Lab1

Ludwig Thaung Elon Brange (ludth852, elobr959)

2019-04-01

Task 1

Let y1, ..., yn|theta - Bern(theta), and assume that you have obtained a sample with s = 14 successes in n = 20 trials. Assume a Beta(alpha0, beta0) prior for theta and let alpha0 = beta0 = 2.

a) Posterior theta|y - Bern(alpha0 + s, beta0 + f).

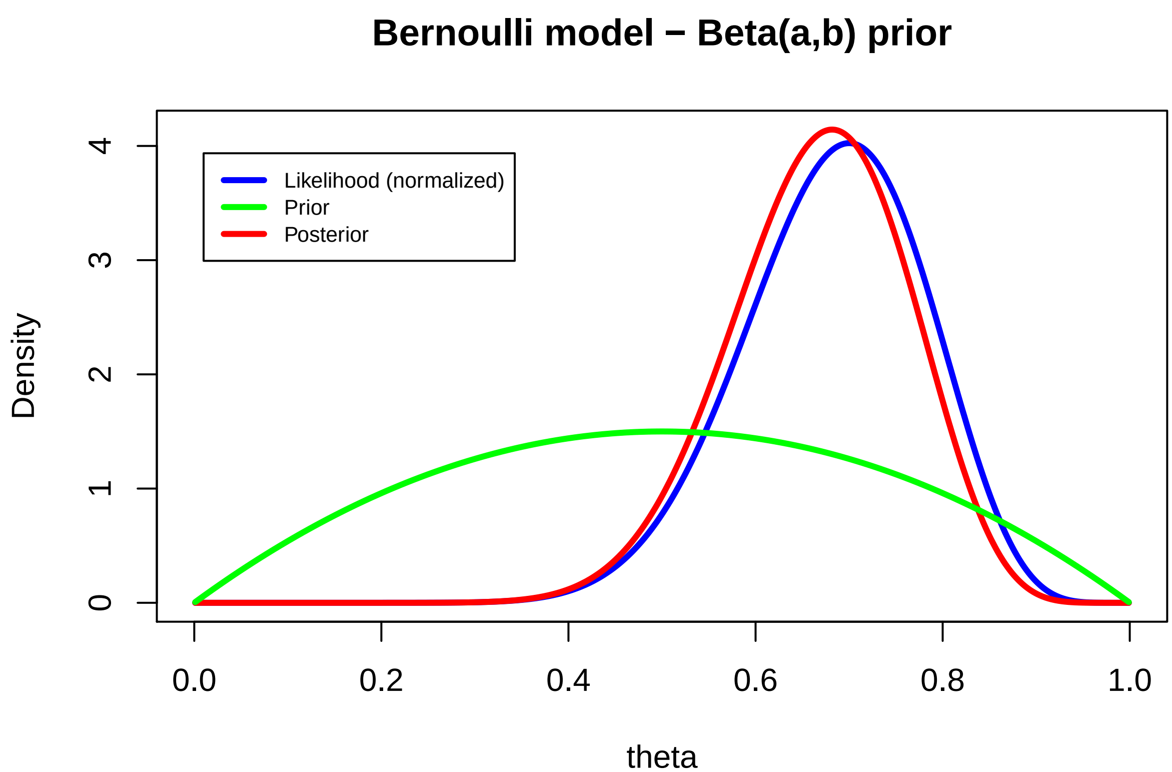
Verify graphically that the posterior mean and standard deviation converges to the true values as the number of random draws grows large.

For 20 draws we get:

```
## [1] "Posterior Mean GT: 0.666666666666667"
## [1] "ground truth std: 0.0942809041582063"
## [1] "std: 0.0923051801582707"
## [1] "Mean: 0.608671325942425"
```

For 10 0000 draws we get:

```
## [1] "Posterior Mean GT: 0.666666666666667"
## [1] "ground truth std: 0.0942809041582063"
## [1] "std: 0.0944912346276031"
## [1] "Mean: 0.666799432655026"
```



b)

Using 10 000 draws we seek to comupte the posterior probability Pr(theta < 0.4|y).

