6.1 Joint, marginal, and conditional distributions

So far we have looked at the distribution of one random variable at a time, but often we care about the relationship between multiple r.v.s in the same experiment. Some examples are:

- Surveys: ask multiple questions to each respondent
- Medicine: take multiple measurements per patient
- Genetic: study multiple genetic markers on a particular disease
- Time series: to study how something evolves over time, we can make a series of measurements over time, and then study the series jointly.

The 3 key concepts to grasp are joint, marginal, conditional distribution. This story will explain them:

Recall that the distribution of r.v X provides information about probability of X falling into any subset of $\mathbb{R}$.

- The joint distribution of r.v.s. X and Y provides information about probability of vector (X, Y) falling into any subset of plane $\mathbb{R}^2$.
- The marginal distribution of X is the individual distribution of X ignoring the value of Y.
- The conditional distribution of X given $Y = y$ is the updated distribution of X after observing $Y = y$.

Discrete

The most general description of the joint distribution of 2 r.v.s is joint CDF.

Definition 6.1.1 (Joint CDF)

The joint CDF of r.v.s X and Y is function $F_{X,Y}$ given by:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

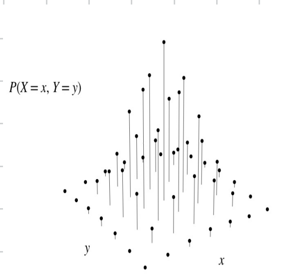
The joint CDF is not a very well-behaved function, it consists of jumps and flats. Therefore, we usually work with joint PMF for discrete r.v.s.

Definition 6.1.2 (Joint PMF)

The joint PMF of r.v.s X and Y is function $P_{X,Y}$ given by:

$$P_{X,Y}(x, y) = P(X = x, Y = y)$$

Note: Just like univariate PMFs must be nonnegative and sum up to 1, a valid joint PMF are also:
- Nonnegative
- Sum up to 1: $\sum_x \sum_y P(X = x, Y = y) = 1$



How to convert from joint distribution to marginal distribution?

From the joint distribution of X and Y, we can get the distribution of X alone (marginal) by summing over all possible values of Y.

Definition 6.1.4 (Marginal PMF)

For discrete r.v.s X and Y, the marginal PMF of X is:

$$P(X = x) = \sum_y P(X = x, Y = y)$$

↳ the equation follows the axioms of probability (summing over disjoint cases)

↳ Can we convert marginal distribution to joint distribution?

If we only know the marginals of X and Y, there is no way to obtain joint distribution of X and Y without further assumptions, those are:
- Are X and Y independent?
- Dependencies, i.e, X and Y could be conditionally independent given r.v Z

How does Y distributed when we observe X?

Instead of using just PMF $P(Y = y)$, which doesn't take into account value of X, we use conditional PMF $P(Y = y \mid X = x)$

Example 6.1.8 (Chicken - egg)

Suppose a chicken lays a random number of eggs, $N \sim \text{Pois}(\lambda)$.

Each egg: $\begin{cases} \text{hatch successfully with probability } p \\ \text{hatch fail with probability } q = 1 - p \end{cases}$

Let $\begin{cases} X \text{ be the number that hatch} \\ Y \text{ be the number do not hatch} \end{cases}$ $\Rightarrow$ so $X + Y = N$

What is the joint PMF of X and Y?

- We try to find joint PMF $P(X=i, Y=j)$
- If we condition on total number of eggs N, X and Y are Binomial:

$$X \mid N=n \sim \text{Bin}(n, p)$$
$$Y \mid N=n \sim \text{Bin}(n, q)$$

- The joint PMF $P(X=i, Y=j)$ is only nonzero if $i + j = n$ ($i+j \neq n$ suggests the hatch and unhatched eggs don't add up to total eggs laid, which is impossible, hence has zero probability). So by LOTP:

$$P(X=i, Y=j) = P(X=i, Y=j \mid N=n) \cdot P(N=n)$$

Since $X=i$, $Y=j$ are the same event given $i+j=n$ and $N \sim \text{Pois}(\lambda)$:

$$P(X=i, Y=j) = P(X=i \mid N=i+j) \cdot P(N=n)$$
$$= \binom{i+j}{i} p^i q^j \cdot \frac{e^{-\lambda} \lambda^{i+j}}{(i+j)!}$$
$$= \frac{e^{-\lambda p} (\lambda p)^i}{i!} \cdot \frac{e^{-\lambda q} (\lambda q)^j}{j!}$$

Conclusion: The join PMF is the product of $\text{Pois}(\lambda p)$ as function of i, and $\text{Pois}(\lambda q)$ as function of j

This means: $\begin{cases} \cdot\ X \text{ and } Y \text{ are independent, since their joint PMF is the product of their marginal PMFs} \\ \cdot\ X \sim \text{Pois}(\lambda p) \text{ and } Y \sim \text{Pois}(\lambda q) \end{cases}$

X and Y are independent because the number of eggs is random and Poisson distributed (special property of the Poisson)

Continuous:

Once we understand the discrete case, it is not harder to consider continuous joint distributions.

