STAT 870 Lecture 7

Consistency of MLE

Cauchy problem: for vector $(x_1, \ldots, x_n)$ we define $h_n(x_1, \ldots, x_n)$ to be that root of

$$\ell'(\theta|x_1, \ldots, x_n) = \sum \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2} = 0$$

which is closest to $g_n(x_1, \ldots, x_n)$ (here g_n is the Borel function used in defining the median above). If $g_n(x_1, \ldots, x_n)$ is exactly midway between two roots which are tied for closest we define h_n to be the root smaller than g_n.

It is possible to prove that this defines a Borel function from $\mathbb{R}^n$ to $\mathbb{R}$. Now define

$$\hat\theta_n = h_n(X_1, \ldots, X_n)$$

I claim that if $X_1, X_2, \ldots$ are iid Cauchy(0) then

$$\hat\theta_n \to 0 \tag{1}$$

almost surely.

To prove this fix $\epsilon > 0$ we prove that $P(C_\epsilon) = 1$ where

$$C_\epsilon = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|\hat\theta_n| \le \epsilon\}$$

Notation:

- Suppress dependence of ℓ on the data.
- Use superscript $'$ to denote differentiation wrt θ.
- Components of ℓ etc:

$$\begin{aligned}
L_i(\theta) &= -\log(f(X_i, \theta)) + \log(f(X_i, \theta_o)) \\
U_i(\theta) &= L_i'(\theta) \\
V_i(\theta(&= L_i''(\theta) \\
W_i(\theta) &= L_i'''(\theta)
\end{aligned}$$

Find event D_ϵ inside C_ϵ with $P(D_\epsilon) = 1$. To define this new event note that if

1. ℓ' has a unique root over $[-3\epsilon, 3\epsilon]$ and
2. that root is actually in $[-\epsilon, \epsilon]$ and
3. $|\tilde{\theta}_n| \le \epsilon$

then the root of ℓ' closest to $\tilde{\theta}_n$ is actually the root in points 1 and 2 and so

$$|\hat{\theta}_n| \le \epsilon$$

Define $D_\epsilon^{(1)}$ to be the event that there is an N such that for all $n \ge N$ and all $|\theta| \le 2\epsilon$ we have

$$\ell''(\theta|X_1, \ldots, X_n) < 0$$

Define $D_\epsilon^{(2)}$ to be the event that there is an N such that for all $n \ge N$

$$\ell''(\epsilon|X_1, \ldots, X_n) < 0$$

and

$$\ell''(-\epsilon|X_1, \ldots, X_n) > 0$$

Finally define $D_\epsilon^{(3)}$ to be the event that there is an N such that for all $n \ge N$

$$|\tilde{\theta}_n| \le \epsilon$$

Already shown: $P(D_\epsilon^{(3)}) = 1$.
Next show $P(D_\epsilon^{(2)}) = 1$. Note that

$$U_k(\epsilon) = \frac{2(X_k - \epsilon)}{1 + (X_k - \epsilon)^2}$$

and

$$U_k(-\epsilon) = \frac{2(X_k + \epsilon)}{1 + (X_k + \epsilon)^2}$$

Then

$$\frac{1}{n}\ell''(\epsilon|X_1, \ldots, X_n) = \overline{U_n(\epsilon)}$$

and

$$\frac{1}{n}\ell''(-\epsilon|X_1, \ldots, X_n) = \overline{U_n(-\epsilon)}$$

Thus $D_\epsilon^{(2)}$ is the event that there is an N such that for all $n \geq N$

$$\overline{U_n(\epsilon)} < 0 \quad \text{and} \quad \overline{U_n(-\epsilon)} > 0$$

$\overline{U_n(-\epsilon)}$, $\overline{U_n(\epsilon)}$ averages of iid variates. Apply SLLN and show

$$\begin{aligned} \mathrm{E}(U_k(\epsilon)) &< 0 \\ \mathrm{E}(U_k(-\epsilon)) &> 0\,. \end{aligned}$$

In fact

$$\begin{aligned} \mathrm{E}(U_k(\epsilon)) &= \frac{-2\pi\epsilon}{\epsilon^2+4} < 0 \\ \mathrm{E}(U_k(-\epsilon)) &= \frac{2\pi\epsilon}{\epsilon^2+4} > 0\,. \end{aligned}$$

Defect: argument not easy to generalize; uses exact computation of moment.

More general tactic: use Jensen's inequality.

If ℓ is smaller at $-\epsilon$ and at ϵ than it is at 0 then there must be a critical point in $[-\epsilon, \epsilon]$, that is, a root of ℓ'.

Define

$$L_i(\theta) = \log(1+X_i^2) - \log(1+(X_i-\theta)^2)\,.$$

Note that $\ell(\theta) - \ell(0) = \sum L_i(\theta)$.

