Joint Distributions: Part One[1]

STA 256: Fall 2018

[1] This slide show is an open-source document. See last slide for copyright information.

Overview

Joint Distributions: The idea

- A single random variable is a measurement conducted on the elements of the sample space.
- More than one measurement can be taken on the same $\omega \in \Omega$.
- For example, X is height, and Y is weight.
- Of course more than two measurements are possible.
- Most real data sets have dozens of measurements on each sampling unit.
- Technically, a pair of jointly distributed random variables is a function from Ω to $\mathbb{R}^2$.

Probability

As with single random variables, the joint probability distribution of a set of random variables comes from the underlying probability distribution defined on the subsets of Ω.

$$P\left((X,Y)\in C\right)=P\left\{\omega\in\Omega:(X(\omega),Y(\omega))\in C\right\}$$

Joint Cumulative Distribution Functions

Whether X and Y are discrete or continuous, their joint distribution is defined by

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

Joint Probability Mass Function

Frequency Function

$$p(x, y) = P(X = x, Y = y)$$

Example

The discrete random variables X and Y have joint distribution

	$x = 1$	$x = 2$	$x = 3$
$y = 1$	$3/12$	$1/12$	$3/12$
$y = 2$	$1/12$	$3/12$	$1/12$

- What is $P(Y = 1)$? $p_y(1) = \frac{3}{12} + \frac{1}{12} + \frac{3}{12} = \frac{7}{12}$
- What is $P(Y = 2)$? $p_y(2) = \frac{1}{12} + \frac{3}{12} + \frac{1}{12} = \frac{5}{12}$
- What is $P(X = 2)$? $p_x(2) = \frac{1}{12} + \frac{3}{12} = \frac{4}{12}$

Marginal distributions

	$x=1$	$x=2$	$x=3$	$p_y(y)$
$y=1$	$3/12$	$1/12$	$3/12$	$7/12$
$y=2$	$1/12$	$3/12$	$1/12$	$5/12$
$p_x(x)$	$4/12$	$4/12$	$4/12$	1.00

Give the marginal distribution of Y.

$$p_y(y) = \begin{cases} \frac{7}{12} & \text{for } y = 1 \\ \frac{5}{12} & \text{for } y = 2 \\ 0 & \text{Otherwise} \end{cases}$$

Notation: $p_{xy}(1,2) = 1/12$

In general

- $p_x(x) = \sum_y p_{xy}(x, y)$
- $p_y(y) = \sum_x p_{xy}(x, y)$
- Two-dimensional, three-dimensional marginals etc. are obtained by summing over the other variables.
- Implicitly, the summation is over values where the joint probability is non-zero.

$$p_x(x) = \sum_{\{y:\, p(x,y)>0\}} p(x, y)$$

Multinomial Distribution

Begin with an example

- A six-sided die is rolled n times.
- The die is not necessarily fair.
- Probabilities are p_j for $j = 1, \ldots, 6$.
- Want probability of n_1 ones, $\ldots$, n_6 sixes.
- The probability of any particular string is $p_1^{n_1}\, p_2^{n_2}\, p_3^{n_3}\, p_4^{n_4}\, p_5^{n_5}\, p_6^{n_6}$.
- How many ways are there to choose n_1 positions for the ones, n_2 positions for the twos, etc.?
- $\binom{n}{n_1 \; \cdots \; n_6} = \frac{n!}{n_1! \; \cdots \; n_6!}$, so

$$P(X_1 = n_1, X_2 - n_2, \ldots, X_6 = n_6) = \binom{n}{n_1 \; \cdots \; n_6} p_1^{n_1} \cdots p_6^{n_6}$$

Multinomial Distribution in General

$$p(n_1, \ldots, n_r) = \begin{cases} \binom{n}{n_1 \cdots n_r} p_1^{n_1} \cdots p_r^{n_r} & \text{for } (n_1, \ldots, n_r) \in S \\ 0 & \text{Otherwise} \end{cases}$$

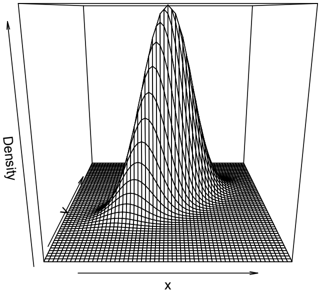
where $(n_1, \ldots, n_r) \in S$ means

$n_j \geq 0$ for $j = 1, \ldots, r$ and
$\sum_{j=1}^{r} n_j = n$.

