Review of Probability Theory

1. 随机试验

一个**随机试验**（或称随机模型）由下面三个要素组成：

- 样本空间（sample space）$\Omega \triangleq \{$所有可能的试验结果$\}$；
- 事件集（set of events）$\mathcal{F}$，满足：
  (1) $\Omega \in \mathcal{F}$ (必然事件)
  (2) $A \in \mathcal{F} \ \Rightarrow \ \bar{A} \in \mathcal{F}$ (补运算封闭)
  (3) $A_i \in \mathcal{F} \ \Rightarrow \ \bigcup_i A_i \in \mathcal{F}$ (可列并运算封闭)
  由此推出不可能事件$\varnothing \in \mathcal{F}$，可列交运算封闭。
  "加、减、乘"运算封闭。
- 概率是一个映射$P : \mathcal{F} \to (-\infty, \infty)$，满足：
  (1) $P(A) \geqslant 0$ (非负性)
  (2) $P(\Omega) = 1$ (规范性)
  (3) $P(\biguplus_i A_i) = \sum_i P(A_i)$ (可列可加性)
  由(2)(3)可以推出$P(\varnothing) = 0$。
  P类似 归一化的"计数、长度、面积、体积"等测度。

1. 随机试验

例子：抛一枚硬币，根据实际情况，可能选择如下模型。

模型1

1. 样本空间：{正,反}
2. 事件集：$\{\varnothing,\{正\},\{反\},\{正,反\}\}$
   对有限样本空间而言，事件集通常取其所有子集。
3. 概率：

$$
\begin{aligned}
P(\{正\}) &= 1/2 \\
P(\{反\}) &= 1/2
\end{aligned}
$$

1. 随机试验

例子：抛一枚硬币，根据实际情况，可能选择如下模型。

模型1

1. 样本空间：{正,反}
2. 事件集：{∅,{正},{反},{正,反}}
   对有限样本空间而言，事件集通常取其所有子集。
3. 概率：

$$\begin{aligned} P(\{\text{正}\}) &= 1/2 \\ P(\{\text{反}\}) &= 1/2 \end{aligned}$$

模型2

1. 样本空间、事件集同模型1。
2. 概率：

$$\begin{aligned} P(\{\text{正}\}) &= 1/4 \\ P(\{\text{反}\}) &= 3/4 \end{aligned}$$

1. 随机试验

例子：抛一枚硬币，根据实际情况，可能选择如下模型。

模型1

1. 样本空间：{正,反}
2. 事件集：{∅,{正},{反},{正,反}}
   对有限样本空间而言，事件集通常取其所有子集。
3. 概率：

$$\begin{aligned} P(\{\text{正}\}) &= 1/2 \\ P(\{\text{反}\}) &= 1/2 \end{aligned}$$

模型2

1. 样本空间、事件集同模型1。
2. 概率：

$$\begin{aligned} P(\{\text{正}\}) &= 1/4 \\ P(\{\text{反}\}) &= 3/4 \end{aligned}$$

模型3

1. 样本空间：{正,反,立}
2. 事件集、概率略去。

2. 概率运算法则

- 若$P(B) > 0$，条件概率$P(A|B) \triangleq P(AB)/P(B)$；
- $P(A \cup B) = P(A) + P(B) - P(AB)$；
  若A与B不交（**互不相容**），则$P(A \cup B) = P(A) + P(B)$；
- $P(AB) = P(A)P(B|A) = P(B)P(A|B)$；
  若A与B**相互独立**，则$P(AB) = P(A)P(B)$；
  不相容，计算并事件的概率简单；独立，计算交事件的概率简单。
- **全概率公式**：若B_i是样本空间Ω的划分，则
  有$P(A) = \sum_i P(B_i)P(A|B_i)$；“分类，分步”
- **贝叶斯公式**：$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}$；
- 和-积算法：加法对应分类，乘法对应分步；
- 计算概率时，建议理清概率模型并画出**概率树**。

