STA431 Assignment 3

(1) $v_3^T C = (0,0,1,0)$

(2) (a) $v_j^T x \sim N(0, \lambda_j)$, because from the formula sheet $Ax \sim N(A\mu, A\Sigma A^T)$, and

$$E(v_j^T x) = v_j^T E(x) = v_j^T 0 = 0$$
$$Var(v_j^T x) = cov(v_j^T x) = v_j^T \Sigma v_j$$
$$= v_j C D C^T v_j \overset{\text{Problem one}}{=} \lambda_j$$

(b) Again using the formula sheet,
$$y = C^T x \sim N_p(C^T 0, C^T C D C^T C) = N_p(0, D)$$

(c) Because the matrix D is diagonal, and for the multivariate normal, zero covariance implies independence.

(3) (a) Set $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Then $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Ax$

is MVN with expected value $A \begin{pmatrix} 1 \\ 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

and covariance matrix $A \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} A^T = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim MVN$

$$E\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$$

$$cov\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

$$y \quad = \quad A \quad x$$

$$④ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N_2(A\mu_x, A\Sigma_x A^T)$$

$$\mu_y = \begin{pmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{pmatrix}, \quad \Sigma_y = A \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} A^T$$

$$= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_2^2 \\ \sigma_1^2 & -\sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{pmatrix}$$

For independence, need $\sigma_1^2 = \sigma_2^2$

⑤ Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim N(\mu \mathbf{1}, \sigma^2 I_n)$

$$\bar{x} = \frac{1}{n}\mathbf{1}^T x \sim N\left(\frac{1}{n}\mathbf{1}^T \mu \mathbf{1}, \frac{1}{n}\mathbf{1}^T \sigma^2 I_n \left(\frac{1}{n}\mathbf{1}^T\right)^T\right)$$
$$= N\left(\frac{1}{n} n \mu, \frac{1}{n^2}\sigma^2 \mathbf{1}^T \mathbf{1}\right)$$
$$= N\left(\mu, \sigma^2 \frac{1}{n^2} n\right) = N\left(\mu, \frac{\sigma^2}{n}\right)$$

⑥ (a) $w - \mu \sim N_p(0, \Sigma)$

(b) $z = \Sigma^{-1/2}(w-\mu) \sim N(\Sigma^{-\frac{1}{2}} 0, \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}})$
$= N(0, I_p)$

(c) Because the covariance matrix is diagonal, and for the multivariate normal, zero covariance implies independence. They are standard normal because $E(z_j) = 0$, $Var(z_j) = 1$

(d) $z^T z = \sum_{j=1}^{p} z_j^2 \sim \chi^2(p)$, because a squared standard normal is $\chi^2(1)$, and if $y_1 \ldots y_p$ are independent with $y_j \sim \chi^2(\nu_j)$, then $\sum_{j=1}^{p} y_j \sim \chi^2(\sum_{j=1}^{p} \nu_j)$. That is, the sum of independent chi-squares is chi-squared.

(7) (a) Parameter is λ

(b) $\Theta = \{\lambda : \lambda > 0\}$

(8) (a) $\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$

(b) $\hat{\Theta} = \left(\bar{x}, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$

(c) $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \hat{\sigma}^2$

(d) $\hat{\Theta} = (94.38, 147.3976)$.

(e) $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{147.3976} = 12.14$

It's okay by invariance. The function

$g(\mu, \sigma^2) = (\mu, \sqrt{\sigma^2})$ is one to one,

since $\sigma^2 > 0$.

(9) This is a normal with $\mu = 0$ and $\sigma^2 = \theta$. Since $\sigma^2 = E(X^2) - \mu^2$,

Try $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \xrightarrow{P} E(X_i^2) = \theta$

by the Law of Large numbers.

(10) No. By the Law of Large Numbers,

$\bar{x}_n \xrightarrow{P} E(X_i) = \alpha\beta = \theta^2 \neq \theta$.

For consistency, must have convergence to the true value in the whole parameter space, and here it only works at one point.

(11) This was done in lecture. See slide 74 in the Estimation slide set.

