PROBABILITY THEORY

Session 2

EXPECTATION AND VARIANCE

Topics:

Expectation and

Variance of a random variable.

Covariance and Variance of

sums of random variables.

Key performance indicators

- In order to completely describe the behavior of a r.v. an entire function – the cdf or pdf/pmf – must be given. In some situations we are interested just in a few parameters that summarize the information provided by these functions.
- For example, when a large collection of data is assembled, we are typically interested not in the individual numbers, but rather in a certain quantities such as the average.

Expectation

- Expectation = expected value = mean = mean value
- Definition. The expectation of X is defined by (if X is a discrete r.v.)

$$E[X] = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i)$$

- The expectation of X is a weighted average of the possible values that X can take on
- If X is a continuous r.v.

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Expectation – using cdf

- If instead of density function we consider distribution function, we can use the following expressions in case when a discrete distribution takes values 0,1,2,3,… and a continuous distribution takes nonnegative values:

$$E[X] = \sum_{k=0}^{\infty} P(X > k)$$

$$E[X] = \int_0^{\infty} (1 - F(x))dx$$

Expectation

- The expected value is defined if the above sum or integral converges absolutely:

$$\sum_i |x_i| p(x_i) < \infty$$

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

- Example: the expectation does not exist for $F(x) = 1 - \frac{1}{x}, \quad x > 1$

Expectation of a function of a random variable

- Suppose that we are given a r.v. X and its probability distribution. Suppose we are interested in finding not the expected value of X, but the expected value of some function of X, g(X).
- g(X) is itself a r.v. with some probability distribution.
- How to obtain E[g(X)]?
  - One way from obtaining distribution for g(X) and definition
  - Another way: introduce a r.v. Y=g(X)

$$\boxed{E[Y] = E[g(X)] = \sum_i g(x_i)p(x_i)}$$

$$\boxed{E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx}$$

Properties of the expected value

- Proposition:

$$\boxed{E[aX+b] = aE[X] + b}$$

$$E[b] = b$$

Intuitive that the expestation of a constant is the constant itself

$$E[aX] = aE[X]$$

$$E[X+b] = E[X] + b$$

We can shift the mean of a r.v.
By adding a constant to it

Proof of linear property of expectation

Discrete case:

$$E[aX+b] = \sum_i (ax_i + b)p(x_i) =$$

$$= a\sum_i x_i p(x_i) + b\sum_i p(x_i) = aE[X] + b$$

Continuous case:

$$E[aX+b] = \int_{-\infty}^{+\infty} (ax+b) f(x)\,dx =$$

$$= a\int_{-\infty}^{+\infty} x f(x)\,dx + b\int_{-\infty}^{+\infty} f(x)\,dx = aE[X] + b$$

Expected value of function of random variables

- Many important results in probability theory concerns sums of random variables.
- We first consider, what it means to add two r.v.
- In general case, we consider a function of two random variables Z=g(X,Y). The problem of finding the mean of Z=g(X,Y) is similar to the problem of finding the mean of a function of a single r.v.:

$$\boxed{E[Z] = E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) p(x_i, y_j)}$$

$$\boxed{E[Z] = E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy}$$

Expectation of the sum of r.v.

- For example, if g(X,Y)=X+Y

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy =$$

$$\int_{-\infty}^{\infty}xf_X(x)dx + \int_{-\infty}^{\infty}yf_Y(y)dy =$$

$$E[X] + E[Y]$$

- It is easy to prove by mathematical induction, that the expected value of the sum of any finite number of random variables is the sum of the expected values of the individual random variables

- Note, that the r.v. do not have to be independent; formula is valid for any kind of r.v.

Expectation of the product of r.vs.

$$Z = X \cdot Y$$

more generally, we consider $Z = g(X,Y) = g_1(X) \cdot g_2(Y)$

$$E[Z] = E[\,g_1(X) \cdot g_2(Y)\,] =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x)\, g_2(y)\, \underbrace{f(x,y)}_{\substack{\| \\ f_X(x) \cdot f_Y(y)}}\, dx\, dy$$

if X and Y are independent

$$= \int_{-\infty}^{+\infty} g_1(x) f_X(x)\, dx \cdot \int_{-\infty}^{+\infty} g_2(y) f_Y(y)\, dy = E[g_1(X)]\, E[g_2(Y)]$$

Prediction of a value of a r.v.

