IC internal seminar

February 20, 2026

Based on work with Peter McCullagh and Daniel Xiang (both Chicago).

Likelihood asymptotics

Model parametrised by $\theta \in \Theta$. Log likelihood $\ell(\theta)$, MLE $\hat{\theta}$.

Standard regularity conditions give

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, v), \quad \text{under } H_0 : \theta = \theta_0 \qquad (1)$$

Violate one regularity condition: Θ bounded, θ_0 at boundary. (1) holds, but boundary forces null probability $\frac{1}{2}$ at θ_0.

Violate two regularity conditions: θ_0 at boundary, and no Fisher information there.

Highly non-standard distribution theory for $\hat{\theta}$ and LR statistic.

Simple model breaking two regularity conditions

Model with bounded parameter space:

$$f_\theta = (1-\theta)f_0 + \theta f_1, \quad 0 \le \theta \le 1$$

If f_0 is standard normal and f_1 has heavier tails, then typically the Fisher information does not exist.

Model arises in Efron's approach to multiple testing.

Scientific context

Two different settings for multiple comparison considerations.

1. Statements probing same scientific null: simultaneous correctness?
   e.g. particle physics: many tests of H_0: "no new physics".
   Emphasis: avoidance of false declarations of discovery.
2. Different scientific null hypotheses: few are false by assumption.
   e.g. genomics: multiple hypotheses $H_{0,i}$: genomic locus i biologically irrelevant. Emphasis: detection of relevant loci.

Second case is more decision theoretic: relates to selective inference.
Some subtle conceptual points, but not for this talk...

Two-component mixture formulation

Reduce to pivotal test statistic at each site. Any statistic is either from f_0 or f_1:

- f_0: standard normal (distn of test stat. under null);
- f_1: heavy-tailed (distn of test stat. across sites when null is violated).

The statistic at an arbitrary site is treated as a draw from $f_\theta = (1-\theta)f_0 + \theta f_1$ on account of the label ('signal' or 'noise') being unknown.

Left vs right boundary

Two boundary cases

- $\theta = 0$: all observations generated from f_0;
- $\theta = 1$: all observations generated from f_1.

Behaviour of $\mathbb{P}_0(\hat{\theta}_n > 0)$ and $\mathbb{P}_1(\hat{\theta}_n < 1)$ as $n \to \infty$ entirely different.

Likelihood geometry

Density function: $f_\theta(x) = (1-\theta) f_0(x) + \theta f_1(x)$.

Log likelihood: $\ell(\theta) = \sum_i \log f_\theta(X_i)$, $0 \le \theta \le 1$.

second derivative: $\ell''(\theta) \le 0$ uniformly in θ (concave).

First derivative: {negative, zero, positive} for $\hat{\theta}_n$ at {left boundary, interior, right boundary}.

Left boundary: $\hat{\theta}_n = 0 \iff \ell'(0) = \sum_{i=1}^{n} (h(X_i) - 1) \le 0$.

Density ratio: $h(X) = f_1(X)/f_0(X)$ governs $\mathbb{P}_0(\hat{\theta}_n > 0)$.

Left boundary event

Density ratio: $h(x) = f_1(x)/f_0(x)$.

Left boundary: $\hat{\theta}_n = 0 \iff \ell'(0) = \sum_{i=1}^n (h(X_i) - 1) \leq 0$.

Equivalent events: $\{\hat{\theta}_n > 0\} = \{\frac{1}{n}\sum_{i=1}^n h(X_i) > 1\}$.

$\hat{\theta}_n > 0 \iff$ sample average of $h(X_i)$ exceeds expected value:

$$\mathbb{E}(h(X)) = \int h(x) f_0(x)\, dx = \int f_1(x)\, dx = 1.$$

Standard boundary inference results

For true θ an interior point of the parameter space,

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, v)$$

under regularity conditions.

Since zero is interior to the extended parameter space $\mathbb{R}$, $\lim \mathbb{P}_0(\hat{\theta}_n > 0) = \frac{1}{2}$ if regularity conditions met.

