对上式右边取最小值求解 $t = 4\epsilon/(b-a)^2$, 然后带入上式可得:

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E[X_i] \geqslant \epsilon\right] \leqslant \exp(-2n\epsilon^2/(b-a)^2).$$

从而完成证明.

7.2.3 Gaussian 和 Sub-Gaussian 随机变量不等式

首先考虑独立同分布的 Gaussian 随机变量:

定理 7.9 设随机变量 $X_1, \ldots, X_n$ 相互独立、且服从 $X_i \sim \mathcal{N}(\mu, \sigma)$, 对任意 $\epsilon > 0$ 有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \geqslant \epsilon\right] = \Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \leqslant -\epsilon\right] \leqslant \frac{1}{2}\exp(-n\epsilon^2/2\sigma^2).$$

证明 对随机变量 $X_i \sim \mathcal{N}(\mu, \sigma)$, 根据正太分布的性质有

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \sim \mathcal{N}(0, \sigma^2/n) \quad \Rightarrow \quad \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu) \sim \mathcal{N}(0, 1).$$

若 $X' \sim \mathcal{N}(0, 1)$, 对任意 $\epsilon > 0$, 根据以前的定理有

$$P(X' \geqslant \epsilon) \leqslant \frac{1}{2}e^{-\epsilon^2/2}.$$

因此得到

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \geqslant \epsilon\right] = \Pr\left[\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu) \geqslant \epsilon\sqrt{n}/\sigma\right] \leqslant \frac{1}{2}\exp(-n\epsilon^2/2\sigma^2),$$

定理得证.

下面定义 Sub-Gaussian 随机变量, 将有界随机变量和 Gaussian 随机变量统一起来:

定义 7.3 对任意 $t \in (-\infty, +\infty)$, 若随机变量 X 满足

$$E[e^{(X-E[X])t}] \leqslant \exp(bt^2/2),$$

则称随机变量 X 是服从参数为 b 的亚高斯 (Sub-Gaussian) 随机变量.

亚高斯随机变量表示随机变量的尾部分布不会比一个高斯分布更严重.

例 7.4 对任意有界的随机变量 $X \in [a, b]$, 根据 Chernoff 引理有

$$E[e^{(X-\mu)t}] \leqslant \exp(t^2(b-a)^2/8),$$

即有界的随机变量是参数为 $(b-a)^2/4$ 的亚高斯随机变量.

例 7.5 如果随机变量 X 服从高斯分布 $\mathcal{N}(\mu,\sigma^2)$, 则有

$$E[e^{(X-\mu)t}]=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{xt}e^{-x^2/2\sigma^2}dx=e^{\sigma^2t^2/2}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-(t\sigma-x/\sigma)^2/2}d(x/\sigma)=e^{\sigma^2t^2/2}.$$

Gaussian 随机变量是参数为 σ^2 的亚高斯随机变量.

由前面的例子可知高斯随机变量和有界的随机变量都是亚高斯随机变量. 根据 Chernoff 方法有

定理 7.10 设 $X_1,\dots,X_n$ 是 n 个独立的且参数为 b 的亚高斯随机变量, 对任意 $\epsilon>0$ 有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\geqslant\epsilon\right]\leqslant\exp(-n\epsilon^2/2b)\quad \text{和}\quad \Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\leqslant-\epsilon\right]\leqslant\exp(-n\epsilon^2/2b).$$

证明 对任意 $t>0$, 根据 Chernoff 方法有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\geqslant\epsilon\right]\leqslant e^{-tn\epsilon}\prod_{i=1}^{n}E[e^{(X_i-\mu)t}]\leqslant e^{-tn\epsilon+nbt^2/2},$$

通过求解上式最小值可得 $t_{\min}=\epsilon/b$, 代入完成证明.

