Introduction to Bayesian Statistics

Finding the posterior distribution

Radu T. Trîmbiţaş

May 19, 2016

1 Introduction

Introduction

- Introductory example: Suppose that we are interested in estimating the proportion of responders to a new therapy for treating a disease that is serious and difficult to cure (such a disease is said to be **virulent**).
  - p – the probability that any single person with the disease responds to the treatment
  - Y – the number of responders in a sample of size n might reasonably be assumed to have a binomial distribution with parameter p.
- In the classical (*frequentist*) approach p has a fixed but unknown value and have discussed point estimators, interval estimators, and tests of hypotheses for this parameter
- *Bayesian approach*: suppose we know that $p \approx 0.25$
- we model the *conditional distribution* of Y given p, $Y|p$, as binomial

$$p\left(Y|p\right) = \binom{n}{y} p^y q^{n-y}, \qquad y = 0, 1, 2, \ldots, n.$$

- Uncertainty about the parameter p is handled by treating it as a random variable and, before observing any data, assigning a *prior distribution* to p.
- Because we know that $0 < p < 1$ and the beta density function has the interval $(0, 1)$ as support, it is convenient to use a beta distribution as a prior for p.
- If $X \sim beta(\alpha, \beta)$, then $\mu = E(X) = \frac{\alpha}{\alpha+\beta}$, a good candidate for p distribution is a beta with $\alpha = 1$ and $\beta = 3$.

- Thus, the density assigned to p is

$$g(p) = \frac{1}{3}(1-p)^2, \qquad 0 < p < 1.$$

- Since we have specified the conditional distribution of $Y|p$ and the distribution of p, we have also specified the joint distribution of (Y, p) and can determine the marginal distribution of Y and the conditional distribution of $p|Y$.
- *After* observing $Y = y$, the *posterior density* of p given $Y = y$, $g^*(p|y)$, can be determined.
- In the next section, we derive a general result that, in our virulent-disease example, implies that the posterior density of p given $Y = y$ is

$$g^*(p|y) = \frac{\Gamma(n+4)}{\Gamma(y+1)\Gamma(n-y+3)} p^y (1-p)^{n-y+2},\ 0 < p < 1,$$

that is a $beta(\alpha, \beta)$ with $\alpha = y + 1$ and $\beta = n - y + 3$.

- This posterior density is the "updated" (by the data) density of p and is the basis for all Bayesian inferences regarding p.
- In the following sections, we describe the general Bayesian approach and specify how to use the posterior density to obtain
  - estimates
  - credible intervals
  - hypothesis tests for p and for parameters associated with other distributions.

2 Bayesian Priors and Posteriors

Bayesian Priors and Posteriors

- $Y_1, Y_2, \ldots, Y_n$ RVs associated with a sample of size n, $L(y_1, y_2, \ldots, y_n|\theta)$ the likelihood of the sample.
- In the Bayesian approach, the unknown parameter θ is viewed to be a random variable with a probability distribution, called the *prior distribution* of θ.
- This prior distribution is specified before any data are collected and provides a theoretical description of information about θ that was available before any data were obtained.
- We will assume that the parameter θ has a continuous distribution with density $g(\theta)$

- Using the likelihood of the data and the prior on θ, it follows that the joint likelihood of $Y_1, Y_2, \ldots, Y_n, \theta$ is

$$f(y_1, y_2, \ldots, y_n, \theta) = L(y_1, y_2, \ldots, y_n|\theta)g(\theta)$$

and that the marginal density or mass function of $Y_1, Y_2, \ldots, Y_n$ is

$$m\,(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta)g(\theta)\mathrm{d}\,\theta.$$

- The posterior density of $\theta|y_1, y_2, \ldots, y_n$:

$$g^*\,(\theta|y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n|\theta)g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta)g(\theta)\mathrm{d}\,\theta}. \tag{1}$$

- The posterior density summarizes all of the pertinent information about the parameter θ by making use of the information contained in the prior for θ and the information in the data.

3 Examples

Examples

Example 1. Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Bernoulli distribution where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$ and assume that the prior distribution for p is $beta(\alpha, \beta)$. Find the posterior distribution for p.

