ML Sampling Theory for non-Gaussian models: Asymptotic Approximation

Surya Tokdar

Frequentist guarantees for non-Gaussian model

- $X \sim f(x|\theta)$, $\theta \in \Theta$
- ML intervals:

$$B_c(x) = \{\theta \in \Theta : \ell_x(\theta) \geq \ell_x(\hat{\theta}_{\text{MLE}}(x)) - c^2/2\}$$

- ML test for $H_0 : \theta \in \Theta_0$

$$\delta_c(x) \longleftrightarrow \text{reject } H_0 \text{ if } B_c \cap \Theta_0 = \emptyset$$

- How to calculate $P_{[X|\theta]}(\theta \in B_c(X))$?
- How to calculate $P_{[X|\theta]}(\Theta_0 \cap B_c(X) = \emptyset)$?

Case by case

1. $f(x|\theta)$ a pmf on a finite set S and Θ too is a finite set.
   - Calculate by complete enumeration
   - See HW 2
2. $X_1, \cdots, X_n \overset{\text{IID}}{\sim} Unif(0, \theta)$, $\theta \in (0, \infty)$
   - Fairly elegant general calculations
   - Can express $\gamma(\theta, A_k)$ in closed form
   - HW 2
3. $X_1, \cdots, X_n \overset{\text{IID}}{\sim} Poi(\mu)$, $\mu > 0$
   - An exact calculation
   - And a very accurate approximation (for large n)
   - Will do this now...

ML sampling theory for Poisson model

- $X_1, \cdots, X_n \overset{\text{IID}}{\sim} Poi(\mu)$.
- $\hat{\mu}_{\text{MLE}}(x) = \bar{x}$, $I_x = n/\bar{x}$.
- ML interval $B_c(x) = \bar{x} \mp c\sqrt{\bar{x}/n}$ and

$$P_{[X|\mu]}(\mu \in B_c(X)) = P_{[X|\mu]}\left(\bar{X} \in \mu + \frac{c^2}{2n} \mp \frac{c}{\sqrt{n}}\sqrt{\mu + \frac{c^2}{4n}}\right)$$

by a simple rearrangement (followed by a square completion)

- But when $X_i \overset{\text{IID}}{\sim} Poi(\mu)$, $T = \sum_{i=1}^n X_i = n\bar{X} \sim Poi(n\mu)$.

Exact coverage of B_c

μ	n	$\gamma(\mu, B_{1.64})$	$\gamma(\mu, B_{1.95})$	$\gamma(\mu, B_{2.58})$
1	10	0.906	0.926	0.970
	25	0.874	0.948	0.977
	100	0.900	0.945	0.988
5	10	0.904	0.949	0.987
	25	0.901	0.949	0.988
	100	0.901	0.948	0.990
10	10	0.894	0.946	0.989
	25	0.901	0.948	0.990
	100	0.900	0.949	0.990

Approximation

- With a standard rearrangement

$$P_{[X|\mu]}(\mu \in B_c(X)) = P_{[X|\mu]}\left(-c \leq \frac{\bar{X} - \mu}{\sqrt{\bar{X}/n}} \leq c\right)$$

- For "large n": $T = \frac{\bar{X}-\mu}{\sqrt{\bar{X}/n}} \overset{\text{approx}}{\sim} N(0, 1)$
  (Normal approx to Poisson + something more)
- So $P_{[X|\mu]}(\mu \in B_c(X)) \approx 2\Phi(c) - 1$.

Exact coverage of B_c vs. approximation

μ	n	$\gamma(\mu, B_{1.64})$		$\gamma(\mu, B_{1.95})$		$\gamma(\mu, B_{2.58})$	
	10	0.906	0.9	0.926	0.95	0.970	0.99
1	25	0.874	0.9	0.948	0.95	0.977	0.99
	100	0.900	0.9	0.945	0.95	0.988	0.99
	10	0.904	0.9	0.949	0.95	0.987	0.99
5	25	0.901	0.9	0.949	0.95	0.988	0.99
	100	0.901	0.9	0.948	0.95	0.990	0.99
	10	0.894	0.9	0.946	0.95	0.989	0.99
10	25	0.901	0.9	0.948	0.95	0.990	0.99
	100	0.900	0.9	0.949	0.95	0.990	0.99

Asymptotic calculations: Basics

- For the Poisson example, a precise statement is:

$$\lim_{n\to\infty} P_{[X|\mu]}(\mu \in B_c(X)) = \lim_{n\to\infty} P_{[X|\mu]}(-c \le T \le c) = 2\Phi(c) - 1$$

- Here n is implicit in both $X = (X_1, \cdots, X_n)$ and T which is derived from X.
- For the second equality to hold for every $c > 0$, it is necessary and sufficient that $T \xrightarrow{d} N(0, 1)$.

