\documentclass{article}

\usepackage{amsmath,amssymb,amsthm}

\usepackage{framed}

\usepackage{ctex}

\usepackage{xcolor}

\usepackage{tcolorbox}

\usepackage{thmtools}

\usepackage{fancyhdr}

\usepackage{graphicx}

\usepackage{tikz}

\usepackage{float}

\usepackage{geometry}

\geometry{a4paper, margin=2cm}

\usepackage{enumitem}

\definecolor{shadecolor}{rgb}{0.9,0.95,1}

% 定理环境

\newtheorem{theorem}{定理}[section]

\newtheorem{definition}{定义}[section]

\newtheorem{lemma}{引理}[section]

\newtheorem{proposition}{命题}[section]

\newtheorem{corollary}{推论}[section]

\newtheorem{remark}{备注}[section]

\newtheorem{exer}{练习}[section]

\newtheorem{que}{问题}[section]

\newtheorem{example}{示例}[section]

% tcolorbox 环境

\newtcolorbox{understandingbox}{

colback=red!5!white,

colframe=red!75!black,

title=理解,

fonttitle=\bfseries,

sharp corners

}

\newtcolorbox{techniqueBox}{

colback=green!5,

colframe=green!70!black,

fonttitle=\bfseries,

title=技术要点,

sharp corners,

breakable

}

\pagestyle{fancy}

\fancyhf{}

\lhead{MIStatlE}

\rhead{高维概率论}

\cfoot{\thepage}

\title{次高斯分布}

\author{MIStatlE}

\date{Last Update: \today}

\begin{document}

\maketitle

\begin{abstract}

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。

\end{abstract}

\section{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。

\section{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\Bigl[|X-\mu|\ge t\Bigr]\le 2\exp\Bigl(-\frac{t^2}{2\sigma^2}\Bigr).

\]

\begin{exer}

证明上述不等式。

\end{exer}

\begin{proof}

令 \(\lambda>0\)。利用 Markov 不等式，有

\[

P[X-\mu\ge t]=P\Bigl[e^{\lambda (X-\mu)}\ge e^{\lambda t}\Bigr]\le e^{-\lambda t}\,\mathbb{E}\Bigl[e^{\lambda (X-\mu)}\Bigr].

\]

由于 \(X\sim N(\mu,\sigma^2)\) 的矩生成函数为

\[

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

\]

故有

\[

P[X-\mu\ge t]\le \exp\Bigl(-\lambda t+\frac{\lambda^2\sigma^2}{2}\Bigr).

\]

取 \(\lambda=\frac{t}{\sigma^2}\) 得

\[

P[X-\mu\ge t]\le \exp\Bigl(-\frac{t^2}{2\sigma^2}\Bigr).

\]

利用正态分布的对称性，同理可得

\[

P[\mu-X\ge t]\le \exp\Bigl(-\frac{t^2}{2\sigma^2}\Bigr).

\]

因此，

\[

P\Bigl[|X-\mu|\ge t\Bigr]\le 2\exp\Bigl(-\frac{t^2}{2\sigma^2}\Bigr).

\]

\end{proof}

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\emph{次高斯分布}。

\begin{shaded}

\begin{definition}[次高斯分布]

若随机变量 \(X\) 的均值为 \(\mu\)，存在 \(\nu>0\) 使得对于所有 \(\lambda\in\mathbb{R}\) 有

\[

\mathbb{E}\Bigl[e^{\lambda(X-\mu)}\Bigr]\le \exp\Bigl(\frac{\lambda^2\nu^2}{2}\Bigr),

\]

则称 \(X\) 为次高斯分布，其参数为 \(\nu\)。

\end{definition}

\end{shaded}

\begin{remark}

\begin{enumerate}

\item 注意参数 \(\nu\) 不一定等于标准差，但它反映了随机变量的集中性。

\item 若随机变量满足上述定义，则有

\[

P[X-\mu\ge t]\le \exp\Bigl(-\frac{t^2}{2\nu^2}\Bigr).

\]

\end{enumerate}

\end{remark}

\begin{example}[高斯分布]

设 \(X\sim N(\mu,\sigma^2)\)，则

\[

\mathbb{E}\Bigl[\exp\bigl(\lambda(X-\mu)\bigr)\Bigr]=\exp\Bigl(\frac{\sigma^2\lambda^2}{2}\Bigr),

\]

因此 \(X\) 为次高斯分布，其参数为 \(\nu=\sigma\)。

\end{example}

\begin{example}[Rademacher 随机变量]

Rademacher 随机变量 \(\varepsilon\) 取值于 \(\{-1,+1\}\) 且取值概率相等。可以证明

\[

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

\]

即 \(\varepsilon\) 为次高斯分布，其参数为 \(\nu=1\)。

\end{example}

\begin{example}[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}\Bigl[e^{\lambda X}\Bigr]\le \exp\Bigl(\frac{\lambda^2(b-a)^2}{8}\Bigr),

\]

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

\end{example}

\begin{proof}

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^2e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}-\Bigl[\frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}\Bigr]^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

\[

\operatorname{Var}[X]=\operatorname{Var}\Bigl[X-\frac{a+b}{2}\Bigr]\le \mathbb{E}\Bigl[\Bigl(X-\frac{a+b}{2}\Bigr)^2\Bigr]\le \frac{(b-a)^2}{4},

\]

从而得到所需的不等式。

\end{proof}

\subsection{Hoeffding 不等式}

利用切尔诺夫方法可以证明如下定理：

\begin{shaded}

\begin{theorem}[Hoeffding 不等式]

设随机变量 \(X\)（其均值为 \(\mu\)）为次高斯分布，参数为 \(\nu\)。则对于任意 \(t>0\) 有

\[

P\Bigl[|X-\mu|>t\Bigr]\le 2\exp\Bigl(-\frac{t^2}{2\nu^2}\Bigr).

