Feynman Liang
STAT 210A
Lecture #20
11/1/2016

1 Review

$X_n \xrightarrow{p} c$ if $\forall \varepsilon > 0 : P(|X_n - c| > \varepsilon) \to 0$.
$X_n \Rightarrow X$ if $\forall x$ where F_X cts $F_{X_n}(x) \to F(x)$.

2 Maximum likelihood estimation

For a generic dominated family $\mathcal{P} = \{p_\theta : \theta \in \Theta\}$ we can define a "natural" estimate for θ

Definition 2.1. The *maximum likelihood estimate* of θ is

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} p_\theta(x) = \arg\max_{\theta \in \Theta} \underbrace{l(\theta; x)}_{\text{log-likelihood}} \tag{2.1}$$

The MLE of $g(\theta)$ is $g(\hat{\theta}_{MLE})$

2.1 MLE in exponential families

For exponential famililities $p_\eta(x) = h(x)\exp\{\eta' T(x) - A(\eta)\}$, the log-likelihood

$$l(\eta; x) = \log h(x) + \eta' T(x) - A(\eta) \tag{2.2}$$

is strictly convex (because $\nabla_\eta^2 A(\eta) = \text{Var}(T) > 0$ TODO: Is this true?).
For MLE $\hat{\eta}$

$$0 = \nabla_\eta l(\hat{\eta}; x) = T(x) - \nabla_\eta A(\hat{\eta}) = T(x) - \mathbb{E}_{\hat{\eta}} T(x) \tag{2.3}$$

$$\implies T(x) = \mathbb{E}_{\hat{\eta}} T(x) \tag{2.4}$$

so MLE in exponential family makes model moments match empirical moments (moment matching).

Example 2.2. Consider $X \sim \text{Pois}(\theta)$, so $p_\theta(x) = \frac{\theta^x p^{-\theta}}{x!}$.
$\mathbb{E}_\theta X = \theta$ so $\hat{\theta} = X$.
The natural parameter $\eta = \log\theta \implies \hat{\eta} = \log\hat{\theta}$.
If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Pois}(\theta)$, then $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i$

Let $\mu(\eta) = \mathbb{E}_\eta T(x) = \nabla_\eta A(\eta)$. Then

$$\hat{\eta} = \mu^{-1}(T(x)) \tag{2.5}$$

For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} p_\eta$

$$l(\eta; x) = \eta \sum_{i=1}^{n} T(x_i) - nA(\eta) + \sum_{i=1}^{n} \log h(x) \tag{2.6}$$

Applying first-order necessary conditions shows that $\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} T(x_i)$ and therefore $\hat{\eta} = \mu^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} T(x_i)\right)$

For 1-dimensional exponential families, asymptotically

$$\sqrt{n}(\hat{\mu} - \mu(\eta)) \Rightarrow N(0, A''(\eta)) \tag{2.7}$$

From the delta method with $\psi = \mu^{-1}$

$$\sqrt{n}(\hat{\eta} - \eta) \Rightarrow N(0, A''(\eta)\psi'(\mu(\eta))^2) \tag{2.8}$$

But $\psi'(\mu(\eta)) = \frac{1}{\mu'(\eta)} = \frac{1}{A''(\eta)}$, so

$$\sqrt{n}(\hat{\eta} - \eta) \Rightarrow N(0, A''(\eta)^{-1}) = N(0, \underbrace{J(\eta)^{-1}}_{\text{Fisher info}}) \tag{2.9}$$

And asymptotically $\text{Var}(\hat{\eta}) \approx (nJ_1(\eta))^{-1} = J(\eta)^{-1}$ which achieves the CRLB.

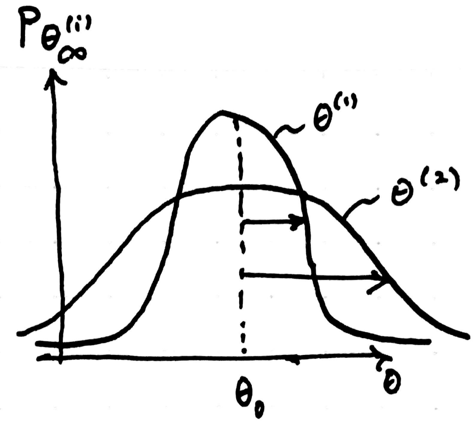
Example 2.3. $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Pois}(\theta)$.
$\hat{\theta} = \bar{X}_n$, $\hat{\eta} = \log \bar{X}_n$.
$\sqrt{n}(\hat{\mu}(X) - \eta) \Rightarrow N(0, \underbrace{e^{-\eta}}_{\text{asymptotic variance}})$
But for finite sample $P(\bar{X}_n = 0) = P(X_1 = 0)^n = e^{-\theta n} > 0$.
So $P(\bar{X}_n = 0) = P(\hat{\eta} = -\infty) > 0 \implies \mathbb{E}[\hat{\eta}] = -\infty$.