3. 随机变量

Definition 1

随机变量X不是（自）变量，而是样本空间Ω到实数集$\mathbb{R}$的映射，即$X:\Omega\to\mathbb{R}$。

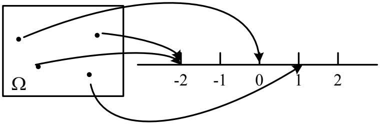


Figure: 随机变量

例子：抛一枚硬币，前面模型1。

$$\begin{aligned} \text{正} &\mapsto 1 \\ \text{反} &\mapsto 0 \end{aligned}.$$

3. 随机变量

- 离散型随机变量X：概率质量函数（pmf）$P_X(x), x \in \mathcal{X}$（$\mathbb{R}$的有限子集或可列子集）满足$P_X(x_i) \geqslant 0$ 且$\sum_i P_X(x_i) = 1$。

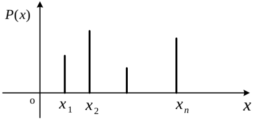


Figure: 离散型随机变量的pmf (又称分布律), "谱线"。

- 连续型随机变量X：概率密度函数（pdf）$f_X(x)$，$x \in \mathcal{X}$（$\mathbb{R}$上的区间$\mathcal{X}$）满足$f_X(x) \geqslant 0$ 和$\int_{\mathbb{R}} f_X(x)\mathrm{d}x = 1$。

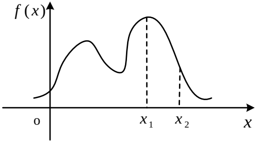


Figure: 连续型随机变量的pdf, "谱密度"。

3. 随机变量

一般随机变量X可由累积分布函数（cdf）$F_X(x) \triangleq P(X \leqslant x), x \in \mathbb{R}$来刻画。

1. 对离散型随机变量X，$F_X(x) = \sum_{x_i \leqslant x} P_X(x_i)$；
2. 对连续型随机变量X，$F_X(x) = \int_{y \leqslant x} f_X(y)\mathrm{d}y$；
3. $F_X(x)$是非递减函数，且
$$\lim_{x\to-\infty} F_X(x) = 0, \qquad \lim_{x\to\infty} F_X(x) = 1$$
4. $F_X(x)$的断点至多可列个；若在x_1处有跳跃，则跳跃的高度恰好是概率$P_X(x_1)$。

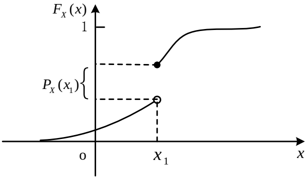
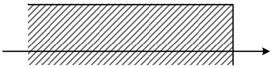


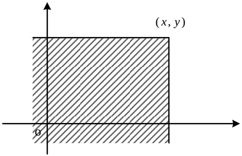
Figure: 随机变量的cdf。

3. 随机变量

1. 复随机变量：$X: \Omega \to \mathbb{C}$.
2. 随机向量：$X: \Omega \to \mathbb{R}^n$.
3. 随机序列：$X: \Omega \to \mathbb{R}^{\mathbb{Z}}$（离散时间随机过程）.
4. 随机过程：$X: \Omega \to \mathbb{R}^{\mathbb{R}} = \{f(t), t \in \mathbb{R}\}$ （函数集）.

一般地，多维随机变量（包括随机序列、随机过程）可由联合分布 $F_{X^n}(\underline{x}) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n)$来刻画。独立同分布（i.i.d.）的随机序列（包括平稳“白”过程）可用一个时刻的分布来描述。





3. 随机变量

二维离散随机变量(X, Y)

- 联合分布律$P_{X,Y}(x, y), x \in \mathcal{X}, y \in \mathcal{Y}$;
- 条件分布律$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$, $P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$;
- 边缘分布律$P_X(x) = \sum\limits_y P_{X,Y}(x, y)$, $P_Y(y) = \sum\limits_x P_{X,Y}(x, y)$。

