

定理6.13 (Bernstein不等式). 设 X_1, \dots, X_n 是独立同分布的随机变量, 其均值为 μ 和方差为 σ^2 , 若存在常数 $b > 0$, 使得对任意正整数 $m \geq 2$ 有 $E[X_i^m] \leq m!b^{m-2}\sigma^2/2$, 那么我们有

$$\Pr \left[\frac{1}{n} \sum_{i=1}^n (X_n - \mu) \geq \epsilon \right] \leq \exp \left(-\frac{n\epsilon^2}{2\sigma^2 + 2b\epsilon} \right).$$

Proof. 对任意 $t > 0$, 根据 Chernoff 方法有

$$\Pr \left[\frac{1}{n} \sum_{i=1}^n (X_n - \mu) \geq \epsilon \right] \leq e^{-nt\epsilon} E \left[\exp \left(\sum_{i=1}^n (X_i - \mu) \right) \right] = e^{-nt\epsilon - n\mu t} (E[e^{tX_1}])^n$$

利用公式 $\ln z \leq z - 1$ 有

$$\ln E[e^{tX_1}] \leq E[e^{tX}] - 1 = \sum_{m=1}^{\infty} E[X^m] \frac{t^m}{m!} \leq t\mu + \frac{t^2\sigma^2}{2} \sum_{m=2}^{\infty} (bt)^{m-2} = t\mu + \frac{t^2\sigma^2}{2(1-bt)}.$$

由此可得

$$\Pr \left[\frac{1}{n} \sum_{i=1}^n (X_n - \mu) \geq \epsilon \right] \leq \exp \left(-nt\epsilon + \frac{t^2n\sigma^2}{2(1-bt)} \right)$$

取 $t = \epsilon/(\sigma^2 + b\epsilon)$ 完成证明. □

课题练习. 给出 Bernstein 不等式的 $1 - \delta$ 表述.

6.6 应用: 随机投影 (Random Projection)

设高维空间 \mathbb{R}^d 有 n 个点 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ (d 非常大, 如 100 万或 1 亿). 处理这样一个高维的问题很难, 实际中的一种解决方案是能否找到一个保距变换: $f: \mathbb{R}^d \rightarrow \mathbb{R}^k$ ($k \ll d$), 使得以较大概率有

$$(1 - \epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \leq \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2 \leq (1 + \epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2.$$

随机投影广泛应用于高维的机器学习问题, 例如最近邻、 k -近邻、降维、聚类等问题.

随机投影可以简单的表示为

$$f(\mathbf{x}) = \mathbf{x}P/c,$$

其中 P 是一个 $d \times k$ 的随机矩阵, 其每个元素之间相互独立, c 为一常数 (根据随机矩阵 P 确定). 下面介绍三种常见的随机矩阵:

- $P = (p_{ij})_{d \times k} \in \mathbb{R}^{d \times k}$, $p_{ij} \sim \mathcal{N}(0, 1)$, 此时 $c = \sqrt{k}$;
- $P = (p_{ij})_{d \times k} \in \{-1, 1\}^{d \times k}$, p_{ij} 为 Rademacher 随机变量, 即 $\Pr(p_{ij} = 1) = \Pr(p_{ij} = -1) = 1/2$, 此时 $c = \sqrt{k}$;
- $P = (p_{ij})_{d \times k} \in \{-1, 0, 1\}^{d \times k}$, 满足 $\Pr(p_{ij} = 1) = \Pr(p_{ij} = -1) = 1/6$ 和 $\Pr(p_{ij} = 0) = 2/3$, 此时 $c = \sqrt{k/3}$. 【主要用于 sparse 投影, 减少计算量】

下面我们重点理论分析 Gaussian 随机变量, 其它随机变量可参考相关资料, 对 Gaussian 随机变量, 这里介绍著名的 Johnson–Lindenstrauss 引理, 简称 JL 引理.