Translate from discrete to continuous for joint distributions

Same as univariate, we follow these steps:

- Change Σ (sum) to $\int$ (integral)
- Change PMF to PDF

Remember that probability of any individual point is 0

CDF of joint distribution:

Joint CDF of X and Y is:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

Definition 6.1.9 (Joint PDF)

If X and Y are continuous with joint CDF $F_{X,Y}$, their joint PDF is the derivative of the joint CDF with respect to x and y:

$$f_{X,Y}(x, y) = \frac{\delta^2}{\delta x \, \delta y} F_{X,Y}(x,y)$$

Valid PDF:

- Nonnegative. $f_{X,Y}(x,y) \geq 0$
- Integrate to 1. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$

Example:

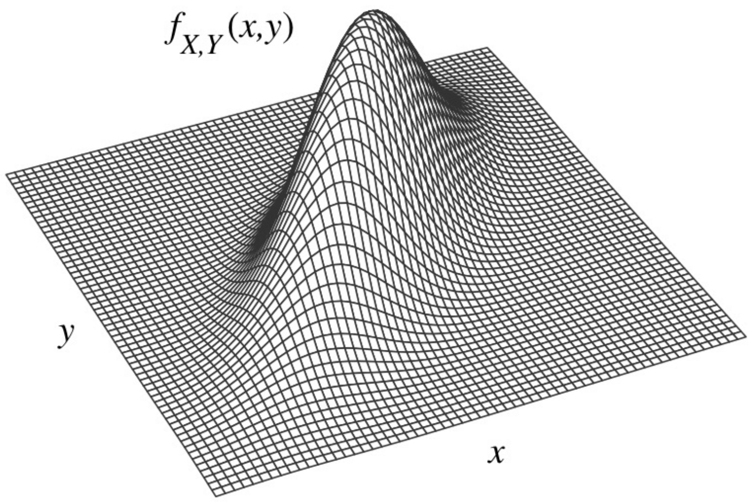
Similar to univariate case, to get the probability of an interval, we integrate the PDF. ie:

$$P(X < 3, 1 < Y < 4) = \int_{1}^{4}\int_{-\infty}^{3} f_{X,Y}(x,y)\, dx\, dy$$

In general, for set $A \subseteq \mathbb{R}^2$:

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\, dx\, dy$$

$\rightarrow$ Probability



Things to note from image of joint PDF of X and Y:

- The height of a single point on the surface does not represent s probability
- The probability of any individual points is 0
- The only way to get nonzero probability is to integrate over a region / are of positive values in the plane
- When we integrating, we are actually calculating the volume under the surface in the region.

In discrete case, we get marginal PMF of X by summing over all possible values of Y.

In continuous case, we get marginal PDF of X by integrating over all possible values of Y.

Definition 6.1.11 (Marginal PDF)

For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the marginal PDF of X:

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

What if we try to find joint PDF of 2 r.v.s from a joint PDF of 4 r.v.s?

If we have joint PDF of X, Y, Z, W and want joint PDF of X, W:

$$f_{X,W}(x, w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z,W}(x, y, z, w)\, dy\, dz$$

Lets consider how to update our distribution for Y after observing the value of X, using conditional PDF.

Definition 6.1.12 (Conditional PDF)

For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the conditional PDF of Y given $X = x$ is:

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

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

Notion 6.1.13:

Remember that we have 3 functions: $f_{X,Y}(x,y)$, $f_X(x)$, $f_{Y|X}(y|x)$ and the relationship between them are:

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

Relationship between joint, marginal, and conditional PDFs:

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

This relationship allow us to continue develop continuous analogs of Bayes rule and LOTP.

Theorem 6.1.15 (Continuous form of Bayes' rule and LOTP)

For continuous r.v.s X and Y:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) \cdot f_Y(y)}{f_X(x)}$$

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

We now have versions of Bayes' rule and LOTP for both discrete and continuous cases. Better yet, we can mix these ups, having 1 continuous r.v and 1 discrete.

Four versions of Bayes' rule:

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

Four versions of LOTP:

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

Finally, let's discuss the independence of continuous r.v.s

Definition 6.1.16 (Independence of Continuous r.v.s)

Random variables X and Y are independent for all x and y if:

$$F_{X,Y}(x,y) = F_X(x)\cdot F_Y(y)$$

If X, Y are continuous with joint PDF $f_{X,Y}$, then:

$$f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$$

And also:

$$f_{Y|X}(y|x) = f_Y(y)$$

Example 8.1.17 (Uniform on a region in the plane)

We will consider the independence of r.v.s X and Y in Uniform distribution on a square and on unit disk.

