Bayesian Econometrics **Summer Quarter 2013**

Instructor: Hedibert Freitas Lopes **1st review session**

Solution

Suppose we are using an Exponential(λ) distribution[1] to model the lifetimes of n items, ie. the lifetimes $t_1, \ldots, t_n$ are conditionally independent and identically distributed Exponential(λ). In addition, suppose $\gamma \sim Gamma(a, b)$ summarizes the prior[2] information regarding λ. It is also relatively straightforward to see that $(s_n|\lambda) \sim Gamma(n, \lambda)$, where $s_n = \sum_{i=1}^n t_i$. Derive

a) The maximum likelihood estimator of λ, ie. $\hat{\lambda}$.

The joint density of $t_1, \ldots, t_n$ given λ is

$$p(t_1, \ldots, t_n|\lambda) = \prod_{i=1}^n \lambda \exp\{-\lambda t_i\} = \lambda^n \exp\{-\lambda s_n\},$$

so the log-likelihood function is $l(\lambda; s_n) = c + n \log \lambda - \lambda s_n$, $l'(\lambda; s_n) = n/\lambda - s_n$ and $l''(\lambda; s_n) = -n/\lambda^2 < 0, \forall \lambda > 0$. Then, the MLE of λ is $\hat{\lambda} = n/s_n$.

b) The standard error of $\hat{\lambda}$ (assuming n is large).

Since $s_n|\lambda \sim G(n, \lambda)$, it is easy to show that $(s_n/n|\lambda) \sim G(n, n\lambda)$ and $\hat{\lambda}|\lambda \sim IG(n, n\lambda)$[3]. Therefore,

$$V(\hat{\lambda}|\lambda) = \frac{n^2}{(n-1)^2(n-2)}\lambda^2 \approx \frac{\lambda^2}{n},$$

for large n. Hence, $V(\hat{\lambda}|\lambda)$ can be approximately estimated by $\hat{\lambda}^2/n = n/s_n^2$. Hence, the standard error of $\hat{\lambda}$ is approximated (for large n) by $\sqrt{n}/s_n$.

c) The 90% confidence interval (L, U) for λ.

Based on the derivation from b) above and the central limit theorem, $\hat{\lambda}|\lambda \sim N(\lambda, \lambda^2/n)$ and an (approximate) 90% confidence interval for λ is

$$\begin{aligned} L &= \hat{\lambda} - 1.644854\sqrt{n}/s_n = \frac{n - 1.644854\sqrt{n}}{s_n} \\ U &= \hat{\lambda} + 1.644854\sqrt{n}/s_n = \frac{n + 1.644854\sqrt{n}}{s_n} \end{aligned}$$

d) The prior predictive[4] for s_n, $p(s_n)$.

The prior predictive is

$$\begin{aligned} p(s_n) &= \int_0^\infty p(s_n|\lambda)p(\lambda)d\lambda \\ &= \int_0^\infty \frac{\lambda^n}{\Gamma(n)} s_n^{n-1} \exp\{-\lambda s_n\} \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp\{-b\lambda\} d\lambda \\ &= \frac{b^a s_n^{n-1}}{\Gamma(a)\Gamma(n)} \int_0^\infty \lambda^{(a+n)-1} \exp\{-(b + s_n)\lambda\} d\lambda \\ &= \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \frac{b^a s_n^{n-1}}{(b + s_n)^{a+n}}. \end{aligned}$$

[1] We say that the random variable T has an exponential distribution (or is an exponential random variable), and write $T \sim Exponential(\lambda)$ if the probability density function (pdf) for T is $p(t|\lambda) = \lambda \exp\{-\lambda t\}$ for $t > 0$ and $\lambda > 0$. $E(t|\lambda) = 1/\lambda$ and $V(t|\lambda) = 1/\lambda^2$.

[2] We say that $\lambda \sim Gamma(a, b)$ if its pdf is $p(\lambda|a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp\{-b\lambda\}$. $E(\lambda|a, b) = a/b$ and $V(\lambda|a, b) = a/b^2$. In fact, $Gamma(1, b) \equiv Exp(b)$.