引理6.5. 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 为 \mathbb{R}^d 空间的 n 个点, 随机矩阵 $P = (p_{ij})_{d \times k} \in \mathbb{R}^{d \times k}$, $p_{ij} \sim \mathcal{N}(0, 1)$ 且每个元素相互独立, 令

$$\mathbf{y}_i = f(\mathbf{x}_i) = \mathbf{x}_i P / \sqrt{k}, \quad i \in [n]$$

将 d 维空间中 n 个点 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 通过随机矩阵 P 投影到 k 维空间. 对任意 $\epsilon \in (0, 1/2)$, 当 $k \geq 8 \log 2n / (\epsilon^2 - \epsilon^3)$ 时至少以 $1/2$ 的概率有

$$(1 - \epsilon) \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \leq \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq (1 + \epsilon) \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \quad (i, j \in [n]).$$

Proof. 下面分三步证明 J-L 引理.

第一步: 对任意非零 $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, 首先证明

$$E \left[\left\| \mathbf{x} P / \sqrt{k} \right\|_2^2 \right] = \|\mathbf{x}\|_2^2,$$

即在期望的情况下, 随机投影变换前后的点到原点的距离相同. 根据 $P = (p_{ij})_{d \times k}$ ($p_{ij} \sim \mathcal{N}(0, 1)$) 有

$$\begin{aligned} E \left[\left\| \frac{\mathbf{x} P}{\sqrt{k}} \right\|_2^2 \right] &= E \left[\sum_{j=1}^k \left(\sum_{i=1}^d \frac{x_i p_{ij}}{\sqrt{k}} \right)^2 \right] = \sum_{j=1}^k \frac{1}{k} E \left[\left(\sum_{i=1}^d x_i p_{ij} \right)^2 \right] \\ &= \sum_{j=1}^k \frac{1}{k} \sum_{i=1}^d x_i^2 = \frac{1}{k} \sum_{j=1}^k \|\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2. \end{aligned}$$

第二步: 对任意非零 $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, 证明

$$\Pr \left[\left\| \frac{\mathbf{x} P}{\sqrt{k}} \right\|_2^2 \geq (1 + \epsilon) \|\mathbf{x}\|_2^2 \right] \leq \exp(-(\epsilon^2 - \epsilon^3)k/4).$$

将矩阵 P 表示为 $P = (P_1, P_2, \dots, P_k)$, 其中 P_i ($i \in [k]$) 是一个 $d \times 1$ 的列向量, 令 $v_j = \mathbf{x} P_j / \|\mathbf{x}\|_2$, 即

$$(v_1, v_2, \dots, v_k) = \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_2} P_1, \frac{\mathbf{x}}{\|\mathbf{x}\|_2} P_2, \dots, \frac{\mathbf{x}}{\|\mathbf{x}\|_2} P_k \right).$$

根据 Gaussian 分布的性质有 $v_j \sim \mathcal{N}(0, 1)$, 且 v_1, v_2, \dots, v_k 是 k 个独立的随机变量. 对任意 $t \in (0, 1/2)$, 根据 Chernoff 方法有

$$\begin{aligned} \Pr \left[\left\| \frac{\mathbf{x} P}{\sqrt{k}} \right\|_2^2 \geq (1 + \epsilon) \|\mathbf{x}\|_2^2 \right] &= \Pr \left[\left\| \frac{\mathbf{x} P}{\|\mathbf{x}\|_2} \right\|_2^2 \geq (1 + \epsilon)k \right] \\ &= \Pr \left[\sum_{j=1}^k v_j^2 \geq (1 + \epsilon)k \right] \leq e^{-(1+\epsilon)kt} \left(E[e^{t \sum_{j=1}^k v_j^2}] \right) = e^{-(1+\epsilon)kt} \left(E[e^{t v_1^2}] \right)^k. \end{aligned}$$

对标准 Gaussian 分布有

$$E[e^{t v_1^2}] = \int_{-\infty}^{+\infty} \frac{e^{t u^2}}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \int_{-\infty}^{+\infty} \frac{e^{-\frac{u^2}{2}(1-2t)}}{\sqrt{2\pi}} du = \frac{1}{\sqrt{1-2t}},$$

