信息论与编码

主讲教师: 马啸
maxiao@mail.sysu.edu.cn

计算机学院(软件学院)
中山大学

2021年春季学期

TA：

王翌昕，wangyx58@mail2.sysu.edu.cn
郭孔靖，guokj@mail2.sysu.edu.cn

成绩组成:

平时50%(考勤5，作业18，随堂12，课程设计15)+期末50%(闭卷考试)

1. 数字通信系统

2. 概率论回顾

数字通信系统

C. E. Shannon, “*A mathematical theory of communication,*” Bell Sys. Tech. Journal, 27, 379-423, 623-656, 1948.

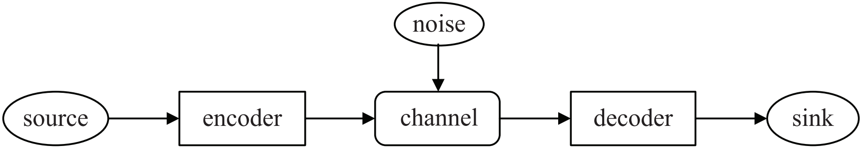


Figure: Block diagram of communication system.

通信：把消息（文本，语音，图像，视频等）从信源传输（或存储）到信宿；
编码/译码：消息→发送信号/接收信号→消息；
信道：发送信号⇝ 接收信号。

1. 用多少二进制数位表示消息是足够的？信源编码定理
2. 添加多少“冗余”可以保证可靠传输？信道编码定理

The distinguishing characteristics of this theory are, first, a great emphasis on probability theory and, second, a primary concern with the encoder and decoder, both in terms of their functional roles and in terms of the existence (or nonexistence) of encoders and decoders that achieve a given level of performance.–Gallager 1968.

The distinguishing characteristics of this theory are, first, a great emphasis on probability theory and, second, a primary concern with the encoder and decoder, both in terms of their functional roles and in terms of the existence (or nonexistence) of encoders and decoders that achieve a given level of performance.–Gallager 1968.

Information theory answers two fundamental questions in communication theory: what is the ultimate data compression (answer: the entropy H), and what is the ultimate transmission rate of communication (answer: the channel capacity C). For this reason some consider information theory to be a subset of communication theory. We argue that it is much more.–Cover and Thomas 1991

The distinguishing characteristics of this theory are, first, a great emphasis on probability theory and, second, a primary concern with the encoder and decoder, both in terms of their functional roles and in terms of the existence (or nonexistence) of encoders and decoders that achieve a given level of performance.–Gallager 1968.

Information theory answers two fundamental questions in communication theory: what is the ultimate data compression (answer: the entropy H), and what is the ultimate transmission rate of communication (answer: the channel capacity C). For this reason some consider information theory to be a subset of communication theory. We argue that it is much more.–Cover and Thomas 1991

Advances in digital communication in the latter half of this century were guided by the lessons of information theory but fueled by the progress in solid state electronics.–Viterbi 1991.

概率论回顾

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：{∅,{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：

$$
\begin{aligned}
Pr\{\text{正}\} &= 1/2 \\
Pr\{\text{反}\} &= 1/2
\end{aligned}
$$

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：

$$\begin{aligned} Pr\{\text{正}\} &= 1/2 \\ Pr\{\text{反}\} &= 1/2 \end{aligned}$$

模型2

1. 样本空间、事件集同模型1。
2. 概率：

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

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

$$\begin{aligned} Pr\{正\} &= 1/2 \\ Pr\{反\} &= 1/2 \end{aligned}$$

模型2

1. 样本空间、事件集同模型1。
2. 概率：

$$\begin{aligned} Pr\{正\} &= 1/4 \\ Pr\{反\} &= 3/4 \end{aligned}$$

模型3

1. 样本空间：{正,反,立}
2. 事件集、概率略去。

掷一均匀骰子

1. 样本空间$\Omega = \{1, 2, 3, 4, 5, 6\}$
2. 取$A = \{1, 3, 5\}$. $\{\Omega, A, \varnothing\}$ 是事件集吗？$\{\Omega, A, \overline{A}, \varnothing\}$ 是事件集吗？
3. 通常情况下，我们对于离散样本空间取所有子集构成的类为事件类，即任意一个子集都视作事件。若样本空间大小为$|\Omega|$, 问共有多少子集？
4. 若样本空间为离散样本空间，则概率的定义通常以定义单点集的概率为基本方式。若骰子是均匀的，则定义

