Week 10: Bivariate continuous random variables

Armenak Petrosyan

Thursday class

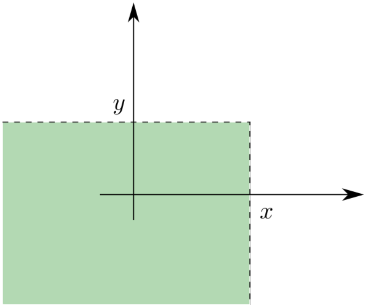
Bivariate cdf

- Let X, Y be two random variables (do not have to be discrete) defined on the same set of outcomes S.

Definition

The joint cdf of X, Y is called the following function

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



Definition

Bivariate random variable (X, Y) is called **continuous** if there exists a function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

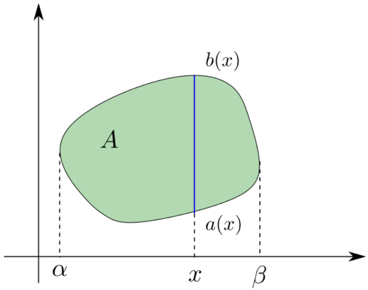
$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\xi, \eta) \, \mathrm{d}\xi \, \mathrm{d}\eta.$$

$f(x, y)$ is called the **joint pdf** of X and Y.

Theorem

For any set $A \subset \mathbb{R} \times \mathbb{R}$,

$$P[(X, Y) \in A] = \iint_A f(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$



Here

$$\iint_A f(x, y)\, \mathrm{d}x\, \mathrm{d}y = \int_{\alpha}^{\beta} \int_{a(x)}^{b(x)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y$$

Theorem

A function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *is the joint pdf of a bivariate continuous random variable if and only if*

1. $f(x, y) \geq 0$,
2. $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1$

Marginal pdf

Definition

For any $x \in \mathbb{R}$ and $y \in \mathbb{R}$ define

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x.$$

$f_X(x)$ and $f_Y(y)$ are called the **marginal pdf** of X and Y correspondingly.

- $f_X(x) \geq 0,\ f_Y(y) \geq 0$
- $$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1, \quad \int_{-\infty}^{\infty} f_Y(y)\,\mathrm{d}y = 1.$$
- $$P(X \leq x) = \int_{-\infty}^{x} f_X(\xi)\,\mathrm{d}\eta, \quad P(Y \leq y) = \int_{-\infty}^{y} f_Y(\eta)\,\mathrm{d}\eta.$$

Independence

Definition

Let (X, Y) be a continuous bivariate random variable. X and Y are called independent if

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

- If X and Y are independent then for any sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B).$$

Problem (4.4-1 in the textbook)

Let $f(x,y) = (3/16)xy^2$, $0 \leq x \leq 2, 0 \leq y \leq 2$, *be the joint pdf of* X *and* Y.

(a) *Find* $f_X(x)$ *and* $f_Y(y)$, *the marginal probability density functions.*

(b) *Are the two random variables independent? Why or why not?*

(c) *Compute the means and variances of* X *and* Y.

(d) *Find* $P(X \leq Y)$.

Solution

(a) *From definition*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_0^2 \frac{3}{16}xy^2\,\mathrm{d}y = \frac{x}{16}\,2^3 - \frac{x}{16}\,0^3 = \frac{x}{2}.$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \int_0^2 \frac{3}{16}xy^2\,\mathrm{d}x = \frac{3y^2}{16}\,\frac{2^2}{2} - \frac{3y^2}{16}\,\frac{0^2}{2} = \frac{3y^2}{8}.$$

Solution

(b) *Independent, because*

$$f_X(x)f_Y(y) = \frac{3xy^2}{16} = f(x,y).$$

(c)

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_0^2 \frac{x^2}{2}\,\mathrm{d}x = \frac{2^3}{6} - \frac{0^3}{6} \boxed{= \frac{4}{3}}.$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x = \int_0^2 \frac{x^3}{2}\,\mathrm{d}x = \frac{2^4}{8} - \frac{0^4}{8} = 2.$$

