3Y03 - Probability and Statistics for Engineering

Extra Worked Examples for Lecture 16.

WS19

The following is a collection of worked examples for the topics covered in Lecture 16, which we would have done at least partially in class if we hadn't lost 2 lectures to the weather. Please work through these independently and ask me (in person or by email) if you have any questions.

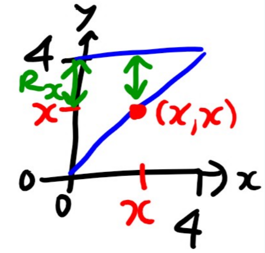
5-1.2 Marginal probability density functions

In Lecture 15 we had the following example of a joint p.d.f. for random variables X and Y:

$$f_{XY}(x,y) = \frac{1}{32}(x+y) \quad , \quad 0 < X < Y < 4.$$

From this we can find the marginal p.d.f.s of X & Y:

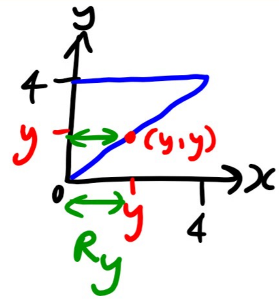
$$f_X(x) = \int_{R_x} f_{XY}(x,y)\,dy = \int_x^4 \frac{1}{32}(x+y)\,dy$$



$$= \left[\frac{1}{32}xy + \frac{1}{64}y^2\right]_x^4 = \frac{x}{8} + \frac{1}{4} - \frac{1}{32}x^2 - \frac{1}{64}x^2$$

$$= -\frac{3x^2}{64} + \frac{x}{8} + \frac{1}{4}$$

$$f_Y(y) = \int_{R_y} f_{XY}(x,y)\,dx = \int_0^y \frac{1}{32}(x+y)\,dx$$



$$= \left[\frac{1}{64}x^2 + \frac{1}{32}xy\right]_0^y = \frac{1}{64}y^2 + \frac{1}{32}y^2 = \frac{3y^2}{64}.$$

This in particular shows us that X and Y are not independent r.v.s (5–1.4) as $f_{XY}(x,y) = \frac{1}{32}(x+y)$

$$\neq f_X(x) f_Y(y)$$

This fact is not surprising though since the domain of f_{XY} is not rectangular – for any fixed x, the possible values that Y can take depends on x, and for any fixed y, the possible values that X can take depends on y.

5.2 Covariance & Correlation Worked Example

<u>Example</u> Suppose X and Y are jointly distributed r.v.s with joint pdf

$$f_{XY}(x,y) = \begin{cases} 6 & \text{if } x^2 < y < x \\ 0 & \text{otherwise.} \end{cases}$$

Find σ_{XY} and ρ_{XY}.

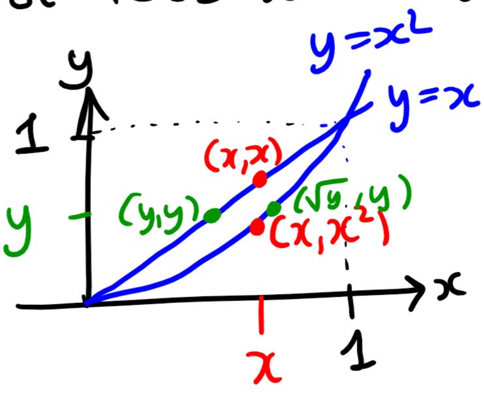
<u>Solution</u> Recall the formulae:

$$\sigma_{XY} = E(XY) - E(X)E(Y) \quad \text{and} \quad \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

and $\sigma_X = \sqrt{E(X^2) - E(X)^2}$ $\quad \sigma_Y = \sqrt{E(Y^2) - E(Y)^2}$

So we need to find $E(X), E(Y), E(X^2), E(Y^2), E(XY)$.

