Homework 4

Senan Hogan-H.

29 March 2018

Question 1.

a. $x = 0, 1, 2.$

```
freq.test <-
  prop.test(3, 12, p=0.5, alt='less', conf.level = 0.95)
freq.test$p.value < 0.05
```

```
## [1] FALSE
```

```
freq.test <-
  prop.test(2, 12, p=0.5, alt='less', conf.level = 0.95)
freq.test$p.value < 0.05
```

```
## [1] TRUE
```

b.

$$H_0 : p \geq .5 \qquad H_1 : p < .5$$

Suppose 3 successes out of 12 were observed.

```
frequentist.test <-
  prop.test(3, 12, p=0.5, alt='less', conf.level = 0.95)

frequentist.test
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  3 out of 12, null probability 0.5
## X-squared = 2.1, df = 1, p-value = 0.07
## alternative hypothesis: true p is less than 0.5
## 95 percent confidence interval:
##  0.0000 0.5287
## sample estimates:
##    p
## 0.25
```

```
frequentist.test$p.value
```

```
## [1] 0.07446
```

```
frequentist.test$p.value < 0.05
```

```
## [1] FALSE
```

```
frequentist.test$p.value < 0.1
```

```
## [1] TRUE
```

Results are significant under frequentist paradigm at the 10% level, maybe publish.

c. Suppose now that the data was distributed according to a negative binomial.

```
new.test <- pnbinom(12-3, size = 12, prob = 0.5, lower.tail = TRUE)
new.test
```

```
## [1] 0.3318
```

```
new.test < 0.05
```

```
## [1] FALSE
```

```
new.test < 0.1
```

```
## [1] FALSE
```

Results are not significant at any standard level under frequentist paradigm, no publishing will be done today.

d. The ambiguity in data collection methods has made me less confident in frequentist tests of significance.

e. Under a Bayesian setting, the likelihood of observing the data given in (b) and (c) change in the same as in a frequentist setting. This means that (b) will be considered less likely than (c) when a (non-stupid) prior is available, giving the same result as the frequentist setting.

Question 2.

Let $y|\theta \sim$Pois(θ) and assume a prior on θ of Gamma(1,1).

a.

$$H_0 : \theta \leq 1 \qquad H_1 : \theta > 1$$

The posterior is given by Gamma$(1 + n\bar{Y}, 1 + n)$.

Consider the cases $y = 3, 4, 5$ are observed separately in a sample size of 1.

```
n <- 1

y <- 3 # data y = 3
sample <- rgamma(10000, 1 + sum(y)/n, 1 + n) # sample from gamma
H0 <- length(sample[sample <= 1])
post_prob_H0 <- H0/length(sample) # How many draws satisfy H0
post_prob_H0
```

```
## [1] 0.1371
```

```
y <- 4 # data y = 4
sample <- rgamma(10000, 1 + sum(y)/n, 1 + n)
H0 <- length(sample[sample <= 1])
post_prob_H0 <- H0/length(sample)
post_prob_H0
```

```
## [1] 0.0546
```

```
y <- 5 # data y = 5
sample <- rgamma(10000, 1 + sum(y)/n, 1 + n)
H0 <- length(sample[sample <= 1])
post_prob_H0 <- H0/length(sample)
post_prob_H0
```

```
## [1] 0.0155
```

b. $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$, with $q_0 = P(H_0) = 0.5$, and $p_1(\theta) = e^{-\theta}$.

$$P(H_0|Y=y) = \frac{P(Y=y|H_0)P(H_0)}{P(Y=y)} =$$

$$\frac{P(\theta=1)P(Y=y|\theta=1)}{P(\theta=1)P(Y=y|\theta=1)+P(\theta\neq 1)P(Y=y|\theta\neq 1)} =$$

$$\frac{P(Y=y|\theta=1)}{P(Y=y|\theta=1)+P(Y=y|\theta\neq 1)} =$$

$$\frac{P(Y=y|\theta=1)}{P(Y=y|\theta=1)+\int_{\theta>0} e^{-\theta}P(Y=y|\theta\neq 1)d\theta} =$$

$$\frac{P(Y=y|\theta=1)}{P(Y=y|\theta=1)+\int_{\theta>0} e^{-\theta}\frac{\theta^{-y}e^{\theta}}{y!}d\theta} =$$

