第五章 极限理论

Probability and Statistics for Computer Scientists

第五章 极限理论

- 马尔可夫和切比雪夫不等式
- 弱大数定律
- 依概率收敛
- 中心极限定理
- 强大数定律

极限理论

设 $X_1, \cdots, X_n$,独立同分布 研究 S_n

$$S_n = X_1 + \cdots + X_n$$

随机变量序列 $M_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$

当 $n \to \infty$ 时, $\mathrm{E}[M_n] = \mu, \quad \mathrm{var}(M_n) = \frac{\sigma^2}{n}.$

另一个随机变量序列 $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}. \quad \mathrm{E}[Z_n] = 0, \quad \mathrm{var}(Z_n) = 1.$

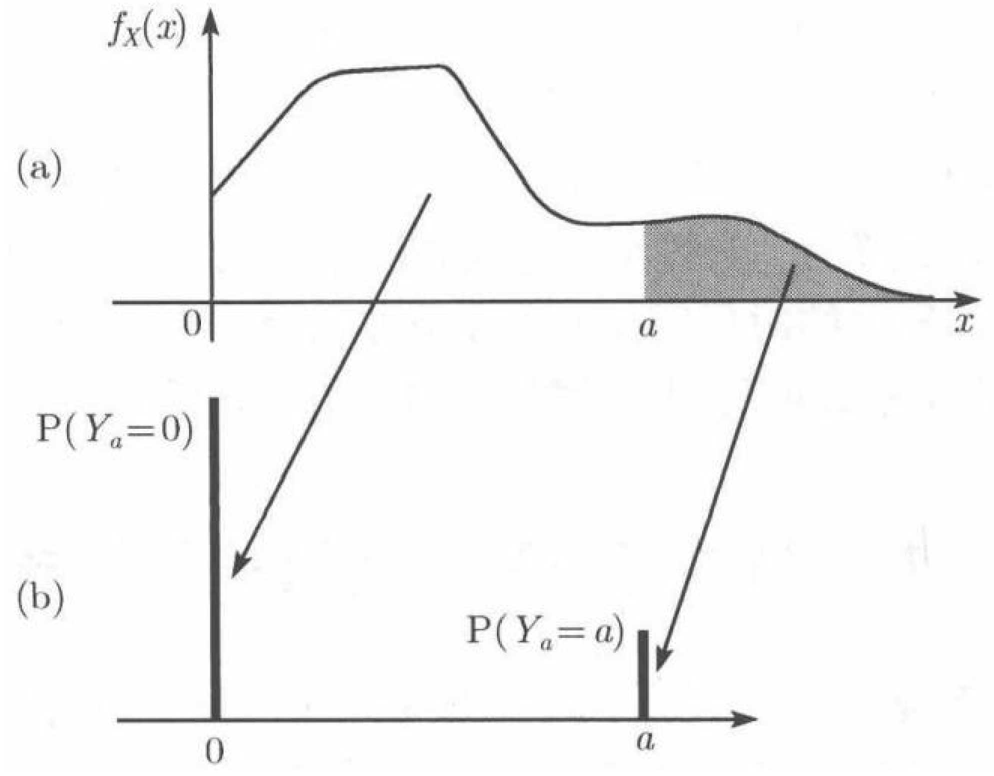
§1 马尔可夫和切比雪夫不等式

1.1 马尔可夫不等式

马尔可夫不等式

设随机变量 X 只取非负值, 则对任意 $a>0$,

$$P(X \geqslant a) \leqslant \frac{E[X]}{a}.$$



§1 马尔可夫和切比雪夫不等式

固定正数 a, 定义随机变量 Y_a,

$$Y_a = \begin{cases} 0, & \text{若 } X < a, \\ a, & \text{若 } X \geqslant a. \end{cases}$$

易知

$$Y_a \leqslant X$$

总成立, 从而

$$\mathrm{E}[Y_a] \leqslant \mathrm{E}[X].$$

另一方面

$$\mathrm{E}[Y_a] = a\mathrm{P}(Y_a = a) = a\mathrm{P}(X \geqslant a),$$

所以

$$a\mathrm{P}(X \geqslant a) \leqslant \mathrm{E}[X]$$

§1 马尔可夫和切比雪夫不等式

例 5.1 设 X 服从 $U[0,4]$ 的均匀分布. 易知 $\mathrm{E}[X]=2$. 由马尔可夫不等式可得

$$\mathrm{P}(X \geqslant 2) \leqslant \frac{2}{2}=1,\quad \mathrm{P}(X \geqslant 3) \leqslant \frac{2}{3}=0.67,\quad \mathrm{P}(X \geqslant 4) \leqslant \frac{2}{4}=0.5.$$

与真实概率进行比较

$$\mathrm{P}(X \geqslant 2)=0.5,\quad \mathrm{P}(X \geqslant 3)=0.25,\quad \mathrm{P}(X \geqslant 4)=0.$$

由马尔可夫不等式给出的上界与真实概率相差非常远.

