arXiv:2603.09314v1 [math.ST] 10 Mar 2026

估计量偏离其目标值超过 ε 的次数的二阶渐近

Nils Lid Hjort 和 Grete Fenstad，奥斯陆大学

ABSTRACT. 设 $\{\theta_n\colon n \geq 1\}$ 是参数 θ的强相合估计量序列，其中 θ_n 基于前 n 个观测值。考虑 Q_ε，即 $|\widehat{\theta}_n - \theta| \geq \varepsilon$的次数。在另一篇论文（Hjort 和 Fenstad，1992）中，我们证明了，在自然的正则条件下，当 $\varepsilon \to 0$时，$\varepsilon^2 Q_\varepsilon$具有一个极限分布，该分布仅取决于 σ，即 $\sqrt{n}(\theta_n - \theta)$的极限分布的标准差。本文探讨了 Q_ε变量之差的一些二阶渐近性质。在 $\mathrm{E}(Q_{1,\varepsilon} - Q_{2,\varepsilon})$ 趋于 $\mathrm{E}Q_{1,\varepsilon}/\mathrm{E}Q_{2,\varepsilon}$趋于 1 的情况下，计算了其极限，由此在渐近相对效率为 1 时引入了“渐近相对亏量”的概念。该概念可用于区分极限分布相同的竞争估计量。因此，在估计正态方差的常用公式中使用分母 $n - \frac{1}{3}$，优于使用 n 和 $n-1$，事实上也优于所有其他选择；例如，它能使 ε错误的期望次数尽可能少。本文选取若干常见的估计问题，利用期望差的极限结果得到了这类结论，并将其与按 Hodges 和 Lehmann 意义下的相应渐近相对亏量分析进行了比较。此外，还得出了一些二阶分布结果。文中证明，ε 倍的 Q_ε差趋于一个变量，该变量与布朗运动相关的某些指数分布有关；Hjort 和 Khasminskii（1993）近期研究了这些分布。

KEY WORDS: 渐近相对亏损、渐近相对效率、估计量比较、损失函数、二阶最优性、布朗运动的总相对时间、 ε-偏差的次数

1. 引言。 假设 $X_1, X_2, \ldots$ 是来自某个分布 F的独立观测值序列，并令 $\theta = \theta(F)$ 为某个感兴趣的参数。假设 θ_n基于前个 n 数据点，且该估计量序列强相合，即 $\theta_n \to \theta$ 几乎必然成立。那么， Q_ε，即 $|\theta_n - \theta| \geq \varepsilon$的情形数，几乎必然是有限的。在自然的正则条件下（包括 $\sqrt{n}(\widehat{\theta}_n - \theta)$ 存在正态 $(0, \sigma^2)$ 极限），Hjort 和 Fenstad（1992，第 7 节）证明了

$$\varepsilon^2 Q_\varepsilon \to_d Q = \mathrm{Leb}\{s \geq 0\colon |W(s)| \geq s/\sigma\} = \int_0^\infty I\{|W(s)| \geq s/\sigma\}\,\mathrm{d}s, \tag{1.1}$$

其中 Leb 表示半直线上的勒贝格测度， $W(.)$ 是布朗运动过程， $I\{\ldots\}$表示示性函数。他们还证明 $\varepsilon^2 \mathrm{E}Q_\varepsilon \to \mathrm{E}Q$，这恰好等于$\sigma^2$。设 $\sqrt{n}(\widehat{\theta}_{n,j} - \theta) \to_d N\{0, \sigma_j^2\}$ 是两个相互竞争的估计量序列，并令 $Q_{j,\varepsilon}$为它们相应的 ε-偏差次数变量。那么，一个自然的渐近相对效率度量是

$$\text{a.r.e.} = \lim_{\varepsilon \to 0} \frac{\mathrm{E}Q_{1,\varepsilon}}{\mathrm{E}Q_{2,\varepsilon}} = \frac{\sigma_1^2}{\sigma_2^2}. \tag{1.2}$$

这也是基于渐近比计算的渐近相对效率的传统公式 n_1/n_2，其中 n_1 和 n_2 是方法1和方法2达到相同均方误差精度水平所需的样本量。Hjort 和 Fenstad（1992）还给出了（1.1）和（1.2）对多维参数及一般距离度量的推广，并在诸如分布函数和密度的非参数估计等更复杂的问题中，得到了 Q_ε-变量的极限结果。

基于 Q_ε的准则（1.2）只是一种“一阶度量”，无法区分具有相同极限分布的估计量序列；在这种情况下，通常 $\varepsilon^2(Q_{1,\,\varepsilon}-Q_{2,\,\varepsilon})\to 0$与$Q_{1,\,\varepsilon}/Q_{2,\,\varepsilon}\to 1$ 依概率成立。本文的目的是为 Q_ε的差建立某种“二阶”理论，从而可用于在具有相同极限分布的一类估计量中挑选出“最佳”估计量。这里的“最佳”是指当 ε 趋于零时， ε-误差的期望次数尽可能小。特别地，我们打算在一些它们并不一致的常见情形中，解决一致最小方差无偏（UMV）估计量、最大似然（ML）估计量与贝叶斯估计量之间的比较问题。

这与霍奇斯和莱曼（1970）的研究思路相近。他们采用渐近相对亏量这一度量，将其表示为样本量差的极限 n_1-n_2 ，适用于 $n_1/n_2\to 1$ 的情形；另见 Lehmann（1983，第5.2章）。第2节给出了基于 aQ_ε 的公式

$$\text{a.r.d.}=\lim_{\varepsilon\to 0}\mathrm{E}(Q_{1,\varepsilon}-Q_{2,\varepsilon}),\tag{1.3}$$

¯，针对均值参数估计这一简单情形，使用贝叶斯类估计量 $\frac{n}{n+c}X_n+\frac{c}{n+c}d$。在这个渐近相对亏损公式中， X_i 的内在偏度自然地发挥作用，而霍奇斯和莱曼的计算中则没有体现这一点。例如，渐近相对亏损公式表明，在熟悉的 $\sum_{i=1}^n(X_i-\bar{X}_n)^2/(n+c)$ 正态方差公式中， $n-\frac{1}{3}$ 是更优的分母选择。第3节讨论了这个例子以及其他与均值参数相关的例子。第4节还收录了几个技术要求更高的问题。例如，我们能够证明，作为正态均值平方的估计量，$(\bar{X}_n)^2+\hat{\sigma}_n^2/n$ 预期会比所有其他形式为 $(\bar{X}_n)^2+d\hat{\sigma}_n^2/n$ 的估计量产生更少的 ε 误差。

