概率论与数理统计

Lecture 6

随机变量的数字特征与极限定理

曾靖

中国科学技术大学统计与金融系



目录

4.1 数学期望及中位数

4.2 方差与矩

4.3 大数定律和中心极限定理

4.1 数学期望及中位数

随机变量的数字特征

- **分布函数**完整刻画随机变量的概率性质。
- **数字特征**是描述随机变量某一方面性质的量。
  - 中心位置：**期望、中位数、众数**。
  - 散布程度：**方差、绝对偏差、极差**。
  - 分布形状与相关程度：**偏度、峰度、相关系数**。

数学期望的引入：分赌金问题

例 1

甲乙两人赌技相同，各出赌金 150 元，约定先胜三局者为胜，取得全部 300 元。现在甲胜 2 局乙胜 1 局的情况下中止，问赌本该如何分？

分赌金问题：期望所得

引入随机变量 X 表示继续赌下去甲的最终所得，则

$$P(X=300)=\frac{3}{4},\qquad P(X=0)=\frac{1}{4}.$$

- 在**平均意义下**，甲的所得是多少?
- 以每个可能值的**概率为权重**，作**加权平均**:

$$300\times\frac{3}{4}+0\times\frac{1}{4}=225(\text{元})\ .$$

- 这个加权平均称为 X 的**数学期望**，记为 $E(X)$。

因此，按期望所得分配，甲得 225 元，乙得 75 元。

数学期望的定义：离散型

定义 4.1　数学期望：离散型

设 X 的分布律为 $P(X = x_i) = p_i$，$i = 1, 2, \ldots$。若

$$\sum_{i=1}^{\infty} |x_i| p_i < \infty,$$

则称

$$EX = \sum_{i=1}^{\infty} x_i p_i$$

为 X 的**数学期望（均值）**。

数学期望的定义：连续型

定义 4.2　数学期望：连续型

设 X 的概率密度为 $f(x)$。若

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty. \tag{1}$$

则称 $EX = \int_{-\infty}^{\infty} x f(x)\, dx$ 为 X 的**数学期望**。若绝对可积条件 (1) 不成立，则称数学期望不存在。

常见分布的数学期望

常见分布的期望

- 二项分布 $X \sim B(n,p)$：$EX = np$
- Poisson 分布 $X \sim P(\lambda)$：$EX = \lambda$
- 均匀分布 $X \sim U[a,b]$：$EX = \frac{a+b}{2}$
- 正态分布 $X \sim N(\mu, \sigma^2)$：$EX = \mu$
- 指数分布 $X \sim Exp(\lambda)$：$EX = 1/\lambda$
- 卡方分布 $X \sim \chi^2_n$：$EX = n$
- t 分布 $X \sim t_n$：$EX = 0$（$n > 1$）

注：柯西分布的期望不存在。

数学期望的性质

以下均假设相应期望存在。

- **线性性质：** 对常数 $c_1, \ldots, c_n$，

$$E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i EX_i.$$

- **独立随机变量之积：** 若 $X_1, \ldots, X_n$ 相互独立，则

$$E(X_1 \cdots X_n) = EX_1 \cdots EX_n.$$

随机变量函数的期望

- **随机变量函数的期望:**

$$Eg(X) = \begin{cases} \sum_i g(a_i)p_i, & \text{离散型,} \\ \int_{-\infty}^{\infty} g(x)f(x)\,dx, & \text{连续型.} \end{cases}$$

条件期望

定义 4.3　条件期望

在给定 $X = x$ 下，Y 的条件期望为

$$E(Y|X=x) = \begin{cases} \int_{-\infty}^{+\infty} y f(y|x)\, dy, & \text{连续型,} \\ \sum_i y_i P(Y = y_i \mid X = x), & \text{离散型.} \end{cases}$$

- 记 $g(x) = E(Y \mid X = x)$，它是关于 x 的**函数**。
- 将 x 换成随机变量 X，得到

$$E(Y \mid X) = g(X).$$

它是随机变量 X 的函数，因此也是一个**随机变量**。

全期望公式

定理 4.1　全期望公式

若 $E|X| < \infty$，则 $E\{E[X \mid Y]\} = EX$。

推导（连续型）：由条件期望的定义，

$$\begin{aligned} E\{E[X \mid Y]\} &= \int_{\mathbb{R}}\int_{\mathbb{R}} x f_{X|Y}(x \mid y) f_Y(y)\, dx\, dy \\ &= \int_{\mathbb{R}} x \left[\int_{\mathbb{R}} f_{X,Y}(x,y)\, dy\right] dx \\ &= \int_{\mathbb{R}} x f_X(x)\, dx = EX. \end{aligned}$$