§1 马尔可夫和切比雪夫不等式

1.2 切比雪夫不等式

切比雪夫不等式

设随机变量 X 的均值为 μ, 方差为 σ^2, 则对任意 $c > 0$,

$$\mathrm{P}(|X - \mu| \geqslant c) \leqslant \frac{\sigma^2}{c^2}.$$

令 $c = k\sigma$, 其中 k 是正数. 切比雪夫不等式的另一个版本是:

$$\mathrm{P}(|X - \mu| \geqslant k\sigma) \leqslant \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}.$$

一个随机变量的取值偏离其均值 k 倍标准差的概率最多 $1/k^2$.

§1 马尔可夫和切比雪夫不等式

考虑非负随机变量 $(X-\mu)^2$. 令 $a=c^2$, 使用马尔可夫不等式,

$$\mathrm{P}((X-\mu)^2 \geqslant c^2) \leqslant \frac{\mathrm{E}[(X-\mu)^2]}{c^2} = \frac{\sigma^2}{c^2}.$$

事件 $|X-\mu|^2 \geqslant c^2$ 等价于事件 $|X-\mu| \geqslant c$, 所以

$$\mathrm{P}(|X-\mu| \geqslant c) = \mathrm{P}(|X-\mu|^2 \geqslant c^2) \leqslant \frac{\sigma^2}{c^2}.$$

§1 马尔可夫和切比雪夫不等式

设 X 是连续型随机变量, 定义函数

$$g(x)=\begin{cases}0, & 若\ |x-\mu|<c\ 时, \\ c^2, & 若\ |x-\mu|\geqslant c\ 时.\end{cases}$$

对任意的 x, $(x-\mu)^2\geqslant g(x)$, 所以

$$\sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2 f_X(x)\mathrm{d}x\geqslant\int_{-\infty}^{\infty}g(x)f_X(x)\mathrm{d}x=c^2\mathrm{P}(|x-\mu|\geqslant c),$$

§1 马尔可夫和切比雪夫不等式

例 5.2 (续例 5.1) 设 X 服从 $U[0,4]$ 的均匀分布. 现在使用切比雪夫不等式来给出事件 $|X-2| \geqslant 1$ 的概率上界. 显然 $\sigma^2 = \frac{4}{3}$, $\mu = 2$, 则

$$\mathrm{P}(|X-2| \geqslant 1) \leqslant \frac{\sigma^2}{1} = \frac{4}{3}.$$

由于概率的值永远不超过 1, 所以这个不等式并不带来任何信息. □

§1 马尔可夫和切比雪夫不等式

现在看另一例子，设 X 服从参数 $\lambda = 1$ 的指数分布，则 $\mathrm{E}[X] = \mathrm{var}(X) = 1$. 对任意的 $c > 1$，使用切比雪夫不等式可得

$$\mathrm{P}(X \geqslant c) = \mathrm{P}(X - 1 \geqslant c - 1) \leqslant \mathrm{P}(|X - 1| \geqslant c - 1) \leqslant \frac{1}{(c-1)^2}.$$

而真实概率是 $\mathrm{P}(X \geqslant c) = \mathrm{e}^{-c}$. 可以看出由切比雪夫不等式给出的上界比较保守.

§1 马尔可夫和切比雪夫不等式

例 5.3 (切比雪夫不等式的上界) 设随机变量 X 取值空间是 $[a,b]$, 现在我们证明 $\sigma^2 \leqslant (b-a)^2/4$. 因此, 如果 σ^2 未知, 我们就可以用上界 $(b-a)^2/4$ 来代替切比雪夫不等式中的 σ^2, 即

$$\mathrm{P}(|x-\mu| \geqslant c) \leqslant \frac{(b-a)^2}{4c^2}, \quad 对任意的 c > 0.$$

§1 马尔可夫和切比雪夫不等式

对任意的常数 γ, 我们有

$$\mathrm{E}[(X-\gamma)^2]=\mathrm{E}[X^2]-2\mathrm{E}[X]\gamma+\gamma^2,$$ 多项式在 $\gamma=\mathrm{E}[X]$ 处达到极小.

$$\sigma^2=\mathrm{E}[(X-\mathrm{E}[X])^2]\leqslant \mathrm{E}[(X-\gamma)^2].$$

令 $\gamma=(a+b)/2$, 可得

$$\sigma^2\leqslant \mathrm{E}\left[\left(X-\frac{a+b}{2}\right)^2\right]=\mathrm{E}[(X-a)(X-b)]+\frac{(b-a)^2}{4}\leqslant\frac{(b-a)^2}{4}$$

§2 弱大数定律

2.1 弱大数定律

独立同分布随机变量序列 $X_1, X_2, \cdots$, 公共分布的均值为 μ, 方差为 σ^2.