从决策论角度，可以用 EQ_ε 来比较不同的估计方法，其中基础损失函数等于整个估计序列中 ε- 误差的数量。第 5 节简要讨论了这一点，并涉及一些贝叶斯考量。

第 2–5 节所阐述的理论和应用只涉及 Q_ε 变量的期望值。这体现在所采用的工具上，即埃奇沃思展开以及对概率作进一步的 Taylor型逼近。第6节简要补充了相关内容，探讨 Q_1分布 二阶性质的 $_\varepsilon-Q_{2,\,\varepsilon}$ 差异。在结构上最简单的均值参数估计情形中，结果表明

$$\varepsilon^2(Q_{1,\varepsilon}-Q_{2,\varepsilon})\to_p 0 \text{ and } Q_{1,\varepsilon}/Q_{2,\varepsilon}\to_p 1 \quad \text{while} \quad \varepsilon(Q_{1,\varepsilon}-Q_{2,\varepsilon})\to_d A-B,$$

例如，其中 A 和 B 是与布朗运动相关的某些总相对时间变量，其分布为指数分布、指数分布的混合，或在零点处的单位点质量；关于这些内容的独立论述，参见 Hjort 和 Khasminskii（1993）。

本文主要报告一项研究，探讨如何区分一阶等价序列这一特定方法。我们并不声称，我们的二阶比较所依据的 a.r.d.判据 (1.3) 在统计学上比其他判据更自然；但它确实可以计算，这一点颇为有趣，而考察它与其他二阶判据相比表现如何也很有意义，因为这类判据还有好几种。除了 Hodges 和 Lehmann 以及本文所做的 a.r.d.计算外，Rao（1962年）等人研究了渐近充分性，Pfanzagl（1973年）等人研究了覆盖概率和中位数无偏性，Ghosh 和 Subramanyam（1974年）等人则研究了均方误差的展开式。Berkson（1980年）对该论文的讨论是进一步了解相关内容的良好来源。最后指出，本文的方法可用于建立某类覆盖^概率的展开结果，即 $p_n=\Pr\{|\theta_n-\theta|\le k/\sqrt{n}\}$。这些结果在某种程度上比第2–5节中的结果更简单，进而可用于比较具有相同极限分布的不同估计量。不过，这类比较取决于 _k 的取值。

2. 均值估计的一般结果。假设 X_1，X_2，X_3，.....是独立同分布的，且 $\mathrm{E}X_i = \xi$、Var $X_i = \sigma^2$，偏度为 $\mathrm{E}(X_i - \xi)^3/\sigma^3 = \gamma$。下面的命题1和命题2给出了最有用且最直接的信息，但出于技术上的考虑，先介绍以下引理更为方便。

LEMMA. 令 a 为正数，并考虑 $Q_\varepsilon(c)$，即在 $n \geq a/\varepsilon^2$ 中满足 $|\frac{n}{n+c}X_n - \xi| \geq \varepsilon$ 的次数。假设 X_i 具有有限四阶矩，且其分布为非格点分布，并记 $\phi(\cdot)$ 为标准正态密度。那么

$$\begin{aligned}\lambda_a(c) &= \lim_{\varepsilon\to 0} \mathrm{E}\{Q_\varepsilon(c) - Q_\varepsilon(0)\} \\ &= 2\Big[\frac{\xi^2}{\sigma^2}c^2 - \Big(2 - \frac{2\gamma}{3}\frac{\xi}{\sigma}\Big)c\Big]\int_{\sqrt{a}/\sigma}^{\infty}\phi(y)\,\mathrm{d}y - \frac{2\gamma}{3}\frac{\xi}{\sigma}\frac{\sqrt{a}}{\sigma}\phi(\sqrt{a}/\sigma)\,c.\end{aligned} \tag{2.1}$$

PROOF: 我们选择将 $Q_\varepsilon(c)$ 写成如下形式

$$Q_\varepsilon(c) = \sum_{n\geq am} I\Big\{\Big|\frac{n}{n+c}\bar{X}_n - \xi\Big| \geq \frac{1}{\sqrt{m}}\Big\}, \tag{2.2}$$

记 $m = 1/\varepsilon^2$。注意，根据强大数定律，$Q_\varepsilon(c)$ 几乎处处有限。考虑 $T_n = \sqrt{n}\,(X_n - \xi)/\sigma$。代入 $X_n = \xi + \sigma T_n/\sqrt{n}$ 后，(2.2) 中的示性函数项可写为 $I\{T_n \leq l(c)$ 或 $T_n \geq r(c)\}$，其中

$$l(c) = -\frac{1}{\sigma}\sqrt{\frac{n}{m}} + \frac{1}{\sqrt{n}}\frac{c\xi}{\sigma} - \frac{1}{\sqrt{nm}}\frac{c}{\sigma}, \quad r(c) = \frac{1}{\sigma}\sqrt{\frac{n}{m}} + \frac{1}{\sqrt{n}}\frac{c\xi}{\sigma} + \frac{1}{\sqrt{nm}}\frac{c}{\sigma}. \tag{2.3}$$

令 $s = n/m$，则有 $l(0) = -u$，$l(c) = -u + b_1/\sqrt{n} - b_2/n$，$r(0) = u$，$r(c) = u + b_1/\sqrt{n} + b_2/n$，其中 $u = \sqrt{s}/\sigma$、$b_1 = c\xi/\sigma$ 和 $b_2 = c\sqrt{s}/\sigma$。由于

$$Q_\varepsilon(c) - Q_\varepsilon(0) = \sum_{n\geq am}\Big[I\big\{T_n \leq l(c) \text{ or } T_n \geq r(c)\big\} - I\big\{T_n \leq l(0) \text{ or } T_n \geq r(0)\big\}\Big] \tag{2.4}$$

其期望值可以写成

$$E\{Q_\varepsilon(c) - Q_\varepsilon(0)\} = \sum_{n\geq am}\Big[\big\{G_n(l(c)) - G_n(l(0))\big\} - \big\{G_n(r(c)) - G_n(r(0))\big\}\Big], \tag{2.5}$$

关于累积分布函数 G_n of T_n 的表达式。

在 满足非格点假设时，G_n of G_n 的分布存在 Cramér–Edgeworth 展开，适用于 m

$$G_n(t) = \Phi(t) - \frac{1}{6}\frac{\gamma}{\sqrt{n}}A(t)\phi(t) + \frac{1}{n}R(t)\phi(t) + O(n^{-3/2}), \tag{2.6}$$

