# Computational Statistics & Probability

Lab 1 - Bayesian Inference

Fall 2022

## 1. Posterior Probability Distributions

Use the following R code to generate a set of `samples` to from which to answer questions about its distribution.

```
p_grid <-seq( from=0 , to=1, length.out=1000 )
prior <- rep( 1 , 1000 )
likelihood <- dbinom( 6, size=9, prob=p_grid )
posterior <-likelihood * prior
posterior <- posterior / sum(posterior)
set.seed(212)
samples <- sample( p_grid , prob=posterior , size=1e4 , replace = TRUE)
```

**a)** How much posterior probability lies below $p = 0.25$?

```
sum(samples < 0.25) / length(samples)
```

```
## [1] 0.0025
```

**b)** How much posterior probability lies above $p = 0.75$?

```
sum(samples > 0.75) / length(samples)
```

```
## [1] 0.2209
```

**c)** How much posterior probability lies between $p = 0.25$ and $p = 0.75$?

```
sum(samples > 0.25 & samples < 0.75) / length(samples)
```

```
## [1] 0.7766
```

**d)** 25% of the posterior probability lies below which value of $p$?

```
quantile(samples, 0.25)
```

```
##       25%
## 0.5435435
```

## 2. Globe Tossing Experiment

**a)** Suppose the globe tossing experiment yielded the following sequence of 15 observations,

$$[W, L, W, W, L, L, W, L, W, L, L, W, L, W, W]$$

where $W$ denotes 'water' and $L$ denotes 'land'. Using grid approximation and the same flat prior as in Question 1, construct the posterior distribution.

```
p_grid <-seq( from=0 , to=1, length.out=1000 )
prior <- rep( 1 , 1000 )
```

```r
likelihood <- dbinom( 8, size=15, prob=p_grid )
posterior <-likelihood * prior
posterior <- posterior / sum(posterior)
set.seed(212)
samples_a <- sample( p_grid , prob=posterior , size=1e4 , replace = TRUE)
```

**b)** Using grid approximation, construct the posterior distribution with a prior that is 0 below $p = 0.5$ and otherwise constant.

```r
p_grid <-seq( from=0 , to=1, length.out=1000 )
prior <- ifelse(p_grid < 0.5, 0, 1)
likelihood <- dbinom( 8, size=15, prob=p_grid )
posterior <-likelihood * prior
posterior <- posterior / sum(posterior)
set.seed(212)
samples_b <- sample( p_grid , prob=posterior , size=1e4 , replace = TRUE)
```

**c)** Explain the difference between model (2a) and (2b).

```r
# Model (b) encodes your prior knowledge that at least one-half of the Earth's
# surface is covered with water, whereas Model (a) encodes that you believe any
# proportion, from all water to all land, is equally plausible
```

**d)** Which prior, (a) or (b), is better and why?

```r
# The more informative prior, (b), gives an estimate that is closer to the true
# value (~ 71%):
mean(samples_a)
```

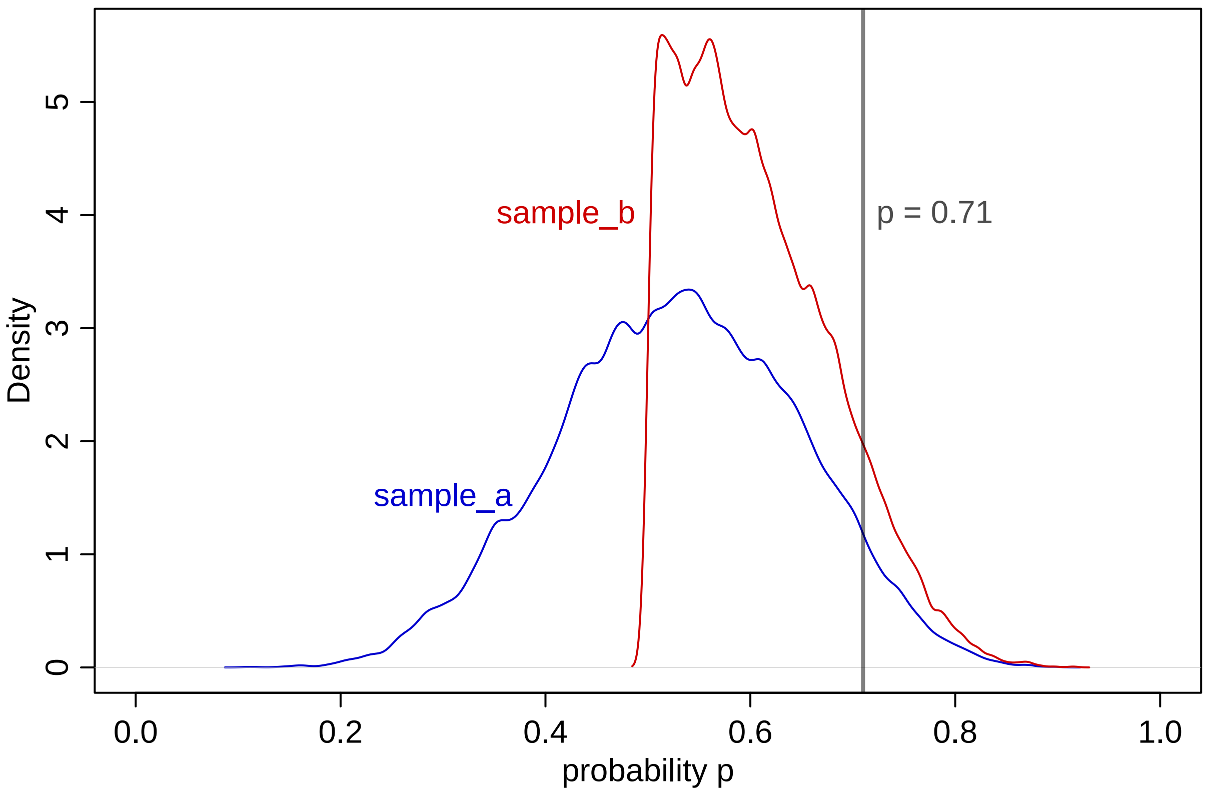
```
## [1] 0.5289651
```

```r
mean(samples_b)
```

```
## [1] 0.6060088
```

```r
# Nevertheless, Model (b) still concentrates more mass between 0.5 and 0.6 than
# we would like. Observing more data would allow either model to converge to the
# true value of $p \approx 0.71$
```

```r
# PLOT
dens(samples_a, xlab="probability p", xlim=c(0,1), ylim=c(0,5.6), col="blue3")
dens( samples_b, add=TRUE, col="red3")
abline( v=0.71, lty=1, col=grau(0.5),lwd=2)
text(0.78, 4, "p = 0.71", col=grau(0.7))
text(0.3, 1.5, "sample_a", col="blue3")
text(0.42, 4, "sample_b", col="red3")
```


sample_b
p = 0.71
sample_a
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
probability p