Solution

Since the Bernoulli pmf could be written as

$$p(y_i|p) = p^{y_i}(1-p)^{1-y_i},\ y_i = 0, 1,$$

the likelihood is $L(y_1, y_2, \ldots, y_n|p)$ is

$$\begin{aligned} L(y_1, y_2, \ldots, y_n|p) &= p(y_1, y_2, \ldots, y_n|p) \\ &= p^{y_1}(1-p)^{1-y_1}p^{y_2}(1-p)^{1-y_2}\cdots p^{y_n}(1-p)^{1-y_n} \\ &= p^{\sum y_n}(1-p)^{n-\sum y_i}, \qquad y_i = 0, 1, p \in (0, 1). \end{aligned}$$

Thus,

$$\begin{aligned} f(y_1, y_2, \ldots, y_n, p) &= L(y_1, y_2, \ldots, y_n|p)g(p) \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i}\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \\ &= \frac{p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}}{B(\alpha, \beta)} \end{aligned}$$

and

$$m(y_1, y_2, \ldots, y_n) = \int_0^1 \frac{p^{\sum y_i + \alpha - 1}(1-p)^{n - \sum y_i + \beta - 1}}{B(\alpha, \beta)} \mathrm{d}\, p$$
$$= \frac{B\left(\sum y_i + \alpha, n - \sum y_i + \beta\right)}{B(\alpha, \beta)}$$

Finally, the posterior density of p is obtained from 1)

$$g^*(p|y_1, y_2, \ldots, y_n) = \frac{\frac{p^{\sum y_i + \alpha - 1}(1-p)^{n - \sum y_i + \beta - 1}}{B(\alpha,\beta)}}{\frac{B(\sum y_i + \alpha, n - \sum y_i + \beta)}{B(\alpha,\beta)}}$$
$$= \frac{p^{\sum y_i + \alpha - 1}(1-p)^{n - \sum y_i + \beta - 1}}{B\left(\sum y_i + \alpha, n - \sum y_i + \beta\right)},$$

i.e. a beta distribution with parameters $\alpha^* = \sum y_i + \alpha$ şi $\beta^* = n - \sum y_i + \beta$.

Examples

Example 2. Consider the virulent-disease scenario and the results of Example 1. Compare the prior and posterior distributions of the Bernoulli parameter p (the proportion of responders to the new therapy) if we chose the values for α and β and observed the hypothetical data given below:

(a) $\alpha = 1$, $\beta = 3$, $n = 5$, $\sum y_i = 2$.

(b) $\alpha = 1$, $\beta = 3$, $n = 25$, $\sum y_i = 10$.

(c) $\alpha = 10$, $\beta = 30$, $n = 5$, $\sum y_i = 2$.

(d) $\alpha = 10$, $\beta = 30$, $n = 25$, $\sum y_i = 10$.

Solution

Notice that both beta priors have mean

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{1}{1+3} = \frac{10}{10+30} = 0.25$$

and that both hypothetical samples result in the same value of the MLEs for p:

$$\widehat{p} = \frac{1}{n} \sum y_i = \frac{2}{5} = \frac{10}{25} = 0.4.$$

As derived in Example 1, if $y_1, y_2, \ldots, y_n$ denote the values in a random sample from a Bernoulli distribution, where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$, and the prior distribution for p is $beta(\alpha, \beta)$, the posterior distribution for p is a beta with $\alpha^* = \sum y_i + \alpha$, $\beta^* = n - \sum y_i + \beta$. Therefore, for the choices in this example,

Distribuţia	n	$\sum y_i$	Parametrii distribuţiei beta		Media	Dispersia
a priori	–	–	$\alpha = 1,$	$\beta = 3$	.2500	.0375
a posteriori	5	2	$\alpha^* = 3,$	$\beta^* = 6$	.3333	.0222
a posteriori	25	10	$\alpha^* = 11,$	$\beta^* = 16$	.4074	.0078
a priori	–	–	$\alpha = 10,$	$\beta = 30$	.2500	.0046
a posteriori	5	2	$\alpha^* = 12,$	$\beta^* = 33$	.2667	.0043
a posteriori	25	10	$\alpha^* = 20,$	$\beta^* = 45$	.3077	.0032

Table 1: Beta a priori şi a posteriori pentru exemplul 2

(a) when the prior is $\beta(1,3)$, $n = 5$, $\sum y_i = 2$, the posterior is beta with $\alpha^* = \sum y_i + \alpha = 2 + 1 = 3$ and $\beta^* = n - \sum y_i + \beta = 5 - 2 + 3 = 6$.

