Week 3 Lab solutions, MAST90125: Bayesian Statistical Learning

Question One

In class, we consider the case where observations followed $y_i|\lambda \sim \text{Exp}(\lambda)$, and assumed the prior distribution of λ was $\text{Ga}(\alpha, \beta)$. You are told the sample mean, $\bar{y}$, was 1.21.

a) Determine the posterior distribution of λ.

To determine the posterior, find the joint distribution of $\lambda, y_1, \ldots, y_n$

$$\begin{aligned}
p(y_1, \ldots, y_n, \lambda) &= p(y_1, \ldots, y_n|\lambda)p(\lambda) \\
&= \left(\prod_{i=1}^{n} \lambda e^{-\lambda y_i}\right) \times \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda\beta} \\
&= \lambda^n e^{-\lambda n\bar{y}} \times \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda\beta} \\
&\propto \lambda^{\alpha+n-1}e^{-\lambda(\beta+n\bar{y})}
\end{aligned}$$

which is the kernel of a Gamma distribution with parameters $a = \alpha + n$ and $b = \beta + n\bar{y}$

b) Calculate the 95 % central and HPD credible intervals for λ for $n = 2, 5, 10, 20, 50$. Let $\alpha = 1, \beta = 0.1$ Please discuss the patterns observed.

The central 95 % credible interval can be determined using the qgamma function. As Gamma is an unimodal distribution when $a \geq 1$, the HPD interval should consist of a single interval. Determining the quantiles can be done using the function optim in R followed by the qgamma function to report the actual interval.

```
a=1;b=0.1
bary=1.21

#Central interval, n=2
n=2;qgamma(c(0.025,0.975),n+a,b+n*bary)
```

```
## [1] 0.2455048 2.8669396
```

```
#HPD interval
mpar<-optim(par=0.005,lower=0,upper=0.05, fn = function(x)
  {qgamma(x+0.95,n+a,b+n*bary)-qgamma(x,n+a,b+n*bary)},method='Brent')
qgamma(c(mpar$par,mpar$par+0.95),a+n,b+n*bary)
```

```
## [1] 0.1204367 2.5401675
```

```
#Central interval, n=5
n=5;qgamma(c(0.025,0.975),n+a,b+n*bary)
```

```
## [1] 0.3580316 1.8972898
```

```r
#HPD interval
mpar<-optim(par=0.005,lower=0,upper=0.05, fn = function(x)
  {qgamma(x+0.95,n+a,b+n*bary)-qgamma(x,n+a,b+n*bary)},method='Brent')
qgamma(c(mpar$par,mpar$par+0.95),a+n,b+n*bary)
```

```
## [1] 0.2858666 1.7665771
```

```r
#Central interval, n=10
n=10;qgamma(c(0.025,0.975),n+a,b+n*bary)
```

```
## [1] 0.4500951 1.5074062
```

```r
#HPD interval
mpar<-optim(par=0.005,lower=0,upper=0.05, fn = function(x)
  {qgamma(x+0.95,n+a,b+n*bary)-qgamma(x,n+a,b+n*bary)},method='Brent')
qgamma(c(mpar$par,mpar$par+0.95),a+n,b+n*bary)
```

```
## [1] 0.4081091 1.4437172
```

```r
#Central interval, n=20
n=20;qgamma(c(0.025,0.975),n+a,b+n*bary)
```

```
## [1] 0.5349519 1.2711267
```

```r
#HPD interval
mpar<-optim(par=0.005,lower=0,upper=0.05, fn = function(x)
  {qgamma(x+0.95,n+a,b+n*bary)-qgamma(x,n+a,b+n*bary)},method='Brent')
qgamma(c(mpar$par,mpar$par+0.95),a+n,b+n*bary)
```

```
## [1] 0.5119065 1.2402229
```

```r
#Central interval, n=50
n=50;qgamma(c(0.025,0.975),n+a,b+n*bary)
```

```
## [1] 0.6266147 1.0877684
```

```r
#HPD interval
mpar<-optim(par=0.005,lower=0,upper=0.05, fn = function(x)
  {qgamma(x+0.95,n+a,b+n*bary)-qgamma(x,n+a,b+n*bary)},method='Brent')
qgamma(c(mpar$par,mpar$par+0.95),a+n,b+n*bary)
```

```
## [1] 0.6166964 1.0758334
```

As expected, the width of the interval decreased as n increased. As the Gamma distribution is not symmetric, the HPD and central intervals were not equivalent. Rather the HPD interval was shifted to the left relative to the Central interval. This is less pronounced as n increases, which in practice is equivalent to stating the Gamma distributions converges to a Normal distribution as $n \to \infty$.