4. 数字特征

Definition 2

设离散型随机变量X的概率质量函数为$P_X(x_k) \triangleq p_k, k = 1, 2, \cdots$。若级数$\sum_{k=1}^{\infty} x_k p_k$绝对收敛，则$X$的**数学期望**定义为：

$$E(X) = \sum_{k=1}^{\infty} x_k p_k.$$

设连续型随机变量的概率密度函数为$f(x)$,若积分$\int_{-\infty}^{\infty} x f(x) \mathrm{d}x$绝对收敛，则$X$的**数学期望**定义为：

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x.$$

思考问题：为什么要求绝对收敛？

4. 数字特征

Definition 3

一般随机变量X的**数学期望**可以表示为以下的Stieltjes积分：

$$E(x) = \int_{-\infty}^{\infty} x \mathrm{d} F_X(x)$$

其中，$F_X(x)$为X的累积分布函数。

Definition 4

设X是一个随机变量，如果$D(X) \triangleq E\left([X - E(X)]^2\right)$存在，则称其为$X$的**方差**。

- $E(X + Y) = E(X) + E(Y)$(不管X与Y是否独立)；
- 若X与Y独立，则$E(XY) = E(X)E(Y)$；
- 若X与Y独立，$D(X + Y) = D(X) + D(Y)$。

5. 随机变量的函数

设X是离散型随机变量，具有概率质量函数$P_X(x)$，而$g(x)$是一个普通函数，则$Y=g(X)$也是一个随机变量，即

$$Y:\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$$

例子：

X	-1	0	1	2
$P_X(x)$	0.2	0.3	0.2	0.3

$Y=X^2$	0	1	4
$P_Y(y)$	0.3	0.4	0.3

$Z=P_X(X)$	0.2	0.3
$P_Z(z)$	0.4	0.6

$L=-\log P_X(X)$	2.32	1.74
$P_L(l)$	0.4	0.6

1. 随机变量"被函数"之后，"谱线"的数目可能减少，高度可能增加。
2. 若$Y=g(X)$，则$E(Y)=E(g(X))$（这个等式的含义是什么？注意不是简单的"等量代入"。）

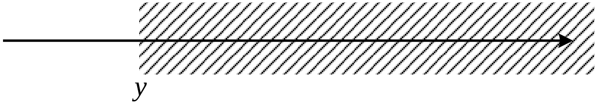
6. 尾部概率

- Markov不等式：线性递减至0
  设非负随机变量X的期望$E(X)$的值有限，则

$$P(X \geqslant y) \leqslant \frac{E(X)}{y} \tag{1}$$

证明：

$$E(X) = \int_0^{+\infty} x dF_X \geqslant \int_y^{+\infty} x dF_X \geqslant y \int_y^{+\infty} dF_X = yP(X \geqslant y)$$



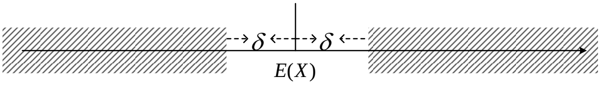
6. 尾部概率

- Chebyshev不等式：平方递减至0

$$P(|X-E(X)|\geqslant\delta)\leqslant\frac{\sigma_X^2}{\delta^2}$$

其中σ_X^2表示X的方差。
证明：对$(X-E(X))^2$应用Markov不等式，

$$\begin{aligned}
& P(|X-E(X)|\geqslant\delta)\\
= \; & P(|X-E(X)|^2\geqslant\delta^2)\\
\leqslant \; & \frac{E(X-E(X))^2}{\delta^2}=\frac{\sigma_X^2}{\delta^2}
\end{aligned}$$



