7.2.2 0/1 随机变量的 Chernoff 不等式

定理 7.8 若相互独立的随机变量 $X_1, X_2, \cdots, X_n$ 满足 $X_i \sim \mathrm{Ber}(p_i)$, 设 $\mu = p_1 + p_2 + \cdots + p_n$, 对任意 $\epsilon > 0$ 有

$$P\left[\sum_{i=1}^n X_i \geqslant (1+\epsilon)\mu\right] \leqslant \left(\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right)^\mu;$$

对任意 $\epsilon \in (0,1)$ 有

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

证明 对任意 $t > 0$, 根据 Chernoff 方法和独立同分布的假设有

$$\begin{aligned}P\left[\sum_{i=1}^n X_i \geqslant (1+\epsilon)\mu\right] &= P\left[\exp\left(t\sum_{i=1}^n X_i\right) \geqslant \exp(t(1+\epsilon)\mu)\right] \\ &\leqslant \exp(-t(1+\epsilon)\mu)E\left[\exp\left(t\sum_{i=1}^n X_i\right)\right] = \exp(-t(1+\epsilon)\mu)\prod_{i=1}^n E\left[\exp\left(tX_i\right)\right].\end{aligned}$$

根据随机变量 $X_i \sim \mathrm{Ber}(p_i)$ 和 $1 + x \leqslant e^x$ 可知

$$E[\exp(tX_i)] = 1 - p_i + p_i e^t = 1 + p_i(e^t - 1) \leqslant \exp(p_i(e^t - 1)),$$

再结合 $\mu = p_1 + p_2 + \cdots + p_n$ 可得

$$P\left[\sum_{i=1}^n X_i \geqslant (1+\epsilon)\mu\right] \leqslant \exp(-t(1+\epsilon)\mu)\exp(\mu(e^t - 1)).$$

由于上述不等式对任意 $t > 0$ 都成立, 可得

$$P\left[\sum_{i=1}^n X_i \geqslant (1+\epsilon)\mu\right] \leqslant \inf_{t>0}\left\{\exp(-t(1+\epsilon)\mu + \mu(e^t - 1))\right\} = \left(\frac{e^\epsilon}{(1+\epsilon)^{(1+\epsilon)}}\right)^\mu,$$

在上面的不等式中当 $t = \ln(1+\epsilon)$ 取得最小值.

对第二个不等式, 只需证明当 $\epsilon \in (0,1)$ 有

$$f(\epsilon) = \ln\left(\frac{e^\epsilon}{(1+\epsilon)^{(1+\epsilon)}}\right) + \frac{\epsilon^2}{3} = \epsilon - (1+\epsilon)\ln(1+\epsilon) + \frac{\epsilon^2}{3} \leqslant 0.$$

易知 $f(0) = 0$ 和 $f(1) < 0$. 当 $\epsilon \in (0,1)$,

$$f'(\epsilon) = -\ln(1+\epsilon) + 2\epsilon/3, \quad f''(\epsilon) = -\frac{1}{1+\epsilon} + \frac{2}{3}.$$

于是得到 $f'(0)=0$, $f'(1)=-0.0265<0$ 和 $f'(1/2)=-0.0721<0$, 由连续函数性质有 $f^{'}(\epsilon)\leqslant 0$, 即函数 $f(\epsilon)$ 在 $[0,1]$ 上单调递减. 当 $\epsilon\geqslant 0$ 时有 $f(\epsilon)\leqslant f(0)=0$, 由此完成证明.

下面给出另一方向的集中不等式, 其证明将留作练习题.

