Likelihood-Neighborhood Calibration: Finite Likelihood-Ratio Secants for Parametric Estimation

Ahmed Jebur Ali[a]
Esraa Abdul Ridha[b]
Fadhil AlSabbagh[c]
Publication date:18 /10 /2026
Keywords: estimating functions; likelihood ratio; Godambe information; Fisher information; Kullback-Leibler divergence; generalized method of moments; robustness; parametric inference.

ABSTRACT

A finite-neighborhood estimating framework for regular parametric models is developed from exact change-of-measure identities between neighboring members of the same statistical family. Let f(x; θ) denote a regular one-parameter density and define the within-family likelihood ratio $R_{\theta,h}(x) = \frac{f(x;\theta+h)}{f(x;\theta)}$.

Whenever the two densities have common support, $E_\theta\left[R_{\theta,h}(X)\right] = 1$ holds exactly for every admissible finite perturbation h. The paper uses these identities systematically as unbiased estimating equations without requiring analytic differentiation of the likelihood. The resulting methodology is called Likelihood-Neighborhood Calibration Estimation (LNCE).

The paper develops both one-sided and symmetric likelihood-neighborhood estimators. The ordinary likelihood score is shown to arise as the zero-neighborhood tangent limit of the proposed likelihood-ratio secants, while the second and higher derivatives of the normalization identity recover the classical Bartlett identities. For a finite one-sided neighborhood, the Godambe information is expressed explicitly in terms of a Pearson χ^2 divergence and a cross-score expectation. The exact relative efficiency is the squared correlation between the likelihood-ratio secant and the likelihood score and is therefore bounded above by one.

To eliminate dependence on the numerical scale of the parameter, an intrinsic version is introduced in which the neighboring models are selected at equal Kullback-Leibler distance from the candidate model. The resulting pair of neighboring probability laws is invariant under smooth one-to-one reparameterization. A symmetric KL-neighborhood likelihood secant converges in L_2 to the standardized score with second-order approximation error. Consequently, its asymptotic efficiency loss is of fourth order in the statistical neighborhood radius, whereas a one-sided neighborhood generally has a second-order efficiency loss.

Multiple neighborhood radii yield an overidentified estimating system. The optimally weighted estimator has Godambe information equal to the squared $L_2(P_\theta)$ norm of the orthogonal projection of the likelihood score onto the span of the likelihood-neighborhood secants. Information increases monotonically as additional neighborhoods are added and converges to Fisher information when the system contains neighborhoods shrinking to zero.

A bounded likelihood-neighborhood construction is proposed for robust estimation. It preserves exact unbiasedness after model centering and retains reparameterization invariance under intrinsic neighborhoods. The associated influence function is bounded whenever the transformed likelihood-ratio contrast is bounded.

Two models admit particularly transparent theory. For normal location with known variance, the symmetric KL-neighborhood estimator has an explicit closed form and relative efficiency $\frac{r^2}{\sinh(r^2)}$, where r is the KL-standardized neighborhood radius. For the exponential rate model, a symmetric relative-neighborhood estimator is shown to have a unique root and fourth-order local efficiency loss. A reproducible Monte Carlo study confirms the local efficiency calculations. Under Gaussian contamination, raw finite likelihood-ratio secants are unstable, whereas the bounded KL-neighborhood estimator remains stable and competitive with standard Huber and density-power-divergence benchmarks.

The results provide a finite-neighborhood estimating-equation counterpart to classical score-based inference and organize finite within-family likelihood-ratio calibration within estimating-function, GMM, and robust-inference theory.

[a]*Faculty of Education, University of Kufa, Iraq, E-Mail: ahmedj.alkhuzai@uokufa.edu.iq*
[b]*Faculty of Education, University of Kufa, Iraq, E-Mail: esraaa.alghanim@uokufa.edu.iq*
[c]*Faculty of Education, University of Kufa, Iraq, E-Mail: fadhilt.alsabbagh@uokufa.edu.iq*

1. Introduction

Likelihood theory is fundamentally local. In a regular parametric model, estimation is ordinarily driven by the score

$$s_\theta(x) = \frac{\partial}{\partial\theta}\log f(x;\theta),$$

which records the infinitesimal response of the log density to a perturbation of the parameter. Fisher information,

$$I(\theta) = E_\theta[s_\theta(X)^2],$$

then measures the local quadratic magnitude of this response.

The score is optimal within broad classes of regular unbiased estimating functions. This principle is central to the theory initiated by Godambe and developed further through quasi-likelihood and estimating-function methods. Likewise, generalized method-of-moments theory provides a systematic framework for combining finite or infinite collections of population moment restrictions.

The present work starts from a different question. Instead of asking how a model changes *infinitesimally* at θ, consider how the model calibrates against another actual member of the same parametric family at a finite distance from θ.

Let P_θ have density $f_\theta(x) = f(x;\theta)$, and let P_ϑ be another member of the same family. Whenever $P_\vartheta \ll P_\theta$, the likelihood ratio

$$\frac{f_\vartheta(X)}{f_\theta(X)}$$

satisfies the exact change-of-measure identity

$$E_\theta\left[\frac{f_\vartheta(X)}{f_\theta(X)}\right] = 1.$$

The identity itself is classical. The proposed use of this identity is to select f_ϑ from the same parametric model as a finite perturbation of the candidate parameter and to use its exact normalization identity as an estimating equation for that candidate parameter.

This turns every neighboring model into an unbiased calibration condition. The ordinary score then appears only as the limiting case in which the neighborhood collapses to zero. This gives the hierarchy

finite neighboring probability laws → likelihood-ratio secants → score tangent.

Finite differences are widely used to approximate derivatives in statistical optimization and sensitivity analysis. LNCE serves a different purpose: its likelihood-ratio moments are exactly unbiased for every fixed admissible neighborhood, and the finite neighborhood is part of the estimating procedure rather than a numerical approximation parameter.

Finite likelihood-ratio comparisons themselves are classical. Test-point likelihood ratios underlie the Barankin and Hammersley-Chapman-Robbins variance bounds (Barankin, 1949; Hammersley, 1950; Chapman & Robbins, 1951), and likelihood-ratio change-of-measure arguments are standard in simulation sensitivity analysis (Reiman & Weiss, 1989). The contribution of LNCE is therefore not the normalization identity or the existence of finite likelihood ratios per se. Rather, it is their systematic use as within-family unbiased Z-estimating equations, together with symmetric and equal-KL neighborhood design, optimal GMM aggregation, and centered bounded transformations for robustness.

The main contributions are summarized as follows. First, a general class of exact likelihood-neighborhood estimating functions is formalized. Second, one-sided likelihood-ratio calibration is connected explicitly with Pearson χ^2 divergence, cross-score expectations, and Godambe information. Third, symmetric likelihood-ratio secants are shown to have a higher-order local efficiency advantage over one-sided secants. Fourth, a parameterization-independent construction is obtained by selecting neighbors at equal Kullback-Leibler distance. Fifth, the relative efficiency loss of an intrinsic

symmetric secant is shown to be fourth order in the statistical radius under the stated smoothness conditions. Sixth, multiple neighborhood radii are combined through optimal GMM, yielding a projection interpretation of the information recovered by finite neighborhoods. Seventh, a centered bounded transformation of likelihood-neighborhood contrasts produces estimators with bounded influence. Finally, exact normal-location and exponential-rate examples and reproducible Monte Carlo experiments illustrate the theory.

2. Likelihood-neighborhood calibration

Let $X_1, \dots, X_n$ be independent and identically distributed according to P_{θ_0}, where $\theta_0 \in \Theta \subset \mathbb{R}$. Assume throughout that Θ is an open interval, $f(x;\theta) > 0$ on a common support $\mathcal{X}$, and the required parameter derivatives exist and may be interchanged with integration.
Write

$$f_\theta(x) = f(x;\theta).$$

The likelihood score is

$$s_\theta(x) = \frac{\partial}{\partial\theta} \log f_\theta(x).$$

The Fisher information is

$$I(\theta) = E_\theta[s_\theta(X)^2].$$

Under regularity,

$$E_\theta[s_\theta(X)] = 0.$$

It is assumed that

$$0 < I(\theta) < \infty$$

throughout a neighborhood of the true parameter.

2.1 Exact calibration identity

Definition 1. Within-family likelihood ratio
For an admissible perturbation h, define

$$R_{\theta,h}(x) = \frac{f_{\theta+h}(x)}{f_\theta(x)}.$$

The ratio compares two members of the same parametric family.

Theorem 1. Exact neighborhood calibration identity
For every θ and every admissible h,

$$E_\theta\big[R_{\theta,h}(X)\big] = 1.$$

Proof.
By direct substitution,

$$E_\theta\big[R_{\theta,h}(X)\big] = \int_{\mathcal{X}} \frac{f_{\theta+h}(x)}{f_\theta(x)} f_\theta(x)\, dx.$$

The factors $f_\theta(x)$ cancel, giving

$$E_\theta\big[R_{\theta,h}(X)\big] = \int_{\mathcal{X}} f_{\theta+h}(x)\, dx.$$

Since $f_{\theta+h}$ is a probability density,

$$\int_{\mathcal{X}} f_{\theta+h}(x)\, dx = 1.$$

Therefore,

$$E_\theta\left[R_{\theta,h}(X)\right] = 1.$$

This exact identity is the basis of the proposed methodology.

