MATH 541a Homework 4

Qilin Ye

March 10, 2022

Problem 1

Let $X_1, \ldots, X_n$ be a random sample of size n.

(a) Suppose X is a discrete random variable and we order the values X takes as $x_1 < x_2 < \ldots$. For $i \geqslant 1$ define $p_i := \mathbb{P}(X \leqslant x_i)$. Show that

$$\mathbb{P}(X_{(j)} \leqslant x_i) = \sum_{k=j}^{n} \binom{n}{k} p_i^k (1-p_i)^{n-k}.$$

(b) Let X be uniformly distributed on $[0,1]$. Show that $X_{(j)}$ is a beta distributed random variable with parameters j and $n-j+1$. Conclude that

$$\mathbb{E}X_{(j)} = \frac{j}{n+1}.$$

(c) Let $a < b$. Let U be the number of indices $1 \leqslant j \leqslant n$ such that $X_j \leqslant a$. Let V be the number of indices $1 \leqslant j \leqslant n$ such that $a < X_j < b$. Show that the vector $(U, V, n-U-V)$ is a multinomial random variable with

$$\mathbb{P}((U,V,n-U-V) = (u,v,n-u-v)) = \frac{n!}{u!v!(n-u-v)!} F_X(a)^u (F_X(b) - F_X(a))^v (1 - F_X(v))^{n-u-v}.$$

Proof. (a) $X_{(j)} \leqslant x_i$ means that among $X_1, \ldots, X_n$, at least j are $\leqslant x_i$ and at most $n-j$ are $\geqslant x_i$. For $k \in [j, n]$, the probability of exactly k less than x_i and $n-k$ greater than x_i follows a binomial distribution:

$$\binom{n}{k} \mathbb{P}(X \leqslant x_i)^k \mathbb{P}(x_i > k)^{n-k} = \binom{n}{k} p_i^k (1-p_i)^{n-k}.$$

Summing over all possible k's we establish our claim.

(b) If X is uniformly distributed on $[0,1]$ then $f_X \equiv 1$ on $[0,1]$ and $F(x) = x$. Therefore

$$f_{X_{(j)}} = \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x)^{n-j}$$

which indeed is consistent with a $(j, n-j+1)$-distributed beta distribution. Therefore

$$\begin{aligned} \mathbb{E}X_{(j)} &= \int_0^1 x f_{X_{(j)}} \, \mathrm{d}x = \frac{n!}{(j-1)!(n-j)!} \int_0^1 c^j (1-x)^{n-j} \, \mathrm{d}x \\ &= \frac{n!}{(j-1)!(n-j)!} \cdot \frac{j!(n-j)!}{(n+1)!} = \frac{j}{n+1}. \end{aligned}$$

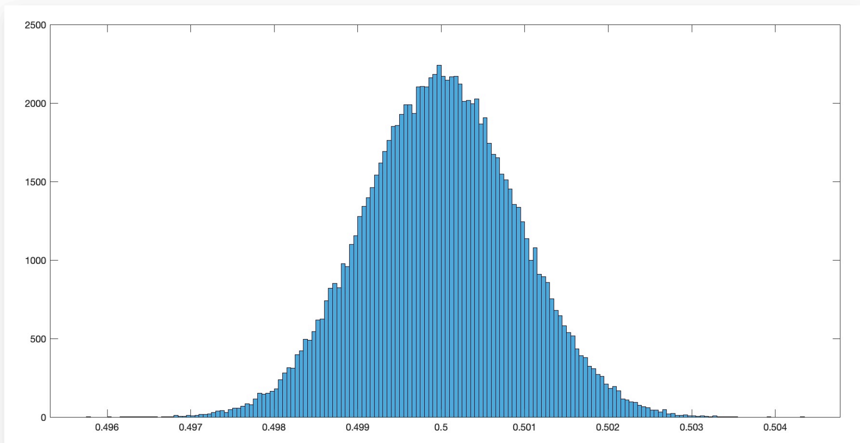
(c) This is because the events $\{X_j \leqslant a\}, \{a < X_j < b\}, \{\text{otherwise}\}$ partition the event space and that the X_i's are i.i.d. Therefore the probability of getting $u, v, n-u-v$ occurrences of each follows a multinomial distribution.

□

Problem 2

Using Matlab, verify that its random number generator agrees with the LLN. For example, average 10^6 samples from the uniform distribution on $[0,1]$ and check how close the sample average is to $1/2$. Also make a histogram and check how close the histogram is to a Gaussian.

