概率论笔记

作者：肖程哲

时间：August 25, 2022

苟日新，日日新，又日新

目录

第 1 章 概率基础 **1**

1.1 概率空间 1

1.1.1 随机事件 1

1.1.2 概率空间 2

1.1.3 概率的性质 3

1.2 古典概型与几何概型 6

1.2.1 古典概型 6

1.2.2 几何概型 6

1.2.3 Bertrand 奇论 8

1.3 条件概率与独立 9

1.3.1 条件概率 9

1.3.2 独立性 12

第 1 章 练习 14

第 2 章 随机变量 **16**

2.1 随机变量的分布 16

2.2 多元随机变量 20

2.2.1 边际分布 20

2.2.2 条件分布 21

2.2.3 独立 22

2.3 随机变量的函数 22

2.3.1 分布函数法 23

2.3.2 Copula 24

2.3.3 概率密度函数法 24

2.3.4 矩母函数法 25

2.3.5 次序统计量 25

第 3 章 随机变量的数值特征 **27**

3.1 期望 27

3.2 均值与方差 28

3.2.1 均值 28

3.2.2 方差 29

3.2.3 协方差 29

3.3 条件期望 30

3.4 矩母函数与特征函数 32

3.4.1 矩 32

3.4.2 矩母函数 32

3.4.3 联合特征函数 33

3.4.4 特征函数 34

3.5 估计与预测 34

3.5.1 delta 法 34

3.5.2 预测 34

3.6 熵与信息 34

3.6.1 费尔希信息量 34

3.7 其他特征 34

3.7.1 相关系数 34

第 3 章 练习 34

第 4 章 常见分布 **35**

4.1 离散分布 35

4.1.1 均匀分布 35

4.1.2 伯努利分布 35

4.1.3 二项分布 36

4.1.4 几何分布 37

4.1.5 负二项分布 38

4.1.6 多项分布 39

4.1.7 泊松分布 40

4.1.8 超几何分布 42

4.2 连续分布 43

4.2.1 均匀分布 43

4.2.2 指数分布 43

4.2.3 伽马分布 45

4.2.4 贝塔分布 46

4.3 正态分布及其导出分布 48

4.3.1 正态分布 48

4.3.2 卡方分布 49

4.3.3 F 分布 50

4.3.4 t 分布 51

4.3.5 柯西分布 52

4.4 各分布间关系 54

第 5 章 概率极限 **56**

5.1 收敛 56

5.2 大数定理 58

5.3 中央极限定理 59

第 A 章 测度论基础 **61**

A.1 可测空间和可测映射 61

A.1.1 集合及其运算 61

A.1.2 集合系 62

第 B 章 组合计数 **63**

第 3 章 随机变量的数值特征

考试重点

- ❑ 重期望公式
- ❑ 重方差公式
- ❑ 切比雪夫不等式

3.1 期望

定义 3.1 (期望)

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 $g:\mathbb{R}^n\to\mathbb{R}$，若

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

存在，则称其为 $g(X)$ 的**期望** (expectation)。

注 当 $F_X(x)$ 在 x_0 初连续可导时，$\mathrm{d}F_X(x_0)=f_X(x_0)\mathrm{d}x$；当 x_0 为其间断点时时，$\mathrm{d}F_X(x_0)=p_X(x_0)\delta(x_0)\mathrm{d}x$。

命题 3.1 (期望的线性性质)

由积分的线性性质可得：均值为随机变量的线性映射，即

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

定理 3.1 (Markov 不等式)

设 $g(x)$ 为随机变量 X 取值的集合上的非负不减函数，且 $E(g(X))$ 存在，则

$$P(X>\varepsilon)\le\frac{E(g(X))}{g(\varepsilon)},\quad \forall\varepsilon>0$$

证明

$$\begin{aligned}P(X>\varepsilon)&=\int_{\varepsilon}^{+\infty}p(x)\mathrm{d}x\le\int_{\varepsilon}^{+\infty}\frac{g(X)}{g(\varepsilon)}p(x)\mathrm{d}x\\&\le\int_{-\infty}^{+\infty}\frac{g(X)}{g(\varepsilon)}p(x)\mathrm{d}x=\frac{E(g(X))}{g(\varepsilon)}\end{aligned}$$

