MORE BAYES

Ex $X_1, \ldots X_n \overset{iid}{\sim} Exp(\lambda)$, Prior $Gamma(\alpha, \beta)$

In place of λ

$$f(\underset{\sim}{x} \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = \lambda^n e^{-\lambda n \bar{x}}$$

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda} \lambda^{\alpha-1} I(\lambda>0)$$

$$\pi(\lambda \mid \underset{\sim}{x}) \propto f(\underset{\sim}{x} \mid \lambda)\, \pi(\lambda)$$

$$= \lambda^n e^{-\lambda n \bar{x}} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\lambda} \lambda^{\alpha-1} I(\lambda>0)$$

$$\propto e^{-(\beta+n\bar{x})\lambda} \lambda^{(\alpha+n)-1} I(\lambda>0)$$

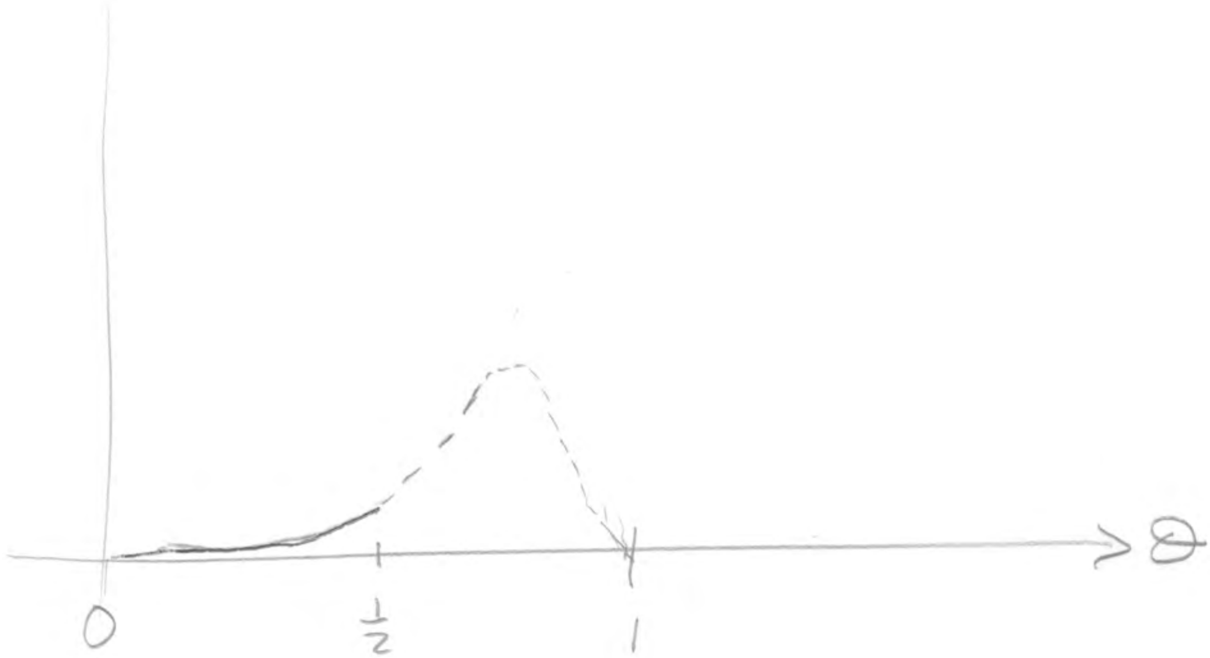
$Gamma(\alpha+n, \beta+n\bar{x})$

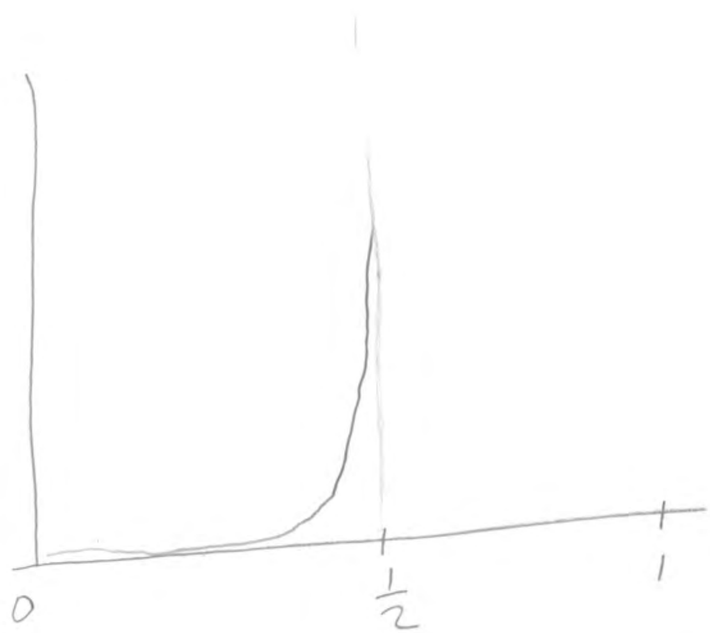
$$\propto \frac{(\beta+n\bar{x})^{\alpha+n}}{\Gamma(\alpha+n)} e^{-(\beta+n\bar{x})\lambda} \lambda^{(\alpha+n)-1} I(\lambda>0)$$

- Another conjugate prior
- The posterior always has the same support as the prior
- Avoid Bone-headed priors

Beware of prior distributions that are not supported on the entire parameter space

Suppose $\Omega = (0,1)$, but $\pi(\theta) = 2 I(0 < \theta < \frac{1}{2})$





Ex Conditionally on $M = m$, $X_1, \dots X_n \overset{iid}{\sim} N(m, \tau_0)$ Precision $\tau_0 = \frac{1}{\sigma^2}$ is known

$M \sim N(\mu, \tau)$ The prior distribution

$$f(\underset{\sim}{x} | m) = \prod_{i=1}^{n} \frac{\tau_0^{1/2}}{\sqrt{2\pi}} e^{-\frac{\tau_0}{2}(x_i - m)^2}$$

$$= \frac{\tau_0^{n/2}}{(2\pi)^{n/2}} e^{-\frac{\tau_0}{2}\sum_{i=1}^{n}(x_i - m)^2}$$