Non-standard inference at left boundary

Density ratio: $h(x) = f_1(x)/f_0(x)$.

Left boundary: $\hat{\theta}_n = 0 \iff \ell'(0) = \sum_{i=1}^n (h(X_i) - 1) \leq 0$.

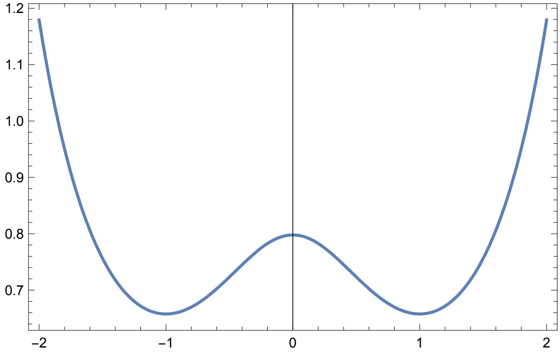
Equivalent events: $\{\hat{\theta}_n > 0\} = \{\frac{1}{n}\sum_{i=1}^n h(X_i) > 1\}$.

$\hat{\theta}_n > 0 \iff$ sample average of $h(X_i)$ exceeds expected value

$$\mathbb{E}(h(X)) = \int h(x) f_0(x)\, dx = \int f_1(x)\, dx = 1.$$

How does $\lim \mathbb{P}_0(\hat{\theta}_n > 0)$ behave if $h(X)$ does not have finite variance?

Plot of $h(x) = f_1(x)/f_0(x)$ for f_0 standard normal, f_1 standard Cauchy



For f_0 standard Gaussian, we consider f_1 such that $h(x)$ is *ultimately monotone*.

Ultimate monotonicity

Definition: For each η sufficiently large $\exists\, \xi$ s.t.

$$\{x : h(x) > \eta\} = \{x : |x| > \xi\}.$$

Implication: Tail behaviour for single $h(X_i)$

$$\begin{aligned} \mathbb{P}_0(h(X) > \eta) &= \mathbb{P}_0(|X| > \xi) \\ &\sim 2\xi^{-1}\phi(\xi) \quad \text{[Mills's approx.]} \\ &= \frac{2f_1(\xi)}{\eta\,\xi} \quad \text{[implicit def. } \phi(\xi) = \eta^{-1} f_1(\xi)\text{]}. \end{aligned}$$

The asymptotic inverse relationship $\xi(\eta) = h^{-1}(\eta)$ determines the null tail behaviour via $f_1(\xi(\eta))/\xi(\eta)$.

Large-sample behaviour of $\hat{\theta}_n$

Equivalent events: $\{\hat{\theta}_n > 0\} = \{\frac{1}{n}\sum_{i=1}^n h(X_i) > 1\}$.

Density ratio $h(X)$: unit mean, infinite variance, $\mathbb{P}_0(h(X) > \eta)$ governed by $f_1(\xi(\eta))/\xi(\eta)$, where $\xi(\eta) = h^{-1}(\eta)$.

Limit distn of $\hat{\theta}_n$: follows non-Gaussian stable law after standardisation.

Stable limit laws

Members of stable class parametrised by $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$ have log characteristic function of form

$$\log \psi(t) = \begin{cases} -t^2 & (\alpha = 2), \\ -|t|\left(1 + i\beta \,\mathrm{sign}(t) \frac{2}{\pi} \log |t|\right) & (\alpha = 1), \\ -|t|^\alpha \left(1 - i\beta \,\mathrm{sign}(t) \tan(\pi\alpha/2)\right) & (\alpha \neq 1). \end{cases}$$

Properties expounded by: Gnedenko & Kolmogorov (1954), Zolatarev (1986), Bingham, Goldie, Teugels (1987).

Stability

For $h(X)$ in domain of attraction of Stable(α, β)

$$B_n^{-1} \sum_{i=1}^{n} h(X_i) - A_n \to_d \text{Stable}(\alpha, \beta), \quad n \to \infty.$$

Stabilising A_n, B_n depend on domain of attraction.