对亚高斯型随机变量, 还可以给出最大值期望的估计:

定理 7.11 设 $X_1,\dots,X_n$ 是 n 个相互独立的、参数为 b 的亚高斯随机变量, 且满足 $E[X_i]=0$, 我们有

$$E\left[\max_{i\in[n]}X_i\right]\leqslant\sqrt{2b\ln n}.$$

证明 对任意 $t>0$, 根据 Jensen 不等式有

$$\begin{aligned}\exp\left(tE\left[\max_{i\in[n]}X_i\right]\right)&\leqslant E\left[\exp\left(t\max_{i\in[n]}X_i\right)\right]\\&= E\left[\max_{i\in[n]}\exp(tX_i)\right]\leqslant\sum_{i=1}^{n}E[\exp(tX_i)]\leqslant n\exp(t^2b/2).\end{aligned}$$

对上式两边同时取对数整理可得

$$E\left[\max_{i\in[n]}X_i\right]\leqslant\frac{\ln n}{t}+\frac{bt}{2}.$$

通过求解上式最小值可得 $t_{\min}=\sqrt{2\ln n/b}$, 代入完成证明.

前面所讲的概率不等式, 可以用另外一种表达形式给出, 这里以定理 7.10 为例: 假设 $X_1,\ldots,X_n$ 是独立的、且参数为 b 的亚高斯随机变量, 对任意 $\epsilon>0$ 有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_i-\mu)\geqslant\epsilon\right]\leqslant\exp(-n\epsilon^2/2b).$$

令 $\delta=\exp(-n\epsilon^2/2b)$, 求解出

$$\epsilon=\sqrt{2b\ln(1/\delta)/n},$$

代入整理可得: 至少以 $1-\delta$ 的概率有下面的不等式成立

$$\frac{1}{n}\sum_{i=1}^n (X_i-\mu)\leqslant\sqrt{\frac{2b}{n}\ln\frac{1}{\delta}}.$$

前面讲的所有不等式都可以采用 $1-\delta$ 的形式描述.

7.3 Bennet 和 Bernstein 不等式

通过考虑随机变量的方差, 可能推导出更紧地集中不等式, 下面介绍两个基于方差的不等式.

定理 7.12 (Bennet不等式) 设 $X_1,\ldots,X_n$ 是独立同分布的随机变量且满足 $X_i-E[X_i]\leqslant 1$, 其均值为 μ 和方差为 σ^2, 我们有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_i-\mu)\geqslant\epsilon\right]\leqslant\exp\left(-\frac{n\epsilon^2}{2\sigma^2+2\epsilon/3}\right).$$

证明 对任意 $t>0$, 根据 Chernoff 方法有

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n (X_i-\mu)\geqslant\epsilon\right]\leqslant e^{-nt\epsilon}E\left[\exp\left(\sum_{i=1}^n t(X_i-\mu)\right)\right]=e^{-nt\epsilon}\left(E[e^{t(X_1-\mu)}]\right)^n.$$

设 $Y=X_1-\mu$, 利用公式 $\ln z\leqslant z-1$ 得到

$$\begin{aligned}\ln E[e^{t(X_1-\mu)}] &= \ln E[e^{tY}]\leqslant E[e^{tY}]-1=t^2E\left[\frac{e^{tY}-tY-1}{t^2Y^2}Y^2\right]\\ &\leqslant t^2E\left[\frac{e^t-t-1}{t^2}Y^2\right]=(e^t-t-1)\sigma^2\end{aligned}$$

这里利用 $tY\leqslant t$ 以及 $(e^z-z-1)/z^2$ 是一个非单调递减的函数. 进一步有

$$e^t-t-1\leqslant\frac{t^2}{2}\sum_{k=0}^{\infty}(t/3)^k=\frac{t^2}{2(1-t/3)}.$$

因此可得

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_n-\mu)\geqslant\epsilon\right]\leqslant\exp\left(-nt\epsilon+\frac{nt^2\sigma^2}{2(1-t/3)}\right).$$

猜出 $t=\epsilon/(\sigma^2+\epsilon/3)$, 带入完成证明.

