概率统计课程第十次作业

2020 年 11 月

1. 利用 chernoff 方法证明: 设 $X_1, X_2, \cdots, X_k$ 是 k 个独立的随机变量, 且 $X_i \sim N(0,1)$, 则有
$$Pr\left(\sum_{i=1}^{k} X_i^2 \geq (1+\epsilon)k\right) \leq \exp(-k(\epsilon^2-\epsilon^3)/4). \tag{1}$$

2. 对 $\forall \overline{x} \in R^d$, 且 $\|\overline{x}\|_2^2 = 1$, $P_1, P_2, \cdots, P_k \in R^d$, $P_i = (P_{i1}, P_{i2}, \cdots, P_{id})$. P 中每个元素服从 $N(0,1)$ 分布, 即 $P_{ij} \sim N(0,1)$. 求证
$$Pr\left(\sum_{j=1}^{k} (\overline{x}P_j^T)^2 \geq (1+\epsilon)k\right) \leq \exp(-k(\epsilon^2-\epsilon^3)/4). \tag{2}$$

3. 若 $\overline{x_1}, \overline{x_2}, \cdots, \overline{x_k} \in R^d$, $\overline{y_i} = f(\overline{x_i})$, f 是关于 $\overline{x_i}$ 的函数. 对任意给定 $i \neq j$, 有
$$Pr\left((1-\epsilon)\|\overline{x_i}-\overline{x_j}\|_2^2 \leq \|\overline{y_i}-\overline{y_j}\|_2^2 \leq (1+\epsilon)\|\overline{x_i}-\overline{x_j}\|_2^2\right) \geq 1-2e^{-k(\epsilon^2-\epsilon^3)/4}. \tag{3}$$
求证
$$Pr\left(\forall i \neq j : (1-\epsilon)\|\overline{x_i}-\overline{x_j}\|_2^2 \leq \|\overline{y_i}-\overline{y_j}\|_2^2 \leq (1+\epsilon)\|\overline{x_i}-\overline{x_j}\|_2^2\right) \geq 1-2k^2e^{-k(\epsilon^2-\epsilon^3)/4}. \tag{4}$$

4. 已知 Berstein 不等式
$$Pr\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu) \geq \epsilon\right] \leq \exp\left(\frac{-n\epsilon^2}{2\sigma^2+2b\epsilon}\right), \tag{5}$$
求其等价 $1-\delta$ 描述.

5. 已知 $X_1, X_2, \cdots, X_n$ 是独立同分布随机变量, $X_i \sim N(\mu, \sigma^2)$, 求 $\mathbb{E}[\max_{i\in 1:n} X_i]$ 的一个上界, 并给出严格证明.

作业上交日期：12 月 3 日课前

学术诚信

允许同学之间的相互讨论，但是署你名字的工作必须由你完成，不允许直接照搬任何已有的材料，必须独立完成作业的书写过程。

在完成作业过程中，对他人工作（出版物、互联网资料）中文本的直接照搬（包括原文的直接摘抄及语句的简单修改等）都将视为剽窃，剽窃者成绩将被取消。对于完成作业中有关键作用的公开资料，应予以明显引用。

如果发现作业之间高度相似将被判定为互相抄袭行为，抄袭和被抄袭双方的成绩都将被取消。因此请主动防止自己的作业被他人抄袭。