2.2 Likelihood-neighborhood estimating functions

Define

$$g_h(x;\theta) = R_{\theta,h}(x) - 1.$$

Then Theorem 1 gives

$$E_\theta[g_h(X;\theta)] = 0.$$

Hence, for every admissible h, g_h is an unbiased estimating function. The sample estimating equation is

$$\frac{1}{n}\sum_{i=1}^{n} g_h\,(X_i;\theta) = 0.$$

Equivalently,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i;\theta+h)}{f(X_i;\theta)} = 1.$$

Definition 2. One-sided LNCE estimator

A root of the preceding equation is called a **one-sided Likelihood-Neighborhood Calibration estimator**, denoted $\hat{\theta}_h$. When more than one root exists, a consistent localization rule or preliminary estimator is required.

2.3 The score as the tangent limit

Suppose f_θ is differentiable. Then

$$f_{\theta+h}(x) = f_\theta(x) + h\frac{\partial f_\theta(x)}{\partial\theta} + O(h^2).$$

Since

$$\frac{\partial f_\theta(x)}{\partial\theta} = f_\theta(x)s_\theta(x),$$

we obtain

$$f_{\theta+h}(x) = f_\theta(x)[1 + hs_\theta(x) + O(h^2)].$$

Dividing by $f_\theta(x)$,

$$R_{\theta,h}(x) = 1 + hs_\theta(x) + O(h^2).$$

Therefore,

$$\lim_{h\to 0}\frac{R_{\theta,h}(x)-1}{h} = s_\theta(x).$$

Define

$$\phi_{\theta,h}(x) = \frac{R_{\theta,h}(x)-1}{h}.$$

Then $\phi_{\theta,h}(x) \to s_\theta(x)$ as $h \to 0$. Thus $\phi_{\theta,h}$ is a finite likelihood-ratio secant, whereas s_θ is the likelihood tangent.

2.4 Higher-order normalization identities

Assume four parameter derivatives exist locally, with a dominating integrable envelope sufficient to justify the fourth-order Taylor remainder and termwise integration. Taylor expansion gives

$$f_{\theta+h} = f_\theta + hf_\theta{}' + \frac{h^2}{2} f_\theta{}'' + \frac{h^3}{6} f_\theta{}''' + O(h^4).$$

Dividing by f_θ,

$$R_{\theta,h} = 1 + h\frac{f_\theta{}'}{f_\theta} + \frac{h^2}{2}\frac{f_\theta{}''}{f_\theta} + \frac{h^3}{6}\frac{f_\theta{}'''}{f_\theta} + O(h^4).$$

Now,

$$\frac{f_\theta{}'}{f_\theta} = s_\theta,$$

and

$$\frac{f_\theta{}''}{f_\theta} = s_\theta^2 + \dot{s}_\theta,$$

where

$$\dot{s}_\theta = \frac{\partial s_\theta}{\partial \theta}.$$

Thus,

$$R_{\theta,h} = 1 + hs_\theta + \frac{h^2}{2}\left(s_\theta^2 + \dot{s}_\theta\right) + O(h^3).$$

Because $E_\theta\left[R_{\theta,h}\right] = 1$ for every h, the coefficient of h gives

$$E_\theta[s_\theta(X)] = 0,$$

while the coefficient of h^2 gives

$$E_\theta[s_\theta(X)^2 + \dot{s}_\theta(X)] = 0.$$

Hence,

$$I(\theta) = -E_\theta[\dot{s}_\theta(X)].$$

Higher derivatives of the exact calibration identity similarly reproduce higher Bartlett-type normalization identities.

3. Information and local efficiency

Throughout this section, all estimating functions under consideration are assumed square-integrable at the parameter value of interest, with finite nonzero variability and sufficient domination to justify differentiation under the expectation whenever the sensitivity identity is invoked.

For any scalar unbiased estimating function $\psi(X;\theta)$ with $E_\theta\psi(X;\theta) = 0$, define its sensitivity

$$A_\psi(\theta) = -E_\theta\left[\frac{\partial}{\partial\theta}\psi(X;\theta)\right]$$

and variability

$$B_\psi(\theta) = E_\theta[\psi(X;\theta)^2].$$

The Godambe information is

$$G_\psi(\theta) = \frac{A_\psi(\theta)^2}{B_\psi(\theta)}.$$

Lemma 1. Sensitivity-score identity
If $E_\theta[\psi(X;\theta)] = 0$ for all θ, then

$$A_\psi(\theta) = E_\theta[\psi(X;\theta)s_\theta(X)].$$

Proof.
Differentiate

$$E_\theta[\psi(X;\theta)] = 0.$$

Then

$$0 = E_\theta\left[\frac{\partial\psi}{\partial\theta}\right] + E_\theta[\psi s_\theta].$$

Therefore,

$$-E_\theta\left[\frac{\partial\psi}{\partial\theta}\right] = E_\theta[\psi s_\theta].$$

Consequently,

$$G_\psi(\theta) = \frac{\{E_\theta[\psi s_\theta]\}^2}{E_\theta[\psi^2]}.$$

3.1 Divergence representation of one-sided LNCE
Take $\psi = g_h = R_{\theta,h} - 1$. Its variance is

$$E_\theta\left[\left(R_{\theta,h} - 1\right)^2\right] = E_\theta\left[R_{\theta,h}^2\right] - 1.$$

But

$$E_\theta\left[R_{\theta,h}^2\right] = \int \frac{f_{\theta+h}(x)^2}{f_\theta(x)}\,dx.$$

Therefore,

$$E_\theta[g_h^2] = \chi^2(P_{\theta+h}\|P_\theta),$$

where Pearson's divergence is

$$\chi^2(P\|Q) = \int \frac{(p-q)^2}{q}.$$

Next,

$$E_\theta[g_h s_\theta] = E_\theta\left[R_{\theta,h}s_\theta\right].$$

Because $R_{\theta,h}f_\theta = f_{\theta+h}$,

$$E_\theta[g_h s_\theta] = E_{\theta+h}[s_\theta(X)].$$

Theorem 2. Exact Godambe information of a one-sided likelihood neighborhood
For $g_h = R_{\theta,h} - 1$,

$$G_h(\theta) = \frac{\{E_{\theta+h}[s_\theta(X)]\}^2}{\chi^2(P_{\theta+h}\|P_\theta)}.$$

Thus finite-neighborhood efficiency is determined by the cross-score expectation of the neighboring model and the Pearson divergence between the neighboring and candidate models.

3.2 Fisher-information bound

Theorem 3. Fisher-information bound for likelihood neighborhoods
For every admissible h,

$$G_h(\theta) \le I(\theta).$$

Equality holds if and only if $R_{\theta,h}(X) - 1$ is proportional almost surely to $s_\theta(X)$.

Proof.
Using the sensitivity-score identity,

$$G_h = \frac{E_\theta[g_h s_\theta]^2}{E_\theta[g_h^2]}.$$

By Cauchy-Schwarz,

$$E_\theta[g_h s_\theta]^2 \le E_\theta[g_h^2] E_\theta[s_\theta^2].$$

Therefore,

$$G_h \le E_\theta[s_\theta^2] = I(\theta).$$

Equality occurs exactly under almost-sure linear dependence.
Consequently,

$$\mathrm{ARE}_h = \frac{G_h(\theta)}{I(\theta)} = \mathrm{Corr}_\theta\big(g_h(X;\theta), s_\theta(X)\big)^2.$$

Finite-neighborhood efficiency is therefore exactly the squared alignment between the likelihood-ratio secant and the score tangent in $L_2(P_\theta)$.

3.3 Small-neighborhood efficiency

From the expansion

$$g_h = h s_\theta + O(h^2),$$

we obtain

$$\frac{g_h}{h} = s_\theta + O(h).$$

Under L_2 smoothness,

$$\frac{g_h}{h} \to s_\theta$$

in $L_2(P_\theta)$. Hence,

$$\lim_{h\to 0} G_h(\theta) = I(\theta).$$

A one-sided finite neighborhood is therefore asymptotically efficient as its radius shrinks to zero.

3.4 Symmetric likelihood-ratio secants

For $h > 0$, define

$$R_{\theta,\mathrm{R}}(x) = \frac{f_{\theta+h}(x)}{f_\theta(x)},$$

and

$$R_{\theta,\mathrm{L}}(x) = \frac{f_{\theta-h}(x)}{f_\theta(x)}.$$

Define the odd likelihood secant

$$\psi_{\theta,h}^{\mathrm{odd}}(x) = \frac{R_{\theta,\mathrm{R}}(x) - R_{\theta,\mathrm{L}}(x)}{2h}.$$

Because both ratios integrate to one,

$$E_\theta\big[\psi_{\theta,h}^{\mathrm{odd}}(X)\big] = 0.$$

Similarly, define the even curvature contrast

$$\psi_{\theta,h}^{\mathrm{even}}(x) = \frac{R_{\theta,\mathrm{R}}(x) + R_{\theta,\mathrm{L}}(x) - 2}{h^2}.$$

Again,

$$E_\theta\left[\psi_{\theta,h}^{\text{even}}(X)\right] = 0.$$

3.5 Tangent and curvature limits

Under the same fourth-order local smoothness and domination conditions used above, the two one-sided expansions may be subtracted term by term.

