概率论与数理逻辑笔记整理V1.0

陈鸿峥

2018.10 *

目录

1 基本概念

1.1 事件与概率

命题 1. 事件的基本运算

1. 分配律：$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. 德摩根律：$\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$

注意概率是对一个集合的函数，有如下定义.

定义 1 (概率). 随机试验E的所有可能结果构成E的样本空间Ω，Ω的子集称为事件，Ω的幂集构成E的事件空间$\mathcal{F}$，记概率函数$\mathbb{P}(\cdot) : \mathcal{F} \mapsto \mathbb{R}$满足：

*Build 20181017

1. 非负性：$\mathbb{P}(A) \geq 0, \forall A \in \mathcal{F}$

2. 规范性：$\mathbb{P}(\Omega) = 1$

3. 可列可加性：$A_1, A_2, \ldots$为两两不相容的事件，$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

由定义可得一些基本性质：

1. $\mathbb{P}(\varnothing) = 0, \mathbb{P}(A) \leq 1$
2. 有限可加性：$A_1, A_2, \ldots$两两不相容，$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$
3. 若$A \subset B$，则$\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A), \mathbb{P}(B) \geq \mathbb{P}(A)$
4. 逆事件概率：$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
5. 容斥原理：$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{\substack{I \subset \{1,\ldots,n\} \\ |I|=k}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)\right)$

定理 1 (实际推断原理)**.** 概率很小的事件在一次试验中实际上几乎是不发生的

1.2 条件概率

定义 2 (条件概率)**.** 设A, B为两个事件，且$\mathbb{P}(A) > 0$，则称

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)}$$

为在事件A发生的条件下事件B发生的条件概率

定理 2 (乘法公式)**.** 设$A_1, A_2, \ldots, A_n$为n个事件，$n \geq 2$，且$\mathbb{P}(A_1 A_2 \cdots A_{n-1}) > 0$，则

$$\mathbb{P}(A_1 A_2 \cdots A_{n-1}) = \mathbb{P}(A_n \mid A_1 A_2 \cdots A_{n-1}) \mathbb{P}(A_{n-1} \mid A_1 A_2 \cdots A_{n-2}) \cdots \mathbb{P}(A_2 \mid A_1) \mathbb{P}(A_1)$$

定义 3 (划分)**.** 两两交为空，所有并为全集

定理 3 (全概率公式)**.** 设试验E的样本空间为S，A为E的事件，$B_1, B_2, \ldots, B_n$为S的一个划分，且$\mathbb{P}(B_i) > 0$，则

$$\mathbb{P}(A) = \mathbb{P}(AB_1) + \mathbb{P}(AB_2) + \cdots + \mathbb{P}(AB_n) = \mathbb{P}(A \mid B_1) \mathbb{P}(B_1) + \mathbb{P}(A \mid B_2) + \cdots + \mathbb{P}(A \mid B_n) \mathbb{P}(B_n)$$

定理 4 (贝叶斯(Bayes)公式)**.** 设试验E的样本空间为S，A为E的事件，$B_1, B_2, \ldots, B_n$为S的一个划分，且$\mathbb{P}(A) > 0, \mathbb{P}(B_i) > 0$，则

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(B_i A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_i) \mathbb{P}(B_i)}{\sum_{i=1}^{n} P(A \mid B_i) \mathbb{P}(B_i)}$$

特别地，当$n=2$时有

$$\mathbb{P}\left(B \mid A\right)=\frac{\mathbb{P}\left(A \mid B\right)\mathbb{P}\left(B\right)}{\mathbb{P}\left(A \mid B\right)\mathbb{P}\left(B\right)+\mathbb{P}\left(A \mid \overline{B}\right)\mathbb{P}\left(\overline{B}\right)}$$