[3] We say that $x \sim IG(a, b)$ if its pdf is $p(x|a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp\{-b/x\}$. $E(x|a, b) = b/(a-1)$, if $a > 1$, and $V(x|a, b) = b^2/((a-1)^2(a-2))$, if $a > 2$.

[4] Also known as *predictive*, *normalizing constant*, or *marginal likelihood*.

e) The posterior distribution for λ, ie. $p(\lambda|s_n)$.

Combining likelihood and prior:

$$\begin{aligned} p(\lambda|s_n) &\propto \lambda^n \exp\{-\lambda s_n\}\lambda^{a-1}\exp\{-b\lambda\} \\ &\propto \lambda^{(a+n)-1}\exp\{-(b+s_n)\lambda\}, \end{aligned}$$

leads to $\lambda|s_n \sim G(a+n, b+s_n)$, so $E(\lambda|s_n) = (a+n)/(b+s_n)$. Again, $E(\lambda|s_n)$ converges to the MLE, $\hat{\lambda} = n/s_n$, when $a = b = 0$ or when n and s_n are much larger than a and b, respectively.

f) The posterior predictive $p(t_{n+1}|s_n)$.

The posterior predictive is

$$\begin{aligned} p(t_{n+1}|s_n) &= \int_0^\infty p(t_{n+1}|\lambda)p(\lambda|s_n)d\lambda \\ &= \int_0^\infty \lambda\exp\{-\lambda t_{n+1}\}\frac{(b+s_n)^{a+n}}{\Gamma(a+n)}\lambda^{(a+n)-1}\exp\{-(b+s_n)\lambda\}d\lambda \\ &= \frac{(b+s_n)^{a+n}}{\Gamma(a+n)}\int_0^\infty \lambda^{(a+n+1)-1}\exp\{-(b+s_n+t_{n+1})\lambda\}d\lambda \\ &= \frac{\Gamma(a+n+1)}{\Gamma(a+n)}\frac{(b+s_n)^{a+n}}{(b+s_n+t_{n+1})^{a+n+1}} \\ &= \left(\frac{b+s_n}{b+s_n+t_{n+1}}\right)^{a+n}\left(\frac{a+n}{b+s_n+t_{n+1}}\right). \end{aligned}$$

Suppose that we observed $n = 50$ items and that $s_{50} = 25$. Also, let the prior hyperparameters of the prior be $a = 1$ and $b = 2$.

g) Compare the MLE summaries from a) and b) to posterior summaries $E(\lambda|s_{50} = 25)$ and $\sqrt{V(\lambda|s_{50} = 25)}$, respectively.

$$\begin{aligned} \hat{\lambda} &= \frac{50}{25} = 2.00 \\ \text{s.e.}(\hat{\lambda}) &= \frac{\sqrt{50}}{25} = 0.2828427 \\ E(\lambda|s_{50} = 25) &= \frac{1+50}{2+25} = 1.888889 \\ \sqrt{V(\lambda|s_{50} = 25)} &= \frac{\sqrt{1+50}}{2+25} = 0.2644973 \\ L &= \frac{50 - 1.644854\sqrt{50}}{25} = 1.534765 \\ U &= \frac{50 + 1.644854\sqrt{50}}{25} = 2.465235 \end{aligned}$$

h) Plot the posterior $p(\lambda|s_{50} = 25)$ and posterior predictive $p(t_{51}|s_{50} = 25)$.

See Figures 1 and 2.

i) What is the posterior probability that λ falls in the 90% confidence interval found in (c)? Or, $Pr(L \leq \lambda \leq U|s_{50} = 25)$?

$$
\begin{aligned}
Pr(1.534765 \leq \lambda \leq 2.465235|s_{50} = 25) &= Pr(\lambda \leq 2.465235|s_{50} = 25) - Pr(\lambda \leq 1.534765|s_{50} = 25) \\
&= 0.9791073 - 0.08293522 \\
&= 89.62\%,
\end{aligned}
$$

which is pretty close to 90%. In fact, the 90% Bayesian creditability interval is $[1.475879; 2.343966]$.

