Applied Bayesian Data Analysis — Chapter 6

Kim Albertsson

CERN and Luleå University of Technology

kim.albertsson@ltu.se

November 14, 2019

Chapter 6

Inferring a Binomial Probability via Exact Mathematical Analysis

Introduction

Bayes' rule:

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{\int p(\mathbf{y}|\theta)p(\theta)\, d\theta}$$

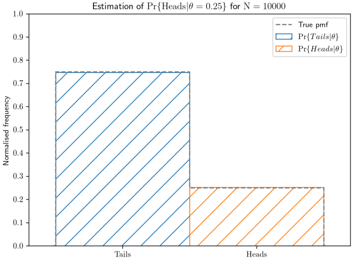
Maths made easier if $p(\mathbf{y}|\theta)p(\theta)$ has same form as $p(\theta|\mathbf{y})$. We will consider a reasonable choice for coin flips.

Bernoulli distribution for $p(\mathbf{y}|\theta)$. Beta distribution for $p(\theta)$. The Beta distribution is a *conjugate prior* to the Bernoulli distribution when used as a likelihood.

The Likelihood Function: Bernoulli Distribution

Analytical form for $p(\mathbf{y}|\theta)$: Bernoulli distribution

$$\begin{aligned}
p(\mathbf{y}|\theta) &= \theta^{\mathbf{y}}(1-\theta)^{1-\mathbf{y}} && (6.1)\\
&= p(y_0, y_1, \ldots|\theta)\\
&\text{assume independence}\\
&= p(y_0|\theta)p(y_1|\theta)p(\ldots|\theta)\\
&= \prod_{y_i \in \mathbf{y}} \theta^{y_i}(1-\theta)^{1-y_i}\\
&= \theta^{\sum y_i}(1-\theta)^{\sum(1-y_i)}\\
&= \theta^{z}(1-\theta)^{N-z} && (6.2)
\end{aligned}$$



A Description of Credibilities: The Beta Distribution

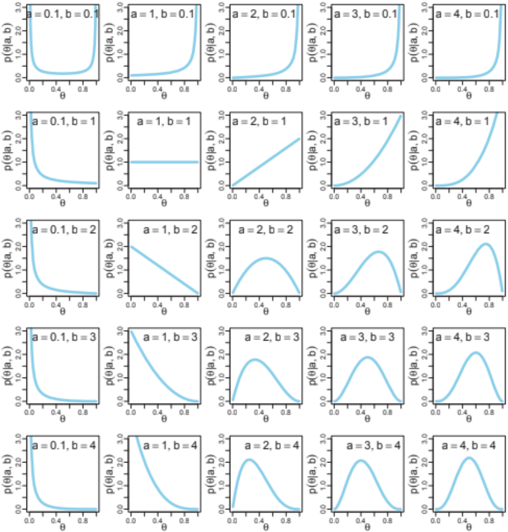
Analytical form for $p(\theta)$: Beta distribution

$$\begin{aligned} p(\theta|a,b) &= \text{beta}(\theta|a,b) \\ &= \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} \\ &= \frac{\theta^{a-1}(1-\theta)^{b-1}}{\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\,d\theta} \end{aligned} \tag{6.3}$$

$B(a,b)$ is the *Beta function*.

Note: Prior and posterior must integrate to 1. Likelihood needs not.

Beta distribution



Interlude: Why a Beta Prior? (I)

Consider the normalised likelihood:

$$\int_0^1 \frac{p(\mathbf{y}|\theta)}{f(\mathbf{y})} \, d\theta = 1$$

$$\int_0^1 \frac{\theta^z (1-\theta)^{N-z}}{f(\mathbf{y})} \, d\theta = 1 \qquad \text{(using 6.2)}$$

Consider the beta distribution:

$$\int_0^1 \frac{\theta^{a-1} (1-\theta)^{b-1}}{B(a, b)} \, d\theta = 1 \qquad (6.3)$$

Interlude: Why a Beta Prior? (II)

Hence:

$$\int_0^1 \frac{\theta^z(1-\theta)^{N-z}}{f(\mathbf{y})}\,d\theta = \int_0^1 \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}\,d\theta$$

Clearly $f = B; a = z+1; b = N-z+1$ and one can interpret $\mathrm{beta}(z+1, N-z+1)$ as representing prior knowledge about past coin flips.

Note: $a = 1; b = 1 \implies z = 0; N - z = 0$, thus $\mathrm{beta}(1,1)$ represents *no* prior observations. The book claims a different thing.