. Square case:

- Let (X, Y) be a random point in the square $\{(x,y) : x, y \in [0,1]\}$
- Joint PDF of X and Y is constant over the square and 0 outside of it:

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } x, y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

- We can derive to these conclusions:
  . X and Y are Unif $(1,0)$ marginally. We can check by computing:

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)\, dy = \int_0^1 1\, dy = 1$$

  - X and Y are independent, since:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

$$\Leftrightarrow \quad 1 = 1 \cdot 1 \qquad \langle \text{always true} \rangle$$

- Unit disk case:

- Let (X, Y) be random point in the unit disk $\{(x,y) : x^2 + y^2 \leq 1\}$
- Joint PDF of X and Y is:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- We can come to these conclusions:
  . X and Y are not independent, since larger $|X|$ will restrict Y in smaller range, we can check by:

$$x^2 + y^2 \leq 1$$

$$\Leftrightarrow \quad -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$$

  - Marginals of X and Y are not Uniform, we can check this by:

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy = \frac{2}{\pi}\sqrt{1-x^2} \quad , -1 \leq x \leq 1$$

Here we can that X and Y are more likely to fall near 0 than ± 1.

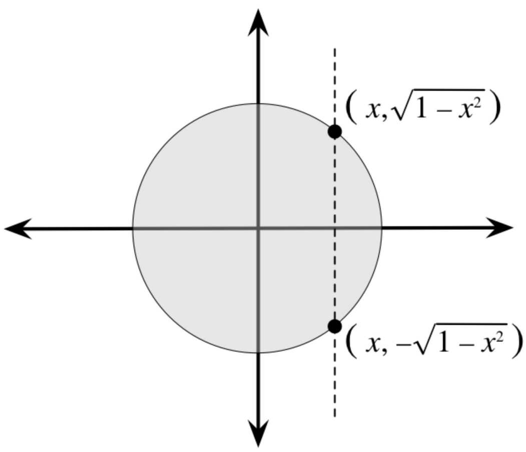


Figure 6.1.18: Bird's-eye view of the Uniform joint PDF on the unit disk. Conditional on $X = x$, Y is restricted to the interval $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$.

- However, the conditional PDF of Y given X is Uniform on interval $\left[-\sqrt{1-x^2}, \sqrt{1-x^2}\right]$, we can check this by:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{1}{\pi}}{\frac{2}{\pi}\sqrt{1-x^2}} = \frac{1}{2\sqrt{1-x^2}}$$

a constant

for $-\sqrt{1-x^2} \leq y \leq \sqrt{1+x^2}$ and 0 otherwise

Final conclusion:

- For a region R on a plane, the Uniform distribution on R is defined to have joint PDF that is constant inside R and 0 outside.
- If R is rectangle $\{(x,y) : a \leq x \leq b, c \leq y \leq d\}$, then X and Y are independent, a vertical slice of rectangle have same area (unlike a disk)
- But if any region where the value of X constrains the value of Y, i.e unit disk $\{(x,y) : x^2 + y^2 \leq 1\}$, then X and Y are not independent.
- However, for the case of unit disk, condition on X (or Y), the condition PDF is Uniform

Definition 6.1.5 (Conditional PMF)

For discrete r.v.s X and Y, the conditional PMF of Y given $X=x$ is:

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

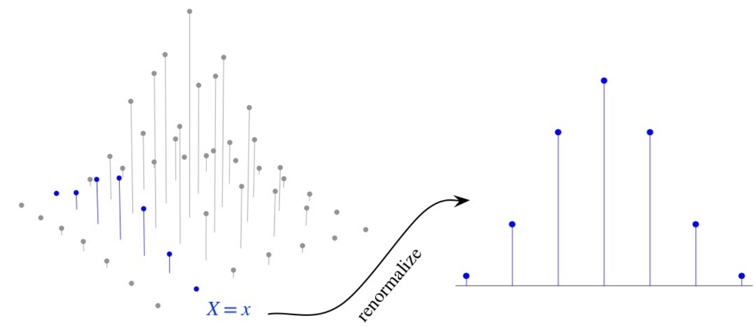
this is viewed as a function of y for fixed x

Relating conditional distribution of Y given X to X given Y:

Using Bayes' rule: $$P(Y=y \mid X=x) = \frac{P(X=x \mid Y=y)\ P(Y=y)}{P(X=x)}$$

Getting marginal PMF from conditional PMF:

Using LOTP: $$P(X=x) = \sum_y P(X=x \mid Y=y)\ P(Y=y)$$



Now with the understanding of joint, marginal, conditional distribution, we can revisit the definition of independence.

Definition 6.1.7 (Independence of Discrete r.v.s)

Random variables X and Y are independent if:

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \quad \text{for all } x, y$$

If X and Y are discrete, this is equivalent to the condition:

$$P(X=x, Y=y) = P(X=x) \cdot P(Y=y) \quad \text{for all } x, y$$

And is also equivalent to the condition:

$$P(Y=y \mid X=x) = P(Y=y) \quad \text{for all } y, x \text{ s.t. } P(X=x) > 0$$