R code for item g), h) and i)

```
# Data summary
n  = 50
sn = 25
# Prior hyperparameters
a = 1
b = 2

# Frequentist results
lambdahat    = n/sn
se.lambdahat = sqrt(n)/sn
L = (n -1.644854*sqrt(n))/sn
U = (n + 1.644854*sqrt(n))/sn

# Bayesian results
Elambda   = (a+n)/(b+sn)
Vlambda   = (a+n)/(b+sn)^2
sd.lambda = sqrt(Vlambda)

# lambda|sn=25 ~ G(a+n,b+sn)
p1 = pgamma(U,a+n,b+sn)
p2 = pgamma(L,a+n,b+sn)
Lb = qgamma(0.05,a+n,b+sn)
Ub = qgamma(0.95,a+n,b+sn)

pdf(file="gamma-posterior.pdf",width=10,height=8)
lambda = seq(0.5,3.5,length=1000)
plot(lambda,dgamma(lambda,a+n,b+sn),xlab=expression(lambda),ylab="density",type="l")
segments(L,0,U,0,lwd=4,col=1)
points(lambdahat,0,pch=16,cex=2,col=1)
segments(Lb,0.05,Ub,0.05,lwd=4,col=2)
points(Elambda,0.05,pch=16,cex=2,col=2)
legend(2.5,1.5,legend=c("MLE","BAYES"),col=1:2,lwd=4,bty="n")
dev.off()

pdf(file="posterior-predictive.pdf",width=10,height=8)
t = seq(0.01,3,length=1000)
plot(t,((b+sn)/(b+sn+t))^(a+n)*((a+n)/(b+sn+t)),xlab=expression(t[n+1]),
       ylab="Density",main="",type="l",ylim=c(0,2))
abline(h=0,lty=3)
dev.off()
```

