STAT 30100 : Mathematical Statistics-1

Winter 2020

•

Homework 1

•

Solutions by

Jinhong Du

12243476

STAT 30100, Homework 1

The first few problems of this assignment cover the handouts `charfn.pdf` and `revmatalg.pdf`, which are available on Canvas and are assumed to be review for you.

1. Prove that the characteristic function of a univariate normal $N(\mu, \sigma^2)$ random variable Z is $\phi_Z(t) = \mathbb{E}[\exp(itZ)] = \exp(it\mu - \frac{t^2\sigma^2}{2})$.

Proof.

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{e}^{j\omega x} \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \cos(\omega x) \mathrm{d}x + i \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \sin(\omega x) \mathrm{d}x$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \cos(\omega x) \mathrm{d}x$$

Now, compute $\mathcal{F}'(\omega)$, and integrate by parts:

$$\mathcal{F}'(\omega) = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2}} x \sin(\omega x) \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(\omega x) \mathrm{d}\left(\mathrm{e}^{-\frac{x^2}{2}}\right)$$
$$= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \sin(\omega x)\Big|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2}} \omega \cos(\omega x) \mathrm{d}x = -\omega \mathcal{F}(\omega)$$

The solution to so obtained ODE, $\mathcal{F}'(\omega) = -\omega \mathcal{F}(\omega)$ is $\mathcal{F}(\omega) = c \exp\left(-\frac{\omega^2}{2}\right)$, and the integration constant is seen to be one from normalization requirement $\mathcal{F}(0) = 1$ of the Gaussian probability density. □

2. Prove (3), (37), (19), (22), (24) on the "Review of Matrix Algebra" handout. For the first part of (19) you may use the determinental expansion $|A| = \sum_i a_{ij} B_{ij}$, where B_{ij} is the (i, j)th cofactor of A. [Hint: The second part of (19) can be shown from the first part by writing the general matrix A as a product of two matrices of the type in the first part, one of which has unit determinant and the other of which has the same determinant as A.]

(3) $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, whenever AB and BA are defined (and square).

(37) $\mathbf{x}^\top A \mathbf{x} = \mathrm{tr}(A\mathbf{x}\mathbf{x}^\top)$.

(19) If $A_{n\times n} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, where A_{11} and A_{22} are square matrices, then if $A_{12} = 0$ (or $A_{21} = 0$), we have $|A| = |A_{11}||A_{22}|$. More generally, we have $|A| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|$.

(22) Let $\lambda_1, \lambda_2 \in \xi(A)$ where A is symmetric, $\lambda_1 \neq \lambda_2$. Then corresponding eigenvectors x_1 and x_2 exist, and $x_1 \perp x_2$ (i.e. $x_1^\top x_2 = 0$; x_1 and x_2 are orthogonal).

(24) (a) H idempotent implies all eigenvalues of H are 0 or 1. (b) If $H_{m\times m}$ is symmetric, then H is idempotent iff all eigenvalues of H are 0 or 1.

Proof. (3) Suppose that $A, B \in \mathbb{R}^{n\times n}$, then

$$\mathrm{tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{n} B_{ji} A_{ij} = \mathrm{tr}(BA).$$

Solution (cont.)

(37)

$$\mathbf{x}^\top A\mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n x_i A_{ij} x_j = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j = \sum_{i=1}^n \sum_{j=1}^n A_{ij} (\mathbf{x}\mathbf{x}^\top)_{ji} = \operatorname{tr}(A\mathbf{x}\mathbf{x}^\top).$$

(19) Suppose that $A_{11} \in \mathbb{R}^{n_1 \times n_1}$ and $A_{22} \in \mathbb{R}^{n_2 \times n_2}$ where $n_1 + n_2 = n$. If $A_{12} = 0$, then

$$\begin{aligned} \left|\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}\right| &= \left|\begin{pmatrix} A_{11} & 0 \\ A_{21} & I_{n_2} \end{pmatrix}\begin{pmatrix} I_{n_1} & 0 \\ 0 & A_{22} \end{pmatrix}\right| \\ &= \left|\begin{pmatrix} A_{11} & 0 \\ A_{21} & I_{n_2} \end{pmatrix}\right| \left|\begin{pmatrix} I_{n_1} & 0 \\ 0 & A_{22} \end{pmatrix}\right|. \end{aligned}$$

