BDA - Assignment 3

Anonymous

Contents

Exercise 1 - Inference for normal mean and deviation

```
library(aaltobda)
data("windshieldy1")
data("windshieldy2")
```

The observations in the windshieldy-data follow a normal distribution with an unknown standard deviation σ and mean μ. We assume standard uniform prior

$$p(\mu, \sigma) \propto \sigma^{-1}$$

1) Likelihood

$$p(y|\mu, \sigma^2) = N(\mu, \sigma^2)$$

2) Prior

$$p(\mu, \sigma) \propto \sigma^{-1}$$

3) Posterior

$$p(\mu|y) = t_{n-1}(\overline{y}, s^2/n)$$

a)

```
mu_point_est = function(data) {
  estimate = mean(data)
  return(estimate)
}
mu_point_est(windshieldy1)
```

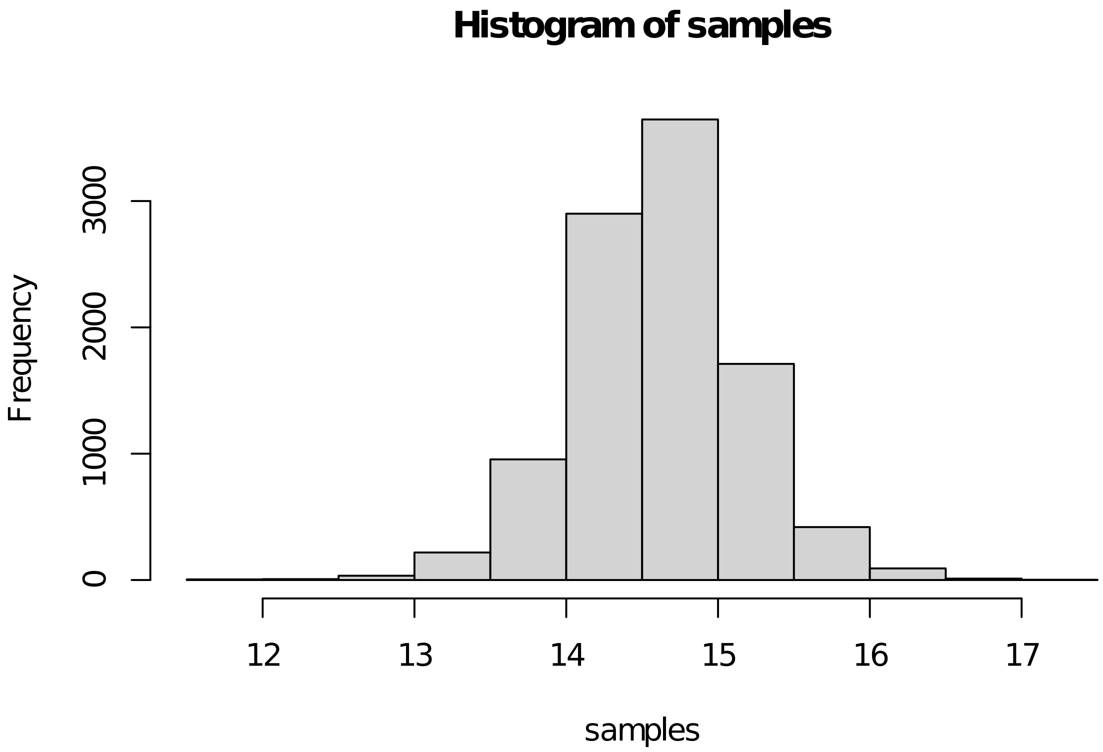
```
## [1] 14.61122
```

```
mu_point_est = function(data, prob) {
  n = length(data)
  mean = mean(data)
  var = (1/(n-1))*sum((data-mean)^2)
```

```
  quantiles = qt(c((1-prob)/2, 1-(1-prob)/2), n-1) * (sqrt(var) / sqrt(n)) + mean(data)
  return(quantiles)
}
mu_point_est(windshieldy1, 0.95)
```

```
## [1] 13.47808 15.74436
```

```
n = length(windshieldy1)
mean = mean(windshieldy1)
var = (1/(n-1))*sum((windshieldy1-mean)^2)
sampleSize = 10000
samples = rtnew(sampleSize, n-1, mean, (sqrt(var) / sqrt(n)))
hist(samples)
```



b)

```
mu_pred_point_est = function(data) {
  estimate = mean(data)
  return(estimate)
}
mu_pred_point_est(windshieldy1)
```