定义 4 (独立性)**.** 对于事件$A_1,\ldots,A_n$，

- 若$\mathbb{P}\left(A_i\cap A_j\right)=\mathbb{P}\left(A_i\right)\mathbb{P}\left(A_j\right),\forall i,j$，则称$\{A_1,\ldots,A_n\}$两两(pairwise)独立
- 若$\mathbb{P}\left(\bigcap_{j\in I}A_j\right)=\prod_{j\in I}\mathbb{P}\left(A_j\right),\forall I\in 2^{[n]}$，其中$2^{[n]}$为$\{A_i\}_{i=1}^n$的所有子集，则称$\{A_1,\ldots,A_n\}$相互(mutually)独立

区分以下两个概念

1. A,B对立(exclusive) $\Leftrightarrow \mathbb{P}(A\cap B)=0$，即不相交(disjoint)
2. A,B独立(independent) $\Leftrightarrow \mathbb{P}(A\cap B)=\mathbb{P}(A)\cdot\mathbb{P}(B)$，即不相关(unrelated)

2 随机变量及其分布

定义 5. *对于离散随机变量X，其概率质量函数(PMF)为$f_X(k)=\mathbb{P}\left(X=k\right)$，分布函数为$F_X(x)=\mathbb{P}\left(X\le x\right)=\sum_{k\le x}f_X(k)$*

定义 6. *对于连续随机变量X，其累积密度函数(CDF)为$F_X(x)=\mathbb{P}\left(X\le x\right)=\int_{-\infty}^{x}f_X(z)\ \mathrm{d}z\int_{-\infty}^{x}f(t)\ \mathrm{d}x$，其中$f_X(x)$为$X$的概率密度函数(PDF)，也即$f_X(x)=\frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$. 一定要注意，$f_X(x)\ne\mathbb{P}\left(X=x\right)$!*

注意积分区间!

定义 7 (期望)**.**

$$\begin{aligned}\mathbb{E}\left(X\right)&=\sum_x xf_X(x) &\quad \mathbb{E}\left(g(X)\right)&=\sum_x g(x)f_X(x)\\ \mathbb{E}\left(X\right)&=\int_{-\infty}^{+\infty}xf_X(x)\ \mathrm{d}x &\quad \mathbb{E}\left(g(X)\right)&=\int_{-\infty}^{+\infty}g(x)f_X(x)\ \mathrm{d}x\end{aligned}$$

期望具有线性性，即

$$\mathbb{E}\left(X+Y\right)=\mathbb{E}\left(X\right)+\mathbb{E}\left(Y\right),\mathbb{E}\left(cX\right)=c\mathbb{E}\left(X\right)$$

定义 8 (方差)**.**

$$\mathrm{Var}\left(X\right)=\sigma^2=\mathbb{E}\left(\left(X-\mathbb{E}\left(X\right)\right)^2\right)=\mathbb{E}\left(X^2\right)-\mathbb{E}\left(X\right)^2\ge 0$$

由方差定义和期望的线性性有

$$\mathrm{Var}\left(aX+b\right)=a^2\mathrm{Var}\left(X\right)$$

注意方差并不是线性的

$$\begin{aligned}\operatorname{Var}(X+Y) &= \mathbb{E}\left((X+Y)^2\right) - (\mathbb{E}(X)+\mathbb{E}(Y))^2 \\ &= \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 + \mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2 + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)\end{aligned}$$

若X, Y独立，则$\operatorname{Cov}(X,Y)=0$

定理 5. 若$X_1,\ldots,X_n$独立，则

$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i)$$

$$\operatorname{Var}\left(\prod_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i)$$

