Stein 方法及其在统计物理中的应用

毕业论文答辩

巩舒阳

2017 级统计班

2021 年 5 月 21 日

山东大学数学学院
1930 School of Mathematics Shandong University

目录

Section 1

基本知识和背景

正态分布的刻画

关于正态分布，我们有如下重要定理：

正态分布的刻画（引理 1.1）

W 服从标准正态分布，当且仅当对于任意满足 $\mathbb{E}[|f'(W)|] < \infty$ 的绝对连续函数 f，我们有

$$\mathbb{E}f'(W) = \mathbb{E}(Wf(W))$$

Remark

我们将用 $|\mathbb{E}(f'(W) - Wf(W))|$ 的大小来衡量近似程度，给出上界。

Stein 方程

下面介绍证明非一致界的重要工具：

Stein 方程的解（引理 1.2）

Stein 方程

$$f'(x) - xf(x) = \mathbf{1}\{x \le z\} - \mathbb{P}(Z \le z)$$

的解可以写作如下形式

$$f_z(x) = \begin{cases} \sqrt{2\pi}e^{x^2/2}\Phi(x)(1-\Phi(z)), & x < z \\ \sqrt{2\pi}e^{x^2/2}\Phi(z)(1-\Phi(x)), & x \ge z \end{cases}$$

Remark

Stein 方程解具有很好的衰减性和光滑性。

一致界

Berry-Esseen 界（Berry(1941)、Esseen(1942)）

我们假设 $\xi_1,\xi_2,\cdots,\xi_n$ 是一列 0 均值相互独立的随机变量，满足 $\sum_{i=1}^n \mathrm{Var}(\xi_i)=1$，令 $W=\sum_{i=1}^n \xi_i$，F 为 W 的分布，Φ 为标准正态分布，有

$$\sup_{x\in\mathbb{R}}|F(x)-\Phi(x)|\leq C\sum_{i=1}^n E|\xi_i|^3$$

Remark

一般地，利用 Stein 方法我们能够获得 $O(\frac{1}{\sqrt{n}})$ 的上界，而且一致界与这里的 x 是无关的。本文中，我们常用技巧有：exchangeable pairs, zero bias, size bias，分别给出三个有关非一致界的定理。

目录

Section 2

指数非一致界（主要结果）

非一致界的提出

Chen and Shao(2001)

X_i 是一列独立但不一定同分布的随机变量，令 $W=\sum_{i=1}^n X_i$，存在一个绝对常数 C，使得

$$|F(x)-\Phi(x)|\le C\sum_{i=1}^n\left\{\frac{\mathbb{E}X_i^2 I(|X_i|>1+|x|)}{(1+|x|)^2}+\frac{\mathbb{E}|X_i|^3 I(|X_i|\le 1+|x|)}{(1+|x|)^3}\right\}$$

Remark

Louis H.Y. Chen 和 Q.M. Shao 主要利用了 Stein 方法和集中不等式方法，获得了一个 $O(|x|^{-2})$ 的非一致界。与之相比，本文主要证明了指数型的非一致界。

Zero bias 分布的非一致界

定理 3.1

令 W 为 0 均值方差为 1 的随机变量，$|W|$ 具有有限的矩母函数且 $\mathbb{E}e^{W^2}<\infty$，假设 W' 具有 W 的 zero bias 分布，且有 $|W'-W|\le\delta$，那么对于任意的常数 $M>0$，存在常数 $C_i, i=1,2,3$，使得

$$|\mathbb{P}(W\le z)-\mathbb{P}(Z\le z)|\le\begin{cases} C_1\delta e^{-\frac{(z+2\delta)^2}{2}}, & z\le -M-2\delta \\ C_2\delta e^{-\frac{z^2}{2}}, & |z|\le M+2\delta \\ C_3\delta e^{-\frac{(z-2\delta)^2}{2}}, & z\ge M+2\delta \end{cases}$$

