Single-parameter models

February 26, 2023

Overview

1. Binomial distribution

2. Informative prior distributions

Estimating a probability from binomial data

- In our previous study, we consider the binomial distribution as a two-parameter distribution. But because the number of trials *n* is easy to be assigned. So here we consider *n* as a condition which is fixed.
- The parameter to be considered here is only the probability of success *p*, so in the rest part about binomial distribution, we use θ to represent the probability *p* as the parameter we want to know.

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

Estimating a probability from binomial data

Example: Estimating the probability of a female birth.

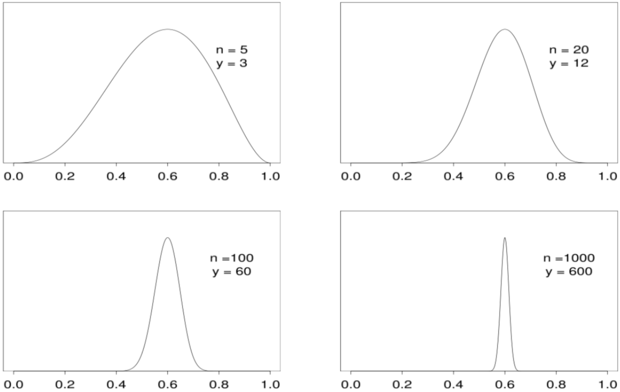
- Historical note shows the proportion of female birth in Europe was less than 0.5 (0.485).
- In order to know the exact proportion, we define the parameter θ to be it.
- Let y be the number of girls in *n* recorded births.
- To perform Bayesian inference in the binomial model to know the posterior probability of θ, $p(\theta|y)$, we must specify a prior distribution for θ.
- So far, for simplicity, we set the prior distribution as a uniform distribution with interval $[0, 1]$. What is the density function?
- Then, the posterior density for θ is

$$p(\theta|y) \propto \theta^y (1-\theta)^{n-y}.$$

Estimating a probability from binomial data

Example: Estimating the probability of a female birth.

- In single-parameter problems, this allows immediate graphical presentation of the posterior distribution. Each graph has the same proportion.



Estimating a probability from binomial data

Example: Estimating the probability of a female birth.

- The previous posterior distribution of θ could be recognized as the unnormalized from of the beta distribution

$$\theta|y \sim \mathrm{Beta}(y+1, n-y+1).$$

- The Appendix A provides more information of related distributions.

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad \theta \in [0,1]$$

$$\mathrm{E}(\theta) = \frac{\alpha}{\alpha+\beta}$$
$$\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
$$\mathrm{mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$$

History Note

Bayes wanted to know the probability $Pr(\theta \in (\theta_1, \theta_2)|y)$, with the assumption that prior is a uniform distribution on $[0, 1]$.

- Then Bayes obtained

$$
\begin{aligned}
\Pr(\theta \in (\theta_1, \theta_2)|y) &= \frac{\Pr(\theta \in (\theta_1, \theta_2), y)}{p(y)} \\
&= \frac{\int_{\theta_1}^{\theta_2} p(y|\theta)p(\theta)d\theta}{p(y)} \\
&= \frac{\int_{\theta_1}^{\theta_2} \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta}{p(y)}.
\end{aligned}
$$

History Note

- Bayes succeeded in evaluating the denominator

$$
\begin{aligned}
p(y) &= \int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta \\
&= \frac{1}{n+1} \quad \text{for } y = 0, \ldots, n.
\end{aligned}
$$

which shows that all possible values of y are equally likely a prior.

- The integration trick of Beta distribution could be used in the calculation.

$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$ $\theta \in [0,1]$	$\mathrm{E}(\theta) = \frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$

Prediction

- For posterior prediction from this model, the outcome of one new trial, denoting as $\tilde{y}$ with the first *n* sample would be

$$
\begin{aligned}
\Pr(\tilde{y} = 1|y) &= \int_0^1 \Pr(\tilde{y} = 1|\theta, y) p(\theta|y) d\theta \\
&= \int_0^1 \theta p(\theta|y) d\theta = \mathrm{E}(\theta|y) = \frac{y+1}{n+2},
\end{aligned}
$$

The result follows the expectation of a beta distribution.