$$= \frac{\tau_0^{n/2}}{(2\pi)^{n/2}} e^{-\frac{\tau_0}{2}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - m)^2}$$

$$= \frac{\tau_0^{n/2}}{(2\pi)^{n/2}} e^{-\frac{\tau_0}{2}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - m)^2\right]}$$

$$= \frac{\tau_0^{n/2}}{(2\pi)^{n/2}} e^{-\frac{\tau_0}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2} e^{-\frac{n\tau_0}{2}(\bar{x} - m)^2}$$

AND $\pi(m) = \frac{\tau^{1/2}}{\sqrt{2\pi}} e^{-\frac{\tau}{2}(m - \mu)^2}$

$$\pi(m | x) \propto f(\underset{\sim}{x} | m)\, \pi(m)$$

$$\propto e^{-\frac{n\tau_0}{2}(\bar{x} - m)^2} e^{-\frac{\tau}{2}(m - \mu)^2}$$

$$= \mathrm{Exp} - \frac{1}{2}\left[n\tau_0(\bar{x}^2 - 2\bar{x}m + m^2) + \tau(m^2 - 2m\mu + \mu^2) \right]$$

$$= \mathrm{Exp} - \frac{1}{2}\left[\underline{n\tau_0\bar{x}^2} - 2\tau_0 n\bar{x}m + n\tau_0 m^2 + \tau m^2 - 2\tau\mu m + \underline{\tau\mu^2} \right]$$

$$= e^{-\frac{1}{2}(n\tau_0\bar{x}^2 + \tau\mu^2)} * \mathrm{Exp} - \frac{1}{2}\left[m^2(\tau + n\tau_0) - 2(\tau\mu + \tau_0 n\bar{x})m \right]$$

$$\propto \mathrm{Exp} - \frac{(\tau + n\tau_0)}{2}\left[m^2 - 2\frac{\tau\mu + \tau_0 n\bar{x}}{\tau + n\tau_0} m + \left(\frac{\tau\mu + \tau_0 n\bar{x}}{\tau + n\tau_0}\right)^2 \right]$$

$$\propto \frac{(\tau + n\tau_0)^{1/2}}{\sqrt{2\pi}} \mathrm{Exp} - \frac{\tau + n\tau_0}{2}\left(m - \frac{\tau\mu + \tau_0 n\bar{x}}{\tau + n\tau_0} \right)^2$$

$$\mathrm{Normal}\left(\frac{\tau\mu + \tau_0 n\bar{x}}{\tau + n\tau_0},\ \tau + n\tau_0 \right)$$