Solution.



Problem 3

Let $X : \Omega \to \mathbb{R}$ be a random variable on Ω equipped with $\mathbb{P}$. For $t \in \mathbb{R}$ define $F(t); = \mathbb{P}(X \leqslant t)$. For $s \in (0,1)$ define

$$Y(s) := \sup\{t \in \mathbb{R} : F(t) < s\}.$$

Then Y is a random variable on $(0,1)$ with uniform probability law on $(0,1)$. Show that X and Y are equal in distribution, i.e., $\mathbb{P}(Y \leqslant t) = F(t)$ for all $t \in \mathbb{R}$.

Proof. Notice that we have another definition for $Y(s)$:

$$Y(s) = \sup\{t \in \mathbb{R} : F(t) < s\} = \inf\{t \in \mathbb{R} : F(t) \geqslant s\}. \tag{1}$$

Furthermore, by definition of supremum and infimum, whether or not the inequalities are strict impose no effect, so $\leqslant$ and $<$, $\geqslant$ and $>$ are freely interchangeable.

Now, given $F : \mathbb{R} \to [0,1]$, the CDF of X, let $Y : \text{Range}(F) \to \mathbb{R}$ be its generalized inverse. Then by (1) we have

$$Y(F(t)) = \inf\{\tilde{t} \in \mathbb{R} : F(\tilde{t}) \geqslant F(t)\} \leqslant t$$

since t is in the set of which the infimum is taken. A symmetric argument for $F(Y(t))$ can be obtained analogously, and thus

$$Y(F(t)) \leqslant t \qquad \text{and} \qquad F(Y(s)) \geqslant s. \tag{2}$$

Also observe that Y is monotone increasing: if $a \leqslant b$ then

$$\{x : F(x) \geqslant b\} \subset \{x : F(x) \geqslant a\}$$

so

$$\inf\{x : F(x) \geqslant b\} = Y(b) \geqslant Y(a) = \inf\{x : F(x) \geqslant a\}. \tag{3}$$

Now we prove $\mathbb{P}(Y \leqslant t) = F(t)$. This is true because on one hand

$$\begin{aligned}
\mathbb{P}(Y \leqslant t) &= \mathbb{P}_{\text{unif}}(\{s \in [0,1] : Y(s) \leqslant t\}) \\
&= \mathbb{P}_{\text{unif}}(\{s \in [0,1] : F(Y(s)) \leqslant F(t)\}) \\
&\leqslant \mathbb{P}_{\text{unif}}(\{x \in [0,1] : s \leqslant F(t)\}) && \text{[By (2)]} \\
&= \int_0^{F(t)} 1 \, \mathrm{d}s = F(t),
\end{aligned}$$

and on the other hand

$$\begin{aligned}
F(t) = \mathbb{P}(X \leqslant t) &= \mathbb{P}_{\text{unif}}(\{s \in [0,1] : s < F(t)\}) \\
&= \mathbb{P}_{\text{unif}}(\{s \in [0,1] : Y(s) < Y(F(t))\}) \\
&\leqslant \mathbb{P}_{\text{unif}}(\{s \in [0,1] : Y(s) < t\}) && \text{[By (2)]} \\
&= \mathbb{P}_{\text{unif}}(\{s \in [0,1] : Y(s) \leqslant t\}) && [\mathbb{P}(Y(s) = t) = 0] \\
&= \int_0^t Y(s) \, \mathrm{d}s = \mathbb{P}(Y \leqslant t).
\end{aligned}$$

□

Problem 4: Box-Muller Algorithm

Let U_1, U_2 be independent variables uniformly distributed in $(0,1)$. Define

$$R := \sqrt{-2 \log U_1}, \Phi := 2\pi U_2, \qquad X := R\cos\Phi, Y := R \sin\Phi.$$

Show that X, Y are independent standard Gaussians.

Then, let $X := (X_1, \ldots, X_n)$ be a vector of i.i.d. standard Gaussians. Let A be an $n \times n$ symmetric positive semidefinite matrix and let $A = RR^T$ be its Cholesky decomposition. Let $e^{(i)}$ be the i^{th} row of R. For $1 \leqslant i \leqslant n$ define $Z_i := \langle X, e^i \rangle$. Show that $\mathbb{E}(Z_i Z_j) = a_{ij}$.

Proof. Notice that the inverse transformations are given by

$$U_1 = \exp\left(-\frac{X^2 + Y^2}{2}\right) \qquad \text{and} \qquad U_2 = \frac{1}{2\pi}\arctan(Y/X).$$

(The first is obtained by taking $X^2 + Y^2$ to cancel out U_2 and the second is by taking Y/X to cancel out U_1.)