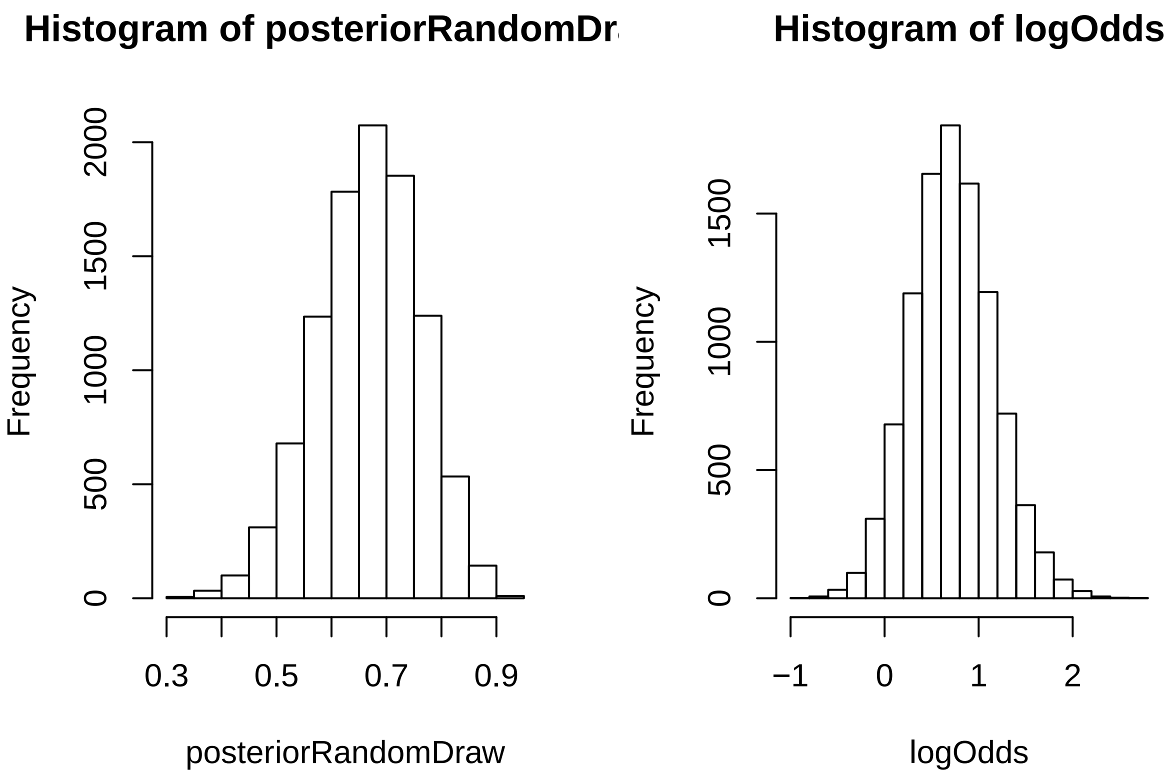
```
## [1] "propability condition with random: 0.0036"
```

```
## [1] "ground truth probability: 0.00397268082810898"
```

Looking at the plot above, the probability for theta < 0.4|y is very small. The simulated value is relatively close to the ground truth. (Note: The further to the left on the tail, the larger sample we will need as the data points become more sparse.)

c)

Computing posterior distribution of the log-odds by simulating 10 000 random draws.

