- 度量了随机变量中所含有的信息量大小。体现随机变量的不确定性程度，熵值越大不确定性就越大.

设 X 为离散随机变量，分布律为

$$P(X = x_k) = p_k, \quad k \in \mathbb{N},$$

则其熵（**Entropy**）定义为

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

如果 X 为连续随机变量，概率密度函数为 $f_X(x)$，则其熵定义为

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \ln(f_X(x))\, dx,$$

例子

- (0-1 分布的熵) 设 X 是 0-1 分布随机变量, 分布律为

$$\mathbb{P}(X=1)=1-\mathbb{P}(X=0)=p,\ 0<p<1,$$

求其熵.

$$H(X)=-[p\log_2(p)+(1-p)\log_2(1-p)]$$

➢ $p=1/2$时, X的熵达到最大 . 此时X 的取值没有任何偏向性, 最难以预测和估计, 因此随机程度最高, 不确定性也最大.

➢ 若$p=0$或者 $p=1$, 则说明X取值是确定的, 没有任何随机性, 其熵为 0 很合理.

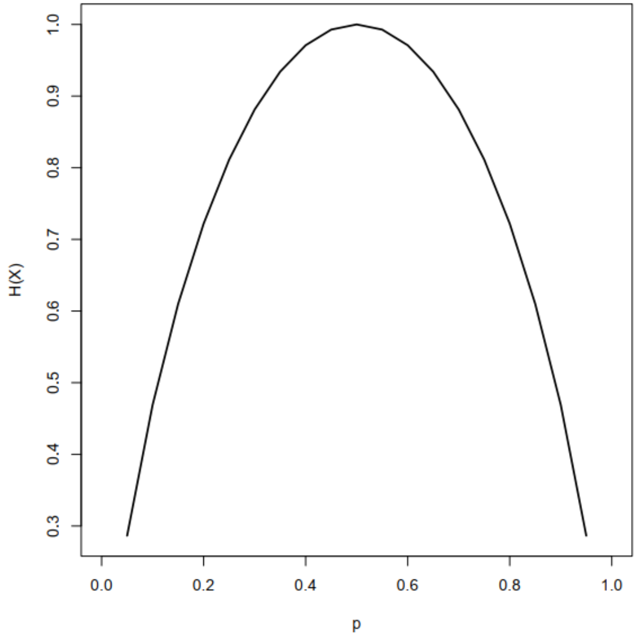


图 **4.5:** 0-1 分布的熵

性质 随机变量 X 的熵 $H(X)$ 有许多性质:

(1) $H(X) \geqslant 0$

(2)如果取有限个值的随机变量 X 的概率分布 $\{p_k, k=1,\cdots,n\}$ 不是离散均匀分布，那么有

$$H(X) \leqslant \log_2(n),$$

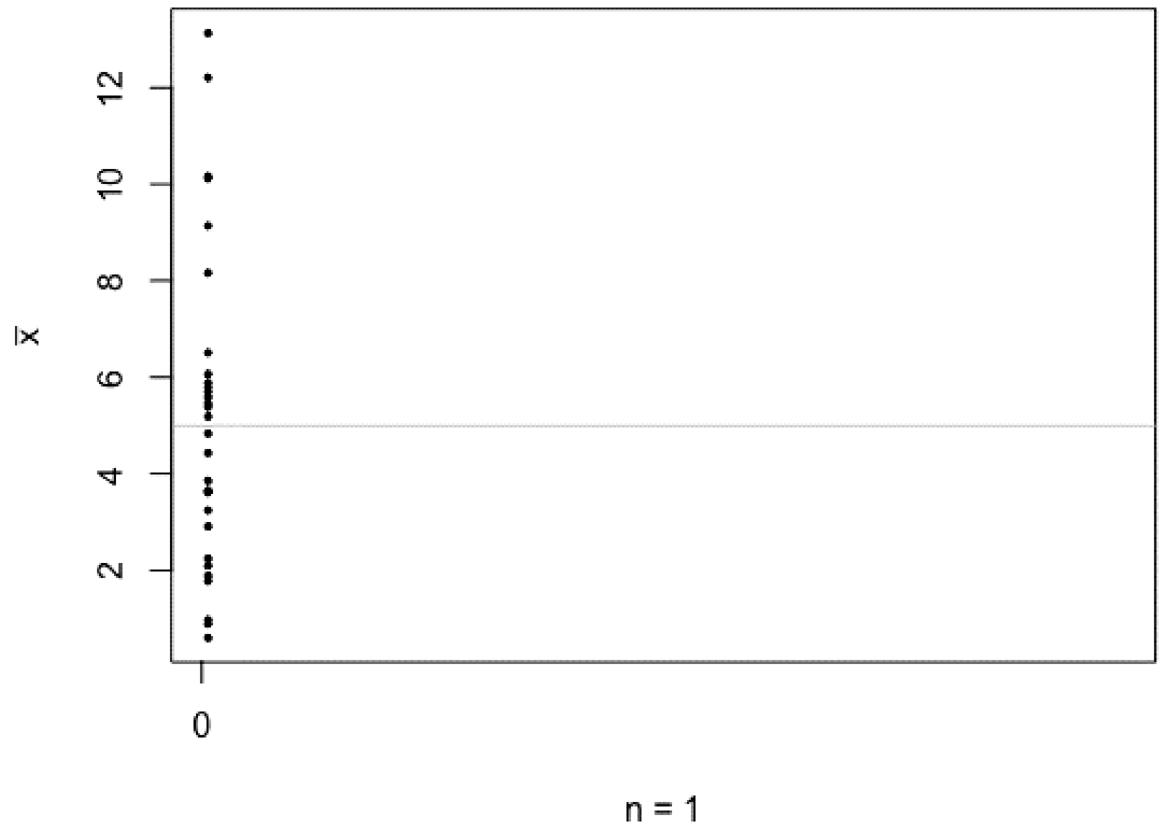
当且仅当 X 为离散均匀分布随机变量时等号成立.

