Bayesian Inference in the Bernoulli Model

Hedibert Freitas Lopes

January 2014

Petrobras

Date	Open	High	Low	Close	Volume	Adj Close	Up(1)/Down(0)
Jan 24, 2014	11.97	11.99	11.60	11.76	24,925,900	11.76	0
Jan 23, 2014	12.51	12.52	12.05	12.16	18,627,600	12.16	0
Jan 22, 2014	12.31	12.55	12.30	12.50	12,350,400	12.50	1
Jan 21, 2014	12.35	12.39	12.10	12.24	22,789,800	12.24	0
Jan 17, 2014	12.53	12.65	12.42	12.48	15,576,700	12.48	0
Jan 16, 2014	12.73	12.80	12.45	12.54	20,579,500	12.54	0
Jan 15, 2014	12.55	12.94	12.51	12.71	34,442,900	12.71	1
Jan 14, 2014	12.45	12.55	12.41	12.42	18,328,000	12.42	0
Jan 13, 2014	12.81	12.86	12.40	12.46	21,429,200	12.46	0
Jan 10, 2014	12.73	12.91	12.65	12.84	21,865,300	12.84	1
Jan 9, 2014	12.65	12.73	12.33	12.52	21,117,400	12.52	0
Jan 8, 2014	12.93	12.96	12.66	12.68	17,711,000	12.68	0
Jan 7, 2014	13.38	13.38	12.82	12.90	19,305,300	12.90	0
Jan 6, 2014	12.97	13.20	12.89	13.16	12,475,200	13.16	1
Jan 3, 2014	13.26	13.37	13.00	13.12	19,993,700	13.12	0
Jan 2, 2014	13.45	13.50	13.20	13.32	18,218,000	13.32	0
Dec 31, 2013	13.66	13.85	13.58	13.78	6,936,800	13.78	-

Source: `http://finance.yahoo.com/q/hp?s=PBR+Historical+Prices`

Sequence of $n = 16$ days: 0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0

Bernoulli trials

Let $x_1, \ldots, x_n$ be iid $Ber(\theta)$, ie. conditionally on θ the binary variables $x_1, \ldots, x_n$ are all independent from one another. The probability function of any x_i (given θ) is

$$p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

$x_i \in \{0,1\}$, $\theta \in [0,1]$ and $i = 1, \ldots, n$. Usually, $x_i = 1$ is labeled a "success" and $x_i = 0$ a "failure".

Let $s_n = \sum_{i=1}^{n} x_i$ be the number of successes out of n trials. Therefore, the joint density of $x_1, \ldots, x_n$ is given by

$$p(x_1, \ldots, x_n|\theta) = \theta^{s_n}(1-\theta)^{n-s_n}$$

for all $(x_1, \ldots, x_n) \in \{0,1\}^n$.

Number of successes

It can be shown that $s_n|\theta \sim Binomial(n, \theta)$, ie.

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

for $s_n \in \{0, 1, \ldots, n\}$.

We observed $s_n = 4$ for $n = 16$.

Likelihood function and MLE

The maximum likelihood estimate of θ, namely $\hat{\theta}_{mle}$, maximizes the log-likelihood function

$$l(\theta; s_n) = \log p(s_n|\theta) = c + s_n \log\theta + (n - s_n)\log(1-\theta),$$

where $c = \log n! - \log s_n! - \log(n - s_n)!$ is constant for all values of $\theta \in [0, 1]$.

Then,

$$\hat{\theta}_{mle} = \frac{s_n}{n}.$$

In our Petrobras example,

$$\hat{\theta}_{mle} = \frac{4}{16} = 0.25 \quad \text{or} \quad 25\%.$$

Frequentist argument

In the limit (when n is large enough),

$$(\hat{\theta}_{mle}|\theta) \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right),$$

and (after some more assumptions) an approximate 95% confidence interval for θ is

$$\hat{\theta}_{mle} \pm 2\sqrt{\frac{(\hat{\theta}_{mle})(1-\hat{\theta}_{mle})}{n}},$$

or

$$(19.59\%; 30.41\%)$$

in our petrobras example.

More assumptions would be need to approximate

$$p(x_{17}|x_1, \ldots, x_{16}).$$

Bayesian approach

Recall that, for $s_n \in \{0, 1, \dots, n\}$,

$$p(s_n|\theta) = \frac{n!}{s_n!(n - s_n)!} \theta^{s_n} (1 - \theta)^{n - s_n}.$$

Let us assume the following prior distribution for θ:

$$\theta \sim Beta(a, a) \qquad \text{for a} \geq 1.$$

If $x \sim Beta(a, b)$, then

$$p(x|a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1 - x)^{b-1},$$

with $E(x) = a/(a + b)$ and $V(x) = ab/(a + b)^2(a + b + 1)$. Also, $\text{mode}(x) = (a - 1)/(a + b - 2)$, if $a, b > 1$.

