Math 3070, Applied Statistics

Section 1

October 4, 2019

Lecture Outline, 10/4

Section 5.1

- Joint Probability Density Functions
- Marginal PDFs
- Independence
- Conditional PDFs

Joint Probability Density

Random variables X and Y are said to have a **joint probability density** (joint pdf) $f(x, y)$ if

$$
\begin{aligned}
P(a \le X \le b, c \le Y \le d) &= P(a \le X \le b \cap c \le Y \le d) \\
&= \int_a^b \int_c^d f(x, y)\ dy\ dx
\end{aligned}
$$

for all real constants $a \le b$ and $c \le d$.
Need $f(x, y) \ge 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\ dy\ dx = 1$

Example, Check PDF

A bank operates a drive-up facility and a walk-up window. On a random day, let X be the proportion of time the drive-up facility is in use, and let Y be the proportion of time the walk-up window is in use. Check that a valid joint pdf for X and Y is given by

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$f(x,y) \ge 0$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dy\ dx = \int_0^1\int_0^1 \frac{6}{5}(x+y^2)\ dy\ dx$$
$$= \frac{6}{5}\int_0^1 \left(xy + \frac{y^3}{3}\right)\Big|_{y=0}^{1}\ dx$$
$$= \frac{6}{5}\int_0^1 \left(x+\frac{1}{3}\right)\ dx = \frac{6}{5}\left(\frac{1}{2}+\frac{1}{3}\right) = 1$$

Marginal Probability Density Function

Let X and Y be random variables with joint pdf $f(x, y)$. The **marginal probability density function** (marginal pdf) of $X = x$ is defined as

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

Similarly, the marginal pdf of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\ dx$$

Note: f_X is simply the pdf of X considered as a random variable on its own. Same as $f_X(x)$ for PDFs.

Example, Marginal PDF

In the bank example, the joint pdf of two random variables X and Y was given by

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the marginal pdf of X.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\ dy \\ &= \int_0^1 \frac{6}{5}(x+y^2)\ dy \\ &= \frac{6}{5}\left(xy + \frac{y^3}{3}\right)\Big|_{y=0}^{1} = \frac{6}{5}\left(x + \frac{1}{3}\right) \end{aligned}$$

Independence

X and Y are **independent** continuous random variables if their joint pdf is the product of the marginal pdfs:

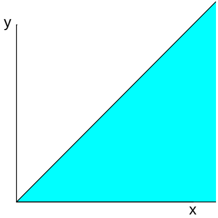
$$f(x, y) = f_X(x) f_Y(y)$$

Their events are also independent,
$P(a < X < b \cap c < Y < d) = P(a < X < b)P(c < Y < d)$.

Example, Independence

A system has two components; assuming their lifetimes X and Y are independent exponential random variables with $E(X) = 10$ and $E(Y) = 20$, what is the probability that the first component outlasts the second?

$$
\begin{aligned}
P(X \geq Y) &= \int_0^\infty \int_y^\infty f(x,y)\, dx\, dy \\
&= \int_0^\infty \int_y^\infty \frac{1}{200} e^{-\frac{x}{10} - \frac{y}{20}}\, dx\, dy \\
&= \frac{1}{200} \int_0^\infty e^{-y/20} \int_y^\infty e^{-x/10}\, dx\, dy \\
&= \frac{1}{200} \int_0^\infty e^{-y/20} \cdot 10 e^{-y/10}\, dy \\
&= \frac{1}{20} \int_0^\infty e^{-\frac{3}{20}y}\, dy = \frac{1}{20} \cdot \frac{20}{3} = \frac{1}{3}
\end{aligned}
$$



Conditional Probability Density Function

Let X and Y be random variables with joint pdf $f(x, y)$. The **conditional probability density function** of Y given X is

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

If X and Y are independent, then the conditional pdf of Y given X is simply the marginal pdf of Y:

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

$$P(a < Y < b|X = x) = \int_a^b f_{Y|X}(y|x)dy$$

Note: The event $X = x$ has zero probability. Typically, this is used when one value is already measured. Note, this would be a different calculation for events involving two random varaibles.

Example, Conditional Probability Density Function

Recall the bank example. X is the proportion of time the drive-up facility is in use and Y is the proportion of time the walk-up window is in use. They have the following joint PDF,

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Given that drive-up facility is always in use, what is the probability that the walk-up facility is in use less than half the time?
We are conditioning on the event that $X = 1$.
We can use the marginal from the previous example.

$$f_X(x) = \frac{6}{5}\left(x + \frac{1}{3}\right)$$

$$f_{Y|X}(Y|1) = \frac{f(1,y)}{f_X(1)} = \frac{\frac{6}{5}(1+y^2)}{\frac{6}{5}(1+\frac{1}{3})} = \frac{1+y^2}{1+\frac{1}{3}} = \frac{3}{4}(1+y^2)$$