Then, the Jacobian for the transformation $(X, Y) \mapsto (U_1, U_2)$ is

$$\begin{vmatrix} \partial U_1/\partial X & \partial U_1/\partial Y \\ \partial U_2/\partial X & \partial U_2/\partial Y \end{vmatrix} = \begin{vmatrix} -\exp()X & -\exp()Y \\ -\frac{1}{2\pi}\frac{1}{1+Y^2/X^2}\frac{Y}{X^2} & \frac{1}{2\pi}\frac{1}{1+Y^2/X^2}\frac{1}{X} \end{vmatrix}$$
$$= \left| -\exp\left(-\frac{X^2+Y^2}{2}\right)\frac{1}{2\pi}\frac{1}{1+Y^2/X^2}\left(1+\frac{Y^2}{X^2}\right)\right|$$
$$= \frac{1}{2\pi}\exp\left(-\frac{X^2+Y^2}{2}\right).$$

Therefore,

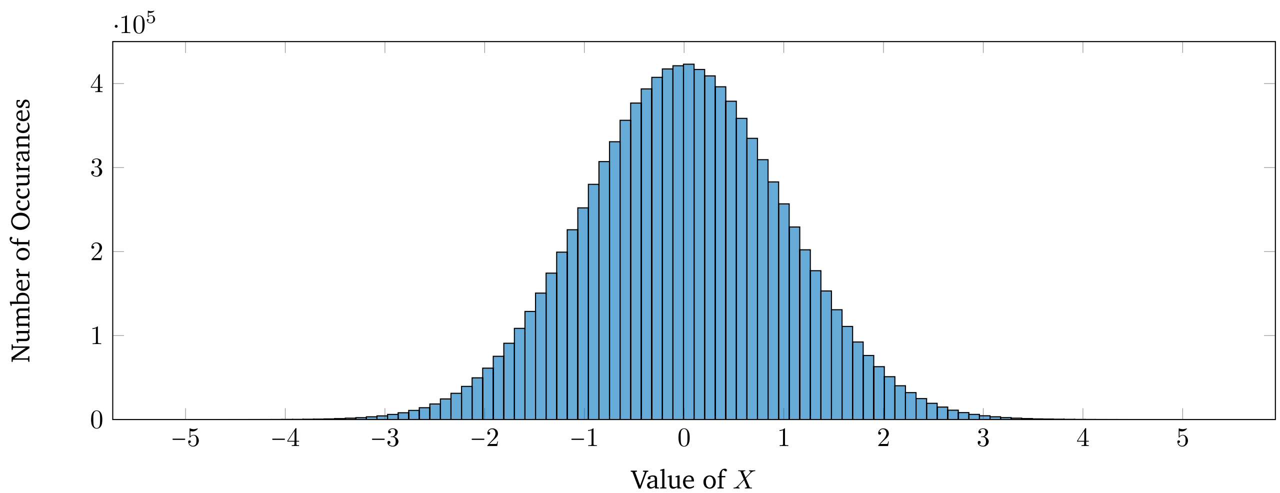
$$f_{X,Y}(x,y) = f_{U_1,U_2}(u_1,u_2)\mathcal{J}(u_1,u_2)$$
$$= 1 \cdot \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}}e^{-y^2/2}.$$

A simple calculation shows that the X-marginal and Y-marginal indeed have the PDFs of a Gaussian, and the claim therefore follows as $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. □

Code and output for X below:

```
U1 = rand(1,10^7);
U2 = rand(1,10^7);
X = sqrt((-2 * log(U1))) .* sin(2*pi*U2);
Y = sqrt((-2 * log(U1))) .* cos(2*pi*U2);

histogram(X,100);
```



Finally,

$$\mathbb{E}(Z_iZ_j) = \mathbb{E}\left\langle X, e^i\right\rangle\left\langle X, e^j\right\rangle = \mathbb{E}\sum_{k,\ell=1}^{n} X_k e_k^i \cdot X_\ell e_\ell^j = \mathbb{E}\sum_{k,\ell=1}^{n} e_k^i e_\ell^j X_k X_\ell = \mathbb{E}\sum_{k=1}^{n} e_k^{(i)} e_k^j = a_{ij}.$$

Problem 6

Let A, B, Ω be sets. Let $u : \Omega \to A$ and $t : \Omega \to B$. Assume that for every $x, y \in \Omega$, if $u(x) = u(y)$ then $t(x) = t(y)$. Show that there exists a function $s : A \to B$ such that $t = s(u)$.

