Probability Theory

Lecture 1: Multivariate Random Variables

Hector Rodriguez-Deniz

STIMA, IDA
Linköping University

Overview Lecture 1

- Course information
- Introduction
- Functions of random variables
- Bivariate random variables
- Multivariate random variables

Course Information

- Course evaluation for 2022: ≈ 4.1
- 6 Lectures, 6 Seminars
- Literature:
  Gut, A. *An intermediate course in probability*. 2nd ed. Springer-Verlag, New York, 2009 (textbook)
- Link for misprints and corrections:
  `http://www2.math.uu.se/~allan/81misprints.pdf`
- Course materials on Lisam

Contents

- Chapter 1: Multivariate random variables
- Chapter 2: Conditioning
- Chapter 3: Transforms
- Chapter 4: Order statistics
- Chapter 5: The multivariate normal distribution
- Chapter 6: Convergence

Examination

- Written exam with max score 20 points and grade limits
  A: 18p, **B**: 16p, **C**: 14p, **D**: 12p, **E**: 10p
- You are **allowed to bring** one sheet A4 with your own hand-written notes (written on both sides) and a pocket calculator.
- Following materials **will be distributed** with exam questions:
  - Table with common formulas and moment generating functions (available on Lisam)
  - Table of integrals (available on Lisam)
  - Table with distributions from Appendix B in textbook (on Lisam).
- Active participation in seminars gives **bonus points** to exam (max. 2). All bonus points can be added to exam result in order to reach E, D or C; 1 bonus point in order to reach B or A.

Random variables

- Ω - **sample space** of an experiment, i. e. space of all elementary events or outcomes
- A - any possible event $\quad \Rightarrow \quad A \subset \Omega$
- If Ω discrete:
  $\Omega = \{\omega_1, \omega_2, ...\}$, where $\omega_i \cap \omega_j = \emptyset$ for $i \neq j$
- **Random variable** X - real-valued function from Ω: $X : \Omega \to \mathbb{R}$
- Example:
  - Roll a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$X(\omega) = \begin{cases} 0 & \text{if } \omega = 1, 2 \text{ or } 3 \\ 1 & \text{if } \omega = 4, 5 \text{ or } 6 \end{cases}$$

Distribution of random variables

- Probabilities of events on sample space Ω imply probability distribution for random variable $X(\omega)$ on Ω
- Probability distribution of X is given by

$$\mathrm{P}(X \in C) = \mathrm{P}(\{\omega : X(\omega) \in C\}),$$

  where $\{\omega : X(\omega) \in C\}$ - event (in Ω) consisting of all outcomes that give values of X in C
- **Discrete random variable** can take only finite or countable number of different values $x_1, x_2, \ldots$
- **Continuous random variable** can take every value in an interval

Discrete random variables

- **Probability (mass) function, pmf**:

$$p(x) = \mathrm{P}(X = x)$$

- Discrete distributions:
  - Bernoulli distribution, $X \sim Be(p)$:

$$p(x) = \begin{cases} p & \text{if x=1} \\ 1 - p & \text{if x=0} \end{cases}$$

  - Poisson distribution, $X \sim Po(\lambda)$

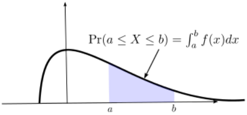
$$p(x) = \frac{\exp(-\lambda) \cdot \lambda^x}{x!}, \qquad \text{for } x = 0, 1, 2, ...$$

  - Binomial distribution, $X \sim \mathrm{Bin}(n, p)$

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \text{ for } x = 0, 1, ..., n$$

Continuous random variables

- **(Probability) density function, pdf**: $f(x)$
  - $f(x) \geq 0$ for all x
  - $\mathrm{P}(a \leq X \leq b) = \int_a^b f(x)dx$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$



- Continuous distributions
  - Uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

  - Normal, or Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Cumulative distribution function

- (Cumulative) **distribution function**, **cdf**:

$$F(x) = \mathrm{P}(X \leq x) \text{ for } -\infty < x < \infty$$

- Cdf is non-decreasing:

$$\text{If } x_1 \leq x_2 \text{ then } F(x_1) \leq F(x_2)$$

- Limits at $\pm\infty$: $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
- For discrete random variables

$$F(x) = \sum_{y \leq x} p(y)$$

- For continuous random variables

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \& \quad \frac{dF(x)}{dx} = f(x)$$

Expected value, variance, moments

- **Expected value** of X:

$$E\left(X\right) = \begin{cases} \sum_{\text{All } x} x \cdot p(x) & , X \textit{ discrete} \\ \int_{-\infty}^{\infty} x \cdot f\left(x\right) dx & , X \textit{ continuous} \end{cases}$$

- **Variance** of X:

$$Var\left(X\right) = E(X - EX)^2 = E\left(X^2\right) - \left(EX\right)^2$$

$$Var\left(X\right) = \begin{cases} \sum_{\text{All } x} (x - E(X))^2 \cdot p(x) & , X \textit{ discrete} \\ \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f\left(x\right) dx & , X \textit{ continuous} \end{cases}$$

- *n*-th moment of X: $E(X^n)$, for $n = 1, 2, 3, \ldots$

FUNCTIONS OF RANDOM VARIABLES

- We know distribution of X, but need distribution of $Y = g(X)$, where $g(\cdot)$ is some function
- Example 1: $Y = a + b \cdot X$, where a and b constants
- Example 2: $Y = 1/X$
- Example 3: $Y = \ln(X)$
- Example 4: $Y = \ln \frac{X}{1-X}$
- X discrete, $Y = g(X)$

pmf 是用来说明离散随机变量的 与 pdf 相区别

$$p_{\mathrm{Y}}(\mathrm{y}) = \mathrm{P}(Y = y) = \mathrm{P}\left[g(X) = y\right] = \sum_{x:g(x)=y} p_X(x)$$

where $p_X(x)$ - pmf for X, $p_Y(y)$ - pmf for Y

Function of continuous random variables

- X - continuous random variable with support (a, b)
- g - function with differentiable inverse $g^{-1} = h$
  - g strictly increasing