Posterior of unknown expected value M

is $N\left(\frac{\tau\mu + \tau_0 n \bar{x}}{\tau + n\tau_0} , \tau + n\tau_0 \right)$

- Another conjugate prior
- What happens as $n \to \infty$

Posterior expected value

Viewing the posterior as a function of the random vector $\underset{\sim}{X}$ (not just conditional on $\underset{\sim}{X} = \underset{\sim}{x}$) and assuming that there is a true value of m, a fixed constant

$$E(M \mid \underset{\sim}{X}) = \frac{\tau\mu + \tau_0 n \bar{X}_n}{\tau + n\tau_0}$$

$$= \frac{\frac{\tau\mu}{n} + \frac{\tau_0 \cancel{n} \bar{X}_n}{\cancel{n}}}{\frac{\tau}{n} + \frac{\cancel{n}\tau_0}{\cancel{n}}} \xrightarrow{p} \frac{\tau_0 m}{\tau_0} = m$$

And

$$Var(M \mid \underset{\sim}{X} = \underset{\sim}{x}) = \frac{1}{\tau + n\tau_0} \longrightarrow 0$$

Credible Intervals: Bayesian Confidence Intervals

An interval that contains a large amount of the posterior probability, like 95%

There are ∞ many such intervals. 3 methods

1) Shortest interval
2) Interval centered at $E(\Theta \mid \underset{\sim}{X} = \underset{\sim}{x})$
3) $\alpha/2$ in each tail

For a symmetric posterior, these are the same

Normal example from Q3A of Assignment 6

$X_1, \ldots X_{10} \overset{iid}{\sim} N(\cancel{\mu}, \sigma_0^2 = \frac{1}{2})$ so $\tau_0 = 2$

m

$\bar{x}_n = 4.88$, 95% $CI = \left(\bar{X} - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right)$

$= (4.44, 5.32)$

We have posterior

$$M \mid \underset{\sim}{X} = \underset{\sim}{x} \sim N\left(\frac{\tau\mu + \tau_0 n\bar{x}}{\tau + n\tau_0}, \tau + n\tau_0\right)$$

Make prior standard normal, so $\mu = 0, \tau = 1$

Posterior is $N\left(\frac{0 + 2*10*4.88}{1 + 10*2}, 1 + 10*2\right)$

$$= N(4.65, 21) \quad \sigma^2 = \frac{1}{21}$$

(Θ is same as M)

$$0.95 = P(-1.96 < Z < 1.96 \mid \underset{\sim}{X} = \underset{\sim}{x})$$

$$= P\left(-1.96 < \frac{\Theta - 4.65}{\sqrt{1/21}} < 1.96 \mid \underset{\sim}{x}\right)$$

$$= P\left(4.65 - \frac{1.96}{\sqrt{21}} < \Theta < 4.96 + \frac{1.96}{\sqrt{21}} \mid \underset{\sim}{x}\right)$$

$$= P(4.22 < \Theta < 5.08 \mid \underset{\sim}{x})$$

95% Credible interval $= (4.22, 5.08)$

Compare Confidence interval $(4.44, 5.32)$

Example with a discrete prior, not obviously conjugate

Toss a fair coin n times, observe X heads, try to estimate n.

What does λ mean?

$$X|n \sim \text{Binomial}(n, \tfrac{1}{2}), \quad N \sim \text{Poisson}(\lambda)$$

(NATURAL FREQUENTIST GUESS (AND MOM) IS $\hat{n} = 2x$ Because $E(X) = n\theta = n\frac{1}{2}$ so set $x = \hat{n} \cdot \frac{1}{2} \Rightarrow \hat{n} = 2x$)

$$\pi(n|x) = \frac{p(x|n)\,\pi(n)}{\sum_m p(x|m)\,\pi(m)}$$

$$= \frac{\binom{n}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{n-x} I(x=0,\dots,n)\ \frac{e^{-\lambda}\lambda^n}{n!} I(n=0,1,\dots)}{\sum_{m=0}^{\infty} \frac{\cancel{m!}}{x!\,(m-x)!}\ \frac{1}{2^m}\ \frac{e^{-\lambda}\lambda^m}{\cancel{m!}} I(m=x,x+1,\dots)}$$

$$= \frac{\text{num}}{\text{denom}} \quad \text{and denom} =$$

$$\sum_{m=x}^{\infty} \frac{1}{x!\,(m-x)!}\ \frac{1}{2^m}\ e^{-\lambda}\lambda^m = \frac{e^{-\lambda}}{x!} \sum_{m=x}^{\infty} \frac{1}{(m-x)!}\left(\frac{\lambda}{2}\right)^m$$

