5 Linear Methods in Regression: A Statistical Revisit

5.1 Introduction to Parameter Estimation Theory / Inverse Problems

Model: $y = f(x; \beta^*)$

x: input

y: output

β^*: ground-truth parameter

Goal of Parameter Estimation Theory / Inverse Problems: find estimators for β^* from observations of (x, y)

Remark: An <u>estimator</u> is a rule for calculating an estimate of the model parameter.

It can be deterministic (see Chapter 1) or stochastic (this chapter).

From now on, all estimators are random variables or random vectors.

5.1.1. Math Prep.

Def: A <u>random variable</u> is a measurable function defined on a probability space.
(Intuitively, it is a variable that can take different values with certain probabilities.)

Remark: There are two types of random variables: discrete and continuous. We focus on continuous random variables.

For a random variable ξ, denote:

$P(\xi \leq a)$: prob. that ξ takes a value $\leq a$. $a \in \mathbb{R}$

$F_\xi(x) \overset{\text{def}}{=} P(\xi \leq x)$: cumulative distribution function (CDF)

$p_\xi(x) \overset{\text{def}}{=} \dfrac{dF_\xi}{dx}(x)$: probability density function (PDF)

$E(\xi) \overset{\text{def}}{=} \int_{-\infty}^{\infty} x\, p_\xi(x)\, dx \in \mathbb{R}$: expectation or mean

$Var(\xi) \overset{\text{def}}{=} E(\xi - E(\xi))^2 \in \mathbb{R}$: variance of ξ

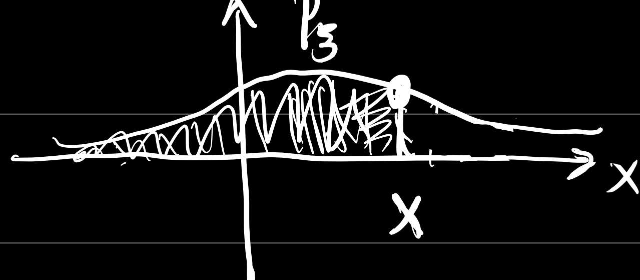
Remark (1) $\int_a^b p_\xi(x)\, dx \overset{\text{funda thm of Calculus}}{=} F_\xi(b) - F_\xi(a)$

$= P(\xi \leq b) - P(\xi \leq a)$

$= P(a < \xi \leq b)$

prob that $a < \xi \leq b$.

(2) $\int_{-\infty}^{\infty} p_\xi(x)\,dx = 1$



(3) $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ $\qquad \lim\limits_{x\to\infty} F_\xi(x) = 1$

Example: (1) ξ is called a Gaussian random variable / normally distributed, denoted by $\xi \sim N(\mu, \sigma^2)$, if

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Then $E(\xi) = \mu$, $\quad Var(\xi) = \sigma^2$

(2) ξ is called a Laplacian random variable / double exponentially distributed, denoted by $\xi \sim L(\mu, b)$, if

$$p_\xi(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

Then $E(\xi) = \mu$, $Var(\xi) = 2b^2$.

Def: A <u>random vector</u> is a vector of random variables.

For a random vector $\xi = (\xi_1, \cdots, \xi_n)^T$, each ξ_i is a random variable. Denote:

$P(\xi_1 \leq a_1, \cdots, \xi_n \leq a_n)$: prob that $\xi_1 \leq a_1, \cdots$

$\xi_n \leq a_n$ simultaneously,

$(a_1, \cdots, a_n) \in \mathbb{R}^n$.

$$F_\xi(x) = F_\xi(x_1, \cdots, x_n) \stackrel{\text{def}}{=} P(\xi_1 \leq x_1, \cdots, \xi_n \leq x_n) \quad : \text{CDF}$$

$$P_\xi(x) = P_\xi(x_1, \cdots, x_n) \stackrel{\text{def}}{=} \frac{\partial^n F_\xi}{\partial x_1 \cdots \partial x_n}(x) \quad : \text{PDF}.$$

$$E(\xi) \stackrel{\text{def}}{=} (E(\xi_1), \cdots, E(\xi_n)) \in \mathbb{R}^n \quad : \text{expectation or mean}$$

$$\mathrm{Cov}(\xi) \overset{\text{def}}{=} E\left[(\xi - E(\xi))(\xi - E(\xi))^T\right]$$

$$= \Big(\underbrace{E\left[(\xi_i - E(\xi_i))(\xi_j - E(\xi_j))\right]}_{(i,j)\text{-entry}} \Big)_{n\times n}$$

In particular, the diagonal entries of $\mathrm{Cov}(\xi)$ are:

$$\big(\mathrm{Cov}(\xi)\big)_{ii} = E\left[(\xi_i - E(\xi_i))^2\right] = \mathrm{Var}(\xi_i)$$

Remark: $\displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_\xi(x)\, dx_1 \cdots dx_n = 1$

Def: The random variables $\xi_1, \cdots, \xi_n$ are called <u>independent</u> if

$$P(\xi_1 \le a_1, \cdots, \xi_n \le a_n) = P(\xi_1 \le a_1) P(\xi_2 \le a_2) \cdots P(\xi_n \le a_n) \quad \forall (a_1, \cdots, a_n) \in \mathbb{R}^n$$

or equivalently:

$$F_{\xi}(x_1, \cdots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n)$$

or equivalently:

$$p_{\xi}(x_1, \cdots, x_n) = p_{\xi_1}(x_1) \cdots p_{\xi_n}(x_n).$$