Single Parameter Models

BDA3

Sydney

03/02/22

Chapter Overview

- Estimating one-dimensional parameters.
- Investigate four fundamental and widely used models: Binomial, Normal, Poisson and Exponential.

Binomial model - Proportion of Female Births

A Binomial model for estimating proportions:

$$p(\theta|y) \propto Bin(y|n,\theta)p(\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}p(\theta)$$

We know the form of the posterior distribution. By matching parameters and substituting a uniform distribution:

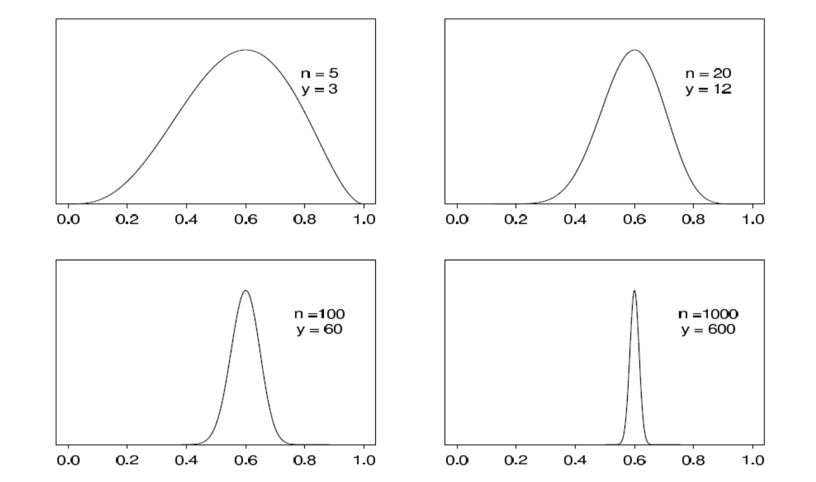
$$p(\theta) = Beta(\alpha = 1, \beta = 1) \propto 1$$

Binomial model - Proportion of Female Birth

- The density of a beta distribution (ignoring the normalisation constant) has the same form as the Binomial.
- We can simply collect terms to read off the shape parameters of the posterior.

$$
\begin{aligned}
p(\theta|y) &\propto \theta^{y}(1-\theta)^{n-y} \\
&= \theta^{y=\alpha-1}(1-\theta)^{n-y=\beta-1} \\
&\propto Beta(y+1, n-y+1)
\end{aligned}
$$

Beta Posteriors



Bayes is a compromise

If you imagine Bayesian computation as some function of the prior distribution, it seems natural to suspect a relation with the posterior distribution.

$$\mathbb{E}[\theta] = \mathbb{E}_y\Big[\mathbb{E}_\theta[\theta|y]\Big] = \int \mathbb{E}[\theta|y]p(y)dy$$

The prior mean is the expected value of the posterior, averaged over the distribution of possible data[1].

[1] This idea forms the basis behind Simulation Based Calibration a self consistency check for Bayesian computation

Bayes is a compromise

The variance of the prior is larger than the average posterior variation.

$$\mathbb{E}\Big[var(\theta|y)\Big] = var(\theta) - var\Big(\mathbb{E}[\theta|y]\Big)$$

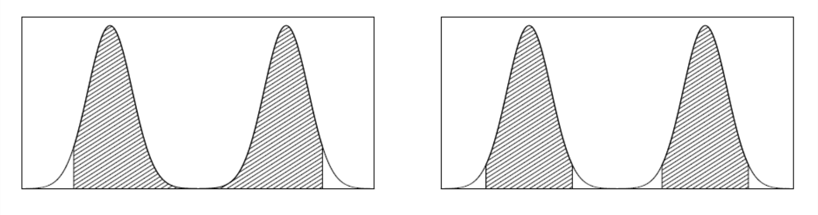
This discrepancy is dependant on the variation of posterior means over the distribution of data:

Summarising posterior inference

The posterior distribution contains all *current* information about θ. We need to consider appropriate ways to convey our posterior result:

- Summary statistics
    - Mean, median, standard deviations etc. computed from theoretical results or more practically though (possibly transformed) posterior samples.
- Intervals
    - IQR, **Highest Density Intervals**, quantiles...

Summarising posterior inference



The HDI gives the smallest interval that contains (100-α)% of the posterior density. For symmetric distributions this is the same as a centered interval containing the same density percentage.

Constructing prior distributions

How can we reflect substantive information in our prior distributions?

- **Population**: We may have some accepted, and objectively true information about the population.
- **State of Knowledge**: We know little about the population in general. We can make informed guesses that include plausible values.

Probability of girl birth

- Take *placenta previa* as a factor that could influence the sex ratio.
- Early studies found that out of a total of 980 births with this condition, 437 of them were female.

