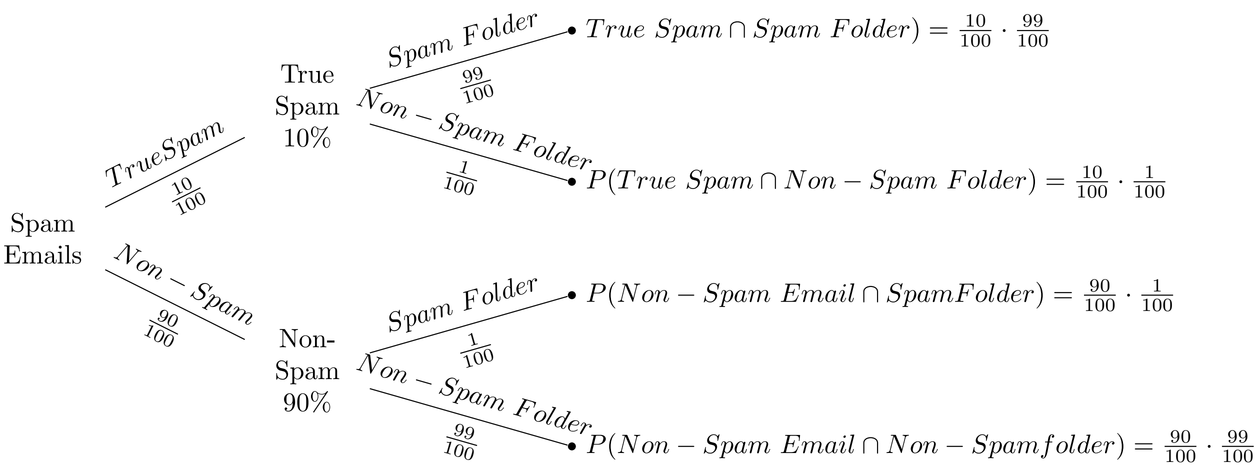
ISyE 6420: Bayesian Statistics	**(Due: 01/24/2020)**

Homework Assignment #1

Instructor: Roshan Joseph — *Name:* Chen-Yang(Jim), Liu *GTID:* 90345****

Problem 1: Spam Email (10 points)



$P(Non-Spam\ Email \mid Spam\ Folder) = \frac{P(Non-Spam\ \cap\ SpamFolder)}{P(Spam\ Folder)} = \frac{0.9\times0.01}{0.1\times0.99+0.9\times0.01} = 1/12 \approx 8.3\%$

Problem 2: HPD and Equal-tail Interval (10 points)

(a) 95% Equal-tail Interval

$$\frac{log(\theta)-0}{\sqrt{\frac{1}{1}}} \sim Z_{distribution}$$

$$\begin{aligned} log(\theta) &\in [-Z_{0.025}\sqrt{\frac{1}{1}},\ Z_{0.025}\sqrt{\frac{1}{1}}] \\ &\implies e^{-Z_{0.025}} < \theta < e^{Z_{0.025}} \\ &\implies e^{-1.96} < \theta < e^{1.96} \\ &\implies e^{-1.96} < \theta < e^{1.96} \\ &\implies 0.14 < \theta < 7.1 \end{aligned}$$

```
# Comparison
> qlnorm(c(0.025, 0.975))
[1] 0.1408635 7.0990714
```

(b) HPD credible intervals of θ

```
> library(HDInterval)
# Create 1e6 numbers random variables
> t <- rlnorm(1e6, meanlog = 0, sdlog = 1)
#Run hdi and mass = 0.95
> hdi(t, credMass = 0.95)
```

```
      lower       upper
0.02568637 5.18462661
attr(,"credMass")
[1] 0.95
```

(c) Compute the width of two intervals

- Width of HPD interval ≈ 5.15
- Width of Equal-tail interval ≈ 6.96

(d) Which method is better?

In my opinion, in this problem, we only discuss only one parameter. And, the width of HPD is less than one of equal-tail. That is, any point within HPD interval has a higher density than any other point outside. We could believe that those points are most likely values of the parameters. And, the width is small, so we could rely on this interval to select possible parameters.

Problem 3: Posterior Distribution (10 points)

Find the posterior distribution of θ.

(a) We need prior and observed data distribution. And then we could calculate posterior distribution.

$$P(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}$$
$$y_i|\theta \sim Exp(\theta) = \theta e^{-\theta y_i}$$

Posterior Distribution:

$$\begin{aligned}
P(\theta|y) &\propto P(y_1,\ldots,y_n|\theta)P(\theta) \\
&= \Pi_{i=1}^n \theta e^{-\theta y_i}\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \\
&= \theta^n e^{-\theta\sum y_i}\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \\
&= \frac{b^a}{\Gamma(a)}\theta^{a+n+-1}e^{-\theta(\sum y_i+b)} \\
&= [\frac{b^a}{\Gamma(a)}\frac{\Gamma(a+n)}{(\sum y_i+b)^{a+n}}]\frac{(\sum y_i+b)^{a+n}}{\Gamma(a+n)}\theta^{a+n-1}e^{-\theta(\sum y_i+b)} \\
&\sim \Gamma(a+n, \sum y_i + b)
\end{aligned}$$

Problem 4: Predictive Distribution (10 points)

Find the posterior predictive distribution of a future observation in problem 3

$$\begin{aligned}
p(y|y_1,\ldots,y_n) &= \int P(y,\theta|y_1,\ldots,y_n)d\theta \\
&= \int P(y|y_1,\ldots,y_n,\theta)P(\theta|y_1,\ldots,y_n)d\theta \\
&= \int P(y|\theta)P(\theta|y_1,\ldots,y_n)d\theta \\
&= \int_0^\infty \theta e^{-\theta y}\Pi_{i=1}^n\frac{(\sum y_i+b)^{a+n}}{\Gamma(a+n)}\theta^{a+n-1}e^{-\theta(\sum y_i+b)} \\
&= constant \times \int_0^\infty \frac{(n(\sum y_i+b)+y)^{an+n^2-n+2}}{\Gamma(an+n^2-n+2)}\theta^{(an+n^2-n+2)-1}e^{-\theta(n(\sum y_i+b)+y)} \\
&\sim \Gamma(an+n^2-n+2, n(\sum y_i+b)+y)
\end{aligned}$$

Problem 5: Posterior Distribution (10 points)

Find the posterior distribution of θ.

$$P(\theta|y) \propto P(y_1, \ldots, y_n|\theta)P(\theta)$$

$$\begin{aligned}
&\propto ba^b\theta^{-(b+1)}I(a \leq \theta)\Pi_{i=1}^n \frac{1}{\theta} I(y_{(n)} \leq \theta) \\
&\propto \theta^{-(b+1+n)}ba^b I(max(y_{(n)}, a) \leq \theta) \\
&\sim Pareto(max(y_{(n)}, a), b+n)
\end{aligned}$$