ESE 520 Probability and Stochastic Processes

Lecture 9

”Expectation (continued). Covariance and correlation.”

Some more examples.

Example 1. Let X be a random variable with $\mathcal{N}(\mu, \sigma^2)$ probability distribution. Then $Var(X) = \sigma^2$.
Indeed:

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = [z = \frac{x-\mu}{\sigma}] =$$

$$\int_{-\infty}^{\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

More generally, for $m = 1, 2, ...$, we have

$$E(X-\mu)^{2m} = ... = \int_{-\infty}^{\infty} \sigma^{2m} z^{2m} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz =$$

$$\sigma^{2m} \int_{-\infty}^{\infty} z^{2m-1} \frac{z}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz =$$

- we use integration by parts here -

$$\sigma^{2m}[z^{2m-1}(-\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2})|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (2m-1) z^{2m-2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz]$$

$$\sigma^{2m}[0 + \int_{-\infty}^{\infty} (2m-1) z^{2m-2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz].$$

As the result, we obtain the following formula:

$$E(X-\mu)^{2m} = (2m-1)(2m-3)....3 \times 1 \sigma^{2m} = (2m-1)!!\sigma^{2m}$$

In particular, it follows then that $Var(X) = E(X-\mu)^2 = \sigma^2$.

Also, if Z has $\mathcal{N}(0, 1)$ distribution, then

$$E(Z^{2m}) = (2m-1)!!$$

Below are some additional "calculation" examples on the material already covered in class.

Example 2. ("maximum and minimum statistics") Let $X_i, i = 1, 2, ..., n$ be a sequence of independent and identically distributed (i.i.d.) random variables with the cdf $F(x)$ (respectively with the pdf $f(x)$). Define

$$Y := \max\{X_1, X_2, ..., X_n\},$$

and

$$Z := \min\{X_1, X_2, ..., X_n\}.$$

Find $F_Y(y)$ and $F_Z(z)$ (respectively $f_Y(y)$ and $f_Z(z)$).

Solution: We find

$$F_Y(y) = P(Y \leq y) = P(\max\{X_1, X_2, ..., X_n\} \leq y) =$$

$$P(X_1 \leq y, ..., X_n \leq y) =$$

- we use then independence of random variables -

$$P(X_1 \leq y)....P(X_n \leq y) = [F(y)]^n.$$

Also, if $f(x) = F_Y'(x)$ exists, we obtain

$$f_Y(y) = F'(y) = nF^{n-1}(y)f(y).$$

Similarly,

$$F_Z(y) = P(Z \leq z) = 1 - P(Z > z) = 1 - P(\min\{X_1, X_2, ..., X_n\} > z) =$$

$$1 - P(X_1 > z, ..., X_n > z) =$$

- we use then independence of random variables -

$$1 - P(X_1 > z)....P(X_n > z) = 1 - [1 - P(X_i \leq z)]^n = 1 - [1 - F(z)]^n.$$

Also, if $f(x) = F'(x)$ exists, we obtain

$$f_Z(z) = F_Z'(y) = n[1 - F(z)]^{n-1}f(z).$$

Example 3. Let X be a random variable having a Gamma probability distribution with parameters $p > 0$ and $\lambda > 0$. In other words, it has the pdf

$$f(x) = \frac{\lambda^p x^{p-1}}{\Gamma(p)}e^{-\lambda x}, x > 0$$

(and 0 otherwise). A typical notation for the Gamma distribution would be then $\Gamma(p, \lambda)$.

Assume X has $\Gamma(p, 1)$ distribution. Show that

$$E(X^n) = \frac{\Gamma(n+p)}{\Gamma(p)}.$$

Solution: We calculate:

$$E(X^n) = \int_0^\infty x^n \frac{x^{p-1}}{\Gamma(p)} e^{-x} dx = \frac{1}{\Gamma(p)} \int_0^\infty x^{(n+p)-1} e^{-x} dx =$$

$$\frac{\Gamma(n+p)}{\Gamma(p)}.$$

Example 4. Let X has a uniform probability distribution on the interval $(0, 1)$ and define $Y := \ln \frac{1}{X}$. Find $f_Y(y)$.

Solution: For $y \leq 0$, we have that $f_Y(y) = 0$ since $Y > 0$.

Assume $y > 0$. Then

$$F_Y(y) = P(Y \leq y) = P(\ln \frac{1}{X} \leq y) = P(\frac{1}{X} \leq e^y) =$$

$$P(X \geq e^{-y}) = 1 - F_X(e^{-y}).$$

As the result, we obtain

$$f_Y(y) = F_Y'(y) = -f_X(e^{-y})(-e^{-y}) = e^{-y}$$

since $f_X(x) = 1$.