代入可得

$$\Pr \left[\left\| \mathbf{x}P/\sqrt{k} \right\|_2^2 \geq (1+\epsilon)\|\mathbf{x}\|_2^2 \right] \leq \left(\frac{e^{-2(1+\epsilon)t}}{1-2t} \right)^{k/2}.$$

上式右边对 t 求最小解得 $t_{\min} = \frac{\epsilon}{2(1+\epsilon)}$, 代入可得

$$\Pr \left[\left\| \mathbf{x}P/\sqrt{k} \right\|_2^2 \geq (1+\epsilon)\|\mathbf{x}\|_2^2 \right] \leq ((1+\epsilon)e^{-\epsilon})^{k/2}.$$

设 $f(\epsilon) = \ln(1+\epsilon)$, 根据 $\epsilon \in (0, 1/2)$ 有

$$f'(\epsilon) = \frac{1}{1+\epsilon}, f''(\epsilon) = -\frac{1}{(1+\epsilon)^2}, f'''(\epsilon) = \frac{2}{(1+\epsilon)^3} \leq 2.$$

根据泰勒中值定理有

$$f(\epsilon) = f(0) + f'(0)\epsilon + \frac{f''(0)\epsilon^2}{2!} + \frac{f'''(\xi)\epsilon^3}{3!} \leq \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^2}{3} \leq \epsilon - \frac{\epsilon^2 - \epsilon^3}{2}.$$

于是得到

$$\Pr \left[\left\| \frac{\mathbf{x}P}{\sqrt{k}} \right\|_2^2 \geq (1+\epsilon)\|\mathbf{x}\|_2^2 \right] \leq e^{-k(\epsilon^2 - \epsilon^3)/4}.$$

同理可证

$$\Pr \left[\left\| \frac{\mathbf{x}P}{\sqrt{k}} \right\|_2^2 \leq (1-\epsilon)\|\mathbf{x}\|_2^2 \right] \leq e^{-k(\epsilon^2 - \epsilon^3)/4}.$$

第三步: 对任意给定 $i \neq j$, 根据第二步的结论可知

$$\begin{aligned} \Pr[\|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \geq (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2] &\leq e^{-k(\epsilon^2 - \epsilon^3)/4}, \\ \Pr[\|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq (1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2] &\leq e^{-k(\epsilon^2 - \epsilon^3)/4}. \end{aligned}$$

由于 $i, j \in [n]$, 因此共有 $n(n-1)$ 对 (i, j) , 根据 Union 不等式有

$$\Pr \left[\exists i \neq j: \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \geq (1+\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \quad \text{或} \quad \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \leq (1-\epsilon)\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 \right] \leq 2n^2 e^{-k(\epsilon^2 - \epsilon^3)/4},$$

设 $2n^2 e^{-k(\epsilon^2 - \epsilon^3)/4} \leq 1/2$, 求解 $k \geq 8 \log 2n/(\epsilon^2 - \epsilon^3)$. 引理得证. \square

7 大数定律及中心极限定理

7.1 大数定律

给定随机变量 X_1, X_2, \dots, X_n , 这些随机变量的均值 (算术平均值) 为

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

当 n 非常大时, 大数定律考虑随机变量的均值是否具有稳定性.

定义7.1 (依概率收敛). 设 $X_1, X_2, \dots, X_n, \dots$ 是一随机变量序列, a 是一常数, 如果对任意 $\epsilon > 0$ 有

$$\lim_{n \rightarrow \infty} \Pr\{|X_n - a| < \epsilon\} = 1 \quad \text{或} \quad \lim_{n \rightarrow \infty} \Pr\{|X_n - a| > \epsilon\} = 0,$$

则称随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 依概率收敛于 a , 记 $X_n \xrightarrow{P} a$.