其中 $A(t) = t^2 - 1$，而 $R(t)$ 是一个五次多项式，参见例如 Barndorff-Nielsen 和 Cox（1989，第4章）。鉴于（2.5），我们需要近似形如 $G_n(t+\delta) - G_n(t)$ 的项，其中 δ 的阶为 $1/\sqrt{n}$。利用泰勒展开可得

$$G_n(t+\delta) - G_n(t) = \phi(t)\delta - \frac{1}{2}t\phi(t)\delta^2 + \frac{1}{6}\frac{\gamma}{\sqrt{n}}B(t)\phi(t)\delta + O(n^{-3/2}),$$

其中 $B(t)\phi(t)$ 是 $-A(t)\Phi(t)$ 的导数，即 $B(t) = t^3 - 3t$，且 O 项关于 t 一致。由此得到

$$\begin{aligned}G_n(-u + b_1/\sqrt{n} - b_2/n) - G_n(-u) &= \phi(u)b_1/\sqrt{n} + \phi(u)\{-b_2 + \tfrac{1}{2}ub_1^2 - \tfrac{1}{6}\gamma b_1 B(u)\}/n, \\ G_n(u + b_1/\sqrt{n} + b_2/n) - G_n(u) &= \phi(u)b_1/\sqrt{n} + \phi(u)\{b_2 - \tfrac{1}{2}ua^2 + \tfrac{1}{6}\gamma b_1 B(u)\}/n.\end{aligned}$$

因此，式 (2.5) 中的被加数除去 $O(n^{-3/2})$ 项后，可表示为

$$\frac{1}{n}\phi(u)\Big\{-2b_2+ub_1^2-\frac{1}{3}\gamma b_1 B(u)\Big\}=\frac{1}{m}\phi\Big(\frac{\sqrt{s}}{\sigma}\Big)\frac{1}{s}\Big\{-2\frac{\sqrt{s}}{\sigma}c+\frac{\xi^2\sqrt{s}}{\sigma^3}c^2-\frac{\gamma}{3}\frac{\xi}{\sigma}B\Big(\frac{\sqrt{s}}{\sigma}\Big)c\Big\}.$$

由于 $\frac{1}{m}\sum_{n/m\ge a}h(n/m)$ 是积分 $\int_a^\infty h(s)\,\mathrm{d}s$ 的黎曼近似，并且当 $m\to\infty$ 时收敛到这一极限，因此最终我们得到 $\mathrm{E}\{Q_\varepsilon(c)-Q_\varepsilon(0)\}$ 收敛到

$$\begin{aligned}\lambda_a(c)&=\int_a^\infty\Big[\frac{\xi^2}{\sigma^3}\frac{1}{\sqrt{s}}c^2-\frac{2}{\sigma}\frac{1}{\sqrt{s}}c-\frac{\gamma}{3}\frac{\xi}{\sigma}\frac{1}{s}B\Big(\frac{\sqrt{s}}{\sigma}\Big)c\Big]\,\phi\Big(\frac{\sqrt{s}}{\sigma}\Big)\,\mathrm{d}s\\&=2\int_{\sqrt{a}/\sigma}^\infty\Big[\frac{\xi^2}{\sigma^2}c^2-2c-\frac{\gamma}{3}\frac{\xi}{\sigma}\frac{B(u)}{u}c\Big]\,\phi(u)\,\mathrm{d}u.\end{aligned}$$

利用 $B(u)/u=u^2-3$ 进行进一步分析，最终可证明 (2.1)。 □

在下面给出的应用中，我们将研究这个极限及类似极限关于 c 的函数形式，以找出预期能使 ε-误差最少的估计量序列。然而，根据 (2.1) 计算出的 c 最优值将取决于 a 这一多少有些任意的数值。一般地，记 $\bar{Q}_{\varepsilon,\,m}(c)$ 为 $\frac{n}{n+c}X_n$ 在 $n\ge m$ 个案例中所犯的 ε-误差数。关于 $\lambda_a(c)$ 的极限结果 (2.1) 与 $Q_{\varepsilon,\,a/\varepsilon^2}(c)$ 有关，因此自然地令 a 趋于零。令 (2.1) 中的 $a\to 0$ 趋于零所得到的极限函数 $\lambda_0(c)$，并不完全等于 $Q_{\varepsilon,\,1}(c)$ 与 $Q_{\varepsilon,\,1}(0)$ 之差的期望的极限，因为余项 $O(\sum_{n\ge am}n^{-3/2})$ 在 $a=1/m$ 时不会趋于零。不过，以下结论成立。

PROPOSITION 1. 令 $Q_\varepsilon(c)$ 为 $\frac{n}{n+c}X_n$ 中 ε-次失误的次数，这次是在 $n\ge a(\varepsilon)/\varepsilon^2$ 中计数，其中 $a(\varepsilon)\to 0$，而 $a(\varepsilon)/\varepsilon^2\to\infty$（例如，与 $a(\varepsilon)=\varepsilon$ 一样）。那么

$$\lambda_0(c)=\lim_{\varepsilon\to0}\mathrm{E}\big\{Q_{\varepsilon,a(\varepsilon)/\varepsilon^2}(c)-Q_{\varepsilon,a(\varepsilon)/\varepsilon^2}(0)\big\}=\frac{\xi^2}{\sigma^2}\,c^2-2\Big(1-\frac{\gamma}{3}\frac{\xi}{\sigma}\Big)\,c. \tag{2.7}$$

有时，将目标 X_n 与某个先验值进行权衡是合理的。贝叶斯学派常使用类型为 $\xi_n(c,d)=\frac{n}{n+c}X_n+\frac{c}{n+c}\xi_0$ 的估计量，参见第 5 节。因此，考虑以下略微扩展的形式

$$Q_\varepsilon(c,d)=\sum_{n\ge a/\varepsilon^2}I\Big\{\Big|\frac{n}{n+c}\bar{X}_n+\frac{c}{n+c}d-\xi\Big|\ge\varepsilon\Big\}.$$

发现（2.3）的恰当推广为

$$l(c,d)=-\frac{1}{\sigma}\sqrt{\frac{n}{m}}+\frac{1}{\sqrt{n}}\frac{c(\xi-d)}{\sigma}-\frac{1}{\sqrt{nm}}\frac{c}{\sigma},\quad r(c,d)=\frac{1}{\sigma}\sqrt{\frac{n}{m}}+\frac{1}{\sqrt{n}}\frac{c(\xi-d)}{\sigma}+\frac{1}{\sqrt{nm}}\frac{c}{\sigma},$$

