Discrete Multivariate Analysis - 579 - HW #5

Alex Towell (atowell@siue.edu)

Problem 1

Consider data from a prospective study on the relationship between daily aspirin use and the onset of heart disease.

	Disease	No Disease	Total
Placebo	28	656	684
Aspirin	18	658	676

Part (a)

Provide an equation for $\sigma^2\left(\log \hat{\text{RR}}\right)$.

$$
\begin{aligned}
\sigma^2\left(\log \hat{\text{RR}}\right) &= \frac{1}{n_{11}} - \frac{1}{n_{1+}} + \frac{1}{n_{21}} - \frac{1}{n_{2+}} \\
&= \frac{1}{28} - \frac{1}{684} + \frac{1}{18} - \frac{1}{658} \\
&\approx 0.883 \, .
\end{aligned}
$$

Part (b)

Provide an equation for $\sigma\left(\log \hat{\text{RR}}\right)$.

$$
\sigma\left(\log \hat{\text{RR}}\right) = \sqrt{\sigma^2(\log \hat{\text{RR}})} \approx 0.297 \, .
$$

Part (c)

Compute a 95% confidence interval for $\log \text{RR}$.

The relative risk RR is defined as

$$
\text{RR} := \frac{\pi_1}{\pi_2}
$$

The MLE of $\log \text{RR}$ is given by

$$
\hat{\text{RR}} = \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{n_{11}}{n_{1+}} \frac{n_{2+}}{n_{21}} = \frac{28}{676} \times \frac{684}{18} \approx 1.574 \, .
$$

By the invariance property of the MLE, if $\hat{\theta}$ is an MLE of θ then $\text{g}(\theta)$ is an MLE of $\text{g}(\hat{\theta})$. Thus,

$$
\log \hat{\text{RR}} = \log \hat{\text{RR}} \approx 0.454 \, .
$$

A confidence interval for $\log \mathrm{RR}$ is thus

$$\mathrm{CI}(\log \mathrm{RR}) = \log \hat{\mathrm{R}}\mathrm{R} \pm z_{\alpha/2}\sigma(\log \hat{\mathrm{R}}\mathrm{R}).$$

Letting $\alpha = 0.05$ and substituting in the values for $\log \hat{\mathrm{R}}\mathrm{R}$ and $\sigma(\log \hat{\mathrm{R}}\mathrm{R})$, we get the result

$$\mathrm{CI}(\log \hat{\mathrm{R}}\mathrm{R}) = 0.454 \pm 1.96 \times 0.297 = [-0.128, 1.036].$$

Part (d)

Compute a 95% confidence interval for RR.

We take the inverse of the logarithm and obtain the result

$$\mathrm{CI}(\mathrm{RR}) = [e^{-0.128}, e^{1.036}] = [0.880, 2.818].$$

Problem 2

A diagnostic test for Covid antibodies is being studied. For a sample of $n = 122$ specimens with antibodies known to be present, the test returned $y = 103$ positive results. For a sample of $n = 401$ specimens absent any antibodies, the test returned $y = 399$ negative results.

Function used to compute interval estimates for this problem.

```
# this program performs Bayesian inference for a binomial probability
#   - y is the number of successes
#   - n is the number of trials
#   - alpha is the sum of the tail probabilities
#
# we also compute a classical interval using normality assumptions of the mle.
binomial_bayesian <- function(y,n,alpha=.05) {
  # compute the maximum likelihood estimator
  mle = y/n
  mle_var = mle*(1-mle)/n
  mle_sd = sqrt(mle_var)

  t0 = qt(1-alpha/2, n-1)
  L = mle - mle_sd*t0
  U = mle + mle_sd*t0

  # define the beta distribution parameters
  a=y+1
  b=n-y+1

  # computing a confidence interval by taking the upper and lower percentiles of
  # the beta distribution computing the median to represent the center of the
  # distribution
  lower = qbeta(alpha/2,a,b)
  median = qbeta(.5,a,b)
  upper = qbeta(1-alpha/2,a,b)
  cat("bayesian estimate: ", c(lower,median,upper),"\n")
  cat("mle confidence interval estimate: ", c(L,mle,U),"\n")

  # create a grid of p for plotting. you may change to a smaller range than from
```

