541A Midterm 1 Solutions[1]

1. Question 1

- Let (X, Y) be a vector in $\mathbf{R}^2$ that is uniformly distributed in the set
$$\{(x, y) \in \mathbf{R}^2 \colon x^2 + y^2 = 1\}.$$
Compute $\mathbf{E}\,|X|$.
- Let $X_1, \ldots, X_n$ be i.i.d. standard Gaussian random variables. (So, each random variable has PDF $t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, $t \in \mathbf{R}$.) Compute
$$\lim_{n\to\infty} \mathbf{P}(X_1^2 + \cdots + X_n^2 \leq n).$$
(Hint: use the Central Limit Theorem. Recall also that $\mathbf{E}X_1^2 = 1$ and $\mathbf{E}X_1^4 = 3$.)

Solution. We can rewrite (X, Y) as $(X, Y) = (\cos\Theta, \sin\Theta)$ where Θ is a uniformly distributed random variable in $[0, 2\pi)$. That is, Θ has density $f(\theta) := \frac{1}{2\pi}$ for all $\theta \in [0, 2\pi)$, so that

$$\mathbf{E}\,|X| = \mathbf{E}\,|\cos\Theta| = \frac{1}{2\pi}\int_0^{2\pi} |\cos\theta|\,\mathrm{d}\theta = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \cos\theta\,\mathrm{d}\theta = \frac{1}{\pi}[\sin\theta]_{-\pi/2}^{\pi/2} = \frac{2}{\pi}.$$

Now, using the CLT, and $\mathrm{Var}(X^2) = \mathbf{E}X^4 - (\mathbf{E}X^2)^2 = 3 - 1 = 2$, we have

$$\lim_{n\to\infty} \mathbf{P}(X_1^2 + \cdots + X_n^2 \leq n) = \lim_{n\to\infty} \mathbf{P}\Big(\frac{X_1^2 + \cdots + X_n^2 - n}{\sqrt{n}\sqrt{2}} \leq 0\Big) = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\mathrm{d}t = \frac{1}{2}.$$

2. Question 2

Recall that a Gaussian density with mean $\mu \in \mathbf{R}$ and standard deviation $\sigma > 0$ as

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \forall\, x \in \mathbf{R}.$$

Write this density as a two-parameter exponential family $\{f_\theta \colon \theta \in \Theta\}$, where $\Theta = \{(\mu, \sigma^2) \colon \mu \in \mathbf{R},\ \sigma > 0\}$, and $\theta = (\theta_1, \theta_2) \in \Theta$. That is, write this density as

$$f_\theta(x) := h(x)\exp\Big(\sum_{i=1}^2 w_i(\theta)t_i(x) - a(w(\theta))\Big), \qquad \forall\, x \in \mathbf{R},$$

for some $t_1, t_2 \colon \mathbf{R} \to \mathbf{R}$, $w_1, w_2 \colon \Theta \to \mathbf{R}$, $w := (w_1, w_2)$, $a \colon \mathbf{R}^2 \to \mathbf{R}$, $h \colon \mathbf{R} \to [0, \infty)$.

Solution. We have

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \exp\Big(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \Big(\frac{\mu^2}{2\sigma^2} + \log\sigma\Big)\Big), \qquad \forall\, x \in \mathbf{R}.$$

Then, we interpret θ as $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2) \in \mathbf{R}^2$, and define

$$t_1(x) := x, \qquad t_2(x) := x^2,$$
$$w_1(\theta) := \frac{\theta_1}{\theta_2} = \frac{\mu}{\sigma^2}, \qquad w_2(\theta) := -\frac{1}{2\theta_2} = -\frac{1}{2\sigma^2},$$

[1]February 25, 2022, © 2022 Steven Heilman, All Rights Reserved.

$$a(w(\theta)) := \frac{\theta_1^2}{2\theta_2} + \frac{1}{2}\log\theta_2 = \frac{\mu^2}{2\sigma^2} + \log\sigma,$$

and $h(x) := \frac{1}{\sqrt{2\pi}}$ for all $x \in \mathbf{R}$. Let $\Theta := \{\theta \in \mathbf{R}^2 \colon \theta_2 > 0\}$, and for any $\theta \in \Theta$, define

$$f_\theta(x) := h(x)\exp\Big(\sum_{i=1}^{2} w_i(\theta)t_i(x) - a(w(\theta))\Big), \qquad \forall\, x \in \mathbf{R}.$$

Then $\{f_\theta \colon \theta \in \Theta\}$ is a two parameter exponential family.

3. Question 3

Let X be a minimal sufficient statistic. Let Z be another sufficient statistic. Let u be a function such that

$$Z = u(X).$$

Show that Z is a minimal sufficient statistic.