Example, Conditional Probability Density Function

$$f_{Y|X}(y|1) = \begin{cases} \frac{3}{4}(1+y^2), & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} P\left(Y < \frac{1}{2} \middle| X = 1\right) &= \int_{-\infty}^{\frac{1}{2}} f_{Y|X}(y|1)dy \\ &= \int_{0}^{\frac{1}{2}} \frac{3}{4}(1+y^2)dy \\ &= \frac{3}{4}\left[y + \frac{y^3}{3}\bigg|_{y=0}^{1/2}\right] \\ &= \frac{3}{4}\left[\frac{1}{2} + \frac{1}{24}\right] \approx 0.40625 \end{aligned}$$

Non-Example, Conditional Probability Density Function

Recall the bank example. X is the proportion of time the drive-up facility is in use and Y is the proportion of time the walk-up window is in use. They have the following joint PDF,

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Given that drive-up facility in use less than half the time, what is the probability that the walk-up facility is in use less than half the time?

Non-Example, Conditional Probability Density Function

$$P(Y < 1/2 | X < 1/2) = \frac{P(Y < 1/2 \cap X < 1/2)}{P(X < 1/2)}$$
$$= \frac{\int_{-\infty}^{1/2} \int_{-\infty}^{1/2} f(x,y) dxdy}{\int_{-\infty}^{1/2} f_X(x) dx}$$

$$\int_{-\infty}^{1/2} \int_{-\infty}^{1/2} f(x,y) dxdy = \int_{0}^{1/2} \int_{0}^{1/2} \frac{6}{5}(x + y^2) dxdy = \frac{1}{5}$$

$$\int_{-\infty}^{1/2} f_X(x) dx = \int_{0}^{1/2} \frac{6}{5}(x + 1/3) dx = \frac{7}{20}$$

$$P(Y < 1/2 | X < 1/2) = \frac{1/5}{7/20} = \frac{4}{7}$$

Joint PDF, Summary

- Joint PDF: $f(x, y)$ such that
  $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\ dy\ dx$
- Marginal PDF: $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$. Is a univariate PDF.
- If X and Y are independent, $f(x, y) = f_X(x) f_Y(y)$.
  Their events are also independent
  $P(a < X < b \cap c < Y < d) = P(a < X < b)P(c < Y < d)$
- Conditional PDF:

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

  Note: the conditioned varible X is fixed, $X = x$. Think of this as a function of y. Different from conditioning on events.
- Can define these with more variables than two.

Example, Sum of Two Random Variables

Suppose that X and Y are independent continuous random variables. Determine the distribution of $Z = X + Y$. Determine the CDF.

$$P(Z < z) = P(X + Y < z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f(x,y)dxdy$$

on the inside integral, use $x = \nu - y \rightarrow dx = d\nu$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(\nu - y, y)d\nu dy = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(\nu - y, y)dyd\nu$$

differentiate to determine $f_Z(z)$

$$\frac{d}{dz}P(Z < z) = \int_{-\infty}^{\infty} f(z - y, y)dy$$

$$f_Z(z) = \int_{-\infty}^{\infty} f(z - y, y)dy$$

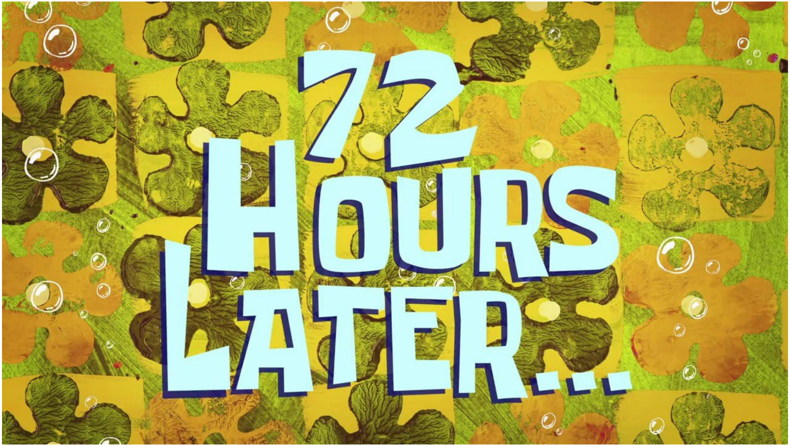
With independence, $f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y)dy$.

Example, Sum of Two Independent Normals

Suppose that $X \sim N(\mu_1, \sigma_1)$ and $Y \sim N(\mu_2, \sigma_2)$ are independent. Determine the distribution of $Z = X + Y$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(\frac{-(z-y-\mu_1)^2}{2\sigma_1^2}\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(\frac{-(y-\mu_2)^2}{2\sigma_2^2}\right) dy \end{aligned}$$

Example, Sum of Two Independent Normals



Example, Sum of Two Independent Normals

$$f(z) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(\frac{-(z - (\mu_1 + \mu_2))}{2(\sigma_1^2 + \sigma_2^2)}\right)$$

$$Z \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

`https://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_variables#Proof_using_convolutions`