全期望公式：地牢问题

例 2

地牢窃贼选择三个门：门 1 走 3 小时获自由；门 2 走 5 小时回地牢；门 3 走 7 小时回地牢。每次等可能选择，求获得自由的平均时间。

解：设 X 为总时间，则 $EX = \frac{1}{3}[3 + (5 + EX) + (7 + EX)]$，解得 $EX = 15$ 小时。

中位数

定义 4.4　中位数

称 m 为随机变量 X 的中位数，如果

$$P(X \le m) \ge \frac{1}{2}, \quad P(X \ge m) \ge \frac{1}{2}.$$

- 中位数受极端值影响小。
- 中位数可能不唯一。例如，若

$$P(X=1) = P(X=2) = P(X=3) = P(X=4) = \frac{1}{4},$$

则 $[2,3]$ 内的任意数都是中位数。

众数

定义 4.5　众数

- 离散型：概率质量函数最大值对应的取值称为众数 m_d。
- 连续型：使密度 $f(x)$ 达到最大值的 x 称为众数 m_d。

- 众数可能不唯一。若密度只有一个峰，则称单峰。

分位数

定义 4.6　分位数

设 $0 < p < 1$，称 Q_p 为 X 的 p 分位数，如果

$$P(X \le Q_p) \ge p, \quad P(X \ge Q_p) \ge 1 - p.$$

- **四分位数**：$Q_{0.25}, Q_{0.5}, Q_{0.75}$ 分别称为第一、第二、第三四分位数，其中 $Q_{0.5}$ 是中位数。
- **内四分位距**（interquartile range, IQR）：

$$IQR = Q_{0.75} - Q_{0.25}.$$

- **百分位数**：$Q_{k/100}$ 称为第 k 百分位数，$k = 1, \ldots, 99$。

正态分布的四分位距

例 3

设 $X \sim N(\mu, \sigma^2)$，求内四分位距 $IQR(X)$。

解：标准正态分布的 0.75 和 0.25 分位数分别约为

$$0.6745 \quad \text{和} \quad -0.6745.$$

因此，

$$\begin{aligned} IQR(X) &= Q_{0.75} - Q_{0.25} \\ &\approx (\mu + 0.6745\sigma) - (\mu - 0.6745\sigma) \\ &= 1.349\sigma. \end{aligned}$$

对数正态分布的中心位置

例 4

设 X 服从对数正态分布，即 $\ln X \sim N(\mu, \sigma^2)$，求 X 的期望、中位数和众数。

解：期望 $e^{\mu+\sigma^2/2}$，中位数 e^{μ}，众数 $e^{\mu-\sigma^2}$。

对数正态分布：期望的计算

令 $Y = \ln X \sim N(\mu, \sigma^2)$，则 $X = e^Y$。因此

$$EX = E(e^Y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(y - \frac{(y-\mu)^2}{2\sigma^2} \right) dy.$$

对指数中的表达式配方：

$$y - \frac{(y-\mu)^2}{2\sigma^2} = \mu + \frac{\sigma^2}{2} - \frac{[y-(\mu+\sigma^2)]^2}{2\sigma^2}.$$

故

$$EX = e^{\mu+\sigma^2/2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[y-(\mu+\sigma^2)]^2}{2\sigma^2}} dy}_{=1} = e^{\mu+\sigma^2/2}.$$

对数正态分布：中位数与众数的计算

中位数： 由 $P(X \le m) = P(Y \le \ln m) = \frac{1}{2}$，得 $\ln m = \mu$，故 $m = e^{\mu}$。

众数： X 的密度为

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0.$$

对数密度求导，得到

$$\frac{d}{dx} \ln f_X(x) = -\frac{1}{x} - \frac{\ln x - \mu}{\sigma^2 x} = \frac{\mu - \sigma^2 - \ln x}{\sigma^2 x}.$$

