Probability Calculus

Anna Janicka

lecture VIII, 20.11.2018

RANDOM VARIABLES – CONT.

RANDOM VECTORS – MULTIDIMENSIONAL RANDOM VARIABLES

Plan for Today

1. Moments of random variables – cont.
2. Sample characteristics
3. Definition of a Random Vector
4. Joint and marginal distributions
5. Discrete and continuous RV
6. Expected values of functions
7. Covariance, correlation
8. Expected value, variance

Moments – reminder

1. Definitions

For $p \in (0, \infty)$, *we define:*
(i) the **absolute moment** *of rank* p *for random variable* X *as* $\mathbb{E}|X|^p$ *(if this value is finite);*
For $p \in \mathbb{N}$, *we define:*
(ii) the **moment** *of rank* p *for random variable* X *as* $\mathbb{E}X^p$ *(provided that the* p*-th absolute moment exists);*
(iii) the **central moment** *of rank* p *for random variable* X *as* $\mathbb{E}(X - \mathbb{E}X)^p$ *(provided that the* p*-th absolute moment exists).*

Moments: skewness, kurtosis

2. Definitions

Let X be a random variable such that $\mathbb{E}|X|^3 < \infty$.

The **skewness** *of X is*

$$\alpha_3 = \frac{\mathbb{E}(X-\mathbb{E}X)^3}{(D^2X)^{3/2}} = \frac{\mathbb{E}(X-\mathbb{E}X)^3}{\sigma_X^3}.$$

Let X be a random variable such that $\mathbb{E}|X|^4 < \infty$.

The **kurtosis** *of X is*

$$\alpha_4 = \frac{\mathbb{E}(X-\mathbb{E}X)^4}{(D^2X)^2} - 3 = \frac{\mathbb{E}(X-\mathbb{E}X)^4}{\sigma_X^4} - 3.$$

3. Example: standard normal distribution

Empirical distributions

1. In reality, we frequently do not know the distributions of random variables, and work with *samples* instead.
2. *Let* $X_1, X_2, \ldots, X_n$ *be random variables with unknown distributions. An* **Empirical distribution (measure)** *for this sample is* $\mu_n(A) = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(A) = \frac{|\{i\leqslant n:\, X_i\in A\}|}{n}$,

Empirical distributions – cont.

3. *An* **empirical distribution function** *of the sample* $X_1, X_2, \ldots, X_n$ *is the function* $F\colon \mathbb{R} \to [0,1]$, *such that*
$$F_n(t) = \mu_n((-\infty, t]) = \frac{|\{i \leqslant n\colon X_i \leqslant t\}|}{n}.$$

this is the CDF of the empirical distribution

4. *A* **Quantile** *of rank p of the sample* $X_1, \ldots, X_n$ *is any number* x_p, *such that*
$$\mu_n((-\infty, x_p]) \geqslant p$$
$$\mu_n([x_p, \infty)) \geqslant 1 - p.$$

Empirical distributions – cont (2)

5. *A* **Sample mean** *for* $X_1, X_2, \ldots, X_n$ *is equal to* $m = \frac{X_1+X_2+\ldots+X_n}{n}$, *i.e. the arithmetic mean of* $X_1, X_2, \ldots, X_n$.

6. *A* **sample variance** *for* $X_1, X_2, \ldots, X_n$ *is equal to* $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$, *where* m *is the sample mean.*

the mean and the variance of the empirical distribution

End of material for the midterm test

to be held on

Friday, December 7th, 6:30 PM

Random vectors

1. A random vector (X_1, X_2, ..., X_n)
2. The joint distribution of a random vector:

The (joint) **distribution of a random vector** $X = (X_1, X_2, \ldots, X_n)$ *is a probability measure* μ_X *defined over* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$*, such that* $\mu_X(A) = \mathbb{P}(X \in A)$.

3. Marginal distributions:

$$\mu_{X_i}(B) = \mathbb{P}(X_i \in B) \text{ for } B \subseteq \mathbb{R},$$

such that for $A = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{i-1} \times B \times \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n-i}$

we have

$$\mathbb{P}(X_i \in B) = \mathbb{P}((X_1, X_2, \ldots, X_n) \in A) = \mu_X(A).$$

Random vectors – cont.

4. Example: joint distribution is more than the aggregate of marginal distributions.
5. Cumulative distribution function:

The **cumulative distribution function** *of a random vector* (X, Y) *is a function* $F_{(X,Y)} : \mathbb{R}^2 \to [0, 1]$*, such that* $F_{(X,Y)}(s, t) = \mathbb{P}(X \leqslant s, Y \leqslant t)$.

6. No simple definitions of quantiles...

Random vectors – types.