定理 7.9 若相互独立的随机变量 $X_1,X_2,\cdots,X_n$ 满足 $X_i\sim \mathrm{Ber}(p_i)$, 设 $\mu=p_1+p_2+\cdots+p_n$, 对任意 $\epsilon\in(0,1)$ 有

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

还可以给出集中不等式另一种等价的表述形式, 这里以独立的随机变量 $X_1,\ldots,X_n$ $(X_i\in[0,1])$ 为例, 根据定理 7.7, 对任意实数 $\epsilon>0$ 有

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

令 $\delta=\exp(-2n\epsilon^2)$, 求解出 $\epsilon=\sqrt{\ln(1/\delta)/2n}$, 代入整理可得: 至少以 $1-\delta$ 的概率有不等式

$$\frac{1}{n}\sum_{i=1}^n X_i-\frac{1}{n}\sum_{i=1}^n E[X_i]\leqslant\sqrt{\frac{1}{2n}\ln\frac{1}{\delta}}.$$

这就是集中不等式的 $1-\delta$ 概率等价性表述, 对前面所讲的任何集中不等式均有等价的表述.

7.3 基于方差的集中不等式

若考虑随机变量的方差, 可能得到更好的集中不等式, 这里介绍基于方差的两种集中不等式.

定理 7.10 (Bennet不等式) 若独立同分布的随机变量 $X_1,\ldots,X_n$ 的期望 $E[X_1]=\mu$ 和方差 $\mathrm{Var}(X_1)=\sigma^2$、且满足 $X_i-\mu\leqslant 1$, 则有

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

证明 对任意 $t>0$, 根据独立同分布假设和 Chernoff 方法有

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

设 $Y = X_1 - \mu$, 则有 $E[Y] = 0$ 和 $E[Y^2] = \sigma^2$. 利用公式 $\ln z \leqslant z - 1$ 可得

$$\ln E[\exp(t(X_1 - \mu))] = \ln E[\exp(tY)] \leqslant E[\exp(tY)] - 1 = t^2 E\left[\frac{\exp(tY) - tY - 1}{t^2Y^2} Y^2\right].$$

再根据 $tY \leqslant t$ 和 $(e^z - z - 1)/z^2$ 是一个非单调递减的函数, 可得

$$\ln E[\exp(t(X_1 - \mu))] \leqslant t^2 E\left[\frac{\exp(t) - t - 1}{t^2} Y^2\right] = (\exp(t) - t - 1)\sigma^2.$$

根据指数函数的 Taylor 展式有

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

由此可得

$$P\left[\frac{1}{n}\sum_{i=1}^{n}(X_n - \mu) \geqslant \epsilon\right] \leqslant \min_{t>0}\left\{\exp\left(-nt\epsilon + \frac{nt^2\sigma^2}{2(1-t/3)}\right)\right\} = \exp\left(-\frac{n\epsilon^2}{2\sigma^2 + 2\epsilon/3}\right).$$

在上面的不等式中当 $t = \epsilon/(\sigma^2 + \epsilon/3)$ 取得最小值.

对于 Bennet 不等式, 令

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

可以给出不等式的另外一种表述: 至少以 $1 - \delta$ 的概率有下面的不等式成立

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

当方法 σ^2 非常小或等于零时, 得到更好的收敛率 $\sum_{i=1}^{n} X_n/n - \mu \leqslant O(1/n)$.

还可以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 7.11 (Bernstein不等式) 设独立同分布的随机变量 $X_1, \ldots, X_n$ 的期望 $E[X_1] = \mu$ 和方差 $\mathrm{Var}(X_1) = \sigma^2$, 若存在常数 $b > 0$, 使得对正整数 $k \geqslant 2$ 有 $E[X_1^k] \leqslant k!b^{k-2}\sigma^2/2$ 成立, 则有

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

证明 对任意 $t>0$, 根据 Chernoff 方法和独立同分布假设有

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

利用公式 $\ln z\leqslant z-1$ 和指数分布的 Taylor 展开式有

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

由此可得

$$P\left[\frac{1}{n}\sum_{i=1}^{n}(X_n-\mu)\geqslant\epsilon\right]\leqslant\inf_{t>0}\left\{\exp\left(-nt\epsilon+\frac{t^2n\sigma^2}{2(1-bt)}\right)\right\}=\exp\left(-\frac{n\epsilon^2}{2\sigma^2+2b\epsilon}\right),$$

在上面的不等式中当 $t=\epsilon/(\sigma^2+b\epsilon)$ 取得最小值.

