Statistics

Notes by Finley Cooper

22nd January 2026

Contents

1 Parametric Estimation

We observe some data $X_1, \dots, X_n$ iid random variables taking values in a sample space $\mathcal{X}$. Let $X = (X_1, \dots, X_n)$. We assume that X_1 belongs to a *statistical model* $\{p(x;\theta) : \theta \in \Theta\}$ with θ unknown. For example $p(x;\theta)$ could be a pdf.

Let's see some examples

(i) Suppose that $X_1 \sim \text{Poisson}(\lambda)$ where $\theta = \lambda \in \Theta = (0, \infty)$.

(ii) Suppose that $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$.

We have some common questions about these statistical models.

(i) We want to give an estimate $\hat{\theta} : \mathcal{X}^n \to \Theta$ of the true value of θ.

(ii) We also want to give an interval estimator $(\hat{\theta}_1(X), \hat{\theta}_2(X))$ of θ.

(iii) Further we want to test of hypothesis about θ. For example we might make the hypothesis that $H_0 : \theta = 0$.

Let's do a quick reivew of IA Probability. Let $X : \Omega \to \mathbb{R}$ be a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So Ω is the sample space, $\mathcal{F}$ is the set of events, and $\mathbb{P} : \mathcal{F} \to [0,1]$ is the probability measure.

The cumulative distribution function (cdf) of X is $F_X(s) = \mathbb{P}\left(X \leq x\right)$. A discrete random variable takes values in a countable set $\mathcal{X}$ and has probability mass function (pmf) given by $p_X(x) = \mathbb{P}\left(X = x\right)$. A continuous random variable has probability density function (pdf) f_X satisfying $P(X \in A) = \int_A f_X(x) \mathrm{d}x$ (for measurable sets A). We say that $X_1, \dots, X_n$ are independent if $\mathbb{P}\left(X_1 \leq x_1, \dots, X_n \leq x_n\right) = \prod_{i=1}^n \mathbb{P}\left(X_i \leq x_i\right)$ for all choices $x_1, \dots, x_n$. If $X_1, \dots, X_n$ have pdfs (or pmfs) $f_{X_1}, \dots, f_{X_n}$, then this is equivalent to $f_X(x) = \prod_{i=1}^n f_{X_i}(x_i)$ for all x_i. The expectation of X is,

$$\mathbb{E}(x) = \begin{cases} \sum_{x \in \mathcal{X}} x p_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) & \text{if } X \text{ is continuous} \end{cases}.$$

The variance of X is $\text{Var}\left(X\right) = \mathbb{E}[(X - \mathbb{E}(X))^2]$. The moment generating function of X is $M(t) = \mathbb{E}[e^{tX}]$ and can be used to generate the momentum of a random varaible by taking derivatives. If two random variables have the same moment generating functions, then they have the same distribution.

The expectation operator is linear and

$$\text{Var}\left(a_1 X_1 + \dots + a_n X_n\right) = \sum_{i,j=1}^n a_i a_j \,\text{Cov}\left(X_i, X_j\right),$$

where $\text{Cov}\left(X_i, X_j\right) = \mathbb{E}[(X_i - \mathbb{E}(X_i)(X_j - \mathbb{E}(X_j))]$. In vector notation writing X as the column vector of X_i and a as the column vector for a_i we get that

$$\mathbb{E}[a^T X] = a^T E[X].$$

Similar for the variance we get that

$$\text{Var}\left(a^T X\right) = a^T \,\text{Var}\left(X\right) a$$

where $\text{Var}\left(X\right)$ is the covariance matrix for X with entries $\text{Cov}\left(X_i, X_j\right)$.

If X is a discrete random variable with pmf $P_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$ and marginal pmf $P_Y(y) = \sum_{x \in X} P_{X,Y}(x,y)$, then the conditional pmf is

$$P_{X|Y}(x \mid y) = \mathbb{P}(X = x \mid Y = y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}.$$

We also have the law of total expectation,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]].$$