The determinental expansion of $\begin{pmatrix} A_{11} & 0 \\ A_{21} & I_{n_2} \end{pmatrix}$ along the nth column is given by $\sum_{i=1}^n a_{in} B_{in} = B_{nn}$ where B_{in} is the (i, n)th cofactor of the matrix. Notice that $B_{nn} = (-1)^{n+n}|A_{-n}|$ is just the determinant of the matrix A_{-n} formed by removing the nth column and row of A, and A_{-n} has the similar form of A that is block diagonal with a identity matrix at the lower right corner. Repeating expand the determinent along the last column, we will have $\left|\begin{pmatrix} A_{11} & 0 \\ A_{21} & I_{n_2} \end{pmatrix}\right| = |A_{11}|$. Analogously, we have $\left|\begin{pmatrix} I_{n_1} & 0 \\ 0 & A_{22} \end{pmatrix}\right| = |A_{22}|$. Therefore, $|A| = |A_{11}||A_{22}|$.

If $A_{21} = 0$, similarly, we have

$$\left|\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}\right| = \left|\begin{pmatrix} A_{11} & 0 \\ 0 & I_{n_2} \end{pmatrix}\right| \left|\begin{pmatrix} I_{n_1} & A_{12} \\ 0 & A_{22} \end{pmatrix}\right| = |A_{11}||A_{22}|.$$

If A_{22} is invertible, then

$$\left|\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\right| = \left|\begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{21} & I_{n_2} \end{pmatrix}\right| \left|\begin{pmatrix} I_{n_1} & A_{12} \\ 0 & A_{22} \end{pmatrix}\right| = |A_{11} - A_{12}A_{22}^{-1}A_{21}||A_{22}|.$$

(22) Since A is symmetric, it can be decomposed untarily as $A = Q\Lambda Q^\top$ where $Q \in \mathbb{R}^n$ is orthonormal and $\Lambda \in \mathbb{R}^{n \times n}$ is a diagonal matrix with diagonal entries being $\lambda_1, \ldots, \lambda_n$. For all $\mathbf{x} \in \mathbb{R}$ and $\lambda \in \mathbb{R}$, $Ax = \lambda x$ is equivalent to $\Lambda(Q^\top x) = \lambda(Q^\top x)$. Let $\lambda = \lambda_1$ and λ_2, then the eigenvector corresponding to λ_1 and λ_2 is given by $x_1 = Qe_1$ and $x_2 = Qe_2$ respectively (unique up to a scaling factor), where $e_i \in \mathbb{R}^n$ is a unit vector with ith entry being 1. $x_1^\top x_2 = e_1^\top Q^\top Q e_2 = e_1^\top e_2 = 0$.

(24) (a) First, we show that for different eigenvalues, the corresponding eigenvectors are different. Otherwise, $Hx = \lambda_1 x = \lambda_2 x$ with $\lambda_1 \neq \lambda_2$, and then all linear combination of λ_1 and λ_2 is also an eigenvalue. However, H can have at most n eigenvalues since the eigenvalues are the roots of a degree-n polynomial.

Suppose that λ is any eigenvalue of H with respect to eigenvector x, i.e., $Hx = \lambda x$. Then $Hx = H^k x = \lambda H x = \lambda^k x$, i.e., λ^k is also a eigenvalue of H with respect to x for all $k = 1, 2, \ldots$. Therefore, $\lambda = \lambda^k$ for all $k \in \mathbb{Z}^+$. So $\lambda = 0$ or 1.

(b) $\Longrightarrow$ Result follows by (a).

$\Longleftarrow$ Sicne H is symmetric, it can be decomposed unitarily as $H = Q\Lambda Q^\top$ with where $Q \in \mathbb{R}^n$ is orthonormal and $\Lambda \in \mathbb{R}^{n \times n}$ is a diagonal matrix with diagonal entries being 0 or 1. Then $\Lambda^2 = \Lambda$ since Λ is a diagonal matrix with diagonal entries being 0 or 1. Since $H^2 = Q\Lambda Q^\top Q\Lambda Q^\top = Q\Lambda^2 Q^\top = Q\Lambda Q^\top = H$, we have H is idempotent.