$$\mathrm{P}\{1\} = \mathrm{P}\{2\} = \cdots = \mathrm{P}\{6\} = \frac{1}{6}$$

2. 概率运算法则

- $P(A \cup B) = P(A) + P(B) - P(AB)$;
  若A与B不交（互不相容），则$P(A \cup B) = P(A) + P(B)$;
- **容斥定理**:若事件集$A_1, ..., A_n$ 为有限集，则有：
  $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) + \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n)$;

例子：在0到200的整数中，求$A = \{$能被3整除的数$\}$，$B = \{$能被5 整除的数$\}$和$C = \{$能被7整除的数$\}$并集的大小。

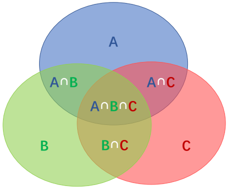


Figure: 3个集合的容斥定理

2. 概率运算法则

- 设A与B为样本空间Ω中的两个事件，其中$P(B) > 0$。那么在事件B发生的条件下，事件A发生的**条件概率**为：

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

问题:讨论P($B|A$)与P(B)的大小关系。

2. 概率运算法则

- **多事件独立**：设有$A_1, A_2, \ldots, A_n (n \geq 2)$ 个事件，如果对于其中任意2个，任意3 个，…，任意n个事件的积事件的概率，都等于各事件概率之积，则称它们相互独立。
- $P(AB) = P(A)P(B|A) = P(B)P(A|B)$；
  若A与B**相互独立**，则$P(AB) = P(A)P(B)$；
  注意，两两独立并不意味着相互独立。

例子：掷两个骰子，记:
事件$A = \{$第一个骰子点数为奇数$\}$
事件$B = \{$第二个骰子点数为奇数$\}$
事件$C = \{$两个骰子点数和为奇数$\}$
则A,B,C两两独立，但不互相独立。

不相容，计算和事件的概率简单；独立，计算积事件的概率简单。

2. 概率运算法则

- **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**：$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{P(A)}=\frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}$；

  例子1：Alice向Bob发送字符0或1，发送0的概率为p_0，传输过程中字符有概率p_e被翻转，求当Bob 收到的字符为1 时，Alice 发送字符为0的概率。

  例子2：甲乙两人网聊。甲向乙随机发送一个数字$X\in\{0,1,2\}$，概率依次为$\frac{1}{2}$，$\frac{1}{4}$，$\frac{1}{4}$。乙收到数字$Y\in\{0,1,2\}$。但传输过程中有可能出错。设错误模型是$Y=X+Z mod 3$，其中$Z\in\{0,1,2\}$ 的分布律是$\frac{7}{8}$，$\frac{1}{16}$，$\frac{1}{16}$。讨论乙如何根据接收的Y推断发送的X，并给出具体的准则。
- 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**。

2. 概率运算法则

例子：Ambiguity in determining the conditioning event.

A local radio station has a promotional scheme underway. Listeners are encouraged to register for a drawing to win tickets to the baseball *World Series*. To number the finalists, 1, 2, 3, 4 and 5, and let A_i be the event that finalist number i is selected among the wining three.