对于 Bennet 不等式, 令

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_n-\mu)\geqslant\epsilon\right]\leqslant\exp(-n\epsilon^2/(2\sigma^2+2\epsilon/3))=\delta,$$

可以给出不等式的另外一种表述: 至少以 $1-\delta$ 的概率有以下不等式成立

$$\frac{1}{n}\sum_{i=1}^{n}X_n\leqslant\mu+\frac{2\ln 1/\delta}{3n}+\sqrt{\frac{2\sigma^2}{n}\ln\frac{1}{\delta}}.$$

当方法 σ^2 非常小, 或趋于 0 时, 得到更紧的收敛率 $\bar{X}_n-\mu\leqslant O(1/n)$.

下面考虑另一种基于方差的不等式, 与 Bennet 不等式不同之处在于约束随机变量的矩:

定理 7.13 (Bernstein不等式) 设 $X_1,\ldots,X_n$ 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 $m\geqslant 2$ 有 $E[X_i^m]\leqslant m!b^{m-2}\sigma^2/2$, 那么我们有

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

证明 对任意 $t>0$, 根据 Chernoff 方法有

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

利用公式 $\ln z\leqslant z-1$ 有

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

由此可得

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

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

例 7.6 给出 Bernstein 不等式的 $1-\delta$ 表述.

7.4 应用: 随机投影 (Random Projection)

设高维空间 $\mathbb{R}^d$ 有 n 个点 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ (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: $f: \mathbb{R}^d \to \mathbb{R}^k$ ($k \ll d$), 使得以较大概率有

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

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

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

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 7.5 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 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

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

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

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

证明 下面分三步证明 J-L 引理.

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

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

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

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

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

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

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

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

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

$$\begin{aligned}&\Pr\left[\left\|\frac{\boldsymbol{x}P}{\sqrt{k}}\right\|_2^2\geqslant(1+\epsilon)\|\boldsymbol{x}\|_2^2\right]=\Pr\left[\left\|\frac{\boldsymbol{x}P}{\|\boldsymbol{x}\|_2}\right\|_2^2\geqslant(1+\epsilon)k\right]\\ &= \Pr\left[\sum_{j=1}^k v_j^2\geqslant(1+\epsilon)k\right]\leqslant e^{-(1+\epsilon)kt}\left(E[e^{t\sum_{j=1}^k v_j^2}]\right)^k=e^{-(1+\epsilon)kt}\left(E[e^{tv_1^2}]\right)^k.\end{aligned}$$

对标准 Gaussian 分布有

$$E[e^{tv_1^2}] = \int_{-\infty}^{+\infty}\frac{e^{tu^2}}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du=\int_{-\infty}^{+\infty}\frac{e^{-\frac{u^2}{2}(1-2t)}}{\sqrt{2\pi}}du=\frac{1}{\sqrt{1-2t}},$$

代入可得

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

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

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

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

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

根据泰勒中值定理有

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

于是得到

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

同理可证

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

第三步: 对任意给定 $i\neq j$, 根据第二步的结论可知

$$\Pr[\|\boldsymbol{y}_i-\boldsymbol{y}_j\|_2^2\geqslant(1+\epsilon)\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2^2]\leqslant e^{-k(\epsilon^2-\epsilon^3)/4},$$
$$\Pr[\|\boldsymbol{y}_i-\boldsymbol{y}_j\|_2^2\leqslant(1-\epsilon)\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2^2]\leqslant e^{-k(\epsilon^2-\epsilon^3)/4}.$$

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

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

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

第 8 章　大数定律及中心极限定理

8.1 大数定律

给定随机变量 $X_1, X_2, \cdots, X_n$, 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 8.1 (依概率收敛) 设 $X_1, X_2, \cdots, X_n, \cdots$ 是一随机变量序列, a 是一常数, 如果对任意 $\epsilon > 0$ 有

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

则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 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 $g\colon \mathbb{R} \to \mathbb{R}$ 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 $g\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ 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 $X_nY_n \xrightarrow{P} ab$.

