ADVANCED BAYESIAN MODELING

BAYESIAN FUNDAMENTALS:
A BINOMIAL EXAMPLE

Reminder:

We specify that a random variable has a named distribution using

$$\cdot \quad \sim \quad name(\cdot)$$

for a marginal distribution and

$$\cdot \mid \cdot \quad \sim \quad name(\cdot)$$

for a conditional distribution.

See BDA3, Tables A.1 and A.2 for naming conventions.

Scenario

You want to start a dog-walking service.

What percentage of households in your community would be potential clients?

You conduct a limited survey of 20 households, finding 2 that would be interested in your service.

Sampling Distribution

- θ = proportion of population with some characteristic
- n = size of a random sample from population
- y = number in sample who have characteristic

Sampling distribution (for a large population):

$$y \mid \theta \sim \mathrm{Bin}(n, \theta)$$

Sampling density (BDA3, Table A.2):

$$p(y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \qquad y = 0, \ldots, n$$

For your survey data,

$$y = 2 \quad \text{responded positively from among} \quad n = 20$$

and θ is the proportion in the entire community who would be interested.

The likelihood:

$$\begin{aligned} p(y = 2 \mid \theta) &= \binom{20}{2} \theta^2 (1 - \theta)^{20-2} \\ &\propto \theta^2 (1 - \theta)^{18} \end{aligned}$$

Remember: Bayes' rule only requires this to be known up to proportionality in θ.

Prior Distribution

You could make a prior for θ based on your best guess (later).

For now, let's be cautious and choose a uniform prior (BDA3, Table A.1):

$$\theta \sim \mathrm{U}(0,1)$$

so that

$$p(\theta) = 1, \qquad 0 < \theta < 1$$

Such a "flat" prior often allows the conclusions to be driven mainly by the data.

Applying Bayes' Rule

The posterior density:

$$
\begin{aligned}
p(\theta \mid y = 2) &\propto 1 \cdot \theta^2 \, (1 - \theta)^{18} \\
&\propto \theta^2 \, (1 - \theta)^{18}, \qquad 0 < \theta < 1
\end{aligned}
$$

Note: This *does* fully define the posterior distribution (even though it is in unnormalized form).

We will avoid computing the normalizing factor.

Scan a table of densities (BDA3, Table A.1) for something similar ...

Find

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}\,(1-\theta)^{\beta-1} \qquad \propto \qquad \theta^{\alpha-1}\,(1-\theta)^{\beta-1}, \qquad 0<\theta<1$$

the density of a $\mathrm{Beta}(\alpha,\beta)$ distribution.

Compare with

$$\theta^2\,(1-\theta)^{18}, \qquad 0<\theta<1$$

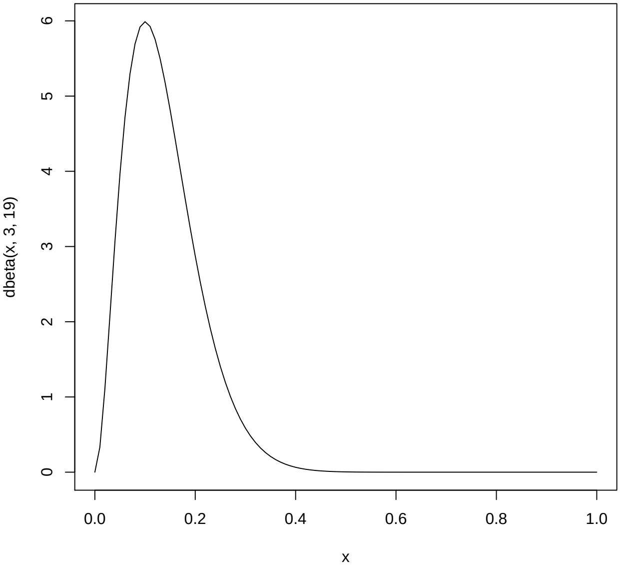
to identify the posterior:

$$\theta \mid y=2 \;\sim\; \mathrm{Beta}(\alpha=3,\beta=19)$$

In the R statistical computing environment, function `dbeta` gives the density of the beta distribution.

Let's view the density curve for the posterior ...

```
> curve(dbeta(x,3,19), 0, 1)
```


dbeta(x, 3, 19)
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
x

Most posterior probability for θ ranges between 0.05 and 0.2.

This is compatible with the "raw" estimate that 10% of households might be interested in the dog-walking service.