7.4 应用案例: 随机投影 (Random Projection)*

设高维空间 $\mathbb{R}^d$ 有 n 个数据点 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$, 这里维度 d 非常大, 例如 $d=10^7$ 或 $d=10^9$, 处理这样的高维数据在存储与计算方面存在着很多挑战. 实际中一种可行的解决方案是找到一种保距变换 $f\colon\mathbb{R}^d\to\mathbb{R}^k$ $(k\ll d)$, 使得以较大概率有

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

保距变换 f 本质上是一个降维的函数, 将数据点从 d 维降为 k 维, 这里 k 非常小, 远远小于 d. 一般通过引入随机矩阵来进行降维, 所以称为随机投影 (Random projection), 常被用于最近邻、k-近邻、降维、聚类等机器学习方法.

随机投影可简单地表示为

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

其中 A 是一个 $d\times k$ 的随机矩阵, 其元素是相互独立的随机数, c 为一个常数, 根据随机矩阵 A 确定. 下面介绍三种常用的随机投影:

- 随机矩阵 $A=(a_{ij})_{d\times k}\in\mathbb{R}^{d\times k}$, 其中每个元素 a_{ij} 通过标准正态分布 $\mathcal{N}(0,1)$ 独立随机采样得到, 此时取 $c=\sqrt{k}$;
- 随机矩阵 $A=(a_{ij})_{d\times k}\in\{-1,1\}^{d\times k}$, 其中每个元素 a_{ij} 通过分布 $P(a_{ij}=1)=P(a_{ij}=-1)=1/2$ 独立随机采样得到, 此时取 $c=\sqrt{k}$;
- 随机矩阵 $A=(a_{ij})_{d\times k}\in\{-1,0,1\}^{d\times k}$, 其中每个元素 a_{ij} 通过分布 $P(a_{ij}=1)=P(a_{ij}=-1)=1/6, P(a_{ij}=0)=2/3$ 独立随机采样得到, 此时取 $c=\sqrt{k/3}$. 这种方法常被用于稀疏化投影, 产生一个稀疏矩阵减少计算量.

本节研究基于标准正态分布 $\mathcal{N}(0,1)$ 的随机投影方法, 其余方法可类似讨论.

引理 7.2 若随机矩阵 $A=(a_{ij})_{d\times k}\in\mathbb{R}^{d\times k}$ 中每个元素 a_{ij} 根据标准正态分布 $\mathcal{N}(0,1)$ 独立同分布采样所得, 对任意向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_d)\in\mathbb{R}^d$ 有

$$E_A\left[\left\|\boldsymbol{x}A/\sqrt{k}\right\|_2^2\right]=\|\boldsymbol{x}\|_2^2.$$

证明 根据独立性和 $a_{ij}\sim\mathcal{N}(0,1)$ 有

$$\begin{aligned}
E_A\left[\left\|\frac{\boldsymbol{x}A}{\sqrt{k}}\right\|_2^2\right] &= E_{a_{ij}\sim\mathcal{N}(0,1)}\left[\sum_{j=1}^k\left(\sum_{i=1}^d\frac{x_ia_{ij}}{\sqrt{k}}\right)^2\right]=\frac{1}{k}\sum_{j=1}^kE_{a_{ij}}\left[\left(\sum_{i=1}^dx_ia_{ij}\right)^2\right]\\
&= \frac{1}{k}\sum_{j=1}^kE_{a_{ij},a_{lj}}\left[\sum_{i,l}x_ix_la_{ij}a_{lj}\right]=\frac{1}{k}\sum_{j=1}^k\sum_{i=1}^dx_i^2=\frac{1}{k}\sum_{j=1}^k\|\boldsymbol{x}\|_2^2=\|\boldsymbol{x}\|_2^2,
\end{aligned}$$

这里利用了标准正态分布的性质 $E[a_{ij}]=0$ 和 $\mathrm{Var}[a_{lj}]=1$.