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$X:\Omega\to\mathbb{R}$，且满足对于任意实数x，$\{X\le x\}$是个事件。

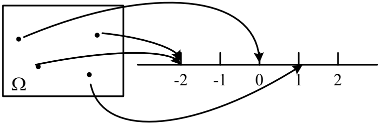


Figure: 随机变量

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$\sum_i P_X(x_i) = 1$。

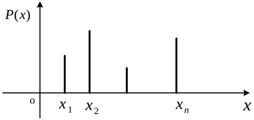


Figure: 离散型随机变量的pmf (又称分布律), "谱线"。

- 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$\int_{\mathbb{R}} f_X(x)\mathrm{d}x = 1$。

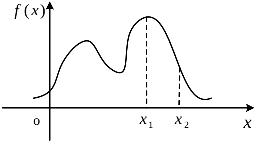


Figure: 连续型随机变量的pdf, "谱密度"。

3. 随机变量

Table 1.1 The PDF, mean, variance and MGF for some common continuous rv s

Name	PDF $f_X(x)$	Mean	Variance	MGF $g_X(r)$
Exponential:	$\lambda \exp(-\lambda x); \quad x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - r}; \quad \text{for } r < \lambda$
Erlang:	$\frac{\lambda^n x^{n-1} \exp(-\lambda x)}{(n-1)!}; \quad x \geq 0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - r}\right)^n; \quad \text{for } r < \lambda$
Gaussian:	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-a)^2}{2\sigma^2}\right)$	a	σ^2	$\exp(ra + r^2\sigma^2/2)$
Uniform:	$\frac{1}{a}; \quad 0 \leq x \leq a$	$\frac{a}{2}$	$\frac{a^2}{12}$	$\frac{\exp(ra) - 1}{ra}$

3. 随机变量

Table 1.2 The PMF, mean, variance and MGF for some common discrete rv s

Name	PMF $\mathsf{p}_M(m)$	Mean	Variance	MGF $\mathsf{g}_M(r)$
Binary:	$\mathsf{p}_M(1)=p;\ \mathsf{p}_M(0)=1-p$	p	$p(1-p)$	$1-p+pe^r$
Binomial:	$\binom{n}{m}p^m(1-p)^{n-m};\ 0\le m\le n$	np	$np(1-p)$	$[1-p+pe^r]^n$
Geometric:	$p(1-p)^{m-1};\ m\ge 1$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^r}{1-(1-p)e^r}$; for $r<\ln\frac{1}{1-p}$
Poisson:	$\frac{\lambda^n\exp(-\lambda)}{n!};\ n\ge 0$	λ	λ	$\exp[\lambda(e^r-1)]$

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$P_X(x_1)$。

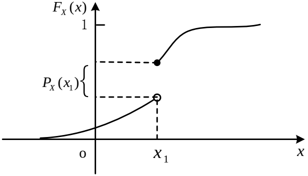
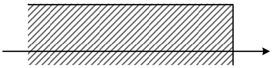


Figure: 随机变量的cdf。

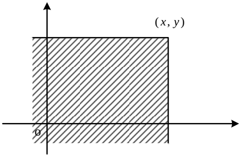
3. 随机变量

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。



一维



二维

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$P_X(x) = \sum_y P_{X,Y}(x, y)$, $P_Y(y) = \sum_x P_{X,Y}(x, y)$。

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$\sum\limits_{k=1}^{\infty} x_k p_k$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，$D(X+Y) = D(X) + D(Y)$。

4. 数字特征

例子

期末某校评优，甲、乙、丙、丁、戊5名同学分别获得“德、智、体、美、劳”单项奖状。不巧的是，辅导员发放奖状时完全随机放乱了。请问，5名同学中平均意义上有几人拿到了与自己匹配的奖状？

