Why Bayes

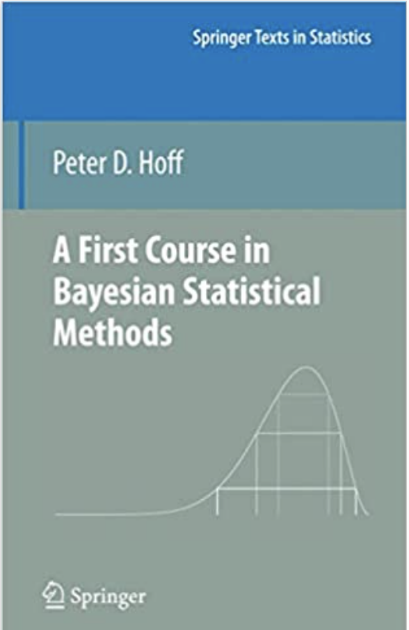
April 22, 2021

Why Bayes?

Why Bayes?

Estimating the probability of a rare event

References


Springer Texts in Statistics
Peter D. Hoff
A First Course in Bayesian Statistical Methods
Springer

Description of problem

- Want to estimate the prevalence of an infectious disease in a small town.
- The higher the prevalence, the more public health precautions will be recommended.
- A small random sample of 20 individuals are checked for infection.

Description of problem

- **Parameter** θ, the fraction of infected individuals in the city.
- **Parameter space**: $\Theta = [0, 1]$
- **Sample**: Y the number of infected individuals in the sample
- **Sample space**: $\mathcal{Y} = \{0, 1, ..., 20\}$

Sampling model

If the value of θ were known, a reasonable sampling model for Y would be

$$Y \mid \theta \sim \text{Binomial}(20, \theta)$$

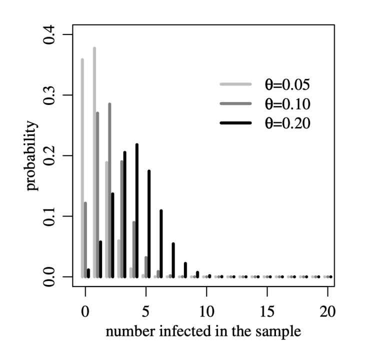


Figure 1: Binomial$(20, \theta)$ distributons for three values of θ.

Prior distribution

Other studies from various parts of the country indicate that the infection rate in comparable cities range from about 0.05 to 0.20, with an average prevalence of 0.10.

Prior distribution

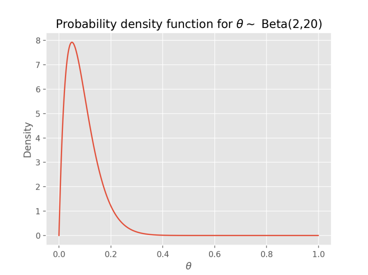
Other studies from various parts of the country indicate that the infection rate in comparable cities range from about 0.05 to 0.20, with an average prevalence of 0.10.

This suggests we use a prior distribution $p(\theta)$ that assigns a substantial amount of probability to the interval $(0.05, 0.20)$.

Prior distribution

Other studies from various parts of the country indicate that the infection rate in comparable cities range from about 0.05 to 0.20, with an average prevalence of 0.10.

This suggests we use a prior distribution $p(\theta)$ that assigns a substantial amount of probability to the interval $(0.05, 0.20)$.



We can encode this prior information using

$$\theta \sim \text{Beta}(2, 20)$$

From prior to posterior

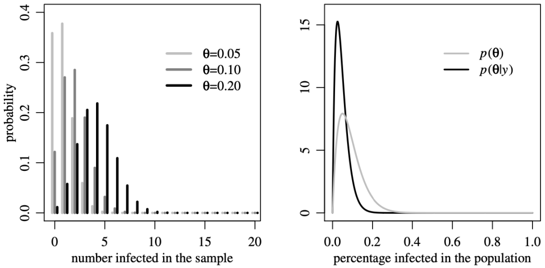


Fig. 1.1. Sampling model, prior and posterior distributions for the infection rate example. The plot on the left-hand side gives binomial$(20, \theta)$ distributions for three values of θ. The right-hand side gives prior (gray) and posterior (black) densities of θ.

