).$$

Ex (**Cont PF**): Let X and Y be cont w/joint PF: $f(x,y) = \begin{cases} 2(1-x) \text{ for } 0 \le x \le 1 \text{ and } 0 \le y \le x, \\ 0 \text{ otherwise.} \end{cases}$

$a.$ Find $P\big(X < \frac{1}{2},\ Y \ge \frac{x}{2}\big)$.

Solution: $P\left(X < \frac{1}{2},\ Y \geq \frac{x}{2}\right) = \int_0^{\frac{1}{2}} \int_{\frac{x}{2}}^{x} 2(1-x)dydx = \int_0^{\frac{1}{2}} \left[2(1-x)y\Big|_{y=\frac{x}{2}}^{y=x}\right]dx$

$$= \int_0^{\frac{1}{2}} \left[2(1-x)x - 2(1-x)\tfrac{x}{2}\right]dx = \int_0^{\frac{1}{2}} x - x^2 dx$$

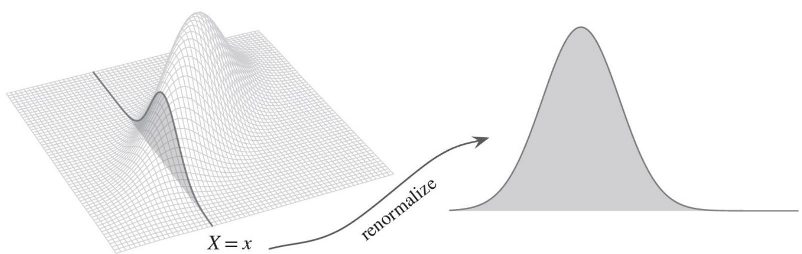
$$= \left[\tfrac{1}{2}x^2 - \tfrac{1}{3}x^3\right]_0^{\frac{1}{2}} = \left(\tfrac{1}{2}\tfrac{1}{4} - \tfrac{1}{3}\tfrac{1}{8}\right) - 0 = \tfrac{1}{12}.$$

b. Find the marginal PF of X.

Solution: $2(1-x)\int_0^x dy = 2(1-x)[y]_0^1 = 2(1-x)(1-0) = 2(1-x)$.

$$f_X(x) = \begin{cases} 2(1-x) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Def (**Cont Conditional PF**). For cont X and Y w/joint PF $f_{X,Y}$, the conditional PF of Y given $X = x$ is
$f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)}$, for all x w/$f_X(x) > 0$. This is considered as a function of y for fixed x.
As a convention, in order to make $f_{Y|X}(y|x)$ well-defined for all x, we let $f_{Y|X}(y|x) = 0$ for all x w/ $f_X(x) = 0$.



Activity 14: Problems 2,3

Thm (**Cont Form of Bayes' Rule and LOTP**). For cont X and Y we have:
Bayes' rule: $f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$, for $f_X(x) > 0$.

LOTP: $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$.

These follow quite trivially from the definition of continuous conditional PF. See the book.

Sometimes one rv is cont and the other is discrete!

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y\|X=x) = \frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x) = \frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X **continuous**	$P(Y=y\|X=x) = \frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x) = \frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

Versions of Bayes' Rule

	Y **discrete**	Y **continuous**
X **discrete**	$\sum_y P(X=x\|Y=y)P(Y=y)$	$\int_{-\infty}^{\infty} P(X=x\|Y=y)f_Y(y)dy$
X **continuous**	$\sum_y f_X(x\|Y=y)P(Y=y)$	$\int_{-\infty}^{\infty} f_{X\|Y}(x\|y)f_Y(y)dy$

Versions of LOTP

Def (**Indep of Cont Rvs**): X and Y are indep if for all x and y, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. (we saw this in §3.8, but this is new notation)

If X and Y are cont w/joint PF $f_{X,Y}$, this is equivalent to the condition $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x,y, and it is also equivalent to the condition $f_{Y|X}(y|x) = f_Y(y)$ for all x,y such that $f_X(x) > 0$.

This relies on the fact that you know the marginal distributions ($f_X(x)$, $f_Y(y)$). If you don't, see the following:

Proposition (**Funtional Rv Indep**): Suppose that joint PF $f_{X,Y}$ factors as $f_{X,Y}(x,y) = g(x)h(y)$ for all x and y, where g and h are nonnegative functions. Then X and Y are indep.

If either g or h is a valid PF, then the other one is a valid PF too and g and h are the marginal PFs of X and Y, respectively. (The analogous result in the discrete case also holds.)

Activity 14: Problem 4

Activity 15

Harvard Video: youtube.com/watch?v=J70dP_AECzQ&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=20

---

§7.2 - 2D LOTUS

Like the single rv version, the 2D LOTUS saves us from having to find the distr of $g(X,Y)$ in order to calculate its expectation. Instead, just having the joint PF of X and Y is enough.

Thm (**2D LOTUS**): Let g be a function from $\mathbb{R}^2$ to $\mathbb{R}$. If X and Y are *discrete*, then $E(g(X,Y)) = \sum_x \sum_y g(x,y)P(X=x, Y=y)$.

If X and Y are *cont* w/joint PF $f_{X,Y}$, then $E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy$.

Ex (**Cont 2D LOTUS**): Let X and Y be cont w/joint PF: $f(x,y) = \begin{cases} \frac{15}{8}x^2y \text{ for } 0 \le y \le 1-x \le 2, \\ 0 \quad \text{otherwise.} \end{cases}$

Set up (but don't evaluate) the expression necessary to calculate $E(XY)$.

$$E(XY) = \int_{-1}^{1}\int_{0}^{1-x}(xy)\left(\tfrac{15}{8}x^2y\right)dydx$$

OR

$$E(XY) = \int_0^2 \int_{-1}^{y} (xy)\left(\frac{15}{8}x^2 y\right) dx dy$$

Harvard Video: youtube.com/watch?v=9vp1Ll2NpRw&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=15

What did we learn?

- Joint/Marginal/Conditional Distr
- Bayes' Rule and LOTP for two rvs
- Independence of Rvs
- 2D LOTUS