Proof. Let $X \subset A$ be the range of u. Then, for $x \in X$ there exists some $\omega \in \Omega$ such that $x = u(\omega)$. Define $s : A \to B$ by $s(x) := t(\omega)$. Then $t(w) = s(u(\omega))$ so the claim is met. Next, if $\omega_1 = \omega_2$, i.e., if $u(\omega_1) = u(\omega_2)$, then by assumption $t(\omega_1) = t(\omega_2)$, so our mapping is well-defined. □

Problem 7

Let $\{f_\theta : \theta \in \Theta\}$ be a k-parameter exponential family $\{f_\theta : \theta \in \Theta, a(w(\theta)) < \infty\}$ of PDFs or PMFs where

$$f_\theta(x) := h(x) \exp\Big(\sum_{i=1}^{k} w_i(\theta) t_i(x) - a(w(\theta))\Big), \qquad \text{for all } x \in \mathbb{R}.$$

For $\theta \in \Theta$, let $w(\theta) := (w_1(\theta), \ldots, w_k(\theta))$. Assume that the following subset of $\mathbb{R}^k$ is k-dimensional:

$$\{w(\theta) - w(\theta') \in \mathbb{R}^k : \theta, \theta' \in \Theta\}.$$

Let $X = (X_1, \ldots, X_n)$ be a random sample of size n from f_θ and define $t : \mathbb{R}^n \to \mathbb{R}^n$ by $t(X) := \sum_{i=1}^{n} (t_1(X_i), \ldots, t_k(X_i))$. Show that $t(X)$ is minimal sufficient for θ.

Proof. We recall the characterization of MSS: a MSS satisfies

$$\text{if } f_\theta(x) = c(x, y) f_\theta(y) \text{ for } c \text{ not depending on } \theta, \qquad \text{then} \qquad t(x) = t(y).$$

Suppose the LHS is satisfied. Looking at the exponential family we see that $\langle w(\theta), t(y)\rangle - \langle w(\theta), t(x)\rangle$ must then be a constant c depending solely on x, y. Therefore, for these fixed x, y, for any $\theta_1, \theta_2 \in \Theta$, we have

$$\langle w(\theta_1), t(y)\rangle - \langle w(\theta_1), t(x)\rangle = \langle w(\theta_2), t(y)\rangle - \langle w(\theta_2), t(x)\rangle$$

so

$$\langle w(\theta_1) - w(\theta_2), t(y) - t(x)\rangle = 0.$$

Since by assumption $\{w(\theta_1) - w(\theta_2) : \theta_1, \theta_2 \in \Theta\}$ is assumed to be k-dimensional, its orthogonal complement is $\{0\}$, meaning that $t(x) = t(y)$. This proves that $t(X)$ is an MSS (sufficiency is immediate following the exponential form). □

Problem 8

Let $\mathbb{P}_1, \mathbb{P}_2$ be two probability laws on $\Omega = \mathbb{R}$. Suppose they induce PDFs f_1, f_2. Show that

$$\sup_{A \subset \mathbb{R}} |\mathbb{P}_1(A) - \mathbb{P}_2(A)| = \frac{1}{2} \int_{\mathbb{R}} |f_1(x) - f_2(x)| \, \mathrm{d}x.$$

Similarly, if $\Omega = \mathbb{Z}$, show that

$$\sup_{A \subset \mathbb{Z}} |\mathbb{P}_1(A) - \mathbb{P}_2(A)| = \frac{1}{2} \sum_{z \in \mathbb{Z}} |\mathbb{P}_1(z) - \mathbb{P}_2(z)|.$$

Proof. Define $S := \{x : f_1(x) > f_2(x)\}$. On one hand

$$0 = \int_{\mathbb{R}} f_1(x) - f_2(x) \, \mathrm{d}x = \int_S f_1(x) - f_2(x) \, \mathrm{d}x + \int_{S^c} f_1(x) - f_2(x) \, \mathrm{d}x,$$

so

$$\int_S f_1(x) - f_2(x) \, \mathrm{d}x = \int_{S^c} f_2(x) - f_1(x) \, \mathrm{d}x.$$