Solution. Let Y be a sufficient statistic. We need to show that Z is a function of Y. Since X is minimal sufficient, there exists a function s such that $X = s(Y)$. By assumption, $Z = u(X)$, so that $Z = u(s(Y))$. That is, Z is a function of Y. Therefore, Z is minimal sufficient.

4. Question 4

Let $X = (X_1, \ldots, X_n)$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the three element set $\{\theta, \theta+1, \theta+2\}$, where $\theta \in \mathbf{R} =: \Theta$. (The distribution of X is then $f_\theta(x) = \prod_{i=1}^{n}(1/3)1_{x_i \in \{\theta,\theta+1,\theta+2\}}$, $\forall\, x \in \mathbf{R}^n$, $\forall\, \theta \in \mathbf{R}$.)

Let Y be a statistic that is sufficient for θ. (Let $t \colon \mathbf{R}^n \to \mathbf{R}^k$ such that $Y = t(X_1, \ldots, X_n)$.)

Is it true that we can write

$$f_\theta(x) = g_\theta(t(x))h(x), \qquad \forall\, \theta \in \Theta, \quad \forall\, x \in \mathbf{R}^n,$$

for some $g_\theta \colon \mathbf{R}^k \to [0,\infty)$, $h \colon \mathbf{R}^n \to [0,\infty)$?

Either prove this factorization can be done, or prove that it cannot be done.

Solution. This factorization can be done, but it does not follow from the statement of the Factorization Theorem, since the assumption we wrote there does not hold, since the set $\cup_{\theta\in\mathbf{R}}\{x \in \mathbf{R}^n \colon f_\theta(x) > 0\}$ is uncountable (this set contains the uncountable set $\cup_{\theta\in\mathbf{R}}\{x \in \mathbf{R}^n \colon x_1 = \cdots = x_n = \theta\} = \{x \in \mathbf{R}^n \colon x_1 = \cdots = x_n\}$.)

We then observe that we can just repeat the proof of one direction of the factorization theorem. Let $x \in \mathbf{R}^n$ and note that

$$f_\theta(x) = \mathbf{P}_\theta(X = x) = \mathbf{P}_\theta(X = x \text{ and } t(X) = t(x)) = \mathbf{P}_\theta(Y = t(x))\mathbf{P}_\theta(X = x | Y = t(x)).$$

By sufficiency, the last quantity does not depend on θ, so $f_\theta(x) = g_\theta(t(x))h(x)$, where $g_\theta(y) := \mathbf{P}_\theta(Y = y)$ for all $y \in \mathbf{R}^k$ and $h(x) := \mathbf{P}(X = x | Y = t(x))$ for all $x \in \mathbf{R}^n$.

5. Question 5

Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$, where $\theta \in \mathbf{R}$ is unknown.

Show that $(X_{(1)}, X_{(n)})$ is minimal sufficient but not complete.

Solution. Minimal sufficiency follows by Theorem 5.8 in the notes. Note that

$$f_\theta(x_1, \ldots, x_n) = 1_{x_1,\ldots,x_n \in [\theta-1/2,\theta+1/2]} = 1_{x_{(1)},x_{(n)} \in [\theta-1/2,\theta+1/2]}.$$

So, $f_\theta(x) = c(x,y) f_\theta(y)$ for all $\theta \in \mathbf{R}$ if and only if, for all $\theta \in \mathbf{R}$, we have the dichotomy that either

- $x_{(1)}, x_{(n)}, y_{(1)}, y_{(n)} \in [\theta - 1/2, \theta + 1/2]$, or
- $x_{(1)}, x_{(n)}, y_{(1)}, y_{(n)} \notin [\theta - 1/2, \theta + 1/2]$.

The last condition holds if and only if $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$. (The converse is clear, and for the forward direction, note e.g. if $x_{(1)} < y_{(1)}$, then choosing $\theta := y_{(1)} + 1/2$, we get $x_{(1)} \notin [\theta - 1/2, \theta + 1/2]$ but $y_{(1)} \in [\theta - 1/2, \theta + 1/2]$.)

Finally, the statistic is not complete, since $X_{(n)} - X_{(1)}$ is ancillary, and it is nonconstant with probability one, so there exists $c \in \mathbf{R}$ such that $\mathbf{E}_\theta[X_{(n)} - X_{(1)} - c] = 0$ for all $\theta \in \mathbf{R}$. (If $U_1, \ldots, U_n$ are i.i.d. uniform in $[-1/2, 1/2]$, then we can write $X_i = U_i + \theta$ for all $1 \leq i \leq n$, so that $X_{(n)} = U_{(n)} + \theta$ and $X_{(1)} = U_{(1)} + \theta$, then $X_{(n)} - X_{(1)} = U_{(n)} - U_{(1)}$, which does not depend on θ.)