- 大数定律, 是概率论中讨论随机变量和的平均值的收敛情况, 是数理统计学中参数估计的理论基础.

- 中心极限定理, 是概率论中讨论随机变量和的分布以正态分布为极限的一组定理, 这组定理是数理统计学和误差分析的理论基础, 指出了大量随机变量近似服从正态分布的条件.

$$\boxed{n \longrightarrow \infty}$$

6.1. 大数定律—引例：χ_5^2 分布



https://yihui.name/animation/example/lln-ani/

定义 4.13 依概率收敛

设 $X_1, X_2, \ldots, X_n, \cdots$ 是一随机变量序列, X 为随机变量, 如果对 $\forall \varepsilon > 0$, 有

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0, \tag{4.23}$$

则称随机变量序列 X_n 依概率趋于随机变量 X, 记为 $X_n \to X$, *in* P, 或 $X_n \xrightarrow{P} X$. ♣

特例： (X为常数c) 考虑随机变量序列$\{X_n\}$，$X_n \sim Exp(n)$，即X_n服从参数为n的指数分布. 证明$X_n \xrightarrow{P} 0$.

定理 4.4 大数定律

设 $X_1, X_2, \ldots, X_n, \cdots$ 是一列独立同分布 (i.i.d.) 随机变量序列, 记它们相同的期望和方差分别为 μ 和 σ^2. 记 $S_n = X_1 + X_2 + \cdots + X_n$, 则对 $\forall\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|S_n/n - \mu| \geq \varepsilon) = 0, \tag{4.24}$$

即： $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$， $\overline{X} \xrightarrow{P} \mu$

- （伯努利大数定律）若以Y_n表示n次伯努利实验中的成功次数，则对任意的常数$\epsilon > 0$，都有

$$\lim_{n\to\infty} P(|p_n - p| \geq \epsilon) = 0$$

其中$p_n = \frac{Y_n}{n}$。

- 频率(依概率)收敛于概率

- (弱)大数定律(weak law of large numbers(WLLN) 依概率收敛,
converges in probability) :

$$\overline{X} \xrightarrow{P} \mu, \qquad \lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right| \geq \epsilon\right) = 0$$

样本均值“依概率收敛”于总体均值

- (强)大数定律(strong law of large numbers(SLLN)几乎处处收敛,
converges almost surely) :

$$\overline{X} \xrightarrow{a.s.} \mu, \qquad P\left(\lim_{n\to\infty} \overline{X} = \mu\right) = 1$$

样本均值“以概率为1收敛”于总体均值

- 设 $X_1, X_2, \dots, X_n, \dots$ 为一列实值随机变量, X 为随机变量, F_n 和 F 分别为随机变量 X_n 和 X 的分布函数. 如果对 F 的所有连续点 $x \in \mathbb{R}$, 有

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

则称 $\{F_n\}$ 弱收敛(converge weakly)于 F，也称 $\{X_n\}$ 依分布收敛(converge in distribution)于 X, 常记为 $X_n \xrightarrow{\mathcal{L}} X$.

例 4.36 设 $X_n \sim U\left(0, \frac{1}{n}\right)$, X 表示退化到 0 的随机变量. 记 $F_n(x)$ 和 $F(x)$ 分别为 X_n 和 X 的分布函数, 则

$$F_n(x) = \begin{cases} 0, & x < 0 \\ nx, & 0 \leq x < 1/n, \\ 1, & x \geq 1/n \end{cases} \qquad F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

因此当 $n \to \infty$ 时对任意固定的 $x \in (-\infty, 0) \cup [1/n, +\infty)$ 有依分布收敛 $F_n(x) \to F(x)$ 成立. 但是在 $x = 0$ 处, 对任意的 n 成立 $F_n(0) \equiv 0$, 而 $F(0) \equiv 1$. 即在点 $x = 0$ 处依分布收敛不成立.

定理 4.5 依概率收敛与依分布收敛关系

设 $X_1, X_2, \ldots$ 为一列实值随机变量, X 为另一随机变量, 则

(1). 如果 $X_n \xrightarrow{P} X$, 则 $X_n \xrightarrow{\mathcal{L}} X$.

(2). 如果 $X_n \xrightarrow{\mathcal{L}} c$, 则 $X_n \xrightarrow{P} c$, 其中 c 为一个常数.

- **依分布收敛一般不能推出依概率收敛.**

例 4.37 设 $X_1, X_2, X_3, \ldots$ 为一列独立同分布的 $B\left(1, \frac{1}{2}\right)$ 随机变量序列. $X \sim B\left(1, \frac{1}{2}\right)$ 为一独立于序列 $\{X_i\}$ 的随机变量. 则 $X_n \xrightarrow{\mathcal{L}} X$. 但是, X_n 不依概率收敛到 X. 事实上, 因为 $|X_n - X|$ 实际上也是服从 $B\left(1, \frac{1}{2}\right)$ 的随机变量且

$$\mathbb{P}\left(|X_n - X| \geq \epsilon\right) = \frac{1}{2}, \quad \text{对 } 0 < \epsilon < 1.$$