2

```
# 0 to 1 to see the distribution better
p = seq(from=max(0,lower*.965),to=min(1,1.035*upper),length.out=100)
# we are computing the beta density at each point in the grid
posterior = dbeta(p,a,b)

ps = rnorm(n=1000000,mean=mle,sd=mle_sd)

# plotting the beta density, higher values of the curve represent stronger data
# evidence
plot(p,posterior,type = "l",col="blue")

lines(density(ps),col="orange")

points(median,0,col="blue",pch="*")
points(c(lower,upper),c(0,0),,col="blue",pch="|")

# let's plot the mle and confidence interval using t-score
points(mle,0,col="orange",pch="*")
points(c(L,U),c(0,0),,col="orange",pch="|")
}
```

Part (a)

Compute a Bayesian/likelihood interval estimate for the sensitivity, $\delta = \Pr(+|A)$.

Given that antibodies are present, the probability that a positive test result is observed is given by the sensitivity δ.

In classical statistics, while the probability δ is not known, the MLE estimator is $\hat{\delta} = \frac{y}{n}$ which is asymptotically normal under regularity conditions.

We compare it with the Bayesian approach, which does not assume normality. The following R function call computes the Bayesian interval estimate for a binomial probability. For comparison, we also compute the confidence interval for the MLE using classical techniques that rely upon the normality assumption of the MLE.

```
binomial_bayesian(y=103,n=122,alpha=.05)
```

```
## bayesian estimate:  0.7693298 0.8405335 0.8977577
## mle confidence interval estimate:  0.7792689 0.8442623 0.9092557
```
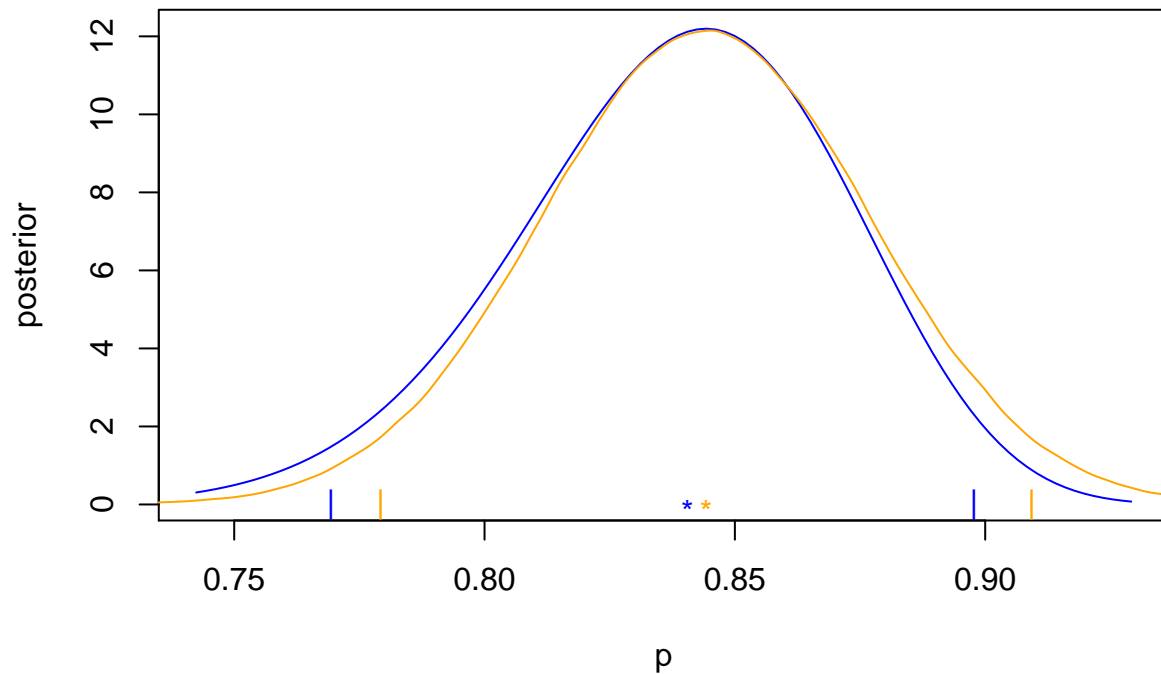
The maximum likelihood estimator $\hat{\delta}$ is 0.844 with an $\alpha = 0.05$ confidence interval $[0.779, 0.909]$.