Set $k = m - x \Leftrightarrow m = x + k$

m	$k = m-x$
∞	∞
x	0

$$\text{denom} = \frac{e^{-\lambda}}{x!} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda}{2}\right)^{x+k}$$

$$= \frac{e^{-\lambda}\left(\frac{\lambda}{2}\right)^{x}}{x!} \sum_{k=0}^{\infty} \frac{(\lambda/2)^k}{k!}$$

$$= \frac{e^{-\lambda}\left(\frac{\lambda}{2}\right)^{x}}{x!} e^{\lambda/2} = \frac{e^{-\lambda/2}\left(\frac{\lambda}{2}\right)^{x}}{x!}$$

Poisson!

So that $\pi(n \mid x) = \dfrac{\text{num}}{\text{denom}}$

$$= \frac{\frac{\cancel{n!}}{\cancel{x!}\,(n-x)!} \frac{1}{2^n} \frac{e^{-\lambda}\lambda^n}{\cancel{n!}} I(n = x, x+1, \ldots)}{\frac{e^{-\lambda/2}\left(\frac{\lambda}{2}\right)^{x}}{\cancel{x!}}}$$

$$= \frac{e^{-\lambda/2}\left(\frac{\lambda}{2}\right)^{n-x}}{(n-x)!} I(n = x, x+1, \ldots)$$

$$\pi(n \mid x) = \frac{e^{-\lambda/2} \left(\frac{\lambda}{2}\right)^{n-x}}{(n-x)!} I(n = x, x+1, \dots)$$

- It adds to one. Set $k = n - x$.

- It's intelligent. The number of times the coin is tossed cannot be less than the number of heads.

- This is the posterior pmf of $N = W + x$, where $W \sim \text{Poisson}\left(\frac{\lambda}{2}\right)$.

$$E(N \mid x) = E(W) + E(x)$$
$$= \frac{\lambda}{2} + x$$

λ is the prior guess at n

Frequentist $\hat{n} = 2x$ so $x = \frac{\hat{n}}{2}$

$$E(N \mid x) = \frac{\lambda}{2} + \frac{\hat{n}}{2}$$

An equal compromise between prior belief and data.

- This is another conjugate prior, sort of...

Big family of potentially shifted Poisson distributions.

The (posterior) Predictive Distribution

$$\int\int f(x|\theta) \cancel{\pi(\theta|x)} d\theta \, dx$$

$$= \int \underbrace{\left(\int f(x|\theta)\, dx\right)}_{=1} \pi(\theta|\underset{\sim}{x})\, d\theta = 1$$

$$f(x|\underset{\sim}{x}) = \int f(x|\theta)\, \pi(\theta|\underset{\sim}{x})\, d\theta = E\left(f(x|\Theta) \,\middle|\, \underset{\sim}{x} \right)$$

- "Averaging out" the parameter with respect to the posterior distribution
- Kind of a Bayesian density estimate.

For the exponential - gamma example,

$$\int_0^\infty \lambda e^{-\lambda x} I(x>0) \frac{(\beta+n\bar{x})^{\alpha+n}}{\Gamma(\alpha+n)} e^{-(\beta+n\bar{x})\lambda} \lambda^{\alpha+n-1} d\lambda \quad \cancel{I(\lambda>0)}$$

$$= \frac{(\beta+n\bar{x})^{\alpha+n}}{\Gamma(\alpha+n)} \frac{\Gamma(\alpha+n+1)}{(\beta+n\bar{x}+x)^{\alpha+n+1}}$$

$$\underbrace{\frac{(\beta+n\bar{x}+x)^{\alpha+n+1}}{\Gamma(\alpha+n+1)} \int_0^\infty e^{-(\beta+n\bar{x}+x)\lambda} \lambda^{\alpha+n+1-1} d\lambda}_{=1} \; I(x>0)$$

$$= \frac{(\beta+n\bar{x})^{\alpha+n}}{(\beta+n\bar{x}+x)^{\alpha+n+1}} \frac{(\alpha+n)\cancel{\Gamma(\alpha+n)}}{\cancel{\Gamma(\alpha+n)}} I(x>0)$$

$$= \frac{(\alpha+n)(\beta+n\bar{x})^{\alpha+n}}{(\beta+n\bar{x}+x)^{\alpha+n+1}} I(x>0)$$