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

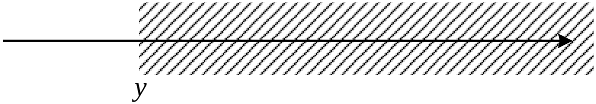
- Markov 不等式：线性递减至0

  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

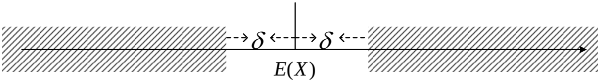
- Chebyshev 不等式：平方递减至0

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

其中σ_X^2表示X的方差。

证明：对$(X-E(X))^2$ 应用Markov不等式，

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



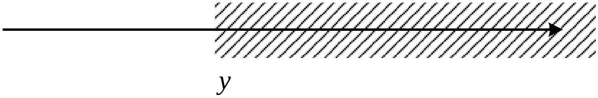
6. 尾部概率

- Chernoff 界：指数递减至0

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

证明： 对e^{sX}应用Markov不等式，

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



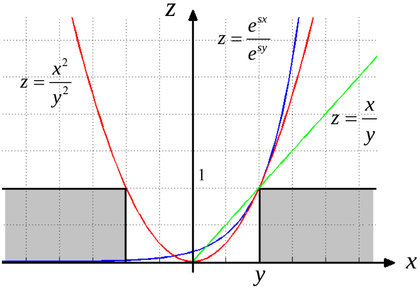
类似地，$P(X \leqslant y) \leqslant E(e^{sX})e^{-sy}$ 对所有 $s \leqslant 0$。

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

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

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

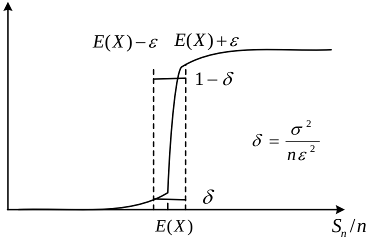
$$Z\leqslant\frac{e^{sX}}{e^{sy}}\Rightarrow P(X\geqslant y)=E(Z)\leqslant\frac{E(e^{sX})}{e^{sy}}.$$

7. 大数定律

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

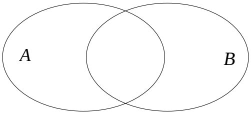
设$X_1, X_2, \dots, X_n$ 是独立同分布（i.i.d.）的随机变量，具有数学期望$E(X)$和方差σ^2，则序列$S_n = \sum_{k=1}^{n} X_k$满足：

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

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)



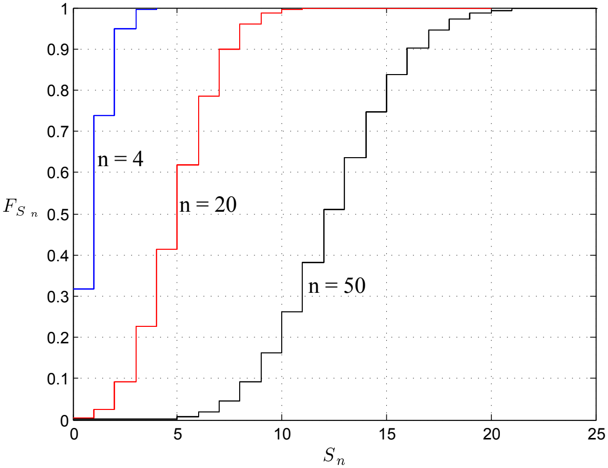
设A，B两个事件的概率均大于3/4。证明A 与B的积事件的概率大于1/2。

思考题：两个小概率事件的和事件也是小概率事件；两个大概率事件的积事件也是大概率事件。

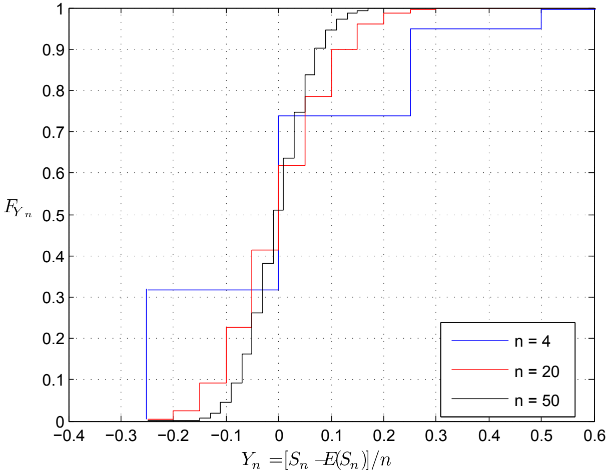
9. 简单事实

1. 必然事件≠ 概率等于1的事件, 不可能事件≠ 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$\mathrm{P}(X \leqslant x) > 0$。(“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$S_n = \sum_{k=1}^{n} X_k$。
   - 一般不说 $S_n \to nE(X)$；
   - 一般不说$S_n \sim \mathcal{N}(nE(X), n\sigma^2)$。
   - 以$P_{X_k}(0) = 3/4, P_{X_k}(1) = 1/4$ 为例，画图说明。