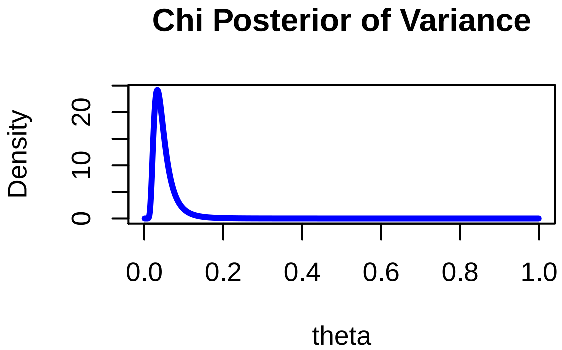


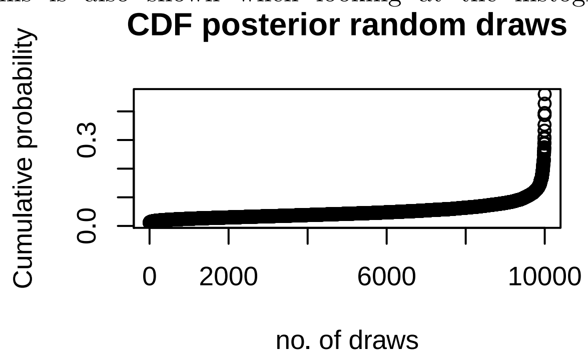

```
##
## Call:
##  density.default(x = logOdds)
##
## Data: logOdds (10000 obs.);  Bandwidth 'bw' = 0.06221
##
##        x                  y
##  Min.   :-1.02163   Min.   :0.0000074
##  1st Qu.:-0.06905   1st Qu.:0.0068038
##  Median : 0.88353   Median :0.0875281
##  Mean   : 0.88353   Mean   :0.2621891
##  3rd Qu.: 1.83611   3rd Qu.:0.4995155
##  Max.   : 2.78869   Max.   :0.9207366
```

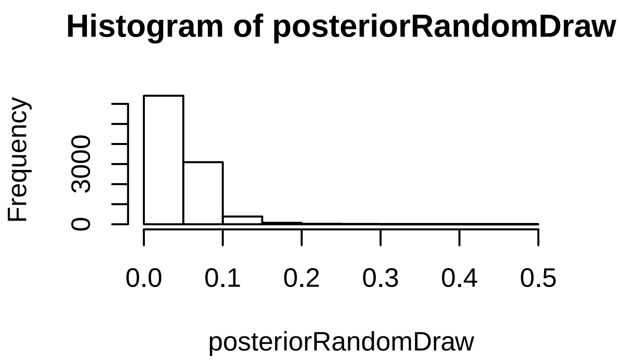
Task 2

Log-normal distribution and the Gini coefficient ###a) Simunaltion 10 000 draws from posterior of variance, assuming mean = 3.5 and comparing with the theoretical Inv Chi square posterior distribution.

"Chi posterior of variance" plot is the basis for comparison. The posterior CDF of the random draws together with the histogram of the random posterior shows that the randoms draws looks as expected in their cdf och histogram plots compared to the theoretical pdf, with the given look och the chi posterior pdf the cdf should have a relatively flat surface for a lot of values as a lot of values will be around 0.5, this is also shown when looking at the histogram.

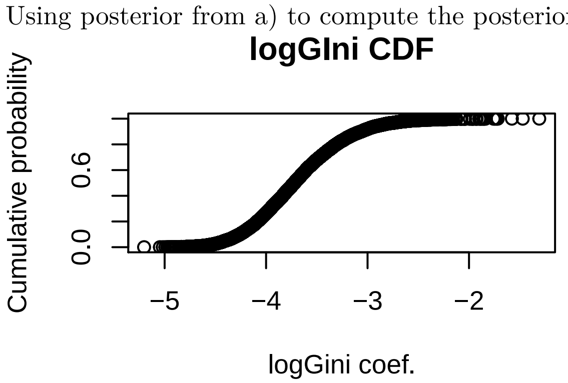


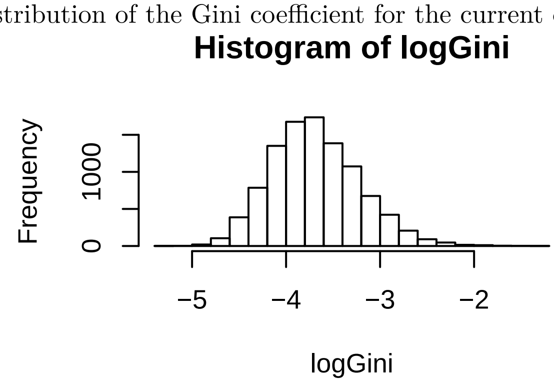




b)

Using posterior from a) to compute the posterior distribution of the Gini coefficient for the current dataset.





c)

