Homework 3

Senan Hogan-H.

22 February 2018

Question 3.

   a. Aim to estimate p in a binomial distribution. Start with prior distribution of $Beta(\alpha, \beta)$, where $\alpha = \beta = 1$.

```r
# definitions for prior
n <- 100
alpha <- beta <- 1

#vectors for final storage
coverage_rate <- p_value <- c()

for (p in seq(0, 1, length.out = 1000)){
  coverage <- c()
  for (i in c(1:1000)){
    # draw from the binomial distribution
    x <- sum(rbinom(n, 1, p))

    # form credible interval from posterior distribution.
    CI <- qbeta(c(0.025, 0.975), alpha + x, beta + n - x)

    # document coverage
    ifelse(CI[1] < p & p < CI[2],
          coverage <- c(coverage, 1),
          coverage <- c(coverage, 0))
  }
  coverage_rate <- c(coverage_rate, mean(coverage))
  p_value <- c(p_value, p)
}

data.frame(coverage_rate, p_value) %>%
  ggplot(aes(x=p_value, y=coverage_rate)) +
  geom_hline(yintercept=0.95, linetype='dashed') +
  geom_line() +
  ylim(0.9,1) +
  labs(x= 'p value', y='Coverage Rate') +
  theme_classic()
```
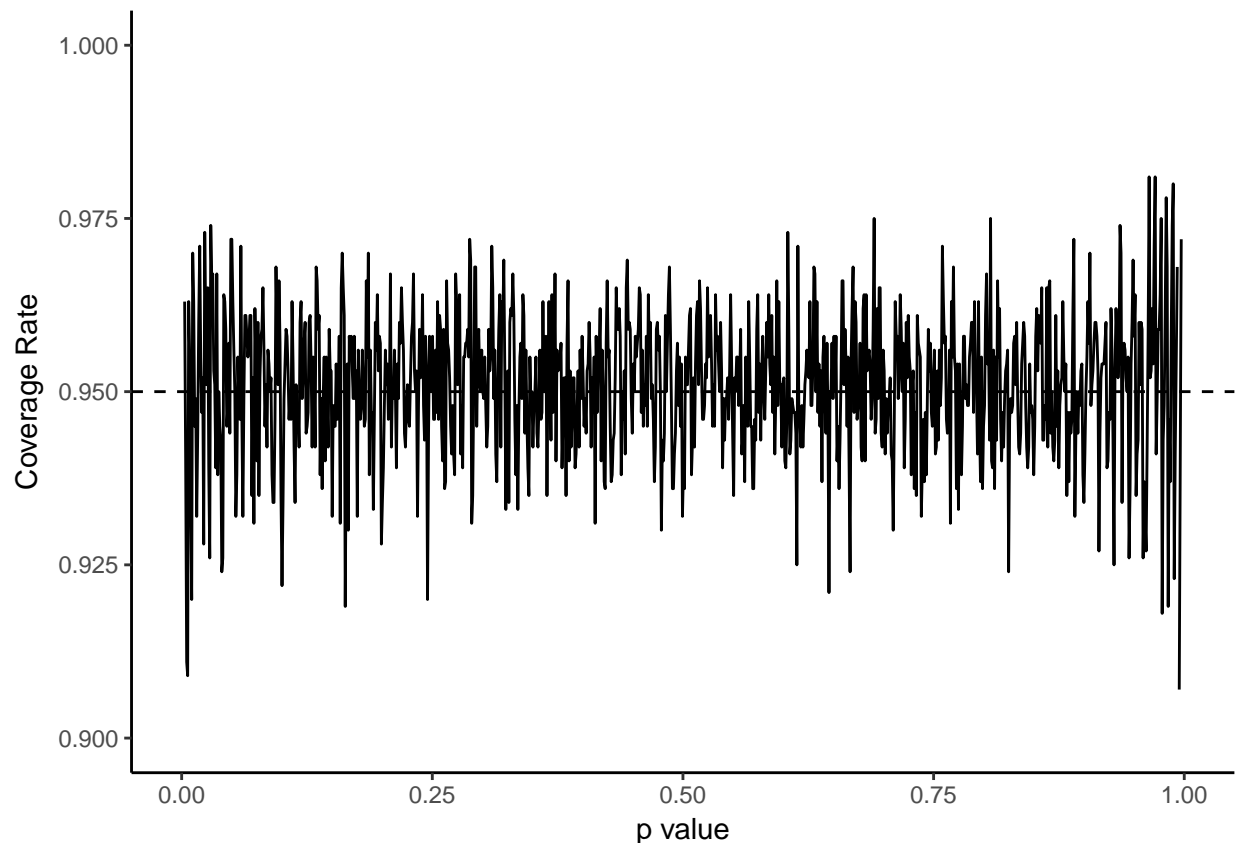
```
## Warning: Removed 2 rows containing missing values (geom_path).
```

The Bayesian credible interval suffers from the same problem to a much lesser extend than the frequentist confidence interval, even from a completely non-informative prior. The main takeway is that the credible interval spends much less time achieving a coverage lower than stated than does the frequentist approach.

   b.

```r
# check average coverage
mean(coverage_rate)
```

```
## [1] 0.9486
```

```r
# Check difference from stated coverage
abs(mean(coverage_rate)-0.95)
```

```
## [1] 0.001391
```

```r
# proportion below stated coverage
below_coverage <- ifelse(
  coverage_rate < 0.95,
  1,
  0)
mean(below_coverage)
```

```
## [1] 0.438
```

The credible interval attains its stated probability, in fact (though barely) at a slighlty higher rate than stated.

This is slighlty dofferent to the sense in which the confidence interval works, nstead referring to coverage probability over a lifetime of experiments.