Covariance and correlation between two random variables

If working with a random vector, two cases might occur: the components of a the vector are independent from each other or not. In the latter case, how to measure the dependence between two components?

This motivates the introduction of the following concept.

Definition 1. The number (if it exist) defined as

$$cov(X, Y) := E[(X - \mu_X)(Y - \mu_Y)] \tag{1}$$

is called the *covariance* between two random variables X and Y.

Furthermore, the number (if it exist)

$$\rho(X, Y) := \frac{cov(X, Y)}{\sigma_X \sigma_Y}. \tag{2}$$

is called the *correlation* (or *correlation coefficient*) between X and Y.

Often $cov(X, Y)$ is used as the *measure of association* between X and Y. If $cov(X, Y) > 0$, then X and Y are said to be *positively associated* and if $cov(X, Y) < 0$, then X and Y are said to be *negatively associated*.

The following properties will explain the role of correlation coefficient:

1) $|\rho(X, Y)| \leq 1$;
2) $|\rho(X, Y)| = 1$ if and only if there exit real numbers $a \neq 0$ and b so that $X = aY + b$ (X and Y are *linearly related* to each other);

Indeed, to verify the above properties, we notice that

$$0 \leq Var[\frac{X - \mu_X}{\sigma_X} \pm \frac{Y - \mu_Y}{\sigma_Y}] = E[\frac{X - \mu_X}{\sigma_X} \pm \frac{Y - \mu_Y}{\sigma_Y}]^2 \tag{3}$$

since

$$E[\frac{X - \mu_X}{\sigma_X} \pm \frac{Y - \mu_Y}{\sigma_Y}] = 0.$$

From another side, since

$$Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y)$$

for any two random variables X and Y, we can write

$$0 \leq Var[\frac{X - \mu_X}{\sigma_X} \pm \frac{Y - \mu_Y}{\sigma_Y}] =$$

$$Var[\frac{X - \mu_X}{\sigma_X}] + Var[\frac{Y - \mu_Y}{\sigma_Y}] \pm 2cov(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}) =$$

$$1 + 1 \pm 2\rho(X, Y) = 2(1 \pm \rho(X, Y))$$

implying 1).

Moreover, if $\rho(X, Y) = \mp 1$, then we see that

$$Var[\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y}] = 0$$

and by (3)

$$E[\frac{X-\mu_X}{\sigma_X} \pm \frac{Y-\mu_Y}{\sigma_Y}]^2 = 0. \tag{4}$$

It follows from (4) that the random variable inside of $E[\cdot]$ must be zero. We have then that

$$\frac{X-\mu_X}{\sigma_X} = \mp \frac{Y-\mu_Y}{\sigma_Y}$$

or

$$X = \mu_X \pm \frac{\sigma_X}{\sigma_Y}(Y - \mu_Y).$$

3) If X and Y are independent random variables, then it follows that $cov(X, Y) = 0$ and $\rho(X, Y) = 0$ (Why?)(if $\rho(X, Y) = 0$, then X and Y are said to be *uncorrelated*).

The converse of property 3) is not true in general except one case when X and Y both have normal distribution (to be seen later why this is the case).

Counterexample: (continued example from before) Let (X_1, X_2) be the coordinates of a point in $\mathbb{R}^2$ where $X_i, i = 1, 2$ both have $\mathcal{N}(0, 1)$ distribution and are independent. Let also (R, θ) be its polar coordinates. Since $R = \sqrt{X_1^2 + X_2^2}$, it is clear that X_1 and R are *not independent*. However, we calculate:

$$cov(X_1, R) = E[(X_1 - \mu_{X_1})(R - \mu_R)] = E(X_1 R) - \mu_R E(X_1) = E(X_1 R) =$$

$$E[R^2 \cos\theta]$$

- we use the fact that R and θ are independent since $f_{R,\theta)}(r, \theta) = f_R(r) f_\theta(\theta)$
-

$$E(R^2)E(\cos\theta) = 0$$

since

$$E(\cos\theta) = \int_0^{2\pi} \cos\theta \frac{1}{2\pi} d\theta = 0.$$

Thus, we have shown that R and X_1 are uncorrelated.

Remark 1. a) The correlation coefficient ρ is a measure of *linear dependence* between two random variables. It is not designed to measure other kinds of dependence that might exist. For that, there are other ways to measure dependencies different from linear one. We do not go into details of those constructions.

b) As we know, if X and Y are independent, then it always holds that

$$E[XY] = E[X]E[Y]. \tag{5}$$

However, the property (5) is not enough to conclude that X and Y are independent. For that, one would need a more stronger version of (5) to be true. One possibility for that would be to assume that

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

holds for all bounded and continuous real-valued functions g and h.