Tutorial 1 : Probability Review

1 Theory

1.1 Basics

- Ω - Sample Space: The set of all possible outcomes in an experiment.
  When rolling a die, Ω = 1,2,3,4,5,6.
- $\mathcal{F}$ - Events Space: A set of subsets of Ω.
  When rolling a die, one option is $\mathcal{F} = \big\{\{1,3,5\}, \emptyset, \Omega, \{2,4,6\}\big\}$.
- P - Probability Measure: A function $P : \mathcal{F} \to [0,1]$ which satisfies
  1. $\forall A \in \mathcal{F} : \quad P(A) \geq 0$.
  2. $P(\Omega) = 1$.
  3. Let $A_1, A_2, ..., \in \mathcal{F}$ be mutually exclusive events (i.e., $A_i \cap A_j = \emptyset \, \forall i \neq j$) , then

$$P\Big\{\bigcup_{i=1}^{\infty} A_i\Big\} = \sum_{i=1}^{\infty} P\{A_i\}.$$

1.2 Random Variable

A random variable (RV) is a function $X : \Omega \to \mathbb{R}$ which satisfies that $\{X \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$. The cumulative distribution function (CDF) of a random variable X is defined as

$$F_X(x) \triangleq P(X \leq x).$$

The CDF $F_X(x)$ is monotonic non-decreasing, right-continuous and satisfies

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1.$$

For a discrete random variable X, the CDF can be written as

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

where $p(x) \triangleq P(X = x)$ is termed the probability mass function (PMF). For a continuous random variable, if there exists a function $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt,$$

then $f_X(x)$ is termed the probability density function (PDF) and it holds that

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a).$$

For brevity, henceforth we omit the subscript X, i.e. $F(x)$ instead $F_X(x)$ and etc.

Expectation of X:

- Discrete variable - $E[X] \triangleq \sum_i x_i p(x_i)$.
- Continuous variable - $E[X] \triangleq \int xp(x)dx$.

Expectation is a linear operation - $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$.

Variance of X:

$$Var(X) \triangleq E[(X - E(X))^2] = E[X^2] - (E[X])^2.$$

Covariance of X and Y:

$$Cov(X, Y) \triangleq E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]E[Y].$$

Notice that $Cov(X, X) = Var(X)$.

1.2.1 Gaussian Random Variable

A Gaussian random variable $X \sim \mathcal{N}\left(0, \sigma^2\right)$ is a continuous random variable which its PDF is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ - \frac{(x-\mu)^2}{2\sigma^2} \Big\},$$

where $\mu = E[X]$ and $\sigma^2 = Var(X)$. A multivariate Gaussian random variable is a vector of Gaussian random variables $\boldsymbol{X} = (x_1, x_2, ..., x_n)^T$ which its PDF is given by

$$f(x_1, x_2, ..., x_n) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\Big\{ - \frac{1}{2} (\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma^{-1}} (\boldsymbol{X} - \boldsymbol{\mu}) \Big\}$$

where $\boldsymbol{\mu} = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix}$ and $\boldsymbol{\Sigma} = E[\boldsymbol{(X - \mu)(X - \mu)}^T] = \begin{bmatrix} Var(x_1) & Cov(x_1, x_2) & \cdots & Cov(x_1, x_n) \\ Cov(x_2, x_1) & Var(x_2) & \cdots & Cov(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_n, x_1) & Cov(x_n, x_2) & \cdots & Var(x_n) \end{bmatrix}.$

The Central Limit Theorem

Let $\{x_1, x_2, ..., x_n\}$ be a sequence of independent and identically distributed (iid) random variables drawn from a distribution of expected value given by μ and finite variance given by σ^2. Define the sample average as

$$S_n \triangleq \frac{x_1 + x_2 + ... + x_n}{n}.$$

By the law of large numbers we have that

$$S_n \underset{n \to \infty}{\to} \mu.$$

The central limit theorem states that as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $N(0, \sigma^2)$:

$$\sqrt{n}(S_n - \mu) \underset{d}{\to} N(0, \sigma^2).$$

1.3 Joint, Marginal, and Conditional Probabilities

<u>**Note**</u>

For brevity, hereafter the definitions are given only for a discrete variable. For the continuous case replace sums with integrals, PMF with PDF and etc.