Using posterior draws from b) to calculate a 95% euqal tail credible interval for the Gini coefficient G. In addition a kernel density estimate of the posterior of G to use that kernel density to compute a 95% HPD intercal for G.

```
##
## Call:
##  density.default(x = middleData)
##
## Data: middleData (9501 obs.);    Bandwidth 'bw' = 0.001584
##
##        x                  y
##  Min.   :0.006222   Min.   : 0.003
##  1st Qu.:0.023336   1st Qu.: 2.707
##  Median :0.040449   Median : 7.944
##  Mean   :0.040449   Mean   :14.594
##  3rd Qu.:0.057563   3rd Qu.:24.845
##  Max.   :0.074677   Max.   :44.743
## [1] "Lower end of interval: 0.0109745745699203"
## [1] "Upper end of interval: 0.0699240331269386"
```

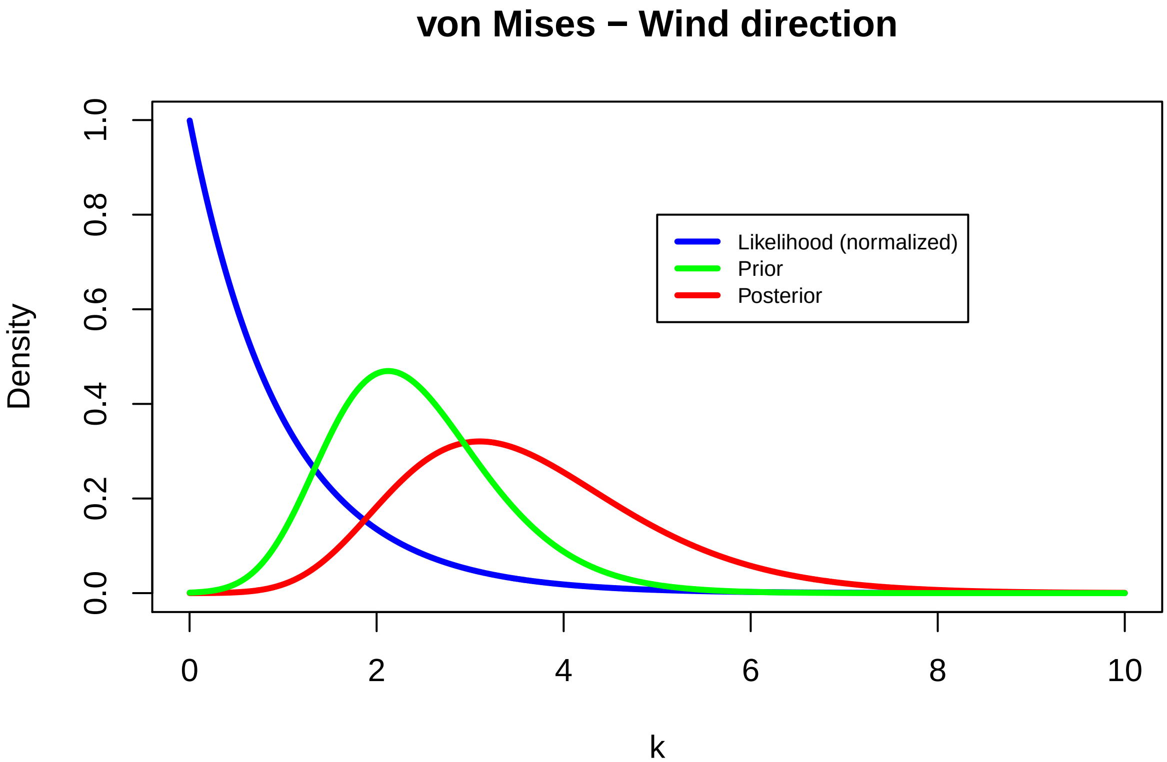
We can see that the interval for using our cutoff-method is more narrow than the interval for using the density-function as the kernel density estimate provides a interval of[0.006348, 0.074790] compared to the cutoff-interval of [0.01114952, 0.06998829].

Task 3

Bayesian inference for the concentration parameter in the von Mises distribution.

a)

Plot the posterior distribution of k (concentration parameter) for the wind direction.



b)

Approximate posterior mode of the concentration parameter k given the information in a).

```
## [1] 2.125
```

Output above shows the approximate mode of posterior, which is expected when comparing to the graph above(the green line).