导数在 $x = e^{\mu - \sigma^2}$ 两侧由正变负，故众数为 $m_d = e^{\mu - \sigma^2}$。

Pop Quiz 1

设 $P(X=0)=P(X=1)=\frac{1}{2}$，且

$$Y \mid (X=0) \sim U(0,2), \qquad Y \mid (X=1) \sim U(0,4).$$

问题 1 求 $E(Y)$。

问题 2 Y 的中位数是否等于 $E(Y)$？求其中位数。

Pop Quiz 1：答案

问题 1 由全期望公式，

$$E(Y) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2 = \frac{3}{2}.$$

问题 2 当 $0 \le y \le 2$ 时，

$$F_Y(y) = \frac{1}{2} \cdot \frac{y}{2} + \frac{1}{2} \cdot \frac{y}{4} = \frac{3y}{8}.$$

令 $F_Y(m) = \frac{1}{2}$，得 $m = \frac{4}{3} \neq E(Y)$。

混合分布的期望可加权平均，中位数一般不能。

4.2 方差与矩

方差的引入与定义

例 5

甲、乙各射击 10 次。甲击中 8、9、10 环分别为 3、1、6 次；乙分别为 2、5、3 次。谁的发挥更稳定?

甲、乙的平均环数分别为 9.3 和 9.1。计算平均平方偏差:

$$\text{甲:}\ \frac{3(8-9.3)^2+1(9-9.3)^2+6(10-9.3)^2}{10}=0.81,$$

$$\text{乙:}\ \frac{2(8-9.1)^2+5(9-9.1)^2+3(10-9.1)^2}{10}=0.49.$$

乙的平均平方偏差更小，这 10 次射击的发挥更稳定。

方差与标准差

定义 4.7　方差与标准差

设 X 为随机变量，若 X 平方可积，则称

$$\mathrm{Var}(X) = E(X - EX)^2 = \sigma^2$$

为 X 的方差，其平方根 σ 称为标准差。

- 方差和标准差是描述随机变量在**均值附近散布程度**的数字特征。

方差的性质

定理 4.2

设 c 为常数，则

1. $0 \leq \mathrm{Var}(X) = EX^2 - (EX)^2$，故 $(EX)^2 \leq E(X^2)$（Jensen 不等式）。
2. $\mathrm{Var}(X) = 0$ 当且仅当存在常数 c，使得 $P(X = c) = 1$。
3. $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$，$\mathrm{Var}(X + c) = \mathrm{Var}(X)$。

方差的性质（续）

定理 4.2　（续）

4. 若 X, Y 独立，a, b 为常数，则

$$\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y).$$

若 $X_1, \ldots, X_n$ 相互独立，则对常数 $a_1, \ldots, a_n$，

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i).$$

5. 对任何常数 c，$\mathrm{Var}(X) \le E(X - c)^2$，等号成立当且仅当 $c = EX$。

常见分布的方差

常见分布的方差

- 二项分布 $X \sim B(n,p)$：$\mathrm{Var}(X) = np(1-p)$
- Poisson 分布 $X \sim P(\lambda)$：$\mathrm{Var}(X) = \lambda$
- 均匀分布 $X \sim U[a,b]$：$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
- 指数分布 $X \sim Exp(\lambda)$：$\mathrm{Var}(X) = 1/\lambda^2$
- 正态分布 $X \sim N(\mu, \sigma^2)$：$\mathrm{Var}(X) = \sigma^2$

马尔可夫不等式

定理 4.3　马尔可夫不等式

若随机变量 $Y \geq 0$，则对任意 $\varepsilon > 0$，

$$P(Y \geq \varepsilon) \leq \frac{EY}{\varepsilon}.$$

引理 4.1　切比雪夫不等式

若 $\mathrm{Var}(X) < \infty$，则对任意 $\varepsilon > 0$，

$$P(|X - EX| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

不等式的应用：调查人数

例 6　教材例 4.23

为了解某地区青少年沉迷于网络游戏的比例 p，进行随机抽样调查。需要调查多少人，才能使调查比例与真实比例之差的绝对值小于 0.5% 的概率不小于 95%?

