

Identifying Parametric Prior Distributions

Stephanie Kovalchik
UCLA, Department of Biostatistics
UseR! 2008 Conference

August 14, 2008

In a Bayesian analysis the statistician must specify prior densities for the model parameters.

In a Bayesian analysis the statistician must specify prior densities for the model parameters.

If he is bold enough to choose an informative prior for the model parameter θ , then this prior should well-represent beliefs about θ .

Why be bold?

If there are experts or historical data that have *accurate* information about the model parameters, then we should use this in choosing a prior as this will make our posterior inferences more precise.

Why be bold?

If there are experts or historical data that have *accurate* information about the model parameters, then we should use this in choosing a prior as this will make our posterior inferences more precise.

This point is well made by Garthwaite, Kadane and O'Hagan: 'An aim of much statistical research is to wring as much from data as we possibly can, but using expert opinion better (or using it at all) could add more information than slight improvement in efficiency through better techniques of data analysis' (p. 698, 2005)

Being bold ...

The challenge for the statistician is that beliefs are most commonly expressed as *typical values*, an average or a set of cumulative probabilities. It is not always clear how to translate these beliefs to a specific density.

We need tools to help in this translation process.

$$\textit{Belief} \rightarrow p(\theta)$$

As a start towards this end, a set of R functions have been written to identify a prior distribution for θ when a continuous parametric density can adequately represent prior beliefs about θ .

Available Densities

- ▶ Normal
- ▶ Beta
- ▶ Gamma
- ▶ Inverse-Gamma
- ▶ Student's T

General Function Syntax

For a prior density f with parameters α, β then the function form is

```
f.prior(args)
```

which returns the vector (α, β) and a plot if desired.

What are the args?

Each function attempts to find a matching prior based on some combination of mean, mode, variance and coverage probabilities.

Essentially, the functions are an R-version of the *modal interval* approach discussed by Garthwaite, Kadane and O'Hagan (2005).

Usage by example

Consider the following problem posed by Jim Albert:

'If p denotes the probability of flipping a head, then your 'best guess' at p is .5. Moreover, you believe that it is highly likely that the coin is close to fair, which you quantify by $P(.44 < p < .56) = .9$ ' (p. 55, 2007).

If we want a Beta prior to reflect our beliefs about p , what are the appropriate parameter settings?

If we want a Beta prior to reflect our beliefs about p , what are the appropriate parameter settings?

If we take 'best guess' to mean the mode, then we could use the following as a first attempt at identifying a suitable Beta.

If we want a Beta prior to reflect our beliefs about p , what are the appropriate parameter settings?

If we take 'best guess' to mean the mode, then we could use the following as a first attempt at identifying a suitable Beta.

```
beta.prior(mode=.5,p=.05,q=.44)
```

If we want a Beta prior to reflect our beliefs about p , what are the appropriate parameter settings?

If we take 'best guess' to mean the mode, then we could use the following as a first attempt at identifying a suitable Beta.

```
beta.prior(mode=.5,p=.05,q=.44)
```

Here q is the specified quantile and p its associated probability.

With this specification, we assume that the density was fairly symmetric. To check that this is reasonable we can use the plot option and mark the 90 % coverage interval.

With this specification, we assume that the density was fairly symmetric. To check that this is reasonable we can use the plot option and mark the 90 % coverage interval.

```
beta.prior(mode=.5,p=.05,q=.44,plot=T)
```

With this specification, we assume that the density was fairly symmetric. To check that this is reasonable we can use the plot option and mark the 90 % coverage interval.

```
beta.prior(mode=.5,p=.05,q=.44,plot=T)
```

So we find that $Beta(93.5, 93.5)$ well represents the prior beliefs about p .

With this specification, we assume that the density was fairly symmetric. To check that this is reasonable we can use the plot option and mark the 90 % coverage interval.

```
beta.prior(mode=.5,p=.05,q=.44,plot=T)
```

So we find that $Beta(93.5, 93.5)$ well represents the prior beliefs about p .

Note that Albert suggests $Beta(100, 100)$ which `curve(dbeta(x, 100, 100), add=T, lty=2)` indicates places slightly more mass for values between (.44, .56) than the prior beliefs warrant.

Again, from Albert, suppose our data are $Y_i|\lambda \sim \text{Poisson}(\lambda)$.

Consider the following beliefs and choice of density for the rate λ .

Density	Belief
Gamma	$E[\lambda] = 3$ and $P(\lambda \leq 2.1) = .25$

Again, from Albert, suppose our data are $Y_i|\lambda \sim \text{Poisson}(\lambda)$.

Consider the following beliefs and choice of density for the rate λ .

Density	Belief
Gamma	$E[\lambda] = 3$ and $P(\lambda \leq 2.1) = .25$
$\log(\lambda) \sim \text{Normal}$	$P(1.94 \leq \lambda \leq 3.81) = .5$

Again, from Albert, suppose our data are $Y_i|\lambda \sim \text{Poisson}(\lambda)$.

Consider the following beliefs and choice of density for the rate λ .

Density	Belief
Gamma	$E[\lambda] = 3$ and $P(\lambda \leq 2.1) = .25$
$\log(\lambda) \sim \text{Normal}$	$P(1.94 \leq \lambda \leq 3.81) = .5$
$\log(\lambda) \sim \text{Normal}$	$E[\lambda] = 3$ and $P(\lambda \geq 8) = .02$

Enhancements

- ▶ Use of effective sample size for Beta distribution where $n_{\text{effective}} = \alpha + \beta$ for $\theta \sim \text{Beta}(\alpha, \beta)$.
- ▶ Transformations for Normal and T distributions so beliefs can be expressed in terms of X though the actual model is $f(X) \sim N(\mu, \sigma^2)$ or $\sim T(\mu, \sigma^2 = 1, \nu)$.

Further Extensions

- ▶ More densities.
- ▶ Identifying the prior density that *best* matches prior conditions rather than requiring an exact match for a minimally sufficient set of arguments.
- ▶ Mixture priors.

Source Code and Documentation

`http://skoval.bol.ucla.edu/R.html`

References



Jim Albert.

Bayesian Computation with R.

Springer, New York, 2007.



Paul H. Garthwaite, Joseph B. Kadane, and Anthony O'Hagan.

Statistical methods for eliciting probability distributions.

JASA, 100:680–700, 2005.