问题: 与数列极限的区别? 下面我们给出依概率的性质:

- 1) 若 $X_n \xrightarrow{P} a$ 且函数 $g: \mathbb{R} \rightarrow \mathbb{R}$ 在 $X = a$ 点连续, 则 $g(X_n) \xrightarrow{P} g(a)$.
- 2) 若 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, 函数 $g: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ 在点 $(X, Y) = (a, b)$ 处连续, 则 $g(X_n, Y_n) \xrightarrow{P} g(a, b)$.

例如: 如果 $X_n \xrightarrow{P} a$ 和 $Y_n \xrightarrow{P} b$, 那么 $X_n + Y_n \xrightarrow{P} a + b$ 和 $X_n Y_n \xrightarrow{P} ab$.

定理7.1 (大数定律). 若随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 满足

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n E[X_i],$$

则称 $\{X_n\}$ 服从大数定律.

大数定理刻画了随机变量的均值 (算术平均值) 依概率收敛于期望的均值 (算术平均值). 下面介绍几种大数定律:

定理7.2 (马尔可夫 Markov 大数定律). 如果随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 满足

$$\frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n X_i \right) \rightarrow 0 \quad n \rightarrow \infty,$$

则 $\{X_n\}$ 服从大数定理.

马尔可夫大数定律不要求随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 相互独立或同分布, 其证明直接通过 Chebyshev 不等式有

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - E[X_i]) \right| \geq \epsilon \right] \leq \frac{1}{n^2 \epsilon^2} \text{Var} \left(\sum_{i=1}^n X_i \right) \rightarrow 0 \quad n \rightarrow \infty.$$

定理7.3 (切比雪夫 Chebyshev 大数定律). 设随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 且存在常数 $c > 0$ 使得 $\text{Var}(X_n) \leq c$, 则 $\{X_n\}$ 服从大数定律.

此处独立的随机变量可以修改为‘不相关随机变量’. 证明直接通过切比雪夫不等式

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - E[X_i]) \right| \geq \epsilon \right] \leq \frac{1}{\epsilon^2 n^2} \text{Var} \left(\sum_{i=1}^n X_i \right) \leq \frac{c}{n\epsilon^2} \rightarrow 0 \quad n \rightarrow \infty.$$

定理7.4 (辛钦 Khintchine 大数定律). 设 $X_1, X_2, \dots, X_n, \dots$ 为独立同分布随机变量序列, 且每个随机变量的期望 $E[X_i] = \mu$ 存在, 则 $\{X_n\}$ 服从大数定律.

辛钦大数定律不要求方差一定存在, 其证明超出了本书范围.

定理7.5 (Bernoulli 大数定律). 设随机变量序列 $X_n \sim B(n, p)$ ($p > 0$), 对任意 $\epsilon > 0$ 有

$$\lim_{n \rightarrow \infty} \Pr \left[\left| \frac{X_n}{n} - p \right| \geq \epsilon \right] = 0,$$

即 $X_n/n \xrightarrow{P} p$.

定理的证明依据二项分布的性质: 独立同分布随机变量 Y_1, Y_2, \dots, Y_n 满足 $Y_i \sim \text{Ber}(p)$, 则

$$X_n = \sum_{i=1}^n Y_i \sim B(n, p).$$

于是得到

$$\lim_{n \rightarrow \infty} \Pr \left[\left| \frac{X_n}{n} - p \right| \geq \epsilon \right] = \lim_{n \rightarrow \infty} \Pr \left[\left| \frac{1}{n} \sum_{i=1}^n Y_i - E[Y_i] \right| \geq \epsilon \right] \leq \frac{1}{\epsilon^2 n^2} \text{Var} \left(\sum_{i=1}^n Y_i \right) = \frac{p(1-p)}{\epsilon^2 n} \rightarrow 0.$$

如何判断随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 满足大数定律:

- 若随机变量独立同分布, 则利用辛钦大数定律查看期望是否存在;
- 对非独立同分布随机变量, 则利用 Markov 大数定律判断方差是否趋于零.

例7.1. 独立的随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 满足 $\Pr\{X_n = n^{1/4}\} = \Pr\{X_n = -n^{1/4}\} = 1/2$. 证明 $\{X_n\}$ 服从大数定律.