Stability

For $h(X)$ in domain of attraction of Stable(α, β)

$$B_n^{-1} \sum_{i=1}^{n} h(X_i) - A_n \to_d \text{Stable}(\alpha, \beta), \quad n \to \infty.$$

Stabilising A_n, B_n depend on domain of attraction.

That each $h(X_i)$ is positive and has finite mean implies that $\beta = 1$ and $\alpha \geq 1$.

For f_1 having heavier tails than Gaussian f_0, $\alpha = 1$ is the most important case.

Domain of attraction (DoA)

Fix $\beta = 1$ (skewness param.). Random variable Z is in the DoA of a stable law with index $0 < \alpha < 2$ iff

$$\bar{F}(z) := \mathbb{P}(Z > z) \sim \frac{2C_\alpha}{z^\alpha L(z^\alpha)}, \quad (z \to \infty)$$

where L is a slowly varying function.

Slow variation: at large scales, scaling immaterial: $\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0.$

Stabilising sequences ($1 < \alpha \leq 2$)

Stabilising sequence B_n obtained from L. For $1 < \alpha \leq 2$, Z has finite mean μ, $A_n = n\mu/B_n$, and

$$B_n^{-1} \sum_{i=1}^{n} Z_i - A_n = \frac{1}{B_n} \sum_{i=1}^{n} (Z_i - \mu) \to_d \text{Stable}(\alpha, \beta)$$

For $\beta = 1$, and ε a RV with stable distribution $P_{\alpha,1}$

$$\lim_{n\to\infty} \mathbb{P}(\bar{Z}_n > \mu) = P_{\alpha,1}(\varepsilon > 0) = 1 - 1/\alpha.$$

Apply with $Z_i = h(X_i)$: $\lim_{n\to\infty} \mathbb{P}_0(\hat{\theta}_n > 0) = 1 - 1/\alpha$.

Recovers standard boundary behaviour for $\alpha = 2$.

Cauchy domain of attraction ($\alpha = 1$)

Most important case. Theory more difficult.

$$\mathbb{P}_0(\hat{\theta}_n > 0) \to 0 \quad \text{but at slow rate}.$$

Rate requires calculation of A_n and B_n.

Stabilising sequences ($\alpha = 1$)

Upper probabilities of form $\bar{F}(z) \sim 2C_1/(zL(z))$. Consider the parametrised family

$$L(z) = \begin{cases} (\beta_0 \log z)^{\delta+1} e^{(\beta_1 \log z)^\gamma} & \beta_1 > 0 \\ (\beta_0 \log z)^{\delta+1} & \beta_1 = 0. \end{cases}$$

Encompasses examples arising in two-component mixture problems. Calculation gives

$$A_n = \frac{n\mu}{B_n} - K_{\delta,\gamma,\beta_1} \frac{n(\log B_n)^{1-\gamma}}{B_n L(B_n)}(1 + o(1)),$$

where $K_{\delta,\gamma,\beta_1} = 2C_1/(\beta_1^\gamma \gamma)$ for $\gamma, \beta_1 > 0$, and $K_{\delta,0,0} = 2C_1/\delta$.

Implication

$$\begin{aligned} \mathbb{P}(\bar{Z}_n > \mu) &\sim P_{1,1}\left(\varepsilon > \frac{n\mu}{B_n} - A_n\right) \\ &\sim 2K_{\delta,\gamma,\beta_1}^{-1}(\log n)^{\gamma-1}/\pi, \quad (n \to \infty). \end{aligned}$$

Apply with $Z_i = h(X_i)$: $\mathbb{P}_0(\hat{\theta}_n > 0)$ approaches zero at rate governed by tail behaviour of distribution of $h(X)$, characterised by δ, γ, β_1.

Regularly varying vs. exponential tails

Recall: the tail behaviour of $h(X)$ is induced by the tail of f_1.