(b) when the prior is $\beta(1,3)$, $n = 25$, $\sum y_i = 10$, the posterior is beta with $\alpha^* = 10 + 1 = 11$ and $\beta^* = 25 - 10 + 3 = 18$.

(c) when the prior is $\beta(10,30)$, $n = 5$, $\sum y_i = 2$, the posterior is beta with $\alpha^* = 2 + 10 = 12$ and $\beta^* = 5 - 2 + 30 = 33$.

(d) when the prior is $\beta(10,30)$, $n = 25$, $\sum y_i = 10$, the posterior is beta with $\alpha^* = 20$ and $\beta^* = 45$.

Recall that the mean and variance of a beta(α,β) distributed random variable are

$$\mu = \frac{\alpha}{\alpha+\beta},\ \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}.$$

The parameters of the previous beta priors and posteriors, along with their means and variances are summarized in Table 1. Figure 1 contains graphs of the beta distributions (priors and posteriors) associated with the beta prior with parameters $\alpha = 1$, $\beta = 3$. Graphs of the beta distributions associated with the *beta*$(10,30)$ prior are given in Figure 2.

In Examples 1 and 2, we obtained posterior densities that, like the prior, are beta densities but with altered (by the data) parameter values.

4 Conjugate Priors and Estimators

Conjugate Priors and Estimators

Definition 3. Prior distributions that result in posterior distributions that are of the same functional form as the prior but with altered parameter values are called *conjugate priors*.

- Any beta distribution is a conjugate prior distribution for a Bernoulli (or a binomial) distribution.

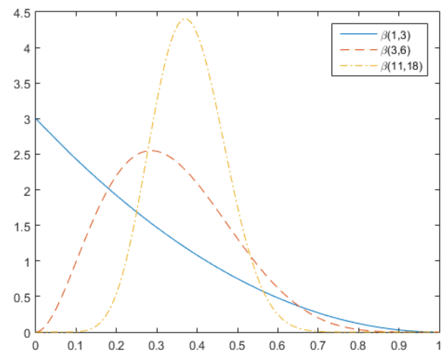

Figure 1: Graficele distribuţiilor beta a priori şi a posteriori din exemplul 1 (a)

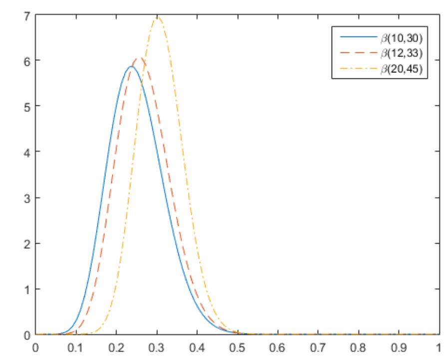

Figure 2: Graficele distribuţiilor beta a priori şi a posteriori din exemplul 2

- For the distributions that we use in this lecture, there are conjugate priors associated with the relevant parameters. These families of conjugate priors are often viewed to be broad enough to handle most practical situations. As a result, conjugate priors are often used in practice.
- Since the posterior is a bona fide probability density function, some summary characteristic of this density provides an estimate for θ.

Definition 4. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with likelihood function $L(y_1, y_2, \ldots, y_n|\theta)$ and let θ have prior density $g(\theta)$. The *posterior Bayes estimator* for $t(\theta)$ is given by

$$\widehat{t(\theta)}_B = E(t(\theta)|Y_1, Y_2, \ldots, Y_n).$$

Example

Example 5. In Example 1, we found the posterior distribution of the Bernoulli parameter p based on a beta prior with parameters (α, β). Find the Bayes estimators for p and $p(1-p)$. [Recall that $p(1-p)$ is the variance of a Bernoulli random variable with parameter p].