The joint cumulative distribution function of a pair of random variables (X, Y) is defined as

$$F(x, y) \triangleq P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j),$$

where $p(x, y) \triangleq P(X = x, Y = y)$ is the joint probability mass function.
The marginal probability mass function of X is given by

$$p(x) \triangleq P(X = x) = \sum_y p(x, y) = \sum_y P(X = x, Y = y).$$

The conditional probability mass function of X given y is defined as

$$p(x|y) = P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)} = \frac{p(x, y)}{p(y)}$$

Law of total probability

$$p(x) = \sum_y p(x,y) = \sum_y P(x|y)p(y).$$

The conditional expectation of X given Y is

$$E[X|Y=y] = \sum_i x_i p(x_i|y).$$

Law of total expectation

$$E[X] = E[E[X|Y]]$$

1.3.1 Bayes' Rule

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)},$$

where the second equality is due to the definition of conditional probability.

1.3.2 Independence and Correlation

Let X and Y be two random variable where $E[X]$, $E[Y]$ and $E[XY]$ exist and finite. X and Y are said to be **uncorrelated** if

$$Cov(X,Y) = E[XY] - E[X]E[Y] = 0.$$

In this case, $Var(X+Y) = Var(X) + Var(Y)$:

$$Var(X+Y) = Cov(X+Y, X+Y) = Cov(X,X) + \underbrace{Cov(X,Y)}_{=0} + \underbrace{Cov(Y,X)}_{=0} + Cov(Y,Y) = Var(X) + Var(Y).$$

X and Y are said to be **statistically independent** if

$$F(x,y) = F(x)F(y).$$

Which implies the following

- $p(x,y) = p(x)p(y)$.
- $p(x|y) = p(x), p(y|x) = p(y)$.
- $E[XY] = E[X]E[Y]$.
- $Cov(X,Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0 \rightarrow X$ and Y are uncorrelated.

2 Practice

Question 1

A blood test for the detection of a certain disease has a probability of 95% of giving a positive result if the patient is indeed sick, and a probability of 1% to get a positive result if the patient is healthy. It is known that the disease exists in 0.5% of the population. What is the probability that a certain patient is sick, if he got a positive result in the blood test?

Solution

Lets use Bayes' rule:

$$P(sick|positive) = \frac{P(positive|sick)P(sick)}{P(positive)} = \frac{0.95 \cdot 0.005}{0.95 \cdot 0.005 + 0.01 \cdot 0.995} = 0.3231$$

Question 2

Let $Z \sim U[0, 2\pi]$, i.e.

$$f_Z(z) = \begin{cases} \frac{1}{2\pi}, & z \in [0, 2\pi] \\ 0, & o.w. \end{cases}$$

Define $X = \cos(Z)$, $Y = \sin(Z)$.

(a) Prove that $E[XY] = 0$.

(b) Prove that $Cov(x, Y) = 0$.

(c) Show that X and Y are statistically dependent.

Solution

(a) Recall that $\cos(\theta) \cdot \sin(\theta) = \frac{1}{2}\sin(2\theta)$. Hence,

$$\begin{aligned} E[XY] &= E[\cos(Z)\sin(Z)] \\ &= E[\frac{1}{2}\sin(2Z)] \\ &= \frac{1}{2}\int_0^{2\pi} \sin(2z) \cdot \frac{1}{2\pi} dz \\ &= \frac{1}{4\pi}\int_0^{2\pi} \sin(2z) dz = 0. \end{aligned}$$

(b) Remember that $Cov(X, Y) = E[XY] - E[X]E[Y]$. Therefore, we need to compute the means of X and Y:

$$\begin{aligned} E[X] = E[\cos(Z)] &= \int_0^{2\pi} \cos(z) \cdot \frac{1}{2\pi} dz = 0 \\ E[Y] = E[\sin(Z)] &= \int_0^{2\pi} \sin(z) \cdot \frac{1}{2\pi} dz = 0. \end{aligned}$$

Overall we get that $Cov(X, Y) = 0 - 0 = 0$, i.e., X and Y are uncorrelated.