- Suppose
$$\lim_{x\to\infty} \frac{\bar{F}_1(tx)}{\bar{F}_1(x)} = t^{-2\delta}, \quad t > 0,$$
i.e., regularly-varying tails with index -2δ. Then
$$\mathbb{P}_0(\hat{\theta}_n > 0) \sim \frac{\delta}{\log n}.$$

- Suppose $-\log \bar{F}_1(x) \sim |x|^{2\gamma}$, $0 < \gamma < 1$, i.e., exponential tails. Then
$$\mathbb{P}_0(\hat{\theta}_n > 0) \sim \frac{\gamma\, 2^{\gamma}}{(\log n)^{1-\gamma}}.$$

Example: Cauchy vs. Laplace

- If f_1 is standard Cauchy,

$$\mathbb{P}_0(\hat{\theta}_n > 0) \sim \frac{1}{2\log n}.$$

- If f_1 is standard Laplace

$$\mathbb{P}_0(\hat{\theta}_n > 0) \sim \frac{1}{(2\log n)^{1/2}}.$$

Intuition: notional signal, if it existed, would be easier to detect if drawn from a distribution with heavier tail: $\mathbb{P}_0(\hat{\theta}_n > 0)$ converges more quickly.

Why does $\alpha = 1$ characterise most cases?

From before:

$$\mathbb{P}_0(h(X) > \eta) \sim \frac{2f_1(\xi)}{\eta\xi}$$

where $\xi(\eta) = h^{-1}(\eta)$.

Enlarge Gaussian tails even slightly e.g. $f_1(x) \sim e^{-|x|^{2\kappa}L(x)}$, $0 < \kappa < 1$,

$$h(x) \sim \exp\left(\frac{x^2}{2} - |x|^{2\kappa}L(x)\right) \sim \exp(x^2/2).$$

Inversion of $h(\xi) = \eta$ gives $\xi(\eta) \sim (2\log\eta)^{1/2}$. Implies $(f_1(\xi)/\xi)^{-1}$ is slowly varying in η.

Left vs. right boundary

Left boundary: $\mathbb{P}_0(\hat{\theta}_n > 0)$ non-standard.

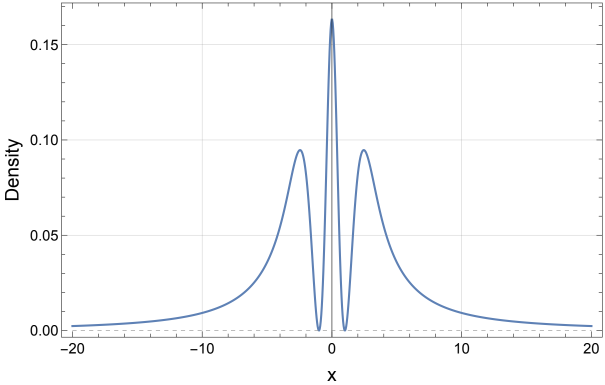
Right boundary: $\mathbb{P}_1(\hat{\theta}_n < 1)$ standard.

Define:
$$\theta_{\mathrm{min}} := \min\{\theta : f_\theta \text{ is a valid pdf}\}$$
$$\theta_{\mathrm{max}} := \max\{\theta : f_\theta \text{ is a valid pdf}\}.$$

When f_1 has heavier tails than f_0, typically $\theta_{\mathrm{min}} = 0$ and $\theta_{\mathrm{max}} > 1$.

Left boundary is real; right boundary is an interior point of an extended parameter space.

Plot of $f_{\theta_{\max}}$ for f_1 standard Cauchy: $\theta_{\max} \approx 2.922$.



Reparametrisation

$$(1-\rho)f_0 + \rho f_{\theta_{\mathrm{max}}}, \quad 0 \leq \rho \leq 1$$

Left: $\theta = 0$ in original parametrisation corresponds to $\rho = 0$ in new parametrisation.

Right: $\theta = 1$ in original parametrisation corresponds to $\rho = 1/\theta_{\mathrm{max}} < 1$ in new parametrisation. Interior point.