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq h(y)) = F_X(h(y))$$

$$\Rightarrow \quad f_Y(y) = f_X(h(y)) \cdot \frac{\partial h(y)}{\partial y}$$

  - g strictly decreasing

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \geq h(y)) = 1 - F_X(h(y))$$

$$\Rightarrow \quad f_Y(y) = -f_X(h(y)) \cdot \frac{\partial h(y)}{\partial y}$$

- For general monotonic g

这个实际上是上述两个的结合

$$f_Y(y) = f_X\left[h(y)\right] \left|\frac{\partial h(y)}{\partial y}\right| \quad \text{for } \alpha<y<\beta$$

where (α, β) is the mapped interval from (a, b), support for Y

Bivariate distributions

- **Joint** (or **bivariate**) **distribution** of two random variables X and Y is collection of all probabilities of form

$$\mathrm{P}\left[(X, Y) \in C\right]$$

- **Joint cumulative distribution function** (joint cdf) of X and Y:

$$F_{X,Y}(x, y) = \mathrm{P}(X \leq x, Y \leq y)$$

- **Discrete** random variables - **joint probability function** (joint pmf):

$$p_{X,Y}(x, y) = \mathrm{P}(X = x, Y = y)$$

- Properties:

  $F_{X,Y}(x, y) = \sum_{x_1 \leq x, y_1 \leq y} p_{X,Y}(x_1, y_1)$
  $\mathrm{P}\left[(X, Y) \in C\right] = \sum_{(x,y) \in C} p_{X,Y}(x, y)$
  $\sum_{All\ (x,y)} p_{X,Y}(x, y) = 1 \quad \& \quad \lim_{x \to \infty, y \to \infty} F_{X,Y}(x, y) = 1$

Bivariate distributions

- **Continuous** random variables - **joint density function**:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

- Properties:

  $F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(r,s) dr\, ds$

  $\mathrm{P}[(X,Y) \in C] = \int\int_C f_{X,Y}(x,y) dx\, dy$

  Example: $C = (0,a) \times (0,a)$, $a > 0$

  $$\mathrm{P}[(X,Y) \in C] = \int_0^a \int_0^a f_{X,Y}(x,y) dx\, dy$$

- X, Y discrete or continuous. $\mathrm{P}(a < X \leq b, c < Y \leq d)$:

$$F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$$

Bivariate normal distribution

- Pdf of **bivariate normal distribution**:

$$(N_1) \qquad f_{X,Y}(x,y) = \frac{1}{2\pi|\Sigma|^{1/2}} \times \exp(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)),$$

where

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \& \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{x,y} \\ \sigma_{x,y} & \sigma_y^2 \end{pmatrix}$$

$\mu_x = E(X)$, $\mu_y = E(Y)$, $\sigma_x^2 = Var(X)$, $\sigma_y^2 = Var(Y)$, $\sigma_{x,y} = Cov(X,Y)$

- Alternatively

$$(N_2) \qquad f_{X,Y}(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_x\sigma_y} \times$$

$$\exp\left(-\frac{1}{2\,(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right)$$

where $\rho = corr(X,Y)$

Marginal distributions

- Given joint distribution of X and Y

  (Marginal) distribution of X:

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

- Marginal probability / density function:

$$p_X(x) = \sum_{All\ y} p_{X,Y}(x, y) \text{ [Discrete case]}$$
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \text{ [Continuous case]}$$

- Interpretation: Marginal distribution for X tells about probability of different values of X, averaged over all possible values of Y

Independent variables

- Two random variables are **independent** if

$$\mathrm{P}(\mathrm{X} \in \mathrm{A}, \mathrm{Y} \in \mathrm{B}) = \mathrm{P}(\mathrm{X} \in \mathrm{A}) \cdot \mathrm{P}(\mathrm{Y} \in \mathrm{B})$$

  for all sets of real numbers A and B
- Two random variables are **independent** iff (if and only if)

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y), \text{ for all } x, y$$

- Two discrete random variables are **independent** iff

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y), \text{ for all } x, y$$

- Two continuous random variables are **independent** iff

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y), \text{ for all } x, y$$

Multivariate distributions

- Obvious extension to more than two random variables, $X_1, X_2, ..., X_n$
- Joint pdf

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$$

- Marginal distribution of X_1

$$f_{X_1}(x_1) = \int_{x_2} \cdots \int_{x_n} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_2 \cdots dx_n$$

- Marginal distribution of x_1 and x_2

$$f_{X_1,X_2}(x_1, x_2) = \int_{x_3} \cdots \int_{x_n} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_3 \cdots dx_n$$

...

Functions of random vectors

对应y1

- Let X be an n-dimensional continuous random variable
- Let X have density $f_X(x)$ on support $S \subset \mathbb{R}^n$
- Let $Y = g(X)$, where $g : S \to T \subset \mathbb{R}^n$ is a bijection (1:1 and onto)
- Let $h = (h_1, h_2, ..., h_n)$ be unique inverse of $g = (g_1, g_2, ..., g_n)$
- Let h be continuously differentiable with Jacobian determinant

对应X1

$$J = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{vmatrix}$$

Theorem

("The transformation theorem") The density of Y is

$$f_Y(y) = \begin{cases} f_X[h_1(y), h_2(y), ..., h_n(y)] \cdot |J|, & y \in T \\ 0, & \textit{otherwise}. \end{cases}$$

Thank you for your attention!