Posterior as compromise between data and prior information

We go through the equations again with more details:

$$\mathrm{E}(\theta) = \mathrm{E}(\mathrm{E}(\theta|y))$$

$$\mathrm{var}(\theta) = \mathrm{E}(\mathrm{var}(\theta|y)) + \mathrm{var}(\mathrm{E}(\theta|y)),$$

- The prior mean of θ is the average of all possible posterior means over the distribution of possible data.
- The posterior variance is on average smaller than the prior variance, by an amount that depends on the variation in posterior means over the distribution of possible data.

Summarizing posterior inference

- Visualization: because we know the posterior function, sampling methods such as MCMC could be used for the visualization.
- Descriptive statistics, especially for the distributions with closed form.
- Posterior uncertainty: intervals. Intervals can also be computed with computer simulations from the posterior distribution. But for the binomial distribution, the interval of θ might be less appropriate than the interval of the odds ratio or even log odds ratio, because of the bounds.

Informative prior distributions

Two interpretations that can be given to prior distributions.

- Population interpretation: the prior distribution represents a population of possible parameter values.
- state of knowledge interpretation: the guiding principle is we must express our knowledge about θ

Typically, the prior distribution should include all plausible values of θ, but the distribution need not be realistically concentrated around the true value.

Binomial example with different prior distributions

- Considered as a function of θ, the likelihood is of the form

$$p(y|\theta) \propto \theta^a (1-\theta)^b.$$

- If the prior density is of the same form, with its own values a and b, then the posterior density will also be of this form.
- We will parameterize such a prior density as

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1},$$

  which is a beta distribution with parameters α and β.

Binomial example with different prior distributions

- Comparing $p(\theta)$ and $p(y|\theta)$ suggests that this prior density is equivalent to $\alpha - 1$ prior successes and $\beta - 1$ prior failures.
- The parameters of the prior distribution are often referred to as hyperparameters.
- By selecting reasonable values of α and β, the posterior density for θ is

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

Binomial example with different prior distributions

- The property that the posterior distribution follows the same parametric form as the prior distribution is called conjugacy.
- The beta prior distribution is a conjugate family for binomial likelihood.
- conjugate family is mathematically convenient in that the posterior distribution follows a known parametric form.

Binomial example with different prior distributions

- With the posterior probability as a beta distribution, the expectation of parameter θ is

$$\mathrm{E}(\theta|y) = \frac{\alpha + y}{\alpha + \beta + n},$$

- the variance of posterior parameter θ is

$$\mathrm{var}(\theta|y) = \frac{(\alpha + y)(\beta + n - y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)} = \frac{E(\theta|y)[1 - E(\theta|y)]}{\alpha + \beta + n + 1}.$$

Check what would happen when n and y increase.

Conjugate prior distributions

Definition

- If $\mathcal{F}$ is a class of sampling distributions $p(y|\theta)$, and $\mathcal{P}$ is a class of prior distributions of θ, then the class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$p(\theta|y) \in \mathcal{P} \text{ for all } p(\cdot|\theta) \in \mathcal{F} \text{ and } p(\cdot) \in \mathcal{P}.$$

- We are most interested in natural conjugate prior families, which arise by taking $\mathcal{P}$ to be the set of all densities having the same functional form as the likelihood.
- Conjugate prior distributions provide computational convenience.

Conjugate distributions, exponential families, and sufficient statistics

- Probability distributions that belong to an exponential family have natural conjugate prior distributions.
- For complete generality in this section, we allow data points y_i and parameters θ to be multidimensional.
- The class $\mathcal{F}$ is an exponential family if all its members have the form,

$$p(y_i|\theta) = f(y_i)g(\theta)e^{\phi(\theta)^T u(y_i)}.$$

- The factors $\phi(\theta)$ and $u(y_i)$are vectors of equal dimension to that of θ.

