Bahadur 渐近有效性在中等偏差概率区域中的表现

Mikhail Ermakov

2025 年 4 月

俄罗斯科学院机械工程问题研究所和圣彼得堡国立大学

对于一系列具有未知参数的独立同分布随机变量，其分布函数来自集合 $\Theta \subset \boldsymbol{R}^d$，我们证明了中偏差概率区域下的 Bahadur 渐近效率下界的一个类似结果。假设条件与 Hajek-Le Cam 局部渐近极小极大下界被证明时所用的假设条件相同。局部 Bahadur 渐近效率的下界是这个下界的一个特例。

1 介绍

令 $X_1, \ldots, X_n$ 为具有概率测度 $\boldsymbol{P}_\theta, \theta \in \Theta$ 的独立同分布随机变量。概率测度 $\boldsymbol{P}_\theta, \theta \in \Theta$ 在集合 S 的 σ-代数 $\mathcal{B}$ 上定义。集合 Θ 是 $\boldsymbol{R}^d$ 的一个开有界子集。参数 θ 的值未知。我们感兴趣的是参数 θ 的估计量的渐近效率的下界。

有两个方法用于研究渐近效率。其中之一是局部渐近极小极大下界。Hajek-Le Cam 的 [7, 9, 8, 11] 局部渐近极小极大下界为估计量 $\hat\theta_n = \hat\theta_n(X_1, \ldots, X_n)$ 提供了一个关于其从参数的真实值 θ 偏差阶数为 $n^{-1/2}$ 的渐近效率下界。在

[1]该研究得到了俄罗斯联邦科学与高等教育部的支持（项目 124041500008-1）。

这种情况下，我们得到了统计估计的风险在参数真实值任意小邻域内的上确界的下界。这样的渐近效率下界不能精确地提供特定参数值下统计估计器风险行为的信息。

Bahadur 渐近效率的下界 [1, 2, 3, 8]，对于统计估计的大偏差概率已被证明，提供了在可能参数值的每个特定点处统计估计风险的下界。在这种情况下，仅假设估计的一致性。

在适度偏差概率区域内，局部渐近极小风险统计估计量的界限已在 [4, 5, 10] 中建立。对该问题的兴趣源于构建置信区间边界是基于估计量分布的尾部，因此可以被视为适度偏差概率问题。对于对数渐近线，在与 Hajek-Le Cam 局部渐近效率下界 [7, 9, 8, 11] 相同的假设 [4, 10] 下，获得了适度偏差概率的局部渐近极小下界。对于统计估计量适度偏差概率的强渐近线，在不太强的额外假设 [4, 5] 下，已经证明了渐近极小风险的下界。

在中等偏差概率区域，可以证明渐近效率的下界，既包括渐近极小极大设置也包括 Bahadur 设置。对于渐近极小极大设置，下界的证明已在 [4, 5, 10] 中给出。本文的目标是在假设与 Hajek-Le Cam 局部渐近有效性相同的条件下，在中等偏差概率区域获得一个类似于 Bahadur 的下界。请注意，直接应用 Bahadur 的证明方法会导致建立局部 Bahadur 渐近有效性的下限时出现显著的附加条件 [6, 8]。局部 Bahadur 渐近有效性的下界是本文在中等偏差概率区域建立的关于渐近效率的下界的特殊情况。对于 $\Theta \subset \boldsymbol{R}^1$，我们还在估计量不应属于该区间的外部的每一侧分别证明了类似于 Bahadur 的下界。这些“单边”下界的多维类比也已获得。

本文组织如下。所有结果和证明都收集在第 2节中。在子节 2.1中，我们介绍了主要结果的证明条件。这个条件与建立局部渐近最小最大 Hajek-Le Cam 风险界相同的条件一致。在子节 2.2中，为了完整展示和更好地理解后续结果，呈现了一个统计估计中等偏差概率的局部渐近最小最大风险界。在子节 2.3中，提供了统计估计器在中等偏差概率区域内的 Bahadur 渐近效率下界的类似物。Bahadur 局部渐近效率的下界是这些情况中的特例。在子节 2.4中，我们给出了 Bahadur 下界的推广到多元情况，涵盖了“单边”下界，并展示了它的证明与传统 Bahadur 渐近效率下的界限证明没有区别

