Binomial Likelihood

Suppose that a r.v. X obeys a Binomial distribution $\mathcal{B}in(p)$, characterized by the following Probability Mass Function (PMF)

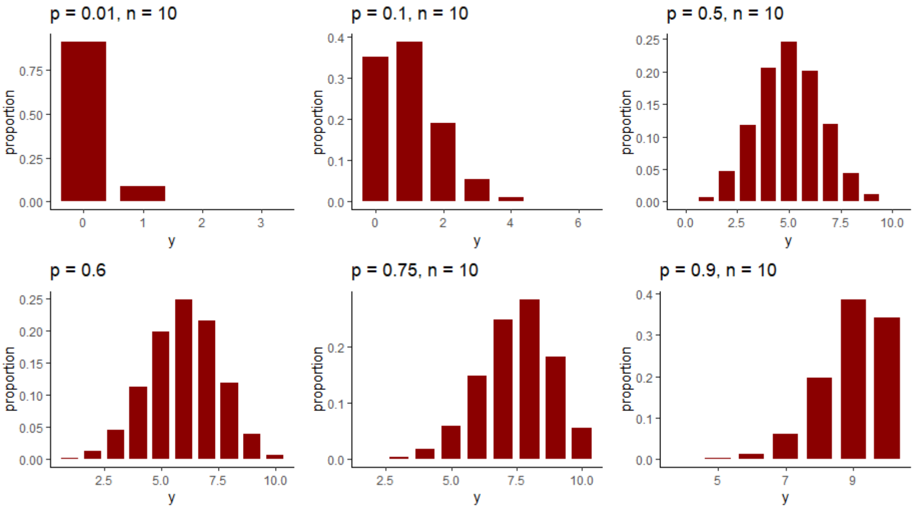
$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

If we observe a sample $x_1, ..., x_n$, then the likelihood is just the product of the individual PMF

$$\begin{aligned} \pi(x_1, ..., x_n \mid p) &= \prod_{i=1}^{n} f_p(x_i) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} \\ &= \prod_{i=1}^{n} \binom{n}{x_i} \; p^{\sum_{i=1}^{n} x_i} \; (1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

where the model parameter p estimated by $\hat{p}$ is the proportion of successes in n independent Bernoulli trials obtained by $\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Some examples of Binomial samples



Beta prior

In Bayesian statistics, to model the uncertainty about p, a convenient prior would be the Beta distribution $Beta(\alpha, \beta)$, since its support is the interval $[0, 1]$. The Beta distribution has the following Probability Density Function (PDF)

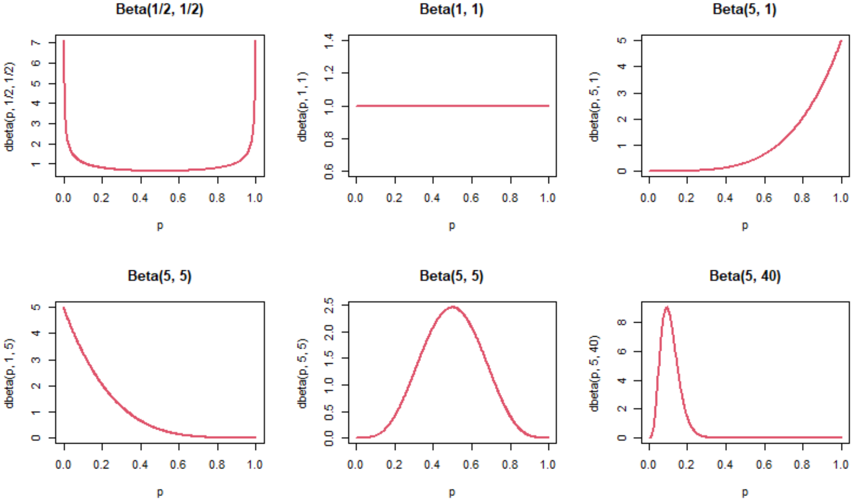
$$\pi(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

where α and β are shape parameters, $B(\alpha, \beta)$ is the Beta function and $\Gamma(\alpha)$ is the Gamma function, defined as $\int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha - 1)!$.

Let $p \sim Beta(\alpha, \beta)$. Then $E[p] = \frac{\alpha}{\alpha+\beta}$ and $var(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

It is a convenient and flexible choice since the Beta distribution can take a wide variety of shapes.

