Introduction to statistical inference 2

Lauri Mehtätalo
University of Eastern Finland
School of Computing

April 17, 2018

II

Contents

Chapter 1

Recap from “Introduction to statistical inference 1”

1.1 Random variable

- Random variable is a function from sample space $\mathcal{S}$ of an experiment to sample space of the random variable $\mathcal{X}$, which is set of real numbers.

$$X : \mathcal{S} \rightarrow \mathcal{X}$$

- The sample space is a set of all possible values random variable can get.
- $\mathcal{X}$ can be
  - an interval of real axis (continuous random variable).

$$\mathcal{X} = [0, 10),\ \mathcal{X} = [0, 10],\ \mathcal{X} = (0, 10)$$

  - An uncountable set of integers (discrete random variable)

$$\mathcal{X} = \{0, 1, 2, \ldots\}$$

  - A countable set of integers or real numbers (discrete random variable)

$$\mathcal{X} = \{0, 1\},\ \mathcal{X} = \{0, 0.5, 1\},\ \mathcal{X} = \{0, 1, \ldots, 10\}$$

- Probabilities associated with each value of X are defined by the cumulative distribution function (cdf for short).

$$F_X(x) = P(X \leq x),\ \text{where } -\infty < x < \infty$$

Note: $F_X(x)$ is a step function if X is discrete.

Note: $F_X(x)$ is a continuous function if X is continuous.

- $F_X(x)$ or $F(x)$ is cdf, if

  1) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
  2) $F(x)$ is non-decreasing
  3) $F(x)$ is right-continuous

- Cdf is useful in calculation of any probabilities; for example

$$P(a < X \leq b) = F(b) - F(a)$$

Note: Be careful with $<$ and $\leq$ when working with discreate random variables.

- The probability density function (pdf for short) is defined for continuous random variable as

$$f_X(x) = F'(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}, \quad -\infty < x < \infty$$

  and

$$\int_{-\infty}^{x} f_X(t)\,\mathrm{d}t = F_X(x)$$

- The probability mass function (pmf for short) is defined for discrete random variables as

$$f_X(x) = P(X = x)$$
$$F_X(x) = \sum_{k=1}^{x} f_X(k)$$

1.2 Transformations of random variable

- Consider a monotonic function $g : \mathcal{X} \to \mathcal{Y}$
- $Y = g(X)$ is also a random variable; function g is called an transformation (muunnos).
- If $g(x)$ is a increasing function of x, then

$$F_Y(y) = F_X(g^{-1}(y))$$

- If $g(x)$ is a decreasing function of x, then

$$F_Y(y) = 1 - F_X(g^{-1}(y))$$

- The pdf of continuous Y is

$$f_Y(y) = F'_Y(y)$$

1.3 Expected values

$$E(X) = \mu_X = \begin{cases} \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} x f(x) & \text{if } X \text{ is discrete} \end{cases}$$

$$E(g(X)) = \begin{cases} \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f(x) & \text{if } X \text{ is discrete} \end{cases}$$

1.4 Variance

$$\begin{aligned} \sigma_X^2 = Var(X) \quad &= E(X - \mu_X)^2 \\ &= E(X^2 - 2X\mu_X + \mu_X^2) \\ &= E(X^2) - E(2X\mu_X) + E(\mu_X^2) \\ &= E(X^2) - \underbrace{2\mu_X \underbrace{E(X)}_{\mu_X}}_{2\mu_X^2} + E(\mu_X^2) \\ &= E(X^2) - \mu_X^2 \end{aligned}$$

$$sd(X) = \sqrt{Var(X)} = \sigma_X$$

1.5 Bivariate random variables

- For two discrete random variables, the joint pmf is defined as

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

- For two continuous random variables, we define the joint pdf $f_{X,Y}(x, y)$ as

$$P((X, Y) \in A) = \iint\limits_A f(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

- The expected value for transformation $g(X, Y) : \mathbb{R}^2 \to \mathbb{R}$ (for example, $g(X, Y) = XY$ or $g(X, Y) = \frac{X}{Y}$) is

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

  if (X, Y) is continuous, and

$$E(g(X, Y)) = \sum_{x,y \in \mathbb{R}^2} g(x, y) f(x, y)$$

  if (X, Y) is discrete.

- The marginal pmf / pdf for X are

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) \quad \text{(pmf)}$$
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y \quad \text{(pdf)}$$

and correspondingly for Y.

- The conditional pmf / pdf are both defined as

$$f(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad \text{(for both doscrete and continuous random variables)}$$

and correspondingly for $x \mid y$.

1.6 Independence

- Random variables are said to be independent ($X \perp\!\!\!\perp Y$) if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

- For independent random variables, conditional distribution $y \mid x$ is

$$f(y \mid x) = f_y(y)$$

regardless of the value of x.

1.7 Covariance

- Covariance measures the linear association between two random variables X and Y

$$\begin{aligned}
cov(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\
&= E(XY) - \mu_X \mu_Y \\
cov(X, Y) &= cov(Y, X) \\
corr(X, Y) &= \rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \qquad \text{where } -1 \leq \rho_{XY} \leq 1
\end{aligned}$$

Note: $cov(X, X) = Var(X)$

Note: If $X \perp\!\!\!\perp Y$, then $cov(X, Y) = 0$, but if $cov(X, Y) = 0$, it does not mean X and Y are necessarily independent.