Taylor expansion yields

$$R_{\theta,\mathrm{R}} = 1 + hs_\theta + \frac{h^2}{2}\frac{f_\theta''}{f_\theta} + \frac{h^3}{6}\frac{f_\theta'''}{f_\theta} + O(h^4),$$

while

$$R_{\theta,\mathrm{L}} = 1 - hs_\theta + \frac{h^2}{2}\frac{f_\theta''}{f_\theta} - \frac{h^3}{6}\frac{f_\theta'''}{f_\theta} + O(h^4).$$

Subtracting,

$$R_{\theta,\mathrm{R}} - R_{\theta,\mathrm{L}} = 2hs_\theta + \frac{h^3}{3}\frac{f_\theta'''}{f_\theta} + O(h^4).$$

Hence,

$$\psi_{\theta,h}^{\text{odd}} = s_\theta + \frac{h^2}{6}\frac{f_\theta'''}{f_\theta} + O(h^3).$$

Therefore,

$$\psi_{\theta,h}^{\text{odd}} = s_\theta + O(h^2),$$

whereas the one-sided normalized secant has only $s_\theta + O(h)$. Similarly,

$$\psi_{\theta,h}^{\text{even}} = s_\theta^2 + \dot{s}_\theta + O(h^2).$$

Thus the odd and even finite neighborhoods approximate the score and likelihood-curvature identity, respectively.

3.6 Fourth-order local efficiency

Let

$$t_\theta = \frac{s_\theta}{\sqrt{I(\theta)}},$$

so that

$$E_\theta[t_\theta^2] = 1.$$

Suppose an unbiased normalized symmetric secant has the L_2 expansion

$$\psi_\varepsilon = t_\theta + \varepsilon^2 b + o_{L_2}(\varepsilon^2).$$

Decompose

$$b = at_\theta + b_\perp,$$

where

$$E_\theta[t_\theta b_\perp] = 0.$$

Theorem 4. Fourth-order local efficiency loss

Under the preceding expansion,

$$\frac{G_{\psi_\varepsilon}}{I(\theta)} = 1 - \varepsilon^4 E_\theta[b_\perp(X)^2] + o(\varepsilon^4).$$

Proof.
Relative Godambe information is the squared correlation with the standardized score tθ. Let cε = Eθ[ψε tθ] and define qε = ψε - cε tθ. Then Eθ[qε tθ] = 0, so the Pythagorean decomposition in L²(Pθ) gives G(ψε)/I(θ) = cε²/{cε² + Eθ(qε²)}.
The assumed expansion implies, in L², cε = 1 + aε² + o(ε²) and qε = ε²b⊥ + rε, where the L² norm of rε is o(ε²). Hence Eθ(qε²) = ε⁴Eθ(b⊥²) + o(ε⁴), whereas cε² = 1 + O(ε²). Expanding the ratio proves the stated fourth-order efficiency loss.
For comparison, if a one-sided secant has expansion $t_\theta + \varepsilon b + o(\varepsilon)$, its efficiency loss is generally $O(\varepsilon^2)$. Symmetry removes the leading efficiency loss.

4. Intrinsic KL neighborhoods and asymptotic theory

A fixed additive perturbation h depends on the numerical scale in which θ is expressed. To construct statistical rather than coordinate-based neighborhoods, let

$$D_{\mathrm{KL}}(P_\theta \| P_\vartheta) = E_\theta\left[\log \frac{f_\theta(X)}{f_\vartheta(X)}\right].$$

For $\varepsilon > 0$, define the two neighboring parameter values

$$\theta_{\varepsilon,\mathrm{L}} < \theta < \theta_{\varepsilon,\mathrm{R}}$$

as solutions of

$$D_{\mathrm{KL}}\left(P_\theta \| P_{\theta_{\varepsilon,\mathrm{R}}}\right) = \frac{\varepsilon^2}{2},$$

and

$$D_{\mathrm{KL}}\left(P_\theta \| P_{\theta_{\varepsilon,\mathrm{L}}}\right) = \frac{\varepsilon^2}{2}.$$

The local relation between KL divergence and Fisher information gives

$$D_{\mathrm{KL}}(P_\theta \| P_{\theta+\delta}) = \frac{1}{2} I(\theta)\delta^2 + O(\delta^3).$$

Hence,

$$\theta_{\varepsilon,\mathrm{R}} = \theta + \frac{\varepsilon}{\sqrt{I(\theta)}} + O(\varepsilon^2),$$

and

$$\theta_{\varepsilon,\mathrm{L}} = \theta - \frac{\varepsilon}{\sqrt{I(\theta)}} + O(\varepsilon^2).$$

Thus ε is a statistical radius measured in local Fisher-information units. This local relation is the usual Fisher-Rao quadratic geometry induced by Kullback-Leibler divergence (Amari & Nagaoka, 2000).

4.1 Existence of intrinsic neighbors

Proposition 1
Suppose $I(\theta) > 0$ and $\vartheta \mapsto D_{\mathrm{KL}}(P_\theta \| P_\vartheta)$ is three times continuously differentiable near $\vartheta = \theta$. Then, for sufficiently small $\varepsilon > 0$, the two equal-KL equations possess unique right and left solutions.
Proof.
At $\vartheta = \theta$, the divergence equals zero. Its first derivative is zero because the score has expectation zero. Its second derivative is $I(\theta) > 0$. Therefore,

$$D_{\mathrm{KL}}(P_\theta \| P_{\theta+\delta}) = \frac{1}{2} I(\theta)\delta^2 + o(\delta^2),$$

so θ is a strict local minimum. Consequently, the divergence is strictly decreasing as ϑ approaches θ from the left and strictly increasing as ϑ moves away from θ to the right. Equivalently, it increases

locally with |ϑ-θ| on either branch. The intermediate value theorem and branchwise strict monotonicity yield existence and uniqueness of the left and right equal-KL solutions.

4.2 Reparameterization invariance

Theorem 5. Coordinate invariance of KL neighborhoods
Let

$$\eta = h(\theta)$$

be any smooth one-to-one reparameterization. The pair of probability distributions

$$\{P_{\theta_{\varepsilon,\mathrm{L}}}, P_{\theta_{\varepsilon,\mathrm{R}}}\}$$

defined by the equal-KL equations is unchanged by reparameterization. If h is increasing, the right and left labels are preserved; if h is decreasing, the labels are interchanged.
Proof.
Kullback-Leibler divergence depends only on the probability measures and not on their coordinate labels. Therefore the set of distributions satisfying

$$D_{\mathrm{KL}}(P_\theta \| P_\vartheta) = \frac{\varepsilon^2}{2}$$

is unchanged when θ and ϑ are replaced by their transformed coordinates. Monotonicity of h determines only which solution is called left or right.
Consequently, the estimating equation based on their odd difference is invariant up to multiplication by -1, which leaves its roots unchanged.

4.3 Intrinsic symmetric likelihood secant

Define

$$R_{\theta,\varepsilon,\mathrm{R}}(x) = \frac{f(x;\theta_{\varepsilon,\mathrm{R}})}{f(x;\theta)},$$

and

$$R_{\theta,\varepsilon,\mathrm{L}}(x) = \frac{f(x;\theta_{\varepsilon,\mathrm{L}})}{f(x;\theta)}.$$

Define the standardized intrinsic likelihood secant

$$\Psi_{\theta,\varepsilon}(x) = \frac{R_{\theta,\varepsilon,\mathrm{R}}(x) - R_{\theta,\varepsilon,\mathrm{L}}(x)}{2\varepsilon}.$$

Since both ratios have expectation one,

$$E_\theta[\Psi_{\theta,\varepsilon}(X)] = 0.$$

4.4 Standardized-score limit

Theorem 6. Intrinsic secant-to-score limit
Under fourth-order L^2 regularity of the likelihood-ratio and equal-KL expansions,

$$\Psi_{\theta,\varepsilon} = \frac{s_\theta}{\sqrt{I(\theta)}} + O_{L_2}(\varepsilon^2).$$

Proof.
Write

$$\delta_{\mathrm{R}} = \theta_{\varepsilon,\mathrm{R}} - \theta, \qquad \delta_{\mathrm{L}} = \theta_{\varepsilon,\mathrm{L}} - \theta.$$

Expansion of the equal-KL equations gives

$$\delta_{\mathrm{R}} = a\varepsilon + b\varepsilon^2 + O(\varepsilon^3),$$

and

$$\delta_{\mathrm{L}} = -a\varepsilon + b\varepsilon^2 + O(\varepsilon^3),$$

where

$$a = I(\theta)^{-1/2}.$$

The likelihood-ratio expansion is

$$R(\delta) = 1 + \delta s_\theta + \frac{\delta^2}{2}\frac{f_\theta''}{f_\theta} + \frac{\delta^3}{6}\frac{f_\theta'''}{f_\theta} + O_{L_2}(\delta^4).$$

Because the two KL solutions have equal first-order magnitudes and equal second-order midpoint displacement, all terms of order ε^2 cancel in $R(\delta_{\mathrm{R}}) - R(\delta_{\mathrm{L}})$. Therefore,

$$R(\delta_{\mathrm{R}}) - R(\delta_{\mathrm{L}}) = 2a\varepsilon s_\theta + O_{L_2}(\varepsilon^3).$$

Division by 2ε yields

$$\Psi_{\theta,\varepsilon} = a s_\theta + O_{L_2}(\varepsilon^2).$$

Since $a = I(\theta)^{-1/2}$, the result follows.
Combining Theorems 4 and 6 gives
1 - Gε(θ)/I(θ) = O(ε⁴).