设调查 n 人，其中 N 人沉迷于网络游戏。将各次调查视为独立试验，则 $N \sim B(n,p)$。要求

$$P\left(\left|\frac{N}{n}-p\right|<0.005\right)\geq 0.95.$$

调查人数：解答

由 $EN = np$、$\mathrm{Var}(N) = np(1-p)$，利用切比雪夫不等式，

$$P\left(\left|\frac{N}{n} - p\right| < 0.005\right) \geq 1 - \frac{np(1-p)}{(0.005n)^2}$$
$$\geq 1 - \frac{1}{4n(0.005)^2},$$

其中利用了 $p(1-p) \leq \frac{1}{4}$。因此，只需

$$n \geq \frac{1}{4 \times 0.05 \times (0.005)^2} = 200\,000.$$

矩

定义 4.8　矩

设 X 为随机变量，c 为常数，r 为正整数。$E[(X-c)^r]$ 称为 X 关于 c 点的 r 阶矩。

- $c=0$：$\alpha_r = EX^r$ 称为 r 阶**原点矩**。
- $c=EX$：$\mu_r = E[(X-EX)^r]$ 称为 r 阶**中心矩**。

- 期望 $EX=\alpha_1$，是**一阶原点矩**。
- 方差 $\mathrm{Var}(X)=E[(X-EX)^2]=\mu_2$，是**二阶中心矩**。

偏度系数

定义 4.9　偏度系数

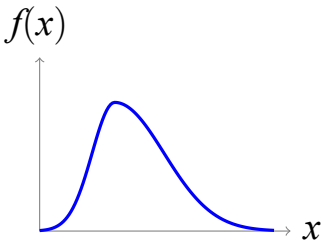
设 $\mu = EX$、$0 < \sigma^2 = \mathrm{Var}(X) < \infty$，且三阶绝对矩有限。称

$$\beta_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}.$$

为 X 的偏度系数。

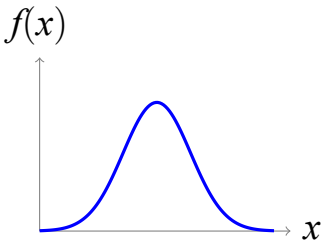
- 偏度刻画分布的不对称程度。

偏度：图形比较



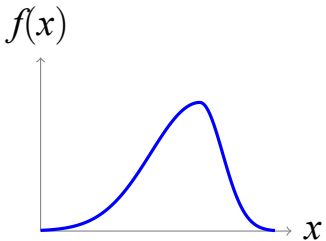
正偏（右偏）

$\beta_1 > 0$



对称

$\beta_1 = 0$



负偏（左偏）

$\beta_1 < 0$

峰度系数

定义 4.10　峰度系数

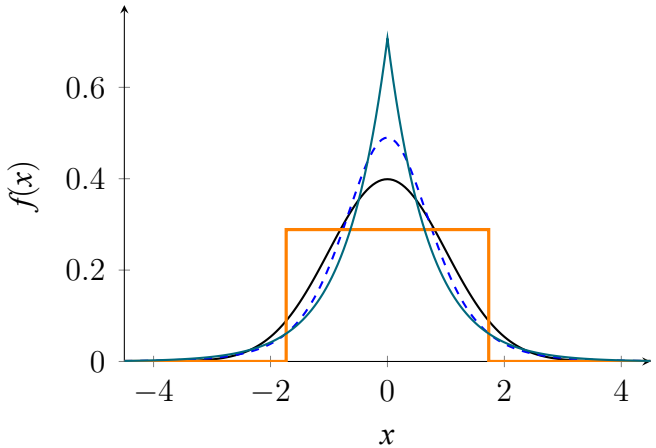
设 $\mu = EX$、$0 < \sigma^2 = \mathrm{Var}(X) < \infty$，且四阶矩有限。称

$$\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4} = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2}.$$

