STA431s23 Assignment 4

1. (a) $\theta = (\beta_0, \beta_1, \beta_2, \mu_{x1}, \mu_{x2}, \phi_{11}, \phi_{22}, \phi_{12}, \psi)$

(b) $\Theta = \{(\beta_0, \beta_1, \beta_2, \mu_{x1}, \mu_{x2}, \phi_{11}, \phi_{22}, \phi_{12}, \psi):$
$-\infty < \beta_j < \infty, -\infty < \mu_{x_j} < \infty, \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}$ positive definite, $\psi > 0 \}$

(c) $\Theta = \{(\text{etc.}) : \beta_1 = \beta_2, \phi_{11} = \phi_{22} = \psi = 1, -1 < \phi_{12} < 1\}$

(d)
$$\begin{array}{c} \begin{array}{ccccccccc} \beta_0 & \beta_1 & \beta_2 & \mu_1 & \mu_2 & \phi_{11} & \phi_{22} & \phi_{12} & \psi \end{array} \\ \underbrace{\begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{L} \end{array} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \mu_{x1} \\ \mu_{x2} \\ \phi_{11} \\ \phi_{22} \\ \phi_{12} \\ \psi \end{pmatrix}}_{\theta} = \underbrace{\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{h}$$

(2) (a) (i) $L\hat{\theta}_n \stackrel{\circ}{\sim} N_r(L\theta, LV_nL^T)$

(ii) $(L\hat{\theta}-L\theta)^T(LV_nL^T)^{-1}(L\hat{\theta}-L\theta) \sim \chi^2(r)$

(iii) $r \times r$

(iv) so that LV_nL^T will have an inverse (the rank of a product is the minimum of the ranks).

(v.) The formula sheet has h in place of $L\theta$. This is because H_0 is $L\theta = h$, and the distribution of W_n is what holds when H_0 is true. Also, $\hat{W}_n$ is used in place of W_n, because you never know the variance.

(b) $a^T\hat{\theta}_n \stackrel{\circ}{\sim} N(a^T\theta, a^TV_na)$

(c) $z = \dfrac{a^T\hat{\theta}_n - a^T\theta}{\sqrt{a^TV_na}} \stackrel{\circ}{\sim} N(0,1)$, so

$$1-\alpha = P_r\left\{-z_{\alpha/2} < \frac{a^T\hat{\theta}_n - a^T\theta}{\sqrt{a^T\hat{V}_na}} < z_{\alpha/2}\right\}$$

$$= P_r\left\{-z_{\alpha/2}\sqrt{a^T\hat{V}_na} < a^T\hat{\theta}_n - a^T\theta < \sqrt{a^T\hat{V}_na}\, z_{\alpha/2}\right\}$$

$$= \cdots = P_r\left\{\underbrace{a^T\hat{\theta}_n - z_{\alpha/2}\sqrt{a^T\hat{V}_na}}_{\text{Lower Limit}} < a^T\theta < \underbrace{a^T\hat{\theta}_n + z_{\alpha/2}\sqrt{a^T\hat{V}_na}}_{\text{Upper Limit}}\right\}$$

Using $\hat{V}_n$ because V_n is unknown.

(2d) $$z = \frac{a^T \hat{\theta}_n - h}{\sqrt{a^T \hat{V}_n a}}$$

$$W_n = \underset{1\times 1}{\underset{\uparrow}{(a^T \hat{\theta}_n - h)^T}} \; \underset{1\times 1}{\underset{\uparrow}{(a^T \hat{V}_n a)^{-1}}} \; \underset{1\times 1}{\underset{\uparrow}{(a^T \hat{\theta}_n - h)}}$$

$$= \frac{(a^T \hat{\theta}_n - h)^2}{a^T \hat{V}_n a} = z^2$$

(3)

$$L(\alpha,\beta) = \prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1}$$

$$= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1} \left(\prod_{i=1}^{n}(1-x_i)\right)^{\beta-1}$$

$$\ell(\alpha,\beta) = \ln L(\alpha,\beta) = n\left(\ln\Gamma(\alpha+\beta) - \ln\Gamma(\alpha) - \ln\Gamma(\beta)\right)$$
$$+ (\alpha-1)\sum_{i=1}^{n} \ln x_i + (\beta-1)\sum_{i=1}^{n} \ln(1-x_i)$$, so

$$-\ell_n(\alpha,\beta) = n\left(\ln\Gamma(\alpha) + \ln\Gamma(\beta) - \ln\Gamma(\alpha+\beta)\right)$$
$$-(\alpha-1)\sum_{i=1}^{n} \ln x_i - (\beta-1)\sum_{i=1}^{n} \ln(1-x_i)$$