Some examples of the Beta family



Beta posterior (1/2)

The posterior distribution for p if X is Binomial and a Beta prior is chosen for p will also have the functional form of a Beta r.v. That is why we say that Beta is the conjugate prior for a Binomial likelihood. Using the Bayes theorem, we have that

$$\pi(p \mid x_1, ..., x_n) = \frac{\pi(x_1, ..., x_n \mid p)\pi(p)}{\pi(x_1, ..., x_n)}$$

and since $\sum_{i=1}^{n} x_i = n\bar{x}$, we have

$$\begin{aligned}
&\prod_{i=1}^{n} \binom{n}{x} \, p^{\sum_{i=1}^{n} x_i} \, (1-p)^{n-\sum_{i=1}^{n} x_i} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\
&= \underbrace{\prod_{i=1}^{n} \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{do NOT depend on } p} \; p^{n\bar{x}}(1-p)^{n-n\bar{x}} p^{\alpha-1}(1-p)^{\beta-1} \\
&\propto p^{n\bar{x}+\alpha-1}(1-p)^{n-n\bar{x}+\beta-1}
\end{aligned}$$

Beta posterior (2/2)

So we find that $p \mid x_1, ..., x_n \propto Beta(n\bar{x} + \alpha, n - n\bar{x} + \beta)$. The posterior mean and variance are given by

$$E[p \mid x] = \frac{n\bar{x} + \alpha}{n + \alpha + \beta} \qquad var(p \mid x) = \frac{(n\bar{x} + \alpha)(n - n\bar{x} + \beta)}{(n + \alpha + \beta)^2(n + \alpha + \beta + 1)}$$

where $n\bar{x} = \sum_{i=1}^{n} x_i$ is the average number of successes in the sample size. Let's take a few examples and plot the likelihood, a possible prior and the posterior, all at once in R.

General remark on Bayesian inference

Unlike in the traditional Frequentist framework, the Bayesian approach views parameters as random variables rather than fixed, unknown quantities. Given a Poisson sample $x_1, ..., x_n$ and a Poisson parameter λ, from the Bayes theorem, we can write

$$\pi(p \mid x_1, ..., x_n) = \frac{\pi(x_1, ..., x_n \mid p)\ \pi(p)}{\pi(x_1, ..., x_n)}$$

Adopting the 'proportional' notation, the constant term in the denominator is dropped so that the above expression is rewritten as $\pi(p \mid x_1, ..., x_n) \propto \pi(x_1, ..., x_n \mid p)\ \pi(p)$

When conjugate models are used (as in the case of a Beta-Binomial model), the posterior distribution can be identified and closed-form quantities of interest like a mean, a variance or quantiles can be computed. Most of the time in practice, the posterior distribution is intractable so that it is necessary to resort to MCMC techniques.

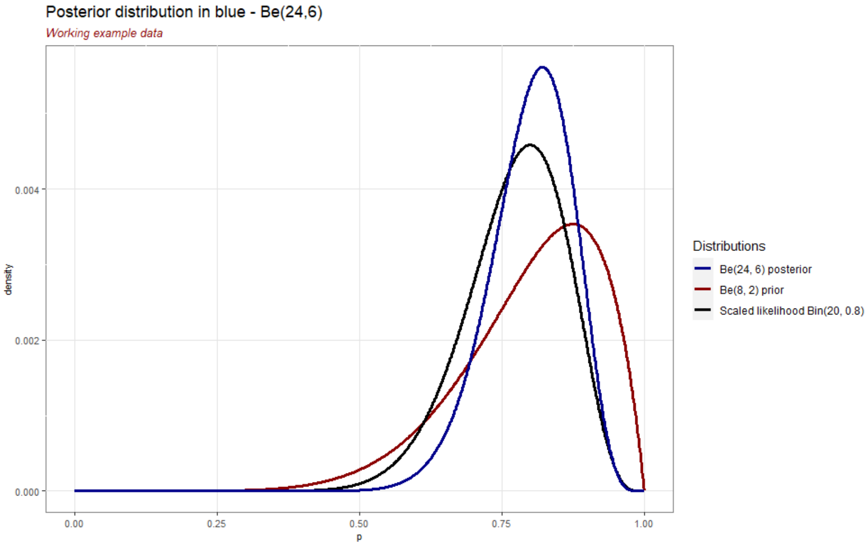
Working example

Suppose that we know that about 80% of the students in a class will pass a mathematics exam. For 20 students , we recorded the results and have the following data at hands ($1 =$ pass, $0 =$ fail).

$$x_i = 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1$$

So assuming a Binomial likelihood with parameter $p = 0.80$, namely a $\mathcal{B}in(20, 0.8)$ likelihood for the data and using a $\mathcal{B}e(8, 2)$ prior, with mean $8/(8 + 2) = 0.8$ and variance $16/((10)^2(11)) \approx 0.0145$, what is the posterior mean and the 95% credible intreval for the model parameter?