定理 8.1 (大数定律) 若随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 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 8.2 (马尔可夫 Markov 大数定律) 如果随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 满足

$$\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \to 0 \qquad n \to \infty,$$

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 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立或同分布, 其证明直接通过 Chebyshev 不等式有

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])\right| \geqslant \epsilon\right] \leqslant \frac{1}{n^2\epsilon^2}\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right) \to 0 \qquad n \to \infty.$$

定理 8.3 (切比雪夫 Chebyshev 大数定律) 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 且存在常数 $c > 0$ 使得 $\mathrm{Var}(X_n) \leqslant c$, 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

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i])\right| \geqslant \epsilon\right] \leqslant \frac{1}{\epsilon^2 n^2}\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right) \leqslant \frac{c}{n\epsilon^2} \to 0 \qquad n \to \infty.$$

定理 8.4 (辛钦 Khintchine 大数定律) 设 $X_1, X_2, \cdots, X_n, \cdots$ 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 8.5 (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

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

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 $Y_1, Y_2, \ldots, Y_n$ 满足 $Y_i \sim \mathrm{Ber}(p)$, 则

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

于是得到

$$\lim_{n\to\infty}\Pr\left[\left|\frac{X_n}{n} - p\right| \geqslant \epsilon\right] = \lim_{n\to\infty}\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}Y_i - E[Y_i]\right| \geqslant \epsilon\right] \leqslant \frac{1}{\epsilon^2 n^2}\mathrm{Var}(\sum_{i=1}^{n}Y_i) = \frac{p(1-p)}{\epsilon^2 n} \to 0.$$

如何判断随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 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 8.1 独立的随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 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.

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

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

再根据

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

由此可得当 $n\to+\infty$ 时有

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

大数定律小结:

- Markov 大数定律: 若随机变量序列 $\{X_i\}$ 满足 $\mathrm{Var}(\sum_{i=1}^{n}X_n)/n^2\to 0$, 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 $\mathrm{Var}(X_i)\leqslant c$, 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 $X_n\sim\mathrm{B}(n,p)$, 有 $X_n/n\xrightarrow{P}p$.

8.2 中心极限定理

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

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

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

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

则称随机变量序列 $Y_1,Y_2,\cdots,Y_n,\cdots$ 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 8.6 设独立同分布的随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 的期望 $E(X_1) = \mu$ 和方差 $Var(X_1) = \sigma^2$, 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

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

随机变量 Y_n 是随机变量 $X_1, X_2, \ldots, X_n$ 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

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

大数定律给出了当 $n \to \infty$ 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 8.2 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 $V_1, V_2, \ldots, V_{20}$ 服从均匀分布 $U(0,10)$, 于是有 $E(V_k) = 5$ 和 $\mathrm{Var}(V_k) = 100/12 = 25/3$. 设 $V = \sum_{k=1}^{20} V_k$, 则有

$$E(V) = 100 \qquad \mathrm{Var}(V) = 500/3.$$

根据中心极限定理近似有

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

根据标准正态分布的分布函数 $\Phi(x)$ 有

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

查表完成证明.

例 8.3 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 $\mathrm{Var}(X_i) = 25$. 设总重量 $X = \sum_{i=1}^n X_i$, 则有 $E(X) = 50n$ 和 $\mathrm{Var}(X) = 25n$. 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

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

根据分布函数的单调性有

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

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 8.1 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:

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

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

例 8.4 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

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

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

例 8.5 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 $\mathrm{Var}(X) = 9$. 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

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

例 8.6 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 $|k/n - p| \leqslant 0.05$ 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

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

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

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

所以 $\Phi\left(\sqrt{n}/10\right) \geqslant 0.95$, 查表解得 $n \geqslant 271$.

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

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

成立, 则有

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

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 $\mathrm{Var}(X_k) = \sigma^2$, 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.