Is this really a density?

$$\int_0^\infty (\alpha+n)(\beta+n\bar{x})^{\alpha+n}(\beta+n\bar{x}+x)^{-(\alpha+n)-1}\,dx$$

$$u = \beta+n\bar{x}+x \qquad du = dx$$

$$\begin{array}{c|c} x & u \\ \hline \infty & \infty \\ 0 & \beta+n\bar{x}+0 \end{array}$$

$$= \int_{\beta+n\bar{x}}^\infty (\alpha+n)(\beta+n\bar{x})^{\alpha+n} u^{-(\alpha+n)-1}\,du$$

$$= \cancel{(\alpha+n)}(\beta+n\bar{x})^{\alpha+n} \frac{u^{-(\alpha+n)}}{-\cancel{(\alpha+n)}}\Bigg|_{\beta+n\bar{x}}^{\infty}$$

$$= (\beta+n\bar{x})^{\alpha+n}(-1)\left(\lim_{u\to\infty}\frac{1}{u^{\alpha+n}} - \frac{1}{(\beta+n\bar{x})^{\alpha+n}}\right)$$

$$= (\beta+n\bar{x})^{\alpha+n}(-1)\left(0 - \frac{1}{(\beta+n\bar{x})^{\alpha+n}}\right)$$

$$= 1$$ <u>OKAY</u>

Posterior predictive density

$$E(f(x|\Theta) \mid \underset{\sim}{x}) = \int f(x|\theta)\,\pi(\theta|x)\,d\theta$$

$$= \frac{(\alpha+n)(\beta+n\bar{x})^{\alpha+n}}{(\beta+n\bar{x}+x)^{\alpha+n+1}}\, I(x>0)$$

What does the posterior predictive density look like?

Is it a good estimate of the actual model density?

"Actual" means that there is a "true" parameter value that exists outside the mind of the statistician.

Maybe in the mind of God.

Example Lecture of Tues Feb 25th, p. 14

Exponential (λ) random sample,
$n=29$, $\bar{x} = 2.5$ so $\hat{\lambda} = 1/\bar{x} = 0.4$
Natural frequentist density estimate is