Specifying a Beta Prior (I)

Hence:

$$a = \mu\kappa \quad \text{and} \quad b = (1-\mu)\kappa \tag{6.5}$$

$$a = \omega(\kappa - 2) + 1 \quad \text{and} \quad b = (1-\omega)(\kappa - 2) + 1 \text{ for } \kappa > 2 \tag{6.6}$$

$$z = \mu\kappa - 1 \quad \text{and} \quad N - z = (1-\mu)\kappa - 1 \text{ for } \kappa > 2$$

$$z = \omega(\kappa - 2) \quad \text{and} \quad N - z = (1-\omega)(\kappa - 2) \text{ for } \kappa > 2$$

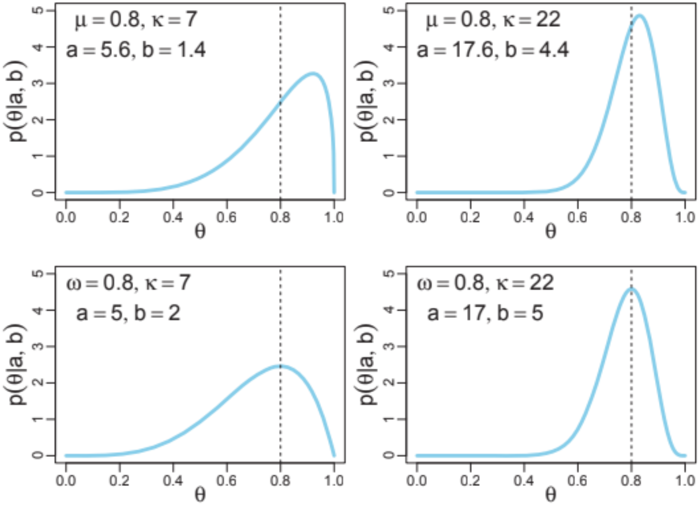
Specifying a Beta Prior (II)

Can we interpret these numbers in terms of physical quantities?

$$
\begin{aligned}
\mu &= \frac{z+1}{N+2} && \text{for } N > 0;\ z > 0 \\
\omega &= \frac{z}{N} && \text{for } N > 1;\ z > 0 \\
\kappa &= N + 2 && \text{for } N > 1;\ z > 0
\end{aligned}
$$

where μ is the mean, ω is the mode, and κ is the "concentration".

μ and ω acually have units of *volume concentration*. κ is probably better interpreted as an inverse spread?


μ = 0.8, κ = 7
a = 5.6, b = 1.4
μ = 0.8, κ = 22
a = 17.6, b = 4.4
ω = 0.8, κ = 7
a = 5, b = 2
ω = 0.8, κ = 22
a = 17, b = 5
p(θ|a, b)
θ
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5

The Posterior Beta

Putting the previous work together gives us the posterior:

$$\begin{aligned}
p(\theta|z,N) &= p(z,N)p(\theta)\frac{1}{p(z,N)} \\
&= \theta^z(1-\theta)^{N-z}\frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}\frac{1}{p(z,N)} \\
&= \frac{\theta^{z+a-1}(1-\theta)^{N-z+b-1}}{B(a,b)p(z,N)} \\
&= \frac{\theta^{z+a-1}(1-\theta)^{N-z+b-1}}{B(z+a,N-z+b)}
\end{aligned}$$

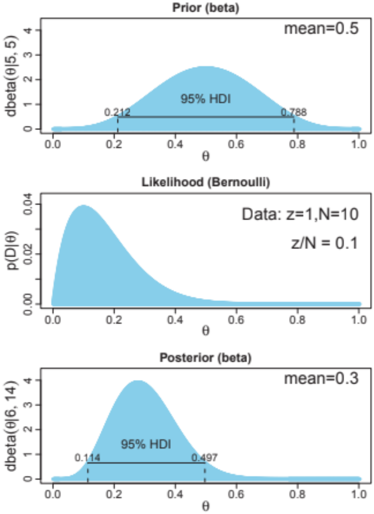
Posterior is compromise of prior and likelihood (I)

Mean of the posterior can be factored:

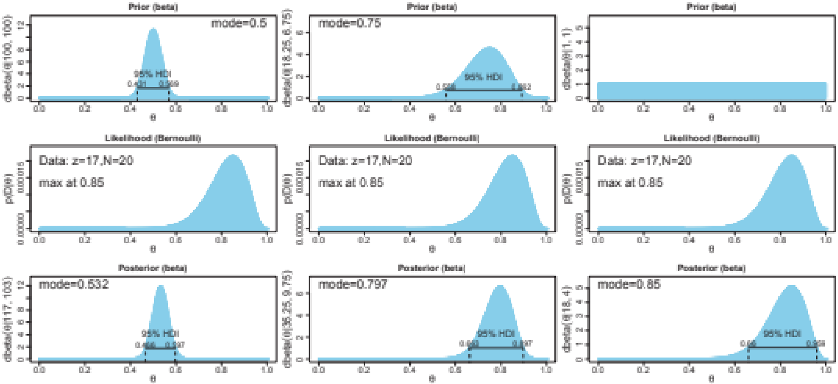
$$\begin{aligned} \mu_{\text{posterior}} &= \mu_{\text{likelihood}} w_{\text{likelihood}} \mu_{\text{prior}} w_{\text{prior}} \\ \frac{z+a}{N+a+b} &= \frac{z}{N} \frac{N}{N+a+b} \frac{a}{a+b} \frac{a+b}{N+a+b} \end{aligned} \tag{6.9}$$

Meaning that the mean of the posterior will be weighted average of the two constituent means. Remember $N_{\text{prior}} = a + b - 2$ (thus this argument is a bit hand-wavy?).

Posterior is compromise of prior and likelihood (II)



Prior knowledge expressed as a beta distribution



Prior knowledge that cannot be expressed as a beta distribution

Suppose categorical mixture of two betas (e.g. coins from two makers).

Posterior no longer beta (but should converge to one given large enough sample?).