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = 2 - \frac{16}{9} \boxed{= \frac{2}{9}}.$$

Solution

(d)

$$
\begin{aligned}
P(X \le Y) &= \iint\limits_{x \le y} f(x,y)\ \mathrm{d}x\,\mathrm{d}y = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{y} f(x,y)\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\limits_{0}^{2} \int\limits_{0}^{y} \frac{3}{16} x y^2\ \mathrm{d}x\,\mathrm{d}y = \int\limits_{0}^{2} \left[\frac{3y^2}{16}\,\frac{y^2}{2} - \frac{3y^2}{16}\,\frac{0^2}{2} \right] \mathrm{d}y \\
&= \int\limits_{0}^{2} \frac{3y^4}{32}\,\mathrm{d}y = \frac{3 \cdot 2^5}{32 \cdot 5} - \frac{3 \cdot 0^5}{32 \cdot 5} \\
&\boxed{= \frac{3}{5}}.
\end{aligned}
$$

- For any $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$E[u(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x,y)\ f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

- **Covariance:**

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y.$$

- Correlation:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}.$$

- Least squares line:

$$y = \rho\frac{\sigma_Y}{\sigma_X}(y-\mu_X) + \mu_Y.$$

- It minimizes the

$$g(a,b) = E[(Y-aX-b)^2].$$

Conditional pdf

Definition

Conditional pdf of Y given $X = x$ is called the pdf defined by

$$h(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Definition

Conditional pdf of X given $Y = y$ is called the pdf defined by

$$g(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Conditional mean and variance

Definition

Conditional mean and variance of Y given $X = x$ are defined

$$E[Y|x] = \int_{-\infty}^{\infty} y\, h(y|x)\, \mathrm{d}y, \quad \mathrm{Var}[Y|x] = \int_{-\infty}^{\infty} (y - E[Y|x])^2\, h(y|x)\, \mathrm{d}y.$$

- $E[X|y], \mathrm{Var}(X|y)$ are defined similarly.
- $m(x) = E[Y|x]$ minimizes (among all functions $m : \mathbb{R} \to \mathbb{R}$)

$$E[(Y - m(X))^2].$$

- If $E[Y|x]$ is a linear function, then

$$E[Y|x] = \rho \frac{\sigma_Y}{\sigma_X}(y - \mu_X) + \mu_Y.$$

Exercise 2

Problem (4.4-17, modified)

Let $f(x,y) = c$, $0 \leq x \leq 4$, $x^2 - x/2 \leq y \leq x^2 + x/2$, *be the joint pdf of* X *and* Y.

(a) *Find* c.

(b) *Sketch the region for which* $f(x,y) > 0$.

(c) *Find* $f_X(x)$, *the marginal pdf of* X.

(d) *Calculate and plot the least squares line.*

(e) *Determine* $h(y|x)$, *the conditional pdf of* Y, *given that* $X = x$.

(f) *Calculate and plot* $E(Y|x)$, *the conditional mean of* Y, *given that* $X = x$.

Solution

(a)

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_{x^2-x/2}^{x^2+x/2} c\,\mathrm{d}y = c(x^2 + x/2) - c(x^2 - x/2) = cx$$

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = \int_{0}^{4} cx\,\mathrm{d}x = c\frac{4^2}{2} - c\frac{0^2}{2} = 8c \Rightarrow \boxed{c = \frac{1}{8}}$$

(b) See below.

(c) $f_X(x) = \frac{x}{8}$.