Using Bayes can we demonstrate how much evidence this provides to the claim :

The proportion of placenta previa births that are female are lower than the female population proportion

Probability of Girl birth given placenta previa

Starting with a uniform prior:

$$p(\theta|y) \propto Bin(y|n,\theta)p(\theta)$$

Using the result from earlier our posterior is a Beta distribution:

$$Beta(y+1, n-y+1) = Beta(438, 544)$$

This gives a posterior mean of 0.446 and an standard deviation of 0.016.

Frame Title

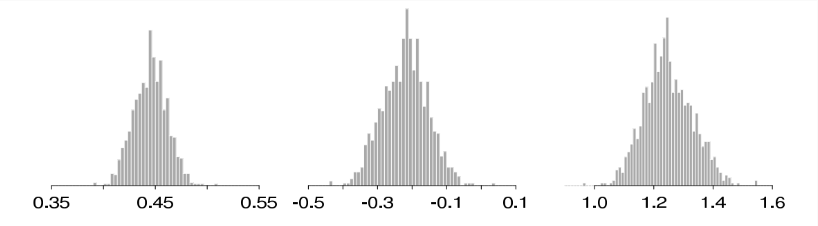
Exact quantiles can be calculated for the Beta distribution, however a normal approximation is also reasonable in this example.

$$0.455 \pm 1.96 \cdot 0.016$$

- In practise we are often limited to approximation of posterior statistics as the analytical result is not available.
- Most of the time we will find ourselves estimating such statistics from **posterior samples**.

Probability of Girl birth given placenta previa

Transforming posterior samples:



θ $\quad$ $\text{logit}(\theta)$ $\quad$ $\frac{(1-\theta)}{\theta}$

Probability of girl birth

- The sensitivity analysis would suggest that the observed results carry some weight for the hypothesis.
- In all analysis shown the population value 0.485 lies outside the HDI

Parameters of the prior distribution		Summaries of the posterior distribution	
$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	[0.415, 0.477]
0.485	2	0.446	[0.415, 0.477]
0.485	5	0.446	[0.415, 0.477]
0.485	10	0.446	[0.415, 0.477]
0.485	20	0.447	[0.416, 0.478]
0.485	100	0.450	[0.420, 0.479]
0.485	200	0.453	[0.424, 0.481]

Normal models

Interpreting the general conjugate result The posterior predictive.

$$p(\theta|y, \sigma, \tau) \propto N(y|\theta, \sigma)N(\theta|\mu, \tau)$$

$$\mathbb{E}[\theta|y] = \frac{\frac{1}{\sigma_j^2}y_j + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}$$
$$\mathbb{V}[\theta|y] = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}$$

The posterior mean is just a weighted average.

Posterior Predictive - Normal

$$
\begin{aligned}
p(\tilde{y}|y) &= \int p(\tilde{y}|\theta)\, p(\theta|y)\, d\theta \\
&\propto \int \exp\left(\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2\right) \exp\left(\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)
\end{aligned}
$$

You could expand out as per usual, however note that the exponent is a quadratic function of both $\tilde{y}$ and θ.

Marginalizing the joint Gaussian distribution is straightforward.

Posterior Predictive - Normal

A more practical way to think about this computation is to use a sampling approach.

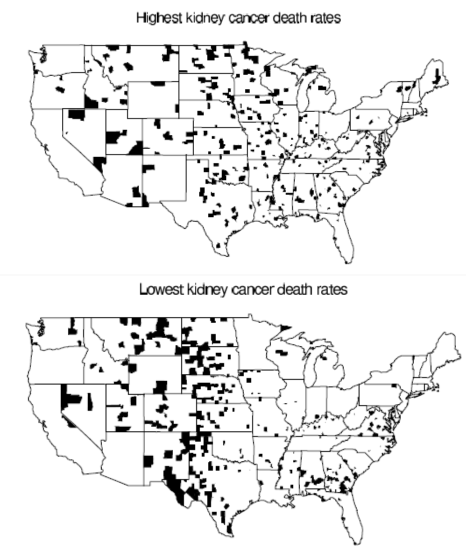
$$\begin{aligned} \theta^s &\sim N(\mu_1, \tau_1^2) \\ \tilde{y} &\sim N(\theta^s, \sigma_1^2) \quad s = 1 \dots N \end{aligned}$$

```
generated quantities{
    real y_tilde;
    y_tilde = normal_rng(theta, sigma);
}
```

A model for kidney cancer

- Low sample sizes can inflate effects out of proportion.
- Consider the following two maps: It appears that certain areas can simultaneously have high and low kidney cancer related deaths.