7. A discrete RV

A random vector (X,Y) *is* **discrete**, *if there exists a countable set* $S \subseteq \mathbb{R}^2$, *such that* $\mu_{(X,Y)}(S) = 1$.

8. Components are also discrete, marginals obtained by summation

9. A continuous RV

A random vector (X,Y) *is* **continuous**, *if there exists a density function, i.e. a function* $g : \mathbb{R}^2 \to [0,\infty)$, *such that for any* $A \in \mathcal{B}(\mathbb{R}^2)$, *we have* $\mu_{(X,Y)}(A) = \iint_A g(x,y)dxdy$.

Random vectors – types (cont.)

10. Examples of continuous RV:
  - drawing from a unit square
  - drawing from a circle
  - a different type of density

11. Marginal distributions of continuous RV:

Let (X, Y) be a random vector with density g. The marginal distributions of X and Y are also continuous, and the respective densities are equal to
$g_X(x) = \int_{\mathbb{R}} g(x,y)dy, \qquad g_Y(y) = \int_{\mathbb{R}} g(x,y)dx.$

Random vectors – types cont (2).

11. Marginal distributions (cont.)

More generally, if an n-dimensional random vector has a joint density function g, then the i-th component is continuous with density g_i, such that
$$g_i(x_i) = \iiint_{\mathbb{R}^{n-1}} g(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$
(the integral is over all variables other than X_i).

12. If marginals are continuous, then the joint distribution need not be.

Characteristics of random vectors

13. Expected values of functions of the components of a RV:

(i) Let (X, Y) be a discrete random vector with support S, and let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a Borel function. Then,
$\mathbb{E}\phi(X,Y) = \sum_{(x,y)\in S} \phi(x,y)\mathbb{P}((X,Y) = (x,y))$
(if the sum converges absolutely).
(ii) Let (X, Y) be a continuous random vector with density g and let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a Borel function. Then,
$\mathbb{E}\phi(X,Y) = \iint_{\mathbb{R}^2} \phi(x,y)g(x,y)dxdy$
(if the expected value exists).

14. Examples

15. Definitions

Let (X, Y) *be a random vector, such that* X *and* Y *have expected values, and such that* $\mathbb{E}|XY| < \infty$. *The* **covariance** *of variables* X *and* Y *is the value* $\mathrm{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$.

If, additionally, the variances of the two random variables exist, and $\mathrm{Var}X > 0$ *and* $\mathrm{Var}Y > 0$, *we may define the (Pearson's)* **correlation coefficient** *of variables* X *and* Y *as* $\rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X \cdot \mathrm{Var}Y}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$.

Covariance and correlation coefficient – cont.

16. Properties:

- invariance to shifts
- bilinearity of the covariance
- variance as a special case
- simplifying formula:

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y.$$

- capture the *linear* relationship, in other cases may be misleading

Correlation coefficient – properties

17. Schwarz inequality

Let $X, Y : \Omega \to \mathbb{R}$ *be random variables such that* $\mathbb{E}X^2 < \infty$ *and* $\mathbb{E}Y^2 < \infty$*. We then have*
$|\mathbb{E}XY| \leqslant (\mathbb{E}X^2)^{1/2}(\mathbb{E}Y^2)^{1/2}$.
Furthermore, we have an equality if and only if there exist two numbers $a, b \in \mathbb{R}$ *not simultaneously equal to zero, such that* $\mathbb{P}(aX = bY) = 1$.

18. Consequences for the correlation coef.

Let $X, Y : \Omega \to \mathbb{R}$ *be random variables with finite nonzero variances. Then* $|\rho(X, Y)| \leqslant 1$*. Furthermore, if* $|\rho(X, Y)| = 1$*, then there exist two numbers* $a, b \in \mathbb{R}$*, such that* $Y = aX + b$.

Expected value and covariance matrix

19. Definitions:

Let (X, Y) *be a two-dimensional random vector.*
Then, we have:
(i) If X *and* Y *have expected values, then the* **expected value** $\mathbb{E}(X, Y)$ *of the vector* (X, Y) *is the vector* $(\mathbb{E}X, \mathbb{E}Y)$.
(ii) If X *and* Y *have variances, then the* **covariance matrix** *of the vector* (X, Y) *is the matrix*

$$\begin{bmatrix} \mathrm{Var}X & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y) & \mathrm{Var}Y \end{bmatrix}.$$

For higher dimensions $(\mathbb{R}^d,\ d \geqslant 3)$, *we have, similarly:*
the expected value is the vector $(\mathbb{E}X_1, \mathbb{E}X_2, \ldots, \mathbb{E}X_d)$,
and the covariance matrix is the matrix $(\mathrm{Cov}(X_i, X_j))_{1 \leqslant i,j \leqslant d}$.