# MATH 541a Homework 7

Qilin Ye

April 22, 2022

**Problem 1**

*Solution.* We first proof a lemma: "$X_n \to X$ in probability if and only if for every subsequence $X_{n_k}$ there exists a further subseqeuence converging almost surely to $X$."

To prove $\Rightarrow$, let $X_{n_k}$ be given. By assumption it converges in probability. For each $k \in \mathbb{N}$, there exists $n_k$ such that

$$\mathbb{P}(|X_{n_k} - X| \geqslant 1/k) \leqslant 2^{-k}.$$

Therefore $\sum_{k=1}^{\infty} \mathbb{P}(|X_{n_k} - X| \geqslant 1/k) < \infty$ so

$$\mathbb{P}(|X_{n_k} - X| \geqslant 1/k \text{ i.o.}) = 0 \implies \mathbb{P}(|X_{n_k} - X| < 1/k \text{ for sufficiently large } k) = 1.$$

This shows $X_{n_k}$ converges to $X$ a.e.

Conversely, it is clear that convergence a.e. implies convergence in probability, so every subsequence of $X_n$ has a further subsequence converging to $X$ in probability. This shows $X_n \to X$ in probability, for if not, there exists $\epsilon > 0$ and $\delta > 0$ and a sequence $X_{n_k}$ with $\mathbb{P}(|X_{n_k} - X| \geqslant \epsilon) \geqslant \delta$, and this sequence cannot have convergent subsequence.

Moving back to the original question, the arguments converge in probability individually, so every subsequence of every argument has a further subsequence converging a.e. Continuity of $f$ implies that every subsequence of the image has a further subsequence converging a.e., which by the iff condition implies that $h(M_{1,n}, \ldots, M_{j,n})$ also converges in probability.

**Problem 2**

*Solution.* (1) The CDF of $X_1$ is

$$\mathbb{P}(X_1 \leqslant x) = \mathbb{P}(\gamma + e^Z \leqslant x) = \mathbb{P}(Z \leqslant \log(x - \gamma)) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\log(x-\gamma)} \exp(-(-t-\mu)^2/2\sigma^2)\, \mathrm{d}t$$

so differentiating gives

$$f_{X_1}(x) = \frac{1}{x-\gamma} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x-\gamma)-\mu)^2}{2\sigma^2}\right).$$

(2) The log-likelihood is

$$-\sum_{i=1}^{n} \log(X_i - \gamma) - n\log(\sigma) - \frac{n}{2}\log(2\pi) - \sum_{i=1}^{n} \frac{(\log(X_i-\gamma)-\mu)^2}{2\sigma^2}.$$

With $\gamma$ known, the $\mu$ and $\sigma$ partials are

$$\sum_{i=1}^{n} \frac{\log(X_i-\gamma)-\mu}{\sigma^2} \qquad \text{and} \qquad -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(\log(X_i-\gamma)-\mu)^2}{\sigma^3}$$

Setting them to zero we have

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\log(X_i-\gamma) \qquad \text{and } \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\log(X_i-\gamma)-\mu)^2.$$

(3) As $\gamma \to X_{(1)}$, we have $M \to -\infty$ and $T \to \infty$. In this case $\Theta$ is not compact.

**Problem 3**

*Proof.* The function $f(w) := \sum_{i=1}^{n}(X_i - \langle x^{(i)}, w\rangle)^2 + c\sum_{i=1}^{n}|w_i| = \|X - Aw\|^2 + c\|w\|_1$, a sum of convex and linear functions and is therefore convex. The derivative is $A^T(Aw-b)+cw$. Therefore, the function indeed attains a global minimum at $w = (A^TA + cI)^{-1}A^TX$. □

**Problem 4**

*Proof.* This is simply brute force computation.

$$\begin{aligned}
\mathbb{E}Z_n &= \frac{n^2}{2}\mathbb{E}Y_n - \frac{(n-1)^2}{n}\sum_{i=1}^{n}\mathbb{E}t_{n-1}(\dots) + \frac{(n-2)^2}{n(n-1)}\sum_{i\neq j}\mathbb{E}t_{n-2}(\dots)\\
&= \frac{n^2\theta}{2} + \frac{na}{2} + \frac{b}{2} + \frac{c}{2n} + \frac{d}{2n^2} + \mathcal{O}(n^{-5})\\
&\quad - (n-1)^2\left(\theta + \frac{a}{n-1} + \frac{b}{(n-1)^2} + \frac{c}{(n-1)^3} + \frac{d}{(n-1)^4} + \mathcal{O}(n^{-5})\right)\\
&\quad + \frac{(n-2)^2}{2}\left(\theta + \frac{a}{n-2} + \frac{b}{(n-2)^2} + \frac{c}{(n-2)^3} + \frac{d}{(n-2)^4} + \mathcal{O}(n^{-5})\right)\\
&= \theta + c\left(\frac{1}{2n} - \frac{1}{n-1} + \frac{1}{2(n-2)}\right) + d\left(\frac{1}{2n^2} - \frac{1}{(n-1)^2} + \frac{1}{2(n-2)^2}\right) + \mathcal{O}(n^{-3})\\
&= \theta + \frac{c}{n^3-3n^2+2n} + d\cdot\frac{3n^2-6n+2}{(n-2)^2(n-1)^2n^2} + \mathcal{O}(n^{-3}) = \theta + \mathcal{O}(n^{-3}).
\end{aligned}$$

