
# 概率论笔记

作者：肖程哲

时间：July 11, 2022

苟日新，日日新，又日新

# 目录

# 第 3 章 随机变量的数值特征

## 3.1 期望值

**定义 3.1**

对于实值随机向量 $X:(\Omega,\mathscr{F},\mathbb{P})\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ 和 (可测) 函数 $g:\mathbb{R}^n\to\mathbb{R}$, 称

$$\mathbb{E}[g(X)]=\int_{\Omega}g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega)=\int_{\mathbb{R}}g(x)\,\mathrm{d}F_X(x)$$

为 $g(X)$ 的**期望** (expectation).

注 当 $F_X(x)$ 在 $x_0$ 出连续可导时, $\mathrm{d}F_X(x_0)=f_X(x_0)\mathrm{d}x$; 当 $x_0$ 为其间断点时时, $\mathrm{d}F_X(x_0)=p_X(x_0)\delta(x_0)\mathrm{d}x$
期望算子 $\mathbb{E}$ 是一个线性泛函, 仅适用于可积 的随机变量.

**定义 3.2**

1. 当 $g(x)=x$ 时, $\mathbb{E}[g(x)]=\mathbb{E}[X]$ 称作 $X$ 的**均值** (mean), 记为 $\mu_X$
2. 当 $g(x)=(x-\mu_X)^2$ 时, $\mathbb{E}[g(x)]=\mathbb{E}[(X-\mathbb{E}[X])^2]$ 称作 $X$ 的**方差** (variance), 记为 $\sigma_X^2$. 其平方根称作 $X$ 的**标准差** (standard deviation), 记为 $\sigma_X$
3. 当 $g(x,y)=(x-\mu_X)(y-\mu_Y)$ 时, $\mathbb{E}[g(X,Y)]=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$ 称作 $X$ 与 $Y$ 的**协方差** (covariance), 记为 $\mathrm{Cov}(X,Y)$ 或 $\sigma_{XY}$.
4. 定义 $X$ 与 $Y$ 的**相关系数** (correlation coefficient) 为:$\sigma_{XY}/(\sigma_X\sigma_Y)$, 记为 $\mathrm{Cor}(X,Y)$ 或 $\rho_{XY}$. 若 $\rho_{XY}=0$, 则称 $X$ 与 $Y$ 不相关

### 3.1.1 均值

注

- 随机变量的均值可看作其加权平均, 权重为其 pdf 或 pmf, 也即其质心. 从大数定律 (5) 的角度看, 也可解释为其长期均值.
- 方差为随机变量距其均值的均方偏差, 刻画了 $X$ 的变动程度
- 随机变量的均值与标准差的单位和其本身的相同, 方差的为其平方

**定理 3.1**

均值为随机变量的线性映射, 即:

$$\mathbb{E}(a+\sum_{i=1}^{n}b_iX_i)=a+\sum_{i=1}^{n}b_i\mathbb{E}(X_i)$$

**定理 3.2 ()**

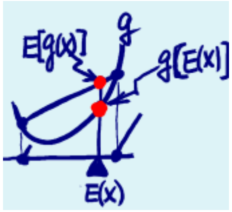
若 $X,Y$ 独立, 则

$$\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(g(X)h(Y))=\mathbb{E}(g(X))\mathbb{E}(h(Y))$$

注 由于 $\mathbb{E}(X/Y)=\mathbb{E}(X)\mathbb{E}(\frac{1}{Y})$, 而 $\mathbb{E}(\frac{1}{Y})\neq\frac{1}{\mathbb{E}(Y)}$, 所以 $\mathbb{E}(X/Y)\neq\mathbb{E}(X)/\mathbb{E}(Y)$

**定理 3.3**

若 $g$ 为下凸 (convex) 函数, 则 $\mathbb{E}[g(X)]\geq g[\mathbb{E}(X)]$; 若 $g$ 为上凸 (concave) 函数, 则 $\mathbb{E}[g(X)]\leq g[\mathbb{E}(X)]$;




一个重要结果是, 若 $g(X) \geq 0$, 则 $\mathbb{E}[g(X)] = 0 \implies g(X) \overset{\text{a.s.}}{=} 0$, 即 $\mathbb{P}\{g(X) = 0\} = 1$. 其证明可通过 **Markov 不等式**

$$\mathbb{P}\{g(X) \geq \varepsilon\} \leq \mathbb{E}[g(X)]/\varepsilon, \quad \forall \varepsilon > 0$$

完成, 其中需要用到概率的连续性, 即 $\lim\limits_{n\to\infty} A_n = A \implies \lim\limits_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

预处理随机变量有两个常用变换:

- **中心化** (centralization) $X \mapsto X - \mathbb{E}X$;
- **标准化** (standardization) $X \mapsto \dfrac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}$.