$$\operatorname{Cov}(X_i, X_j) = 0,\ i \neq j$$

$Y=g(X)$的随机变量分布一般通过求分布函数后求导得到其概率密度函数

2.1 常见的离散分布

1. 伯努利分布 **Bernoulli(p)**（二项分布的特殊情形）

$$f_X(k) = \begin{cases} 1-p & k=0 \\ p & k=1 \end{cases}$$

$$\mathbb{E}(X) = p$$

2. 二项分布 **Binomial(n,p)**

$$f_X(k) = \binom{n}{k} p^k (1-p)^n - k,\ 0 \le k \le n$$

$$\mathbb{E}(X) = n \cdot p$$

e.g. 扔n次硬币扔到k次正面（做实验n次，记录成功的次数）

3. 几何分布 **Geometric(p)**（负二项分布的特殊情形）

$$f_X(k) = (1-p)^k \cdot p, k \ge 0$$

$$\mathbb{E}(X) = \frac{1-p}{p}$$

e.g. 扔k次反面直至扔到正面（做实验直到你成功，记录失败的次数）

4. 负二项分布 **NegetiveBinomial(t,p)**

$$
\begin{aligned}
f_X(k) &= \binom{k+t-1}{t-1} p^t (1-p)^k,\ k \geq 0 \\
\mathbb{E}(X) &= t \cdot \frac{1-p}{p}
\end{aligned}
$$

e.g. 扔k次反面直到有t个正面（做实验直到你获得t次成功，记录失败次数）

5. 超几何分布 **HyperGeometric(N,n,M)**

$$
\begin{aligned}
f_X(k) &= \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} \\
\mathbb{E}(X) &= n \frac{M}{N}
\end{aligned}
$$

e.g. M个产品中有N个次品，检查n次得到k个次品

6. 泊松分布 **Poisson(λ),**$\lambda > 0$

$$
\begin{aligned}
f_X(k) &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\
\mathbb{E}(X) &= \lambda
\end{aligned}
$$

$X \sim B(n,p)$，若$p = \dfrac{\lambda}{n}$，且n非常大，则

$$
\begin{aligned}
\mathbb{P}\left(X = k\right) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 + \frac{-\lambda}{n}\right)^{n-k} \\
&\approx \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}
\end{aligned}
$$

注意泊松分布具有无记忆性(memoryless)，即

$$
\mathbb{P}\left(X \geq a \mid X \geq b\right) = \mathbb{P}\left(X \geq a - b\right)
$$

2.2 常见的连续分布

1. 均匀分布 **Uniform(a,b),**$a < b$

$$
\begin{aligned}
f_X(x) &= \frac{1}{b-a},\ a \leq x \leq b \\
\mathbb{E}\left(X\right) &= \frac{a+b}{2} \\
\mathrm{Var}\left(X\right) &= \frac{(b-a)^2}{12}
\end{aligned}
$$

2. 指数分布 **Exponential(λ),**$\lambda > 0$

$$f_X(x) = \lambda \mathrm{e}^{-\lambda x},\ x \geq 0$$
$$\mathbb{E}\left(X\right) = \frac{1}{\lambda}$$

与泊松分布类似，同样具有无记忆性

3. 正态分布 **Normal(**μ, σ^2**)**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{-(x-\mu)^2}{2\sigma^2}}$$
$$\mathbb{E}\left(X\right) = \mu$$

算平方

$$\begin{aligned}
&\left(\int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-x^2}{2}}\,\mathrm{d}x\right)^2 \\
=& \int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-x^2}{2}}\,\mathrm{d}x \int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-x^2}{2}}\,\mathrm{d}y \\
=& \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-(x^2+y^2)}{2}}\,\mathrm{d}x\,\mathrm{d}y \\
=& \int_{0}^{2\pi}\int_{0}^{+\infty} r\mathrm{e}^{\frac{-r^2}{2}}\ \mathrm{d}r\ \mathrm{d}\theta \\
=& \int_{0}^{2\pi}\int_{0}^{+\infty} \mathrm{e}^{-u}\ \mathrm{d}u\ \mathrm{d}\theta \\
=& \int_{0}^{2\pi} 1\ \mathrm{d}\theta \\
=& 2\pi
\end{aligned}$$

若$X \sim N(\mu, \sigma^2)$，则$Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$标准正态分布