Solutions

In Example 1, we found the posterior density of p to be a beta density with parameters $\alpha^* = \sum y_i + \alpha$ and $\beta^* = n - \sum yi + \beta$:

$$g^*(p|y_1, y_2, \ldots, yn) = \frac{1}{B(\alpha^*, \beta^*)} p^{\alpha^*-1} (1-p)^{\beta^*-1},\ 0 < p < 1.$$

The estimate for p is the posterior mean of p. From our previous study of the beta distribution, we know that

$$\begin{aligned}
\widehat{p}_B = E(p|y_1, y_2, \ldots, y_n) &= \frac{\alpha^*}{\alpha^* + \beta^*} \\
&= \frac{\sum y_i + \alpha}{\sum y_i + \alpha + n - \sum yi + \beta} = \frac{\sum y_i + \alpha}{n + \alpha + \beta}
\end{aligned}$$

Similarly,

$$\begin{aligned}
[\widehat{p(1-p)}]_B &= E(p(1-p)|y_1, y_2, \ldots, y_n) \\
&= \int_0^1 p(1-p) \frac{1}{B(\alpha^*, \beta^*)} p^{\alpha^*-1} (1-p)^{\beta^*-1} \,\mathrm{d}\,p \\
&= \frac{B(\alpha^*+1, \beta^*+1)}{B(\alpha^*, \beta^*)} \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \cdot \frac{\Gamma(\alpha^*+1)\Gamma(\beta^*+1)}{\Gamma(\alpha^*+\beta^*+2)} \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \cdot \frac{\alpha^*\beta^*\Gamma(\alpha^*)\Gamma(\beta^*)}{(\alpha^*+\beta^*+1)(\alpha^*+\beta^*)\Gamma(\alpha^*+\beta^*)}
\end{aligned}$$

$$[\widehat{p(1-p)}]_B = \frac{\alpha^* \beta^*}{(\alpha^* + \beta^*)\,(\alpha^* + \beta^*)}$$
$$= \frac{(\sum y_i + \alpha)\,(n - \sum y_i + \beta)}{(n + \alpha + \beta)\,(n + \alpha + \beta + 1)}$$

So, the Bayes estimators for p and $p(1-p)$ are

$$\widehat{p}_B = \frac{\sum y_i + \alpha}{n + \alpha + \beta}, \qquad [\widehat{p(1-p)}]_B = \frac{(\sum y_i + \alpha)\,(n - \sum yi + \beta)}{(n + \alpha + \beta + 1)\,(n + \alpha + \beta)}.$$

Bayes Estimators and Sufficient Statistics

- We write $\widehat{p}_B$ in Example 5 as

$$\begin{aligned}
\widehat{p}_B &= \frac{\sum y_i + \alpha}{n + \alpha + \beta} \\
&= \frac{n}{n + \alpha + \beta} \frac{\sum y_i}{n} + \frac{\alpha + \beta}{n + \alpha + \beta} \cdot \frac{\alpha}{\alpha + \beta} \\
&= \frac{n}{n + \alpha + \beta} \overline{Y} + \frac{\alpha + \beta}{n + \alpha + \beta} \left(\frac{\alpha}{\alpha + \beta} \right),
\end{aligned}$$

  that is, it is the weighted mean of $\overline{Y}$ (MLE for p) and the prior distribution associated to p.

- For larger sample size the weight of p decreases while the weight of sample mean increases. Since $E(\overline{Y}) \neq p$, the Bayes estimator for p is biased. Bayes estimators are biased in general.

- Estimators obtained in Example 5 are both functions of the sufficient statistic $\sum Y_i$. This is no coincidence since a Bayes estimator is always a function of a sufficient statistic, a result that follows from the factorization criterion.

- If U is a sufficient statistic for θ based on a random sample Y_1, Y_2, $\ldots$, Y_n, then
$$L(y_1, y_2, \ldots, y_n | \theta) = k(u, \theta) h(y_1, y_2, \ldots, y_n),$$
  where $k(u, \theta)$ is a function only of u and θ, and $h(y_1, y_2, \ldots, y_n)$ does not depend on θ.

- In addition [6], the function $k(u, \theta)$ can (but need not) be chosen to be the probability mass or density function of the statistic U. In accord with the notations in this chapter, we write the conditional density of $U|\theta$ as $k(u|\theta)$.