4.5 Consistency and asymptotic normality

Let $\psi_\varepsilon(x;\theta)$ denote either a fixed admissible neighborhood moment or the intrinsic symmetric moment. Define
The following are standard one-dimensional Z-estimation conditions; see, for example, van der Vaart (1998) and Newey and McFadden (1994).

$$m_\varepsilon(\theta;\theta_0) = E_{\theta_0}[\psi_\varepsilon(X;\theta)].$$

Assume:
A1. Θ is compact or the estimator is localized to a compact neighborhood.
A2. $m_\varepsilon(\theta;\theta_0) = 0$ has the unique solution $\theta = \theta_0$.
A3. $\psi_\varepsilon(x;\theta)$ is continuous in θ almost surely.
A4. On the localization set, |ψε(x; θ)| is dominated by a Pθ₀-integrable envelope, so the class obeys a uniform law of large numbers.
A5. The map θ ↦ ψε(x; θ) is differentiable on a neighborhood N of θ₀; ∂θψε(x; θ) is almost surely continuous in θ, and supθ∈N |∂θψε(X; θ)| has a Pθ₀-integrable envelope. Define

$$A_\varepsilon = -E_{\theta_0}[\partial_\theta \psi_\varepsilon(X;\theta_0)]$$

and assume Aε is finite and nonzero. The derivative class then obeys a uniform law of large numbers on N.
A6. The variance at the truth is finite:

$$B_\varepsilon = E_{\theta_0}[\psi_\varepsilon(X;\theta_0)^2] < \infty.$$

Theorem 7. Consistency

Any localized sequence of exact roots, or approximate roots whose sample moment converges to zero in probability, of

$$\frac{1}{n}\sum_{i=1}^{n} \psi_\varepsilon\,(X_i;\theta) = 0$$

satisfies

$$\hat{\theta}_\varepsilon \xrightarrow{p} \theta_0.$$

Proof.
Uniform convergence gives

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \psi_\varepsilon\,(X_i;\cdot) - m_\varepsilon(\cdot;\theta_0) \right\|_\infty \xrightarrow{p} 0.$$

Identification, localization, and continuity then imply convergence of any such root sequence to the unique population zero θ₀.

Theorem 8. Asymptotic normality
Under A1–A6,

$$\sqrt{n}\big(\hat{\theta}_\varepsilon - \theta_0\big) \xrightarrow{d} N\left(0, \frac{B_\varepsilon}{A_\varepsilon^2}\right).$$

Equivalently,

$$\mathrm{Avar}\big(\hat{\theta}_\varepsilon\big) = \frac{1}{nG_\varepsilon}.$$

Proof.
By Theorem 7, θε is consistent. The mean-value expansion of the sample estimating equation around θ₀, together with the uniform law of large numbers for ∂θψε, gives

$$0 = \frac{1}{n}\sum_{i=1}^{n} \psi_\varepsilon\,(X_i;\theta_0) - A_\varepsilon\big(\hat{\theta}_\varepsilon - \theta_0\big) + o_p\big(n^{-1/2}\big).$$

Therefore,

$$\sqrt{n}\big(\hat{\theta}_\varepsilon - \theta_0\big) = A_\varepsilon^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_\varepsilon\,(X_i;\theta_0) + o_p(1).$$

The central limit theorem completes the proof.

4.6 Shrinking neighborhoods and asymptotic efficiency

Let

$$\varepsilon_n \to 0.$$

Normalize the moment so that

$$\Psi_{\theta,\varepsilon_n} \to I(\theta)^{-1/2} s_\theta$$

in L²(Pθ). Assume, uniformly for θ near θ₀, stochastic differentiability of the triangular-array estimating equations, convergence of their sensitivities, local identification, and a Lindeberg condition for the normalized moments. Then the standard triangular-array Z-estimator argument yields

$$\sqrt{n}\big(\hat{\theta}_{\varepsilon_n} - \theta_0\big) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

Thus

$$\varepsilon_n \to 0 \quad \Rightarrow \quad \text{first-order Fisher efficiency.}$$

Unlike an ordinary numerical finite-difference approximation, no finite-ε population bias is introduced: every neighborhood equation remains exactly unbiased at the true parameter.

5. Multiple neighborhoods and robustification

Choose

$$0 < \varepsilon_1 < \cdots < \varepsilon_m.$$

Define the moment vector

$$\mathbf{\Psi}(x;\theta) = \begin{pmatrix} \Psi_{\theta,\varepsilon_1}(x) \\ \vdots \\ \Psi_{\theta,\varepsilon_m}(x) \end{pmatrix}.$$

Then

$$E_\theta[\mathbf{\Psi}(X;\theta)] = \mathbf{0}.$$

Define

$$\overline{\mathbf{\Psi}}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Psi}(X_i;\theta).$$

A GMM estimator is

$$\hat{\theta}_m = \text{argmin}_\theta \overline{\mathbf{\Psi}}_n(\theta)^\top W_n \overline{\mathbf{\Psi}}_n(\theta).$$

Define

$$S = E_{\theta_0}[\mathbf{\Psi}\mathbf{\Psi}^\top],$$

and

$$c = E_{\theta_0}[\mathbf{\Psi} s_{\theta_0}].$$

Assume S is positive definite. If the chosen moments are linearly dependent, redundant directions should be removed; equivalently, the projection statements below can be formulated with the Moore-Penrose generalized inverse.

The optimal weight is

$$W_{\text{opt}} = S^{-1}.$$

Its Godambe information is

$$G_m = c^\top S^{-1} c.$$

5.1 Projection interpretation

Let

$$\mathcal{H}_m = \text{span}\{\Psi_{\varepsilon_1}, \dots, \Psi_{\varepsilon_m}\} \subset L_2(P_{\theta_0}).$$

Let $\Pi_m s$ denote the orthogonal projection of the likelihood score onto $\mathcal{H}_m$.

Theorem 9. Secant projection information

$$\Pi_m s_{\theta_0} = c^\top S^{-1} \mathbf{\Psi}.$$

Moreover,

$$G_m = E_{\theta_0}\left[\left(\Pi_m s_{\theta_0}\right)^2\right].$$

Therefore,

$$G_m \le I(\theta_0).$$

Proof.

The least-squares coefficient vector for projecting s onto $\mathbf{\Psi}$ is

$$S^{-1}E[\mathbf{\Psi} s] = S^{-1}c.$$

Hence,

$$\Pi_m s = c^\top S^{-1} \mathbf{\Psi}.$$

Its squared norm is

$$c^\top S^{-1} E[\mathbf{\Psi}\mathbf{\Psi}^\top] S^{-1} c = c^\top S^{-1} c.$$

Because orthogonal projection cannot increase L_2 norm,

$$E[(\Pi_m s)^2] \le E[s^2].$$

5.2 Monotonicity with additional neighborhoods

If

$$\mathcal{H}_m \subset \mathcal{H}_{m+1},$$

then

$$G_m \le G_{m+1} \le I.$$

Thus adding a genuinely new likelihood-neighborhood direction cannot decrease optimally weighted information. More precisely, consider nested spaces Hm and suppose that for some index j(m) ≤ m the corresponding radius εj(m) tends to zero. Theorem 6 then implies that Hm contains a sequence of normalized secants converging in L^2 to the standardized score, and therefore

$$G_m \to I.$$

Hence a collection of finite likelihood secants can recover full likelihood information.

5.3 Likelihood-neighborhood efficiency spectrum

For a scalar intrinsic symmetric neighborhood, define

$$\mathcal{E}_\theta(\varepsilon) = \frac{G_{\theta,\varepsilon}}{I(\theta)}.$$

Then

$$0 \le \mathcal{E}_\theta(\varepsilon) \le 1,$$

and

Eθ(0) := 1 (continuous extension).

The function $\varepsilon \mapsto \mathcal{E}_\theta(\varepsilon)$ is called the **likelihood-neighborhood efficiency spectrum**. It quantifies how rapidly finite changes in the model cease to align with the local likelihood tangent. For intrinsic symmetric neighborhoods,

1 - Eθ(ε) = O(ε⁴).

5.4 Robust likelihood-neighborhood calibration

Robust M-estimation and divergence-based robust estimation provide natural benchmarks for this construction (Huber, 1964; Basu et al., 1998). LNCE robustification differs in that it bounds a finite within-family log-likelihood contrast and then recenters that contrast under the candidate model.