推论 3.1

若 $g(X)\ge 0$，则

$$\mathbb{E}[g(X)]=0\Longleftrightarrow\mathbb{P}\{g(X)=0\}=1\Longleftrightarrow g(X)\overset{\text{a.s.}}{=}0$$

命题 3.2 (独立变量的期望)

若 X,Y 独立，则

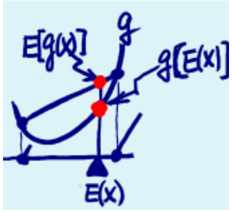
$$\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$$

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

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 $\mathbb{E}(X/Y)\neq\mathbb{E}(X)/\mathbb{E}(Y)$

命题 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 $\mathbb{E}[g(X)] \leq g[\mathbb{E}(X)]$;



3.2 均值与方差

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 μ_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 σ_X^2。其平方根称作 X 的**标准差** (standard deviation), 记为 σ_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 σ_{XY}

3.2.1 均值

随机变量的均值可看作其加权平均，权重为其 pdf 或 pmf，结果为其函数图像上的形心。从大数定律 (第??节) 的角度看，也可解释为其长期均值。

例题 3.1 在一个人数为 N 的人群中普查某种疾病，为此要抽验 N 个人的血。如果将每个人的血分别检验，则共需检验 N 次。为了能减少工作量，一位统计学家提出一种方法：按 k 个人一组进行分组，把同组 k 个人的血样混合后检验，如果这混合血样呈阴性反应，就说明此 k 个人的血都呈阴性反应，此 k 个人都无此疾病，因而这 k 个人只要检验 1 次就够了，相当于每个人检验 $1/k$ 次，检验的工作量明显减少了。如果这混合血样呈阳性反应，就说明此 k 个人中至少有一人的血呈阳性反应，则再对此 k 个人的血样分别进行检验，因而这 k 个人的血要检验 $1+k$ 次，相当于每个人检验 $1+1/k$ 次，这时增加了检验次数。假设该疾病的发病率为 p，且得此疾病相互独立。试问此种方法能否减少平均检验次数?

解 令 X 为该人群中每个人需要的验血次数，则 X 的分布列为 所以每人平均验血次数为

x	$1/k$	$1+1/k$
P	$(1-p)^k$	$1-(1-p)^k$

$$E(X) = \frac{1}{k}(1-p)^k + (1+\frac{1}{k})[1-(1-p)^k] = 1-(1-p)^k + \frac{1}{k}$$

由此可知，只要选择 k 使

$$1-(1-p)^k + \frac{1}{k} < 1,\ 即,\ (1-p)^k > \frac{1}{k}$$

就可减少验血次数，而且还可适当选择 k 使其达到最小。

3.2.2 方差

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。如果随机变量 X 的数学期望存在,其方差不一定存在；而当 X 的方差存在时，则 $E(X)$ 必定存在，其原因在于 $|x| \le x^2 + 1 (0 \le x^2 - 2|x| + 1)$ 总是成立的（若 $E(X^2)$ 收敛，则 $E(|X|)$ 收敛）。

命题 3.4

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

命题 3.5

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

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

推论 3.2

若 $X_1, \cdots, X_n$ 相互独立, 则:

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

定理 3.2 (Chebyshev 不等式)

设随机变量 X 的均值与方差分别为:μ, σ^2, 则:

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

证明 设 $F(x)$ 为 X 的分布函数, 令 $R = \{x : |x-\mu| > t\}$

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

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 3.3

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

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)$X \mapsto \dfrac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}$.