It is also clear from above that if the additional assumptions are met then $Z_n$ is unbiased. □

## Fall 2011 Qual #1

*Solution.* (1) Using linearity,

$$\mathbb{E}\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}Y_i = \mathbb{E}Y_1 = \int_0^\theta y\cdot 2y/\theta^2\,\mathrm{d}y = \frac{2\theta}{3}.$$

Ont he other hand,

$$\mathbb{E}Y_1^2 = \int_0^\theta y^2\cdot 2y/\theta^2\,\mathrm{d}y = \frac{\theta^2}{2}.$$

Therefore

$$\operatorname{var}(\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(Y_i) = \frac{1}{n}(\mathbb{E}Y_1^2 - (\mathbb{E}Y_1)^2) = \frac{1}{n}(\theta^2/2 - 4\theta^2/9) = \theta^2/(18n).$$

## Fall 2012 Qual

**Problem 1**

(1) Let $X_1, \dots, X_n$ be i.i.d. Poisson with parameter $\lambda$. Find the MOM estimator and the MLE of $\lambda$.

(2) Is MLE unbiased? Is it efficient?

(3) Given an example of a distribution where the MOM and MLE are different.

*Solution.* (1) It is well-known that Poisson($\lambda$) has expected value $\lambda$. This gives the MOM estimator. For MLE, the likelihood is given by

$$\ell(\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum x_i} \prod_{i=1}^{n} (x_i!)^{-1}$$

so the log-likelihood is

$$-n\lambda + \sum_{i=1}^{n} x_i \log \lambda + C.$$

Taking first derivative gives $-n + \sum x_i / \lambda$ which gives the critical point $\lambda = \sum x_i / n$. Since the second derivative is negative, the sample mean is indeed the MLE.

(2) It is certainly unbiased as shown in (1). The Fisher information for Poisson is

$$I_X(\lambda) = -\mathbb{E}\left(\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} f_\lambda(X)\right) = -\mathbb{E}\left(-X/\lambda^2\right) = \frac{1}{\lambda}$$

so

$$I_{X_1,\dots,X_n}(\lambda) = \frac{n}{\lambda}.$$

Comparing this with

$$\operatorname{var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{n}{n^2} \operatorname{var}(X_1) = \frac{\lambda}{n}.$$

We see that the MLE indeed achieves the Cramér-Rao lower bound.

(3) Let $X_1, \dots, X_n$ be i.i.d. from uniform $[\theta, \theta + 1]$ where $\theta$ is unknown and the parameter to estimate. The likelihood is $1_{\theta \leqslant X_{(1)} \leqslant X_{(n)} \leqslant \theta + 1}$ so the MLE can be anything in $[X_{(n)} - 1, X_{(1)} + 1]$. On the other hand the first moment is $\theta + 0.5$ so the MOM estimator is $\sum X_i / n - 0.5$.

**Problem 2**

(1) Prove that for any collection of random varaibles $X_1, \dots, X_k$,

$$\operatorname{var}\left(\sum_{i=1}^{k} X_i\right) \leqslant k \sum_{i=1}^{k} \operatorname{var}(X_i).$$

(2) Construct an example with $\geqslant 2$ where equality holds.

*Proof.* (1)

$$\operatorname{var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i,j=1}^{k} \operatorname{cov}(X_i, X_j) \leqslant \sum_{i,j=1}^{k} \sqrt{\operatorname{var}(X_i)}\sqrt{\operatorname{var}(X_j)}$$
$$= \left(\sum_{i=1}^{k} \sqrt{\operatorname{var}(X_i)}\right)^2 = \left(\sum_{i=1}^{k} 1 \cdot \sqrt{\operatorname{var}(X_i)}\right)^2 \leqslant k \sum_{i=1}^{k} \operatorname{var}(X_i).$$

Note that both $\leqslant$ are given by Cauchy-Schwarz.

(2) For equality to hold we want both $\leqslant$ to be $=$. For the first one, we need $\operatorname{cov}(X_i, X_j) = \operatorname{var}(X_i)\operatorname{var}(X_j)$, which implies $X_i$'s need to be multiples of each other. The second $=$ then requires that the coefficients must agree with $(1, 1, \ldots, 1)$, i.e., $X_1 = X_2 = \ldots = X_k$.

□