4. χ^2分布

3 多维随机变量及其分布

3.1 边缘分布

定义 9. *二维随机变量*(X, Y)*的分布函数/联合(joint)分布函数定义如下*

$$F(x,y) = \mathbb{P}\left(X \leq x, Y \leq y\right) = \sum_{x_i \leq x}\sum_{y_i \leq y} p_{ij} = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\ \mathrm{d}u\ \mathrm{d}v$$

其中，$f(x,y)$*为*X, Y*的联合密度函数*

进而，对于离散型随机变量变量有，

$$\mathbb{P}\left(x_1 < X \leq x_2, y_1 < Y \leq y_2\right) = F(x_2,y_2) - F(x_2,y_1) + F(x_1,y_1) - F(x_1,y_2)$$

离散型随机变量

$$\mathbb{P}((X,Y)\in G)=\iint_G f(x,y)\ \mathrm{d}x\ \mathrm{d}y$$

若$f(x,y)$在(x,y)连续，则$\dfrac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)$

定义 10 (边缘(marginal)分布)**.**

$$\begin{aligned}F_X(x)&=\mathbb{P}(X\le x,Y<\infty)=\lim_{y\to\infty}F(x,y)=F(x,\infty)\\&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}f(x,y)\ \mathrm{d}y\right]\mathrm{d}x\\&=\int_{-\infty}^{\infty}f_X(x)\ \mathrm{d}x\end{aligned}$$

注意积分区间不一定是从负无穷到正无穷，而是概率有$(0,1)$之间值的区域

定义 11 (条件概率密度与分布函数)**.**

$$\begin{aligned}f_{X|Y}(x\mid y)&=\frac{f(x,y)}{f_Y(y)}\\F_{X|Y}(x\mid y)&=\mathbb{P}(X\le x\mid Y=y)=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\ \mathrm{d}x\end{aligned}$$

定义 12 (相互独立)**.**

$$F(x,y)=F_X(x)F_Y(y)$$
$$f(x,y)=f_X(x)f_Y(y)$$

3.2 随机变量的函数的分布

$$f_{X+Y}(z)=\int_{-\infty}^{\infty}f(z-y,y)\ \mathrm{d}y=\int_{-\infty}^{\infty}f(x,z-x)\ \mathrm{d}x$$

$$f_{Y/X}(z)=\int_{-\infty}^{\infty}|x|f(x,xz)\ \mathrm{d}x=f_{XY}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f(x,\frac{z}{x})\,\mathrm{d}x$$

$$F_{\max}(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z)$$

$$F_{\min}(z)=1-[1-F_{X_1}(z)][1-F_{X_2}(z)]\cdots[1-F_{X_n}(z)]$$

4 大数定律

定理 6 (切比雪夫(Chebyshev)不等式)**.**

$$\mathbb{P}(|X-\mathbb{E}(X)|\ge\lambda)\le\frac{\mathbb{E}(|X-\mathbb{E}(X)|)}{\lambda}$$

定理 7 (弱大数定律). 若$X_1, X_2, ...$为独立随机变量且同等分布，$\mathbb{E}(X_i) = \mu, \mathrm{Var}(X_i) = \sigma^2$，则

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) = 0, \forall \varepsilon > 0$$

定理 8 (强大数定律). 若$X_1, X_2, ...$为独立随机变量且同等分布，$\mathbb{E}(X_i) = \mu, \mathrm{Var}(X_i) = \sigma^2$，则

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1$$

5 统计

定义 13 (估计). $X_1, X_2, \ldots, X_n$为独立随机变量，从有参数$\mu, \sigma, \theta, \ldots$的分布$f$中得到，对参数$\theta$的估计是函数$T(X_1, \ldots, X_n)$，称$T$是期望*(expected)*估计，若

$$\mathbb{E}(T(X_1, \ldots, X_n)) = \theta$$

合适*(probable)*的估计，若

$$\mathbb{P}(|T(X_1, \ldots, X_n) - \theta| > \varepsilon) \to 0,\ n \to \infty$$