Remark

上式可以以指数下降，这里的常数 C_i 依赖于 W，不依赖于 z。

定理 3.1 的证明思路

证明思路

首先通过 zero bias 分布的定义（$\sigma^2\mathbb{E}f'(W') = \mathbb{E}(Wf(W))$），我们很容易将该问题转换为求 $|\mathbb{E}\big(Wf(W) - W'f(W')\big)|$ 的上界，显然根据 Stein 方程解的性质 $xf(x)$ 在 $\mathbb{R}$ 上是分段可导的，所以分别考虑事件 $\{z \le W\}$ 和 $\{z > W\}$ 上的积分，在其中一个上使用 Lagrange 中值定理，另一个使用 Chebyshev 不等式，分 6 种情况考虑。

exchangeable pairs 的非一致界

定理 3.2

若 (W,W') 为 0 均值方差为 1 的可交换对，且满足

$$\mathbb{E}(W-W'|W)=\lambda(W-R)，其中\lambda\in(0,1)$$

$|W|$ 具有有限的矩母函数且 $\mathbb{E}e^{W^2}<\infty$，且有 $|W'-W|\le\delta$。那么对于任意常数 $M>0$，存在仅依赖于 δ,W,M 但不依赖于 z 的常数 C，使得下式成立

$$|\mathbb{P}(W\le z)-\mathbb{P}(Z\le z)|\le\begin{cases}C(B'+\frac{\delta^3}{2\lambda}+\frac{\delta^2}{2\lambda}+\|R\|_2)e^{-\frac{z^2}{2}}, & |z|\ge M\\ C(B'+1.5\delta+\frac{0.41\delta^3}{\lambda}+\frac{\sqrt{2\pi}}{4}\|R\|_2)e^{-\frac{z^2}{2}}, & |z|<M\end{cases}$$

其中 $B'=\|\mathbb{E}\left(1-\frac{(W-W')^2}{2\lambda}|W\right)\|_2$。

定理 3.2 的证明思路

证明思路

分为以下三个部分分别进行估计:

$$\left|\mathbb{E}\Big(f'(W)(1-\frac{\Delta^2}{2\lambda})\Big)\right|$$

,

$$\left|\mathbb{E}\Big(\frac{f'(W)\left(W'-W\right)^2-\left(f\left(W'\right)-f(W)\right)\left(W'-W\right)}{2\lambda}\Big)\right|$$

,

$$\mathbb{E}\left(Rf(W)\right)$$

与定理 3.1 相似，根据 Lagrange 中值定理和 Taylor 展开在两个事件上分类讨论。

size bias 的非一致界

定理 3.3

若 Y^s 具有 Y–size bias 分布，且满足 $|Y^s - Y| \le A$，$|W|$ 具有有限矩母函数且 $\mathbb{E}e^{W^2} < \infty$，那么对于任意的常数 $M > 0$，不依赖于 z 的常数 C，使得下式成立

$$|\mathbb{P}(W \le z) - \mathbb{P}(Z \le z)| \le \begin{cases} C\Big(D' + \frac{\mu A}{\sigma^2} + \frac{\mu A^2}{\sigma^3}\Big)e^{-\frac{z^2}{2}}, & |z| \ge M \\ C\Big(2D + \frac{6\mu A^2}{\sigma^3}\Big)e^{-\frac{z^2}{2}}, & |z| < M \end{cases}$$

这里 $D' = \|\mathbb{E}(1 - \frac{\mu}{\sigma}(W^s - W)|W)\|_2$

定理 3.3 的证明思路

证明思路

分为以下两部分证明

$$E\left|f'(W)\left(1-\frac{\mu}{\sigma}(W^s-W)\right)\right|$$

,

$$\frac{\mu}{\sigma}E\left|\left(f(W^s)-f(W)-f'(W)(W^s-W)\right)\right|$$

与之前类似，根据根据 Lagrange 中值定理和 Taylor 展开在两个事件上分类讨论。

目录

山东大学数学学院
1930 School of Mathematics Shandong University

Section 3

Stein 方法在统计物理中的应用

Combinatorial CLT on Conjugacy Classes

定理 4.2

设 $n \geq 5$ 且 $\{a_{ij}\}_{i,j=1}^n$ 为一对称矩阵，π 为一随机置换并在各 cycle type 上保持均匀分布，而且没有固定点（即 $\pi(k) \neq k$），我们令 $W = \frac{Y - \mathbb{E}Y}{\sigma_\rho}$，对任意常数 $M > 0$，我们有，

$$|\mathbb{P}(W \leq z) - \mathbb{P}(Z_\rho \leq z)|$$
$$\leqslant \begin{cases} C_1 40 \max_{i\neq j}|a_{ij} - 2a_{io} + a_{oo}| e^{-\frac{(z+2\delta)^2}{2}} \sum_{c\in\mathcal{N}_n} \frac{\rho_c}{\sigma_c}, & z \leq -M - 2\delta \\ C_2 40 \max_{i\neq j}|a_{ij} - 2a_{io} + a_{oo}| e^{-\frac{z^2}{2}} \sum_{c\in\mathcal{N}_n} \frac{\rho_c}{\sigma_c}, & |z| < M + 2\delta \\ C_3 40 \max_{i\neq j}|a_{ij} - 2a_{io} + a_{oo}| e^{-\frac{(z-2\delta)^2}{2}} \sum_{c\in\mathcal{N}_n} \frac{\rho_c}{\sigma_c}, & z \geq M + 2\delta \end{cases}$$