Working example: posterior



Posterior quantities obtained from direct sampling

```
# Posterior mean, posterior variance and 95% Credible Interval including the
    sample median
xbar = p1
alpha_prior = 2; beta_prior = 8
alpha_posterior = ((alpha1 + n*mean(data1))) # 24
beta_posterior = (n - n*mean(data1) + beta1) # 6

pmean = alpha_posterior / (alpha_posterior + beta_posterior)
pmean
# [1] 0.8

pvariance = (alpha_posterior *beta_posterior) / ((alpha_posterior + beta_
    posterior)^2 + (alpha_posterior + beta_posterior + 1) )
pvariance
# [1] 0.1546724

# 95% Cedible Interval obtained by direct sampling (simulation)
set.seed(2023)
round(quantile(rbeta(n = 10^8, alpha_posterior, beta_posterior), probs = c
    (0.025, 0.5, 0.975)),4)
#   2.5%    50%  97.5%
# 0.6423 0.8067 0.9201

# Posterior mean obtained from direct sampling
set.seed(2023)
mean(rbeta(n = 10^8, alpha_posterior, beta_posterior))
# [1] 0.7999991
```

Working example: in conclusion

So the theoretical posterior mean is given by

$$E[p] = \frac{n\bar{x} + \alpha}{n + \alpha + \beta} = \frac{20 * 0.8 + 8}{20 + 8 + 2} = 24/30 = 0.8$$

By direct sampling, using 10^8 number of simulations, the posterior sample mean is 0.79999

By direct sampling, a 95% Credible Interval is given by

$$[0.6423, 0.9201]$$

So, combining modeling and simulations, we are now able to generalize and infer to the whole population of students taking this exam those values from a sample of size 20.

Jeffreys prior: definition

Let us first give the general definition of the Jeffreys' prior $p_J(\theta)$:

$$p_J(\theta) = \sqrt{\mathcal{I}(\theta)}$$

where $\mathcal{I}(\theta)$ is the Fisher information of the sample. The Fisher information is defined as follows:

$$\mathcal{I}(\theta) = -E_\theta \left[\frac{\partial^2 ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta^2} \right]$$

or equivalently

$$\mathcal{I}(\theta) = var_\theta \left(\frac{\partial ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta} \right) = E_\theta \left[\left(\frac{\partial^2 ln\mathcal{L}(\theta \mid \mathbf{x})}{\partial\theta^2} \right)^2 \right]$$

The first derivative of the log-likelihood function with respect to the model parameter $\frac{\partial ln\mathcal{L}(\theta|\mathbf{x})}{\partial\theta}$ is sometimes referred to as the score function.

Likelihood and log-likelihood functions for a Binomial model

The likelihood function, denoted and log-likelihood functionfor a Binomial model are respectively given by

$$\begin{aligned}\mathcal{L}(p \mid \mathbf{x}) &= \prod_{i=1}^{n} f_p(x_i) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \\ &= \prod_{i=1}^{n} \binom{n}{x_i} \; p^{\sum_{i=1}^{n} x_i} \; (1-p)^{n-\sum_{i=1}^{n} x_i}\end{aligned}$$

$$\begin{aligned}&l(p \mid \mathbf{x}) = ln(\mathcal{L}(p \mid \mathbf{x})) \\ &\qquad = \sum_{i=1}^{n} \; ln\left(\binom{n}{x_i}\right) + \sum_{i=1}^{n} x_i \; ln(p) + \left(n - \sum_{i=1}^{n} x_i\right) \; ln(1-p)\end{aligned}$$

Score function and second derivative of the log-likelihood function for a Binomial model

The score function, denoted $\frac{\partial ln\mathcal{L}(p|\mathbf{x})}{\partial p}$ and second derivative of the log-likelihood function, denoted $\frac{\partial^2 ln\mathcal{L}(p|\mathbf{x})}{\partial p^2}$ for a Binomial model are given by

$$\frac{\partial ln\mathcal{L}(p \mid \mathbf{x})}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\frac{\partial^2 ln\mathcal{L}(p \mid \mathbf{x})}{\partial p^2} = -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)^2}$$