If we count the number of people (in a random sample) in r different occupational categories, the multinomial is a reasonable model for the counts.

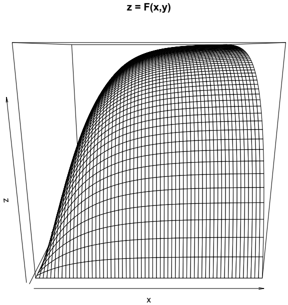
Continuous Jointly Distributed Random Variables

- Joint density of (X, Y) is not a curve, but a surface.



- Probability is volume rather than area.
- This is multivariable calculus.
- We need a quick lesson.

Partial Derivatives



- Think of holding x fixed at some value, disregarding all other points.
- Literally slice the surface with a plane at x.
- The cut mark on the surface is a function of y.
- It's just $F(x, y)$ treating x as a fixed constant.
- You can differentiate that function.

Vocabulary: "Partial derivatives"

- Consider a function of several variables, like $f(x_1, x_2, x_3)$.
- Differentiate with respect to one of the variables, treating the others as fixed constants.
- Call the result a *partial derivative*.

Notation for partial derivatives

- $\frac{\partial}{\partial x_2} f(x_1, x_2, x_3)$ or $\frac{\partial f}{\partial x_2}$ means differentiate $f(x_1, x_2, x_3)$ with respect to x_2, holding x_1 and x_3 constant.
- $\frac{\partial^2}{\partial x_1 \partial x_2} f(x_1, x_2, x_3)$ or $\frac{\partial^2 f}{\partial x_1 \partial x_2}$ means first differentiate with respect to x_2 holding x_1 and x_3 constant, and then differentiate the result with respect to x_1, holding x_2 and x_3 constant.
- When the derivatives are continuous functions, order of partial differentiation does not matter.
- $\frac{\partial^2}{\partial x_1^2} f(x_1, x_2, x_3)$ or $\frac{\partial^2 f}{\partial x_1^2}$ means differentiate twice with respect to x_1, holding x_2 and x_3 constant.

Example: $g(x_1, x_2) = x_1^2 e^{7x_2}$

$$\begin{aligned}
\frac{\partial g}{\partial x_1} &= 2x_1 e^{7x_2} \\
\frac{\partial^2 g}{\partial x_1 \partial x_2} &= \frac{\partial}{\partial x_1} x_1^2 7 e^{7x_2} \\
&= 14 x_1 e^{7x_2} \\
\frac{\partial^2 g}{\partial x_2^2} &= \frac{\partial}{\partial x_2} x_1^2 7 e^{7x_2} \\
&= 7x_1^2 \frac{\partial}{\partial x_2} e^{7x_2} \\
&= 49 x_1^2 e^{7x_2}
\end{aligned}$$

Multiple integration

$\int\int_A f(x,y)\,dx\,dy$ is the volume under the surface $f(x,y)$, over the region A in the x,y plane.

$$\int_a^b \int_c^d f(x,y)\,dx\,dy = \int_a^b \left(\int_c^d f(x,y)\,dx \right) dy$$

Recipe:

- Do the inner integral first, integrating from c to d, and treating y as a fixed constant.
- Then integrate the resulting function of y, from a to b.
- This yields volume under the surface $f(x,y)$, sitting over the region defined by $c < x < d$ and $a < y < b$.

Multiple integration can be pretty mechanical

$$\int_a^b \left(\int_c^d f(x,y)\,dx \right) dy$$

- Do the innermost integral first and work your way out, treating the other variables as constants at each step.
- If you are integrating over finite intervals, switch order of integration freely.
- If the quantity being integrated is non-negative, you may switch order of integration and the result is the same, even if the answer is “infinity.” Thank you, Mr. Fubini.
- There is one thing you often need to watch out for.