\]

\end{theorem}

\end{shaded}

\begin{proof}

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。

\end{proof}

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：

\begin{proposition}

设 \(X\_1,\dots,X\_n\) 为独立的次高斯随机变量，其参数分别为 \(\nu\_1^2,\dots,\nu\_n^2\)。则和 \(\sum\_{k=1}^nX\_k\) 也是次高斯分布，其参数为

\[

\nu=\sqrt{\sum\_{k=1}^n\nu\_k^2}.

\]

\end{proposition}

\begin{proof}

利用独立性有

\[

\mathbb{E}\Bigl[e^{\lambda\sum\_{i=1}^n (X\_i-\mu\_i)}\Bigr]=\prod\_{i=1}^n\mathbb{E}\Bigl[e^{\lambda (X\_i-\mu\_i)}\Bigr]\le \prod\_{i=1}^n\exp\Bigl(\frac{\lambda^2\nu\_i^2}{2}\Bigr)

=\exp\Bigl(\frac{\lambda^2\sum\_{i=1}^n\nu\_i^2}{2}\Bigr).

\]

\end{proof}

这立即推出如下结论：

\begin{shaded}

\begin{theorem}[一般 Hoeffding 不等式]

设 \(X\_k\) (\(k=1,\dots,n\)) 为独立随机变量，其均值为 \(\mu\_k\) 且满足次高斯条件，参数为 \(\nu\_k\)。则对于所有 \(t\ge 0\) 有

\[

P\Bigl[\Bigl|\sum\_{k=1}^{n}(X\_k-\mu\_k)\Bigr|>t\Bigr]\le 2\exp\Bigl(-\frac{t^2}{2\sum\_{k=1}^n\nu\_k^2}\Bigr).

\]

\end{theorem}

\end{shaded}

\begin{example}

若 \(X\_k\) (\(k=1,\dots,n\)) 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\Bigl[\Bigl|\sum\_{k=1}^{n}(X\_k-\mu\_k)\Bigr|>t\Bigr]\le 2\exp\Bigl(-\frac{2t^2}{n(b-a)^2}\Bigr).

\]

\end{example}

\begin{example}

回顾伯努利变量的例子，Hoeffding 不等式给出

\[

P[S\_n>\alpha n]=P\Bigl[\sum\_{k=1}^{n}(X\_k-p)\ge (\alpha-p)n\Bigr]\le \exp\Bigl(-\frac{(\alpha-p)^2n}{2}\Bigr).

\]

\end{example}

\begin{remark}

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。

\end{remark}

\subsection{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：

\begin{enumerate}[label=(\arabic\*)]

\item 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\Bigl(c\_1\lambda^2\nu^2\Bigr).

\]

\item 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\Bigl(-\frac{t^2}{c\_2\nu^2}\Bigr).

\]

\item 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}.

\]

\item 存在常数 \(c\_4>0\) 使得

\[

\mathbb{E}\left[\exp\Bigl(\frac{X^2}{c\_4\nu^2}\Bigr)\right]\le e.

\]

\end{enumerate}

下面给出这些条件之间的简要证明：

\subsubsection\*{(1) $\Rightarrow$ (2)}

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

\[

P(X\ge t)\le e^{-\lambda t}\mathbb{E}[e^{\lambda X}]\le 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\ge t)\le \exp\Bigl(-\frac{t^2}{4c\_1\nu^2}\Bigr).

\]

同理，由于 \(X\) 为零均值且对称，

\[

P(|X|\ge t)\le 2\exp\Bigl(-\frac{t^2}{4c\_1\nu^2}\Bigr).

\]

令 \(c\_2=4c\_1\)即可。

\subsubsection\*{(2) $\Rightarrow$ (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|\ge t)\,dt.

\]

将 (2) 式代入，

\[

\mathbb{E}[|X|^p]\le 2p\int\_0^\infty t^{p-1}\exp\Bigl(-\frac{t^2}{c\_2\nu^2}\Bigr)\,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}\le C\nu\sqrt{p},

\]

其中 \(C\) 为仅依赖于 \(c\_2\) 的常数，从而证明 (3)。

\subsubsection\*{(3) $\Rightarrow$ (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 \(p\ge 1\)，\(\mathbb{E}[|X|^p]\le (c\_3\nu\sqrt{p})^p\)。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)。

\subsubsection\*{(1) $\Leftrightarrow$ (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）。

\vspace{1em}

综上所述，(1)-(4) 四个条件互相等价。

\section{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。

\section{参考文献}

\begin{enumerate}

\item Vershynin, R. (2018). \textit{High-Dimensional Probability: An Introduction with Applications in Data Science}. Cambridge University Press.

\item Wainwright, M. J. (2019). \textit{High-Dimensional Statistics: A Non-Asymptotic Viewpoint}. Cambridge University Press.

\item Boucheron, S., Lugosi, G., \& Massart, P. (2013). \textit{Concentration Inequalities: A Nonasymptotic Theory of Independence}. Oxford University Press.

\end{enumerate}

\end{document}