Lecture 25. MLE norm, Emp Proc, Final practice

Monday, December 11, 2017 9:20 AM

Final Date and Time: Thursday 1:30 - 3:30, M4318

Exam is cumulative as far as knowledge is concerned. However, problems will come from topics discussed in lectures 12 - 25. These include

- Convergence
- Laws of large numbers
- Limit theorems
- M / Z Estimation
- MLE

Exam is closed book

Two 2-sided help-sheets are allowed

Sample practice exam:

1. State and prove the Borel-Cantelli TH

2. State and prove the Cramer-Wold device

3. Let $\{X_i\}$ be i.i.d. with zero mean and unit variance
   Find the asymptotic distribution of

$$Y_n = \frac{X_1 + \dots + X_n}{\sqrt{X_1^2 + \dots + X_n^2}}$$

4. Let X follow a Power-Series distribution with pmf $f(x;\theta) = a(x)\,\theta^x / f(\theta)$.

where $f(\theta) = \sum_{x=1}^{\infty} a(x)\theta^x$ converges for $0 < \theta \leq R$

for some R, $a > 0$.

(a) Prove that MLE of θ, $\hat{\theta}_n$ satisfies the equation

$$\mathbb{E} X = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(b) Provide a weak convergence statement

(c) Use the result to show asymptotic normality and consistency of $\hat{p}_n$ of a negatively Binomial distribution $\overline{Bi}\ (r, p)$

Note: $\overline{Bi}$ has expectation $\frac{rp}{1-p}$

variance $\frac{rp}{(1-p)^2}$

$$P(X = k) = C_{r+k-1}^{k}\, p^k (1-p)^r$$

(5) $\rceil \{X_i\}$ is a sequence of r.v.s

$\rceil$ There exist r.v. X and a sequence of interers $n_i > 0$:

$$\max_{n_{k-1} < m \leq n_k} \left| X_m - X_{n_{k-1}} \right| \xrightarrow[k \to \infty]{} 0 \quad \text{a.s.}$$

and $X_{n_k} \to X$ a.s.

Prove: $X_n \xrightarrow[\text{a.s.}]{} X$

MLEs

$$\ell_n = \sum_{i=1}^{n} \log \underbrace{f_\theta(X_i)}_{\text{pdf}}$$

$$\max_\theta \ell_n \Rightarrow \hat{\theta}_n \quad \text{MLE}$$

$$M_n(\theta) = \mathbb{P}_n \log \frac{f_\theta}{f_{\theta^*}}$$

Model identifiability $\Rightarrow$ θ^* true parameter value

unique solution to $\max M_\theta$

$M_n \xrightarrow[p]{} M_\theta$ uniformly ; $M_\theta = \mathbb{P} m_\theta$

Asymptotic variance of MLEs

From M-estimation theory

Asymptotic variance is

$$\frac{\mathbb{P}(m'_{\theta^*})^2}{(\mathbb{P}(m''_{\theta^*}))^2}$$

$\psi_\theta = m'_\theta$ (MLE as a Z-estimation)

Normalization condition

$$\mathbb{P}(1) = 1 = \int f_\theta \, dx$$

$$\int f'_\theta \, dx = \int f''_\theta \, dx = 0$$

$$m_\theta = \log f_\theta$$

$$m'_\theta = \frac{f'_\theta}{f_\theta}$$

$$m''_\theta = \frac{f''_\theta}{f_\theta} - \left(\frac{f'_\theta}{f_\theta}\right)^2$$

$$\mathbb{P}\, m'_{\theta^*} = \int \frac{f'_{\theta^*}}{f_{\theta^*}} \cdot f_{\theta^*} \, dx = 0$$

Expected value of the score is zero at the true param. value

$$\mathbb{P}\left(m''_{\theta^*}\right) = \underbrace{\int f''_{\theta^*} \, dx}_{=0 \text{ by deriv. of normalization condition}} - \mathbb{P}\left(m'_{\theta^*}\right)^2$$

$$\underbrace{\mathbb{P}\left(m'_{\theta^*}\right)^2}_{\text{variance of the true score}} = -\underbrace{\mathbb{P}\left(m''_{\theta^*}\right)}_{\text{negative Hessian}} = I_{\theta^*}$$

$\Rightarrow$

Asymptotic variance $\dfrac{I_{\theta^*}}{\left[I_{\theta^*}\right]^2} = I_{\theta^*}^{-1}$

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \rightsquigarrow \mathcal{N}\left(0, I_{\theta^*}^{-1}\right)$$

Under reg. conditions Observed information matrix is a consisten estimate of I_{θ^*}

End of material that goes into the final exam

Overview of empirical process theory

Empirical Distribution function

$$\mathbb{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \le t)$$

$n\mathbb{F}_n \sim \mathrm{Bin}(n F(t))$ $\Rightarrow$ unbiased

↳ CDF

LLN $\Rightarrow$ consistency $\mathbb{F}_n(t) \xrightarrow[p]{} F(t)$, $\forall t$