对先前的论证作相应调整并重复，即可证明（2.7）的以下推广。

PROPOSITION 2. 如上所述，令 ε 和 $a(\varepsilon)=\varepsilon$ 趋于零。那么

$$\lambda_0(c,d)=\lim_{\varepsilon\to0}\mathrm{E}\big\{Q_\varepsilon(c,d)-Q_\varepsilon(0,0)\big\}=\frac{(\xi-d)^2}{\sigma^2}\,c^2-2\Big(1-\frac{\gamma}{3}\frac{\xi-d}{\sigma}\Big)\,c. \tag{2.8}$$

下文 4C 节末尾给出了 (2.7) 和 (2.8) 的进一步推广。我们将这些称为 a.r.d.公式，参见 (1.3) 及其中的相关说明。我们还指出，证明 (2.1)、(2.7) 和 (2.8) 另有一种方法，即采用分布分析，并沿用第6节简要讨论的布朗运动近似。有关这一替代证明的细节见技术报告 Hjort和Fenstad（1993），可向作者索取。

REMARK. 在这种情形下，还可以计算霍奇斯和莱曼的渐近相对亏损，其定义为 $n_0(c,d) - n_0$ 的极限，其中 $n_0(c,d)$ 是为达到 $\mathrm{E}\{\widehat{\xi}_n(c,d) - \xi\}^2 = \mathrm{E}\{\widehat{\xi}_{n_0}(0,0) - \xi\}^2$ 所需的样本量；相关讨论见 Lehmann（1983，第 5.2 节）。那么，对于所有 (c,d)，$n_0(c,d)/n_0(0,0) \to 1$，但可以证明

$$\mathrm{a.r.d.}_{\mathrm{hl}}(c,d) = \lim_{n_0\to\infty}\{n_0(c,d) - n_0(0,0)\} = \frac{(\xi-d)^2}{\sigma^2}c^2 - 2c. \tag{2.9}$$

a.r.d. 公式 (2.8) 与此不同之处在于，偏度 γ 也以自然的方式参与其中。 □

3. 一些应用。

3A. 正态均值。设 X_i 服从正态分布（θ,1）。考虑估计量序列 $\theta_n(c,d) = \frac{n}{n+c}X_n + \frac{c}{n+c}d$，并令 $Q_\varepsilon(c,d)$ 表示 ε-未命中的次数。这符合第2节的一般框架，其中 $\xi=\theta$、$\sigma=1$ 和 $\gamma=0$。因此

$$\lambda_0(c,d) = \lim_{\varepsilon\to 0}\mathrm{E}_\theta\big\{Q_\varepsilon(c,d) - Q_\varepsilon(0,0)\big\} = (\theta-d)^2c^2 - 2c.$$

若对 θ 按照某个分布取平均，该分布的“先验均值”为 $\mathrm{E}\theta = \theta_0$，“先验方差”为 $\mathrm{Var}\ \theta = \tau^2$，则可得 $\{\tau^2 + (\theta_0 - d)^2\}c^2 - 2c$；当 d 取为 θ_0 且 c 取为 $1/\tau^2$ 时，该式达到最小值。这为使用……提供了新的、独立的动机

$$\theta_n^* = \frac{n}{n+1/\tau^2}\bar{X}_n + \frac{1/\tau^2}{n+1/\tau^2}\theta_0 \tag{3.1}$$

在存在此类先验知识的情况下，并且与正态模型中常见的贝叶斯计算以及精算统计学中所谓的可信度公式相一致。请注意，若真实的 θ 参数位于 θ_0 的 $\sqrt{2}\tau$ 范围之内，则 θ_n^* 预计会比 $\bar{X}_n$ 犯更少的 ε 误差。另请注意，就先验权重函数的选择而言，只有 θ_0 和 τ^2 是重要的。关于风险和平均风险的进一步讨论见第 5 节。

3B. 指数分布均值。设 $X_i \sim \mathrm{Exp}(1/\theta)$，并考虑 $Q_\varepsilon(c)$，即 $|\theta_n(c)/\theta - 1| \ge \varepsilon$ 的次数，其中 $\theta_n(c) = \frac{n}{n+c}X_n$。这可以写成 $|\frac{n}{n+c}Y_n - 1| \ge \varepsilon$，其中诸 Y_i 来自 e^{-y}。这与第2节中 $\xi=1$、$\sigma=1$、$\gamma=2$ 的情形相同。所关注的极限是 $c^2 - \frac{2}{3}c$，最佳值为 $c_0 = \frac{1}{3}$。采用 $c=0$ 的最大似然解预计会多犯 19 个 ε 误差 [原文如此]，而在平方误差损失下使用 $c=1$ 的最佳估计量预计会多犯 49 个 ε 误差。

3C. 正态方差。接下来考虑 $\widehat{\sigma}_N^2(c) = \sum_{i=1}^N (Y_i - \bar{Y}_N)^2/(N-1+c)$，利用数据 Y_1、……、Y_N 估计方差 σ^2，其中数据服从正态分布（μ，σ^2）。研究次数 $|\widehat{\sigma}_N^2(c)/\sigma^2 - 1| \ge \varepsilon$。这可以写作 $|\frac{n}{n+c}(\chi_n^2/n) - 1| \ge \varepsilon$，其中将 n 记作 $N-1$，即自由度。这与第2节中的情形再次相同，不过这次由 χ_1^2 个变量扮演 X_i 的角色。它们具有 $\xi=1$、$\sigma=\sqrt{2}$ 和 $\gamma = 2\sqrt{2}$。要研究的极限是 $\frac{1}{2}c^2 - \frac{2}{3}c$，其最小值在 $c_0 = \frac{2}{3}$ 时取得。因此

$$\widehat{\sigma}_N^2 = \frac{1}{N-(1/3)}\sum_{i=1}^N (Y_i - \bar{Y}_N)^2 \tag{3.2}$$

可以预期，ε 错误的次数会最少！另请参见下文第4C节。

3D. 二项概率。令 Y_n 服从二项分布 (n,p)，并令 $Q_\varepsilon(c,d)$ 统计 $(Y_n + cd)/(n+c)$ 的 ε 次未命中。注意，Y_n 是若干个 X_i 的和，其均值为 p、方差为 pq，偏度为 $\gamma = (q-p)/(pq)^{1/2}$，其中 $q = 1-p$。因此，(2.8) 中的风险差极限变为

$$\mathrm{risk}(c,d) = \frac{(p-d)^2}{pq}c^2 - 2c - \frac{2}{3}\frac{(p-q)(p-d)}{pq}c.$$

在研究了带有"先验猜测"$d=\frac{1}{2}$的情形后，可以证明 $p_n^*=(Y_n+2/3)/(n+4/3)$是二阶极小极大估计量序列，其意义在于使最大可能风险最小化。无论潜在的 p如何，它似乎都可以预期比 Y_n/n 少犯2.667个 ε误差!