Raw likelihood ratios can be unbounded, causing extreme observations to dominate finite-neighborhood moments. Let

$$b_c: \mathbb{R} \to \mathbb{R}$$

be bounded and continuously differentiable. A convenient choice is

$$b_c(u) = c\tanh(u/c).$$

For an intrinsic pair, define log likelihood ratios

$$\ell_{\theta,\varepsilon,\mathrm{R}}(x) = \log R_{\theta,\varepsilon,\mathrm{R}}(x), \qquad \ell_{\theta,\varepsilon,\mathrm{L}}(x) = \log R_{\theta,\varepsilon,\mathrm{L}}(x).$$

Define the raw bounded contrast

$$H_{\theta,\varepsilon,c}(x) = b_c\left(\ell_{\theta,\varepsilon,\mathrm{R}}(x)\right) - b_c\left(\ell_{\theta,\varepsilon,\mathrm{L}}(x)\right).$$

Its exact model mean is

$$m_{\theta,\varepsilon,c} = E_\theta\left[H_{\theta,\varepsilon,c}(X)\right].$$

Definition 3. Robust likelihood-neighborhood moment

$$\psi^R_{\theta,\varepsilon,c}(x) = H_{\theta,\varepsilon,c}(x) - m_{\theta,\varepsilon,c}.$$

By construction,

$$E_\theta\left[\psi^R_{\theta,\varepsilon,c}(X)\right] = 0.$$

Thus bounded transformation does not sacrifice exact model unbiasedness when the centering expectation is evaluated exactly. In a generic model, mθ,ε,c may require numerical quadrature or Monte Carlo integration; such numerical error is computational rather than a population bias of the defining moment and should be controlled in implementation.

5.5 Invariance of robust intrinsic LNCE

Because KL-neighboring probability laws are invariant to reparameterization and likelihood ratios are Radon-Nikodym derivatives between those same laws, $R_{\theta,\varepsilon,\mathrm{R}}$ and $R_{\theta,\varepsilon,\mathrm{L}}$ are coordinate-independent. Hence $\psi^R_{\theta,\varepsilon,c}$ is unchanged under an increasing reparameterization and changes only sign under an orientation-reversing one. Its estimating root is therefore parameterization equivariant.

5.6 Bounded influence

Let

$$A_R = -E_{\theta_0}\left[\partial_\theta \psi^R_{\theta_0,\varepsilon,c}(X)\right].$$

Assume $A_R \neq 0$. The influence function is

$$\mathrm{IF}(x) = A_R^{-1}\psi^R_{\theta_0,\varepsilon,c}(x).$$

Theorem 10. Bounded influence

If |bc(u)| ≤ c for all u and AR ≠ 0, then the influence function is uniformly bounded.
|IF(x)| ≤ 4c/|AR| < ∞ for all x.

Proof.

Each transformed log-ratio lies in [-c,c], so Hθ,ε,c lies in [-2c,2c]. Its model mean also lies in [-2c,2c], hence the centered moment has absolute value at most 4c. Multiplication by $|AR|^{-1}$ gives the displayed uniform bound.

6. Exact model examples

Consider

$$X \sim N(\mu, \sigma^2),$$

where σ is known. Define

$$Z = \frac{X-\mu}{\sigma}.$$

Then $Z \sim N(0,1)$. The score is

$$s_\mu(X) = \frac{X-\mu}{\sigma^2} = \frac{Z}{\sigma}.$$

Fisher information is

$$I(\mu) = \frac{1}{\sigma^2}.$$

6.1 Normal location: exact likelihood ratio

Take

$$h = r\sigma.$$

Then

$$R_{\mu,\mathrm{R}}(x) = \frac{f(x;\mu+h)}{f(x;\mu)} = \exp\left[\frac{h(x-\mu)}{\sigma^2} - \frac{h^2}{2\sigma^2}\right].$$

Since $h = r\sigma$,

$$R_{\mathrm{R}}(Z) = \exp\left(rZ - \frac{r^2}{2}\right).$$

Likewise,

$$R_{\mathrm{L}}(Z) = \exp\left(-rZ - \frac{r^2}{2}\right).$$

For equal-variance normal distributions,

$$D_{\mathrm{KL}}[N(\mu, \sigma^2) \| N(\mu + h, \sigma^2)] = \frac{h^2}{2\sigma^2}.$$

Hence r is exactly the intrinsic KL radius ε.

6.2 One-sided normal estimator

The calibration equation is

$$\frac{1}{n}\sum_{i=1}^{n} \exp\left[\frac{h(X_i - \mu)}{\sigma^2} - \frac{h^2}{2\sigma^2}\right] = 1.$$

Rearranging and taking logarithms yields

$$\hat{\mu}_{\mathrm{R},h} = \frac{\sigma^2}{h} \log\left[\frac{1}{n}\sum_{i=1}^{n} e^{hX_i/\sigma^2}\right] - \frac{h}{2}.$$

As $h \to 0$,

$$\hat{\mu}_{\mathrm{R},h} \to \overline{X}.$$

6.3 Exact one-sided normal efficiency

Let

$$r = \frac{h}{\sigma}.$$

The asymptotic variance is

$$\mathrm{Avar}\left(\hat{\mu}_{\mathrm{R},h}\right) = \frac{\sigma^2}{n} \frac{e^{r^2} - 1}{r^2}.$$

Therefore,

$$\mathrm{ARE}_{\mathrm{R}}(r) = \frac{r^2}{e^{r^2} - 1}.$$

Its expansion is

$$\mathrm{ARE}_{\mathrm{R}}(r) = 1 - \frac{r^2}{2} + O(r^4).$$

Thus the one-sided efficiency loss is second order.

6.4 Symmetric normal estimator

The odd calibration equation is

$$\sum_{i=1}^{n} [R_{\mathrm{R}}(X_i) - R_{\mathrm{L}}(X_i)] = 0.$$

Using the exact normal ratios, one obtains

$$\hat{\mu}_{\mathrm{sym},r} = \frac{\sigma}{2r} \log \left[\frac{\sum_{i=1}^{n} e^{rX_i/\sigma}}{\sum_{i=1}^{n} e^{-rX_i/\sigma}} \right].$$

As $r \to 0$,

$$\hat{\mu}_{\mathrm{sym},r} \to \overline{X}.$$

6.5 Exact symmetric normal efficiency

Use the normalized secant

$$\psi_r(Z) = \frac{R_{\mathrm{R}}(Z) - R_{\mathrm{L}}(Z)}{2r}.$$

Then

$$\psi_r(Z) = e^{-r^2/2} \frac{\sinh(rZ)}{r}.$$

Now,

$$E[Z\psi_r(Z)] = 1,$$

and

$$E[\psi_r(Z)^2] = \frac{\sinh(r^2)}{r^2}.$$

Therefore,

$$\mathrm{Avar}\left(\hat{\mu}_{\mathrm{sym},r}\right) = \frac{\sigma^2}{n} \frac{\sinh(r^2)}{r^2}.$$

The relative efficiency is

$$\mathrm{ARE}_{\mathrm{sym}}(r) = \frac{r^2}{\sinh(r^2)}.$$

Since

$$\sinh(r^2) = r^2 + \frac{r^6}{6} + O(r^{10}),$$

we obtain

$$\mathrm{ARE}_{\mathrm{sym}}(r) = 1 - \frac{r^4}{6} + O(r^8).$$

This exact normal result illustrates the fourth-order efficiency theorem.

6.6 One-sided versus symmetric efficiency

Table 1. Exact normal-location relative efficiencies for one-sided and symmetric LNCE.

r	One-sided ARE	Symmetric ARE
0.25	0.96908	0.99935
0.50	0.88020	0.98966
0.75	0.74498	0.94915
1.00	0.58198	0.85092

The advantage of symmetric neighborhood calibration becomes substantial even for moderate radii.

6.7 Robust normal KL-neighborhood estimator

For the normal model,

$$\log R_{\mathrm{R}}(Z) = rZ - \frac{r^2}{2},$$

and

$$\log R_{\mathrm{L}}(Z) = -rZ - \frac{r^2}{2}.$$

Using $b_c(u) = c\tanh(u/c)$, define

$$\psi_{r,c}(Z) = \frac{c}{2r}\left[\tanh\left(\frac{rZ - r^2/2}{c}\right) - \tanh\left(\frac{-rZ - r^2/2}{c}\right)\right].$$

By symmetry of Z,

$$E\left[\psi_{r,c}(Z)\right] = 0.$$

Moreover,

$$\frac{\partial \psi_{r,c}}{\partial z} = \frac{1}{2}\left[\operatorname{sech}^2\left(\frac{rz - r^2/2}{c}\right) + \operatorname{sech}^2\left(\frac{-rz - r^2/2}{c}\right)\right] > 0.$$

As a function of μ, every contribution to the location estimating equation is strictly decreasing because $\partial\psi r,c/\partial z > 0$ and $z = (x-\mu)/\sigma$. Moreover, the sample estimating function tends to c/r as $\mu \to -\infty$ and to -c/r as $\mu \to \infty$. Continuity therefore gives existence, and strict monotonicity gives uniqueness of the root.

Also $|\psi r,c(z)| \leq c/r$ for every z, so the estimating function is bounded and the corresponding influence function is bounded whenever its sensitivity is nonzero.

For

$$r = 0.75, \qquad c = 1,$$

numerical Gaussian integration gives

$$\mathrm{ARE} = 0.96557.$$

The standardized gross-error sensitivity is approximately

$$1.950.$$

Thus the bounded estimator retains approximately 96.6% first-order normal efficiency while having bounded influence.