Proof. 根据题意可得 $E[X_i] = 0$, 以及 $\text{Var}(X_i) = E[X_i^2] = i^{1/2}$, 根据 Chebyshev 不等式和独立性有

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq \epsilon \right] \leq \frac{1}{n^2 \epsilon^2} \text{Var} \left(\sum_{i=1}^n X_i \right) = \frac{1}{n^2 \epsilon^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{\epsilon^2} \frac{1}{n^2} \sum_{i=1}^n i^{1/2} \leq \frac{1}{\epsilon^2 \sqrt{n}}$$

再根据

$$\sum_{i=1}^n i^{1/2} \leq \sum_{i=1}^n \int_i^{i+1} i^{1/2} dx \leq \sum_{i=1}^n \int_i^{i+1} x^{1/2} dx = \int_1^{n+1} x^{1/2} dx = 2((n+1)^{3/2} - 1)/3$$

由此可得当 $n \rightarrow +\infty$ 时有

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq \epsilon \right] \leq \frac{2((n+1)^{3/2} - 1)/3}{\epsilon^2 n^2} \rightarrow 0$$

□

大数定律小结:

- Markov 大数定律: 若随机变量序列 $\{X_i\}$ 满足 $\text{Var}(\sum_{i=1}^n X_i)/n^2 \rightarrow 0$, 则满足大数定律;
- Chebyshev 大数定律: 若独立随机变量序列 $\{X_i\}$ 满足 $\text{Var}(X_i) \leq c$, 则满足大数定律;
- Khintchine 大数定律: 若独立同分布随机变量序列 $\{X_i\}$ 期望存在, 则满足大数定律;
- Bernoulli 大数定律: 对二项分布 $X_n \sim B(n, p)$, 有 $X_n/n \xrightarrow{P} p$.

7.2 中心极限定理

对独立的随机变量序列 $X_1, X_2, \dots, X_n, \dots$, 我们考虑标准化后随机变量

$$Y_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)}{\sqrt{\text{Var}(\sum_{i=1}^n X_i)}}$$

的极限分布是否为服从正态分布. 首先介绍依分布收敛.

定义7.2. 设随机变量 Y 的分布函数为 $F_Y(y) = \Pr(Y \leq y)$, 以及随机变量序列 $Y_1, Y_2, \dots, Y_n, \dots$ 的分布函数分别为 $F_{Y_n}(y) = \Pr(Y_n \leq y)$, 如果

$$\lim_{n \rightarrow \infty} \Pr[Y_n \leq y] = \Pr[Y \leq y], \quad \text{即} \quad \lim_{n \rightarrow \infty} F_{Y_n}(y) = F_Y(y),$$

则称随机变量序列 $Y_1, Y_2, \dots, Y_n, \dots$ 依分布收敛于 Y , 记 $Y_n \xrightarrow{d} Y$.

下面介绍独立同分布中心极限定理, 又被称为林德贝格-勒维 (Lindeberg-Lévy) 中心极限定理”:

定理7.6. 设独立同分布的随机变量 $X_1, X_2, \dots, X_n, \dots$ 的期望 $E(X_1) = \mu$ 和方差 $\text{Var}(X_1) = \sigma^2$, 则

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

前面介绍标准正态分布的分布函数为 $\Phi(x)$, 则上述中心极限定理等价于

$$\lim_{n \rightarrow \infty} \Pr[Y_n \leq y] = \Phi(y).$$

随机变量 Y_n 是随机变量 X_1, X_2, \dots, X_n 的标准化, 其极限服从标准正态分布. 当 n 足够大时近似有 $Y_n \sim \mathcal{N}(0, 1)$, 中心极限定理的变形公式为

$$\sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(n\mu, n\sigma^2), \quad \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/n).$$

大数定律给出了当 $n \rightarrow \infty$ 时随机变量平均值 $\frac{1}{n} \sum_{i=1}^n X_i$ 的趋势, 而中心极限定理给出了 $\frac{1}{n} \sum_{i=1}^n X_i$ 的具体分布.