3.2.3 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 3.3

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

定理 3.4

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

定理 3.5

独立是不相关的**充分条件**, 但不是必要条件

定理 3.6

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

证明

定理 3.7

平移与缩放随机变量都不影响其协方差, 即:

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

3.3 条件期望

定义 3.3

若 $h(Y)$ 在给定 $X=x$ 下的条件分布 (定义2.12) 的数学期望存在，则定义其为**条件期望**如下:

$$E(h(Y)|X=x)=\begin{cases}\sum_y h(y)p_{Y|X}(y|x), & \text{离散情况}\\ \int_{-\infty}^{+\infty}h(y)f_{Y|X}(y|x)dy & \text{连续情况}\end{cases}$$

注 条件期望 $\mathbb{E}_{Y|X}(Y|x)$ 是关于给定变量 x 的函数，不随对应变量 Y 本身变动，但其单位与对应变量 Y 相同，可看作一条在 (X,Y) 平面的曲线。

定理 3.8

若随机变量 X,Y 独立，则:

$$\mathbb{E}_{Y|x}(Y|x)=\mathbb{E}_Y(Y)$$

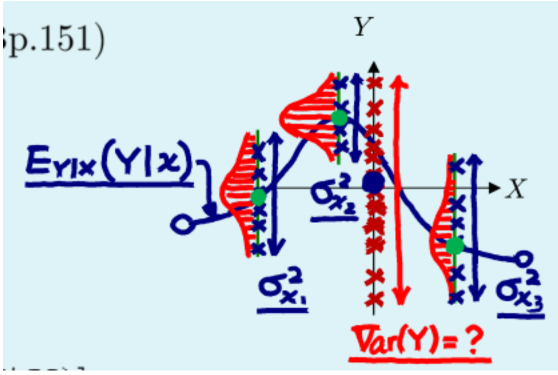
证明 由2.2可知，若随机变量 X,Y 独立，则

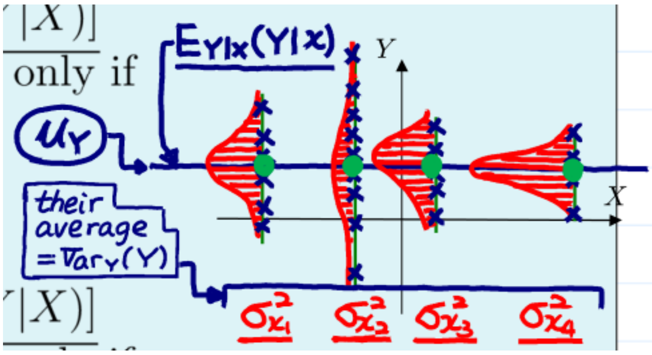
$$p_{Y|X}(y|x)=p_Y(y)$$
$$f_{Y|X}(y|x)=f_Y(y)$$

所以 $\mathbb{E}_{Y|x}(Y|x)=\mathbb{E}_Y(Y)$

由直观感受亦可知：若 X,Y 独立，则 X 不通过任何与 Y 相关的信息，其条件期望亦当与原期望相同。

笔记 令 $g(x)=\mathbb{E}(h(Y)|X=x)$，则 $g(X)$ 是随机变量 X 的变换，也是随机变量，记为 $\mathbb{E}(h(Y)|X)$





定理 3.9 (重期望公式)

$$\mathbb{E}_X[\mathbb{E}_{Y|X}(h(Y)|X)] = \mathbb{E}_Y[h(Y)]$$

证明

定理 3.10

联合期望公式：

$$E_{X,Y} = E_X E_{Y|X} = E_Y E_{X|Y}$$

泛化情况：

$$E_{X,Y}[h(X,Y)] = E_X E_{Y|X}[h(X,Y)|X] = E_Y E_{X|Y}[h(X,Y)|Y]$$

定理 3.11 (重方差公式)

随机变量 Y 的方差可作如下分解：

$$\mathrm{Var}_Y(Y) = \mathrm{Var}_X[E_{Y|X}(Y|X)] + E_X[\mathrm{Var}_{Y|X}(Y|X)]$$

