次高斯分布

MIStatlE

Last Update: 2025 年 3 月 24 日

摘要

本文介绍了次高斯分布的基本定义、性质及其在集中不等式中的应用。首先通过正态分布尾部的不等式推导出次高斯分布的概念，并展示了 Hoeffding 不等式的证明过程。最后，讨论了次高斯分布的四个等价表征，并给出了它们之间等价性的证明。本文为统计推断和机器学习中的集中现象分析提供了理论工具。

1 引言

在概率论和统计学中，研究随机变量的尾部行为对于理解数据波动、集中现象以及误差分析具有重要意义。特别是在高维数据分析和大样本理论中，传统的极限定理往往无法给出有效的非渐近性概率界，而次高斯分布正好描述了随机变量尾部以指数速度衰减的特性。次高斯分布的定义基于矩生成函数的上界，这一性质在推导 Hoeffding 不等式和其他集中不等式时起到了关键作用。本文旨在介绍次高斯分布的基本概念、主要性质及其等价表征，同时展示其在集中不等式中的应用。

2 次高斯分布与 Hoeffding 不等式

设 $X \sim N(\mu, \sigma^2)$ 为正态随机变量，则有

$$P\Big[|X-\mu| \geq t\Big] \leq 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

练习 2.1. 证明上述不等式。

证明. 令 $\lambda > 0$。利用 Markov 不等式，有

$$P[X-\mu \geq t] = P\Big[e^{\lambda(X-\mu)} \geq e^{\lambda t}\Big] \leq e^{-\lambda t}\,\mathbb{E}\Big[e^{\lambda(X-\mu)}\Big].$$

由于 $X \sim N(\mu, \sigma^2)$ 的矩生成函数为

$$\mathbb{E}\Big[e^{\lambda(X-\mu)}\Big] = \exp\Big(\frac{\lambda^2\sigma^2}{2}\Big),$$

故有

$$P[X-\mu \geq t] \leq \exp\Big(-\lambda t + \frac{\lambda^2\sigma^2}{2}\Big).$$

取 $\lambda = \frac{t}{\sigma^2}$ 得

$$P[X-\mu \geq t] \leq \exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

利用正态分布的对称性，同理可得

$$P[\mu - X \geq t] \leq \exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

因此，

$$P\Big[|X-\mu| \geq t\Big] \leq 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

□

这表明正态分布的尾部以指数形式衰减。具有类似尾部性质的分布称为次高斯分布。

定义 2.1 (次高斯分布)**.** 若随机变量 X 的均值为 μ，存在 $\nu > 0$ 使得对于所有 $\lambda \in \mathbb{R}$ 有

$$\mathbb{E}\Big[e^{\lambda(X-\mu)}\Big] \leq \exp\Big(\frac{\lambda^2\nu^2}{2}\Big),$$

则称 X 为次高斯分布，其参数为 ν。

备注 2.1. *1.* 注意参数 ν 不一定等于标准差，但它反映了随机变量的集中性。

2. 若随机变量满足上述定义，则有

$$P[X-\mu \geq t] \leq \exp\Big(-\frac{t^2}{2\nu^2}\Big).$$

示例 2.1 (高斯分布)**.** 设 $X \sim N(\mu, \sigma^2)$，则

$$\mathbb{E}\Big[\exp\big(\lambda(X-\mu)\big)\Big] = \exp\Big(\frac{\sigma^2\lambda^2}{2}\Big),$$

因此 X 为次高斯分布，其参数为 $\nu = \sigma$。

示例 2.2 (Rademacher 随机变量)**.** *Rademacher* 随机变量 ε 取值于 $\{-1, +1\}$ 且取值概率相等。可以证明

$$\mathbb{E}\Big[e^{\lambda\varepsilon}\Big] = \frac{1}{2}\Big(e^{-\lambda} + e^{\lambda}\Big) = \exp\Big(\frac{\lambda^2}{2}\Big),$$