Anti-voter Model

定理 4.4

令 $\{\boldsymbol{X}^{(t)}: t=0,1,\cdots\}$ 为一个图 $\mathscr{G}=(\mathscr{V},\mathscr{E})$ 上的具有平稳分布的马氏链。这个图满足如下性质：(1) 具有 n 个顶点；(2) 是 r-regular 的；(3) 既不是 $n-cycle$ 也不是二分（bipartite）的。我们令 U 表示 $\boldsymbol{X}$ 上的符号和，即 $U=\sum_{v\in\mathscr{V}} X_v$。
令 $W=U/\sigma$，$\sigma^2=Var(U)$，$Q=\sum_{v\in\mathscr{V}}\sum_{w\in N_v} X_v X_w$，则对任意的常数 $M>0$，存在不依赖于 z 的常数 C，使得下式成立

$$
\begin{aligned}
&|\mathbb{P}(W\le z)-\mathbb{P}(Z\le z)|\\
\leqslant &\begin{cases} C\left(\dfrac{\sqrt{Var(Q)}}{2r\sigma^2}+\dfrac{2n}{\sigma^3}+\dfrac{n}{\sigma^2}\right)e^{-\frac{z^2}{2}}, & |z|\ge M\\ C\left(\dfrac{\sqrt{Var(Q)}}{2r\sigma^2}+\dfrac{3}{\sigma}+\dfrac{1.64n}{\sigma^3}\right)e^{-\frac{z^2}{2}}, & |z|<M \end{cases}
\end{aligned}
$$

Quadratic Forms

定理 4.5

令 $X_1,\cdots,X_n$ 为一列 0 均值，方差为 1 的随机变量，有界 ($|X_i|\le A, i=1,2,\cdots,n$) 的独立同分布随机变量，$A=\{a_{ij}\}_{i,j=1}^n$ 为对角线为 0 对
称矩阵，令 $W_n=\sum_{1\le i\neq j\le n}a_{ij}X_iX_j$，$\sigma_n^2=2\sum_{i,j=1}^n a_{ij}^2$，存在 C，C_0，使得

$$|P(W\le z)-P(Z\le z)|$$

$$\le\begin{cases} C\left(\dfrac{C_0A^4}{\sigma_n^2}\sqrt{\sum_{i=1}^n\left(\sum_{j=1}^n a_{ij}^2\right)^2+\sum_{i=1}^n\sum_{j=1}^n\left(\sum_{k=1}^n a_{ik}a_{jk}\right)^2}+\dfrac{16nA^6}{\sigma_n^3}\left(\max_{1\le i\le n}\left(\sum_{j=1}^n|a_{ij}|\right)\right)^3+\dfrac{4nA^4}{\sigma_n^2}\left(\max_{1\le i\le n}\left(\sum_{j=1}^n|a_{ij}|\right)\right)^2\right)e^{-\frac{z^2}{2}}, & |z|\ge M \\ C\left(\dfrac{C_0A^4}{\sigma_n^2}\sqrt{\sum_{i=1}^n\left(\sum_{j=1}^n a_{ij}^2\right)^2+\sum_{i=1}^n\sum_{j=1}^n\left(\sum_{k=1}^n a_{ik}a_{jk}\right)^2}+\dfrac{6A^2}{\sigma_n}\max_{1\le i\le n}\left(\sum_{j=1}^n|a_{ij}|\right)+\dfrac{13.12nA^6}{\sigma_n^3}\left(\max_{1\le i\le n}\left(\sum_{j=1}^n|a_{ij}|\right)\right)^3\right)e^{-\frac{z^2}{2}}, & |z|<M \end{cases}$$