(c) Assume by contradiction that X and Y are statistically independent. Then,

$$E[X^2Y^2] = E[X^2]E[Y^2].$$

However,

$$\begin{aligned} E[X^2] &= \int_0^{2\pi} \cos^2(z) \cdot \frac{1}{2\pi} dz = \frac{1}{4\pi}\int_0^{2\pi} (1 + \cos(2z)) dz = \frac{1}{2}. \\ E[Y^2] &= \int_0^{2\pi} \sin^2(z) \cdot \frac{1}{2\pi} dz = \frac{1}{4\pi}\int_0^{2\pi} (1 - \cos(2z)) dz = \frac{1}{2}. \\ E[X^2Y^2] &= E[\cos^2(Z)\sin^2(Z)] = E\Big[\Big(\frac{\sin(2Z)}{2}\Big)^2\Big] = \frac{1}{4}E\Big[\frac{1 - cos(4Z)}{2}\Big] = \frac{1}{8}. \end{aligned}$$

Hence, $E[X^2Y^2] \neq E[X^2]E[Y^2]$.

<u>Another approach</u> is to see that $X^2 + Y^2 = 1$, hence, given $Y = y$, X could be only $\pm\sqrt{1 - y^2}$ which implies that the PDF of X is different than the conditional PDF of X given $Y = y$.

Question 3

A roulette game has 37 slots, out of which 18 are black, 18 are red, and 1 green (you cannot put money on the green slot). A gambler wins if he correctly guessed the color of the slot where the ball ends up. When a gambler wins, he gets the same amount of money as he put (any successful bet on 1$ earns the gambler an extra 1$). Assume the gambler bets on 1$ at each round.

(a) Calculate the mean and variance of the profit X (sum of winnings) after $n = 100$ bets.

(b) Calculate the probability that for $n = 100$ rounds, the gambler has positive profit.

Solution

(a) The profit is a binomial variable $X \sim B(n, p)$ where $n = 100$ and $p = \frac{18}{37} \approx 0.48$. The mean and variance are given by

$$\begin{aligned} E[X] &= np = 100 \cdot 0.48 = 48, \\ Var(X) &= np(1-p) = 100 \cdot 0.48 \cdot 0.52 = 24.96. \end{aligned}$$

(b) The profit will be positive for $X \geq 50$, the probability for that is

$$P(X \geq 50) = \sum_{k=50}^{100} \binom{100}{k} p^k (1-p)^{100-k} \approx 0.3.$$

Question 4

(a) Prove the Markov inequality: if X is a non-negative random variable with mean $E[X] = \mu$, then for all $a > 0$ it holds that:

$$P(X \geq a) \leq \frac{\mu}{a}.$$

Hint: Define another random variable Y which equals 1 if $X \geq a$ and 0 otherwise (such a variable is called an indicator). Calculate the mean of Y.

(b) Use Markov inequality to prove Chebychev inequality: if X is a random variable with mean $E[X] = \mu$ and variance $Var[X] = \sigma^2$, then for all $a > 0$ it holds that

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Solution

(a) Define the following indicator

$$Y = \begin{cases} 1, & X \geq a, \\ 0, & \text{otherwise.} \end{cases}$$

The mean of Y is $E[Y] = 0 \cdot P(X < a) + 1 \cdot P(X \geq a) = P(X \geq a)$. In addition, notice that

$$aY \leq X.$$

Since expectation is a monotonically increasing function, we can apply it on both sides of the inequality

$$\begin{aligned} & E[aY] \leq EX \\ \Rightarrow \quad & aE[Y] \leq \mu \\ \Rightarrow \quad & P(X \geq a) \leq \frac{\mu}{a}. \end{aligned}$$

(b) Define $Z = (X - \mu)^2$. Z is a non-negative random variable with mean

$$E[Z] = E[(X - \mu)^2] = \sigma^2.$$

Hence, according to Markov inequality we have that

$$\begin{aligned} & P(Z \geq a^2) \leq \frac{\sigma^2}{a^2}. \\ \Rightarrow \quad & P\Big((X-\mu)^2 \geq a^2\Big) \leq \frac{\sigma^2}{a^2}. \end{aligned}$$

Finally, we use the fact that $P\Big((X - \mu)^2 \geq a^2\Big) = P\Big(|X - \mu| \geq a\Big)$ to get Chebychev inequality

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$