□

3. Suppose the real matrix H is $n \times n$ symmetric and idempotent and projects R^n onto a linear subspace L, i.e. $L = \text{range}(H) = \{x \in \mathbb{R}^n \text{ s.t. } x = Hy \text{ for some } y \in \mathbb{R}^n\}$. Let $L^\perp$ be the orthogonal complement of L, i.e. $L^\perp = \{x \in \mathbb{R}^n \text{ s.t. } x^\top y = 0 \text{ for all } y \in L\}$.

(a) Show that $I - H$ is symmetric and idempotent.

Proof.

$$(I-H)^\top = I^\top - H^\top = I - H$$
$$(I-H)^2 = I - 2H + H^2 = I - 2H + H = I - H$$

□

(b) Show that $I - H$ projects $\mathbb{R}^n$ onto $L^\perp$, i.e. show $L^\perp = \text{range}(I - H)$.

Proof. Since $\forall\ z \in \mathbb{R}^n$, $z = z_1 + z_2$ where $z_1 \in L$, $z_2 \in L^\perp$ and $z_1^\top z_2 = 0$. Also, we have $z = (H + I - H)z = Hz + (I - H)z$, $Hz \in L$. So, $Hz = z_1$.
Since

$$(Hz)^\top (I-H)z = z^\top H(I-H)z = z^\top (H-H)z = 0$$

and the orthogonal decomposition is unique, we have $z_2 = (I - H)z$.
Since this holds for all $z \in \mathbb{R}$, we have $L^\perp = \text{range}(I - H)$.

□

4. Consider the lognormal random variable X with pdf

$$f(x) = \frac{1}{\sqrt{2\pi}x} e^{-\frac{(\log x)^2}{2}}, \qquad 0 \leq x < \infty.$$

(a) Show that $\mathbb{E}X^r = e^{\frac{r^2}{2}}$, $r = 0, 1, \ldots$, so all moments of X are finite.

Proof. For $r = 0, 1, \ldots,$

$$\begin{aligned}
\mathbb{E}X^r &= \int_0^\infty x^r \frac{1}{\sqrt{2\pi}x} e^{-\frac{1}{2}\log^2 x} \mathrm{d}x \\
&\xlongequal{y=\log x} \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{ry} e^{-\frac{1}{2}y^2} \mathrm{d}y \\
&= \frac{e^{\frac{1}{2}r^2}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(y-r)^2} \mathrm{d}y \\
&= e^{\frac{r^2}{2}}
\end{aligned}$$

□

(b) Show that the moment generating function of X does not exist, i.e. there does not exist $h > 0$ such that $\mathbb{E}e^{tX}$ is finite for all t in $-h < t < h$.

Proof.

$$\begin{aligned}\mathbb{E}e^{tX} &= \int_0^\infty e^{tx}\frac{1}{\sqrt{2\pi}x}e^{-\frac{1}{2}\log^2 x}\mathrm{d}x \\ &= \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi}xe^{\frac{1}{2}\log^2 x}}\mathrm{d}x\end{aligned}$$

For $t > 0$, since

$$\lim_{x\to\infty}\frac{e^{tx}}{\sqrt{2\pi}xe^{\frac{1}{2}\log^2 x}} = \infty,$$

we have $\int_0^\infty \frac{e^{tx}}{\sqrt{2\pi}xe^{\frac{1}{2}\log^2 x}}\mathrm{d}x = \infty$. Therefore $\mathbb{E}e^{tX} = \infty$ for $t > 0$. $\square$

5. (Problem 3.29 in Casella and Berger) For each family in

(a) normal family with either parameter μ or σ known;

(b) gamma family with either parameter α or β known or both unknown;

(c) beta family with either parameter α or β known or both unknown;

(d) Poisson family;

(e) negative binomial family with r known, $0 < p < 1$,

describe the natural parameter space.