CLT $\Rightarrow$ $\sqrt{n}(\mathbb{F}_n - F) \rightsquigarrow \mathcal{N}(0, F(t)(1-F(t)))$

Random functions of t are called Stochastic Processes

Glivenko - Cantelli Theorem

Uniform Law of Large Numbers for $\mathbb{F}_n$

$\|\cdot\|_\infty$ Norm

$$\|\mathbb{F}_n - F\|_\infty := \underbrace{\sup_t |\mathbb{F}_n(t) - F(t)|}_{\text{Kolmogorov-Smirnov statistic}}$$

TH Glivenko-Cantelli:

Let X_i be i.i.d. r.v.s

$\Rightarrow$

$\|\mathbb{F}_n - F\|_\infty \xrightarrow[a.s.]{} 0$

↳ uniform convergence statement is a functional norm

Proof:

Take $\varepsilon > 0$

Let t_i are points where F jumps by more than ε

Then, for $t_{i-1} \le t < t_i$, we have

$$F_n(t_{i-1}) - F(t_{i-1}) - \varepsilon \le F_n(t) - F(t) \le F_n(t_i-) - F(t_i-) + \varepsilon$$

F can have at most a countable set of jumps $< \infty$
b/c F is monotonic and bounded
b/c jumps @ t_i are more than ε, and F is bounded $\Rightarrow$ at most finite # of t_is
$\Rightarrow$ any pointwise result over a finite number of points is a uniform result over those points

$\Downarrow$

b/c $F_n(t_{i-1}) - F(t_{i-1}) \xrightarrow{a.s.} 0$
$F_n(t_i-) - F(t_i-) \xrightarrow{a.s.} 0$

$\Rightarrow \sup_t |F_n - F| < \varepsilon$

$\Rightarrow P(\|F_n - F\|_\infty \le \varepsilon) \to 1$ $\square$

$(\mathbb{E}_n - \mathbb{E}) f$

DF $\sqrt{n}(\mathbb{P}_n - \mathbb{P}) f = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n f(X_i) - \int f(x)\, dF(x)\right)$

is called an empirical process evaluated at f

TH Donsker
$\mathbb{G}_n := \sqrt{n}(\mathbb{F}_n - F) \rightsquigarrow \mathbb{G}_F$

$\mathbb{G}_F$ is a Brownian Bridge with covariance
$\mathrm{Cov}(\mathbb{G}_F(x), \mathbb{G}_F(y)) = F(\min(x,y)) - F(x)F(y)$

DF Brownian Motion and Bridge

W(t) is ↑ or Wiener Process
when

$W(0) = 0$

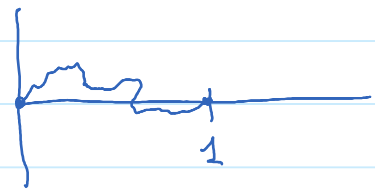
independent increments $dW(t)$

$\Delta_s^t W$ depends only on $t-s$

↓ increment on $[s,t]$

$W(t) \sim \mathbb{N}(0,t)$

Bridge is $W(t)$ given that $W(1) = 0$



↓

$W(t) - W(1)$

(DF) Class of $\overset{\wedge}{\text{measur.}}$ functions f is called Glivenko-Cantelli (GC)

$\mathcal{G}$

if $\sup_{f \in \mathcal{GC}} |(\mathbb{P}_n - \mathbb{P})f| \xrightarrow[a.s.]{} 0$

(DF) A class $\mathbb{D}_P$ is called Donsker if (P-Donsker)

$\mathbb{G}_n f \rightsquigarrow \mathbb{G}_P$ where

$\mathbb{G}_P$ is zero-mean Gaussian

w. cov.

$$\mathbb{E}(\mathbb{G}_p f\ \mathbb{G}_p g) = \mathbb{P}(fg) - \mathbb{P}f \cdot \mathbb{P}g$$

Donsker and Glivenko - Cantelli will be fullfilled when entropy of $\mathcal{C}$ is not growing too fast as $\varepsilon \to 0$ (see below)

DF Bracketing number $N_{[\,]}$ is the min number of balls in L_p norm, centered around functions $f \in \mathcal{C}$, with a radius of $\varepsilon > 0$ are needed to cover all of $\mathcal{C}$.

DF Entropy $= \log N_{[\,]}$

DF Bracketing integral $J_{[\,]} = \int_0^\delta \sqrt{\log N_{[\,]}}\, d\varepsilon$

TH $\exists\, N_{[\,]} < \infty, \ \forall \varepsilon > 0 \Rightarrow \mathcal{C}$ is $\mathbb{GC}$

TH $\exists\, J_{[\,]} < \infty \Rightarrow \mathcal{C}$ is $\mathbb{D}_P$

Note: $N_{[\,]} \to \infty$ slower than $\frac{1}{\varepsilon^2} \Rightarrow \mathbb{D}_P$