3 Asymptotic relative efficiency

Definition 3.1. Suppose $\hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}$ are asymptotically Normal with $\sqrt{n}(\hat{\theta}_n^{(i)} - \theta) \Rightarrow N(0, \sigma_i^2)$. Then the *asymptotic relative efficiency (ARE)* of $\hat{\theta}_n^{(2)}$ wrt $\hat{\theta}_n^{(1)}$ is

$$\sigma_1^2/\sigma_2^2 \tag{3.1}$$

and we say that $\hat{\theta}_n^{(2)}$ is $\frac{\sigma_1^2}{\sigma_2^2}$% as efficient as $\hat{\theta}_n^{(1)}$



Interpretation: If $\sigma_1^2/\sigma_2^2 = \gamma < 1$, then $\hat{\theta}^{(2)}(X_1, \ldots, X_n)$ and $\hat{\theta}^{(1)}(X_1, \ldots, X_{\lfloor n\gamma \rfloor})$ have the same asymptotic distribution, so using $\hat{\theta}^{(2)}$ is like throwing away data.

Example 3.2 (Sample mean vs sample median)**.** Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x-\theta)$ with f symmetric about 0, $f(0) > 0$.

Keener 8.4 shows (sketch: $P_\theta(\tilde{X}_n \leq x) = \#\{X_i \leq x\} \sim \text{Binom}(n, F_\theta(x))$) that the sample median $\tilde{x}_n$ has

$$\sqrt{n}(\tilde{x}_n - \theta) \Rightarrow N(0, (4f(0)^2)^{-1}) \tag{3.2}$$

and the sample mean

$$\sqrt{n}(\bar{x}_n - \theta) \Rightarrow N(0, \text{Var}(x_i)) \tag{3.3}$$

(Laplace) Say $f(x) = \frac{1}{2}e^{-|x|}$. Then

$$\frac{\sigma^2(\tilde{X}_n)}{\sigma^2(\bar{X}_n)} = \frac{\frac{1}{4(1/2)^2}}{2} = 0.5 \tag{3.4}$$

(Cauchy) For Cauchy distribution, sample mean $\bar{X}$ doesn't converge but sample median still converges to Normal density.

Example 3.3 (Log-Normal)**.** We have X_i such that

$$Z_i = \log X_i \sim N(\mu, \sigma^2) \tag{3.5}$$

σ^2 known, want to estimate $\mu = g(\theta) = \mathbb{E}_\theta[X_i] = e^{\theta+\sigma^2/2}$

Sample Mean:

$$\sqrt{n}(\bar{X}_n - \mu) \Rightarrow N(0, \text{Var}(X_i)) = N(0, (e^{\sigma^2} - 1)e^{2\mu-\sigma^2}) \tag{3.6}$$

MLE:

$$\hat{\theta} = \bar{Z}_n \tag{3.7}$$

$$\implies \hat{\mu} = g(\hat{\theta}) = e^{\bar{Z}_n+\sigma^2/2} \tag{3.8}$$

$$\frac{d}{dz}e^{z+\sigma^2/2} = e^{z+\sigma^2/2} \tag{3.9}$$

So asymptotically

$$\sqrt{n}(\hat{\mu}_n - \mu) \Rightarrow N(0, \sigma^2 e^{2\mu+\sigma^2}) \tag{3.10}$$

$$\implies \text{ARE} = \frac{\sigma^2}{e^{\sigma^2} - 1} < 1 \tag{3.11}$$

σ	1	3	10
ARE	0.58	0.0011	3.7×10^{-42}

In general: $\hat{\mu} \xrightarrow{\text{P}} e^{\mathbb{E}[\log x_i] + \frac{\sigma^2}{2}}$