Conjugate distributions, exponential families, and sufficient statistics

- The vector $\phi(\theta)$ is called the 'natural parameter' of the family $\mathcal{F}$.
- With i.i.d. observations $y = (y_1, \ldots, y_n)$ likelihood is

$$p(y|\theta) = \left(\prod_{i=1}^{n} f(y_i)\right) g(\theta)^n \exp\left(\phi(\theta)^T \sum_{i=1}^{n} u(y_i)\right).$$

Conjugate distributions, exponential families, and sufficient statistics

- For all n and y, this has a fixed form (as a function of θ),

$$p(y|\theta) \propto g(\theta)^n e^{\phi(\theta)^T t(y)}, \quad \text{where } t(y) = \sum_{i=1}^{n} u(y_i).$$

- the quantity $t(y)$ is said to be a sufficient statistic for θ, because the likelihood for θ depends on the data y only through the value of $t(y)$.

Conjugate distributions, exponential families, and sufficient statistics

- If the prior density is specified as

$$p(\theta) \propto g(\theta)^{\eta} e^{\phi(\theta)^T \nu},$$

- The posterior density is

$$p(\theta|y) \propto g(\theta)^{\eta+n} e^{\phi(\theta)^T (\nu+t(y))},$$

which shows that this choice of prior density is conjugate.

Example. Probability of a girl birth given placenta previa

- As a specific example of a factor that may influence the sex ratio, we consider the maternal condition placenta previa, an unusual condition of pregnancy in which the placenta is implanted low in the uterus, obstructing the fetus from a normal vaginal delivery. An early study concerning the sex of placenta previa births in Germany found that of a total of 980 births, 437 were female. How much evidence does this provide for the claim that the proportion of female births in the population of placenta previa births is less than 0.485, the proportion of female births in the general population?

Example. Probability of a girl birth given placenta previa

Analysis using a uniform prior distribution

- Posterior: $\mathit{Beta}(438, 544)$
- Mean of the Beta distribution: $438/(438+544) \approx 0.446$
- Median (approximately): $(438-1/3)/(544+438-2/3) \approx 0.446$
- Standard deviation: $\sqrt{\frac{438*544}{(438+544)^2*(438+544+1)}} \approx 0.016$
- With normal approximation, we can have the 95% posterior interval is $[0.415, 0.477]$.

The formulas comes from Appendix A. And we will talk about normal approximation later when talking Hierarchical models.

Example. Probability of a girl birth given placenta previa

Analysis using a uniform prior distribution

- Another method we can use is simulation, especially when the direct calculation on the posterior distribution is not feasible.
- We can draw 1000 samples from $Beta(438, 544)$, then estimate the 95% posterior interval by taking the 25th and 976th quantiles.
- We can also use the approximation method to estimate mean, median, and variance. Then, again, use normal approximation to find the interval.

Analysis using a uniform prior distribution

- As we discussed previously, the probability might be not that appropriate for normal approximation, but a logit transformation $log(\frac{\theta}{1-\theta})$ can make it more appropriate.

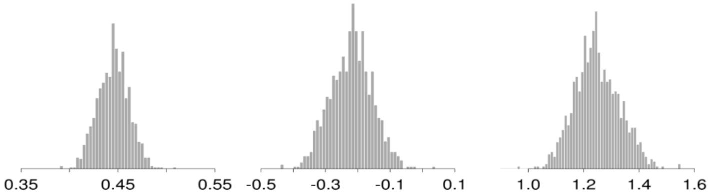


Figure 2.3 *Draws from the posterior distribution of (a) the probability of female birth, θ; (b) the logit transform, logit(θ); (c) the male-to-female sex ratio, $\phi = (1-\theta)/\theta$.*

Example. Probability of a girl birth given placenta previa

Analysis using different conjugate prior distributions.

- We can change the hyperparameters of the conjugate prior distribution to check the impacts from different prior information.

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	$[0.424, 0.481]$