为 X 的**峰度系数**。

- 峰度系数用于刻画尾部的轻重，通常与正态分布的峰度 3 作比较。

峰度：图形比较



- 正态：峰度 3
- t_5：峰度 $9 > 3$
- 均匀：峰度 $1.8 < 3$
- 拉普拉斯：峰度 $6 > 3$

矩母函数

定义 4.11　矩母函数

随机变量 X 的矩母函数（MGF）定义为

$$M_X(s) = E[e^{sX}].$$

若存在 $a > 0$ 使 $M_X(s)$ 对所有 $s \in [-a, a]$ 有限，则称矩母函数存在。

- 矩母函数存在时，可用来判断两个随机变量是否**同分布**。

矩母函数的性质

定理 4.4

若存在 $c > 0$，使 X, Y 的矩母函数在 $[-c, c]$ 上有限且相等，则它们的分布相同。

引理 4.2　独立随机变量之和

若 $X_1, \ldots, X_n$ 相互独立，则在各矩母函数均有限的 s 处，

$$M_{X_1+\cdots+X_n}(s) = M_{X_1}(s) \cdots M_{X_n}(s).$$

矩母函数与原点矩

引理 4.3　由矩母函数求原点矩

若 $M_X(s)$ 在零点的某个邻域内有限，则对任意正整数 k，

$$E(X^k) = M_X^{(k)}(0) = \frac{d^k}{ds^k} M_X(s)\bigg|_{s=0}.$$

- 对矩母函数求 k 阶导数，再令 $s = 0$，即得 k 阶原点矩。
- 特别地，

$$EX = M_X'(0), \qquad E(X^2) = M_X''(0).$$

协方差

定义 4.12　协方差

若 X, Y 平方可积，称

$$\mathrm{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

为 X 与 Y 的协方差。

- 方差是协方差的特例：

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X).$$

协方差的性质

设 X, Y 均平方可积，a, b, c, d 为实数。

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$。
2. $\mathrm{Cov}(X, Y) = E(XY) - (EX)(EY)$。
3. 平移与双线性性质：

$$\begin{aligned}\mathrm{Cov}(aX + b, cY + d) &= ac\,\mathrm{Cov}(X, Y),\\ \mathrm{Cov}(aX + bY, cX + dY) &= ac\,\mathrm{Var}(X)\\ &\quad + (ad + bc)\,\mathrm{Cov}(X, Y) + bd\,\mathrm{Var}(Y).\end{aligned}$$

4. 若 X, Y 独立，则 $\mathrm{Cov}(X, Y) = 0$。一般地，

$$[\mathrm{Cov}(X, Y)]^2 \le \mathrm{Var}(X)\,\mathrm{Var}(Y) \quad \text{（柯西–施瓦茨不等式）}.$$

等号成立当且仅当存在不全为零的 u, v，使 $uX + vY$ 以概率 1 为常数。

相关系数

定义 4.13　相关系数

若 $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$，称

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

为 X 与 Y 的相关系数。$\rho_{X,Y} = 0$ 时称不相关。

- 相关系数通过除以两个随机变量的标准差，**消除计量单位的影响**。

不相关与独立

定理 4.5

若 X, Y 的方差有限且相互独立，则它们不相关；反之不成立。

- 若 (X, Y) 服从**联合正态分布**，则 X, Y 独立当且仅当它们不相关。

回顾：方差的可加性

设各 X_i 的方差有限，a_i 为常数。利用 $\mathrm{Var}(S) = \mathrm{Cov}(S, S)$ 和协方差的**双线性性质**，

$$\begin{aligned}\mathrm{Var}\left(\sum\nolimits_i a_i X_i\right) &= \mathrm{Cov}\left(\sum\nolimits_i a_i X_i, \sum\nolimits_j a_j X_j\right) \\ &= \sum\nolimits_i \sum\nolimits_j a_i a_j \,\mathrm{Cov}(X_i, X_j) \\ &= \sum\nolimits_i a_i^2 \,\mathrm{Var}(X_i) + 2 \sum\nolimits_{i<j} a_i a_j \,\mathrm{Cov}(X_i, X_j).\end{aligned}$$

- 若各 X_i **两两不相关**，则交叉项为零，$\mathrm{Var}(\sum_i a_i X_i) = \sum_i a_i^2 \,\mathrm{Var}(X_i)$。
- 方差可加性只需**两两不相关**，不必要求相互独立。

不相关与独立：单位圆例题

例 7

设 (X, Y) 服从单位圆内的均匀分布，则 X, Y 不相关但不独立。

解：联合密度 $f(x,y) = \frac{1}{\pi}$ $(x^2 + y^2 \leq 1)$。边缘密度 $f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$，$f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$。计算得 $EX = EY = 0$，$E(XY) = 0$，故 $\mathrm{Cov}(X, Y) = 0$，不相关。但 $f(x,y) \neq f_X(x)f_Y(y)$，故不独立。