(12) $\hat{\sigma}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x}\bar{y})$

$= \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y}\right)$

$= \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i - 2n\bar{x}\bar{y} + n\bar{x}\bar{y}\right)$

$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$

By the Law of Large numbers and continuous mapping, this converges to

$E(X_i Y_i) - E(X_i)E(Y_i) = Cov(X_i, Y_i)$

$= \sigma_{xy}$

13 Yes. By the Law of Large numbers, $\bar{x}_n \xrightarrow{P} \mu$ and $\bar{y}_n \xrightarrow{P} \mu$. Then by continuous mapping (and the stack theorem),

Don't have to say this

$$t_n = \alpha \bar{x}_n + (1-\alpha)\bar{y}_n$$

$$\xrightarrow{P} \alpha\mu + (1-\alpha)\mu = \mu$$

14

(a) $\theta = (\beta, \mu_x, \sigma_x^2, \sigma_\varepsilon^2)$

(b) $\Theta = \{(\beta, \mu_x, \sigma_x^2, \sigma_\varepsilon^2): -\infty < \beta < \infty, -\infty < \mu_x < \infty, \sigma_x^2 > 0, \sigma_\varepsilon^2 > 0\}$

(c) $\begin{pmatrix} x_i \\ \varepsilon_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \right)$

(d) $\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \begin{pmatrix} x_i \\ \varepsilon_i \end{pmatrix}$

$$E\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \begin{pmatrix} \mu_x \\ 0 \end{pmatrix} = \begin{pmatrix} \mu_x \\ \beta \mu_x \end{pmatrix}$$

$$\mathrm{Cov}\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_x^2 & 0 \\ \beta\sigma_x^2 & \sigma_\varepsilon^2 \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \beta^2\sigma_x^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

(e) $E(y_i) = E(\beta x_i + \varepsilon_i) = \beta \mu_x$

$\mathrm{Var}(y_i) = \mathrm{Var}(\beta x_i + \varepsilon_i) = \beta^2 \sigma_x^2 + \sigma_\varepsilon^2$

$$\mathrm{Cov}(x_i, y_i) = \mathrm{Cov}(x_i, \beta x_i + \varepsilon_i)$$
$$= \beta\,\mathrm{Cov}(x_i, x_i) + \mathrm{Cov}(x_i, \varepsilon_i)$$
$$= \beta \sigma_x^2$$

(14f)

Using $E\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \mu_x \\ \beta\mu_x \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and

$$\mathrm{cov}\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \beta^2\sigma_x^2+\sigma_e^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

(i) $\beta = \dfrac{\sigma_{12}}{\sigma_{11}}$ so

$$\hat{\beta}_2 = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{11}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$\xrightarrow{p} \dfrac{\sigma_{12}}{\sigma_{11}}$ By continuous mapping and Law of Large numbers

$\quad\;\; \shortparallel$

$\quad\;\; \beta$

(ii) $\beta = \dfrac{\mu_2}{\mu_1}$, so $\hat{\beta}_1 = \dfrac{\bar{y}}{\bar{x}}$

As long as $\mu_1 = \mu_x \neq 0$,

$$\hat{\beta}_1 \xrightarrow{p} \frac{\mu_2}{\mu_1} = \frac{\beta\mu_x}{\mu_x} = \beta$$ By continuous mapping and LLN.

(iii) It fails at all points in the parameter space where $\mu_x = 0$, because the function $g(x,y) = y/x$ is not continuous at $x = 0$.

14

(f iv)

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(\beta x_i + \varepsilon_i)$$

$$= \beta \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$$

$$= \beta \bar{x}_n + \bar{\varepsilon}_n \quad \text{, and}$$

$$\hat{\beta}_1 = \frac{\bar{y}_n}{\bar{x}_n} = \frac{\beta \bar{x}_n + \bar{\varepsilon}_n}{\bar{x}_n}$$

$$= \beta + \frac{\bar{\varepsilon}_n}{\bar{x}_n}$$

Now $\bar{\varepsilon}_n \sim N(0, \frac{\sigma_\varepsilon^2}{n})$, $\bar{x}_n \sim N(\mu_x, \frac{\sigma_x^2}{n})$

So $\dfrac{\bar{\varepsilon}_n}{\sigma_\varepsilon/\sqrt{n}} \sim N(0,1)$ and

If $\mu_x = 0$, $\dfrac{\bar{x}_n - 0}{\sigma_x/\sqrt{n}} \sim N(0,1)$

And they are independent. So,

$$\hat{\beta}_1 = \beta + \underbrace{\frac{\bar{\varepsilon}_n / \sigma_\varepsilon/\sqrt{n}}{\bar{x}_n / \sigma_x/\sqrt{n}}}_{\text{Cauchy, ratio of independent standard normal RVs.}} \left(\frac{\sigma_x}{\sigma_\varepsilon}\right)$$

So $\hat{\beta}_1 \sim$ Cauchy, but scaled and shifted. Expected value does not exist, and ~~it~~ its distribution does not change with n.