6.8 Exponential rate model

Consider

$$X \sim \mathrm{Exp}(\lambda),$$

with density

$$f(x;\lambda) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

The score is

$$s_\lambda(x) = \frac{1}{\lambda} - x.$$

The Fisher information is

$$I(\lambda) = \frac{1}{\lambda^2}.$$

6.9 One-sided exponential calibration

For a fixed additive neighborhood $h > 0$,

$$R_{\lambda,h}(x) = \frac{(\lambda + h)e^{-(\lambda+h)x}}{\lambda e^{-\lambda x}}.$$

Thus,

$$R_{\lambda,h}(x) = \frac{\lambda + h}{\lambda} e^{-hx}.$$

The calibration equation is

$$\frac{\lambda + h}{\lambda}\left[\frac{1}{n}\sum_{i=1}^{n} e^{-hX_i}\right] = 1.$$

Let

$$\overline{Y}_h = \frac{1}{n}\sum_{i=1}^{n} e^{-hX_i}.$$

Then

$$(\lambda + h)\overline{Y}_h = \lambda,$$

so

$$\hat{\lambda}_h = \frac{h\overline{Y}_h}{1 - \overline{Y}_h}.$$

As $h \to 0$,

$$e^{-hX_i} = 1 - hX_i + O(h^2),$$

and therefore

$$\hat{\lambda}_h \to \frac{1}{\overline{X}},$$

the exponential MLE.

6.10 Exact one-sided exponential efficiency

Write

$$r = \frac{h}{\lambda}.$$

Then

$$E[e^{-hX}] = \frac{1}{1 + r},$$

and

$$E[e^{-2hX}] = \frac{1}{1 + 2r}.$$

Delta-method calculation gives

$$\mathrm{Avar}\left(\hat{\lambda}_h\right) = \frac{\lambda^2}{n}\frac{(1 + r)^2}{1 + 2r}.$$

Therefore,

$$\mathrm{ARE}_{\mathrm{R}}^{\mathrm{Exp}}(r) = \frac{1 + 2r}{(1 + r)^2}.$$

Again, ARE $\to 1$ as $r \to 0$.

6.11 Symmetric relative neighborhoods for the exponential rate

For a scale-equivariant construction, take

$$\lambda_{\mathrm{R}} = \lambda(1 + r),$$

and

$$\lambda_{\mathrm{L}} = \lambda(1 - r),$$

where

$$0 < r < 1.$$

The two likelihood ratios are

$$R_{\mathrm{R}}(x) = (1 + r)e^{-r\lambda x},$$

and

$$R_{\mathrm{L}}(x) = (1 - r)e^{r\lambda x}.$$

Define

$$\psi_r(x; \lambda) = \frac{R_{\mathrm{R}}(x) - R_{\mathrm{L}}(x)}{2r}.$$

Then

$$E_\lambda[\psi_r(X; \lambda)] = 0.$$

6.12 Uniqueness of the exponential symmetric root

Proposition 2

For every sample containing at least one strictly positive observation, equation

$$\sum_{i=1}^{n} \psi_r\,(X_i; \lambda) = 0$$

has exactly one solution $\lambda > 0$.

Proof.

Differentiate one contribution:

$$\frac{\partial}{\partial \lambda}\psi_r(x; \lambda) = -\frac{x}{2}\left[(1 + r)e^{-r\lambda x} + (1 - r)e^{r\lambda x}\right].$$

For $x > 0$ this is strictly negative, so the sample estimating function is strictly decreasing in λ.

At $\lambda \downarrow 0$,

$$R_{\mathrm{R}} \to 1 + r, \qquad R_{\mathrm{L}} \to 1 - r,$$

so

$$\psi_r \to 1.$$

Therefore the sample function is positive near zero. As $\lambda \to \infty$, the term $-(1 - r)e^{r\lambda x}$ dominates for every positive observation, so the sample function tends to $-\infty$. Continuity and strict monotonicity imply exactly one root.

6.13 Exact exponential symmetric efficiency

Assume

$$0 < r < \frac{1}{2},$$

so the second moments exist. Direct calculation gives

$$E[\psi_r^2] = \frac{1 - 2r^2}{1 - 4r^2}.$$

Also,

$$-E\left[\frac{\partial \psi_r}{\partial \lambda}\right] = \frac{1}{\lambda(1-r^2)}.$$

Consequently,

$$\mathrm{Avar}\left(\hat{\lambda}_r\right) = \frac{\lambda^2}{n}\frac{(1-r^2)^2(1-2r^2)}{1-4r^2}.$$

Therefore,

$$\mathrm{ARE}_{\mathrm{sym}}^{\mathrm{Exp}}(r) = \frac{1-4r^2}{(1-r^2)^2(1-2r^2)}.$$

Expansion near zero gives

$$\frac{n\mathrm{Avar}\left(\hat{\lambda}_r\right)}{\lambda^2} = 1 + 5r^4 + 18r^6 + O(r^8).$$

Again, the leading loss under a symmetric neighborhood is fourth order.

7. Monte Carlo study

A Monte Carlo experiment was performed to examine finite-sample behavior of likelihood-neighborhood estimators. The target model was

$$X_i \sim N(0,1).$$

The following estimators were compared: the sample mean, the sample median, the one-sided normal LNCE estimator with r = 0.5, the symmetric normal LNCE estimator with r = 0.5, the bounded intrinsic KL-LNCE estimator with r = 0.75 and c = 1, the Huber location M-estimator with clipping constant k = 1.345, and the normal density-power-divergence (DPD) estimator with α = 0.2.

For known σ = 1, the Huber estimator solves $\Sigma\, \psi_k(X_i-\mu) = 0$ with $\psi_k(z) = \max(-k, \min(z,k))$. The DPD location estimator solves $\Sigma\, (X_i-\mu)\exp\{-\alpha(X_i-\mu)^2/2\} = 0$. These two procedures provide established robustness benchmarks (Huber, 1964; Basu et al., 1998).

For each configuration, 5000 Monte Carlo samples were generated. Sample sizes were

$$n = 50,\ 100,\ 250,\ 500.$$

For every estimator, empirical bias, root mean squared error, and scaled variance $n\widehat{\mathrm{Var}}(\hat{\mu})$ were calculated.

7.1 Simulation under the normal model

Table 2. Finite-sample performance under $N(0,1)$

n	Estimator	Bias	RMSE	n Var
50	Mean	-0.00103	0.14164	1.00298
50	Median	-0.00097	0.17360	1.50684
50	One-sided LNCE (r=0.5)	-0.00637	0.14991	1.12165
50	Symmetric LNCE (r=0.5)	-0.00108	0.14226	1.01183
50	Robust KL-LNCE	-0.00099	0.14409	1.03809
50	Huber M (k=1.345)	-0.00086	0.14523	1.05458
50	DPD (α=0.2)	-0.00094	0.14457	1.04492
100	Mean	0.00217	0.09880	0.97571
100	Median	0.00339	0.12482	1.55674
100	One-sided LNCE (r=0.5)	-0.00154	0.10626	1.12887
100	Symmetric LNCE (r=0.5)	0.00198	0.09927	0.98499

n	Estimator	Bias	RMSE	n Var
100	Robust KL-LNCE	0.00252	0.10068	1.01297
100	Huber M (k=1.345)	0.00250	0.10144	1.02847
100	DPD (α=0.2)	0.00257	0.10101	1.01963
250	Mean	-0.00055	0.06365	1.01278
250	Median	-0.00155	0.07966	1.58565
250	One-sided LNCE (r=0.5)	-0.00158	0.06799	1.15491
250	Symmetric LNCE (r=0.5)	-0.00060	0.06393	1.02162
250	Robust KL-LNCE	-0.00058	0.06499	1.05579
250	Huber M (k=1.345)	-0.00052	0.06557	1.07484
250	DPD (α=0.2)	-0.00053	0.06522	1.06333
500	Mean	0.00095	0.04513	1.01795
500	Median	0.00003	0.05638	1.58938
500	One-sided LNCE (r=0.5)	0.00023	0.04800	1.15185
500	Symmetric LNCE (r=0.5)	0.00100	0.04532	1.02640
500	Robust KL-LNCE	0.00085	0.04603	1.05925
500	Huber M (k=1.345)	0.00087	0.04643	1.07745
500	DPD (α=0.2)	0.00086	0.04620	1.06706

The theoretical scaled asymptotic variances are 1 for the mean, $\pi/2 = 1.57080$ for the median, 1.13610 for one-sided LNCE with r = 0.5, 1.01045 for symmetric LNCE with r = 0.5, approximately 1.03565 for robust KL-LNCE, 1.05263 for the Huber estimator with k = 1.345, and 1.04316 for the DPD estimator with $\alpha = 0.2$. The Monte Carlo results closely follow these predictions as n increases.

7.2 Contamination experiment

Robustness was investigated under

$$X \sim (1-\epsilon)N(0,1) + \epsilon N(0,25).$$

The contamination levels were

$$\epsilon = 0.05$$

and

$$\epsilon = 0.10.$$

The true center remained zero.

Table 3. Monte Carlo RMSE under contamination

n	ε	Mean	Median	One-sided	Symmetric	Robust KL	Huber	DPD
100	.05	0.14860	0.13103	1.26729	0.76282	0.11169	0.11093	0.10847
250	.05	0.09385	0.08281	1.34067	0.71624	0.07020	0.06990	0.06861
100	.10	0.18296	0.13680	1.88578	0.96823	0.12103	0.11902	0.11519
250	.10	0.11415	0.08503	1.97096	0.83503	0.07346	0.07226	0.06990

The raw likelihood-ratio estimators are highly unstable under gross-error contamination because exponential likelihood ratios give extreme observations very large leverage. The bounded intrinsic KL-LNCE estimator remains stable and is close to the Huber and DPD benchmarks; in these particular Gaussian-contamination settings, DPD has the smallest RMSE, followed closely by Huber and robust KL-LNCE. This comparison is intentionally included to separate stability of the proposed construction from claims of universal robust optimality.

