Homework Set #6 Solutions

(1.) $\hat{\pi} \approx N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$, $g(\hat{\pi}) \approx N\left(g(\pi), [g'(\pi)]^2 \frac{\pi(1-\pi)}{n}\right)$

$$g(t) = \log\left(\frac{t}{1-t}\right) = \log t - \log(1-t)$$

$$g'(t) = \frac{1}{t} + \frac{1}{1-t} = \frac{1}{t(1-t)}$$

$$\sigma^2\left(\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right)\right) = \left[\frac{1}{\pi(1-\pi)}\right]^2 \cdot \frac{\pi(1-\pi)}{n} = \underline{\frac{1}{\pi(1-\pi)} \cdot \frac{1}{n}}$$

(2.) retrospective study:

alcohol	cancer	no cancer
high consumption	71	82
low consumption	60	441

(a) $\hat{\sigma}(\log \hat{\theta}) = \left(\frac{1}{n_{11}} + \frac{1}{n_{21}} + \frac{1}{n_{12}} + \frac{1}{n_{22}}\right)^{\frac{1}{2}}$

(b) The equation for $\hat{\sigma}(\log \hat{\theta})$ does not depend on the sampling scheme. When asymptotic normality holds, replacing parameters with statistics automatically invokes the likelihood principle.

(c) $\log \hat{\theta} = \log\left(\frac{71 \cdot 441}{82 \cdot 60}\right) = \log(6.364) = 1.8507$

$$\hat{\sigma}(\log \hat{\theta}) = \left(\frac{1}{71} + \frac{1}{82} + \frac{1}{60} + \frac{1}{441}\right)^{\frac{1}{2}} = 0.2126$$

95% CI for $\log \theta = 1.8507 \pm 1.96(0.2126)$

$= 1.8507 \pm 0.4167 = \underline{[1.434, 2.2674]}$

(2) continued

(d) $$\hat{\gamma} = \frac{n_{11}n_{22} - n_{21}n_{12}}{n_{11}n_{22} + n_{21}n_{22}} = \frac{71(441) - 82(60)}{71(441) + 82(60)}$$

$$= \frac{31311 - 4920}{31311 + 4920} = .73$$

we estimate that there is a large size positive association between alcohol consumption and the onset of esophagus cancer.

(3.) refer to cross-sectional ~~data~~ study of the association between Husband's Rating and Wife's Rating

Interval estimate for $\gamma = [.100, .506]$, $\gamma_{med} = .315$

```
> library(dirmult)
> dimR = 4
> dimC = 4
> counts = c(7,7,2,3,2,8,3,7,1,5,4,9,2,8,9,14)
> dir.post = counts+1
> simN = 10000
> gamma = rep(NA,simN)
> sim = 1
> for (sim in 1:simN){
+   p = rdirichlet(1,dir.post)
+   probs = matrix(p,nrow = dimR,byrow = TRUE)
+
+   con = 0
+   i=1
+   j=1
+   for (i in 1:(dimR-1)) {
+     for (j in 1:(dimC-1)) {
+       sub = 0
+       h=i+1
+       while (h <= dimR) {
+         k=j+1
+         while (k <= dimC) {
+           sub = sub + probs[h,k]
+           k=k+1
+         }
+         h=h+1
+       }
+       con = con + probs[i,j]*sub
+       j=j+1
+     }
+     i=i+1
+   }
+
+     dis = 0
+   i=1
+   j=1
+   for (i in 1:(dimR-1)) {
+     for (j in 2:dimC) {
+       sub = 0
+       h=i+1
+       while (h <= dimR) {
+         k=j-1
+         while (k >= 1) {
+           sub = sub + probs[h,k]
+           k=k-1
+         }
+         h=h+1
+       }
+       dis = dis + probs[i,j]*sub
+       j=j+1
+     }
+     i=i+1
+   }
+
+   gamma[sim] = (con-dis)/(con+dis)
+
+   sim = sim+1
+ }
> hist(gamma,probability = TRUE)
> points(density(gamma),type = 'l')
>
> quantile(gamma,c(.025,.25,.5,.75,.975))
     2.5%       25%       50%       75%     97.5%
0.1005011 0.2436756 0.3152427 0.3841113 0.5056300
```