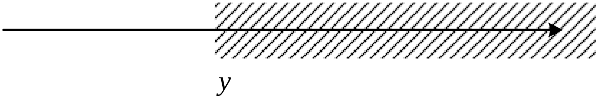
6. 尾部概率

- Chernoff界：指数递减至0

$$P(X \geqslant y) \leqslant E(e^{sX})e^{-sy} \quad 对所有\ s \geqslant 0$$

证明： 对e^{sX}应用Markov不等式，

$$P(X \geqslant y) = P\left(e^{sX} \geqslant e^{sy}\right) \leqslant \frac{E(e^{sX})}{e^{sy}}, \quad s \geqslant 0.$$



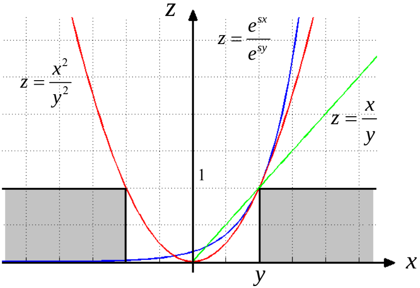
类似地，$P(X \leqslant y) \leqslant E(e^{sX})e^{-sy}$ 对所有 $s \leqslant 0$。

思考题：如何给出$P(X \geqslant x, Y \leqslant y)$的一个指数型上界？

6. 尾部概率

定义示性函数（indicator function）

$$Z(\omega) = \begin{cases} 1, & X(\omega) \geqslant y \ (y > 0) \\ 0, & \text{其他} \end{cases}$$



又$\Pr(X \geqslant y) = \Pr(Z = 1) = E(Z)$，可得

$$Z \leqslant \frac{X}{y} \Rightarrow P(X \geqslant y) = E(Z) \leqslant \frac{E(X)}{y}$$

$$Z \leqslant \frac{X^2}{y^2} \Rightarrow P(|X| \geqslant y) = E(Z) \leqslant \frac{E(X^2)}{y^2}$$

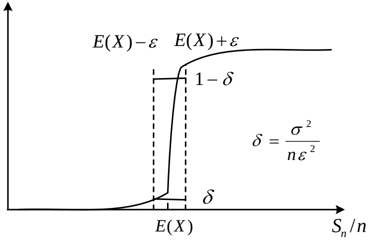
$$Z \leqslant \frac{e^{sX}}{e^{sy}} \Rightarrow P(X \geqslant y) = E(Z) \leqslant \frac{E(e^{sX})}{e^{sy}}.$$

7. 大数定律

设$X_1, X_2, \dots, X_n, \dots$是独立同分布 (i.i.d.)的，且具有数学期望$E(X)$和方差$\sigma^2$，则序列$S_n = \sum_{k=1}^{n} X_k$满足(Chebyshev不等式)：

$$P(|S_n/n - E(X)| \geqslant \varepsilon) \leqslant \sigma^2/n\varepsilon^2$$

1. 统计平均在概率意义上趋于集合平均，即$S_n/n \xrightarrow{P} E(X)$;
2. 随着n的增大，S_n/n的分布函数趋于一个阶跃函数，阶跃点在$E(X)$。



8. 中心极限定理

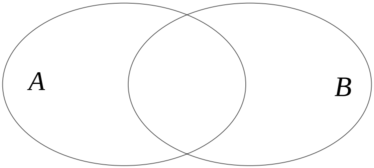
设$X_1, X_2, ..., X_n$是独立同分布（i.i.d.）的随机变量，具有数学期望$E(X)$和方差σ^2，则序列$S_n = \sum_{k=1}^{n} X_k$满足：

$$\lim_{n\to\infty} P\left(\frac{S_n - nE(X)}{\sqrt{n}\sigma} \leqslant y\right) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