- Then, because $h(y_1, y_2, \ldots, y_n)$ is not a function of θ,

$$
\begin{aligned}
g^*(\theta|y_1, y_2, \ldots, y_n) &= \frac{L(y_1, y_2, \ldots, y_n|\theta)g(\theta)}{\int\limits_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta)g(\theta)\mathrm{d}\,\theta} \\
&= \frac{k(u|\theta)h(y_1, y_2, \ldots, y_n)g(\theta)}{\int\limits_{-\infty}^{\infty} k(u|\theta)h(y_1, y_2, \ldots, y_n)g(\theta)\mathrm{d}\,\theta} \\
&= \frac{k(u|\theta)g(\theta)}{\int\limits_{-\infty}^{\infty} k(u|\theta)g(\theta)\mathrm{d}\,\theta}.
\end{aligned}
$$

- Therefore, in cases where the distribution of a sufficient statistic U is known, the posterior can be determined by using the conditional density of $U|\theta$.

Example

Example 6. Let Y_1, Y_2, ... , Y_n denote a random sample from a normal population with unknown mean μ and known variance σ_o^2. The conjugate prior distribution for μ is normal with known meand η and known variance δ^2. Find the posterior distribution and the Bayes estimator for μ.

Solution

Since $U = \sum Y_i$ is a sufficient statistics for μ şi and is $N(n\mu, n\sigma_o^2)$ distributed,

$$
L(u|\mu) = \frac{1}{\sqrt{2\pi n\sigma_o^2}} \exp\left[\frac{1}{2n\sigma_o^2}(u - n\mu)^2\right], \qquad \mu \in \mathbb{R},
$$

and the joint density of U and μ is

$$
\begin{aligned}
f(u, \mu) &= L(u|\mu)g(\mu) \\
&= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}} \exp\left[-\frac{1}{2n\sigma_o^2}(u - n\mu)^2 - \frac{1}{2\delta^2}(\mu - \eta)^2\right], \ u,\mu \in \mathbb{R}.
\end{aligned}
$$

We rewrite the exponent as:

$$
\begin{aligned}
&-\frac{1}{2n\sigma_o^2}(u - n\mu)^2 - \frac{1}{2\delta^2}(\mu - \eta)^2 \\
&= -\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2(u - n\mu)^2 + n\sigma_o^2(\mu - \eta)^2\right] \\
&= -\frac{1}{2n\sigma_o^2\delta^2}\left[n^2\mu^2\delta^2 + n\mu^2\sigma_o^2 - 2nu\mu\delta^2 - 2n\mu\eta\sigma_o^2 + n\eta^2\sigma_o^2 + u^2\delta^2\right] \\
&= -\frac{1}{2n\sigma_o^2\delta^2}\left[n\left(\sigma_o^2 + n\delta^2\right)\mu^2 - 2n\left(u\delta^2 + \eta\sigma_o^2\right)\mu + n\eta^2\sigma_o^2 + u^2\delta^2\right]
\end{aligned}
$$

$$= -\frac{1}{2\sigma_o^2\delta^2}\left[\left(\sigma_o^2 + n\delta^2\right)\mu^2 - 2\left(u\delta^2 + \eta\sigma_o^2\right)\mu\right]$$
$$- \frac{1}{2n\sigma_o^2\delta^2}\left[n\eta^2\sigma_o^2 + u^2\delta^2\right]$$
$$= -\frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left[\mu^2 - 2\frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2}\mu + \left(\frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2}\right)^2\right]$$
$$- \frac{1}{2n\sigma_o^2\delta^2}\left[n\eta^2\sigma_o^2 + u^2\delta^2 - \left(\frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2}\right)^2\right].$$

Finally:

$$- \frac{1}{2n\sigma_o^2}\left(u - n\mu\right)^2 - \frac{1}{2\delta^2}\left(\mu - \eta\right)^2 =$$
$$- \frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2}\right)^2 - \frac{1}{2\left(n\delta^2 + \sigma_o^2\right)}\left(u - n\eta\right)^2.$$