$D_\epsilon^{(4)}$ is the event $\exists N$ such that $\forall n \geq N$

$$\{\sum L_i(\epsilon) < 0, \sum L_i(-\epsilon) < 0\}\,.$$

Define

$$\mu(\epsilon) = \mathrm{E}(L_i(\epsilon))$$

SLLN shows $P(D_\epsilon^{(5)}) = 1$ where $D_\epsilon^{(5)}$ is the event

$$\{\sum L_i(\epsilon)/n \to \mu(\epsilon), \sum L_i(-\epsilon)/n \to \mu(-\epsilon)\}$$

Claim: for all $\epsilon \neq 0$ $\mu(\epsilon) < 0$.

If so then

$$D_\epsilon^{(5)} \subset D_\epsilon^{(4)}$$

and so $P(D_\epsilon^{(4)}) = 1$.

To prove the claim we apply Jensen's inequality:

Proposition 1 (Jensen) *Suppose Y is a random variable and ϕ is a function which is convex on an interval (a,b) with $P(a < Y < b) = 1$ Assume $\mathrm{E}(|Y|) < \infty$. Then*

$$\phi(\mathrm{E}(Y)) \le \mathrm{E}(phi(Y))$$

If ϕ is strictly convex then the inequality is strict unless $\mathrm{Var}(Y) = 0$.

Jargon: ϕ is convex if for each x, y and $\lambda \in (0,1)$

$$\phi(\lambda x + (1-\lambda)y) \le \lambda\phi(x) + (1-\lambda)\phi(y)$$

We call ϕ strictly convex if the inequality is strict.

If ϕ is twice differentiable and $\phi'' \ge 0$ then ϕ is convex; a strict inequality shows ϕ is strictly convex.

Apply Jensen's inequality with $\phi(x) = -\log(x)$ to $Y = g(X)/f(X)$ where X has density f and g is some other density. Then

$$\mathrm{E}\{-\log(Y)\} > -\log\{\mathrm{E}(Y)\}$$

But the latter is

$$\log\left\{\int \frac{g(x)}{f(x)} f(x)dx\right\} = \log(1) = 0$$

Technically: interval (a,b) is $(0,\infty)$. The assumption

$$P(0 < Y < \infty) = 1$$

deserves some discussion. If $f(x) = 0$ for some places where $g(x)$ is not 0 then

$$\mathrm{E}\left\{\frac{g(X)}{f(X)}\right\} = \int g(x) 1(f(X) > 0) dx$$

which might be less than 1. This just makes the inequality stronger, however.

The other technical detail is that $g(x)$ might be 0 some places where $f(x)$ is not 0. This might mean $P(Y = 0) > 0$. On the event $Y = 0$ we will agree to take $-\log(Y) = \infty$ and conclude

$$\mathrm{E}\{-log(Y)\} = \infty$$

In any case we find

$$\mathrm{E}\{-log(Y)\} > 0$$

or

$$\mathrm{E}[\log\{g(X)\} - \log\{f(X)\}] < 0\,.$$

Applied to our Cauchy problem we have shown $\mu(\theta) < 0$ for all $\theta \neq 0$. Hence $P(D_\epsilon^{(4)}) = 1$.

Finally we consider $D_\epsilon^{(1)}$. Up to now we have been able to make do with an arbitrary ϵ. In this case, however, the result holds only for small $\epsilon > 0$. First consider

$$\frac{1}{n}\ell''(0|X_1, \ldots, X_n)$$

According to the strong law of large numbers this converges almost surely to

$$\mathrm{E}\left\{\frac{2(1-X^2)}{(1+X^2)^2}\right\} = -\frac{1}{2} < 0$$

Now you can check that

$$\left|\frac{1}{n}\ell'''(\theta|X_1, \ldots, X_n)\right| < 4\,.$$

(In fact each term in ℓ''' may be shown to be bounded by $3/2 + \sqrt{2}$.) As a result

$$\frac{1}{n}\left|\ell''(\theta|X_1, \ldots, X_n) - \ell''(0|X_1, \ldots, X_n)\right| \leq 4|\theta|$$

Pick $\epsilon > 0$ so that $4\epsilon < \pi/2$. If B is the event

$$\frac{1}{n}\ell''(0|X_1, \ldots, X_n) \to -\frac{1}{2}$$

and ω is in B then there is an N such that for $n \geq N$ we have

$$\frac{1}{n}\ell''(\theta|X_1, \ldots, X_n) < 0$$

for all $|\theta| < \epsilon$.