Likelihood × prior

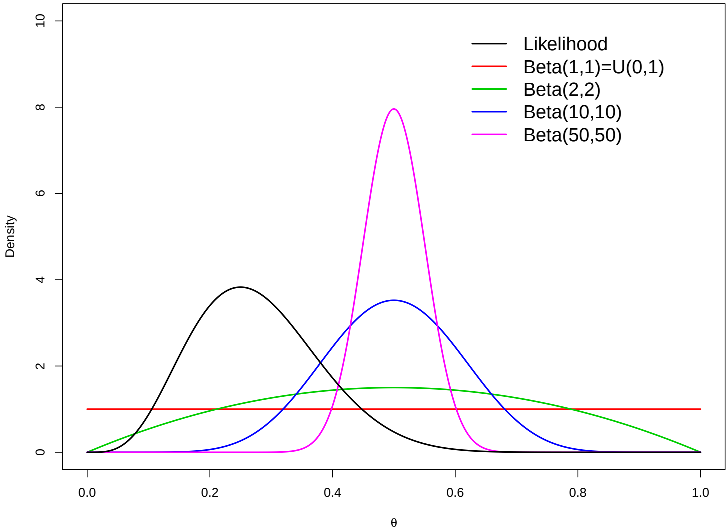
The likelihood function

$$
\begin{aligned}
L(\theta|s_n) &\propto \theta^{s_n}(1-\theta)^{n-s_n} \\
&\propto \theta^{(s_n+1)-1}(1-\theta)^{(n-s_n+1)-1},
\end{aligned}
$$

resembles a $Beta(s_n+1, n-s_n+1)$.

In our Petrobras example, it resembles a $Beta(5, 13)$ with mean 29.41% and standard deviation 10.74%. Obviously, the mode is 25%.

Likelihood and priors $Beta(a, a)$



Posterior

The posterior is also a Beta distribution, since

$$p(\theta|s_n) \propto \theta^{s_n}(1-\theta)^{n-s_n}\theta^{a-1}(1-\theta)^{a-1},$$

leads to

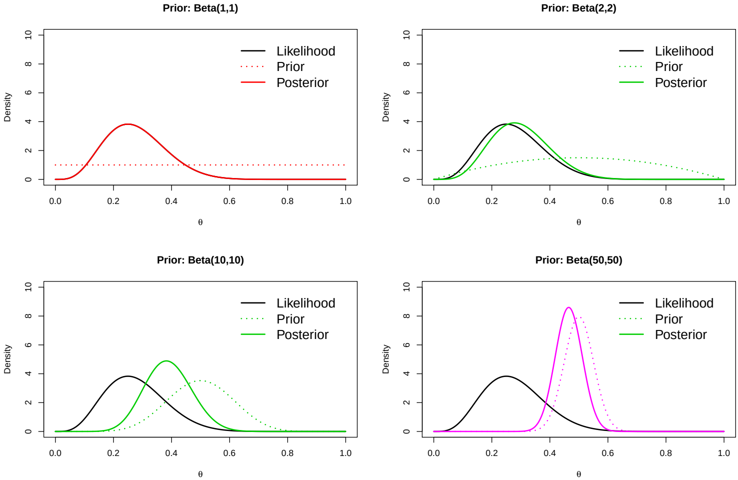
$$(\theta|s_n) \sim Beta(a+s_n, a+n-s_n).$$

For example,

$$E(\theta|s_n) = \frac{a+s_n}{2a+n}.$$

The unbiased (also ML) estimator $\hat{\theta} = s_n/n$ is recovered when $a = 0$ or when a is much smaller than s_n and n. In these cases, $V(\theta|s_n)$ approximates $\hat{\theta}(1-\hat{\theta})/n$.

Posterior distributions



Posterior predictive

$$
\begin{aligned}
p(x_{n+1}|s_n) &= \int_0^1 p(x_{n+1},\theta|s_n)d\theta = \int_0^1 p(x_{n+1}|\theta)p(\theta|s_n)d\theta \\
&= \int_0^1 \theta^{x_{n+1}}(1-\theta)^{1-x_{n+1}} \frac{\Gamma(2a+n)}{\Gamma(a+s_n)\Gamma(a+n-s_n)}\theta^{a+s_n-1}(1-\theta)^{a+n-s_n-1}d\theta \\
&= \frac{\Gamma(2a+n)}{\Gamma(a+s_n)\Gamma(a+n-s_n)}\int_0^1 \theta^{(a+s_n+x_{n+1})-1}(1-\theta)^{(a+n-s_n+1-x_{n+1})-1}d\theta \\
&= \frac{\Gamma(2a+n)}{\Gamma(a+s_n)\Gamma(a+n-s_n)} \frac{\Gamma(a+s_n+x_{n+1})\Gamma(a+n-s_n+1-x_{n+1})}{\Gamma(2a+n+1)} \\
&= \frac{\Gamma(2a+n)}{\Gamma(2a+n+1)} \frac{\Gamma(a+s_n+x_{n+1})}{\Gamma(a+s_n)} \frac{\Gamma(a+n-s_n+1-x_{n+1})}{\Gamma(a+n-s_n)} \\
&= \left\{ \begin{array}{ll} (a+n-s_n)/(2a+n) & x_{n+1}=0 \\ (a+s_n)/(2a+n) & x_{n+1}=1 \end{array} \right. .
\end{aligned}
$$

In our Petrobras example,

$$
Pr(x_{17}=1|s_{16}=4) = \frac{a+4}{2a+16}.
$$

3379 days of Petrobras

$$
\begin{aligned}
(x_1, \ldots, x_{16}) &= (0011101110110011) & s = 10 \\
(x_{17}, \ldots, x_{32}) &= (1010000011001100) & s = 6 \\
&\vdots \\
(x_{3348}, \ldots, x_{3363}) &= (0100110100110101) & s = 8 \\
(x_{3364}, \ldots, x_{3379}) &= (0010001001000100) & s = 4
\end{aligned}
$$

Converging posteriors: data dominates the prior

As observations accumulate over time the posterior distribution of θ, ie. $Beta(a + s_n, b + n - s_n)$ is updated via Bayes' Theorem. Below are the percentiles 2.5%, 50% and 97.5% of $p(\theta|s_n)$, $n \geq 1$.