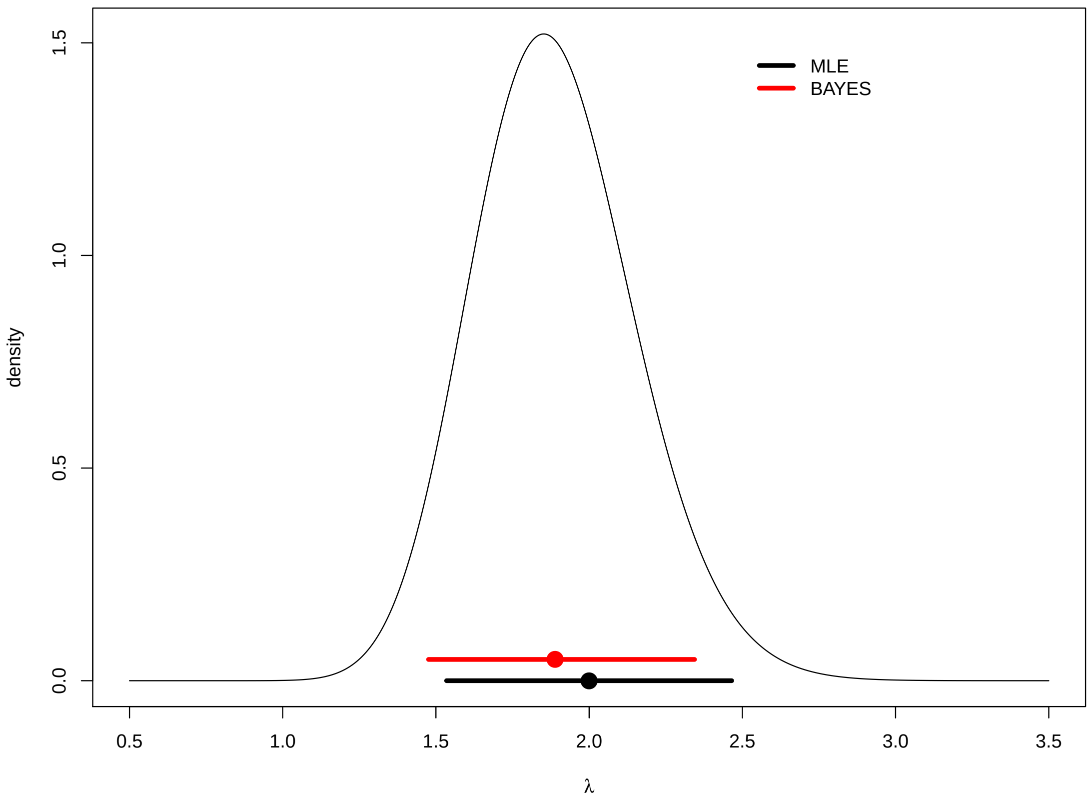

Figure 1: Posterior distribution of λ, $p(\lambda|s_{50} = 25)$.

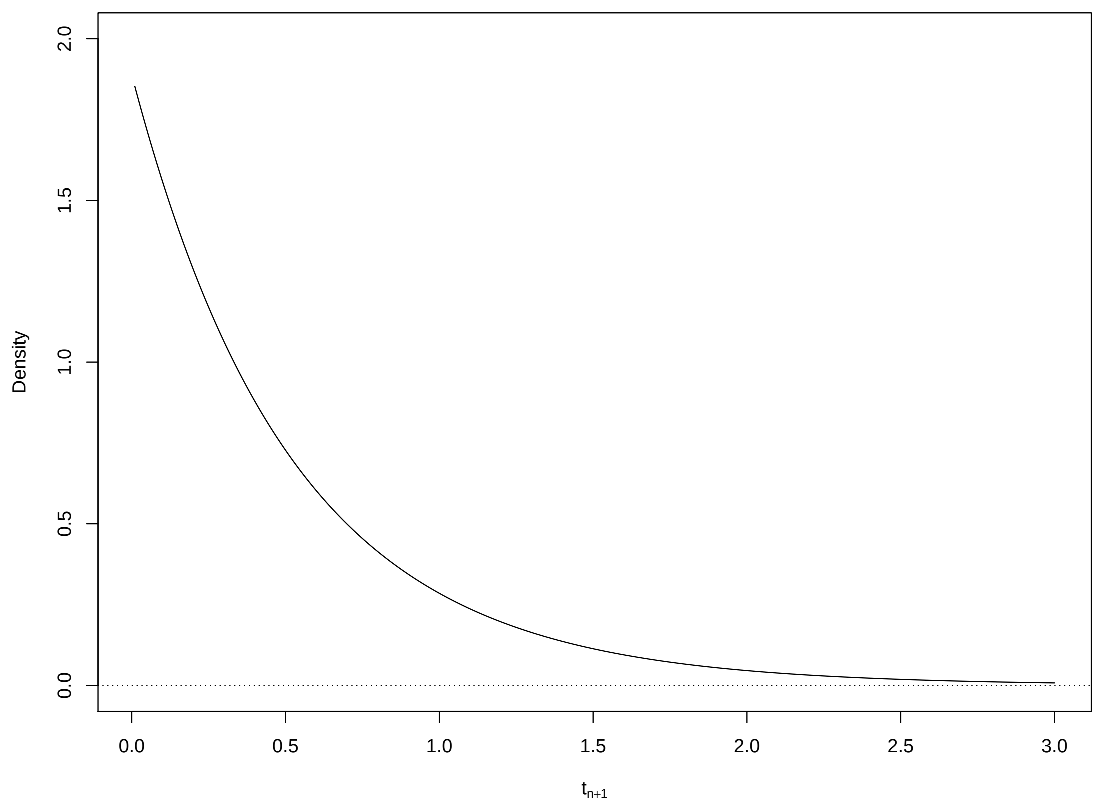

Figure 2: Posterior predictive distribution, $p(t_{51}|s_{50} = 25)$.