6.4 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6.9. 设随机变量 $X_1,\dots,X_n$ 相互独立、且服从 $X_i\sim\mathcal{N}(\mu,\sigma)$, 对任意 $\epsilon>0$ 有

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

Proof. 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

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

因此得到

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

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

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

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

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

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6.5. 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6.10. 设 $X_1,\ldots,X_n$ 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

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

Proof. 对任意 $t>0$, 根据 Chernoff 方法有

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

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6.11. 设 $X_1,\ldots,X_n$ 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

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

Proof. 对任意 $t>0$, 根据 Jensen 不等式有

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

对上式两边同时取对数整理可得

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

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 6.10 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

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

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

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

前面讲的所有不等式都可以采用 $1-\delta$ 的形式描述.

6.5 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6.12 (Bennet不等式)**.** 设 $X_1,\dots,X_n$ 是独立同分布的随机变量且满足 $X_i - E[X_i] \leq 1$, 其均值为 μ 和方差为 σ^2, 我们有

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

Proof. 对任意 $t>0$, 根据 Chernoff 方法有

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

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

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

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

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

因此可得

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

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

对于 Bennet 不等式, 令

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

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

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

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

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

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

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 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n$ (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

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

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

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

其中 P 是一个 $d\times k$ 的随机矩阵, 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 $\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n$ 为 $\mathbb{R}^d$ 空间的 n 个点, 随机矩阵 $P=(p_{ij})_{d\times k}\in\mathbb{R}^{d\times k}$, $p_{ij}\sim\mathcal{N}(0,1)$, 令

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

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

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

Proof. 下面分三步证明 J-L 引理.

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

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

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

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

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

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

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 $v_j=\bar{x}P_j/\|\bar{x}\|_2$, 即

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

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

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

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

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

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

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

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

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

根据泰勒中值定理有

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

于是得到

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

同理可证

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

第三步： 对任意给定 $i \neq j$, 根据第二步的结论可知

$$\Pr[\|y_i - y_j\|_2^2 \ge (1+\epsilon)\|x_i - x_j\|_2^2] \le e^{-k(\epsilon^2-\epsilon^3)/4},$$
$$\Pr[\|y_i - y_j\|_2^2 \le (1-\epsilon)\|x_i - x_j\|_2^2] \le e^{-k(\epsilon^2-\epsilon^3)/4}.$$

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

$$\Pr\left[\exists i \neq j \colon \|y_i - y_j\|_2^2 \ge (1+\epsilon)\|x_i - x_j\|_2^2 \quad \text{或} \quad \|y_i - y_j\|_2^2 \le (1-\epsilon)\|x_i - x_j\|_2^2\right] \le 2n^2 e^{-k(\epsilon^2-\epsilon^3)/4},$$

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