然而，这里存在一个障碍： X_i 的分布是格点分布，Cramér 条件不成立，因此需要经过精细调整的修正项，才能使类似 (2.6) 的公式成立。Kolassa 和 McCullagh (1990) 的方法最终可能会给出类似 (2.8) 的、适用于格点情形的公式，但我们没有继续探究这一点。要避免这里与精确 a.r.d. 计算相关的困难，一个可能的技巧是将 X_i的概率质量 p 均匀分散到某个极小的 $[1-\eta,1+\eta]$ 上。这样，第2节的结果便适用，从而得到针对 p的最佳估计序列。最终令 $\eta\to 0$，便可选出最佳的 $(Y_n+2/3)/(n+4/3)$ 序列。最后请注意，Beta$(2/3,\ 2/3)$ 是 p的最不利先验，其意义如上文计算所示。

4. 更多示例。本节旨在为几个技术上要求更高的估计问题建立结果。在每种情况下，为了确定期望Q_ε差异的极限，都必须对第2节中使用的论证作出一些较为费力的修改。

4A. 正态模型中的均值平方：方差已知。设 X_i 服从正态分布（$\xi,\ \sigma^2$），其中 σ已知，并假设要估计 $\theta=\xi^2$。最大似然解为（$\bar{X}_n$）^{2},，它可能会高估；一致最小方差解为（$\bar{X}_n$）$^2-\sigma^2/n$。因此，考察 $\widehat{\theta}_n$（d）$=$（$\bar{X}_n$）$^2-d\sigma^2/n$ 及其对应的Q_ε（d），即 $n\ge a$（ε）$/\varepsilon^2$中 ε次未命中的数量，其中 a（ε）像（2.7）中那样趋于零。我们打算证明

$$\lambda_0(d)=\lim_{\varepsilon\to 0}\mathrm{E}\{Q_\varepsilon(d)-Q_\varepsilon(0)\}=(\tfrac{1}{4}d^2+\tfrac{1}{2}d)\frac{\sigma^2}{\xi^2}. \tag{4.1}$$

有必要对第2节的论述作一些修改。 Q_ε(d)中的示性函数求和项在 $|(\xi+T_n\sigma_n/\sqrt{n})^2-d\sigma^2/n-\xi^2|\ge 1/\sqrt{m}$时为1，其中 $T_n=\sqrt{n}(\bar{X}_n-\xi)/\sigma$。这可以写成

$$T_n^2+2(\xi/\sigma)\sqrt{n}T_n\ge d+\sqrt{n}\sqrt{s}/\sigma^2\quad\text{or}\quad T_n^2+2(\xi/\sigma^2)\sqrt{n}T_n\le d-\sqrt{n}\sqrt{s}/\sigma^2,$$

其中 $s=n/m$再次出现。令 $b(d)<r(d)$为第一个不等式的两个根，令 $c(d)<l(d)$为第二个不等式的两个根。现在假设 ξ 为正数。那么 $b(d)<c(d)<l(d)<0<r(d)$，而示性函数项为

$$I\{T_n\le b(d)\text{ or }T_n\ge r(d)\text{ or }c(d)\le T_n\le l(d)\}.$$

事实证明， $b(d)$和 $c(d)$的量级都约为 $-2(\xi/\sigma)\sqrt{n}$，这在左侧偏得太远，不具有重要性，因此我们可以集中考察 $I\{T_n\le l(d)$或 $T_n\ge r(d)\}$。我们需要计算

$$\mathrm{E}\{Q_\varepsilon(d)-Q_\varepsilon(0)\}\doteq\sum_{n\ge am}\Big[\big\{\Phi(l(d))-\Phi(l(0))\big\}-\big\{\Phi(r(d))-\Phi(r(0))\big\}\Big], \tag{4.2}$$

参见 (2.4) 和 (2.5)。坚定的头脑会发现

$$\begin{aligned}l(d)&=-\sqrt{n}\frac{\xi}{\sigma}+\sqrt{n}\frac{\xi}{\sigma}\Big(1-\frac{1}{\sqrt{n}}\frac{\sqrt{s}}{\xi^2}+\frac{1}{n}\frac{d\sigma^2}{\xi^2}\Big)^{1/2}\\&\doteq-\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}+\frac{1}{\sqrt{n}}\Big(\frac{1}{2}\frac{d\sigma}{\xi}-\frac{1}{8}\frac{s}{\xi^3\sigma}\Big)+\frac{1}{n}\Big(\frac{1}{4}d\sqrt{s}\frac{\sigma}{\xi^3}-\frac{1}{16}\frac{s^{3/2}}{\xi^5\sigma}\Big)\\&=-u+\delta_1(d)/\sqrt{n}+\delta_2(d)/n,\end{aligned}$$

$$
\begin{aligned}
r(d) &= -\sqrt{n}\frac{\xi}{\sigma} + \sqrt{n}\frac{\xi}{\sigma}\Big(1 + \frac{1}{\sqrt{n}}\frac{\sqrt{s}}{\xi^2} + \frac{1}{n}\frac{d\sigma^2}{\xi^2}\Big)^{1/2} \\
&\doteq \frac{1}{2}\frac{\sqrt{s}}{\xi\sigma} + \frac{1}{\sqrt{n}}\Big(\frac{1}{2}\frac{d\sigma}{\xi} - \frac{1}{8}\frac{s}{\xi^3\sigma}\Big) + \frac{1}{n}\Big(-\frac{1}{4}d\sqrt{s}\frac{\sigma}{\xi^3} + \frac{1}{16}\frac{s^{3/2}}{\xi^5\sigma}\Big) \\
&= u + \delta_1(d)/\sqrt{n} - \delta_2(d)/n.
\end{aligned}
$$

现在可以对 (4.2) 中的各项计算泰勒近似。除$O(n^{-3/2})$ 项外，结果为

$$
\begin{aligned}
&2\phi(u)\{\delta_2(d) - \delta_2(0)\}/n + \{\tfrac{1}{2}\phi'(-u) - \tfrac{1}{2}\phi'(u)\}\{\delta_1(d)^2 - \delta_1(0)^2\}/n \\
&\quad = \frac{1}{ms}\phi\Big(\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\Big)\Big[\frac{1}{2}d\sqrt{s}\frac{\sigma}{\xi^3} + \Big(\frac{1}{4}\frac{d^2\sigma^2}{\xi^2} - \frac{1}{8}\frac{ds}{\xi^4}\Big)\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\Big].
\end{aligned}
$$