Therefore,

$$f(u,\mu) = \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{1}{2n\sigma_o^2}\left(u - n\mu\right)^2 - \frac{1}{2\delta^2}\left(\mu - \eta\right)^2\right]$$
$$= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2}\right)^2\right]\cdot$$
$$\exp\left[-\frac{1}{2\left(n\delta^2 + \sigma_o^2\right)}\left(u - n\eta\right)^2\right]$$

and

$$m(u) =$$
$$\frac{e^{-\frac{1}{2\left(n\delta^2+\sigma_o^2\right)}(u-n\eta)^2}}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2}\right)^2\right]\mathrm{d}\,\mu$$
$$= \frac{e^{-\frac{1}{2\left(n\delta^2+\sigma_o^2\right)}(u-n\eta)^2}}{\sqrt{2\pi n\left(n\delta^2 + \sigma_o^2\right)}}\int_{-\infty}^{\infty}\frac{\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{u\delta^2+\eta\sigma_o^2}{n\delta^2+\sigma_o^2}\right)^2\right]\mathrm{d}\,\mu}{\sqrt{\frac{2\pi n\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}.$$

The above integral is the integral of a normal pdf; its integral is 1. The marginal density of U is $N(n\eta, n^2\delta^2 + n\sigma_o^2)$. Further, the posterior density of μ given

$U = u$ is

$$g^*(\mu|u) = \frac{f(u,\mu)}{m(u)}$$
$$= \frac{1}{\sqrt{\frac{2\pi\sigma_o^2}{n\delta^2+\sigma_o^2}}} \exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{u\delta^2+\eta\sigma_o^2}{n\delta^2+\sigma_o^2}\right)^2\right],$$

that is normal with mean and variance

$$\mu^* = \frac{u\delta^2+\eta\sigma_o^2}{n\delta^2+\sigma_o^2}, \qquad \sigma^{*2} = \frac{\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}.$$

It follows the Bayesian estimator for μ is

$$\widehat{\mu}_B = \frac{\delta^2 U + \sigma^2\eta}{n\delta^2+\sigma_o^2} = \frac{n\delta^2}{n\delta^2+\sigma_o^2}\overline{Y} + \frac{\sigma^2}{n\delta^2+\sigma_o^2}\eta.$$

Weighted mean of MLE, $\overline{Y}$ and prior mean η. If n increases, the weight of $\overline{Y}$ increases, while the weight of η decreases.

5 Bayesian Credible Intervals

Bayesian Credible Intervals

- In the Bayesian context, the parameter θ is a random variable with posterior density function $g^*(\theta)$. If we consider the interval (a, b), the posterior probability that the random variable θ is in this interval is

  $$P^*(a \le \theta \le b) = \int_a^b g^*(\theta)\mathrm{d}\,\theta$$

  If the posterior probability $P^*(a \le \theta \le b) = 1 - \alpha$, we said that (a, b) is a $100(1-\alpha)\%$ credible interval for θ.

Example 7. A manufacturer of gunpowder has developed a new powder, which was tested in eight shells. The resulting muzzle velocities, in feets per second, were as follows:

3005	2925	2935	2965
2995	3005	2937	2905

Find a 95% confidence interval for the true average velocity for shells of this type if muzzle velocities are approximately normally distributed. The manufacturer claims that $\mu \ge 3280.84$ feet/s. Do the sample data provide sufficient evidence to contradict the manufacturer's claim at 0.025 level of significance?

Solution. See `bayes/bulletsfps.pdf`. □

Example

In Example 7, that muzzle velocities were normally distributed with unknown mean μ and unknown variance σ^2. Suppose we wish to find a credicle Bayes interval for μ and the muzzle velocities are with a high probability within şi credem că 30 feet/s of their mean μ. Since for a normal population $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$, it is reasonable to assume that the distribution of muzzle velocities is $N(\mu, \sigma_o^2)$ such that $2\sigma_o = 30$, i.e. $\sigma_o^2 = 225$. If, prior to observing any data, we believed that there was a high probability that $\mu \in (2700, 2900)$, we might choose to use a conjugate normal prior for μ, $N(\eta, \delta^2)$ such that $\eta - 2\delta = 2700$ and $\eta + 2\delta = 2900$, i.e. $\eta = 2800$ and $\delta^2 = 50^2 = 2500$. Note that we have assumed considerably more knowledge of muzzle velocities than we did in Example 7 where we assumed only the normality (with unknown variance). To use the additional information we will use the initial sample. Use the general form for the posterior density for $\mu|u$ developed in Example 6 to give a 95% credible interval for μ.