- The best predictor of a r.v., in terms of minimizing its mean square error, is its expectation.

for any constant c

$$E[(X-c)^2] = E[(X-\mu+\mu-c)^2] =$$

$$= E[(X-\mu)^2] + 2(\mu-c)E[X-\mu] + (\mu-c)^2 \geq$$

$$\geq E[(X-\mu)^2]$$

since $E[X-\mu] = E[X] - \mu = 0$

Variance

- Motivation: The variations are given by r.v. D=X-E[X]. D can take on positive and negative values. We are only interested in magnitude → we work with |D| or D^2.
- Definition: variance of r.v. X is defined as

$$\boxed{Var(X) = E[(X - E(X)^2)]}$$

- This expression can be simplified:

$$\boxed{Var(X) = E[X^2] - E[X]^2}$$

- Definition: Standard deviation of r.v. X is $\boxed{Std(X) = \sqrt{(Var(X))}}$
- Example: Variance of an Indicator of an event

Another formula for variance

$$Var(X) = E[(X - E(X))^2] =$$

$$= E[X^2 - 2E[X]X + E[X]^2] =$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2 =$$

$$= E[X^2] - 2E[X]^2 + E[X]^2 =$$

$$= E[X^2] - E[X]^2$$

Variance

- Some useful identities concerning variances:

$$\boxed{Var(aX + b) = a^2 Var(X)}$$

$$Var(b) = 0$$

$$Var(X + b) = Var(X)$$

Derivation of formula for *Var(aX+b)*

$$Var(aX+b) = E[(aX+b-E[aX+b])^2] =$$

$$= E[(aX+b-aE[X]-b)^2] = E[(aX-aE[X])^2] =$$

$$= a^2 E[(X-E[X])^2] = a^2 Var(X)$$

Variance of a sum of random variables

- Question:
- We have seen that the expectation of a sum of r.v. is equal to the sum of their expectation.
- Is the corresponding statement for variances true?

- No, it’s not true. Example showing this:

$$\mathrm{Var}(X+X) = \mathrm{Var}(2X) = 4\,\mathrm{Var}(X) \neq \mathrm{Var}(X) + \mathrm{Var}(X)$$

- If r.vs. are independent, then it is true!

Covariance

- Definition. Covariance of 2 r.v. X and Y is defined by

$$\boxed{Cov(X,Y) = E[(X - E[X])(Y - E[Y])]}$$

- This expression can be simplified:

$$\boxed{Cov(X,Y) = E[XY] - E[X]E[Y]}$$

Simplification of the expression for covariance

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])] =$$

$$= E[XY - XE[Y] - YE[X] + E[X]E[Y]] =$$

$$= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] =$$

$$= E[XY] - E[X]E[Y]$$

Properties of covariance

1. $Cov(X, Y) = Cov(Y, X)$
2. $Cov(X, X) = Var(X)$
3. $Cov(aX, Y) = aCov(X, Y)$
4. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
5. If X and Y are independent, $Cov(X, Y) = 0$
6. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

   (thus, $Var(X + Y) = Var(X) + Var(Y)$ only in case of independent r.v.)

Property no 4 - proof

$$Cov(X+Y, Z) = E[(X+Y)Z] - E[X+Y]E[Z] =$$

$$= E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z] =$$

$$= Cov(X, Z) + Cov(Y, Z)$$

- Generalizing, using induction principle we can prove for any *n* and any *m*:

$$Cov\left(\sum_{i=1}^{n} X_i , \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$$

Property no 5 - proof

$$\mathrm{Cov}(X,Y) = E[\underbrace{(X-E[X])}_{g_1(X)}\underbrace{(Y-E[Y])}_{g_2(Y)}] =$$

$$= E[g_1(X)]\cdot E[g_2(Y)] = \underbrace{E[X-E[X]]}_{=0}\cdot\underbrace{E[Y-E[Y]]}_{=0} = 0$$

$$E[X-E[X]] = E[X]-E[X] = 0$$

- If X and Y are independent, then covariance is zero (X and Y are uncorrelated).
- It is possible that X and Y are uncorrelated, but not independent.