The point estimator of δ using the Bayesian approach, the 50%-percentile, is 0.841 with an interval estimate $[0.769, 0.898]$.

The maximum likelihood estimator and its corresponding density plot and confidence interval is plotted in *orange* and the Bayesian estimator is plotted in *blue*.
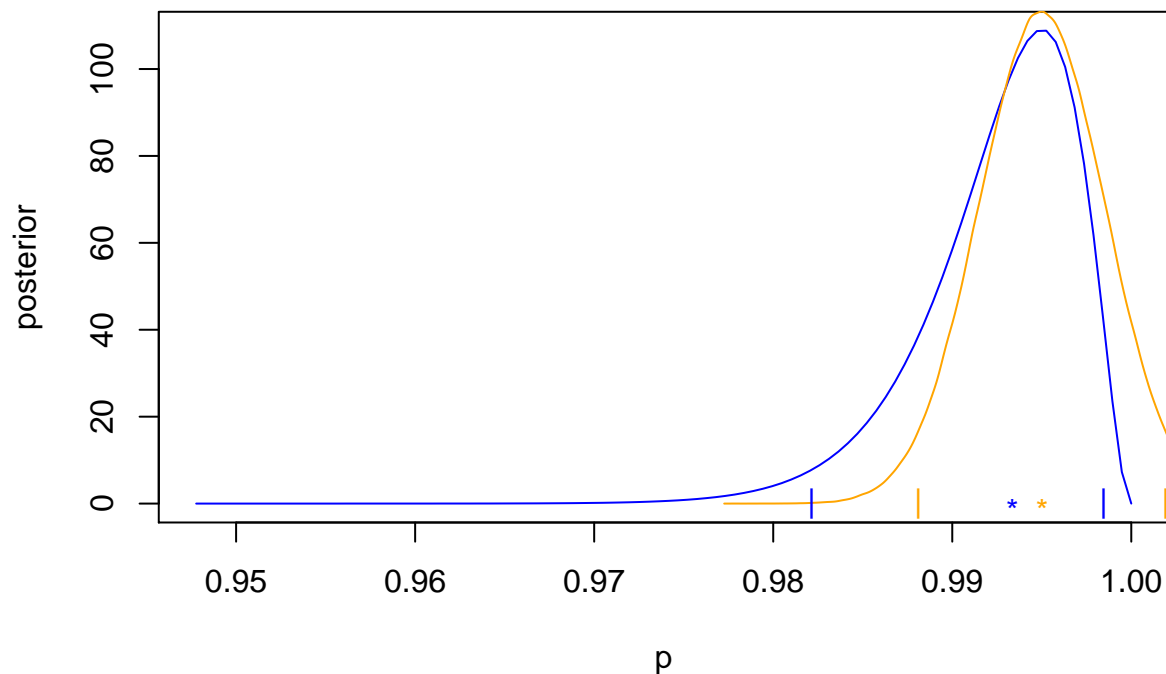
Part (b)

Compute a Bayesian/likelihood interval estimate for the sensitivity, $\gamma = \Pr(+|A)$.

We do the same thing as in part (A).

```
binomial_bayesian(y=399,n=401,alpha=.05)
```

```
## bayesian estimate:  0.9821445 0.9933537 0.9984584
## mle confidence interval estimate:  0.9880966 0.9950125 1.001928
```

The maximum likelihood estimator $\hat{\delta}$ is 0.844 with an $\alpha = 0.05$ confidence interval $[0.982, 1.002]$

The point estimator of δ using the Bayesian approach, the 50%-percentile, is 0.993 with a $\alpha = 0.05$ interval estimate $[0.982, 0.998]$.

The maximum likelihood estimator and its corresponding density plot and confidence interval is plotted in *orange* and the Bayesian estimator is plotted *blue*.