Question Two

Prove: Assume $\tilde{y}$ and y are independent given θ. Using two different ways to prove that

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \quad \text{as } \tilde{y} \perp y|\theta.$$

Proof 1.

$$p(\tilde{y}|y) = \int \frac{p(\tilde{y},\theta,y)}{p(y)}d\theta = \int \frac{p(\tilde{y}|\theta,y)p(\theta,y)}{p(y)}d\theta = \int p(\tilde{y}|\theta,y)p(\theta|y)d\theta = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

Proof 2.

$$p(\tilde{y}|y) = \int \frac{p(\tilde{y},\theta,y)}{p(y)}d\theta = \int \frac{p(\tilde{y},y|\theta)p(\theta)}{p(y)}d\theta = \int \frac{p(\tilde{y}|\theta)p(y|\theta)p(\theta)}{p(y)}d\theta = \int \frac{p(\tilde{y}|\theta)p(y,\theta)}{p(y)}d\theta = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

Question Three

Prove: Assume the posterior density function $p(\theta|y_1,...,y_n)$ is infinitely differentiable at its MAP estimator $\hat{\theta}_{\text{MAP}}$, and the derivatives of any order of $p(\theta|y_1,...,y_n)$ is upper bounded by a positive constant M, where $y_1,...,y_n$ are independent given θ. Prove that, as $n \to \infty$, the posterior should converge to a normal distribution whose variance is $(\frac{-d^2\log(p(\theta|y_1,...y_n))}{d\theta^2}|_{\theta=\hat{\theta}_{\text{MAP}}})^{-1}$. Namely, we have

$$p(\theta|y_1,\dots y_n) \to \mathcal{N}(\hat{\theta}_{\text{MAP}}, I(\hat{\theta}_{\text{MAP}})^{-1}),$$

where $I(\hat{\theta}_{\text{MAP}}) = \frac{-d^2\log(p(\theta|y_1,...y_n))}{d\theta^2}|_{\theta=\hat{\theta}_{\text{MAP}}}$.

Proof. In class, we have shown that $\hat{\theta}_{\text{MAP}}$ is consistent and distributed normally (i.e., normal distribution here) using the property of MLE. Thus, it means that

$$|\hat{\theta}_{\text{MAP}} - \theta_{\text{MAP}}| \propto \frac{1}{\sqrt{n}}.$$

Then, let us go back to

$$p(\theta|y_1,...,y_n) = e^{\log(p(\theta|y_1,...,y_n))} = \exp\left(\log(p(\theta|y_1,...,y_n))|_{\theta=\theta_0} + \sum_{i=1}^{\infty}\frac{d^i\log(p(\theta|y_1,...,y_n))/d\theta^i\big|_{\theta=\theta_0}(\theta-\theta_0)^i}{i!}\right).$$

When $i > 2$, we have

$$\sum_{i=3}^{k}\frac{d^i\log(p(\theta|y_1,...,y_n))}{d\theta^i}\Big|_{\theta=\hat{\theta}_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^i/i! < \frac{MC^3(k-2)}{3!n^{\frac{3}{2}}} \propto \frac{1}{\sqrt{n}}.$$

We will prove the above inequation. Because $\hat{\theta}_{\text{MAP}}$ is consistent, we know the Taylor series will converge. Thus, there exists a finite constant c such that

$$c\sum_{i=3}^{k}\frac{d^i\log(p(\theta|y_1,...,y_n))}{d\theta^i}\Big|_{\theta=\hat{\theta}_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^i/i! < c\sum_{i=3}^{k}\frac{d^3\log(p(\theta|y_1,...,y_n))}{d\theta^3}\Big|_{\theta=\hat{\theta}_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^3/3!,$$

$$c\sum_{i=3}^{k}\frac{d^3\log(p(\theta|y_1,...,y_n))}{d\theta^3}\Big|_{\theta=\hat{\theta}_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^3/3! < c(k-2)\frac{d^3\log(p(\theta|y_1,...,y_n))}{d\theta^3}\Big|_{\theta=\hat{\theta}_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^3/3!,$$

$$c(k-2)\frac{d^3\log(p(\theta|y_1,...,y_n))}{d\theta^3}\Big|_{\theta=\hat{\theta}_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^3/3! < c(k-2)M\Big(\frac{C}{\sqrt{n}}\Big)^3/3! = \frac{cMC^3(k-2)}{3!n^{\frac{3}{2}}} \propto \frac{1}{\sqrt{n}}.$$

Then, as n goes to infinity, we have

$$\begin{aligned}
p(\theta|y_1,...,y_n) &= \exp\left(\log(p(\theta|y_1,...,y_n))|_{\theta=\theta_{\text{MAP}}} + \sum_{i=1}^{\infty}\frac{d^i\log(p(\theta|y_1,...,y_n))/d\theta^i\big|_{\theta=\theta_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^i}{i!}\right)\\
&= \exp\left(\log(p(\theta|y_1,...,y_n))|_{\theta=\theta_{\text{MAP}}} + \frac{d^2\log(p(\theta|y_1,...,y_n))/d\theta^2\big|_{\theta=\theta_{\text{MAP}}}(\theta-\hat{\theta}_{\text{MAP}})^2}{2}\right)\\
&= p(\hat{\theta}_{\text{MAP}}|y_1,...,y_n)\exp\left(\frac{(\theta-\hat{\theta}_{\text{MAP}})^2}{2(d^2\log(p(\theta|y_1,...,y_n))/d\theta^2\big|_{\theta=\theta_{\text{MAP}}})^{-1}}\right)
\end{aligned}$$