Pop Quiz 2

设 X, Y 相互独立，且各以概率 $\frac{1}{2}$ 取 -1 和 1。记

$$U = X + Y, \qquad V = X - Y.$$

问题 1 下列说法中，正确的是：

(A) U, V 相互独立。

(B) U, V 不相关，但不独立。

(C) U, V 的协方差不为零。

问题 2 求 $\mathrm{Var}(2U - V)$。

Pop Quiz 2：答案

问题 1 (B) 不相关，但不独立。

$$\mathrm{Cov}(U, V) = \mathrm{Var}(X) - \mathrm{Var}(Y) = 0.$$

但 $P(U = 0) = P(V = 0) = \frac{1}{2}$，而 $P(U = 0, V = 0) = 0$。

问题 2 因为 $2U - V = X + 3Y$，由独立性，

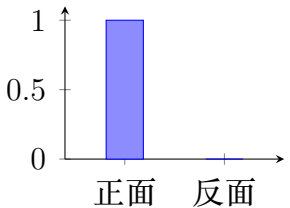
$$\mathrm{Var}(2U - V) = 1 + 9 = 10.$$

熵：从热力学到信息论

- **起源：**熵的概念起源于热力学，后来在统计力学中得到概率解释。
- **信息熵：**香农（Shannon）于 1948 年建立信息熵理论，用来度量离散随机变量的**不确定性**。
- 熵也是得知随机变量取值时获得的**平均信息量**：越难预测，熵越大；越容易预测，熵越小。

熵与不确定性：哪枚硬币更难预测？

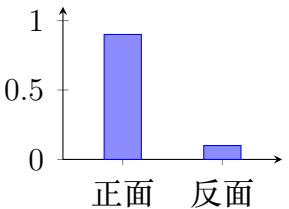
总是正面



$P(\text{正面}) = 1$

$H = 0$

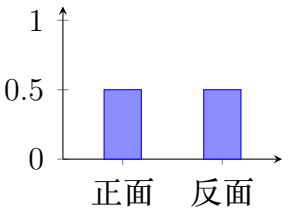
偏向正面



$P(\text{正面}) = 0.9$

$H \approx 0.469$

公平硬币



$P(\text{正面}) = 0.5$

$H = 1$

从左到右：结果越来越难预测，熵越来越大。

熵

定义 4.14　熵

设 X 为离散随机变量，分布律 $P(X=x_k)=p_k$，则熵定义为（约定 $0\log 0=0$）

$$H(X)=-\sum_{k=1}^{\infty}p_k\log_2(p_k).$$

若 X 为连续随机变量，其微分熵定义为（密度为 $f_X(x)$）

$$h(X)=-E[\ln f_X(X)]=-\int_{-\infty}^{\infty}f_X(x)\ln f_X(x)\,dx.$$

熵的性质

- 对于**离散型随机变量**，$H(X) \geq 0$。
- 若 X 取 n 个值，则 $H(X) \leq \log_2 n$，等号成立当且仅当为离散均匀分布。

注：对于连续型随机变量，熵 $H(X)$ 不一定非负。例如，$X \sim U(0, \frac{1}{2})$ 时，$H(X) = \ln \frac{1}{2} < 0$。

常见分布的熵

离散型： 使用 $\log_2$。

- 两点分布 $X \sim B(1, p)$：

$$H(X) = -p \log_2 p - (1-p) \log_2 (1-p).$$

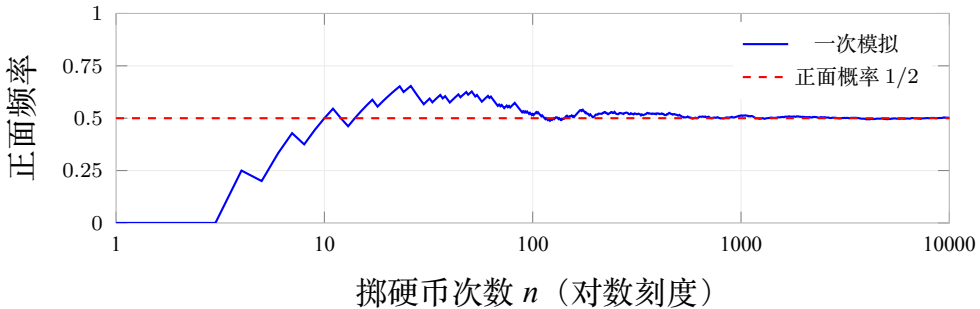
连续型： 微分熵使用自然对数 ln。

- 正态分布 $X \sim N(\mu, \sigma^2)$：$H(X) = \dfrac{1}{2} \ln(2\pi e \sigma^2)$。
- 指数分布（率参数 $\lambda > 0$）：$H(X) = 1 - \ln \lambda$。
- 均匀分布 $X \sim U(a, b)$：$H(X) = \ln(b-a)$。