(a) If μ is known,

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} = e^{-\frac{\mu^2}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{x\frac{\mu}{\sigma^2}+x^2(-\frac{1}{2\sigma^2})}$$

then $t_1(x) = x, t_2 = x^2$, $\eta = \begin{pmatrix}\frac{\mu}{\sigma^2} & -\frac{1}{2\sigma^2}\end{pmatrix}^\top$ and $\mathcal{H} = \{\eta = (\eta_1,\eta_2)^\top \in \mathbb{R}^2 : \mu\eta_1 > 0, \eta_2 < 0\}$.
If σ is known,

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}\cdot e^{-\frac{\mu^2}{2\sigma^2}}\cdot e^{x\frac{\mu}{\sigma^2}}$$

then $t_1(x) = x$, $\eta = \begin{pmatrix}\frac{\mu}{\sigma^2}\end{pmatrix}$ and $\mathcal{H} = \mathbb{R}$.
(b) For $\alpha,\beta\in\mathbb{R}^+$, $f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$. If α is known,

$$f(x;\alpha,\beta) = x^{\alpha-1}\cdot\frac{\beta^\alpha}{\Gamma(\alpha)}\cdot e^{x(-\beta)}$$

then $t_1(x) = x$, $\eta = (-\beta)$ and $\mathcal{H} = \mathbb{R}^-$.
If β is known,

$$f(x;\alpha,\beta) = e^{x(-\beta)}\cdot\frac{\beta^\alpha}{\Gamma(\alpha)}\cdot e^{\ln x(\alpha-1)}$$

then $t_1(x) = \ln x$, $\eta = (\alpha-1)$ and $\mathcal{H} = (-1,\infty)$.

Solution (cont.)

If both are unknown,

$$f(x;\alpha,\beta) = 1 \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot e^{\ln x(\alpha-1)+x(-\beta)}$$

then $t_1(x) = \ln x$, $t_2(x) = x$, $\eta = \begin{pmatrix}\alpha - 1 & -\beta\end{pmatrix}^\top$ and $\mathcal{H} = \{\eta = \begin{pmatrix}\eta_1 & \eta_2\end{pmatrix} \in \mathbb{R}^2 : \eta_1 > -1, \eta_2 < 0\}$.

(c) For $\alpha, \beta \in \mathbb{R}^+$, $f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. If α is known,

$$f(x;\alpha,\beta) = x^{\alpha-1} \cdot \frac{1}{B(\alpha,\beta)} \cdot e^{\ln(1-x)\cdot(\beta-1)}$$

then $t_1(x) = \ln(1-x)$, $\eta = (\beta - 1)$ and $\mathcal{H} = (-1, \infty)$.

If β is known,

$$f(x;\alpha,\beta) = (1-x)^{\beta-1} \cdot \frac{1}{B(\alpha,\beta)} \cdot e^{\ln x\cdot(\alpha-1)}$$

then $t_1(x) = \ln x$, $\eta = (\alpha - 1)$ and $\mathcal{H} = (-1, \infty)$.

If both are unknown,

$$f(x;\alpha,\beta) = 1 \cdot \frac{1}{B(\alpha,\beta)} \cdot e^{\ln x\cdot(\alpha-1)+\ln(1-x)\cdot(\beta-1)}$$

then $t_1(x) = \ln x$, $t_2(x) = \ln(1-x)$, $\eta = \begin{pmatrix}\alpha - 1 & \beta - 1\end{pmatrix}^\top$ and $\mathcal{H} = \{\eta = \begin{pmatrix}\eta_1 & \eta_2\end{pmatrix} \in \mathbb{R}^2 : \eta_1 > -1, \eta_2 > -1\}$.

(d) For $\lambda \in \mathbb{R}^+$,

$$f(x;\lambda) = \frac{\lambda^x}{e^{-\lambda}} x! = x! \cdot e^{-\lambda} \cdot e^{x \ln \lambda},$$

then $t_1(x) = x$, $\eta = (\ln \lambda)$ and $\mathcal{H} = \mathbb{R}$.

(e) For $p \in (0,1)$, if r is known,

$$f(x;r,p) = \binom{x+r-1}{x} p^r (1-p)^x = (-1)^x \binom{-r}{x} \cdot p^r \cdot e^{x\ln(1-p)},$$

then $t_1(x) = x$, $\eta = (\ln(1-p))$ and $\mathcal{H} = \mathbb{R}^-$.