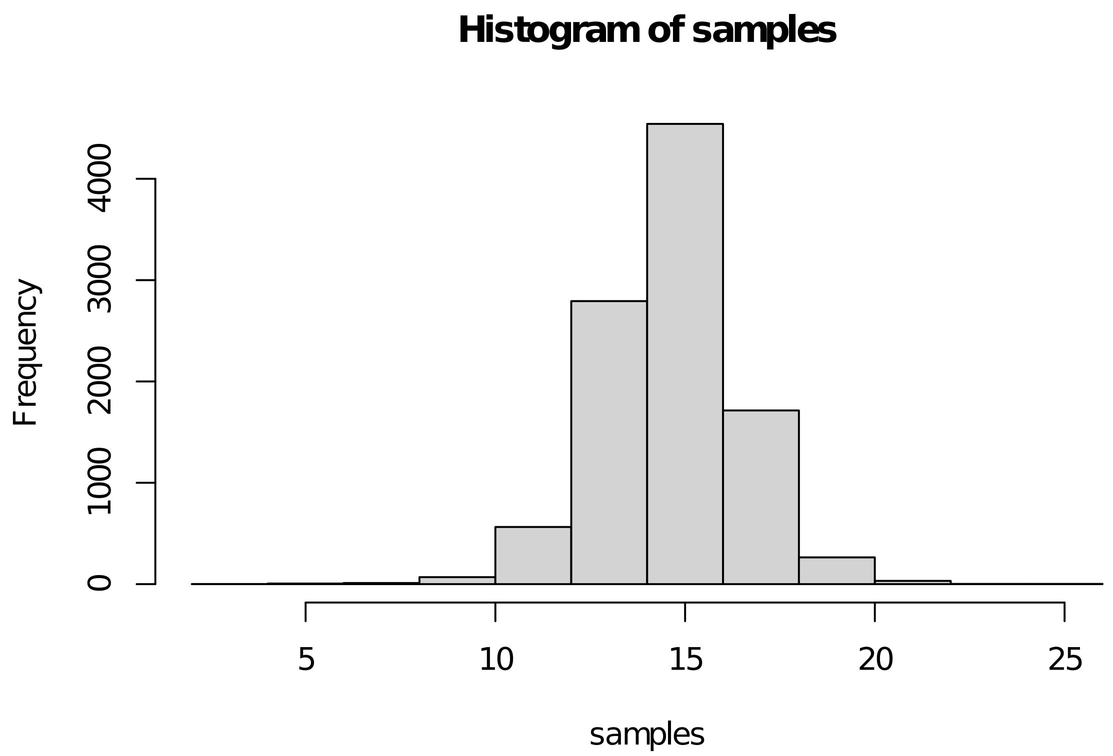
```
## [1] 14.61122
```

```
mu_pred_interval = function(data, prob) {
  n = length(data)
  mean = mean(data)
  var = (1/(n-1))*sum((data-mean)^2)
```

```
  interval = qt(c((1-prob)/2, 1-(1-prob)/2), n-1) * sqrt(1+1/n)*sqrt(var) + mean
  return(interval)
}
mu_pred_interval(windshieldy1, 0.95)
```

```
## [1] 11.02792 18.19453
```

```
n_pred = length(windshieldy1)
mean_pred = mean(windshieldy1)
var_pred = (1/(n_pred-1))*sum((windshieldy1-mean_pred)^2)
sampleSize = 10000
samples = rtnew(sampleSize, n_pred-1, mean_pred, sqrt(1+1/n_pred)*sqrt(var_pred))
hist(samples)
```



Exercise 2 - Inference for the difference between proportions

1.1) Likelihood p0

$$p(y_0|p_0) = \binom{674}{39} p_0^{39}(1-p_0)^{674-39}$$

1.2) Likelihood p1

$$p(y_1|p_1) = \binom{680}{22} p_1^{22}(1-p_1)^{680-22}$$

2.1) Prior p0

$$p(p_0) = Beta(1, 1)$$

2.2) Prior p1

$$p(p_1) = Beta(1, 1)$$

3.1) Posterior p0

$$p(p_0|y_0) = Beta(1 + 39, 1 + 674 - 39) = Beta(40, 636)$$

3.2) Posterior p1

$$p(p_1|y_1) = Beta(1 + 22, 1 + 680 - 22) = Beta(23, 659)$$

a)