即 ε 为次高斯分布，其参数为 $\nu = 1$。

示例 2.3 (有界随机变量)**.** 设 X 为零均值随机变量，其取值范围为 $[a, b]$。可以证明

$$\mathbb{E}\Big[e^{\lambda X}\Big] \leq \exp\Big(\frac{\lambda^2(b-a)^2}{8}\Big),$$

因此 X 为次高斯分布，其参数为 $\nu = \frac{b-a}{2}$。

证明. 令 $\psi(\lambda) = \log \mathbb{E}[e^{\lambda X}]$，注意到 $\psi(0) = \psi'(0) = 0$。又有

$$\psi''(\lambda) = \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \Big[\frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}\Big]^2,$$

这可解释为在新的概率测度 $d\mathbb{Q} = \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]} d\mathbb{P}$ 下的方差。由于

$$\mathrm{Var}[X] = \mathrm{Var}\Big[X - \frac{a+b}{2}\Big] \leq \mathbb{E}\Big[\Big(X - \frac{a+b}{2}\Big)^2\Big] \leq \frac{(b-a)^2}{4},$$

从而得到所需的不等式。 □

2.1 Hoeffding 不等式

利用切尔诺夫方法可以证明如下定理：

定理 2.1 (Hoeffding 不等式)**.** 设随机变量 X（其均值为 μ）为次高斯分布，参数为 ν。则对于任意 $t > 0$ 有

$$P\Big[|X-\mu| > t\Big] \leq 2\exp\Big(-\frac{t^2}{2\nu^2}\Big).$$

证明. 直接将次高斯的矩生成函数界代入切尔诺夫不等式，并对参数 $\lambda > 0$ 取最优值即可得到结论。 □

因为矩生成函数对于独立随机变量具有张量化性质，有如下命题：

命题 2.1. *设 $X_1, \ldots, X_n$ 为独立的次高斯随机变量，其参数分别为 $\nu_1^2, \ldots, \nu_n^2$。则和 $\sum_{k=1}^n X_k$ 也是次高斯分布，其参数为*

$$\nu = \sqrt{\sum_{k=1}^n \nu_k^2}.$$

证明. 利用独立性有

$$\mathbb{E}\left[e^{\lambda \sum_{i=1}^n (X_i - \mu_i)}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{\lambda (X_i - \mu_i)}\right] \le \prod_{i=1}^n \exp\left(\frac{\lambda^2 \nu_i^2}{2}\right) = \exp\left(\frac{\lambda^2 \sum_{i=1}^n \nu_i^2}{2}\right).$$

□

这立即推出如下结论：

定理 2.2 (一般 Hoeffding 不等式). *设 X_k $(k = 1, \ldots, n)$ 为独立随机变量，其均值为 μ_k 且满足次高斯条件，参数为 ν_k。则对于所有 $t \ge 0$ 有*

$$P\left[\left|\sum_{k=1}^n (X_k - \mu_k)\right| > t\right] \le 2\exp\left(-\frac{t^2}{2\sum_{k=1}^n \nu_k^2}\right).$$

示例 2.4. *若 X_k $(k = 1, \ldots, n)$ 为独立随机变量，且 $\mathbb{E}[X_k] = \mu_k$，并且满足 $a \le X_k \le b$，则有*

$$P\left[\left|\sum_{k=1}^n (X_k - \mu_k)\right| > t\right] \le 2\exp\left(-\frac{2t^2}{n(b-a)^2}\right).$$

示例 2.5. *回顾伯努利变量的例子，Hoeffding 不等式给出*

$$P[S_n > \alpha n] = P\left[\sum_{k=1}^n (X_k - p) \ge (\alpha - p)n\right] \le \exp\left(-\frac{(\alpha-p)^2 n}{2}\right).$$