例7.2. 设一电压接收器同时接收到 20 个独立同分布的信号电压 V_k ($k \in [20]$), 且 $V_k \sim U(0, 10)$, 求电压和大于 105 的概率.

解. 根据题意可知独立同分布的随机变量 V_1, V_2, \dots, V_{20} 服从均匀分布 $U(0, 10)$, 于是有 $E(V_k) = 5$ 和 $\text{Var}(V_k) = 100/12 = 25/3$. 设 $V = \sum_{k=1}^{20} V_k$, 则有

$$E(V) = 100 \quad \text{Var}(V) = 500/3.$$

根据中心极限定理近似有

$$\frac{V - E(V)}{\sqrt{\text{Var}(V)}} = \frac{V - 100}{\sqrt{500/3}} \sim \mathcal{N}(0, 1).$$

根据标准正态分布的分布函数 $\Phi(x)$ 有

$$\Pr(V \geq 105) = \Pr\left(\frac{V - 100}{\sqrt{500/3}} \geq \frac{105 - 100}{\sqrt{500/3}}\right) = \Pr\left(\frac{V - 100}{\sqrt{500/3}} \geq 0.387\right) = 1 - \Phi(0.387).$$

查表完成证明. □

例7.3. 某产品装箱, 每箱重量是随机的, 假设其期望是 50 公斤, 标准差为 5 公斤. 若最大载重量为 5 吨, 问每车最多可装多少箱能以 0.997 以上的概率保证不超载?

解. 假设最多可装 n 箱不超重, 用 X_i 表示第 i 箱重量 ($i \in [n]$), 有 $E(X_i) = 50$ 和 $\text{Var}(X_i) = 25$. 设总重量 $X = \sum_{i=1}^n X_i$, 则有 $E(X) = 50n$ 和 $\text{Var}(X) = 25n$. 由中心极限定理近似有

$$(X - 50n)/\sqrt{25n} \sim \mathcal{N}(0, 1).$$

根据标准正态分布的分布函数 $\Phi(x)$ 有

$$\Pr(X \leq 5000) = \Pr\left(\frac{X - 50n}{\sqrt{25n}} \leq \frac{5000 - 50n}{\sqrt{25n}}\right) = \Phi\left(\frac{5000 - 50n}{\sqrt{25n}}\right) > 0.977 = \Phi(2).$$

根据分布函数的单调性有

$$\frac{1000 - 10n}{\sqrt{n}} > 2 \implies 1000n^2 - 2000n + 1000^2 > 4n.$$

求解可得 $n > 102.02$ 或 $n < 98.02$, 根据由题意可知 $n = 98$. □

下面介绍另一个中心极限定理: 棣莫弗-拉普拉斯 (De Moivre-Laplace) 中心极限定理:

推论7.1. 设随机变量 $X_n \sim B(n, p)$, 则

$$Y_n = \frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

由此中心极限定理可知: 当 n 非常大时随机变量 $X_n \sim B(n, p)$ 满足 $X_n \overset{\text{近似}}{\sim} \mathcal{N}(np, np(1-p))$, 从而有如下近似估计:

$$\Pr[X_n \leq y] = \Pr\left[\frac{X_n - np}{\sqrt{np(1-p)}} \leq \frac{y - np}{\sqrt{np(1-p)}}\right] \approx \Phi\left(\frac{y - np}{\sqrt{np(1-p)}}\right).$$

针对上式, 可以考虑三种问题: i) 已知 n 和 $\Pr[X_n \leq y]$, 求 y ; ii) 已知 n 和 y , 求 $\Pr[X_n \leq y]$; iii) 已知 y 和 $\Pr[X_n \leq y]$, 求 n . 下面看三个例子:

例7.4. 车间有 200 台独立工作的车床, 每台工作的概率为 0.6, 工作时每台耗电 1 千瓦, 至少供电多少千瓦才能以 99.9% 的概率保证正常生产.