$$\hat{\lambda} e^{-\hat{\lambda} x} I(x>0) = 0.4 e^{-0.4x} I(x>0)$$

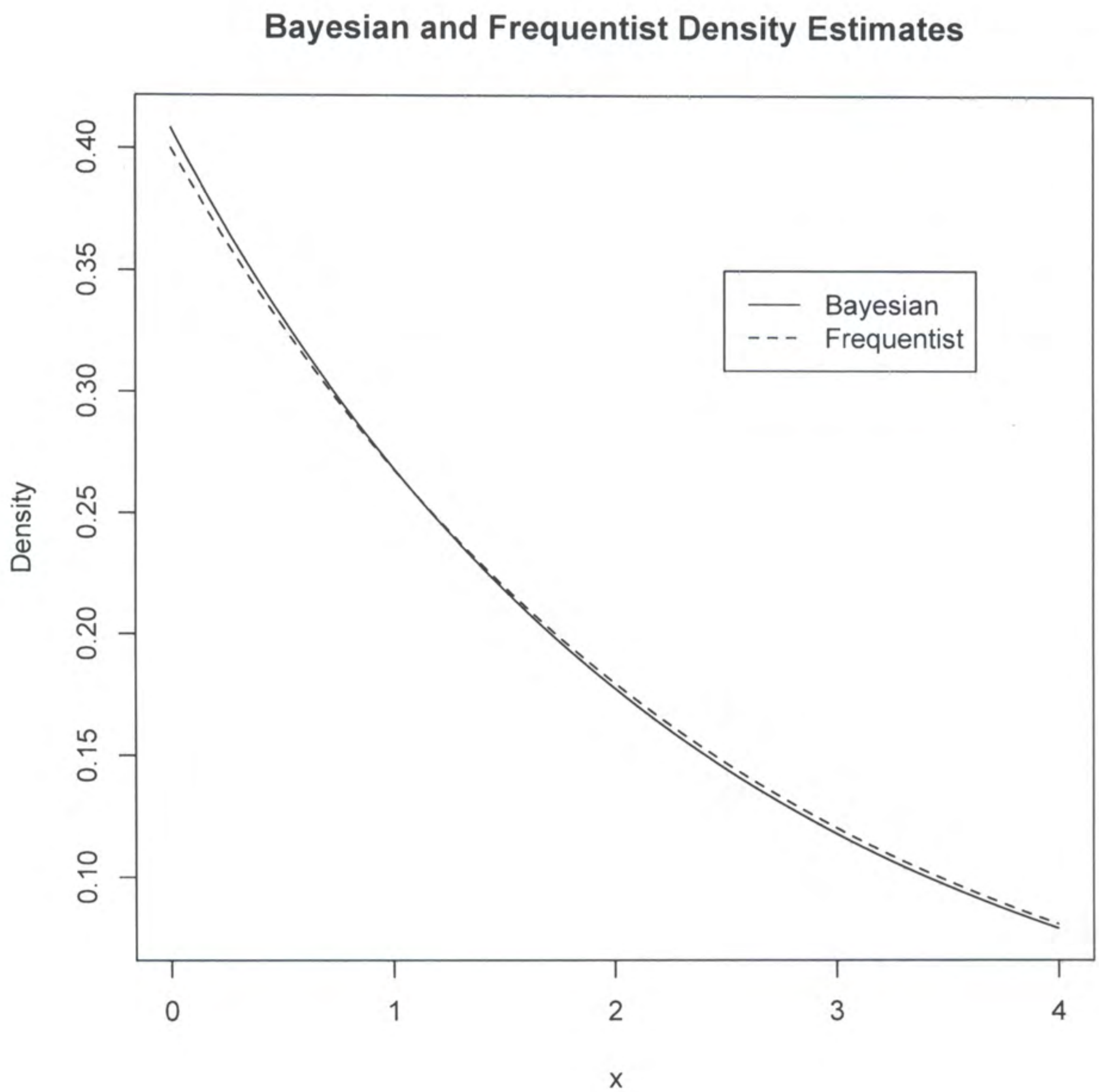
compare

$$\frac{(\alpha+n)(\beta+n\bar{x})^{\alpha+n}}{(\beta+n\bar{x}+x)^{\alpha+n+1}} I(x>0)$$