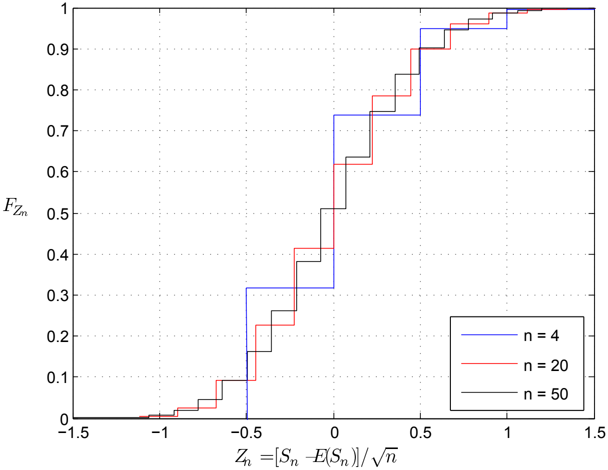
9. 简单事实



9. 简单事实


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

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_{Z_n}
−1.5
−1
−0.5
0
0.5
1
1.5
$Z_n = [S_n - E(S_n)]/\sqrt{n}$
n = 4
n = 20
n = 50

Homework

Exercise 1: I choose a number uniformly at random from the range [1, 1,000,000]Using the inclusion-exclusion principle, determine the probability that the number chosen is divisible by one or more of 4, 6, and 9.

Exercise 2: I have a fair coin and a two- headed coin. I choose one of the two coins randomly with equal probability and flip it. Given that the flip was heads, what is the probability that I flipped the two-headed coin?

Homework

Exercise 3: The following problem is known as the Monty Hall problem, after the host of the game show "Let's Make a Deal ". There are three curtains. Behind one curtain is a new car and behind the other two are goats. The game is played as follows. The contestant chooses the curtain that she thinks the car is behind. Monty then opens one of the other curtains to show a goat. (Monty may have more than one goat to choose from; in this case, assume he chooses which goat to show uniformly at random.) The contestant can then stay with the curtain she originally chose or switch to the other unopened curtain. After that, the location of the car is revealed, and the contestant wins the car or the remaining goat. Should the contestant switch curtains or not, or does it make no difference?

Homework

Exercise 4: Suppose that we roll ten standard six-sided dice. W hat is the probability that their sum will be divisible by 6, assuming that the rolls are independent? (Hint: Use the principle of deferred decisions, and consider the situation after rolling all but one of the dice.)

Exercise 5: Let X be a random variable with distribution function $F_x(x)$. Find the distribution function of the following random variables.

a) The maximum of n IID random variables with distribution function $F_x(x)$.

b) The minimum of n IID random variables with distribution $F_x(x)$.

c) The difference of the random variables defined in (a) and (b); assume X has a density $f_x(x)$.

Homework

Exercise 6: Let $X_1, X_2, ..., X_n, ...$ be a sequence of independent identically distributed (IID) continuous random variables with the common probability density function $f_X(x)$; note that $P(X = \alpha) = 0$ for all α and that $P(X_1 = X_2) = 0$.

a) Find $P(X_1 \leq X_2)$ (give a numerical answer, not an expression; no computation is required and a one or two line explanation should be adequate).

b) Find $P(X_1 \leq X_2; X_1 \leq X_3)$ (in other words, find the probability that X_1 is the smallest of X_1, X_2, X_3; again, think–don't compute).

谢谢！