[3, 8]。

对于多维参数 $\theta \in \Theta \subset \boldsymbol{R}^d$,$d > 1$，我们将用希腊字母 $\theta, \tau, \phi, \ldots$ 表示向量或向量函数。我们用 $\mathbf{1}(A)$ 表示事件 A 的指示器。

2 主要结果

2.1 主要条件

假设概率测度 $\boldsymbol{P}_\theta$,$\theta \in \Theta$ 关于在同一 σ-代数 $\mathcal{B}$ 上的集合 S 的概率测度 ν 是绝对连续的，并且具有密度

$$f(x, \theta) = \frac{d\boldsymbol{P}_\theta}{d\nu}(x), \quad x \in S. \tag{2.1}$$

对于任何 $\theta, \theta_0 \in \Theta$，记 $\boldsymbol{P}^a_{\theta\theta_0}$ 和 $\boldsymbol{P}^s_{\theta\theta_0}$ 为概率测度 $\boldsymbol{P}_\theta$ 关于概率测度 $\boldsymbol{P}_{\theta_0}$ 的绝对连续部分和奇异部分。

对于所有 $\theta_0, \theta_0 + \tau \in \Theta$ 定义函数

$$g(x, \tau) = g(x, \theta_0, \theta_0 + \tau) = \Big(\frac{f(x, \theta_0 + \tau)}{f(x, \theta_0)}\Big)^{1/2} - 1 \tag{2.2}$$

对于所有 x 属于测度 $\boldsymbol{P}^a_{\theta_0+\tau,\theta_0}$ 的支撑集，并且在其他情况下等于零。

我们说统计实验 $\mathcal{E} = \{(S, \mathcal{B}), \boldsymbol{P}_\theta, \theta \in \Theta\}$ 在点 $\theta_0 \in \Theta$ 处具有有限的费舍尔信息，如果存在向量函数 $\phi : S \to \boldsymbol{R}^d$，使得我们有

$$\int_S (g(x, \tau) - \tau^T \phi)^2 \, d\, P_{\theta_0} = o(|\tau|^2), \quad \boldsymbol{P}^s_{\theta_0, \theta_0+\tau}(S) = o(|\tau|^2) \tag{2.3}$$

作为 $\tau \to 0$ 且矩阵分量

$$I(\theta_0) = 4 \int_S \phi\, \phi^T \, d\, \boldsymbol{P}_{\theta_0}. \tag{2.4}$$

取有限值。

矩阵 $I(\theta_0)$ 被称为费舍尔信息矩阵。

设我们有一个序列为 $u_n > 0$,$u_n \to 0$,$n u_n^2 \to \infty$，表示为 $n \to \infty$。

我们说估计量 $\hat{\theta}_n = \theta_n(X_1, \ldots, X_n)$ 参数 $\theta \in \Theta$ 是 u_n-一致的，如果对于任何 $\theta_0 \in \Theta$ 存在一个邻域 U 的 θ，使得对于任何 $\delta > 0$，我们有

$$\lim_{n\to\infty} \sup_{\theta\in U} \boldsymbol{P}_\theta(|\hat{\theta}_n - \theta| > \delta u_n) = 0. \tag{2.5}$$

2.2 局部渐近最小最大下界

局部渐近极小最大下界风险在中等偏差概率区域内的不需任何一致性条件。

定理 2.1 令统计实验 $\mathcal{E} = \{(S, \mathcal{B}), \boldsymbol{P}_\theta, \theta \in \Theta\}$ 在点 $\theta_0 \in \Theta$ 具有有限的费舍尔信息。那么，对于任何估计器 $\hat{\theta}_n$，在点 $\theta_0, \theta_n = \theta_0 + 2u_n \in \Theta$ 处，我们有

$$\liminf_{n\to\infty} \sup_{\theta=\theta_0,\theta_n} (nu_n^2 I(\theta)/2)^{-1} \log \boldsymbol{P}_\theta(|\hat{\theta}_n - \theta| > u_n) \geq -1. \tag{2.6}$$

对于证明，我们考虑假设检验问题 $H_0 : \theta = \theta_0$ 与备择假设 $H_n : \theta = \theta_0 + v_n, v_n = 2u_n$。定义检验 $K_n = K_n(X_1, \ldots, X_n) = \mathbf{1}(\hat{\theta}_n - \theta_0 > u_n)$。记 $\alpha(K_n)$ 和 $\beta(K_n)$ 分别为检验 K_n 的第一类和第二类错误概率。根据 [4] 中的定理 2.2，如果 $\alpha(K_n) < c < 1$ 和 $\beta(K_n) < c < 1$，则有

$$\limsup_{n\to\infty} (nv_n^2)^{-1/2} (|2\log\alpha(K_n)|^{1/2} + |2\log\beta(K_n)|^{1/2}) \leq 1 \tag{2.7}$$