Solution

This is a special case of Example 6 with

$$n = 8, \qquad u = \sum yi = 23672, \qquad \sigma_0^2 = 225, \qquad \eta = 2800, \qquad \delta^2 = 2500.$$

In Example 6, we have proven that the posterior density of $\mu|u$ is $N(\eta^*, \delta^{*2})$ where

$$\eta^* = \frac{u\delta^2 + \eta\sigma_o^2}{n\delta^2 + \sigma_o^2} = \frac{2500 \cdot 23672 + 2800 \cdot 225}{8 \cdot 2500 + 225} = 2957.23$$

$$\delta^{*2} = \frac{\sigma_o^2 \delta^2}{n\delta^2 + \sigma_o^2} = \frac{225 \cdot 2500}{8 \cdot 2500 + 225} = 27.81.$$

Remind that for $W \sim N(\mu_W, \sigma_W^2)$ we have

$$P(\mu_W + z_{\alpha/2}\sigma_W \leq W \leq \mu_W + z_{1-\alpha/2}\sigma_W) = 1 - \alpha.$$

It follows that a 95% credible interval for μ is

$$(\eta^* - 1.96\delta^*, \eta^* + 1.96\delta^*) = (2946.89, 2967.57).$$

Example

Example 8. If $Y_1, Y_2, \ldots, Y_n$ is a random sample from an exponentially distributed population with density $f(y|\theta) = \theta e^{-\theta y}, 0 < y$ and the conjugate gamma prior (with parameters α and β) for θ was employed, then the posterior density for θ is a gamma density with parameters $\alpha^* = n + \alpha$ and $\beta^* = \beta/(\beta \sum yi + 1)$. Assume that an analyst chose $\alpha = 3$ and $\beta = 5$ as appropriate parameter values for the prior and that a sample of size $n = 10$ yielded that $\sum y_i = 1.26$. Construct 90% credible intervals for θ and the population mean $\mu = 1/\theta$.

Solution

In this Example

$$n = 10, \qquad u = \sum y_i = 1.26, \qquad \alpha = 3, \qquad \beta = 5.$$

The posterior density of θ is a gamma density with α^* and β^* given by

$$\alpha^* = n + \alpha = 13$$
$$\beta^* = \frac{\beta}{\beta \sum y_1 + 1} = \frac{5}{5 \cdot 1.26 + 1} = 0.685.$$

We will find two values a and b such that

$$P^*(a \le \theta \le b) = .90.$$

This (a, b) will be a 90% credible interval for θ. Further, since

$$a \le \theta \le b \Longleftrightarrow 1/b \le 1/\theta \le 1/a,$$

We need the 0.05 and 0.95 quantiles of $\gamma(\alpha^*, \beta^*)$. See `Bayes/gammaBayes.pdf`

6 Bayesian Tests of Hypotheses

Bayesian Tests of Hypotheses

- Tests of hypotheses can also be approached from a Bayesian perspective.
- If we are interested in testing that the parameter θ lies in one of two sets of values, Ω_0 and Ω_a, we can use the posterior distribution of θ to calculate the posterior probability that θ is in each of these sets of values.
- When testing $H_0 : \theta \in \Omega_0$ versus $H_a : \theta \in \Omega_a$, one often-used approach is to compute the posterior probabilities $P^*(\theta \in \Omega_0)$ and $P^*(\theta \in \Omega_a)$ and accept the hypothesis with the higher posterior probability. That is, for testing $H_0 : \theta \in \Omega_0$ versus $H_a : \theta \in \Omega_a$,

accept H_0 if $P^*(\theta \in \Omega_0) > P^*(\theta \in \Omega_a)$,

accept H_a if $P^*(\theta \in \Omega_a) > P^*(\theta \in \Omega_0)$.