用语言描述即，$\frac{S_n - nE(X)}{\sqrt{n}\sigma}$的分布函数趋于正态分布函数（注意，前者可能没有密度函数）。

9. 简单事实

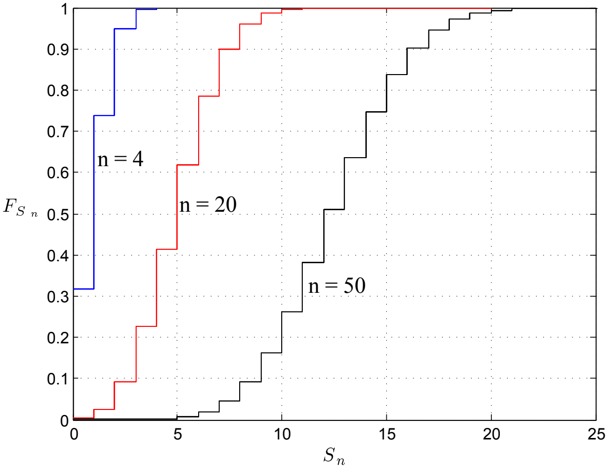
- $P(A \cup B) \geqslant \max\{P(A), P(B)\}$
- $P(AB) \leqslant \min\{P(A), P(B)\}$
- $P(A \cup B) \leqslant P(A) + P(B)$ (并集上限)



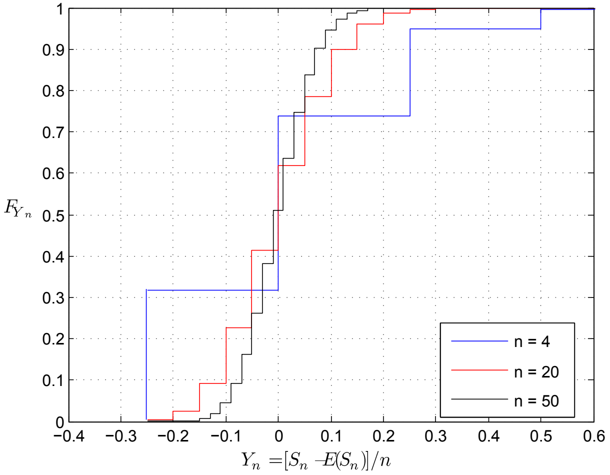
9. 简单事实

1. 必然事件$\neq$ 概率等于1的事件，不可能事件$\neq$ 概率等于0的事件；若样本空间是离散的，可以不加区分这些细微差别。
2. 若随机变量X的数学期望是$E(X)$，则一定存在$x \leqslant E(X)$，使$\Pr(X \leqslant x) > 0$。（"抽屉"原理）
3. 独立同分布的$X_1, X_2, \cdots, X_n, \cdots$，具有数学期望$E(X)$和方差$\sigma^2$，设$S_n = \sum_{k=1}^{n} X_k$。
   - 不能说 $S_n \to nE(X)$；
   - 也不能说$S_n \sim \mathcal{N}(nE(X), n\sigma^2)$。
   - 以$P_{X_k}(0) = 3/4, P_{X_k}(1) = 1/4$为例，画图说明。

9. 简单事实

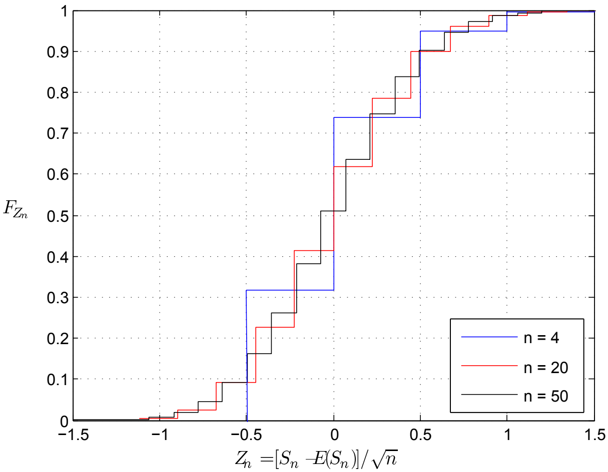


9. 简单事实



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
F_{Y_n}
−0.4
−0.3
−0.2
−0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
$Y_n = [S_n - E(S_n)]/n$
n = 4
n = 20
n = 50

9. 简单事实



谢谢！