First let's draw the domain of f_{XY} : $x^2 < y < x$



To find all of the required expectations, we can use f_{XY}.

$$E(\boxed{X}) = \int_0^1 \int_y^{\sqrt{y}} \boxed{x} \cdot 6 \, dx\, dy = \int_0^1 \left[3x^2\right]_y^{\sqrt{y}} dy = \int_0^1 3y - 3y^2 \, dy$$

$$= \left[\frac{3y^2}{2} - y^3\right]_0^1 = \frac{3}{2} - 1 = \frac{1}{2}.$$

$$E(\boxed{Y}) = \int_0^1 \int_y^{\sqrt{y}} \boxed{y} \cdot 6 \, dx \, dy = \int_0^1 \left[6xy \right]_y^{\sqrt{y}} dy = \int_0^1 6y^{3/2} - 6y^2 \, dy$$

$$= \left[\frac{12y^{5/2}}{5} - 2y^3 \right]_0^1 = \frac{12}{5} - 2 = \frac{2}{5}.$$

$$E(\boxed{X^2}) = \int_0^1 \int_y^{\sqrt{y}} \boxed{x^2} \cdot 6 \, dx \, dy = \int_0^1 \left[2x^3 \right]_y^{\sqrt{y}} dy = \int_0^1 2y^{3/2} - 2y^3 \, dy$$

$$= \left[\frac{4y^{5/2}}{5} - \frac{y^4}{2} \right]_0^1 = \frac{4}{5} - \frac{1}{2} = \frac{3}{10}.$$

$$E(\boxed{Y^2}) = \int_0^1 \int_y^{\sqrt{y}} y^2 \cdot 6 \, dx \, dy = \int_0^1 \left[6y^2 x \right]_y^{\sqrt{y}} dy = \int_0^1 6y^{5/2} - 6y^3 \, dy$$

$$= \left[\frac{12y^{7/2}}{7} - \frac{3y^4}{2} \right]_0^1 = \frac{12}{7} - \frac{3}{2} = \frac{3}{14}.$$

$$E(\boxed{XY}) = \int_0^1 \int_y^{\sqrt{y}} \boxed{xy} \cdot 6 \, dx \, dy = \int_0^1 \left[3x^2 y \right]_y^{\sqrt{y}} dy = \int_0^1 3y^2 - 3y^3 \, dy$$

$$= \left[y^3 - \frac{3y^4}{4} \right]_0^1 = 1 - \frac{3}{4} = \frac{1}{4}.$$

Notice how every single calculation of expectation had the same form : $E(\underline{h(X,Y)}) = \iint_{\text{whole region}} \underline{h(x,y)} f_{XY}(x,y)\,dx\,dy$

$$= \int_0^1 \int_y^{\sqrt{y}} \underline{h(x,y)} \cdot 6\,dx\,dy$$

Now $\sigma_{XY} = E(XY) - E(X)E(Y) = \frac{1}{4} - \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{20} = \underline{\underline{0.05}}$.

This tells us that there is some (perhaps small) positive linear relationship between X and Y.

In order to calculate ρ_{XY} we need first to calculate:

$$\sigma_X = \sqrt{E(X^2) - E(X)^2} = \sqrt{\frac{3}{10} - \left(\frac{1}{2}\right)^2} = \frac{1}{\sqrt{20}}$$

$$\sigma_Y = \sqrt{E(Y^2) - E(Y)^2} = \sqrt{\frac{3}{14} - \left(\frac{2}{5}\right)^2} = \sqrt{\frac{19}{350}}.$$

Now $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \, \sigma_Y} = \dfrac{1/20}{1/\sqrt{20}\sqrt{\frac{19}{350}}} = \dfrac{\cancel{\sqrt{20}}\sqrt{350}}{\sqrt{20}\sqrt{19}} = \sqrt{\dfrac{35}{38}} = \underline{\underline{0.96}}$.

The value of ρ_{XY} is 0.96 which is close to 1, so we can see that X and Y are positively correlated.

5.4 Linear Functions of Random Variables

— Key Examples

For the sake of having all the key examples in one place, here is repeated the example from class:

① <u>Average</u> of a collection of (independent) r.v.s. $X_1, \dots, X_n$ with common mean μ & variance σ^2.