Variance of a sum of r.vs.

What shows covariance?

- The covariance is an indicator of the relationship between two r. vs. and shows the joint variability of two r.vs.
- If two variables tend to vary together (when one is above its expectation, then the other is also above its expectation), then covariance is positive. And negative otherwise.

Correlation coefficient

- Instead of covariance, we can work with dimensionless quality:
- Correlation coefficient of X and Y

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{(Var(X))}\sqrt{(Var(Y))}}$$

- Values of correlation coefficient lie between –1 and 1.

The nth moment of r.v. X

- The mean value is sometimes called the first moment of X.
- Definition. The nth moment of X is defined by

$$E[X^n] = \sum_i x_i^n p(x_i)$$

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx$$

Moment generating function

- Definition. The moment generating function of a r.v. X is defined for all values t by

$$\varphi(t) = E[e^{tX}] = \sum_i e^{tx_i} p(x_i)$$

$$\varphi(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

- This function is called moment generating function, because all moments of X can be obtained by successively differentiating it.
- There is a one-to-one correspondence between the moment generating function and the distribution function of a r.v.: the mgf uniquely determines the distribution and vice versa.

How to obtain E[X] and Var(X) from moment generating function?

$$\varphi'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] =$$

$$= E[Xe^{tX}]$$

$$\varphi'(0) = E[X]$$

$$\varphi''(t) = \frac{d}{dt} E[Xe^{tX}] = E\left[\frac{d}{dt}(Xe^{tX})\right] =$$

$$= E[X^2 e^{tX}]$$

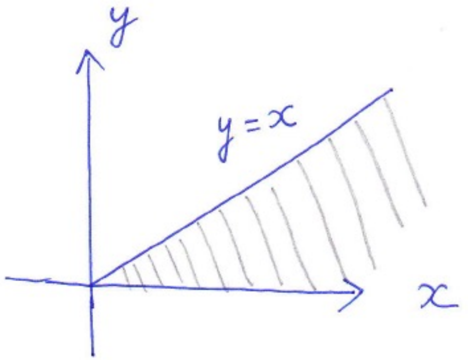
$$\varphi''(0) = E[X^2]$$

$$Var(X) = \varphi''(0) - \varphi'(0)^2$$

Example: covariance

joint pdf for X and Y:

$$f_{X,Y}(x,y) = \begin{cases} c e^{-x} e^{-y}, & 0 \leq y < x < \infty \\ 0, & \text{otherwise} \end{cases}$$



The constant c is found from the normalization condition:

$$1 = \int_0^{+\infty} \int_0^{x} c e^{-x} e^{-y} \, dy \, dx = \frac{c}{2} \quad \Rightarrow \quad c = 2$$

Example: covariance (ctnd)

The marginal pdfs:

$$f_X(x) = \int_0^x 2e^{-x}e^{-y}\,dy = 2e^{-x}(1-e^{-x}), \quad x \in [0,\infty)$$

$$f_Y(y) = \int_0^\infty 2e^{-x}e^{-y}\,dx = 2e^{-2y}, \quad y \in [0,\infty)$$

$$E[X] = \frac{3}{2}, \quad Var(X) = \frac{5}{4}$$

$$E[Y] = \frac{1}{2}, \quad Var(Y) = \frac{1}{4}$$

Example: covariance (ctnd)

$$E[XY] = \int_0^\infty \int_0^x xy \cdot 2e^{-x}e^{-y}\,dy\,dx =$$

$$= \int_0^\infty 2xe^{-x}(1-e^{-x}-xe^{-x})\,dx = 1$$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 1 - \frac{3}{4} = \frac{1}{4}$$

$$\mathrm{Corr}(X,Y) = \frac{1/4}{\sqrt{5/4}\,\sqrt{1/4}} = \frac{1}{\sqrt{5}}$$