Recall convergence in law

- $T_1, T_2, \cdots$ an infinite sequence of random variables in $\mathbb{R}^d$
- f a pdf on $\mathbb{R}^d$
- T_n is said to converge in law to f if for every interval $A \in \mathbb{R}^d$

$$\lim_{n\to\infty} P(T_n \in A) = \int_A f(z)dz$$

and we write $T_n \xrightarrow{d} f$.
- If $Z \sim f$ then we also write $T_n \xrightarrow{d} Z$.
- So for large n we can approximate $P(T_n \in A)$ by $P(Z \in A)$.

Recall Central Limit Theorem

- Suppose $X_1, X_2, \cdots$ are IID with some pdf/pmf $f(x)$
- Assume $\mu = \mathbb{E}X_i$ and $\Sigma = \mathbb{V}\text{ar}X_i$ are finite
- Then $T_n = \sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N_d(0, \Sigma)$
- Also write $T_n \sim AN_d(0, \Sigma)$, or, $\bar{X} \sim AN_d(\mu, \frac{1}{n}\Sigma)$.
- General notation: write $T_n \sim AN_d(\mu_n, \Sigma_n)$ if

$$B_n^{-1}(T_n - \mu_n) \xrightarrow{d} N_d(0, I_d)$$

where $\Sigma_n = B_n B_n^T$.

Asymptotic normality of MLE

- Model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} g(x_i|\theta), \theta \in \Theta$, g is regular
- MLE $\hat{\theta}_{\text{MLE}}(x)$, curvature I_x.
- Fix $\theta \in \Theta$. If $X_i \overset{\text{IID}}{\sim} g(x_i|\theta)$ then

$$\hat{\theta}_{\text{MLE}}(X) \sim AN_d(\theta, I_X^{-1})$$

- A very useful result!

Continuous mapping theorem

If $T_n \xrightarrow{d} Z$ and $g(t)$ is a continuous function then $g(T_n) \xrightarrow{d} g(Z)$

Coverage probabilities of ML intervals

- Suppose $\hat{\theta}_{\text{MLE}}(X) \sim AN_d(\theta, I_X^{-1})$
- Take $\eta = a^T\theta$: a continuous function of θ
- Recall $B_c(x) = a^T\hat{\theta}_{\text{MLE}}(x) \mp c\sqrt{a^T I_x^{-1} a}$ so

$$P_{[X|\theta]}(\eta \in B_c(X)) = P_{[X|\theta]}(-c \leq T \leq c)$$

  where $T = \frac{a^T\hat{\theta}_{\text{MLE}}(X) - a^T\theta}{\sqrt{a^T I_X^{-1} a}}$
- By cont. map thm., $T \sim AN_1(0,1)$ and so

$$P_{[X|\theta]}(-c \leq T \leq c) = 2\Phi(c) - 1.$$

When is MLE asymptotically normal?

- Not for every model. Not for $X_i \overset{\text{IID}}{\sim} Unif(0,\theta)$. (HW 2)
- You need some regularity of the pdfs collected under the model
- To begin with, they need to be positive on the same set and need to be differentiable in the parameter
- Usually holds for exponential family models

Exponential family result

- Model: $X_i \overset{\text{IID}}{\sim} g(x_i|\theta) = h(x_i)e^{\xi(\theta)^T T(x_i) - B(\theta)}$, $\theta \in \Theta \subset \mathbb{R}^d$
- Assumptions
  1. Θ is an open set
  2. $\xi(\theta)$ is one-to-one, two times differentiable with continuous derivatives
  3. There is no vector b such that $b^T T(X_i)$ is a constant number for any $X_i \sim g(x_i|\theta)$
- Then, $\hat{\theta}_{\text{MLE}}(X) \sim AN_d(\theta, I_X^{-1})$ whenever $X_i \overset{\text{IID}}{\sim} g(x_i|\theta)$.
- Will see a proof in a special case but first two very useful probability results

Slutsky's theorem

- Recall convergence in probability: $Y_n \overset{p}{\to} Y$ if for every $\epsilon > 0$, $P(\|Y_n - Y\| > \epsilon) \to 0$.
- Suppose $T_n \overset{d}{\to} Z \in \mathbb{R}^d$ and suppose B_n, $n = 1, 2, \cdots$ are $q \times d$ random matrices such that $B_n \overset{p}{\to} B$ a fixed matrix. Then $B_n T_n \overset{d}{\to} BZ$.

A use of Slutsky's theorem

- Normal approximation to Poisson: if $X_1, \cdots, X_n \overset{\text{IID}}{\sim} Poi(\mu)$ then

$$Z_n = \frac{\bar{X} - \mu}{\sqrt{\mu/n}} \overset{d}{\to} N(0,1)$$

- By WLLN $\bar{X} \overset{p}{\to} \mu$
- Slutksy's theorem: $T_n = \frac{\sqrt{\mu}}{\sqrt{\bar{X}}} Z_n = \frac{\bar{X} - \mu}{\sqrt{\bar{X}/n}} \overset{d}{\to} N(0,1)$.