定义样本均值

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

则

$$\mathrm{E}[M_n] = \frac{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]}{n} = \frac{n\mu}{n} = \mu.$$

再运用独立性可得

$$\mathrm{var}(M_n) = \frac{\mathrm{var}(X_1 + \cdots + X_n)}{n^2} = \frac{\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

$\mathrm{P}(|M_n - \mu| \geqslant \epsilon) \leqslant \dfrac{\sigma^2}{n\epsilon^2}$, 对任意的 $\epsilon > 0$ 成立.

§2 弱大数定律

2.1 弱大数定律

弱大数定律

设 $X_1, \cdots, X_n, \cdots$ 独立同分布, 其公共分布的均值为 μ, 则对任意的 $\epsilon > 0$, 当 $n \to \infty$ 时,

$$\mathrm{P}(|M_n - \mu| \geqslant \epsilon) = \mathrm{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geqslant \epsilon\right) \to 0.$$

可以证明: 当 n 充分大时, 频率以很大的概率落在 p 的 ϵ 邻域里.

§2 弱大数定律

例 5.4 (概率与频率) 在某个试验中, 考虑一个随机事件 A. 记 $p = \mathrm{P}(A)$ 为事件 A 发生的概率. 现在假定在 n 次独立重复的试验中, 记 M_n 为 n 次独立重复试验中事件 A 发生的次数占总试验次数 n 的比例, M_n 通常称为事件 A 的频率. 注意到

$$M_n = \frac{X_1 + \cdots + X_n}{n},$$

其中 $X_i = 1$ 表示事件 A 发生, 否则 $X_i = 0$. 特别地有 $\mathrm{E}[X_i] = p$. 运用弱大数定律可以证明：当 n 充分大时, 频率以很大的概率落在 p 的 ϵ 邻域里. 也就是说频率是 p 的一个很好的估计. 换句话说, 可以将事件 A 发生的频率解释为概率 p. □

§3 依概率收敛

3.1 数列的收敛

数列的收敛

设 $a_1, a_2, \cdots$ 是一实数数列, a 为一实数, 如果对任意的 $\epsilon > 0$, 存在正整数 n_0, 使得对所有的 $n \geqslant n_0$, 都有

$$|a_n - a| \leqslant \epsilon,$$

则称数列 a_n 收敛于 a, 记为 $\lim\limits_{n\to\infty} a_n = a$.

§3 依概率收敛

3.2 依概率收敛

依概率收敛

设 $Y_1, Y_2, \cdots$ 是随机变量序列 (不必相互独立), a 为一实数, 如果对任意的 $\epsilon > 0$, 都有

$$\lim_{n\to\infty} \mathrm{P}(|Y_n - a| \geqslant \epsilon) = 0,$$

则称 Y_n 依概率收敛于 a.

§3 依概率收敛

依概率收敛定义还可以这样描述：

对任意的 $\epsilon > 0$ 和 $\delta > 0$, 存在 n_0, 使得对所有的 $n \geqslant n_0$, 都有

$$P(|Y_n - a| \geqslant \epsilon) \leqslant \delta.$$

ϵ 为精度, δ 为置信水平

给定精度和置信水平下, 在 n 充分大时, Y_n 等于 a.

§3 依概率收敛

设 $Y \sim Exp(\lambda)$. 任意正整数 n. $Y_n = \frac{Y}{n}$. (Y_n 序列不独立)

对 $\forall \varepsilon > 0$.

$$P(|Y_n - 0| \geq \varepsilon) = P(Y_n \geq \varepsilon) = P(Y \geq n\varepsilon) = e^{-n\varepsilon}.$$

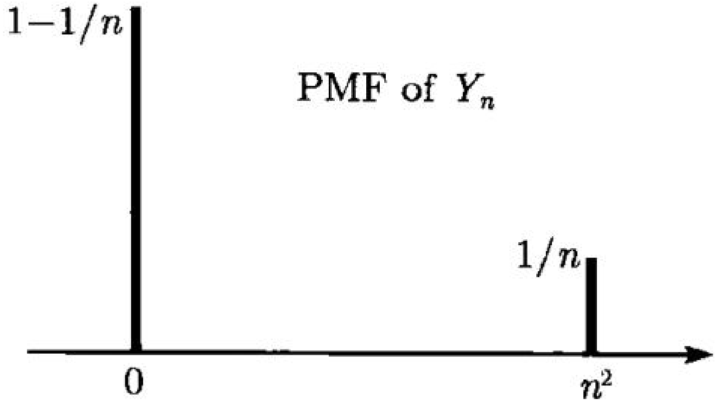
$$\lim_{n\to\infty} P(|Y_n - 0| \geq \varepsilon) = \lim_{n\to\infty} e^{-n\varepsilon} = 0.$$

$\therefore$ Y_n 依概率收敛于0.

