Probability Theory and Statistics

Ruben Traicevski

January 2024

Contents

1 Notation

2 Basic Notions

2.1 Sample Space

- All possible outcomes from a random experiment, which will form a set called the *Sample Space* denoted by Ω
  E.g $\Omega = [H, T]$ Is the Sample Space for flipping a coin

  A random number between 0,1 corresponds to $\Omega = (0, 1)$ which is the entire segment between 0 and 1
- All subsets of the sample space Ω form a set denoted 2^Ω
- $|\Omega|$ denotes the cardinal of Ω which is the number of elements
- This means the number of elements in a subset,2^Ω of Ω is given by $2^{|\Omega|} = n$. Then 2^Ω has 2^n elements

2.2 Events and Probability

The set of parts 2^Ω satisfies the following properties

- It contains the empty set $\emptyset$
- If it contains set A, then it also contains its complement $\bar{A} = \frac{\Omega}{A}$
- It is closed with regard to unions, meaning a union of sets belonging to 2^Ω also belongs to 2^Ω

Any subset F of 2^{Ω} that satisfies the previous three properties is called a σ-field. The sets belonging to F are called events. *This means the complement of event, or union of events is also an event*
The chance of occurrence of an event is measured by a probability function $P : F \to [0, 1]$ which satisfies the following two properties,

- $P(\Omega) = 1$
- For any mutually disjoint events $\in$ F,

$$P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$$

The triplet (Ω, F, P) is called a probability space

3 Statistics

3.1 Random Variables

The σ-field provides F the knowledge about which events are possible on the considered probability space (Ω, F, P). A random variable X is a function which is defined as $X : \Omega \to \mathbb{R}$
Given any two numbers $a, b \in \mathbb{R}$, then all the states for which X takes values between a and b forms a set that is an event $\in F$

$$\{\omega \in \Omega : a < X(\omega) < b\} \in F$$

3.2 Integration in Probability Measure

Let Ω_i be a partition of Ω, then each Ω_i is an event and its associated probability is $P(\Omega_i$
Now consider the *characteristic function* of a set $A \subset \Omega$ which is defined by,

$$\chi_A = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

A *Simple Function* is a sum of characteristic functions $f = \Sigma_i^n c_i \chi_{\Omega_i}$ where $c_i \in \mathbb{R}$
This means $f(\omega) = c_k for \omega \in \Omega_k$
The integral of the simple function f is defined by

$$\int_{\Omega} f dP = \Sigma_i^n c_i P(\Omega_i)$$

- Linearity: For any two Random Variables X and Y and a,b $\in \mathbb{R}$

$$\int_{\Omega} (aX + bY) dP = a \int_{\Omega} X dP + b \int_{\Omega} Y dP$$

- Positivity: If X ≤ 0 then,

$$\int_{\Omega} X dP \leq 0$$

3.3 Two Convergence Theorems

3.4 Distribution Functions

Let X be a random variable on the probability space (Ω, F, P). The *(Cumulative) Distribution Function* of X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = P(\omega; X(\omega) \leq x)$$

The distribution function is non decreasing and satisfies the following limits,

$$\lim_{h \to -\infty} F_X(x) = 0$$

$$\lim_{h \to +\infty} F_X(x) = 1$$

We will now define the *probability density function of X*, $p(x)$

$$\frac{d}{dx} F_X(x) = p(x)$$

Also note the following important relationship between the distribution function, probability and probability density function of the random variable X

$$F_X(x) = P(X \leq x) = \int_{X \leq x} dP(\omega) = \int_{-\infty}^{x} p(u)du$$

The probability density function, $p(x)$ has the following properties

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(u)du = 1$$

From the Fundamental Theorem of Calculus

$$P(a < X < b) = P(\omega; a < X(\omega) < b) = \int_a^b p(x)dx$$

3.5 Independence

3.6 Expectation

A random variable $X : \Omega \rightarrow \mathbb{R}$ is called integrable if

$$\int_\Omega |X(\omega)dP(\omega) = \int_\mathbb{R} |x|p(x)dx < \infty$$

The *Expectation* of an integrable random variable X is defined by

$$E[X] = \int_\Omega |X(\omega)dP(\omega) = \int_\mathbb{R} xp(x)dx$$

In general, for any measurable function g, we have

$$E[g(X)] = \int_\Omega |g(X(\omega))dP(\omega) = \int_\mathbb{R} g(x)p(x)dx$$

The expectation operator is linear,

$$E[cX] = xE[X]$$

$$E[X+Y] = E[X] + E[Y]$$

If two random variables are independent and integrable,

$$E[XY] = E[X]E[Y]$$

The covariance of two random variables is defined by

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

$$Var(X) = Cov(X,X) = E[(X - \mu_X)^2]$$

If X and Y are independent, the covariance is zero. However, the converse is not necessarily true.
The covariance can be standardized by useing correlation,

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X, \sigma_Y}$$

$$-1 \leq \rho(X,Y) \leq 1$$

3.7 Change of Measure in an Expectation

3.8 Moment Generating Function

Given a random variable X and it's probability density function $p(x)$, then the moment generating function is given by

$$m_X(t) = E[e^{tX}] = \int e^{tx} p(x) dx$$

the moments of a function are certain quantitative measures related to the shape of the function's graph. If the function is a probability distribution, then the first moment is the expected value, the second central moment is the variance, the third standardized moment is the skewness, and the fourth standardized moment is the kurtosis.
The nth moments of X, given by $\mu_n = E[X^n]$ are generated by the derivatives of $m_X(t)$ about $t = 0$,

$$\frac{d^n m_X(t)}{dt^n} = \mu_n$$

There is also relation to the Laplace Transform,

$$m_X(t) = \mathcal{L}[f_X](-t)$$

3.9 Sums of Random Variables

4 Distributions

4.1 Normal Distribution

A random variable X is said to have a normal distribution if its probability density function is given by,

$$p(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$E[x] = \mu \qquad Var[X] = \sigma^2 \tag{1}$$

If $X \sim N(\mu, \sigma^2)$, then $Y = aX + b$ is also normally distributed with

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

The cumulative distribution function is given by,

$$F_X(x) = \frac{1}{\sqrt{2\sigma^2\pi}} \int_{-\infty}^{x} exp(-\frac{(t-\mu)^2}{2\sigma^2} dt$$

4.2 Log-Normal Distribution

Let X be a normally distributed random variable, then $Y = e^X$ is said to be log normal distributed with $x > 0$

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}$$

$$E[x] = e^{\mu + \frac{\sigma^2}{2}} \qquad Var[X] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \tag{2}$$

4.3 Multivariate Normal Distribution

4.4 Gamma Distribution

Given parameters $a > 0, b > 0$, then the Gamma density function is,

$$p(x) = \frac{x^{a-1}e^{-x/b}}{b^a\Gamma(a)}$$

Where $\Gamma(a)$ is the gamma function,

$$\Gamma(a) = \int_0^\infty y^{a-1}e^{-y}$$

When a is an integer, $\Gamma(a)$ is now $\Gamma(n) = (n-1)!$

$$E[X] = ab \qquad Var[X] = ab^2 \tag{3}$$