进一步仔细推敲并关注细节，便可得到（4.1）。

当 $\xi < 0$ 时，类似的论证会得出相同的答案 $(\frac{1}{4}d^2 + \frac{1}{2}d)\ \sigma^2/\xi^2$。我们可以得出结论，选择 $d_0 = -1$ 最佳：在所考虑的估计量类别中，估计量 $(\bar{X}_n)^2 + \sigma^2/n$ 具有二阶最优性，即在 $\varepsilon \to 0$ 的极限下，其期望的 ε- 误差次数最少。最大似然解 $(\bar{X}_n)^2$ 预计会多产生 $\frac{1}{4}\sigma^2/\xi^2$ 个 ε- 误差，而一致最小方差无偏估计量 $(\bar{X}_n)^2 - \sigma^2/n$ 预计会多产生 σ^2/ξ^2 个 ε-误差。

一个自然的推广是纳入所有形如 $\widehat{\theta}_n(c,\ d) = (\frac{n}{n+c}\bar{X}_n)^2 - d\sigma^2/n$ 的估计量。Hjort和Fenstad（1993）给出了相关分析，并以 ξ^2/σ^2 和 σ^2/ξ^2 的“先验平均”为依据，求得 c 和 d 的最佳可能取值；另请参见下文第 5 节。Hjort和Fenstad（1993）还分析了特殊情形 $\xi = 0$，在此情形下渐近行为有所不同。

4B. 正态模型中的均值平方：未知方差。 推广到未知 σ的情形很重要，因为该参数通常是 未知的，这当然如此，但也因为稍后给出的方法可用于说明如何在其他双参数及多参数情形中解决类似问题。

自然要研究的估计量类是 $\theta_n^*(d) = (\bar{X}_n)^2 - d\widehat{\sigma}_n^2/n$，其中我们取 $\widehat{\sigma}_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$ 为无偏版本。同样， $d = 0$给出最大似然解，而 $d = 1$给出一致最小方差无偏估计量。记 $\widehat{\sigma}_n^2 = \sigma^2 Z_n$，其中 Z_n 与 $\bar{X}_n$ 独立，且服从$\chi^2_{n-1}/(n-1)$分布，从而事实上 $\theta_n^*(d) = \widehat{\theta}_n(dZ_n)$，采用4A中的记号。那里的论证表明， $Q_\varepsilon^*(d)$，即 $|\theta_n^*(d) - \xi^2| \geq \varepsilon$的情形数，本质上统计的是 $T_n \leq l(dZ_n)$或 $T_n \geq r(dZ_n)$的情形。(4.2)的类似式变为

$$
\sum_{n \geq am} \Big\{ \big[\Pr\{T_n \leq l(dZ_n)\} - \Pr\{T_n \leq l(0)\}\big] - \big[\Pr\{T_n \leq r(dZ_n)\} - \Pr\{T_n \leq r(0)\}\big] \Big\}.
$$

随着 $m = 1/\varepsilon^2$ 增大，计算该和的极限的最简单方法是先对 Z_n 的取值进行条件化，使用 4A 的结果，然后对 Z_n 的分布 $g_n(z_n)$ 进行积分。完成这些步骤后可得

$$
\sum_{n \geq am} \int_0^\infty \frac{1}{ms}\phi\Big(\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\Big)\Big[\frac{1}{2}d\sqrt{s}z_n\frac{\sigma}{\xi^3} + \frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\Big(\frac{1}{4}d^2z_n^2\frac{\sigma^2}{\xi^2} - \frac{1}{8}dz_n\frac{s}{\xi^4}\Big)\Big]\, g_n(z_n)\,\mathrm{d}z_n.
$$

但是，由于 $EZ_n = 1$ 和 $EZ_n^2 = 1 + 2/(n-1)$，最终该极限等于之前的答案 $(\frac{1}{4}d^2 + \frac{1}{2}d)\ \sigma^2/\xi^2$。我们可以得出结论： $(\bar{X}_n)^2 + \widehat{\sigma}_n^2/n$ 是最佳选择。

将这一结果与使用霍奇斯和莱曼亏损所得的相应结果进行比较很有意思。对$\theta_n^*(d)$的均方误差进行计算可得，$\mathrm{a.r.d.}_{\mathrm{hl}}$，即该

样本量差 $n_0(d) - n_0(0)$变为$(\frac{1}{4}d^2 - \frac{1}{2}d)\ \sigma^2/\xi^2$，而按照这一准则的最佳估计量是一致最小方差解$(\bar{X}_n)^2 - \hat{\sigma}_n^2/n$。

4C. 正态模型中的标准差。 接下来我们介绍例3C的一个变体。我们不再计算 $|\hat{\sigma}_N^2(c)/\sigma^2 - 1| \geq \varepsilon$的次数，而是使用 σ的自然尺度来惩罚误差，并研究

$$Q_\varepsilon(c) = \sum_{N \geq a/\varepsilon^2} I\{|\hat{\sigma}_N(c)/\sigma - 1| \geq \varepsilon\}. \tag{4.3}$$

二阶渐近方法中更精细的方面表明，对于 c的最佳取值，答案与3C中的不同。

为了找到这次的最佳取值，我们对（2.7）和（2.8）作一个不大但很有用的推广。假设 $h(\xi)$是一个光滑递增变换，并考虑 $Q_\varepsilon^*(c,\ d)$，即在 $n \geq 1/\varepsilon$ 中（见（2.7））， $|h(\xi_n(c,\ d)) - h(\xi)| \geq \varepsilon$ 的次数，其中 $\xi_n(c,\ d) = (nX_n + cd)/(n+c)$如第2节命题2所述。我们可以沿用前面的论证，补充进一步的泰勒分析，并证明

$$\lambda_0^*(c,d) = \lim_{\varepsilon \to 0} \mathrm{E}\{Q_\varepsilon^*(c,d) - Q_\varepsilon^*(0,0)\} = \frac{(\xi - d)^2}{\sigma^2}c^2 + \left\{-2 + \frac{2\gamma}{3}\frac{\xi - d}{\sigma} + \frac{k_2}{k_1^2}(\xi - d)\right\}c, \tag{4.4}$$

其中 $k_1 = (h^{-1})'(h(\xi))$，且 $k_2 = (h^{-1})''(h(\xi))$。必要的细节见 Hjort和Fenstad（1993）。根据基本规则，还可将 k_2/k_1^2 写成 $-h''(\xi)/h'(\xi)$。特殊情形 $h(\xi) = \xi$ 满足 $k_2 = 0$，并还原为 (2.8)。