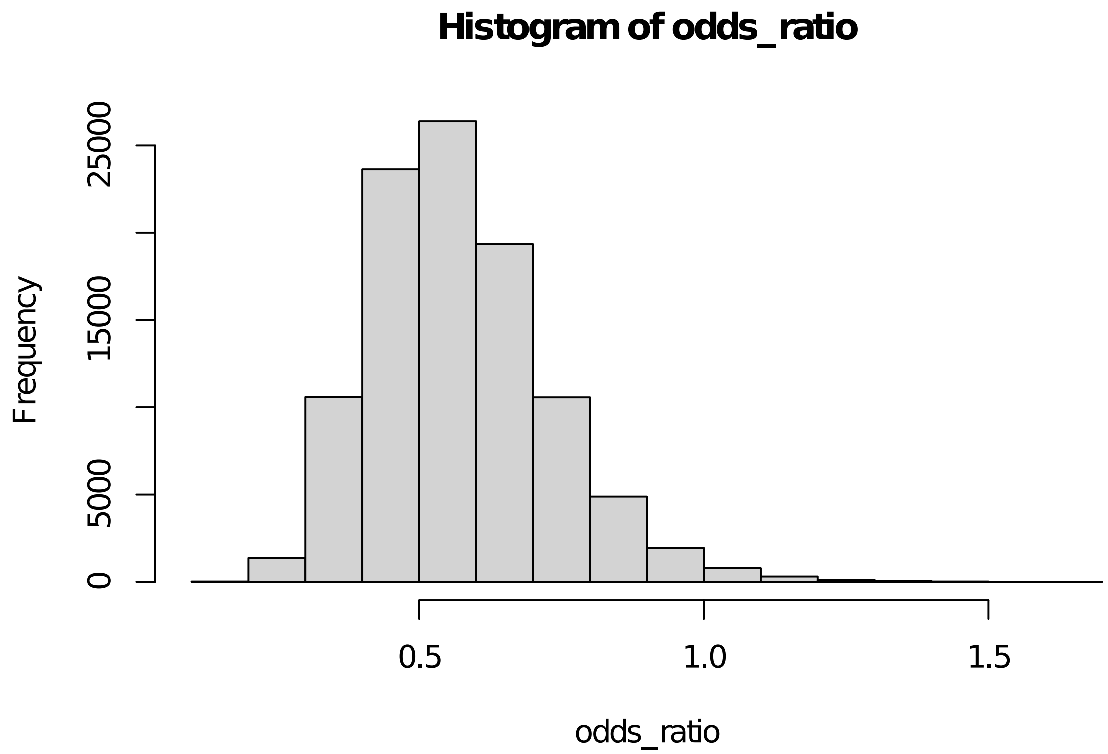
```
set.seed(4711)
sampleSize = 100000
p0 = rbeta(sampleSize, 40, 636)
p1 = rbeta(sampleSize, 23, 659)

posterior_odds_ratio_point_est <- function(p0, p1) {
  odds_ratio = (p1/(1-p1))/(p0/(1-p0))
  estimate = mean(odds_ratio)
  return(estimate)
}
posterior_odds_ratio_point_est(p0 = p0, p1 = p1)
```

```
## [1] 0.570978
```

```
posterior_odds_ratio_interval <- function(p0, p1, prob) {
  odds_ratio = (p1/(1-p1))/(p0/(1-p0))
  interval1 = quantile(odds_ratio, (1-prob)/2)
  interval2 = quantile(odds_ratio, 1-((1-prob)/2))
  return(c(interval1, interval2))
}
posterior_odds_ratio_interval(p0 = p0, p1 = p1, prob = 0.9)
```

```
##        5%       95%
## 0.3507020 0.8535565
```

```
odds_ratio = (p1/(1-p1))/(p0/(1-p0))
hist(odds_ratio)
```



b)

Let's try to change the priors for p_0 and p_1.

```
set.seed(4711)
sampleSize = 100000
p0_test1 = rbeta(sampleSize, 40+700, 636+300)
p1_test1 = rbeta(sampleSize, 23+700, 659+300)
p0_test2 = rbeta(sampleSize, 40+1600, 636+2000)
p1_test2 = rbeta(sampleSize, 23+1600, 659+2000)

estimate1 = posterior_odds_ratio_point_est(p0 = p0_test1, p1 = p1_test1)
estimate2 = posterior_odds_ratio_point_est(p0 = p0_test2, p1 = p1_test2)
c(estimate1, estimate2)
```

```
## [1] 0.9562978 0.9820616
```

After changing the priors to hugely different values, the point estimate changes drastically from ~0.57 to ~0.96 and ~0.98. This indicates that our inference is not performing well, because the prior should not have this large of an effect on the estimate.