4.3 大数定律和中心极限定理

掷硬币实验

独立地重复掷一枚公平硬币，记录前 n 次中正面朝上的**频率**。



- 令正面为 $X_i = 1$、反面为 $X_i = 0$，则频率就是样本均值 $\overline{X}_n$。

依概率收敛

定义 4.15　依概率收敛

如果对任何 $\varepsilon > 0$，都有

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0,$$

则称 $\{\xi_n\}$ 依概率收敛到 ξ，记为 $\xi_n \xrightarrow{p} \xi$。

弱大数定律

定理 4.6　弱大数定律

设 $\{X_n\}$ 为独立同分布随机变量序列，具有公共期望 μ 和有限方差 σ^2，则

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} \mu.$$

弱大数定律

证明：利用切比雪夫不等式。

由期望的线性性质及 $X_1, \ldots, X_n$ 的独立性，

$$E\overline{X} = \frac{1}{n}\sum_{k=1}^{n} EX_k = \mu, \qquad \mathrm{Var}(\overline{X}) = \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}(X_k) = \frac{\sigma^2}{n}.$$

对任意 $\varepsilon > 0$，由切比雪夫不等式，

$$P(|\overline{X} - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\overline{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \quad (n \to \infty).$$

因此，$\overline{X} \xrightarrow{p} \mu$。

依分布收敛

定义 4.16　依分布收敛

设 $X_1, X_2, \ldots$ 为一列实值随机变量，称其依分布收敛到 X，如果对 F 的所有连续点 $x \in \mathbb{R}$ 有

$$\lim_{n\to\infty} F_n(x) = F(x).$$

记为 $X_n \xrightarrow{\mathcal{L}} X$ 或 $X_n \xrightarrow{d} X$。

两种收敛的关系

定理 4.7

1. 若 $X_n \xrightarrow{p} X$，则 $X_n \xrightarrow{\mathcal{L}} X$。（反之不一定成立）
2. 若 $X_n \xrightarrow{\mathcal{L}} c$（常数），则 $X_n \xrightarrow{p} c$。

中心极限定理

定理 4.8　独立同分布中心极限定理

设 $\{X_n\}$ 为独立同分布随机变量序列，具有公共期望 μ 和方差 $0 < \sigma^2 < \infty$。记样本均值 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$，则

$$\sqrt{n}\,(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

即对任意 $x \in \mathbb{R}$，

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} \le x\right) = \Phi(x).$$

棣莫弗-拉普拉斯定理

定理 4.9　de Moivre–Laplace 定理

设 $X_1, \ldots, X_n$ 相互独立且 $P(X_i = 1) = 1 - P(X_i = 0) = p$ $(0 < p < 1)$。记样本均值 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$，则

$$\frac{\sqrt{n}\,(\overline{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1).$$

- 令 $S_n = \sum_{i=1}^{n} X_i = n\overline{X}_n$，整理可得等价形式：

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

中心极限定理应用示例

例 8

某选拔考试有 100 道选择题，每题 4 个答案，只有 1 个正确。考生随机答题。求：

1. 至少答对 25 题的概率；
2. 答对 20–40 题的概率；
3. 至少答对 36 题的概率。

中心极限定理：应用解答

令 $S_n = \sum_{i=1}^{n} X_i$，其中 X_i 为第 i 题得分。由中心极限定理，

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

本题 $n = 100$、$p = 0.25$，故 $np = 25$，$\sqrt{np(1-p)} = \sqrt{18.75} \approx 4.33$。由正态近似，得

1. $P(S_{100} \geq 25) \approx 1 - \Phi\left(\frac{25-25}{4.33}\right) \approx 0.5$。
2. $P(20 \leq S_{100} \leq 40) \approx \Phi\left(\frac{40-25}{4.33}\right) - \Phi\left(\frac{20-25}{4.33}\right) \approx 0.876$。
3. $P(S_{100} \geq 36) \approx 1 - \Phi\left(\frac{36-25}{4.33}\right) \approx 0.0055$。