Lightbulb Process

定理 4.7

令 Y 为时刻终止时刻 n 时亮着的灯泡数目，μ 和 σ^2 为它的期望和方差，令 $W=\frac{Y-\mu}{\sigma}$，对于任意偶数 n，$M>0$，存在不依赖于 z 常数 C，我们有：

$$|\mathbb{P}(W\le z)-\mathbb{P}(Z\le z)|\le\begin{cases}C\left(\frac{\mu}{\sigma^2}\Psi+\frac{2\mu}{\sigma^2}+\frac{4\mu}{\sigma^3}\right)e^{-\frac{z^2}{2}}, & |z|\ge M\\ C\left(\frac{24\mu}{\sigma^3}+\frac{2\mu}{\sigma^2}\Psi\right)e^{-\frac{z^2}{2}}, & |z|<M\end{cases}$$

这里，Ψ 满足

$$\Psi\le\frac{1}{2\sqrt{n}}+\frac{1}{2n}+e^{-\frac{n}{2}}$$

Patterns in Graphs and Permutations

定义（Dependency neighborhood）

考虑一类随机变量 $\mathbf{X}=\{X_\alpha,\alpha\in\mathscr{A}\}$。若 $\mathscr{B}_\alpha\subset\mathscr{A}$ 满足：X_α 和 $\{X_\beta:\beta\notin\mathscr{B}_\alpha\}$ 相互独立。那么称 $\mathscr{B}_\alpha$ 为α 的相依邻域。

定理 4.9

令 $\mathbf{X}=X_\alpha,\alpha\in\mathscr{A}$ 为一类在 $[0,M/2]$ 取值的随机变量，令 $Y=\sum_{\alpha\in\mathscr{A}}X_\alpha$，$\mu=\sum_{\alpha\in\mathscr{A}}EX_\alpha$，$\sigma^2=Var(Y),p_\alpha=\frac{EX_\alpha}{\sum_{\beta\in\mathscr{A}}EX_\beta}$，$\bar{p}=\max_{\alpha\in\mathscr{A}}p_\alpha$，$b=\max_{\alpha\in\mathscr{A}}|\mathscr{B}_\alpha|$ 令 $\mathbf{X}^\alpha$ 具有 $\mathbf{X}$ 的 α-size biased 分布，若存在 $\mathscr{F}\supset\sigma(Y)$ 和 $\mathscr{D}\subset\mathscr{A}\times\mathscr{A}$，使得：若 $(\alpha_1,\alpha_2)\notin\mathscr{D}$，那么对于任意的 $(\beta_1,\beta_2)\in\mathscr{B}_{\alpha_1}\times\mathscr{B}_{\alpha_2}$，$cov\left(E\left(X_{\beta_1}^{\alpha_1}-X_{\beta_1}\middle|\mathscr{F}\right),E\left(X_{\beta_2}^{\alpha_2}-X_{\beta_2}\middle|\mathscr{F}\right)\right)=0$。令 $W=\frac{(Y-\mu)}{\sigma}$，对于

Patterns in Graphs and Permutations

续定理 4.9

任意的 $M>0$，我们有不依赖于 z 的 C，它满足

$$|\mathbb{P}(W\le z)-\mathbb{P}(Z\le z)|\le\begin{cases}C\Big(\frac{\mu\bar{p}bM\sqrt{|\mathscr{D}|}}{\sigma^2}+\frac{\mu bM}{\sigma^2}+\frac{\mu b^2M^2}{\sigma^3}\Big)e^{-\frac{z^2}{2}}, & |z|\ge M\\ C\Big(\frac{6\mu b^2M^2}{\sigma^3}+\frac{2\mu\bar{p}bM\sqrt{|\mathscr{D}|}}{\sigma^2}\Big)e^{-\frac{z^2}{2}}, & |z|<M\end{cases}$$

目录

Section 4

创新和不足

创新

创新

- 计算出非一致的 Berry-esseen 界

创新

创新

- 计算出非一致的 Berry-esseen 界
- 将非一致界应用到模型中，可应用型强

创新

创新

- 计算出非一致的 Berry-esseen 界
- 将非一致界应用到模型中，可应用型强
- 非一致收敛速度是可调的

创新

创新

- 计算出非一致的 Berry-esseen 界
- 将非一致界应用到模型中，可应用型强
- 非一致收敛速度是可调的
- 克服了特征函数法的缺陷

不足

不足

- 很多情况下，常数 C 和 n 有关

不足

不足

- 很多情况下，常数 C 和 n 有关
- 兼顾非一致和一致收敛速度

不足

不足

- 很多情况下，常数 C 和 n 有关
- 兼顾非一致和一致收敛速度
- 不知是否能构造出 zero bias，size bias，exchangeable pair