(d)

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_0^4 x\frac{x}{8}\,\mathrm{d}x = \frac{1}{8}\frac{4^3}{3}\boxed{= \frac{8}{3}}.$$

$$\mu_Y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^4 \int_{x^2-x/2}^{x^2+x/2} y\frac{1}{8}\,\mathrm{d}y\,\mathrm{d}x$$

$$= \frac{1}{8}\int_0^4 [\frac{1}{2}(x^2+x/2)^2 - \frac{1}{2}(x^2-x/2)^2]\,\mathrm{d}y\,\mathrm{d}x$$

$$= \frac{1}{8}\int_0^4 x^3\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{8}\frac{4^4}{4}\boxed{= 8}$$

(d)

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x = \int_0^4 x^2 \frac{x}{8}\,\mathrm{d}x = \frac{1}{8}\frac{4^4}{4} = 8.$$

$$\mathrm{Var}(X) = E[X^2] - \mu_X^2 = 8 - \frac{64}{9} \boxed{= \frac{8}{9}}$$

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^4 \int_{x^2-x/2}^{x^2+x/2} xy\frac{1}{8}\,\mathrm{d}y\,\mathrm{d}x$$

$$= \frac{1}{8}\int_0^4 x[\frac{1}{2}(x^2+x/2)^2 - \frac{1}{2}(x^2-x/2)^2]\,\mathrm{d}y\,\mathrm{d}x$$

$$= \frac{1}{8}\int_0^4 x^4\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{8}\frac{4^5}{5} = \frac{128}{5}$$

$$\mathrm{Cov}(X,Y) = E[XY] - \mu_X\mu_Y = \frac{128}{5} - \frac{64}{3} \boxed{= \frac{64}{15}}$$

(d)

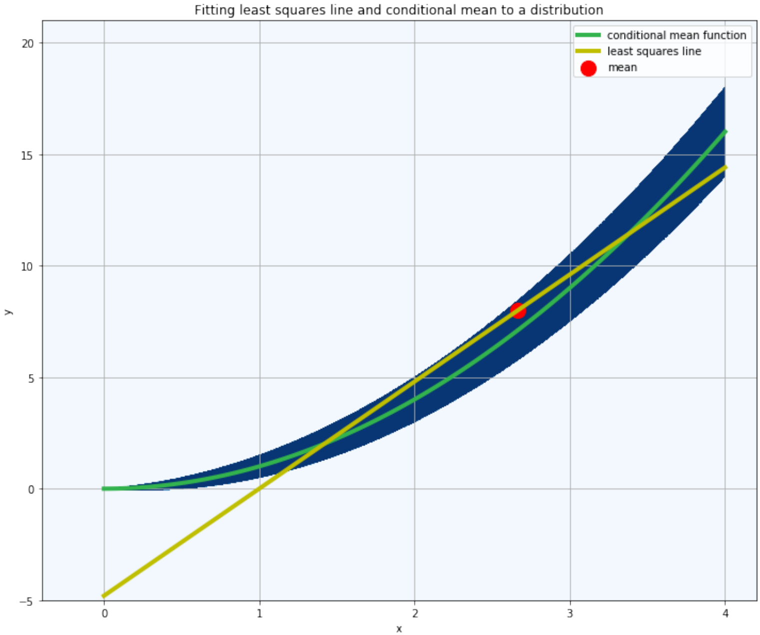
$$\boxed{y = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(x-\mu_X)+\mu_Y} = \frac{128}{5}\frac{9}{8}(x-\frac{8}{3})+8 = \frac{144}{5}(x-\frac{8}{3})+8$$

(e)

$$h(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{c}{cx} = \frac{1}{x}.$$

(f)

$$\begin{aligned} E[Y|x] &= \int_{-\infty}^{\infty} y\; h(y|x)\,\mathrm{d}y = \int_{x^2-x/2}^{x^2+x/2} y\frac{1}{x}\,\mathrm{d}y \\ &= \frac{1}{x}[\frac{1}{2}(x^2+x/2)^2 - \frac{1}{2}(x^2-x/2)^2] \\ &= \frac{1}{x}x^3 \\ &\boxed{= x^2}. \end{aligned}$$



Fitting least squares line and conditional mean to a distribution
conditional mean function
least squares line
mean
20
15
10
5
0
−5
y
0
1
2
3
4
x