解. 设工作的车床数为 X , 则 $X \sim B(200, 0.6)$. 设至少供电 y 千瓦. 根据棣莫弗-拉普拉斯中心定理近似有 $X \sim \mathcal{N}(120, 48)$, 进一步有

$$\Pr(X \leq y) \geq 0.999 \Rightarrow \Pr\left(\frac{X - 120}{\sqrt{48}} \leq \frac{y - 120}{\sqrt{48}}\right) \approx \Phi\left(\frac{y - 120}{\sqrt{48}}\right) \geq 0.999 = \Phi(3.1).$$

所以有 $\frac{y-120}{\sqrt{48}} \geq 3.1$, 求解可得 $y \geq 141$. □

例7.5. 系统由 100 个相互独立的部件组成, 每部件损坏率为 0.1, 至少 85 个部件正常工作系统才能运行, 求系统运行的概率.

解. 设 X 是损坏的部件数, 则 $X \sim B(100, 0.1)$, 有 $E(X) = 10$ 和 $\text{Var}(X) = 9$. 根据棣莫弗-拉普拉斯中心定理近似有 $X \sim \mathcal{N}(10, 9)$, 求系统运行的概率为

$$\Pr(X \leq 15) = \Pr\left(\frac{X - 10}{\sqrt{9}} \leq \frac{15 - 10}{\sqrt{9}}\right) \approx \Phi(5/3).$$

□

例7.6. 一次电视节目调查中调查 n 人, 其中 k 人观看了电视节目, 因此收看比例 k/n 作为电视节目收视率 p 的估计, 要以 90% 的概率有 $|k/n - p| \leq 0.05$ 成立, 需要调查多少对象?

解. 用 X_n 表示 n 个调查对象中收看节目的人数, 则有 $X_n \sim B(n, p)$. 根据棣莫弗-拉普拉斯中心定理近似有 $(X_n - np)/\sqrt{np(1-p)} \sim \mathcal{N}(0, 1)$, 进一步有

$$\begin{aligned} \Pr\left[\left|\frac{X_n}{n} - p\right| \leq 0.05\right] &= \Pr\left[\frac{|X_n - np|}{n} \leq 0.05\right] = \Pr\left[\frac{|X_n - np|}{\sqrt{np(1-p)}} \leq \frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right] \\ &= \Phi\left(\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right) \end{aligned}$$

对于标准正太分布函数有 $\Phi(-\alpha) = 1 - \Phi(\alpha)$ 以及 $p(1-p) \leq 1/4$, 于是有

$$\Pr\left[\left|\frac{X_n}{n} - p\right| \leq 0.05\right] = 2\Phi\left(\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 > 2\Phi(\sqrt{n}/10) - 1 > 0.9.$$

所以 $\Phi(\sqrt{n}/10) \geq 0.95$, 查表解得 $n \geq 271$. □

对独立不同分布的随机变量序列, 有李雅普诺夫 (Lyapunov) 中心极限定理:

定理7.7. 设独立随机变量 $X_1, X_2, \dots, X_n, \dots$ 的期望 $E[X_n] = \mu_n$ 和方差 $\text{Var}(X_n) = \sigma_n^2 > 0$. 记 $B_n^2 = \sum_{k=1}^n \sigma_k^2$, 若存在 $\delta > 0$, 当 $n \rightarrow \infty$ 时有

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^n E[|X_k - \mu_k|^{2+\delta}] \rightarrow 0$$

成立, 则有

$$Y_n = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n E(X_k)}{\sqrt{\text{Var}(\sum_{k=1}^n X_k)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

中心极限定理小结:

- 独立同分布中心极限定理: 若 $E[X_k] = \mu$ 和 $\text{Var}(X_k) = \sigma^2$, 则 $\sum_{k=1}^n X_k \xrightarrow{d} \mathcal{N}(n\mu, n\sigma^2)$;
- 棣莫弗-拉普拉斯中心极限定理: 若 $X_k \sim B(k, p)$, 则 $X_k \xrightarrow{d} \mathcal{N}(np, np(1-p))$;
- 独立不同分布中心极限定理: 李雅普诺夫定理.