Fundamental Probability and Statistics

"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know," Donald Rumsfeld

Probability Theory

Probability Space: $(\Omega, \mathcal{F}, P)$

Ω – Sample Space – set of all possible outcomes of an experiment

$\mathcal{F}$ – σ-field of subsets of Ω that contains all events of interest

$P : \mathcal{F} \to [0,1]$: probability or measure that satisfies

(i) $P(\emptyset) = 0$

(ii) $P(\Omega) = 1$

(iii) $A_i \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$ implies $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Example: Toss possibly biased coin once

$\Omega = \{H, T\}$

$\mathcal{F} = \{\emptyset, H, T, \Omega\}$

Take

$P(\emptyset) = 0\ ,\ P(H) = p\ ,\ P(T) = 1 - p\ ,\ P(\Omega) = 1$

Note: Fair coin if p = 1/2

Probability Theory

Example: Two coins tossed possibly multiple times and outcome is ordered pair

$$\Omega = \{(H,H),(T,H),(H,T),(T,T)\}$$

$$\mathcal{F} = \{\emptyset,(H,H),(T,H),(H,T),(T,T),\Omega,\{(H,H),(T,H)\},\cdots\}$$

Let

$$A = \{(H,H),(T,H)\}$$

$$B = \{(H,H),(H,T)\}$$

Then

$$P(A) = \frac{1}{2}\ ,\ P(B) = \frac{1}{2}$$

$$P(A \cap B) = \frac{1}{4}\ ,\ P(A \cup B) = \frac{3}{4}$$

Definition: Events A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

Random Variables and Distributions

Definition: A random variable is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega \,|\, X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$.

Definition: The value of a random variable X at a point $\omega \in \Omega$

$$x = X(\omega)$$

is a realization of X.

Definition: Associated with every random variable X is a cumulative distribution function $F : \mathbb{R} \to [0, 1]$ given by

$$\begin{aligned} F(x) &= P\{\omega \in \Omega \,|\, X(\omega) \leq x\} \\ &= P(X) \leq x \end{aligned}$$

Example: Take $\Omega = \{(H, H), (T, H), (H, T), (T, T)\}$ and define $X(\omega)$ as number of heads

$$X(H, H) = 2$$
$$X(H, T) = X(T, H) = 1$$
$$X(T, T) = 0$$

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{4} & \text{if} \quad 0 \leq x < 1 \\ \frac{3}{4} & \text{if} \quad 1 \leq x < 2 \\ 1 & \text{if} \quad x \geq 2 \end{cases}$$

Distributions and Densities

Definition: A random variable X is discrete if it takes values in a countable subset $\{x_1, x_2, \cdots\}$, only, of $\mathbb{R}$.

Definition: X is continuous if its distribution function can be expressed as

$$F(x) = \int_{-\infty}^{x} f(s)ds \ , \ x \in \mathbb{R}$$

for some integrable function $f : \mathbb{R} \to [0, \infty)$ where f is called the probability density function (PDF) of X.

Definition: The probability mass function of a discrete random variable X is a function $f : \mathbb{R} \to [0, 1]$ given by $f(x) = P(X = x)$.

PDF Properties:

(i) $f(x) \geq 0$

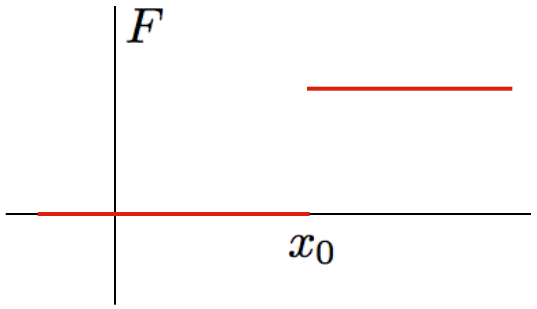
(ii) $\int_{\mathbb{R}} f(x)dx = 1$

(iii) $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$

Density Properties

Example:

$$X(x) = \begin{cases} 0 & , \quad x < x_0 \\ 1 & , \quad x \geq x_0 \end{cases}$$



$$F(x) = \begin{cases} 0 & , \quad x < x_0 \\ 1 & , \quad x \geq x_0 \end{cases} \Rightarrow f(x) = \delta(x - x_0)$$