R work for Question 3

```
> # Q3: Numerical MLE for beta
>
> rm(list=ls())
> bdata = scan("https://www.utstat.toronto.edu/brunner/openSEM/data/beta24.data.txt")
Read 500 items
> # Beta minus log likelihood
> bmll = function(ab,xx)
+     {
+     nn = length(xx); a = ab[1]; b = ab[2]
+     value = nn*(lgamma(a)+lgamma(b) - lgamma(a+b)) -
+             (a-1)*sum(log(xx)) - (b-1)*sum(log((1-xx)))
+     return(value)
+     } # End of function bmll
>
> bsearch = optim(par=c(1,1), fn = bmll,
+                 method = "L-BFGS-B", lower = c(0,0), hessian=TRUE, xx=bdata)
> bsearch

$par
[1] 1.956054 4.026869

$value
[1] -184.783

$counts
function gradient
      12       12

$convergence
[1] 0

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

$hessian
          [,1]      [,2]
[1,] 240.64932 -90.94232
[2,] -90.94232  49.90191
```

```
> # (a) MLE
> thetahat = bsearch$par
> thetahat # (alphahat, betahat)
[1] 1.956054 4.026869

> # (b) Likelihood ratio test of H0: beta = 2 alpha
> # Search restricted parameter space
> bmll0 = function(beta,datta) bmll(c(beta/2,beta), xx=datta)
> bsearch0 = optim(par=1, fn = bmll0, method = "L-BFGS-B", lower = 0, datta=bdata)
> bsearch0

$par
[1] 3.943671

$value
[1] -184.458

$counts
function gradient
       9        9

$convergence
[1] 0

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

>
> # (c) Test H0: mu = 2.1. Reject if |z| > 1.96
> z = (muhat-2.1)/se_muhat; z
[1] -2.844819

> Gsq = 2 * (bsearch0$value - bsearch$value)
> dfree=1
> pval = 1-pchisq(Gsq,dfree)
> c(Gsq,dfree,pval)
[1] 0.6498473 1.0000000 0.4201673

>
> # (c) Wald test
> source("https://www.utstat.toronto.edu/brunner/openSEM/fun/Wtest.txt")
> L = rbind(c(2,-1))
> Vhat = solve(bsearch$hessian)
> Wtest(L,Tn=thetahat,Vn=Vhat)
        W        df   p-value
0.6436909 1.0000000 0.4223774

>
> # (d) CI for 2 alpha - beta
> a = rbind(2,-1)
> se = as.numeric(sqrt(t(a)%*%Vhat%*%a)); se
[1] 0.1430393

> est = as.numeric(t(a)%*%thetahat); est
[1] -0.114761

> # (est/se)^2 # Wald stat
> lower95 = est - 1.96*se; upper95 = est + 1.96*se
> c(lower95, upper95)
[1] -0.3951180  0.1655961
```

R work for Question 4

```
> # Q4: Simulation from simple regression through the origin
>
> rm(list=ls())
> # Set sample size and parameter values
> n = 1000; beta = 1 ; mux = 0 ; sigmasqx = 2 ; sigmasqepsilon = 3
>
> # (a) Simulate from the model
> set.seed(9999)
> x = rnorm(n,mux,sqrt(sigmasqx))
> epsilon = rnorm(n,0,sqrt(sigmasqepsilon))
> y = beta*x + epsilon
>
> # (b) Estimate beta
> betahat1 = mean(y)/mean(x)
> betahat2 = var(x,y) / var(x)
> c(betahat1,betahat2)

[1] 25.870249  1.048945


> lm(y~x) # Checking betahat2

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    0.06713      1.04894
```

(5) (a) $E(\hat{\beta} \mid X) = E\{(X^T X)^{-1} X^T y \mid X\}$

$= (X^T X)^{-1} X^T E\{y \mid X\} = (X^T X)^{-1} X^T E(X\beta + \varepsilon \mid X)$

$= (X^T X)^{-1} X^T (X\beta + 0) = (X^T X)^{-1} X^T X \beta = \beta$

conditionally unbiased

(b) $E(\hat{\beta}) = E(E\{\hat{\beta} \mid X\}) = E(\beta) = \beta$

(c) $P(F > f_c) = \sum_x \cdots \sum P\{F > f_c \mid \mathcal{X} = x\} P(\mathcal{X} = x)$

$$= \sum_x \cdots \sum \alpha \, P(\mathcal{X} = x)$$

$$= \alpha \sum_x \cdots \sum P(\mathcal{X} = x)$$

$$= \alpha \cdot 1 = \alpha$$

(6) (a) $\ell(\theta_1, \theta_2) = \log L(\theta_1, \theta_2)$

$$= \log\left(\prod_{i=1}^{n} g_{\theta_1}(y_i | x_i)\, h_{\theta_2}(x_i)\right)$$

$$= \log\left(\prod_{i=1}^{n} g_{\theta_1}(y_i | x_i) \prod_{i=1}^{n} h_{\theta_2}(x_i)\right)$$

$$= \sum_{i=1}^{n} \log g_{\theta_1}(y_i | x_i) + \sum_{i=1}^{n} \log h_{\theta_2}(x_i)$$

The left-hand term is the log likelihood for a conditional model. It is clear that if you differentiate with respect to any element of θ_1, the second term equals zero and has no effect. As a result, the MLE $\hat{\theta}_1$ is the same for the conditional and unconditional models.

(b) By the work above $L(\theta) = L_1(\theta_1) L_2(\theta_2)$, so

$$G^2 = -2 \log \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = -2 \log \frac{L_1(\hat{\theta}_{1,0})\, \cancel{L_2(\hat{\theta}_2)}}{L_1(\hat{\theta}_1)\, \cancel{L_2(\hat{\theta}_2)}}$$

Because H_0 places no restriction on θ_2

This is the likelihood ratio test statistic for the conditional model.