

# VIP Refresher: Probabilities and Statistics

Afshine AMIDI and Shervine AMIDI

October 13, 2018

翻译: 朱小虎

## 概率和组合导引

□ **样本空间** – 一个实验的所有可能结果的集合称为实验的样本空间，记作  $S$ 。

□ **事件** – 样本空间的任何子集  $E$  被称为一个事件。即，一个事件是一个包含可能结果的集合。如果该实验的结果包含在  $E$  内，那么我们称  $E$  发生。

□ **概率论公理** – 对每个事件  $E$ ，我们记  $P(E)$  为事件  $E$  出现的概率。

$$(1) \quad 0 \leq P(E) \leq 1 \quad (2) \quad P(S) = 1 \quad (3) \quad P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

□ **置换** – 一个置换是从  $n$  个对象的池子中按照给定次序安置  $r$  个对象。这样的安置的数目由  $P(n, r)$  表示，定义为：

$$P(n, r) = \frac{n!}{(n-r)!}$$

□ **组合** – 一个组合是从  $n$  个对象的池子中无序安置  $r$  个对象。这样的安置的数目由  $C(n, r)$  表示，定义为：

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}$$

注：对  $0 \leq r \leq n$ ，我们有  $P(n, r) \geq C(n, r)$

## 条件概率

□ **贝叶斯规则** – 对事件  $A$  和  $B$  满足  $P(B) > 0$ ，我们有：

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

注：我们有  $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$

□ **分划** – 令  $\{A_i, i \in [1, n]\}$  对所有  $i$ ， $A_i \neq \emptyset$ 。我们称  $\{A_i\}$  为一个分划，当有：

$$\forall i \neq j, A_i \cap A_j = \emptyset \quad \text{et} \quad \bigcup_{i=1}^n A_i = S$$

注：对任意在样本空间中的事件  $B$  我们有  $P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$ 。

□ **贝叶斯规则的扩展形式** – 令  $\{A_i, i \in [1, n]\}$  为样本空间的一个分划，我们有：

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

□ **独立** – 两个事件  $A$  和  $B$  是独立的当且仅当我们有：

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

## 随机变量

□ **随机变量** – 一个随机变量，通常记作  $X$ ，是一个将在一个样本空间中的每个元素映射到一个实值的函数。

□ **累积分布函数 (CDF)** – 累积分布函数  $F$ ，是单调不减的，其

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

定义为：

$$F(x) = P(X \leq x)$$

注：我们有  $P(a < X \leq b) = F(b) - F(a)$ 。

□ **概率密度函数 (PDF)** – 概率密度函数  $f$  是  $X$  取值在两个相邻随机变量的实现间的概率。

□ **PDF 和 CDF 的关系** – 这里是离散和连续场景下的重要性质。

类型	CDF $F$	PDF $f$	PDF 的性质
(D)	$F(x) = \sum_{x_i \leq x} P(X = x_i)$	$f(x_j) = P(X = x_j)$	$0 \leq f(x_j) \leq 1$ and $\sum_j f(x_j) = 1$
(C)	$F(x) = \int_{-\infty}^x f(y)dy$	$f(x) = \frac{dF}{dx}$	$f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$

□ **方差** – 随机变量的方差通常记作  $\text{Var}(X)$  或者  $\sigma^2$ ，是分布函数的扩散性的一个度量函数。定义如下：

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

□ **标准差** – 随机变量的标准差，通常记作  $\sigma$ ，是分布函数扩散性的一个和实际随机变量值单位相当的度量函数。定义如下：

$$\sigma = \sqrt{\text{Var}(X)}$$

□ **分布的期望和矩** – 这里是期望值  $E[X]$ 、一般期望值  $E[g(X)]$ 、第  $k$  阶矩  $E[X^k]$  和特征函数  $\psi(\omega)$  在离散和连续场景下的表达式：

Case	$E[X]$	$E[g(X)]$	$E[X^k]$	$\psi(\omega)$
(D)	$\sum_{i=1}^n x_i f(x_i)$	$\sum_{i=1}^n g(x_i) f(x_i)$	$\sum_{i=1}^n x_i^k f(x_i)$	$\sum_{i=1}^n f(x_i) e^{i\omega x_i}$
(C)	$\int_{-\infty}^{+\infty} x f(x) dx$	$\int_{-\infty}^{+\infty} g(x) f(x) dx$	$\int_{-\infty}^{+\infty} x^k f(x) dx$	$\int_{-\infty}^{+\infty} f(x) e^{i\omega x} dx$

Remarque: on a  $e^{i\omega x} = \cos(\omega x) + i \sin(\omega x)$ .