Region of integration

$$\int_a^b \left(\int_c^d f(x,y)\, dx \right) dy$$

- If the function $f(x, y)$ is a case function that is zero for some values of x and y, you need to take care that you are integrating over the correct region.
- You may need to sketch the region of integration.

Example

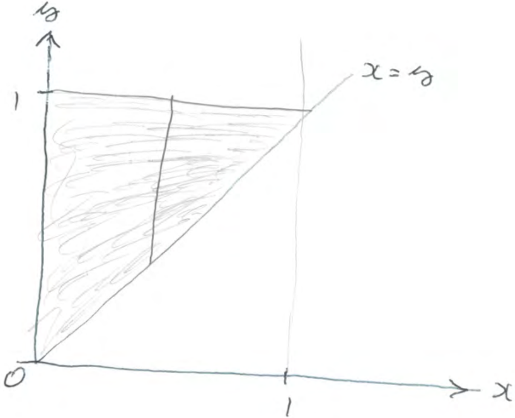
$$f(x, y) = \begin{cases} xy^2 & \text{for } x < y \\ 0 & \text{elsewhere} \end{cases}$$

Find $\int_0^1 \int_0^1 f(x, y)\, dy\, dx$.

$\int_0^1 \int_0^1 xy^2\, dy\, dx = \frac{1}{6}$, but that's not the right answer.

$f(x, y)$ only equals xy^2 for $x < y$.

Sketch the region of integration



As x goes from 0 to 1, y goes from x to 1.

$$\int_0^1 \int_0^1 f(x,y)\,dy\,dx = \int_0^1 \int_x^1 xy^2\,dy\,dx$$

The calculation

$$
\begin{aligned}
\int_0^1 \int_0^1 f(x,y)\,dy\,dx &= \int_0^1 \int_x^1 xy^2\,dy\,dx \\
&= \int_0^1 x \int_x^1 y^2\,dy\,dx \\
&= \int_0^1 x \left.\frac{y^3}{3}\right|_x^1 dx \\
&= \frac{1}{3}\int_0^1 x(1-x^3)\,dx \\
&= \frac{1}{3}\int_0^1 (x-x^4)\,dx \\
&= \frac{1}{3}\left.\left(\frac{x^2}{2}-\frac{x^5}{5}\right)\right|_0^1 \\
&= \frac{1}{3}\left(\frac{1}{2}-\frac{1}{5}\right) = \frac{1}{10}
\end{aligned}
$$

And not $\frac{1}{6}$. More examples will be given.

Joint CDFs

Let the continuous random variables X and Y have joint density function $f(x, y)$. Then

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t)\, dt\, ds$$

The notation extends to larger numbers of variables.

Fundamental Theorem of Calculus

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

At points where the derivatives exist and $f(x, y)$ is continuous.

Marginal distributions and densities

Integrate out the other variable(s)

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dy$$

$$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x,y)\,dx$$

Analogous to $p_x(x) = \sum_y p_{xy}(x,y)$

Marginal cumulative distribution functions

Show $\lim_{y\rightarrow\infty} F_{xy}(x,y)) = F_x(x)$. Use Fubini's Theorem, which says you can always switch order of integration as long as what you are integrating is non-negative.

$$
\begin{aligned}
\lim_{y\rightarrow\infty} F_{xy}(x,y)) &= \lim_{y\rightarrow\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f_{xy}(s,t)\,dt\,ds \\
&= \lim_{y\rightarrow\infty} \int_{-\infty}^{y} \int_{-\infty}^{x} f_{xy}(s,t)\,ds\,dt \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{xy}(s,t)\,ds\,dt \\
&= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{xy}(s,t)\,dt\,ds \\
&= \int_{-\infty}^{x} f_x(s)\,ds \\
&= F_x(x)
\end{aligned}
$$

Don't just move limits through integrals.

Copyright Information