为分析与 (4.3) 相关的问题，我们可以令 $h(\xi) = \xi^{1/2}$, $d = 0$，并在最后使用 $\xi = 1$。由此得到

$$\lambda_0(c) = \lim_{\varepsilon \to 0} \mathrm{E}\{Q_\varepsilon(c) - Q_\varepsilon(0)\} = c^2/\tau^2 + ((2/3)\gamma/\tau - 3/2)c,$$

用标准差 τ 和偏度 γ表示。在特定的 χ_1^2 情形下，我们得到 $\lambda_0(c) = \frac{1}{2}c^2 - \frac{1}{6}c$，其最小值在 $c_0 = \frac{1}{6}$时取得。由此可得，估计量序列$\tilde{\sigma}_N = \{(N - (5/6))^{-1}\sum_{i=1}^N (Y_i - \bar{Y}_N)^2\}^{1/2}$ 预计会使类型为$|\hat{\sigma}_N/\sigma - 1| \geq \varepsilon$的错误最少。我们还使用 (4.4) 展示预计能产生最少对数尺度上 ε-错误的估计量序列。令 $Q_\varepsilon^*(c)$ 统计 $|\log \hat{\sigma}_N^2(c) - \log \sigma^2| \geq \varepsilon$ 的次数。可以证明，其极限为 $\lambda_0^*(c) = \frac{1}{2}c^2 - (\frac{2}{3} - e^{-1})c$，并在 $c_0 = \frac{2}{3} - e^{-1}$ 时达到最小值。因此，使用分母 $N - 1 + c_0$，即 $\sigma_N^* = \{(N - 0.695)^{-1}\sum_{i=1}^N (Y_i - \bar{Y}_N)^2\}^{1/2}$，预计对于 σ^p，在对数尺度上会使错误最少，此结论对 p 的任意取值均成立。

Remark. 公式（3.2）结合 $N - 1/3$ 以及最近推导出的两个 $N - 5/6$ 和 N –0.695，为各种统计原理得出的方差估计公式 $\sum_{i=1}^N (Y_i - \bar{Y}_n)^2/(N - 1 + c)$中的分母名单增添了三个新成员。该名单包括：(i) 基于正态分布或非参数的最大似然估计量，其 $1/N$。(ii) 在正态或非参数情形下，方差最小的 σ^2 无偏估计量，其 $1/(N-1)$。(iii) 若希望 $\mathrm{E}\hat{\sigma}_N(c) = \sigma$，即在自然尺度上满足无偏性，则需要 $\mathrm{E}(\chi_n^2)^{1/2}/(n+c)^{1/2} = 1$；再次使用 $n = N - 1$，其近似解为$c = -\frac{1}{2}$，从而得到 $1/(N - \frac{3}{2})$。(iv) 另一种可能是最小化均方误差 $\mathrm{E}\sigma(\hat{\sigma}_N^2 - \sigma^2)^2$。最优常数为 $1/(N+1)$。对于 σ^2，这也给出了平方误差损失下的最优不变估计量。(v) 使用自然尺度来解决这个问题也很自然，即最小化 $\mathrm{E}\sigma(\hat{\sigma}_N - \sigma)^2$。同样可以对 σ一致地求解，解为 $c = n^2/\{\mathrm{E}(\chi_n^2)^{1/2}\}^2 - n$，其中 1 2 是一个近似值。最优分母为 $N - \frac{1}{2}$，同时也给出了针对 σ、平方误差损失下的最优不变估计量。(vi) 根据威尔逊–希尔费蒂公式，χ_n^2 的中位数近似为 $n - 2/3 + (4/27)/n$。因此，分母 $N - \frac{5}{3} + \frac{4}{27N} \doteq N - \frac{5}{3}$ 给出

近似中位数无偏估计量；它高估和低估的次数相同。(vii) 由于 $\mathrm{E}\log(\chi^2_n) = \log 2 + \psi(n/2) \doteq \log n - (\frac{1}{n} + \frac{1}{3n^2})$，可得 $\tilde{\sigma}$ 采用分母 $N-2+\frac{1}{6N} \doteq N-2$ 时，在对数尺度上无偏。这一性质很好，一方面是因为尺度的对数是一个自然量，另一方面是因为对于 p 的任意取值，$\tilde{\sigma}^p$ 对 σ^p 都具有对数无偏性。(viii) 在模糊先验下的贝叶斯解（$\log\sigma$ 在数轴上服从均匀分布）中，若损失是在 σ^2 尺度上的平方误差，则使用 $N-3$；若损失是在 $1/\sigma^2$ 尺度上的平方误差，则使用 $N-1$。

5. 损失函数与贝叶斯解。

5A. 决策论框架。谁能明白自己的过错（诗篇19:12）？谁又能数算它们？我们本质上是在使用一种略非标准的损失函数 L_ε，它对给定的估计序列 $\{\theta_n\}$ 进行计数

$$L_\varepsilon\big[\theta, \{\widehat{\theta}_n : n \ge 1\}\big] = \sum_{n \ge a(\varepsilon)/\varepsilon^2} I\{|\widehat{\theta}_n - \theta| \ge \varepsilon\}, \tag{5.1}$$

在全部 $n \ge a(\varepsilon)/\varepsilon^2$ 情形中 ε 误差的个数，其中 ε 非常小。相应的风险函数为

$$R_\varepsilon(\theta) = \mathrm{E}_\theta L_\varepsilon\big[\theta, \{\widehat{\theta}_n : n \ge 1\}\big] = \sum_{n \ge a(\varepsilon)/\varepsilon^2} \Pr_\theta\{|\widehat{\theta}_n - \theta| \ge \varepsilon\}. \tag{5.2}$$

在某些情形下，该级数发散，而差分风险 $R_{1,\varepsilon}(\theta) - R_{2,\varepsilon}(\theta)$（比方说）可能构成一个收敛级数。第2节中求得的 $\lambda_a(c)$ 与 $\lambda_0(c)$ 函数实际上是 $R_{c,\varepsilon} - R_{0,\varepsilon}(\theta)$ 当 $\varepsilon \to 0$ 时的极限，其中 $R_{c,\varepsilon}$ 指的是 $\theta_n(c)$ 序列的风险函数。接下来，可以相对于参数空间上的一个权函数求得平均风险，人们可以选择或不选择以贝叶斯方式将其解释为先验分布，最后相对于所考虑的估计量类最小化该表达式。

这就是例 3A 和 4A 中部分实施的方案。也可以采用一种更直接、更一般的方法，避免将范围限制在某一给定类别的估计量中，具体如下，前提是先验分布完全指定。设 $\pi(.)$ 为先验分布，$\pi_n(.) = \pi_n(.|x_1, \ldots, x_n)$ 为第 n 步的后验分布。那么

$$\int R_\varepsilon(\theta)\pi(\theta)\,\mathrm{d}\theta = \sum_{n \ge a/\varepsilon^2} \int \cdots \int \pi_n\{|\theta - \widehat{\theta}_n| \ge \varepsilon\}\, f_n(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