The Delta theorem

Theorem. *Suppose* $\sqrt{n}(T_n - \mu) \overset{d}{\to} N_d(0, \Sigma)$. *If* $g(t) : \mathbb{R}^d \to \mathbb{R}^q$ *has a continuous first derivative* $\dot{g}(t)$ *(a* $d \times q$ *matrix) then*

$$\sqrt{n}(g(T_n) - g(\mu)) \overset{d}{\to} N_q(0, \dot{g}(\mu)^T \Sigma \dot{g}(\mu)).$$

Proof. Mean value theorem $\implies g(T_n) = g(\mu) + \dot{g}(S_n)^T(T_n - \mu)$ for some S_n between T_n and μ. Rearranging,

$$\sqrt{n}(g(T_n) - g(\mu)) = \sqrt{n}\dot{g}(S_n)^T(T_n - \mu)$$

Because $\sqrt{n}(T_n - \mu) \overset{d}{\to} N_d(0,1)$ implies $T_n \overset{p}{\to} \mu$ (why?), we have $S_n \overset{p}{\to} \mu$. By continuity of $\dot{g}$, $\dot{g}(S_n) \overset{p}{\to} \dot{g}(\mu)$. The rest follows from Slutsky's theorem.

A special case: Canonical exponential family

- Model: $X_1, \cdots, X_n \overset{\text{IID}}{\sim} h(x_i)e^{\xi^T T(x_i) - A(\xi)}$
- Parameter space
  $\xi \in \mathcal{E} = \{\xi : A(\xi) = \log \int h(x_i)e^{\xi^T T(x_i)} dx_i < \infty\}$
- Assumptions:
  1. $\mathcal{E}$ is open
  2. $a^T T(X_i)$ is not a constant for any a

Some properties

- Call $T_i = T(X_i)$
- $\mathbb{E}_{[X_i|\xi]} T_i = \dot{A}(\xi)$, $\mathbb{V}\text{ar}_{[X_i|\xi]} T_i = \ddot{A}(\xi)$
- Assumption 2 implies $\ddot{A}$ is positive definite, so $A(\xi)$ is strictly convex over $\mathcal{E}$.
- This means $\dot{A}(\xi)$ is one-to-one and has an inverse $g(t)$ with a continuous derivative

MLE

- Log-likelihood function

$$\ell_x(\xi) = \text{const} + \xi^T \sum_{i=1}^{n} T(x_i) - nA(\xi)$$
$$= \text{const} + n\xi^T \bar{T}(x) - nA(\xi)$$

  with $\bar{T}(x) = \frac{1}{n} \sum_{i=1}^{n} T(x_i)$,
- So MLE solves $\dot{A}(\hat{\xi}_{\text{MLE}}(x)) = \bar{T}(x)$, i.e., $\hat{\xi}_{\text{MLE}}(x) = g(\bar{T}(x))$
- Also, $I_X = n\ddot{A}(\hat{\xi}_{\text{MLE}}(x))$.

Asymptotic normality

- Fix a θ and suppose $X_i \overset{\text{IID}}{\sim} h(x_i)e^{\xi^T T(x_i) - A(\xi)}$
- By CLT $\sqrt{n}(\bar{T}(X) - \dot{A}(\xi)) \xrightarrow{d} N_d(0, \ddot{A}(\xi))$
- Hence, by the Delta theorem

$$\sqrt{n}(\hat{\xi}_{\text{MLE}}(X) - \xi) \xrightarrow{d} N_d(0, \Sigma_\xi = \dot{g}(\dot{A}(\xi))^T \ddot{A}(\xi) \dot{g}(\dot{A}(\xi)))$$

- But, because $g(t)$ is the inverse of $\dot{A}(\xi)$ we must have

$$\dot{g}(t) = \{\ddot{A}(g(t))\}^{-1}$$

  and so $\Sigma_\xi = \{\ddot{A}(\xi)\}^{-1}$.
- Therefore, $\sqrt{n}(\hat{\xi}_{\text{MLE}}(X) - \xi) \sim AN_d(0, \{\ddot{A}(\xi)\}^{-1})$

The final piece

- The last result implies $\hat{\xi}_{\text{MLE}}(X) \xrightarrow{p} \xi$,
- So $I_X/n = \ddot{A}(\hat{\xi}_{\text{MLE}}(X)) \xrightarrow{p} \ddot{A}(\xi)$ because $\ddot{A}$ is continuous
- Therefore,

$$\sqrt{n}(\hat{\xi}_{\text{MLE}}(X) - \xi) \sim AN_d(0, nI_X^{-1})$$

  by Slutksy's theorem
- Rearrange to get $\hat{\xi}_{\text{MLE}}(X) \sim AN_d(\xi, I_X^{-1})$