On the other hand,

$$\begin{aligned} \int_{\mathbb{R}} |f_1(x) - f_2(x)| \, \mathrm{d}x &= \int_S f_1(x) - f_2(x) \, \mathrm{d}x + \int_{S^c} f_2(x) - f_1(x) \, \mathrm{d}x \\ &= 2 \int_S |f_1(x) - f_2(x)| \, \mathrm{d}x \geqslant 2 \left| \int_S f_1(x) - f_2(x) \, \mathrm{d}x \right|. \end{aligned}$$

It is clear that if $B \subset \mathbb{R}$ and $B \neq S$ then either B contains extra parts on which $f_1 \leqslant f_2$ or misses parts on which $f_1 > f_2$ (or both). This would lead to the integral having even smaller (absolute) value. Therefore

$$\sup_{A \subset \mathbb{R}} |\mathbb{P}_1(A) - \mathbb{P}_2(A)| \leqslant \frac{1}{2} \int_{\mathbb{R}} |f_1(x) - f_2(x)| \, \mathrm{d}x$$

whereas the supremum is attained by E. The second case follows by replacing $\mathrm{d}x$ by a counting measure. □

Problem 9

Find a statistic Y that is complete and nonconstant but not sufficient.

Solution. Consider $t(X_1, \ldots, X_n) := X_1$ where X_i are i.i.d. Bernoulli with $0 < p < 1$. It is complete because if $\mathbb{E}_p f(X_1) = 0$ for all p, then $pf(0) + (1-p)f(1) = 0$ for all $p \in (0,1)$. This means $f(0) = f(1) = 0$. However it is not sufficient since

$$\mathbb{P}((X_1, \ldots, X_n) = (x_1, \ldots, x_n) \mid X_1 = x_1) = \mathbb{P}((X_2, \ldots, X_n) = (x_2, \ldots, x_n)) = \prod_{i=2}^{n} p^{x_i} (1-p)^{1-x_i}$$

which still depends on p.

Problem 10

This exercise shows that a complete sufficient statistic might not exist.
Let $X_1, \ldots, X_n$ be a random sample of size n from the uniform distribution on $\{\theta, \theta+1, \theta+2\}$ where $\theta \in \mathbb{Z}$.

(1) Show that $Y := (X_{(1)}, X_{(n)})$ is minimal sufficient for θ.

(2) Show that Y is not complete by considering $X_{(n)} - X_{(1)}$.

(3) Using minimal sufficiency that any sufficient statistic for θ is not complete.

Proof. (1) We use the proportion coefficient characterization of a MSS. Suppose that for all $\theta \in \mathbb{Z}$ we have $x_1, ..., x_n, y_1, ..., y_n$ such that $f_\theta(x) = c(x, y) f_\theta(y)$ where $x := (x_1, ..., x_n)$, $y := (y_1, ..., y_n)$, and $c(x, y)$ does not depend on θ.

For such $x \in \mathbb{Z}^n$, there exist exactly $3 - (\max x_i - \min x_i)$ solutions of θ for which $f_\theta(x)$ is nonzero. (For example if $\max x_i = \min x_i + 1$ then θ can only be $\min x_i - 1$ or $\min x_i$.) Letting x, y vary, we must have $\max x_i - \min x_i = \max y_i = \min y_i$ if the equation holds for all θ: for example if $\max x_i < \max y_i$, then if $\theta := \max y_i$ we see $f_\theta(y) > 0 = f_\theta(x)$. This shows that $(X_{(n)}, X_{(1)}) = (Y_{(n)}, Y_{(1)})$ under such assumptions.

Conversely, if $(X_{(n)}, X_{(1)}) = (Y_{(n)}, Y_{(1)})$ then we simply reverse the argument. Hence $(X_{(n)}, X_{(1)})$ is MSS.

(2) $X_{(n)} - X_{(1)}$ cancels out the θ when making subtraction so its distribution does not depend on θ. That means $\mathbb{E}_\theta(X_{(n)} - X_{(1)})$ is just some constant, which we call c. Then $\mathbb{E}_\theta(X_{(n)} - X_{(1)} - c) = 0$ whereas $X_{(n)} - X_{(1)}$ is not identically zero, showing that $X_{(n)} - X_{(1)}$ is not complete.

(3) If Z is sufficient for θ, then by MSS there exists a function φ with $(X_{(n)}, X_{(1)}) = \varphi(Z)$. To use (2) we define $f(x, y) := y - x$. Then $\mathbb{E}_\theta(f(\varphi(Z)) - c) = 0$ whereas $f \circ \varphi$ is not identically 0. This shows that there does not exist a complete sufficient statistic for θ, thus completing our proof. □