§3 依概率收敛

例 5.8 考虑离散随机变量序列 Y_n, 其分布列为

$$P(Y_n = y) = \begin{cases} 1 - \dfrac{1}{n}, & \text{若 } y = 0, \\ \dfrac{1}{n}, & \text{若 } y = n^2, \\ 0, & \text{其他}. \end{cases}$$



随机变量 Y_n 的分布列

对任意的 $\epsilon > 0$, 有 $\lim\limits_{n\to\infty} P(|Y_n| \geqslant \epsilon) = \lim\limits_{n\to\infty} \dfrac{1}{n} = 0.$

另一方面, 当 $n \to \infty$ 时, $E[Y_n] = n^2/n = n \to \infty.$

§4 中心极限定理

中心极限定理

设 $X_1, X_2\cdots$ 是独立同分布的随机变量序列，序列的每一项的均值为 μ，方差为 σ^2. 记

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}.$$

则 Z_n 的分布函数的极限分布为标准正态分布函数

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-z^2/2}\mathrm{d}z,$$

即

$$\lim_{n\to\infty} \mathrm{P}(Z_n \leqslant x) = \Phi(x), \quad \text{对任意的 } x \text{ 成立}.$$

§4 中心极限定理

4.1 基于中心极限定理的近似

基于中心极限定理的正态近似

令 $S_n = X_1 + \cdots + X_n$, 其中 $X_1, \cdots, X_n$ 是独立同分布, 均值为 μ, 方差为 σ^2 的随机变量序列. 当 n 充分大时, 概率 $\mathrm{P}(S_n \leqslant c)$ 可以通过将 S_n 视为正态随机变量来近似计算. 步骤如下:

(1) 计算 S_n 的均值 $n\mu$ 和方差 $n\sigma^2$;

(2) 计算归一化后的值 $z = (c - n\mu)/(\sqrt{n}\sigma)$;

(3) 计算近似值

$$\mathrm{P}(S_n \leqslant c) \approx \Phi(z),$$

其中 $\Phi(z)$ 可从标准正态分布函数表查得.

§4 中心极限定理

例 5.11 (选举问题) 现在重新考虑例 5.5 的选举问题. 设对 n 个选民进行调查, 记录下他们赞成某候选人的比例 M_n,

$$M_n = \frac{X_1 + \cdots + X_n}{n},$$

其中 X_i 是被调查的第 i 个选民的态度, $X_i = 1$ 表示选民 i 支持某候选人, $X_i = 0$ 表示选民 i 反对某候选人. 假设 p 是这个候选人在全体选民中的支持率, 则 X_i 是服从参数为 p 的伯努利随机变量. 故 M_n 的均值为 p, 方差为 $p(1-p)/n$. 利用中心极限定理, M_n 近似服从正态分布.

§4 中心极限定理

计算概率 $\mathrm{P}(|M_n - p| \geqslant \epsilon)$, ϵ 是估计精度, 由正态分布的对称性, 可得

$$\mathrm{P}(|M_n - p| \geqslant \epsilon) \approx 2\mathrm{P}(M_n - p \geqslant \epsilon).$$

假设 $M_n - p$ 有最大的方差, 即当 $p = 1/2$ 时, 方差为 $1/(4n)$.

$$z = \frac{\epsilon}{\sqrt{1/(4n)}},$$

所以

$$\mathrm{P}(M_n - p \geqslant \epsilon) \leqslant 1 - \Phi(z) = 1 - \Phi(2\epsilon\sqrt{n}).$$

§4 中心极限定理

现在考虑另一个问题. 如果希望估计 M_n 与真值 p 的差距为 0.01 之内的概率至少是 0.95, 则样本容量 n 应该多大?