6. For each of the following families:

   (i) Verify that it is a k-parameter exponential family and specify the value of k. Is it a curved or a full exponential family?

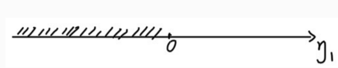
   (ii) Specify the original parameter space (where the original parameter is taken to be θ in each case).

   (iii) Specify and sketch a graph of the natural parameter space. (If the family is a curved exponential family, we could call this the curved parameter space.)

   (a) $N(\theta, \theta)$.

$$f(x;\theta,\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\theta)^2}{2\theta}} = e^x \cdot \frac{e^{-\frac{\theta}{2}}}{\sqrt{2\pi\theta}} \cdot e^{x^2(-\frac{1}{2\theta})}$$

$k = 1$, $t_1(x) = x^2$. It is a full exponential family. The original parameter space is $\mathbb{R}^+$. $\eta = (-\frac{1}{2\theta})$ so the natural parameter space is $\mathcal{H} = \mathbb{R}^-$.

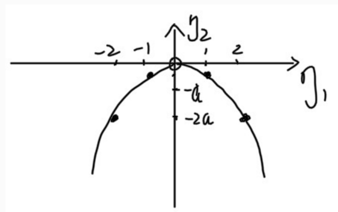


(b) $N(\theta, a\theta^2)$.

For $a > 0$,

$$f(x; \theta, a\theta^2) = \frac{1}{\sqrt{2\pi a\theta^2}} e^{-\frac{(x-\theta)^2}{2a\theta^2}} = e^{-\frac{1}{2a}} \cdot \frac{1}{\sqrt{2\pi a\theta^2}} \cdot e^{x\frac{1}{a\theta} + x^2\left(-\frac{1}{2a\theta^2}\right)}$$

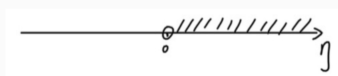
$k = 2$, $t_1(x) = x, t_2(x) = x^2$. It is a curved exponential family. The original parameter space is $\mathbb{R}\backslash\{0\}$. $\eta = \left(\frac{1}{a\theta}, -\frac{1}{2a\theta^2}\right)^\top$ so the natural parameter space is $\mathcal{H} = \{\eta = (\eta_1, \eta_2)^\top \in \mathbb{R}^2 : \eta_2 = -\frac{a}{2}\eta_1^2\}$.



(c) Gamma$(\theta, \frac{1}{\theta})$.

$$f(x; \theta, \tfrac{1}{\theta}) = \frac{\theta^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\theta x} = 1 \cdot \frac{\theta^\theta}{\Gamma(\theta)} \cdot e^{(\ln x - x)\cdot\theta} \cdot \frac{1}{x}$$

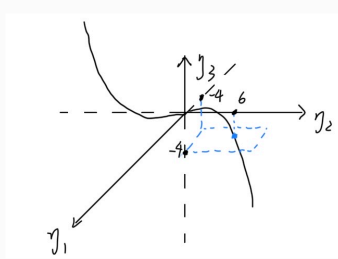
$k = 1$, $t_1(x) = \ln x - x$. It is a full exponential family. The original parameter space is $\mathbb{R}^+$. $\eta = (\theta)$ so the natural parameter space is $\mathbb{R}^+$.



(d) $f(x) = C\exp[-(x-\theta)^4]$, where C is a normalizing constant, $-\infty < x < \infty$.

$$f(x; \theta) = e^{-x^4} \cdot Ce^{-\theta^4} \cdot e^{x(4\theta^3) + x^2(-6\theta^2) + x^3(4\theta)}$$

$k = 3$, $t_1(x) = x, t_2(x) = x^2, t_3(x) = x^3$. It is a curved exponential family. The original parameter space is $\mathbb{R}$. $\eta = \left(4\theta^3, -6\theta^2, 4\theta\right)^\top$ so the natural parameter space is $\mathcal{H} = \{\eta = (\eta_1, \eta_2, \eta_3)^\top \in \mathbb{R}^3 : \eta_1 = \frac{1}{16}\eta_3^3, \eta_2 = \frac{3}{8}\eta_3^2, \eta_3 \in \mathbb{R}\}$.