Prior

$$\theta \sim \text{Beta}(2, 20)$$
$$\mathbb{E}[\theta] = 0.09$$
$$\text{mode}[\theta] = 0.05$$
$$P(\theta < 0.10) = 0.64$$
$$P(0.05 < \theta < 0.20) = 0.66$$

Posterior

$$\theta \mid \{Y = 0\} \sim \text{Beta}(4, 20)$$
$$\mathbb{E}[\theta \mid \{Y = 0\}] = 0.048$$
$$\text{mode}[\theta \mid \{Y = 0\}] = 0.025$$
$$P(\theta < 0.10 \mid \{Y = 0\}) = 0.93$$

Sensitivity analysis

Suppose we consider beliefs represented by $\text{Beta}(a, b)$ distributions for (a, b) other than $(2, 20)$.

If $\theta \sim \text{Beta}(a, b)$, then $\theta \mid Y = y \sim \text{Beta}(a + y, b + n - y)$.

The posterior expectation is

$$\begin{aligned}
\mathbb{E}[\theta \mid Y = y] &= \frac{a + y}{a + b + n} \\
&= \frac{n}{a + b + n}\frac{y}{n} + \frac{a + b}{a + b + n}\frac{a}{a + b} \\
&= \frac{n}{w + n}\bar{y} + \frac{w}{w + n}\theta_0
\end{aligned}$$

where $\theta_0 = a/(a + b)$ is the prior expectation of θ and $w = a + b$.

Sensitivity analysis

Suppose we consider beliefs represented by Beta(a, b) distributions for (a, b) other than $(2, 20)$.

If $\theta \sim \text{Beta}(a, b)$, then $\theta \mid Y = y \sim \text{Beta}(a + y, b + n - y)$.

The posterior expectation is

$$\begin{aligned}
\mathbb{E}[\theta \mid Y = y] &= \frac{a + y}{a + b + n} \\
&= \frac{n}{a + b + n}\frac{y}{n} + \frac{a + b}{a + b + n}\frac{a}{a + b} \\
&= \frac{n}{w + n}\bar{y} + \frac{w}{w + n}\theta_0
\end{aligned}$$

where $\theta_0 = a/(a + b)$ is the prior expectation of θ and $w = a + b$.

So the posterior expectation is a compromise between the prior expectation θ_0 and sample mean $\bar{y}$. The weights on each depend on the sample size, n, and our prior confidence in this guess, w.

Sensitivity analysis

If someone provides us with a prior guess θ_0 and degree of confidence w, then we can approximate their prior beliefs about θ with

$$Beta\left(a = w\theta_0, \quad b = w(1-\theta_0)\right)$$

And their posterior beliefs are represented with

$$Beta\left(a = w\theta_0 + y, \quad b = w(1-\theta_0) + n - y\right)$$

Sensitivity analysis

If someone provides us with a prior guess θ_0 and degree of confidence w, then we can approximate their prior beliefs about θ with

$$Beta\left(a = w\theta_0, \quad b = w(1-\theta_0)\right)$$

And their posterior beliefs are represented with

$$Beta\left(a = w\theta_0 + y, \quad b = w(1-\theta_0) + n - y\right)$$

We can compute such a posterior distribution for a wide range of θ_0 and w values to perform a *sensitivity analysis*, an exploration of how posterior information is affected by differences in prior opinion.

Sensitivity analysis

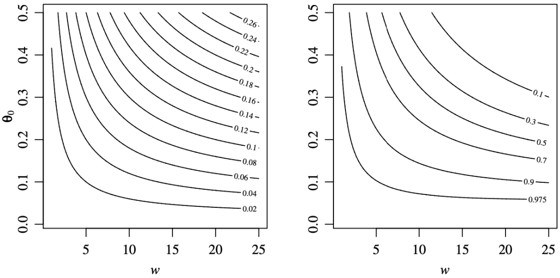


Fig. 1.2. Posterior quantities under different beta prior distributions. The left- and right-hand panels give contours of $E[\theta|Y=0]$ and $\Pr(\theta < 0.10|Y=0)$, respectively, for a range of prior expectations and levels of confidence.

The second plot may be of use if, e.g., city officials would like to recommend a vaccine to the general public unless they were reasonably sure the current infection rate was less than 0.10.