(2.7) 的证明基于将局部渐近正态性的某个版本扩展到中等偏差概率区域，并且由于 Neyman-Pearson 引理，(2.7) 在检验正态分布的类似假设的情况下是有效的。

我们有

$$\alpha(K_n) \leq \boldsymbol{P}_{\theta_0}(|\theta_n - \theta_0| > u_n) \tag{2.8}$$

和

$$\beta(K_n) = \boldsymbol{P}_{\theta_n}(\hat{\theta}_n - \theta_0 < u_n) = \boldsymbol{P}_{\theta_n}(\hat{\theta}_n - \theta_0 - 2u_n < -u_n) = \boldsymbol{P}_{\theta_n}(|\hat{\theta}_n - \theta_n| > u_n). \tag{2.9}$$

通过 (2.7) - (2.9)，我们得到 (2.6)。

2.3 中等偏差概率区域中的 Bahadur 效率

让我们给出一个在中等偏差概率区域的 Bahadur 效率下界的类似物 $d=1$。

定理 2.2 设统计实验 $\mathcal{E}=\{(S,\mathcal{B}),\boldsymbol{P}_\theta,\theta\in\Theta\}$ 在点 $\theta_0\in\Theta\subset\boldsymbol{R}^1$ 处具有有限的费希尔信息。令估计量 $\hat\theta_n$ 是 u_n 一致的。然后，对于任意估计量 $\hat\theta_n$，我们有

$$\liminf_{n\to\infty}(nu_n^2I(\theta)/2)^{-1}\log\boldsymbol{P}_\theta(\,|\hat\theta_n-\theta|>u_n)\geq -1. \tag{2.10}$$

此外，我们有

$$\liminf_{n\to\infty}(nu_n^2I(\theta)/2)^{-1}\log\boldsymbol{P}_\theta(\,\hat\theta_n-\theta>u_n)\geq -1. \tag{2.11}$$

为了证明我们考虑假设检验问题 $H_0:\theta=\theta_0$ 与备择假设 $H_n:\theta=\theta_0+v_n$，$v_n=ru_n$，$r>1$。定义检验 $K_n=K_n(X_1,\ldots,X_n)=\mathbf{1}(|\hat\theta\_\theta_0|>u_n)$。

由于估计量 $\hat\theta_n$ 是 u_n-一致的，我们可以实现 (2.7) 并得到

$$\limsup_{n\to\infty}(nv_n^2)^{-1/2}(|2\log\boldsymbol{P}_{\theta_0}(|\hat\theta_n-\theta_0|>u_n)|^{1/2}+|2\log\boldsymbol{P}_{\theta_0+v_n}(|\hat\theta_n-\theta_0|<u_n)|^{1/2})\leq 1. \tag{2.12}$$

由于估计量 $\hat\theta_n$ 是 u_n-一致的，则

$$\lim_{n\to\infty}\boldsymbol{P}_{\theta_0+v_n}(|\hat\theta_n-\theta_0|<u_n)=0. \tag{2.13}$$

由 (2.19) 和 (2.13)，我们得到

$$\liminf\sup_{n\to\infty}(nr^2u_n^2)^{-1/2}(|2\log\boldsymbol{P}_{\theta_0}(|\hat\theta_n-\theta_0|>u_n)|^{1/2}\leq 1. \tag{2.14}$$

由于选择 $r>1$ 是任意的，我们得到 (2.10)。

不等式 (2.11) 的证明是类似的。只需将 (2.7) 应用于测试 $K_n=\mathbf{1}(\hat\theta_n-\theta_0>u_n)$。此推理被省略。

定理 2.2的多维版本如下。在以下的定理 2.3和 2.5中，渐近效率在不同的方向上变化。这是这些定理的主要区别特征。

令 V 为 $\boldsymbol{R}^d$ 和 $0\in V$ 中的有界凸开集。记 $\bar V$ 为集合 V 的补集。

定理 2.3 令统计实验 $\mathcal{E} = \{(S, \mathcal{B}), \boldsymbol{P}_\theta, \theta \in \Theta\}$ 在点 $\theta_0 \in \Theta \subset \boldsymbol{R}^d$ 处具有有限的费希尔信息。令估计量 $\hat{\theta}_n$ 为 u_n 一致。然后，对于任意估计量 $\hat{\theta}_n$，我们有

$$\liminf_{n\to\infty} (nu_n^2)^{-1} \log \boldsymbol{P}_\theta(\,\hat{\theta}_n - \theta \in u_n \bar{V}\,) \geq -\frac{1}{2} \inf_{\tau \in \bar{V}} \tau^T I(\theta) \tau \tag{2.15}$$