引理 7.3 若随机矩阵 $A=(a_{ij})_{d\times k}\in\mathbb{R}^{d\times k}$ 中每个元素 a_{ij} 根据标准正态分布 $\mathcal{N}(0,1)$ 独立同分布采样所得, 对任意向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_d)\in\mathbb{R}^d$ 和实数 $\epsilon\in(0,1/2)$ 有

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

证明 首先将矩阵 A 表示为 $A=(A_1;A_2;\ldots;A_k)$, 其中 A_i $(i\in[d])$ 是一个 $d\times 1$ 的列向量, 令 $v_j=\boldsymbol{x}A_j/\|\boldsymbol{x}\|_2$, 即

$$\frac{\boldsymbol{x}A}{\|\boldsymbol{x}\|_2}=\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}A_1,\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}A_2,\ldots,\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}A_k\right)=(v_1,v_2,\ldots,v_k).$$

因为元素 a_{ij} 相互独立且服从 $\mathcal{N}(0,1)$, 根据正太分布的性质可知 $v_j\sim\mathcal{N}(0,1)$, 且随机变量 $v_1,v_2,\ldots,v_k$ 相互独立, 于是有

$$P_A\left[\left\|\frac{\boldsymbol{x}A}{\sqrt{k}}\right\|_2^2\geqslant(1+\epsilon)\|\boldsymbol{x}\|_2^2\right]=P_A\left[\left\|\frac{\boldsymbol{x}A}{\|\boldsymbol{x}\|_2}\right\|_2^2\geqslant(1+\epsilon)k\right]=P_{v_1,\cdots,v_k}\left[\sum_{j=1}^kv_j^2\geqslant(1+\epsilon)k\right].$$

对任意 $t\in(0,1/2)$, 根据 Chernoff 方法有

$$P_{v_1,\cdots,v_k}\left[\sum_{j=1}^kv_j^2\geqslant(1+\epsilon)k\right]\leqslant e^{-(1+\epsilon)kt}\left(E[e^{t\sum_{j=1}^kv_j^2}]\right)^k=e^{-(1+\epsilon)kt}\left(E[e^{tv_1^2}]\right)^k.$$

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

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

上式右边对 t 求最小解出 $t = \epsilon/2(1+\epsilon)$, 代入可得

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

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

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

由此完成第一个不等式的证明, 同理可证第二个不等式.

根据引理 7.3 可以推导出著名的 Johnson – Lindenstrauss 引理, 简称 J-L 引理.

定理 7.12 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 为 $\mathbb{R}^d$ 空间中的点, 随机矩阵 $A = (a_{ij})_{d\times k} \in \mathbb{R}^{d\times k}$, 其中元素 a_{ij} 相互独立且满足 $a_{ij} \sim \mathcal{N}(0, 1)$. 设

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

即将 d 维空间中 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 通过随机矩阵 A 投影到 k 维空间. 对任意实数 $\epsilon \in (0, 1/2)$, 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 给定 $i, j \in [n]$ 且满足 $i \neq j$, 根据引理 7.3 可知

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

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

由于 $i, j \in [n]$, 因此共有 $n(n-1)$ 种不同的 $i \neq j$, 根据布尔不等式有

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

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

7.5 大数定律

给定一个随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$, 大数定律考虑随机变量均值 $\sum_{i=1}^n X_i/n$ 是否随着 n 的增加而趋于一个稳定点, 即随机变量均值的稳定性.

定义 7.1 (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, 如果对任意实数 $\epsilon > 0$ 有

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

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

可以发现随机变量序列趋于一个稳定点与数列的极限有本质的不同. 下面给出依概率的一些性质:

- 若 $X_n \xrightarrow{P} a$ 且函数 $g\colon \mathbb{R} \to \mathbb{R}$ 在 $X = a$ 点连续, 则 $g(X_n) \xrightarrow{P} g(a)$;
- 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, 若 $X_n \xrightarrow{P} a$ 和 $Y_n \xrightarrow{P} b$, 则有 $X_n + Y_n \xrightarrow{P} a + b$ 和 $X_nY_n \xrightarrow{P} ab$.

定理 7.13 (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.

根据依概率收敛的定义, 可以给出大数定律的等价条件

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E[X_i]\right| > \epsilon\right) = 0.$$

大数定理刻画了随机变量的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 7.14 (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.