Probability and Statistics

Alejandro Campos

March 20, 2025

Contents

1 Events and Probabilities

- Ω is a set of outcomes ω of a stochastic experiment. It is referred to as the sample space.
- $\mathcal{A}$ is a set of events A. An event A is a subset of the sample space Ω. $\mathcal{A}$ is a σ-algebra on Ω, which means
  - $\mathcal{A}$ is non-empty
  - If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$, where $A^c = \Omega \backslash A$
  - If $A_1, A_2, ..., A_n \in \mathcal{A}$ then $\bigcup_{n=1}^{\infty} A_n$.
- The function $P : \mathcal{A} \to [0, 1]$ is a probability measure, if
  - $P(\Omega) = 1$

- $P(A^c) = 1 - P(A)$
- $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ for $A_i \cap A_j = \emptyset$ if $i \neq j$.

- Conditional probability, Baye's rule

2 Distribution Functions

2.1 Random Variable

- A random variable (RV) is a function $X : \Omega \to \mathbb{R}$ such that $\{\omega \in \Omega : X(\omega) \leq a\}$ is an event for each $a \in \mathbb{R}$.
- Discrete RV: the image of X is finite or countably infinite.
- Continuous RV: the image of X is uncountably infinite.

2.2 Discrete

- Cumulative Distribution Function:

$$F(x_j) = P(\{\omega \in \Omega : X(\omega) \leq x_j\}) \tag{1}$$

- Probability Density Function:

$$f(x_j) = P(\{\omega \in \Omega : X(\omega) = x_j\}) \tag{2}$$

- Interchange:

$$F(x_j) = \sum_{i \leq j} f(x_i) \tag{3}$$

$$f(x_j) = F(x_j) - F(x_{j-1}) \tag{4}$$

2.3 Continuous

- Cumulative Distribution Function:

$$F(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \tag{5}$$

- Probability Density Function:

$$f(x_j) = \lim_{\Delta x \to 0} \frac{P(\{\omega \in \Omega : x - \Delta x < X(\omega) \leq x\})}{\Delta x} \tag{6}$$

- Interchange:

$$F(x) = \int_{-\infty}^{x} f(x) dx \tag{7}$$

$$f(x) = \frac{dF}{dx} \tag{8}$$

2.4 Joint Distribution Functions

Definitions, marginal, conditional, independent, uncorrelated.

3 Moments

- Expectation $E[Q(X)]$ for discrete random variable

$$E[Q(X)] = \sum_{i \in I} Q(x_i) f(x_i) \tag{9}$$

  where $I = \{0, \pm 1, \pm 2, ...\}$ and $Q : \mathbb{R} \to \mathbb{R}$.

- Expectation $E[Q(X)]$ for continuous random variable

$$E[Q(X)] = \int_{-\infty}^{\infty} Q(x) f(x) dx \tag{10}$$

  where $Q : \mathbb{R} \to \mathbb{R}$.

- nth raw moments are defined by $E(X^n)$.
- nth central moments are defined by $E\{[X - E(X)]^n\}$
- The mean μ is given by $\mu = E(X)$.
- The variance σ^2 is given by $\sigma^2 = \mathrm{Var}(X) = E\{[X - E(X)]^2\}$.
- Some properties of variance:
  - For a being a constant

$$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X) \tag{11}$$

  - For X_1, X_2, X_3, ..., X_n being uncorrelated

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \tag{12}$$

4 Random sequences

4.1 Convergence

4.2 Law of Large Numbers

Given the sequence

$$A_n = \frac{1}{n} S_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) \tag{13}$$

where the random variables X_1, X_2, X_3, ..., are independent and identically distributed, each with mean μ, then the Law of Large Numbers states that

$$A_n \to \mu \quad \text{as} \quad n \to \infty. \tag{14}$$

4.3 Central Limit Theorem

Suppose $\{X_1, X_2, ...\}$ is a sequence of i.i.d. random variables, with $E[X_j] = \mu$ and $\text{Var}[X_j] = \sigma^2$ for all j. Then as n approaches infinity

$$Z_n = \frac{A_n - \mu}{\sigma/\sqrt{n}} \tag{15}$$

converges in distribution to a standard Gaussian random variable. That is, A_n becomes normally distributed, with mean μ and variance σ^2/n.

5 Statistical tests

5.1 Confidence Intervals

Consider an interval whose center A_n is random, and has a given deterministic width $2b$. Given a specific value for b, there is a probability that such interval will enclose an unknown but deterministic parameter μ, and is denoted by $P(A_n - b \leq \mu \leq A_n + b)$. For example, if the width of the interval is infinite, then the probability that such interval will enclose μ is 1, whereas if its length is very small, then the probability that it will enclose μ would be very low. Such intervals associated with a given probability are referred to as confidence intervals.

Consider a standard normal RV Z. We can find numbers $-z$ and z between which Z lies with probability $1 - \alpha$, that is

$$P(-z \leq Z \leq z) = 1 - \alpha. \tag{16}$$

The way we find such value z is by noting that the above requires $P(Z > z) = P(Z < -z) = \alpha/2$, because the distribution for Z is symmetric. Thus, $P(Z \leq z) = 1 - \alpha/2$. Given an α, we can use a table for a standard normal distribution to obtain the value of z. Assume A_n is normally distributed, which is a fair assumption due to the Central Limit Theorem, and has mean μ and variance σ^2/n. Thus, $(A_n - \mu)/(\sigma/\sqrt{n})$ will have a standard normal distribution as $n \to \infty$, which allows us to write

$$P(-z \leq \frac{A_n - \mu}{\sigma/\sqrt{n}} \leq z) = 1 - \alpha. \tag{17}$$

Rewriting the inequality above, we obtain

$$P(A_n - z\frac{\sigma}{\sqrt{n}} \leq \mu \leq A_n + z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha. \tag{18}$$

The above is thus the probability that the interval with center A_n and width $2z\sigma/\sqrt{n}$ will enclose μ. Such probability is equal to $1-\alpha$, and hence the interval is referred to as a $100(1-\alpha)\%$ interval.

6 Stochastic Process

Consider the discrete-time stochastic process that consists of measuring the height of a random student in a classroom every minute that passes by. The outcome w could be a vector consisting of the students that were picked, for example

$$w = \begin{pmatrix} \text{George} \\ \text{Paul} \\ \text{Monica} \end{pmatrix} \tag{19}$$

The time dependence of the stochastic process would be as follows. For $t = 1$, $X_t(1, w)$ picks the first element of w and compares against a table that tells the of height such student, in this case George. For $t = 2$, $X_t(2, w)$ picks the second element of w and compares against the table to obtain the corresponding height, in this case the height of Paul. And so on for $t = 3$. This thus shows how at each of the different times we have a different random variable, each of which acting on different elements of w. This example also shows how we could chose a different w, namely

$$w = \begin{pmatrix} \text{Hilary} \\ \text{Sam} \\ \text{John} \end{pmatrix} \tag{20}$$

to obtain a different sample path of the stochastic process.