Fisher information and Jeffrey's prior for a Binomial model (1/3)

So we want

$$\mathcal{I}(p) = -E\left[\frac{\partial^2 ln \mathcal{L}(p \mid \mathbf{x})}{\partial p^2}\right]$$

and we note that

$$E\left[\sum_{i=1}^{n} x_i\right] = E[n\overline{x}] = nE[\overline{x}] = np$$

Fisher information and Jeffrey's prior for a Binomial model (2/3)

$$
\begin{aligned}
\mathcal{I}(p) &= \frac{E\left[\sum_{i=1}^{n} x_i\right]}{p^2} + \frac{E\left[n - \sum_{i=1}^{n} x_i\right]}{(1-p)^2} \\
&= \frac{np}{p^2} + \frac{n - np}{(1-p)^2} \\
&= \frac{np(1-p)^2 + (n-np)p^2}{p^2(1-p)^2} \\
&= \frac{np - 2np^2 + np^3 + np^2 - np^3}{p^2(1-p)^2} \\
&= \frac{np(1-p)}{p^2(1-p)^2} \\
&= \frac{n}{p(1-p)} \propto \frac{1}{p(1-p)} \propto p^{-1}(1-p)^{-1}
\end{aligned}
$$

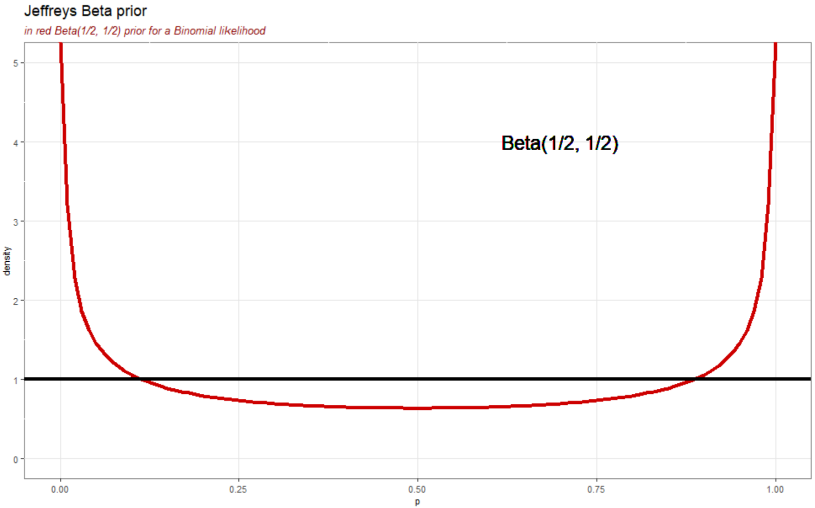
Fisher information and Jeffrey's prior for a Binomial model (3/3)

And we conclude that

$$p_J(p) = \sqrt{\mathcal{I}(p)} \propto p^{-1/2}(1-p)^{-1/2}$$

which is the Beta distribution $Beta(1/2, 1/2)$

Beta Jeffreys prior



Back to the working example (1/2)

Getting back to our spams in mailbox data

$$x_i = 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1$$

Assuming a Binomial likelihood for the data and using the Jeffreys prior, namely a $\mathcal{B}e(1/2, 1/2)$, what is the posterior mean and the 95% credible intreval for the model parameter ?

Back to the working example (2/2)

From the Bayes Theorem, we know that

$$p(p \mid x) = \frac{p(x \mid p)\ p(p)}{p(x)} \propto p(x \mid p)\ p(p)$$

In our case, we have that

$$\begin{aligned} p(p \mid x) &\propto \underbrace{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}_{\text{Binomial Likelihood}}\ \underbrace{p(p)}_{\text{Prior}} \\ &\propto p^{16}(1-p)^4\ p^{-1/2}(1-p)^{-1/2} \\ &\propto p^{15.5}\ (1-p)^{3.5} \end{aligned}$$

and we recognize the functional form of a Gamma density, that is $\mathcal{B}e(\alpha = 16.5, \beta = 4.5)$. The posterior mean is given by $\alpha/(\alpha+\beta) = 16.5/21 = 0.7857$.

Further reading and code

The R Project for Statistical Computing:
https://www.r-project.org/

Accessing the R code:
https://github.com/JRigh/Beta-Binomial-Example-in-R