Example:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \ , \ -\infty < x < \infty$$

$$F(x) = \int_{-\infty}^{x} f(s)ds = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

where

$$\text{erf}(x) = \frac{2}{\pi}\int_0^x e^{-s^2} ds$$

Note: $X \sim N(\mu, \sigma^2)$

Density Properties

Additional Properties:

(i) Mean, first moment or expected value

$$\mu = E(X) = \int_{\mathbb{R}} x f(x) dx$$

(ii) n^{th} moment

$$E(X^n) = \int_{\mathbb{R}} x^n f(x) dx$$

(iii) Second central moment (difference between X and μ)

$$\sigma^2 = \text{var}(X) = E[(X-\mu)^2] = \int_{\mathbb{R}} (x-\mu)^2 f(x) dx$$

Note:

$$\sigma^2 = E(X^2) - \mu^2$$

(iv) The covariance of X and Y is

$$\begin{aligned} \text{cov}(X,Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Note: X and Y are uncorrelated if $\text{cov}(X,Y) = 0$

Multivariate Distributions

Note: Important for longitudinal data

$$X = (X_1, X_2, \cdots, X_n)$$

$$\Rightarrow X : \Omega \to \mathbb{R}^n \text{ Random Vector}$$

Joint CDF: $F : \mathbb{R}^n \to [0,1]$ by

$$\begin{aligned} F(x_1, \cdots, x_n) &= P\{w \in \Omega \,|\, X_j(\omega) \leq x_i \,,\, j = 1, \cdots, n\} \\ &= P(X \leq x) \end{aligned}$$

Joint Density (if it exists): $f : \mathbb{R}^n \to [0, \infty)$

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(s_1, \cdots, s_n) d_{s_1} \cdots d_{s_n}$$

Example: Let $X \sim N(\mu, V)$

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |V|}} \exp\left[-\frac{1}{2}(x-\mu)V^{-1}(x-\mu)^T\right]$$

Multivariate Distributions

Example: Let $X \sim N(\mu, V)$

$$f(x) = \frac{1}{\sqrt{(2\pi)^n |V|}} \exp\left[-\frac{1}{2}(x-\mu)V^{-1}(x-\mu)^T\right]$$

Note:

- $E(X) = \mu$ so $E(X_j) = \mu_j$
- $V = (v_{ij})$ is the covariance matrix since $v_{ij} = \text{cov}(X_i, X_j)$. This is often written

$$V = E((X-\mu)^T(X-\mu))$$

Note:

$$V = \text{cov}(X) = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \text{cov}(X_1, X_n) & \text{cov}(X_2, X_n) & \cdots & \text{var}(X_n) \end{bmatrix}$$

Multivariate Distributions

Definition: The marginal distribution functions of X and Y are

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f(u,y)dy \right) du \end{aligned}$$

with a similar definition for $F_Y(y)$.

Definition: Marginal density function of X

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

Definition: X and Y are independent if and only if

$$F(x,y) = F_X(x)F_Y(y)$$

or

$$f(x,y) = f_X(x)f_Y(y)$$

Note: X and Y are independent $\Rightarrow$ $\text{cov}(X,Y) = 0$ $\Leftrightarrow$ X and Y are uncorrelated

Estimators and Estimates

Definition: An estimator is a function or procedure for deriving an estimate from observed data. An estimator is a random variable whereas an estimate is a real number.

Example: Suppose we want to estimate the variance $\mathrm{var}(Y) = \sigma^2$ of a random variable Y using a sample of n independent observations. Consider two statistics.

$$S^2 = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu)^2$$

$$T^2 = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \bar{Y})^2 \text{ where } \bar{Y} = \frac{1}{n}[Y_1 + Y_2 + \cdots + Y_n]$$

Now

$$E(S^2) = \frac{1}{n}\sum_{j=1}^{n} E(Y_j - \mu)^2 = \frac{1}{n}\sum_{j=1}^{n}\sigma^2 = \sigma^2 \text{ but } \mu \text{ is often unknown}$$

$$E(T^2) = \frac{n-1}{n}\sigma^2 \text{ so biased}$$

Unbiased estimator: $\mathcal{T}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$