### 3.1.2 方差

**定理 3.4**

$\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mu_X)^2] = E(X^2) - \mu_X$

**定理 3.5**

$$\mathrm{Var}(a + bX) = b^2\,\mathrm{Var}(X)$$

**定理 3.6**

$$\mathrm{Var}(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} b_i^2\,\mathrm{Var}(X_i) + \mathbf{b}^{\mathrm{T}}\Sigma\mathbf{b}$$

其中 $\Sigma$ 为协方差矩阵, $\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$

**定理 3.7**

若 $X_1, \cdots, X_n$ 相互独立, 则:

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

考虑**均方误差** (mean squared error)

$$\mathrm{MSE}(X;\theta) = \mathbb{E}[|X - \theta|^2], \quad \theta \in \mathbb{R},$$

通过**方差偏差分解** (variance-bias decomposition)

$$\mathrm{MSE}(X;\theta) = \mathrm{Var}(X) + |\mathbb{E}X - \theta|^2$$

可以说明 $\theta \mapsto \mathrm{MSE}(X;\theta)$ 在 $\mathbb{E}X$ 处取到最小值 $\mathrm{Var}(X)$.

投影 (projection) 和正交分解的思想在各种内积空间中应用广泛, 这里是 $\mathbb{E} = \mathrm{proj}_{\mathbb{R}}$, 概率论中关于子事件域 $\mathscr{G}$ (随机元 $X$, resp.) 的条件期望几何直观是 $\mathrm{proj}_{\mathscr{G}}$ ($\mathrm{proj}_{\sigma(X)}$, resp.), 线性模型 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 中拟合值为 $\hat{\mathbf{y}} = \mathrm{proj}_{\mathrm{Col}(\mathbf{X})}\,\mathbf{y}$.

**定理 3.8 (Chebyshev 不等式)**

设随机变量 $X$ 的均值与方差分别为:$\mu,\sigma^2$, 则:

$$\mathbb{P}(|X-\mu|>t)\le\frac{\sigma^2}{t^2}$$

**证明** 设 $f(x)$ 为 $X$ 的概率密度函数, 令 $R=\{x:|x-\mu|>t\}$

$$\begin{aligned}\mathbb{P}(|x-\mu|>t)&=\int_R 1\cdot f(x)\mathrm{d}x\le\int_R\frac{(x-\mu)^2}{t^2}f(x)\mathrm{d}x\\&\le\int_{-\infty}^{\infty}\frac{(x-\mu)^2}{t^2}f(x)\mathrm{d}x\\&=\frac{\sigma^2}{t^2}\end{aligned}$$

**注** 若令 $t=k\sigma$, 则 $\mathbb{P}(|X-\mu|>k\sigma)\le\frac{1}{k^2}$, 即标准差可代表随机变量偏离均值的概率单位距离.

**推论 3.1**

$$\mathrm{Var}(X)=0\implies P(X=\mu)=1$$

### 3.1.3 协方差

协方差代表了 $X$ 与 $Y$ 之间的联合变化倾向, 或者说他们间的相关程度, 但其间未必有 因果关系.

**定理 3.9**

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

**定理 3.10**

$$\mathrm{Cov}(a+\sum_{i=1}^n b_iX_i,c+\sum_{j=1}^m d_jY_j)=\sum_{i=1}^n\sum_{j=1}^m b_id_j\,\mathrm{Cov}(X_i,Y_i)=\mathbf{b}^{\mathrm{T}}\Sigma\mathbf{d}$$

**定理 3.11**

独立是不相关的**充分条件**, 但不是必要条件

**定理 3.12**

$-1\le\rho_{XY}\le 1$, 当且仅当 $X$ 与 $Y$ 间为线性关系时取等号

**证明**

**定理 3.13**

平移与缩放随机变量都不影响其协方差, 即:

$$|\mathrm{Cov}(a+bX,c+dY)|=|Cor(X,Y)|$$

## 3.2 矩母函数与特征函数

## 3.3 熵与信息