A model for kidney cancer



A model for kidney cancer

$$\theta_j \sim \text{Gamma}(\alpha = 20, \beta = 430,000)$$
$$y_j \sim \text{Poisson}(10n_j \cdot \theta_j)$$

We multiply by 10 as the data is counts over a 10 year period θ is then interpreted as a yearly rate.

This gives us a Gamma posterior

$$p(\theta_j|y_j) \sim \text{Gamma}(20 + y_j, 430,000 + 10n_j)$$

A model for kidney cancer

The takeaway points:

- The prior dominates in counties with low population.
- For counties with large sample sizes the data wins out.

This is clear from how the posterior mean changes between the two cases.

$$\mathbb{E}[\theta_j|y_j] = \frac{20 + y_j}{430000 + 10n_j}$$

$$\mathbb{V}[\theta_j|y_j] = \frac{20 + y_j}{(430000 + 10n_j)^2}$$

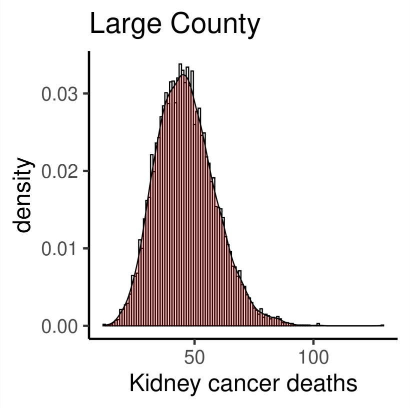
Prior Predictive Distribution

The prior predictive distribution for this example is the Negative Binomial:

$$y_j \sim NB(\alpha, \beta/10n_j)$$

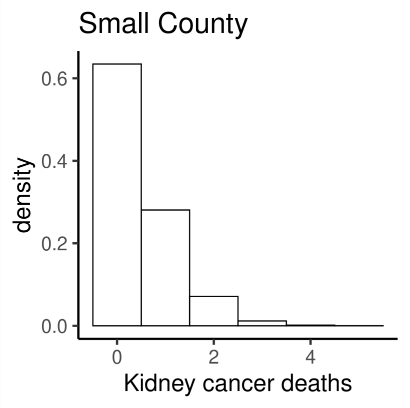
Prior Predictive Distribution

Population : $1,000,000$



Prior Predictive Distribution

Population : $10,000$



Non-informative Prior Distributions

- Reference priors, vague, flat, diffuse and non-informative...
- Priors the with a *minimal* role in the posterior distribution.

Non-informative Prior Distributions

- Searching for a vague posterior in cases where the likelihood is dominant is misguided: the choices should not matter.
- A non informative density for one parameterisation may well be informative for another parameterisation[2]. On what scale should the **principle of insufficient reason** apply?

[2]This is the motivation behind a Jefferys' prior

Improper priors

- A prior will typically be a well defined distribution – it integrates to one.
- Improper priors are prior distributions that don't have a finite integral :

$$p(\theta) \propto 1 \rightarrow \int_{\theta \in \Theta} p(\theta) d\theta = \infty$$

Improper priors can lead to improper posteriors!

Improper priors - Example

Assume a normal with known mean and an improper prior for the standard deviation.

$$p(\sigma) \propto 1$$

Given a single observation the posterior is proportional to:

$$p(\sigma|\theta, y) \propto \frac{\exp c/\sigma^2}{\sigma}$$

This is a divergent integral.

$$\int_0^\infty \frac{\exp c/\sigma^2}{\sigma} = \infty$$

Improper priors - Example

However with just one more observation[3] the integral becomes finite.

$$\int_0^\infty \frac{\exp c/\sigma^2}{\sigma^2} = \frac{\sqrt{\pi}}{2\sqrt{c}}$$

This is interesting: data has updated our prior into a well behaved distribution (after renormalisation).

[3] Example taken from *https://algassert.com/post/1630*

Weakly Informative Priors

- Intentionally provide weaker information than actually available.
- Constraining a parameter to plausible values but very weakly. E.g births weakly described by a normal such as $N(0.5, 0.1)$.

Weakly Informative Priors

Weakly informative priors are often easier to think about if we have an idea about the scale of our parameter:

- **Unit scale** : A $N(0, 1)$ for a scale parameter is weakly informative.
- **Log scale**: Multiplicative ranges where we expect our population value to be.

Here is one idea for quantifying the informativeness of a prior[4]

$$c[f|y] = 1 - \frac{\mathbb{V}_{post}[f|y]}{\mathbb{V}_{prior}[f|y]}$$

A practical discussion about suitable priors can be found here[5]

[4] *https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html*

[5] *https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations*

Questions Next Week

- Question 11: Cauchy Distribution
- Question 19: Exponential Distribution