Since $y|\theta \sim$Pois(θ).

```
y <- c() # data to be defined
likelihood <- function(theta) ((1/factorial(y))*(theta^y)*exp(-2*theta))

y <- 3 # data y = 3
likelihood_H0 <- pgamma(y, 1, 1, lower.tail = F)
likelihood_H1 <- integrate(likelihood, 0, Inf)$value
post_prob_H0 <- likelihood_H0/(likelihood_H0 + likelihood_H1)
post_prob_H0
```

```
## [1] 0.4434
```

```
y <- 4 # data y = 4
likelihood_H0 <- pgamma(y, 1, 1, lower.tail = F)
likelihood_H1 <- integrate(likelihood, 0, Inf)$value
post_prob_H0 <- likelihood_H0/(likelihood_H0 + likelihood_H1)
post_prob_H0
```

```
## [1] 0.3695
```

```
y <- 5 # data y = 5
likelihood_H0 <- pgamma(y, 1, 1, lower.tail = F)
likelihood_H1 <- integrate(likelihood, 0, Inf)$value
post_prob_H0 <- likelihood_H0/(likelihood_H0 + likelihood_H1)
post_prob_H0
```

```
## [1] 0.3013
```

Question 3.

a. $H_0: p = 0.5 \qquad H_1: p > 0.5$.

Via the central limit theorem, reject the null if $\hat{p} > 0.5 + 1.645\sqrt{\frac{.25}{n}}$. Consider the case of a coin with $p = 0.501$.

$\hat{p} \sim N(0.501, \sqrt{\frac{0.501(1-0.501)}{n}}) = N(0.501, \sqrt{\frac{0.249999}{n}}) = N(0.501, \frac{0.499999}{\sqrt{n}})$.

It follows that $\hat{p}$ is 95% above the value $0.501 - 1.645\frac{0.499999}{\sqrt{n}}$.

If the null is rejected with 95% certainty, then $0.501 - 1.645\frac{0.499999}{\sqrt{n}} \geq 0.5 + 1.645\sqrt{\frac{.25}{n}}$.

$\Rightarrow 0.501 - 0.5 \geq \frac{1.645}{\sqrt{n}}(0.5 + 0.499999)$

$\Rightarrow 0.001 \geq \frac{1.645}{\sqrt{n}}(0.999999)$

$\Rightarrow \sqrt{n} \geq \frac{1.645}{(0.001)(0.999999)} = 1645$

$\Rightarrow n \geq 2,706,030$

b.

$$H_0 : p < 0.55 \qquad H_1 : p \geq 0.55$$

Observed data $X = 527$ for $n = 1000$.

Suppose $p|H_0 \sim N(0.5, 0.2)$, non-stupid prior.

Suppose $p|H_1 \sim Beta(5, 1)$, non-stupid prior assigning little probability below 0.55 (seen in the following code).

$$BF_{10} = \frac{P(y|H_1)}{P(y|H_0)} = \frac{\int P(y|p)f_1(p)dp}{\int P(y|p)f_0(p)dp}$$

```r
pbeta(0.55, 5, 1, lower.tail = F)
```

```
## [1] 0.9497
```

```r
n <- 1000
X <- 527
p <- 0.5 # define prior H0
sd <- 0.2 # define prior H0
alpha <- 5 # define prior H1
beta <- 1 # define prior H1

sample_H0 <- rnorm(100000, p, sd)
p_H0 <- dbinom(X, size = n, prob = sample_H0)
likelihood_H0 <- mean(na.omit(p_H0))
likelihood_H0
```

```
## [1] 0.001978
```

```r
sample_H1 <- rbeta(100000, alpha, beta)
p_H1 <- dbinom(X, size = n, prob = sample_H1)
likelihood_H1 <- mean(na.omit(p_H1))
likelihood_H1
```

```
## [1] 0.0003672
```

```r
Bayes_factor <- likelihood_H1/likelihood_H0
Bayes_factor
```

```
## [1] 0.1856
```

```r
Bayes_factor < 10^0
```

```
## [1] TRUE
```

The data supports the null hypothesis, that $p < 0.55$, since $BF_{10} < 10^0$.

c.

```r
freq.test <-
  prop.test(X, n, p=0.55, alt='greater', conf.level = 0.95)
freq.test
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  X out of n, null probability 0.55
## X-squared = 2, df = 1, p-value = 0.9
## alternative hypothesis: true p is greater than 0.55
## 95 percent confidence interval:
##  0.5005 1.0000
## sample estimates:
##     p
## 0.527
```

```r
freq.test$p.value
```

```
## [1] 0.9237
```

```r
freq.test$p.value < 0.05
```

```
## [1] FALSE
```

Fail to reject null hypothesis.