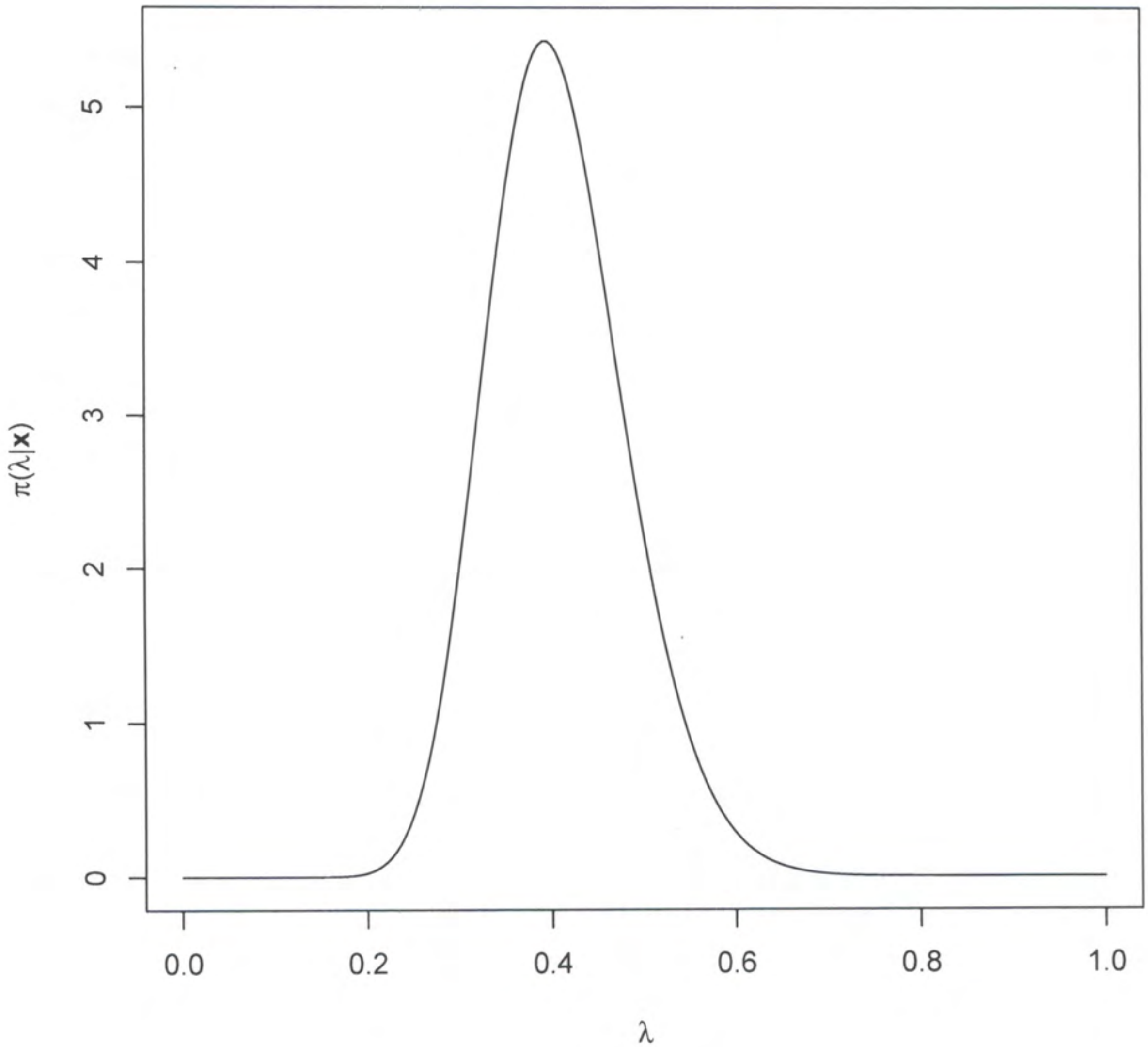
For simplicity let parameters of the prior gamma distribution be $\alpha=\beta=1$
Standard exponential. Predictive density is

$$\frac{(1+29)\left[1+(29)(2.5)\right]^{1+29}}{(1+(29)(2.5)+x)^{31}} I(x>0)$$

$$= \frac{2.92 * 10^{57}}{(73.5+x)^{31}} I(x>0)$$


Bayesian and Frequentist Density Estimates
Density
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0
1
2
3
4
x
Bayesian
Frequentist

Posterior Density of λ



$$\int f(x|\theta)\,\pi(\theta|\underset{\sim}{x})\,d\theta$$

Posterior Predictive Density

$$\int f(x \mid \theta)\, \pi(\theta \mid \underset{\sim}{x})\, d\theta$$

Prior Predictive Density

$$\int f(x \mid \theta)\, \pi(\theta)\, d\theta$$

Posterior Density

$$\frac{f(\underset{\sim}{x} \mid \theta)\, \pi(\theta)}{\int f(\underset{\sim}{x} \mid \theta)\, \pi(\theta)\, d\theta}$$

Denominator is a prior predictive joint density

$$\int f(\underset{\sim}{x} \mid \theta)\, \pi(\theta)\, d\theta = \int \frac{f(\underset{\sim}{x}, \theta)}{\cancel{\pi(\theta)}}\, \cancel{\pi(\theta)}\, d\theta$$

$= f(\underset{\sim}{x})$ marginal of the data

Which is better, Bayesian or Frequentist?

- They fight, especially about the prior.
- Frequentists say prior is subjective, unscientific.
- Bayesians reply the prior allows them to use knowledge they have, rather than starting from zero with every study of a particular topic.

  Bayesian methods can keep learning.
  Frequentist methods are like robots, mind-wiped after every task.

In my opinion the Bayesians win all the philosophic arguments, but they are still losing the war.