□ **随机变量的变换** – 令变量  $X$  和  $Y$  由某个函数联系在一起。记  $f_X$  和  $f_Y$  分别为  $X$  和  $Y$  的分布函数，我们有：

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

□ **莱布尼兹积分法则** – 令  $g$  为  $x$  和  $c$  的函数， $a, b$  是可能依赖于  $c$  的边界。我们有：

$$\frac{\partial}{\partial c} \left( \int_a^b g(x) dx \right) = \frac{\partial b}{\partial c} \cdot g(b) - \frac{\partial a}{\partial c} \cdot g(a) + \int_a^b \frac{\partial g}{\partial c}(x) dx$$

□ **切比雪夫不等式** – 令  $X$  为随机变量期望值为  $\mu$ 。对  $k, \sigma > 0$ ，我们有下列不等式：

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

## 联合分布随机变量

□ **条件密度** –  $X$  关于  $Y$  的条件密度通常记作  $f_{X|Y}$ ，定义如下：

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

□ **独立性** – 两个随机变量  $X$  和  $Y$  被称为独立的当我们有：

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

□ **边缘密度和累积分布** – 从联合密度概率函数  $f_{XY}$ ，我们有：

类型	边缘密度函数	累积函数
(D)	$f_X(x_i) = \sum_j f_{XY}(x_i, y_j)$	$F_{XY}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f_{XY}(x_i, y_j)$
(C)	$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$	$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dx' dy'$

□ **协方差** – 我们定义两个随机变量  $X$  和  $Y$  的协方差，记作  $\sigma_{XY}^2$  或者更常见的  $\text{Cov}(X, Y)$ ，如下：

$$\text{Cov}(X, Y) \triangleq \sigma_{XY}^2 = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

□ **相关性** – 记  $\sigma_X, \sigma_Y$  为  $X$  和  $Y$  的标准差，我们定义随机变量  $X$  和  $Y$  的相关性，记作  $\rho_{XY}$ ，如下：

$$\rho_{XY} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}$$

注：对任何随机变量  $X, Y$ ，我们有  $\rho_{XY} \in [-1, 1]$ 。

□ **主要的分布** – 这里是主要需要记住的分布：

类型	分布	PDF	$\psi(\omega)$	$E[X]$	$\text{Var}(X)$
(D)	$X \sim \mathcal{B}(n, p)$ Binomial	$P(X = x) = \binom{n}{x} p^x q^{n-x}$ $x \in \llbracket 0, n \rrbracket$	$(pe^{i\omega} + q)^n$	$np$	$npq$
	$X \sim \text{Po}(\mu)$ Poisson	$P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$ $x \in \mathbb{N}$	$e^{\mu(e^{i\omega} - 1)}$	$\mu$	$\mu$
(C)	$X \sim \mathcal{U}(a, b)$ Uniforme	$f(x) = \frac{1}{b-a}$ $x \in [a, b]$	$\frac{e^{i\omega b} - e^{i\omega a}}{(b-a)i\omega}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
	$X \sim \mathcal{N}(\mu, \sigma)$ Gaussien	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $x \in \mathbb{R}$	$e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$	$\mu$	$\sigma^2$
	$X \sim \text{Exp}(\lambda)$ Exponentiel	$f(x) = \lambda e^{-\lambda x}$ $x \in \mathbb{R}_+$	$\frac{1}{1 - \frac{i\omega}{\lambda}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

## 参数估计

□ **随机采样** – 一个随机采样是  $n$  个和  $X$  独立同分布的随机变量  $X_1, \dots, X_n$  的集。

□ **估计器** – 估计器是一个用来推断一个统计模型中未知参数值的关于数据的函数。

□ **偏差** – 估计器  $\hat{\theta}$  的偏差定义为  $\hat{\theta}$  分布的期望值和真实值间的差距，即：

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

注：估计器被称为无偏的当我们有  $E[\hat{\theta}] = \theta$ 。

□ **中央极限定理** – 令一个随机采样  $X_1, \dots, X_n$  满足一个给定分布均值  $\mu$  方差  $\sigma^2$ ，我们有：

$$\bar{X}_{n \rightarrow +\infty} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$