Sensitivity analysis

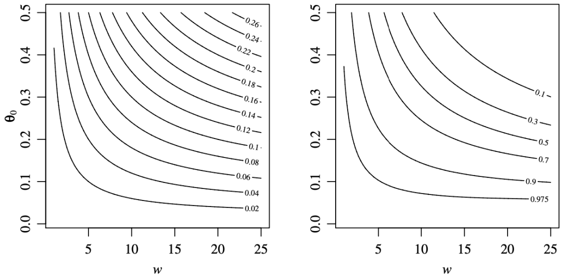


Fig. 1.2. Posterior quantities under different beta prior distributions. The left- and right-hand panels give contours of $\mathrm{E}[\theta|Y=0]$ and $\Pr(\theta < 0.10|Y=0)$, respectively, for a range of prior expectations and levels of confidence.

The second plot may be of use if, e.g., city officials would like to recommend a vaccine to the general public unless they were reasonably sure the current infection rate was less than 0.10.

A high degree of certainty (say 97.5%) is only achieved by people who already thought the infection rate was lower than the average of other cities.

Comparison to non-Bayesian methods

A 95% confidence interval for population proportion θ is the *Wald interval*, given by

$$\bar{y} \pm 1.96\sqrt{\bar{y}(1-\bar{y})/n}$$

Comparison to non-Bayesian methods

A 95% confidence interval for population proportion θ is the *Wald interval*, given by

$$\bar{y} \pm 1.96\sqrt{\bar{y}(1-\bar{y})/n}$$

The interval has *correct asymptotic frequentist coverage*, meaning that if n is large, then with probability approximately equal to 95%, Y will take on a value y such that the above interval contains θ.

Comparison to non-Bayesian methods

A 95% confidence interval for population proportion θ is the *Wald interval*, given by

$$\bar{y} \pm 1.96\sqrt{\bar{y}(1-\bar{y})/n}$$

The interval has *correct asymptotic frequentist coverage*, meaning that if n is large, then with probability approximately equal to 95%, Y will take on a value y such that the above interval contains θ.

For our sample in which $\bar{y} = 0$, the Wald confidence interval comes out to be just a single point: 0.

Comparison to non-Bayesian methods

A 95% confidence interval for population proportion θ is the *Wald interval*, given by

$$\bar{y} \pm 1.96\sqrt{\bar{y}(1-\bar{y})/n}$$

The interval has *correct asymptotic frequentist coverage*, meaning that if n is large, then with probability approximately equal to 95%, Y will take on a value y such that the above interval contains θ.

For our sample in which $\bar{y} = 0$, the Wald confidence interval comes out to be just a single point: 0.

In fact, the 99.99% Wald interval also comes out to be zero.

Comparison to non-Bayesian methods

People have suggested alternatives to avoid this type of behavior.

Comparison to non-Bayesian methods

People have suggested alternatives to avoid this type of behavior.

The "adjusted" Wald interval suggested by Agresti and Coull (1998) is given by

$$\begin{aligned} &\widehat{\theta} \pm \sqrt{\widehat{\theta}(1-\widehat{\theta})/n}, \quad \text{where} \\ &\widehat{\theta} = \frac{n}{n+4}\bar{y} + \frac{4}{n+4}\frac{1}{2} \end{aligned}$$

Comparison to non-Bayesian methods

People have suggested alternatives to avoid this type of behavior.

The "adjusted" Wald interval suggested by Agresti and Coull (1998) is given by

$$\widehat{\theta} \pm \sqrt{\widehat{\theta}(1-\widehat{\theta})/n}, \quad \text{where}$$
$$\widehat{\theta} = \frac{n}{n+4}\bar{y} + \frac{4}{n+4}\frac{1}{2}$$

While not motivated as such, the interval is clearly related to Bayesian inference: $\widehat{\theta}$ is equivalent to the posterior mean for θ under a Beta$(2,2)$ prior, which represents weak prior information centered around $\theta = 1/2$.

Comparison to non-Bayesian methods

Compared to the post-hoc "adjustment" approach, the Bayesian formalism provides

- Reasonable conclusions which fall naturally out of the framework
- Flexibility to other choice of priors than $\text{Beta}(2, 2)$
- Sensitivity analysis to consider the sets of conclusions that would be reached by people with different priors.
- Simultaneous access to various functionals of the posterior – not just $\mathbb{E}[\theta \mid Y = y]$ but also $\mathbb{P}[\theta < 0.10 \mid Y = 0]$.

Extensions: Hierarchical models