The average $\overline{X} = \frac{1}{n}(X_1 + \dots + X_n)$.

$$= \frac{1}{n} X_1 + \dots + \frac{1}{n} X_n.$$

So $E(\overline{X}) = \frac{1}{n} E(X_1) + \dots + \frac{1}{n} E(X_n)$

$$= \frac{1}{n}\mu + \dots + \frac{1}{n}\mu = \frac{1}{n} \cdot n \cdot \mu = \mu.$$

If $X_1, \dots, X_n$ are independent, then

$$V(\overline{X}) = \frac{1}{n^2} V(X_1) + \dots + \frac{1}{n^2} V(X_n)$$

$$= \frac{1}{n^2}\sigma^2 + \dots + \frac{1}{n^2}\sigma^2 = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

So $\overline{X}$ has mean μ, variance $\frac{\sigma^2}{n}$.

(2) Sum of (independent) random variables $X_1, \ldots, X_n$ with common mean μ and variance σ^2.

The sum $X = X_1 + \ldots + X_n$

has mean
$$E(X) = E(X_1) + \ldots + E(X_n)$$
$$= \mu + \ldots + \mu = n\mu$$

and, if $X_1, \ldots, X_n$ are independent, then
$$V(X) = V(X_1) + \ldots + V(X_n)$$
$$= \sigma^2 + \ldots + \sigma^2 = n\sigma^2.$$

So X has mean $n\mu$ and variance $n\sigma^2$.

Important Fact If $X_1, \ldots, X_k$ are independent, normally distributed r.v.s then a linear combination of them $Y = c_1 X_1 + \ldots + c_n X_n$ is also normally distributed.

So if $X_i \sim N(\mu_i, \sigma_i^2)$, then

$$E(Y) = c_1 E(X_1) + \ldots + c_n E(X_n)$$
$$= c_1 \mu_1 + \ldots + c_n \mu_n$$

and $V(Y) = c_1^2 V(X_1) + \ldots + c_n^2 V(X_n)$

$$= c_1^2 \sigma_1^2 + \ldots + c_n^2 \sigma_n^2 .$$

So $Y \sim N(c_1\mu_1 + \ldots + c_n\mu_n \,,\, c_1^2\sigma_1^2 + \ldots + c_n^2\sigma_n^2)$.

What if ① $Y = \bar{X}$ or ② $Y = X$ in the Special Cases above?

i.e. $\boxed{\text{If } X_1, \ldots, X_n \sim N(\mu, \sigma^2) \text{ are independent, then ① } \bar{X} \sim N(\mu, \tfrac{\sigma^2}{n}) \text{ ② } X \sim N(n\mu, n\sigma^2).}$

Pay close attention to these! ① especially we will use very often in the statistics part of the course.

Worked Example for 5.4 Linear Functions of R.Vs

Example Suppose that we have drinks bottles filled to a mean volume of 591 ml, with standard deviation 5 ml. Suppose that the volumes of the bottles are independent normal r.v.s.

(a) What is the probability that one bottle has less than 585 ml?

(b) Now 10 bottles are measured. What is the probability that their average volume is less than 585 ml?

Solution (a) Set X_1 = volume of one bottle

$$\sim N(\underset{\mu}{591}, \underset{\sigma^2}{25}).$$

Part (a) is not new material:

we need to find $P(X_1 < 585) = P\left(Z < \frac{585-591}{5}\right)$

$$= P(Z < -1.2) = 1 - P(Z < 1.2)$$

$$= 1 - 0.88493 = 0.11507.$$

(b) Set X_i = volume of bottle i $(i = 1, \ldots, 10)$.

$$\sim N(591, 25).$$

We want $P(\bar{X} < 585)$ where $\bar{X} = \frac{1}{10}(X_1 + \ldots + X_{10})$

$$\sim N(\underset{\mu}{591}, \underset{\frac{25}{10}}{\overset{\shortparallel}{2.5}})$$

$$P(\bar{X} < 585) = P\left(Z < \frac{585-591}{\sqrt{2.5}}\right)$$

$$= P\left(Z < \frac{-6}{1.58}\right) = P(Z < -3.79)$$

$$= 1 - P(Z < 3.79)$$

$$= 1 - 0.999925$$

$$= 0.000075.$$

Notice how the probability found in part (b) [that the average volume of 10 bottles is < 585ml] is much lower than the probability found in part (a) [that one bottle's volume is < 585ml]. This is because of the much lower standard deviation for $\bar{X}$ than for any individual X_i $\leftarrow \sigma = 5$

$\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{10}} = 1.58.$