其中 f_n 是以 $\pi(.)$ 为先验分布所得的边缘密度。对于给定的 ε，最佳解 $\widehat{\ }$ 是将每一项最小化，也就是说，对每个给定的 $x_1, \ldots, x_n$，选择使 $\pi_n\{|\theta - \theta_n| \le \varepsilon\}$ 最大的 θ_n。当 ε 趋于零而 n 固定时，这意味着使用后验众数。然而这里的权衡更为微妙，因为要先对 n 求和，再令 ε 趋于零。$\widehat{\ }$ 可以将 θ_n 视为 $\theta_n^* + u/\sqrt{n}$，其中 θ_n^* 是 θ 的后验均值。因此，需要对 u 最小化 $\pi_n\{|\sqrt{n}(\theta - \theta_n^*) - u| \ge \sqrt{s}\}$，其中 $m = 1/\varepsilon^2$ 和 $s = n/m$ 的定义与前文各节相同。要进行完整分析，就需要再次使用埃奇沃思展开。特别要注意的是，后验分布的偏度在这里起着重要作用。

5B. 正态均值。再次考虑 3A 的情形：其中指出，对任意一阶矩为 θ_0、方差为 τ^2 的先验权重函数，(3.1) 中的 θ_n^* 在样本均值的所有线性函数中最优。现设 θ 被赋予具有这些参数的正态先验。则 $\pi_n(.)$ 亦为正态，其均值等于 (3.1) 中的 θ_n^*，方差为 $\tau_n^2 = \tau^2/(n\tau^2 + 1)$。正态密度的 $\widehat{\ }$ 对称性表明，$\pi_n\{|\theta - \theta_n| \le \varepsilon\}$ 在 $\theta_n = \theta_n^*$ 处达到最大。因此，(3.1) $\widehat{\ }$ 不仅是线性类 $\theta_n(c,d)$ 中的最优解，而且在平均 EQ_ε 准则下，只要所用先验确为 $N\{\theta_0, \tau^2\}$，它就是所有估计量中的最优者。这一点对每一个正的 ε 以及在极限中均成立。

6. 分布意义下的二阶极限。 本文工作的出发点是 Hjort 和 Fenstad（1992）的一阶结果 (1.1)，它将 $\varepsilon^2 Q_\varepsilon$ 的分布与布朗运动 $W(s)$ 在 $|W(s)| \le s/\sigma$ 区域之外所花费的时间联系起来。对于具有相同极限分布的两个估计量序列，典型情形是 $\varepsilon^2(Q_{1,\varepsilon} - Q_{2,\varepsilon}) \to_p 0$与 $Q_{1,\varepsilon}/Q_{2,\varepsilon} \to_p 1$，并且在本文中，我们针对不同情形给出了 $Q_{1,\varepsilon} - Q_{2,\varepsilon}$ 的均值的极限公式。最后一节简要给出这一差值的若干二阶分布方面，从而提供关于当 $\varepsilon \to 0$ 时会发生什么的更多信息。在若干情形下，该分布在极限中与布朗运动沿两条边界线 $\pm s/\sigma$ 所花费的时间有关。关于布朗运动以及一般扩散过程的此类分布，见 Hjort 和 Khasminskii（1993）的研究。

为了说明典型情形，我们回到第2节的情形。令 c 和 ξ为正数，使得在（2.3）中 $l(c) > l(0)$ 且 $r(c) > r(0)$。由（2.4）可得

$$Q_\varepsilon(c) - Q_\varepsilon(0) = \sum_{n \ge a/\varepsilon^2} \Big[I\big\{l(0) \le T_n \le l(c)\big\} - I\big\{r(0) \le T_n \le r(c)\big\} \Big]. \tag{6.1}$$

这是由相当罕见的 1 和 -1 构成的无穷和。我们此前的工作仅限于研究这一差值的期望值，因此仍处于概率与埃奇沃思展开的范畴。要找出（6.1）中差值的极限分布，需要进行另一类细致分析。相关细节见 Hjort 和 Khasminskii（1993，第6.3节），更完整的论述则见 Hjort和Fenstad（1993）的技术报告。归根结底，

$$\varepsilon\{Q_\varepsilon(c) - Q_\varepsilon(0)\} \to_d A - B, \tag{6.2}$$

例如，其中 A 和 B 是两个变量，分别表示布朗运动沿着 s/σ 和 $-s/\sigma$所经历的时间。在（6.1）中，当 a 为固定正数时， A 和 B 各自都是指数分布与零点处单位点质量的混合分布，且二者相依。特别是，此时 $A - B$以正概率取零。若 $a = a$趋于零（ε），例如（6.1）中的 $a = \varepsilon$，则 A 和 B都服从均值为 $c\xi$、互相关为 $-\frac{1}{3}$ 的指数分布。

参考文献

Barndorff-Nielsen, O.E. 和 Cox, D.R.（1989）。统计学中使用的渐近技术。Chapman and Hall，伦敦。Berkson, J.（1980）。最小卡方，而非最大似然！（附讨论）。《统计年鉴》**8**，457–487。Ghosh, J.K. 和 Subramanyam, K.（1974年）。最大似然估计量的二阶效率。*Sankhy¯a, Ser.A* **36**，325–258。Hodges, J. 和 Lehmann, E.L.（1970）。亏量。*Ann. Math.Statist.* **41**，783–801。Hjort, N.L. 和 Fenstad, G.（1992）。关于估计量偏离其目标值超过 ε 时的最后时刻及发生次数。《统计年鉴》**20**，469–489。Hjort, N.L. 和 Fenstad, G.（1993）。估计量偏离其目标值超过 ε时的发生次数的二阶渐近。统计研究报告，奥斯陆大学数学系。Hjort, N.L. 和 Khasminskii, R.Z.（1993）。关于扩散过程沿直线所经历的时间。*Stoch. Proc.and their Appl.* **47**，229–247。Kolassa, J.E. 和 McCullagh, P.（1990）。格点分布的埃奇沃思级数。《统计年鉴》**18**，981–985。

Lehmann, E.L. (1983). 点估计理论。 Wiley，纽约。Pfanzagl, J. (1973). 与最小对比估计量相关的渐近展开。《统计年鉴》**1**, 993–1026. Rao, C.R. (1962). 大样本下的有效估计量和最优推断程序（附讨论）。《皇家统计学会杂志》，*B*辑 **24**, 46–72.