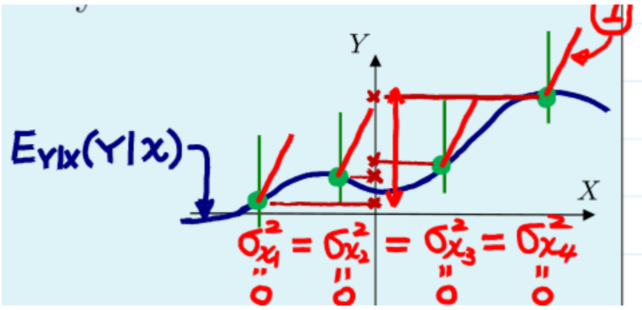
推论 3.4

$$\mathrm{Var}_Y(Y) \geq E_X[\mathrm{Var}_{Y|X}(Y|X)]$$

当且仅当 $E_{Y|X}(Y|X) = E_Y(Y)$ 时取等号

推论 3.5

$$\mathrm{Var}_Y(Y) \geq \mathrm{Var}_X[E_{Y|X}(Y|X)]$$



当且仅当 $\mathrm{Var}_{Y|X}(Y|X)=0$，即 $\mathbb{E}_{Y|X}(Y|X)=Y$ 时取等号

3.4 矩母函数与特征函数

3.4.1 矩

定义 3.4

对于随机变量 X, 定义其 k 阶**矩** (moment) 为 $E(X^k)$, 记为 μ_k; 定义其 k 阶**中心矩** (central moment) 为 $E((X-\mu_X)^k)$, 记为 υ_k;

易知矩与中心矩间存在以下关系:

$$\upsilon_k=\sum_{i=0}^{k}\binom{k}{i}\mu_i(-\mu_X)^{n-i}$$

$$\mu_k=\sum_{i=0}^{k}\binom{k}{i}\upsilon_i(\mu_X)^{n-i}$$

特别的有:

$$E(X)=\mu_1$$

$$\mathrm{Var}(X)=\upsilon_2=\mu_1^2-2\mu_1\cdot\mu_1+\mu_2=\mu_2-\mu_1^2$$

3.4.2 矩母函数

定义 3.5

对于随机变量 X, 若下式期望存在:

$$M_X(t)=\mathbb{E}(e^{tX})$$

则称其为**矩母函数** (moment generating function, mgf)。

注 此表达式等价于对概率质量函数或密度函数作 Laplace 变换, 当 t 取某些特定值时, 可能不存在. (若 $t=0$ 则永远存在)

定理 3.12

若当 t 属于一个包含零点的开区间时, 矩母函数一直存在, 则其唯一对应一个概率分布.

定理 3.13

若当 t 属于一个包含零点的开区间时，矩母函数一直存在，则：

$$M_X^{(k)}(0) = E(X^k)$$

注 借此可方便地计算各阶矩，故称为矩母函数. 反过来，若已知各阶矩，通过 Tayler 展开 $M_X(t) = \sum_{k=0}^{\infty} \frac{M_X^{(k)}(0)}{k!} t^k$ 可还原矩母函数，进而得出概率分布.

命题 3.7

若 a, b 为常数，则

$$M_{a+bX}(t) = e^{at} M_X(bt)$$

定理 3.14

若 X, Y 独立，则

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

泛化情况：若 $X_1, \cdots, X_n$ 相互独立，则

$$M_{X_1+\cdots+X_n} = \prod_{i=1}^{n} M_{X_i}(t)$$

3.4.3 联合特征函数

定义 3.6

对于随机变量 $X_1, \cdots, X_n$，若下式期望存在：

$$M_{X_1 \cdots X_n}(t_1, \cdots, t_n) = \mathbb{E}(e^{t_1 X_1 + \cdots + t_n X_n})$$