7.3 Scaled variances under contamination

Table 4. $n\widehat{Var}$ under contamination

n	ε	Mean	Median	One-sided	Symmetric	Robust KL	Huber	DPD
100	.05	2.2082	1.7162	114.1689	58.1872	1.2475	1.2306	1.1766
250	.05	2.2021	1.7138	256.1893	128.2178	1.2319	1.2215	1.1768
100	.10	3.3474	1.8715	188.9006	93.7459	1.4647	1.4165	1.3268
250	.10	3.2577	1.8074	358.4453	174.2819	1.3486	1.3049	1.2211

The contrast is substantial. Finite unbounded likelihood ratios can be efficient under the model but extremely nonrobust under contamination. Bounding the likelihood-neighborhood contrast restores stability, although classical Huber and DPD procedures can be slightly more efficient for the specific Gaussian-contamination mechanism considered here. The LNCE tuning parameters therefore govern a distinct tradeoff among nominal efficiency, finite-neighborhood sensitivity, and robustness rather than guaranteeing dominance over established robust estimators.

8. Extensions, diagnostics, and computation

The parameter ε has a statistical interpretation fundamentally different from a numerical finite-difference step. Under intrinsic KL calibration, ε specifies the separation between two actual probability distributions.

Small ε gives local score-like inference, whereas moderate ε gives finite-model-change calibration. Under bounded transformation, ε also controls the scale at which likelihood changes begin to interact with the bounding function.

Thus the neighborhood radius governs a three-way tradeoff among local efficiency, finite-neighborhood sensitivity, and robustness.

8.1 Model diagnostics from overidentifying neighborhoods

When several neighborhood radii are used but only one parameter is estimated, the remaining moment restrictions can be used to assess model fit. If $\theta = \theta_0$, then every neighborhood satisfies

$$E_{\theta_0}\left[\Psi_{\varepsilon_j}(X;\theta_0)\right] = 0.$$

Therefore a fitted model should simultaneously calibrate all neighboring probability laws. With m > 1, let $\hat{S}$ denote a consistent estimator of S evaluated at the fitted parameter. The feasible optimally weighted GMM overidentification statistic is

$$J_n = n\overline{\boldsymbol{\Psi}}_n(\hat{\theta}_m)^{\top}\hat{\hat{S}}^{-1}\overline{\boldsymbol{\Psi}}_n(\hat{\theta}_m).$$

Under standard conditions and correct specification,

$$J_n \xrightarrow{d} \chi^2_{m-1}.$$

Thus LNCE supplies a model diagnostic based on whether the fitted model correctly predicts its own finite neighboring likelihood ratios.

8.2 Multiparameter extension

Let

$$\boldsymbol{\theta} \in \mathbb{R}^p.$$

For a neighboring parameter $\boldsymbol{\vartheta}$, the fundamental identity remains

$$E_{\boldsymbol{\theta}}\left[\frac{f(X;\boldsymbol{\vartheta})}{f(X;\boldsymbol{\theta})}\right] = 1.$$

A finite set of neighboring models $\boldsymbol{\vartheta}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\vartheta}_m(\boldsymbol{\theta})$ therefore produces an m-dimensional moment vector.

Locally, choose directions $v_1, \ldots, v_k$ normalized by Fisher information:

$$v_j^{\top} I(\boldsymbol{\theta}) v_j = 1.$$

Symmetric likelihood-ratio secants along $\boldsymbol{\theta} \pm \varepsilon v_j$ converge to

$$v_j^{\top} \mathbf{s}_{\boldsymbol{\theta}}.$$

With p Fisher-orthonormal directions spanning the tangent space, the collection converges to the full score vector. An intrinsic construction may instead select neighboring probability distributions by a divergence constraint and directional condition. A complete multiparameter treatment, particularly concerning optimal choices of finite directions, is left for separate development.

8.3 Computational advantages and limitations

The proposed estimator requires likelihood evaluations at neighboring parameter values but does not require analytic differentiation of the likelihood. This can be useful when $f(x;\theta)$ is evaluable while $\partial_\theta \log f(x;\theta)$ is inconvenient or unstable.

The method nevertheless differs from ordinary numerical differentiation because the likelihood-ratio normalization equation remains an exact population identity for finite radii.

Several limitations must be acknowledged. First, common support is important. If support changes with the parameter, the elementary normalization identity may require correction for singular components. Second, raw likelihood ratios require finite second moments for conventional root-n inference. Third, finite neighborhoods can be strongly nonrobust. Fourth, a single neighborhood need not globally identify the parameter. Fifth, construction of equal-KL neighbors may require numerical root finding. Sixth, model centering of a bounded contrast may require numerical integration, whose error must be controlled. Seventh, in high-dimensional parameter spaces, selection of a useful finite set of neighborhood directions is nontrivial.

8.4 Relation to existing methods

Finite likelihood ratios at test points have a long history in statistical lower-bound theory. Barankin (1949), Hammersley (1950), and Chapman and Robbins (1951) use likelihood ratios between Pθ and alternative parameter values to obtain variance bounds for unbiased estimators. Those results constrain estimator variance; LNCE instead uses selected within-family likelihood ratios themselves as unbiased estimating moments. The normalization identity and finite ratio are therefore classical, while the estimating-equation construction and its neighborhood design are the objects developed here.

Likelihood-ratio change-of-measure methods are also established tools for sensitivity analysis in simulation (Reiman & Weiss, 1989). Their principal target is differentiation of expectations with respect to a parameter. LNCE has a different target: for each finite admissible neighborhood it retains the exact normalization moment and solves that moment for the unknown parameter. Thus the neighborhood radius is part of the statistical estimating rule, not merely a numerical differentiation step.

LNCE is also distinct from semiparametric density-ratio models based on separate populations (Qin, 1998) and from Hyvärinen score matching (Hyvärinen, 2005), which differentiates the log density with respect to the observation and minimizes a score-matching objective. The equal-KL neighborhood construction uses the local Fisher-Rao geometry of the model (Amari & Nagaoka, 2000), but uses that geometry specifically to select finite estimating directions.

Once the neighborhood moments are constructed, their combination belongs to standard estimating-function and GMM theory (Godambe, 1960; Heyde, 1997; Hansen, 1982; Carrasco & Florens, 2000). Likewise, the bounded version should be viewed alongside classical robust M-estimation and divergence-based robust estimation (Huber, 1964; Basu et al., 1998). The distinctive LNCE contribution is the systematic pipeline: choose neighboring members of the same parametric family, form exact within-family likelihood-ratio calibration moments, optionally symmetrize them or choose

them intrinsically by equal KL distance, aggregate them, and, when needed, bound and recenter the finite contrast.

9. Discussion

The likelihood score is usually treated as a primitive object in parametric inference. The present framework provides another perspective. The score can be recovered from the family itself by examining ratios between finite neighboring probability laws:

$$s_\theta = \lim_{h\to 0} \frac{1}{h}\left[\frac{f_{\theta+h}}{f_\theta} - 1\right].$$

Thus the score is the tangent of a more general collection of finite likelihood-ratio secants.

This perspective creates statistical objects that disappear when the neighborhood is immediately sent to zero. One is the efficiency spectrum $\mathcal{E}_\theta(\varepsilon)$. Another is a collection of overidentifying finite-model calibration restrictions. A third is a tunable robustification mechanism.

The normal and exponential examples reveal the local pattern established by the preceding expansions. One-sided neighborhoods generally incur a second-order efficiency loss. Under the stated smoothness and symmetry conditions, symmetric neighborhoods remove that leading distortion and have fourth-order local efficiency loss.

For normal location,

$$\mathrm{ARE}_{\mathrm{sym}}(r) = \frac{r^2}{\sinh(r^2)} = 1 - \frac{r^4}{6} + O(r^8).$$

For the exponential symmetric relative neighborhood,

$$\frac{n\mathrm{Avar}\left(\hat{\lambda}_r\right)}{\lambda^2} = 1 + 5r^4 + O(r^6).$$

The robustness analysis also exposes a limitation of finite likelihood ratios. A secant extending toward a more distant model can amplify atypical observations much more strongly than a local score. Bounding the finite contrast is therefore important when contamination is plausible.

This leads to a useful distinction: finite likelihood neighborhoods provide information beyond the local tangent representation, but finite likelihood ratios require explicit robustness control when contamination is plausible.

10. Conclusions

This paper develops a systematic parametric estimating-equation framework from the exact identity

$$E_\theta\left[\frac{f(X;\vartheta)}{f(X;\theta)}\right] = 1.$$

When ϑ is deliberately chosen as a neighboring member of the same parametric family, this identity becomes an exact estimating equation. The resulting method, Likelihood-Neighborhood Calibration Estimation, provides a finite-neighborhood extension of ordinary score-based inference.

For a one-sided neighborhood,

$$g_h(x;\theta) = \frac{f(x;\theta+h)}{f(x;\theta)} - 1$$

is unbiased. Its Godambe information is

$$G_h(\theta) = \frac{\{E_{\theta+h}[s_\theta(X)]\}^2}{\chi^2(P_{\theta+h}\|P_\theta)}.$$

Its relative efficiency is exactly the squared correlation between the likelihood-ratio secant and the score.