7. (Problem 3.39 in Casella and Berger) Consider the Cauchy family defined in Section 3.3. This family can be extended to a location-scale family yielding pdfs of the form

$$f(x; \mu, \sigma) = \frac{1}{\sigma\pi\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)}, \qquad -\infty < x < \infty$$

The mean and variance do not exist for the Cauchy distribution. So the parameters μ, and σ^2 are not the mean and variance. But they do have important meaning. Show that if X is a random variable with a Cauchy distribution with parameters μ, and σ^2, then:

(a) μ is the median of the distribution of X, that is, $\mathbb{P}(X \geq \mu) = \mathbb{P}(X \leq \mu) = \frac{1}{2}$.

(b) $\mu + \sigma$ and $\mu + \sigma$ are the quartiles of the distribution of X, that is, $\mathbb{P}(X \geq \mu + \sigma) = \mathbb{P}(X \leq \mu - \sigma) = \frac{1}{4}$ (Hint: Prove this first for $\mu = 0$ and $\sigma = 1$ and then use Exercise 3.38.)

Proof. (a) Since $f(2\mu - x; \mu, \sigma) = \frac{1}{\sigma\pi\left(1+\left(\frac{2\mu-x-\mu}{\sigma}\right)^2\right)} = \frac{1}{\sigma\pi\left(1+\left(\frac{x-\mu}{\sigma}\right)^2\right)} = f(x; \mu, \sigma)$, we have $f(x; \mu, \sigma)$ is symmetrical about $x = \mu$. Therefore,

$$\mathbb{P}(X \geq \mu) = \int_{-\infty}^{\mu} f(x; \mu, \sigma)\mathrm{d}x = \int_{-\infty}^{\mu} f(2\mu - x; \mu, \sigma)\mathrm{d}x = \int_{\mu}^{+\infty} f(x; \mu, \sigma)\mathrm{d}x = \mathbb{P}(X \leq \mu).$$

Since $\mathbb{P}(X \geq \mu) + \mathbb{P}(X \leq \mu) = 1$, we have $\mathbb{P}(X \geq \mu) = \mathbb{P}(X \leq \mu) = \frac{1}{2}$.
(b) For $\mu = 0$ and $\sigma = 1$, $f(x; 0, 1) = \frac{1}{\pi(1+x^2)}$. Since $f(x; \mu, \sigma)$ is symmetrical about $x = 0$, it follows that

$$\mathbb{P}(X_{0,1} \geq 1) = \mathbb{P}(X_{0,1} \leq -1) = \int_{1}^{\infty} \frac{1}{\pi(1+x^2)}\mathrm{d}x = \frac{1}{\pi} \arctan x\Big|_1^{\infty} = \frac{1}{4}$$

Since $f(x; \mu, \sigma)$ is in a location-scale family and $Z_{\mu,\sigma} = \frac{Z_{0,1}-\mu}{\sigma}$, from Exercise 3.38, we have $\mathbb{P}(X_{\mu,\sigma} \geq \mu + \sigma) = \frac{1}{4}$. By symmetry $\mathbb{P}(X_{\mu,\sigma} \geq \mu - \sigma) = \frac{1}{4}$. $\square$

8. Suppose X is a continuous, real-valued random variable distributed according to a one-parameter exponential family having density

$$f(x; \eta) = h(x)c(\eta)e^{\eta t(x)}$$

with natural parameter $\eta \in \mathcal{H}$, where $\mathcal{H}$ is an interval in $\mathbb{R}^1$. Note: Do not use Casella and Bergers Theorem 3.4.2 to solve this problem.

(a) Express $\frac{\partial \log[f(x;\eta)]}{\partial\eta}$ in terms of $t(x)$ and $\mathbb{E}[t(X)]$.