我们假设最坏的情况发生, 此时 M_n 的方差达到最大,

$$2-2\Phi(2\cdot 0.01\cdot\sqrt{n})\leqslant 0.05,$$

即

$$\Phi(2\cdot 0.01\cdot\sqrt{n})\geqslant 0.975.$$

$$n\geqslant\frac{1.96^2}{4\cdot(0.01)^2}=9\ 604.$$

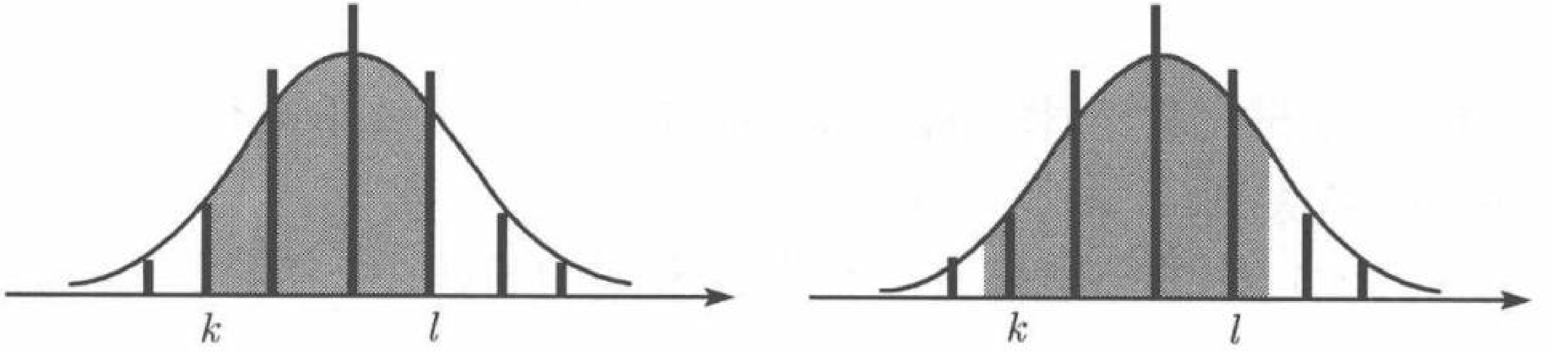
§4 中心极限定理

4.2 二项分布的棣莫弗-拉普拉斯近似

二项分布的棣莫弗 – 拉普拉斯近似

设 S_n 是服从参数为 n 和 p 的二项分布, n 充分大, k 和 l 是非负整数, 则

$$\mathrm{P}(k \leqslant S_n \leqslant l) \approx \Phi\left(\frac{l+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right).$$



(a) (b)

§5 强大数定律

5.1 强大数定律

强大数定律

设 $X_1, X_2, \cdots, X_n$ 是均值为 μ 的独立同分布随机变量序列，则样本均值 $M_n = (X_1 + X_2 + \cdots + X_n)/n$ 以概率 1 收敛于 μ，即

$$\mathrm{P}\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right) = 1.$$

§5 强大数定律

5.2 以概率1收敛

以概率 1 收敛

设 $Y_1, Y_2, \cdots$ 是某种概率模型下的随机变量序列 (但不必独立), c 是某个实数, 如果

$$\mathrm{P}(\lim_{n\to\infty} Y_n = c) = 1,$$

则称 Y_n 以概率 1 (或几乎处处) 收敛于 c.

§5 强大数定律

例 5.14 设 $X_1, X_2, \cdots$ 是独立随机变量序列, X_i 的公共分布是区间 $[0,1]$ 中的均匀分布. 令 $Y_n = \min\{X_1, \cdots, X_n\}$. 下面证明 Y_n 以概率 1 收敛于 0.

Y_n 是非增的, 即对所有的 n, $Y_{n+1} \leqslant Y_n$. 序列 Y_n 有下界 0 极限记为 Y.

固定 $\epsilon > 0$, 如果 $Y \geqslant \epsilon$, 则对所有的 i 都有 $X_i \geqslant \epsilon$,

$$\mathrm{P}(Y \geqslant \epsilon) \leqslant \mathrm{P}(Y_1 \geqslant \epsilon, \cdots, Y_n \geqslant \epsilon) = (1-\epsilon)^n.$$

$$\mathrm{P}(Y \geqslant \epsilon) \leqslant \lim_{n\to\infty} (1-\epsilon)^n = 0.$$

故 $\mathrm{P}(Y > 0) = 0$, 从而 $\mathrm{P}(Y = 0) = 1$. 所以 Y_n 以概率 1 收敛于 0.

作业

9. 假设计算机系统每天至少出现一次死机的概率为 5%, 而且在不同天里, 出现死机的事件是相互独立的. 求在 50 天之内计算机至少有 45 天没有死机的概率.

(a) 试用二项分布的正态近似方法来计算.

(b) 试用二项分布的泊松近似方法来计算.