此外，对于任何一个顶点在原点且内部非空的锥体 K，存在

$$\liminf_{n\to\infty} (nu_n^2)^{-1} \log \boldsymbol{P}_\theta(\,\hat{\theta}_n - \theta \in u_n \bar{V} \cap K) \geq -\frac{1}{2} \inf_{\tau \in \bar{V} \cap K} \tau^T I(\theta) \tau. \tag{2.16}$$

$\theta \in \Theta \subset \boldsymbol{R}^d$ 的情况与 $\theta \in \Theta \subset \boldsymbol{R}^1$ 类似，也归结为在两个点上估计一个参数的问题。

以下局部渐近 Bahadur 效率的下界是从定理 2.3推导出来的。我们称估计量 $\hat{\theta}_n$ 局部一致收敛，如果对于任意的 $\theta_0 \in \Theta$，存在点 θ_0 的一个邻域 U，使得对于任意的 $\varepsilon > 0$，有

$$\lim_{n\to\infty} \sup_{\theta \in U} \boldsymbol{P}_\theta(\,|\hat{\theta}_n - \theta| > \varepsilon) = 0. \tag{2.17}$$

定理 2.4 设统计实验 $\mathcal{E} = \{(S, \mathcal{B}), \boldsymbol{P}_\theta, \theta \in \Theta\}$ 在点 $\theta \in \Theta \subset \boldsymbol{R}^1$ 处具有有限的费舍尔信息。设估计量 $\hat{\theta}_n$ 局部一致收敛。那么，对于任意估计量 $\hat{\theta}_n$，我们有。

$$\liminf_{u\to 0} \lim_{n\to\infty} (nu^2)^{-1} \log \boldsymbol{P}_\theta(\,\hat{\theta}_n - \theta \in u\bar{V}) \geq -\frac{1}{2} \inf_{\tau \in \bar{V}} \tau^T I(\theta) \tau. \tag{2.18}$$

此外，对于任何以零为顶点且具有非空内部的锥体 K，存在

$$\liminf_{u\to 0} \lim_{n\to\infty} (nu^2)^{-1} \log \boldsymbol{P}_\theta(\,\hat{\theta}_n - \theta \in u\bar{V} \cap K) \geq -\frac{1}{2} \inf_{\tau \in \bar{V} \cap K} \tau^T I(\theta) \tau. \tag{2.19}$$

局部一致收敛的要求替换了通常用来证明 Bahadur 渐近效率下界的相合性要求。然而，在一致性条件下，证明（2.16）需要为概率测度族 $\boldsymbol{P}_\theta$,$\theta \in \Theta$ 引入额外的正则性条件（参见 Th. 9.3, Ch.1 在 [8] 和 [6] 中的内容）。

注意函数 g 在 $\boldsymbol{L}_2$ 中的可微性要求 (2.3) 可以被 [4] 中关于函数 g 本身行为的较弱条件 (2.5)-(2.7) 所取代。

2.4 Bahadur 渐近效率的多维下界

估计量 $\hat{\theta}_n$ 称为参数 $\theta \in \Theta$ 的一致估计量，如果对于任意的 $\theta \in \Theta$，对于任意的 $\varepsilon > 0$，我们有

$$\lim_{n\to\infty} \boldsymbol{P}_\theta(\,|\hat{\theta}_n - \theta| > \varepsilon) = 0. \tag{2.20}$$

定理 2.5 令 $\hat{\theta}_n$ 为参数 $\theta \in \Theta$ 的一致估计量。令 $\theta_0 \in \Theta$。令 $\Omega \subset \boldsymbol{R}^d$ 为开集，使得 $0 \notin \Omega$ 和 $\theta_0 + \Omega \subset \Theta$。