Suppose that $\mathcal{H} = [\eta_1, \eta_2]$.
Since $\log f(x; \eta) = \log h(x) + \log c(\eta) + \eta t(x)$, we have

$$\frac{\partial \log[f(x;\eta)]}{\partial\eta} = \frac{c'(\eta)}{c(\eta)} + t(x). \tag{8.1}$$

Since $c(\eta)$ is the normalizing factor that is larger than 0, we have $\frac{1}{c(\eta)} = \int_{\mathbb{R}} h(x)e^{\eta t(x)}\mathrm{d}x$. Differentiating yields with respect to η,

$$-\frac{c'(\eta)}{c^2(\eta)} = \frac{\partial}{\partial\eta}\int_{\mathbb{R}} h(x)e^{\eta t(x)}\mathrm{d} = \int_{\mathbb{R}} h(x)\frac{\partial e^{\eta t(x)}}{\partial\eta}\mathrm{d} = \int_{\mathbb{R}} h(x)t(x)e^{\eta t(x)}\mathrm{d}x,$$

where the second equality comes from Lebesgue dominated convergence theorem since

(i) $h(x)e^{\eta t(x)}$ is integrable of x for each $\eta \in \mathcal{H}$;

(ii) $\frac{\partial}{\partial\eta}h(x)e^{\eta t(x)} = h(x)t(x)e^{\eta t(x)}$ exists for all $x \in \mathbb{R}$;

Solution (cont.)

(iii) $|h(x)e^{\eta t(x)}| \leq \max\{h(x)e^{\eta_1 t(x)}, h(x)e^{\eta_2 t(x)}\} \triangleq f_0(x)$ which is the dominated function.

Then,

$$\frac{c'(\eta)}{c(\eta)} = -\int_{\mathbb{R}} t(x)f(x;\eta)\mathrm{d}x = -\mathbb{E}[t(X)]. \tag{8.2}$$

Therefore,

$$\frac{\partial \log[f(x;\eta)]}{\partial \eta} = t(x) - \mathbb{E}[t(X)].$$

(b) Find $\mathbb{E}\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right]$.

$$\mathbb{E}\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right] = \mathbb{E}\left\{t(X) - \mathbb{E}[t(X)]\right\} = \mathbb{E}[t(X)] - \mathbb{E}[t(X)] = 0.$$

(c) Express $\mathrm{Var}\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right]$ in terms of $\frac{\partial^2 \log[f(x;\eta)]}{\partial \eta^2}$.

$$\begin{aligned}
\mathrm{Var}\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right] &= \mathbb{E}\left\{\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right]^2\right\} - \left\{\mathbb{E}\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right]\right\}^2 \\
&= \mathbb{E}\left\{\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right]^2\right\} \\
&= \mathbb{E}\left\{[t(X) - \mathbb{E}[t(X)]]^2\right\} \\
&= \mathrm{Var}[t(X)]. \qquad (8.5)
\end{aligned}$$

From (8.1), we have

$$\frac{\partial^2 \log[f(x;\eta)]}{\partial \eta^2} = \frac{\partial^2 \log c(\eta)}{\partial \eta^2}. \tag{8.5}$$

From (8.2), we have

$$\begin{aligned}
\frac{\partial^2 \log c(\eta)}{\partial \eta^2} &= -\frac{\partial}{\partial \eta}\int_{\mathbb{R}} t(x)f(x;\eta)\mathrm{d}x \\
&= -\int_{\mathbb{R}} t(x)\frac{\partial}{\partial \eta}f(x;\eta)\mathrm{d}x \\
&= -\int_{\mathbb{R}} t(x)[c'(\eta) + c(\eta)t(x)]e^{\eta t(x)}\mathrm{d}x \\
&= -\frac{c'(\eta)}{c(\eta)}\mathbb{E}[t(X)] - \mathbb{E}[t^2(X)] \\
&= \{\mathbb{E}[t(X)]\}^2 - \mathbb{E}[t^2(X)] \\
&= -\mathrm{Var}[t(X)]. \qquad (8.5)
\end{aligned}$$

Therefore, combining (8.5), (8.5) and (8.5), we have

$$\mathrm{Var}\left[\frac{\partial \log[f(X;\eta)]}{\partial \eta}\right] = -\frac{\partial^2 \log[f(x;\eta)]}{\partial \eta^2}$$