This proves that for all $3\epsilon < \pi/8$

$$P(D^{(1)}(\epsilon)) = 1\,.$$

We have now shown that for $\epsilon < \pi/24$

$$P(D^{(1)}(\epsilon) \cap D_\epsilon^{(2)} \cap D_\epsilon^{(3)}) = 1$$

For ω in this event we have that there is an N such that

$$|\hat{\theta}_n| \leq \epsilon$$

for all $n \geq N$. This establishes the result.

General Case

Consider parametric family

$$\{f(x|\theta); a < \theta < b\}$$

Let θ_o be true value of θ

Let A_ϵ be the event: $\exists N$ such that $\forall n \geq N$ ℓ has a local maximum on the interval $[\theta_o - \epsilon, \theta_o + \epsilon]$.

We have proved quite generally that

$$P(A_\epsilon) = 1 \tag{2}$$

Add assumption

$$\ell \text{ has a continuous derivative} \tag{\textbf{A}}$$

Let B_ϵ be the event that there is an N such that for all $n \geq N$ there is a critical point of ℓ in $(\theta_o - \epsilon, \theta_o + \epsilon)$ which is a local maximum of ℓ we have proved

$$P(B_\epsilon) = 1 \tag{3}$$

The event $B = \cap_\epsilon B_\epsilon$ then has probability 1. On this event there is a sequence of roots of the likelihood equations which is consistent.

Remaining problem: prove, under general conditions, that there is probably only one root near θ_o.

Consider event that ℓ' is monotone on $[\theta_o - \epsilon, \theta_o + \epsilon]$. Previous proof based on showing that next derivative was negative at θ_o and did not change much over a small enough interval.

Behaviour at θ_o is essentially the behaviour of

$$\frac{1}{n}\sum V_i(\theta_o)$$

which converges almost surely to

$$\mathrm{E}(L_1''(\theta_o))$$

I claim this is negative for regular families.

Begin with

$$1 = \int f(x,\theta)dx$$

Differentiating with respect to θ gives

$$\begin{aligned} 0 &= \frac{d}{d\theta}\int f(x,\theta)dx \\ &= \lim_{\epsilon\to 0}\int \frac{f(x,\theta+\epsilon)-f(x,\theta)}{\epsilon}dx \end{aligned}$$

In order to take the limit inside the integral sign we must prove that for any sequence $\epsilon_n \to 0$

$$\lim \int \frac{f(x,\theta+\epsilon_n)-f(x,\theta)}{\epsilon_n}dx$$

$$= \int \frac{\partial}{\partial\theta}f(x,\theta)dx$$

Normally: apply dominated convergence theorem. If f is continuously differentiable wrt θ then difference quotient is exactly

$$\frac{\partial}{\partial\theta}f(x,\theta_n^*)$$

where θ_n^* depends on both n and x.

One tactic: compute

$$M(x,\epsilon) = \sup\{|\frac{\partial}{\partial\theta}f(x,\theta)|; |\theta-\theta_o| \le \epsilon\}$$

and show that

$$\int M(x,\epsilon)dx < \infty$$

to apply dominated convergence.

Assuming dominated convergence theorem applies:

$$\begin{aligned} 0 &= \int \frac{\partial}{\partial\theta}f(x,\theta)dx \\ &= \int \frac{\partial \log f(x,\theta)}{\partial\theta}f(x,\theta)dx \\ &= \mathrm{E}(U_k(\theta_o)) \end{aligned}$$

Differentiating again and again passing limits through integrals gives

$$\mathrm{E}(U_k^2(\theta_o)) = -\mathrm{E}(V_k(\theta_o))$$

This shows that

$$\frac{1}{n}\ell''(\theta_o) \to \mathrm{E}(V_k(\theta_o)) < 0$$

almost surely.

Next we consider

$$\frac{1}{n}\ell''(\theta) - \ell''(\theta_o)$$

For a three times continuously differentiable ℓ there is a θ_n^* (which is random but between θ_o and θ) such that this difference is

$$\frac{\theta-\theta_o}{n}\sum W_i(\theta_n^*)$$

Define

$$M_i(\epsilon) = \sup\{|W_i(\theta)| : |\theta - \theta_o| \le \epsilon\}$$

The M_i are iid. If for some $\epsilon > 0$ the M_i are integrable then SLLN shows

$$\limsup\left|\frac{1}{n}\left\{\ell''(\theta) - \ell''(\theta_o)\right\}\right| \le \epsilon \mathrm{E}(M_1(\epsilon))$$

almost surely. RHS of inequality can be made arbitrarily small by choosing ϵ small enough.

Pick ϵ so small that bound strictly smaller than

$$I(\theta_o) \equiv -\mathrm{E}(V_k(\theta_o)).$$

Then

$$P(E_\epsilon) = 1$$

where E_ϵ is event $\exists N$ st $\forall n \ge N$ we have ℓ' is monotone decreasing on $[\theta_o - \epsilon, \theta_o + \epsilon]$.