Another Prior

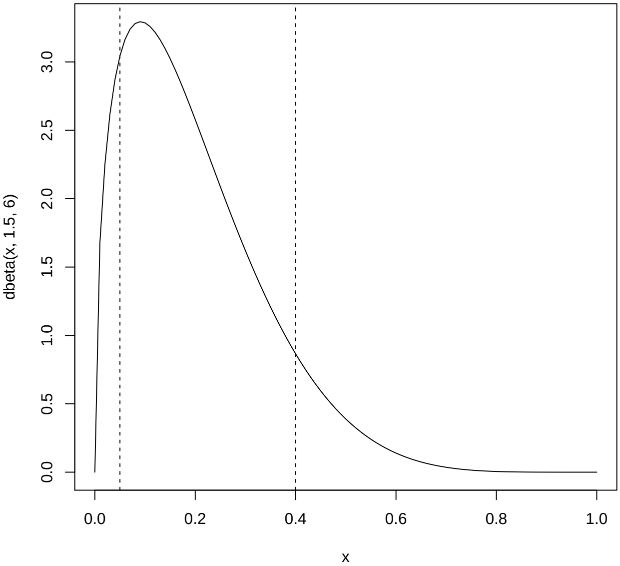
Let's try a more **informative** prior – one that incorporates prior beliefs (possibly subjective).

Suppose you were fairly certain that between 5% and 40% of households would be interested.

With some trial and error, you find that a $\mathrm{Beta}(\alpha = 1.5, \beta = 6)$ distribution gives roughly 80% probability to the range $(0.05, 0.4)$:

```
> pbeta(c(0.05,0.4), 1.5, 6)  # beta distribution function
[1] 0.1128297 0.9049012

> curve(dbeta(x,1.5,6), 0, 1)
> abline(v=c(0.05,0.4), lty=2)
```


dbeta(x, 1.5, 6)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
x

For the $\mathrm{Beta}(\alpha = 1.5, \beta = 6)$ prior,

$$
\begin{aligned}
p(\theta) &\propto \theta^{1.5-1}\,(1-\theta)^{6-1} \\
&\propto \theta^{0.5}\,(1-\theta)^{5}, \qquad 0 < \theta < 1
\end{aligned}
$$

so (by Bayes' rule) the posterior

$$
\begin{aligned}
p(\theta \mid y = 2) &\propto \theta^{0.5}\,(1-\theta)^{5} \cdot \theta^{2}\,(1-\theta)^{18} \\
&\propto \theta^{2.5}\,(1-\theta)^{23}, \qquad 0 < \theta < 1
\end{aligned}
$$

is a $\mathrm{Beta}(\alpha = 3.5, \beta = 24)$ distribution.

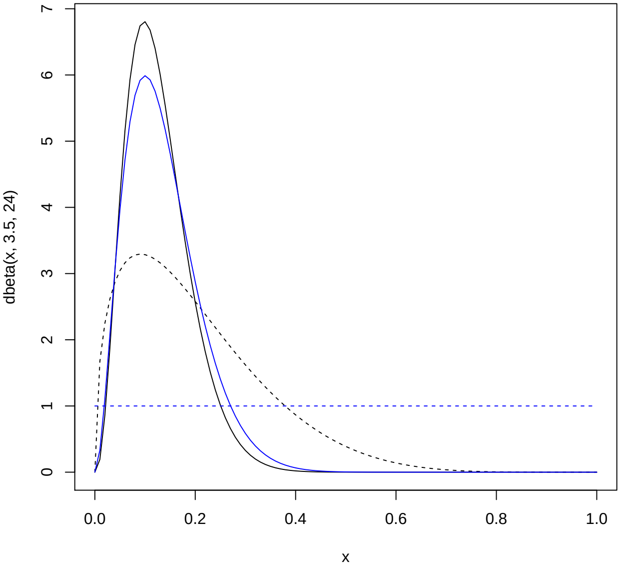
Notice: Using a beta prior produced a beta posterior.

This is **conjugacy**: A family of priors is **conjugate** for a sampling distribution if the posterior is always of that same family.

While not necessary, a conjugate prior is often convenient.

Let's view our priors and posteriors:

```
curve(dbeta(x,3.5,24), 0, 1)  # posterior (informative prior)
curve(dbeta(x,1.5,6), 0, 1, add=TRUE, lty=2)  # informative prior
curve(dbeta(x,3,19), 0, 1, add=TRUE, col="blue")  # posterior (flat prior)
curve(dbeta(x,1,1), 0, 1, add=TRUE, col="blue", lty=2)  # flat prior
```



dbeta(x, 3.5, 24)
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0
x

Note:

- Both posteriors are very similar (driven mainly by the data).
- The more informative prior led to a slightly more concentrated posterior.

Next: How can you use the posterior?