Exercise 3 - Inference for the difference between normal means

1.1) Likelihood μ_1

$$p(y_1|\mu_1, \sigma_1^2) = N(\mu_1, \sigma_1^2)$$

1.2) Likelihood μ_2

$$p(y_2|\mu_2, \sigma_2^2) = N(\mu_2, \sigma_2^2)$$

2.1) Prior μ_1

$$p(\mu_1, \sigma_1^2) = (\sigma_1^2)^{-1}$$

2.2) Prior μ_2

$$p(\mu_2, \sigma_2^2) = (\sigma_2^2)^{-1}$$

3.1) Posterior μ_1

$$p(\mu_1|y_1) = t_{n_1-1}(\overline{y}_1, s_1^2/n_1)$$

3.2) Posterior μ_2

$$p(\mu_2|y_2) = t_{n_2-1}(\overline{y}_2, s_2^2/n_2)$$

a)

Draw samples from both distributions

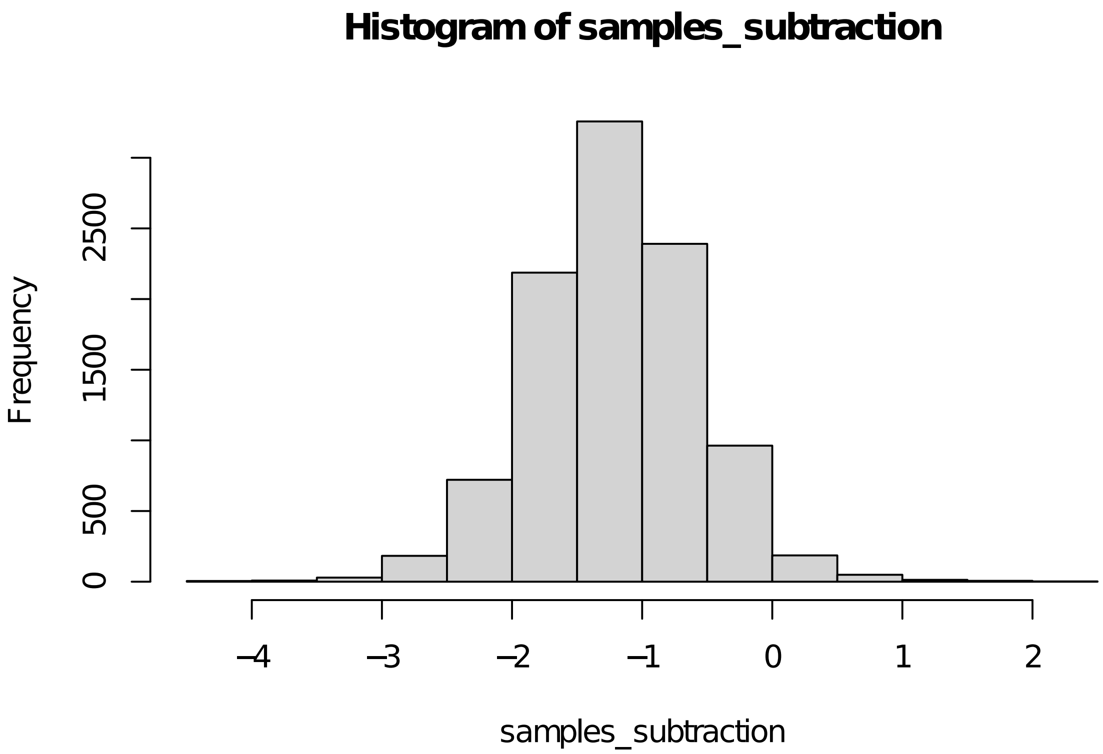
```
sampleSize = 10000

n_mu1 = length(windshieldy1)
mean_mu1 = mean(windshieldy1)
var_mu1 = (1/(n_mu1-1))*sum((windshieldy1-mean_mu1)^2)
samples_mu1 = rtnew(sampleSize, n_mu1-1, mean_mu1, (sqrt(var_mu1) / sqrt(n_mu1)))

n_mu2 = length(windshieldy2)
mean_mu2 = mean(windshieldy2)
var_mu2 = (1/(n_mu2-1))*sum((windshieldy2-mean_mu2)^2)
samples_mu2 = rtnew(sampleSize, n_mu2-1, mean_mu2, (sqrt(var_mu2) / sqrt(n_mu2)))

samples_subtraction = samples_mu1 - samples_mu2
mean(samples_subtraction)
```

```
## [1] -1.2103
```

```
quantile_lower = quantile(samples_subtraction, 0.025)
quantile_higher = quantile(samples_subtraction, 0.975)
c(quantile_lower, quantile_higher)
```

```
##        2.5%       97.5%
## -2.46052113  0.01155611
```

```
hist(samples_subtraction)
```



b)

Both mu_1 and mu_2 are continuous variables and thus is their subtraction μ_d also continuous. The probability that a continuous variable is exactly a discrete value is zero. Thus the probability that $\mu_d = 0$ is zero.