Problem 3

> The clinical trial for the Moderna vaccine resulted in $n_I = 196$ cases of symptomatic infection, with $y_C = 185$ cases from the control group, and $y_V = 11$ cases from the vaccine group. Compute a Bayesian/likelihood interval estimate for the vaccine efficacy, VE.

The interval estimate is computed by the following R code.

```
# this program performs a Bayesian inference for vaccine efficacy
# enter the number of infections in the vaccine group
# and the number of infections in the control group
v = 11
c = 185

# define the beta distribution parameters
a = v+1
b = c+1

# simulate a very large number of draws from the posterior distribution on p=P(V|I)
# for each simulated p, compute the value for vaccine efficacy VE
p = rbeta(1000000,a,b)
efficacy = 1-p/(1-p)

# create a plot of the posterior distribution on VE
hist(efficacy,xlab="Efficacy",probability = TRUE)
```
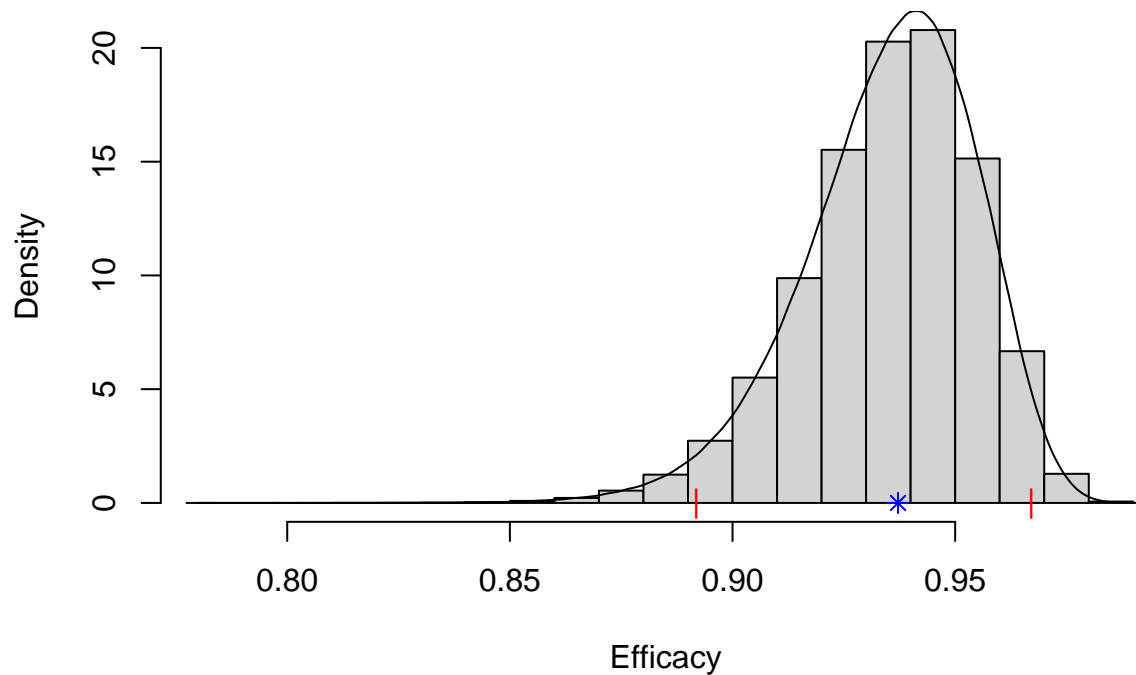
```
points(density(efficacy),type = 'l')

pt_est = quantile(efficacy,c(.5))
points(pt_est,0,col="blue",pch=8)

# compute percentiles for a 95% interval estimate
interval_est = quantile(efficacy,c(.025,.975))
points(interval_est,c(0,0),col="red",pch="|")
```

Histogram of efficacy


```
print(pt_est)
```

```
##       50%
## 0.9371664
```

```
print(interval_est)
```

```
##      2.5%     97.5%
## 0.8918323 0.9670740
```

We see that the density plot is not quite normal (e.g., non-symmetric). On the plot, we plotted the 50% in blue and the 95% interval estimate in red.