A symmetric neighborhood,

$$\frac{f_{\theta+h}/f_\theta - f_{\theta-h}/f_\theta}{2h},$$

approximates the likelihood score with second-order rather than first-order error. When neighbors are selected at equal Kullback-Leibler distance,

$$D_{\mathrm{KL}}\big(P_\theta \| P_{\theta_{\varepsilon,\mathrm{R}}}\big) = D_{\mathrm{KL}}\big(P_\theta \| P_{\theta_{\varepsilon,\mathrm{L}}}\big) = \frac{\varepsilon^2}{2}.$$

the construction becomes intrinsic to the statistical model and invariant to smooth reparameterization. The intrinsic secant satisfies

$$\Psi_{\theta,\varepsilon} = I(\theta)^{-1/2} s_\theta + O_{L_2}(\varepsilon^2),$$

which yields

$$1 - \mathrm{ARE}(\varepsilon) = O(\varepsilon^4).$$

Multiple neighborhoods form an exact overidentified moment system whose optimal information is the L_2 projection of the score onto the span of the likelihood secants. A bounded transformation of finite log-likelihood-ratio contrasts preserves exact centering and produces bounded influence.

The framework therefore links likelihood ratios, score functions, divergences, estimating equations, and robustness.

The central conclusion is that the ordinary score can be viewed both as a derivative and as the zero-radius limit of exact finite within-family calibration moments. The finite-neighborhood family is therefore a legitimate source of estimating equations, with its efficiency and robustness governed by the choice and transformation of neighboring probability laws.

Appendix A. One-sided normal asymptotic variance

Let

$$Y_i = \exp\left(\frac{hX_i}{\sigma^2}\right).$$

Then

$$\hat{\mu}_{\mathrm{R},h} = \frac{\sigma^2}{h}\log\overline{Y} - \frac{h}{2}.$$

Under $X \sim N(\mu, \sigma^2)$,

$$E(Y) = \exp\left(\frac{h\mu}{\sigma^2} + \frac{h^2}{2\sigma^2}\right),$$

and

$$E(Y^2) = \exp\left(\frac{2h\mu}{\sigma^2} + \frac{2h^2}{\sigma^2}\right).$$

Therefore,

$$\frac{\mathrm{Var}(Y)}{E(Y)^2} = e^{h^2/\sigma^2} - 1.$$

The delta method gives

$$n\mathrm{Var}\big(\hat{\mu}_{\mathrm{R},h}\big) = \frac{\sigma^4}{h^2}\big(e^{h^2/\sigma^2} - 1\big).$$

With $r = h/\sigma$,

$$n\mathrm{Var}\big(\hat{\mu}_{\mathrm{R},h}\big) = \sigma^2\frac{e^{r^2} - 1}{r^2}.$$

Appendix B. Symmetric normal asymptotic variance

Let

$$R_{\mathrm{R}} = e^{rZ-r^2/2}, \qquad R_{\mathrm{L}} = e^{-rZ-r^2/2}.$$

Then

$$\psi_r = \frac{R_{\mathrm{R}} - R_{\mathrm{L}}}{2r} = e^{-r^2/2}\frac{\sinh(rZ)}{r}.$$

Since

$$E[Ze^{rZ}] = re^{r^2/2}$$

and

$$E[Ze^{-rZ}] = -re^{r^2/2},$$

we have

$$E[Z\psi_r] = 1.$$

Also,

$$\sinh^2(rZ) = \frac{\cosh(2rZ) - 1}{2},$$

and

$$E[\cosh(2rZ)] = e^{2r^2}.$$

Therefore,

$$E[\sinh^2(rZ)] = \frac{e^{2r^2} - 1}{2},$$

so

$$E[\psi_r^2] = \frac{e^{-r^2}}{r^2}\frac{e^{2r^2} - 1}{2} = \frac{\sinh(r^2)}{r^2}.$$

Thus the relative Godambe information is

$$\frac{r^2}{\sinh(r^2)}.$$

Appendix C. Exponential symmetric-neighborhood calculations

Let

$$Y = \lambda X,$$

so that $Y \sim \mathrm{Exp}(1)$. The two ratios are

$$R_{\mathrm{R}} = (1+r)e^{-rY},$$

and

$$R_{\mathrm{L}} = (1-r)e^{rY}.$$

For $r < 1/2$,

$$E[R_{\mathrm{R}}^2] = \frac{(1+r)^2}{1+2r},$$

and

$$E[R_{\mathrm{L}}^2] = \frac{(1-r)^2}{1-2r}.$$

Also,

$$R_{\mathrm{R}}R_{\mathrm{L}} = 1 - r^2.$$

Thus,

$$E[(R_{\mathrm{R}} - R_{\mathrm{L}})^2] = \frac{(1+r)^2}{1+2r} + \frac{(1-r)^2}{1-2r} - 2(1-r^2).$$

Simplification gives

$$E[(R_{\mathrm{R}} - R_{\mathrm{L}})^2] = \frac{4r^2(1-2r^2)}{1-4r^2}.$$

Therefore, for $\psi_r = (R_{\mathrm{R}} - R_{\mathrm{L}})/(2r)$,

$$E[\psi_r^2] = \frac{1-2r^2}{1-4r^2}.$$

Next,

$$\frac{\partial R_{\mathrm{R}}}{\partial \lambda} = -(1+r)rxe^{-r\lambda x},$$

while

$$\frac{\partial R_{\mathrm{L}}}{\partial \lambda} = (1-r)rxe^{r\lambda x}.$$

Therefore,

$$-E\left[\frac{\partial \psi_r}{\partial \lambda}\right] = \frac{1}{2}\left[(1+r)E\left(Xe^{-r\lambda X}\right) + (1-r)E\left(Xe^{r\lambda X}\right)\right].$$

Now,

$$E\left(Xe^{-r\lambda X}\right) = \frac{1}{\lambda(1+r)^2},$$

and

$$E\left(Xe^{r\lambda X}\right) = \frac{1}{\lambda(1-r)^2}.$$

Hence,

$$A_r = \frac{1}{\lambda(1-r^2)}.$$

Consequently,

$$\mathrm{Avar}\left(\hat{\lambda}_r\right) = \frac{1}{n}\frac{E[\psi_r^2]}{A_r^2},$$

which yields the variance formula stated in Section 6.13.

Appendix D. Simulation design

For every Monte Carlo replication under the correctly specified model,

$$X_1, \ldots, X_n \overset{iid}{\sim} N(0,1).$$

For contaminated samples,

$$X_i = \begin{cases} N(0,1), & \text{with probability } 1-\epsilon, \\ N(0,25), & \text{with probability } \epsilon. \end{cases}$$

The one-sided and symmetric LNCE estimators were evaluated from their closed-form normal expressions. The robust KL-LNCE and Huber estimating equations were solved by bracketing/bisection to numerical tolerance 10^{-10}. The DPD fixed-point iteration was initialized at the sample median and stopped when successive iterates differed by less than 10^{-10}; no convergence failures occurred in the reported runs. All tables use 5000 independent replications generated with NumPy 2.3.5 using the default PCG64 generator and fixed seed 20261004. Monte Carlo standard

errors of the RMSE for the stable estimators (mean, median, robust KL-LNCE, Huber, and DPD) are below 0.0019 in all reported contamination settings.

References

1. Amari, S., & Nagaoka, H. (2000). Methods of Information Geometry. Providence, RI: American Mathematical Society.
2. Barankin, E. W. (1949). Locally best unbiased estimates. The Annals of Mathematical Statistics, 20, 477-501. doi:10.1214/aoms/1177729943.
3. Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. Biometrika, 85, 549-559. doi:10.1093/biomet/85.3.549.
4. Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. Econometric Theory, 16, 797-834. doi:10.1017/S0266466600166010.
5. Chapman, D. G., & Robbins, H. (1951). Minimum variance estimation without regularity assumptions. The Annals of Mathematical Statistics, 22, 581-586. doi:10.1214/aoms/1177729548.
6. Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics, 31, 1208-1211. doi:10.1214/aoms/1177705693.
7. Hammersley, J. M. (1950). On estimating restricted parameters. Journal of the Royal Statistical Society: Series B, 12, 192-229. doi:10.1111/j.2517-6161.1950.tb00056.x.
8. Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50, 1029-1054. doi:10.2307/1912775.
9. Heyde, C. C. (1997). Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. New York: Springer.
10. Huber, P. J. (1964). Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35, 73-101. doi:10.1214/aoms/1177703732.
11. Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research, 6, 695-709.
12. Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. The Annals of Mathematical Statistics, 22, 79-86. doi:10.1214/aoms/1177729694.
13. Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle & D. L. McFadden (Eds.), Handbook of Econometrics (Vol. 4, pp. 2111-2245). Elsevier. doi:10.1016/S1573-4412(05)80005-4.
14. Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. Biometrika, 85, 619-630. doi:10.1093/biomet/85.3.619.
15. Reiman, M. I., & Weiss, A. (1989). Sensitivity analysis for simulations via likelihood ratios. Operations Research, 37, 830-844. doi:10.1287/opre.37.5.830.
16. Small, C. G., & Wang, J. (2003). Numerical Methods for Nonlinear Estimating Equations. Oxford: Oxford University Press.
17. Van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511802256.