Likelihood ratio: failure of standard argument

Conditional on $\hat{\theta}_n > 0$ the log likelihood-ratio statistic for testing $\theta = 0$ is

$$\Lambda_n = 2\{\ell(\hat{\theta}_n) - \ell(0)\}.$$

Conventional arguments for limit distribution based on Taylor expansion fail: first derivative $\ell'(0)$ does not have finite variance, so there is no notion of Fisher information.

But large-sample distribution is close in simulations to $\kappa_n \chi^2_1$, where $\kappa_n \to 1$ is a Bartlett correction factor.

Likelihood ratio approximation

Conditional on $\hat{\theta}_n > 0$, Λ_n has the approximation

$$\Lambda_n = 2\sum_{k\geq 2} R^k/k + o_p(1)$$

where $R = n\bar{Z}_n/Z_{(n)}$, $Z_i = h(X_i) - 1$.

Rough intuition: first-order approx

$$2\{\ell(\hat{\theta}_n) - \ell(0)\} \approx 2\ell_n'(0)\hat{\theta} = 2n\bar{Z}\hat{\theta}.$$

On the event $\hat{\theta}_n > 0$, $\hat{\theta}_n$ has the scale of $1/Z_{(n)}$. Extend to higher order.

Likelihood ratio limit distribution

Conditional on $\hat{\theta}_n > 0$, Λ_n has the approximation

$$\Lambda_n = 2 \sum_{k \geq 2} R^k / k + o_p(1)$$

where $R = n\bar{Z}_n / Z_{(n)}$, $Z_i = h(X_i) - 1$.

Limit distribution requires a new limit theorem for the conditional limit distribution of $\bar{Z}_n$ and $Z_{(n)}$ conditional on $\bar{Z}_n > 0$.

Likelihood ratio: limit distribution

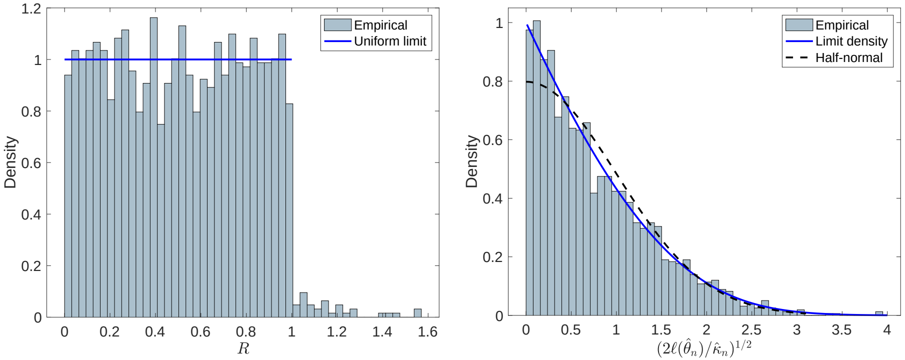


Figure: Histogram of $R = n\bar{Z}_n/Z_{(n)}$ (left) and $(2\ell(\hat{\theta}_n)/\hat{\kappa}_n)^{1/2}$ (right) for 2000 simulations of the Gauss-Cauchy mixture model with $n = 10^7$ observations restricted to samples for which $\bar{Z}_n > 0$.

Summary

- Left boundary probability $\mathbb{P}_0(\hat{\theta}_n > 0)$ depends on tail behaviour of f_1 via $h(x) = f_1(x)/f_0(x)$.
- Infinite variance of $h(X)$ invalidates standard theory.
- $h(X)$ in the skew-Cauchy domain of attraction is the default in Gaussian mixtures.
- $\mathbb{P}_0(\hat{\theta}_n > 0)$ inherits different rates for different tails of f_1 via the stabilising sequences A_n and B_n for $\sum_i h(X_i)$.
- Likelihood ratio limit conditional on $\hat{\theta}_n > 0$ is not χ_1^2.

The end

Thank you for your attention

Battey, H., McCullagh, P. and Xiang, D. (2025+). Non-standard boundary behaviour in two-component mixture models.