则称其为**联合矩母函数** (joint moment generating function, joint mgf)。

注 此处为多元 函数

命题 3.8

$$M_{X_i}(t_i) = M_{X_1 \cdots X_n}(0, \cdots, t_i, \cdots, 0)$$

定理 3.15

当且仅当：

$$M_{X_1 \cdots X_n}(t_1, \cdots, t_n) = \prod_{i=1}^{n} M_{X_i}(t_i)$$

时，$X_1, \cdots, X_n$ 相互独立

注 与累计函数、密度函数、质量函数 的情况类似，变量相互独立等价于联合函数可拆分为边缘函数的乘积

定理 3.16

$$\frac{\partial^{r_1+\cdots+r_n}}{\partial t_1^{r_1} \cdots \partial t_n^{r_n}} M_{X_1 \cdots X_n}(t_1, \cdots, t_n) = E(X_1^{r_1} \cdots X_n^{r_n})$$

3.4.4 特征函数

由于有时矩母函数可能不存在，为避免此缺陷，构造出与之特性类似的特征函数。

定义 3.7

对于随机变量 X, 定义其**特征函数** (characat function, chf) 为：

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))$$

对于随机变量 $X_1, \cdots, X_n$, 定义其**联合特征函数** (joint characat function, joint chf) 为：

$$\phi_{X_1\cdots X_n}(t_1, \cdots, t_n) = \mathbb{E}(e^{i(t_1X_1+\cdots+t_nX_n)})$$

注 此表达式等价于对概率质量函数或密度函数作 Fourier 变换

命题 3.9

随机变量的特征函数总是存在

命题 3.10

若矩母函数存在，则其与特征函数之间满足关系：

$$\phi_X(t) = M_X(it)$$

定理 3.17

特征函数可通过以下逆变换得到分布：

离散	$p_X(x) = \lim_{T\to\infty} \int_{-T}^{T} e^{-itx}\phi_X(t)dt$
连续	$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_X(t)dt$

3.5 估计与预测

3.5.1 delta 法

3.5.2 预测

3.6 熵与信息

3.6.1 费尔希信息量

3.7 其他特征

3.7.1 相关系数

定义 3.8

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 ρ_{XY}. 若 $\rho_{XY} = 0$, 则称 X 与 Y 不相关

错题记录

1. （茆 2.2.7）对一批产品进行检查，如查到第 a 件全为合格品，就认为这批产品合格；若在前 a 件中发现不合格品即停止检查，且认为这批产品不合格。设产品的数量很大，可认为每次查到不合格品的概率都是 p。问每批产品平均要查多少件?
2. （茆 2.2.17）设随机变量 X 的概率密度函数为
$$p(x)=\begin{cases}\frac{1}{2}\cos\frac{x}{2}, & 0\le x\le\pi;\\ 0, & \text{其他}.\end{cases}$$
对 X 独立重复观察 4 次，Y 表示观察值大于 $\pi/3$ 的次数，求 Y^2 的数学期望。
3. （茆 2.2.19 作为结论记住，22 题是它的应用，思考一下什么场合下用这个公式解题比较方便）设 X 为仅取非负整数的离散随机变量, 若其数学期望存在, 证明
$$E(X)=\sum_{k=1}^{+\infty}P(X\ge k)$$
4. （茆 2.2.20）设连续随机变量 X 的分布函数为 $F(x)$，且数学期望存在，证明
$$E(X)=\int_0^{+\infty}[1-F(x)]\,\mathrm{d}x-\int_{-\infty}^{0}F(x)\,\mathrm{d}x$$
5. （茆 2.2.21）设 X 是非负连续随机变量，若 $E(X^n)$ 存在，证明：
   (a). $E(X)=\int_0^{\infty}P(X>x)\mathrm{d}x$
   (b). $E(X^n)=\int_0^{\infty}nx^{n-1}P(X>x)\mathrm{d}x$
6. （茆 2.2.22）甲、乙两人进行象棋比赛，每局甲胜的概率为 p，乙胜的概率为 $q=1-p$。比赛进行到有一人连胜两局为止，求平均比赛局数。
7. 设随机变量 X 的分布函数为
$$F(x)=\begin{cases}0, & x<-1\\ \frac{1}{2}+\frac{1}{\pi}\arcsin x, & -1\le x\le 1\\ 1, & x\ge 1\end{cases}$$
试求 $E(X)$。