备注 2.2. *证明一般 Hoeffding 不等式的关键在于对数矩生成函数对于独立随机变量的张量化性质。*

2.2 次高斯分布的等价表征证明

设 X 为零均值随机变量，下列条件被证明是等价的：

(1) 存在常数 $c_1 > 0$ 使得对所有 $\lambda \in \mathbb{R}$ 有

$$\mathbb{E}\left[e^{\lambda X}\right] \le \exp\left(c_1 \lambda^2 \nu^2\right).$$

(2) 存在常数 $c_2 > 0$ 使得对所有 $t \ge 0$ 有

$$P\left(|X| \ge t\right) \le 2\exp\left(-\frac{t^2}{c_2 \nu^2}\right).$$

(3) 存在常数 $c_3 > 0$ 使得对所有 $p \ge 1$ 有

$$\|X\|_{L^p} := \left(\mathbb{E}[|X|^p]\right)^{1/p} \le c_3 \nu \sqrt{p}.$$

(4) 存在常数 $c_4 > 0$ 使得

$$\mathbb{E}\left[\exp\left(\frac{X^2}{c_4 \nu^2}\right)\right] \le e.$$

下面给出这些条件之间的简要证明：

(1) ⇒ (2)

利用 Chernoff 不等式，对于任意 $\lambda > 0$ 和 $t > 0$：

$$P(X \geq t) \leq e^{-\lambda t}\mathbb{E}[e^{\lambda X}] \leq e^{-\lambda t}\exp(c_1\lambda^2\nu^2).$$

令 $\lambda = \frac{t}{2c_1\nu^2}$（可证明这是最优选取），得到

$$P(X \geq t) \leq \exp\Big(-\frac{t^2}{4c_1\nu^2}\Big).$$

同理，由于 X 为零均值且对称，

$$P(|X| \geq t) \leq 2\exp\Big(-\frac{t^2}{4c_1\nu^2}\Big).$$

令 $c_2 = 4c_1$ 即可。

(2) ⇒ (3)

利用分布函数与矩的关系：

$$\mathbb{E}[|X|^p] = \int_0^\infty p\, t^{p-1}P(|X| \geq t)\, dt.$$

将 (2) 式代入，

$$\mathbb{E}[|X|^p] \leq 2p\int_0^\infty t^{p-1}\exp\Big(-\frac{t^2}{c_2\nu^2}\Big)\, dt.$$

令 $u = \frac{t^2}{c_2\nu^2}$ 则 $du = \frac{2t}{c_2\nu^2}\, dt$，经过积分计算可得

$$\mathbb{E}[|X|^p]^{1/p} \leq C\nu\sqrt{p},$$

其中 C 为仅依赖于 c_2 的常数，从而证明 (3)。

(3) ⇒ (1)

利用幂级数展开有

$$\mathbb{E}\left[e^{\lambda X}\right] = \sum_{p=0}^{\infty}\frac{\lambda^p}{p!}\mathbb{E}[X^p].$$

由 (3) 知，对于 $p \geq 1$，$\mathbb{E}[|X|^p] \leq (c_3\nu\sqrt{p})^p$。利用 Stirling 公式可以证明该级数被某个 $\exp(c'\lambda^2\nu^2)$ 所控制，从而得到 (1)。

(1) ⇔ (4)

注意到 (1) 表示 X 的矩生成函数满足二次型上界，而 (4) 则要求 X^2 的矩生成函数在某一点有界。利用对数变换及二阶展开，可以证明 (1) 和 (4) 是等价的，具体证明见相关文献（例如 Vershynin 的书中有详细讨论）。

综上所述，(1)-(4) 四个条件互相等价。

3 总结

本文介绍了次高斯分布的定义及其在集中不等式中的重要应用。通过对正态分布尾部行为的分析，我们展示了如何利用矩生成函数与切尔诺夫方法推导出 Hoeffding 不等式，为统计推断和机器学习中的集中性分析提供了有力工具。同时，我们讨论了次高斯分布的四个等价表征，并给出了它们之间等价性的简要证明，这些结果在实际问题中具有广泛的应用意义。

4 参考文献

1. Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.

2. Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press.

3. Boucheron, S., Lugosi, G., & Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.