15

(a) $L(\mu, \sigma^2) = \sigma^{-n}(2\pi)^{-n/2}$

$$\times \exp -\frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma^2} + \frac{(\bar{x}-\mu)^2}{\sigma^2}\right)$$

(b) $L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$

$$= \sigma^{-n}(2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$= \sigma^{-n}(2\pi)^{-\frac{n}{2}}$$
$$\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\right)$$

$$= \sigma^{-n}(2\pi)^{-\frac{n}{2}}$$
$$\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((x_i-\bar{x})^2 + 2(\bar{x}-\mu)(x_i-\bar{x}) + (\bar{x}-\mu)^2\right)\right)$$

$$= \sigma^{-n}(2\pi)^{-n/2}$$
$$\times \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + 2(\bar{x}-\mu)\underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{=0} + n(\bar{x}-\mu)^2\right)\right)$$

$$= \sigma^{-n}(2\pi)^{-n/2}\exp -\frac{1}{2\sigma^2}\left(n\hat{\sigma}^2 + n(\bar{x}-\mu)^2\right)$$

$$= \sigma^{-n}(2\pi)^{-n/2}\exp\left(-\frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma^2} + \frac{(\bar{x}-\mu)^2}{\sigma^2}\right)\right)$$

Same as (a)

15

(c) For any σ^2,

$$L(\mu, \sigma^2) = \sigma^{-n}(2\pi)^{-\frac{n}{2}} \exp -\frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma^2} + \frac{(\bar{x}-\mu)^2}{\sigma^2}\right)$$

will be larger when $\frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma^2} + \frac{(\bar{x}-\mu)^2}{\sigma^2}\right)$

is smaller. Minimum occurs when

$(\bar{x}-\mu)^2 = 0 \iff \mu = \bar{x}$.

(16) To find the MLE, maximize

$$L(\theta, \mu) = \prod_{i=1}^{n} \frac{\theta e^{\theta(x_i-\mu)}}{(1+e^{\theta(x_i-\mu)})^2}$$

$$= \frac{\theta^n e^{\theta \sum_{i=1}^{n}(x_i-\mu)}}{\prod_{i=1}^{n}(1+e^{\theta(x_i-\mu)})^2}$$

$$= \frac{\theta^n e^{\theta(n\bar{x}-n\mu)}}{\prod_{i=1}^{n}(1+e^{\theta(x_i-\mu)})^2}$$

Or equivalently, maximize

$\ln L(\theta, \mu)$

16 Continued

$$\ln L(\theta,\mu) = \ln\left(\frac{\theta^n e^{n\theta(\bar{x}-\mu)}}{\prod_{i=1}^{n}\left(1+e^{\theta(x_i-\mu)}\right)^2}\right)$$

$$= n\ln\theta + n\theta(\bar{x}-\mu) - \sum_{i=1}^{n} 2\ln\left(1+e^{\theta(x_i-\mu)}\right)$$

Or minimize the minus Log likelihood

$$2\sum_{i=1}^{n}\ln\left(1+e^{\theta(x_i-\mu)}\right) - n\ln\theta - n\theta(\bar{x}-\mu)$$

R work for Question 16

```
> # MLE for Mystery Distribution
> rm(list=ls()); options(scipen=999)
> mystery = scan("https://www.utstat.toronto.edu/brunner/openSEM/data/mystery2.data.txt")
Read 200 items
> # Mystery minus log likelihood function
> mmll = function(param,x)
+     {
+     theta = param[1]; mu = param[2]
+     n = length(x); xbar = mean(x)
+     value = 2 * sum(log(1 + exp(theta*(x-mu)))) -
+             n*log(theta) - n*theta*(xbar-mu)
+     return(value)
+ } # End of function mmll
>
> msearch = optim(par=c(1,0), fn = mmll,
+                 method = "L-BFGS-B", lower = c(0,-Inf), hessian=TRUE, x=mystery)
> msearch

$par
[1] 2.849444 1.978644

$value
[1] 192.4589

$counts
function gradient
      14       14

$convergence
[1] 0

$message
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"

$hessian
           [,1]        [,2]
[1,] 34.5927653   0.6156322
[2,]  0.6156322 549.5388601

> # (a) MLE
> thetahat = msearch$par[1]; muhat = msearch$par[2]
> c(thetahat,muhat)
[1] 2.849444 1.978644

> # (b) Confidence interval for theta
> Vhat = solve(msearch$hessian)
> se = sqrt(diag(Vhat))
> se_thetahat = se[1]; se_muhat = se[2]
> low95 = thetahat - 1.96*se_thetahat
> upr95 = thetahat + 1.96*se_thetahat
> c(low95,upr95)
[1] 2.516196 3.182693
>
> # (c) Test H0: mu = 2.1. Reject if |z| > 1.96
> z = (muhat-2.1)/se_muhat; z
[1] -2.844819
```