然后，对于任意的 $\tilde{\theta} \in \theta_0 + \Omega$，我们有

$$\lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_{\theta_0}(\,\hat{\theta}_n - \theta_0 \in \Omega) \geq -\int_S \log \frac{f(x,\tilde{\theta})}{f(x,\theta_0)}\, f(x,\tilde{\theta})\, \nu(d\,x). \tag{2.21}$$

证明类似于 [3, 8]。记 $-K$ 为（2.21）的右侧。由詹森不等式，（2.21）的右侧是非正的。

我们定义指示函数 $\lambda_n = \lambda_n(\hat{\theta}_n - \theta_0) = 1$，如果 $\hat{\theta}_n - \theta_0 \in \Omega$ 和 $\lambda_n = \lambda_n(\hat{\theta}_n - \theta_0) = 0$，如果 $\hat{\theta}_n - \theta_0 \notin \Omega$。

我们将 $r = n(K + \delta)$ 与 $\delta > 0$ 放在一起。

表示

$$G_n = G_n(X_1, \ldots, X_n, \theta_0, \tilde{\theta}) = \prod_{j=1}^{n} \frac{f(X_j, \tilde{\theta})}{f(X_j, \theta_0)}. \tag{2.22}$$

我们有

$$\begin{aligned} &\boldsymbol{P}_{\theta_0}(\hat{\theta}_n - \theta_0 \in \Omega) = \boldsymbol{E}_{\theta_0}\lambda_n \\ &\geq \boldsymbol{E}_{\theta_0}(\lambda_n \mathbf{1}(G_n < \exp\{r\})) \geq \exp\{-r\}\boldsymbol{E}_{\tilde{\theta}}\Big\{\lambda_n \mathbf{1}(G_n < \exp\{r\})\Big\} \\ &\geq \exp\{-r\}(\boldsymbol{P}_{\tilde{\theta}}(\hat{\theta}_n - \theta_0 \in \Omega) - \boldsymbol{P}_{\tilde{\theta}}(G_n > \exp\{r\})). \end{aligned} \tag{2.23}$$

由于 $\hat{\theta}_n$ 是一致估计量，我们有

$$\lim_{n\to\infty} \boldsymbol{P}_{\tilde{\theta}}(\hat{\theta}_n - \theta_0 \in \Omega) \tag{2.24}$$

根据大数定律，我们有

$$\lim_{n\to\infty} \boldsymbol{P}_{\tilde{\theta}}\Big(\frac{1}{n}|G_n - n\,K| > \delta/2\Big) = 0. \tag{2.25}$$

由 (2.23) -2.25)，我们得到 (2.21).

[1] Bahadur, R.R. (1960). Asymptotic efficiency of tests and estimates. Sankhya. 22 229 – 252.

[2] Bahadur, R.R. Rates of convergence of estimates and test statistics. Ann.Math.Stat. 38(1967) 303-324

[3] Bahadur, R.R., Gupta J.C. and Zabel S.L. Large deviations of tests and estimates. In Asymptotic Theoryof Statistical Tests and Estimation.(ed. I.M. Chakravati) (1980) 33-64 Academic NY

[4] Ermakov, M.S. (2003). Asymptotically efficient statistical inference for moderate deviation probabilities. Theory Probab. Appl. 48(4) 676 – 700.

[5] Ermakov, M.S. The sharp lower bounds of asymptotic efficiency of estimators in the zone of moderate deviation probabilities. Electronic Journal of Statistics 6(2012) 2150-2184.

[6] Fu J.C. On a theorem of Bahadur on rate of convergence of point estimators. Ann.Stat. 4(1973)745-749)

[7] Hajek, J. (1972). Local asymptotic minimax and admisibility in estimation. Proc.Sixth Berkeley Symp. on Math. Statist. and Probab. 1 175 – 194, Berkeley, California Univ. Press

[8] Ibragimov, I.A. and Hasminskii, R.Z. (1981). Statistical Estimation: Asymptotic Theory. Berlin: Springer.

[9] Le Cam, L. (1972). Limits of Experiments Berlin: Springer. Proc.Sixth Berkeley Symp. on Math. Statist. and Probab. 1 245 – 261, Berkeley, California Univ. Press

[10] Radavichius, M. (1991). From asymptotic efficiency in minimax sense to Bahadur efficiency. New Trends in Probab. and Statist. V. Sazonov and T. Shervashidze (Eds) Vilnius, VSP/Mokslas 1 629 − 635

[11] Van der Vaart, A.W. (1998). Asymptotic Statistics. Cambridge, UK: Cambridge Univ. Press