Example

Example 9. In Example 14, we obtained a 95% credible interval for the mean muzzle velocity associated with shells prepared with a reformulated gunpowder. We assumed that the associated muzzle velocities are normally distributed with mean μ and variance $\sigma_o^2 = 225$ and that a reasonable prior density for μ is normal with mean $\eta = 2800$ and variance $\delta^2 = 2500$. Use the data from Example 7 to obtain that the posterior density for μ is normal with mean $\eta^* = 2957.23$ and standard deviation $\delta^* = 5.274$. Conduct the Bayesian test for

$$H_0 : \mu \le 2950 \qquad \text{vs.} \qquad H_a : \mu > 2950.$$

Solution

In this case, if Z has a standard normal distribution,

$$\begin{aligned}P^*(\theta \in \Omega_0) &= P^*(\mu \leq 2950)\\ &= P\left(Z \leq \frac{2950-\eta^*}{\delta^*}\right) = P\left(Z \leq \frac{2950-2957.23}{5.274}\right)\\ &= P(Z \leq -1.3709) = 0.0853.\end{aligned}$$

and $P^*(\theta \in \Omega_a) = P^*(\mu > 2950) = 1 - P^*(\mu \leq 2950) = 1 - 0.0853 = 0.9147$. Thus, we see that the posterior probability of H_a is much larger than the posterior probability of H_0 and our decision is to accept $H_a : \mu > 2950$.

Comments

- Again, we note that if a different analyst uses the same data to conduct a Bayesian test for the same hypotheses but different values for any of η, δ^2 şi σ_o^2 , she will obtain posterior probabilities of the hypotheses that are different than those obtained in Example 9.
- Thus, different analysts with different choices of values for the prior parameters might reach different conclusions.
- In the frequentist settings, the parameter θ has a fixed but unknown value, and any hypothesis is either true or false. If $\theta \in \Omega_0$, then the null hypothesis is certainly true (with probability 1), and the alternative is certainly false.
- The only way we could *know* whether or not $\theta \in \Omega_0$ is if *we knew the true value of* θ. If this were the case, conducting a test of hypotheses would be superfluous. For this reason, the frequentist makes no reference to the probabilities of the hypotheses but focuses on the probability of a type I error, α, and the power of the test, $power(\theta) = 1 - \beta(\theta)$.
- Conversely, the frequentist concepts of size and power are not of concern to an analyst using a Bayesian test.

Example

Example 10. In Example 8, we obtained credible intervals for θ and the population mean μ based on a random sample $Y_1, Y_2, \ldots, Y_n$, from an exponentially distributed population with density $f(y|\theta) = \theta e{-}^{\theta y}, 0 < y$. Using a conjugate gamma prior for θ with parameters $\alpha = 3$ and $\beta = 5$, we obtained that the posterior density for θ is a gamma density with parameters $\alpha^* = 13$ and $\beta^* = .685$. Conduct a Bayesian test for

$$H_0 : \mu > .12 \qquad \text{vs} \qquad H_a : \mu \leq .12.$$

Solution

Since the mean of the exponential distribution is $\mu = 1/\theta$, the hypotheses are equivalent to $H_0 : \theta < 1/0.12 = 8.333$ versus $H_a : \theta \geq 0$. Because the posterior density for θ is a gamma density with parameters $\alpha^* = 13$ and $\beta^* = .685$, then

$$P^*(\theta \in \Omega_0) = P^*(\theta < 0.833)$$
$$P^*(\theta \in \Omega_a) = P^*(\theta \geq 0.833).$$

In R *Bayes/Bayestestgamma.pdf*. The posterior probability of H_a is somewhat larger than the posterior probability of H_0. It is up to the analyst to decide whether the probabilities are sufficiently different to merit the decision to accept $H_a : \mu \leq 0.12$.

7 An Alternative Method

An Alternative Method

- There is a shortcut to finding the all-important posterior density for θ. If $L(y_1, y_2, \ldots, y_n|\theta)$ denotes the conditional likelihood of the data and θ has a continuous prior density $g(\theta)$, then the posterior density of θ is

$$g^*(\theta|y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n|\theta)g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta)g(\theta)\,\mathrm{d}\,\theta}.$$

- Notice that the denominator on the right hand side of the expression depends on $y_1, y_2, \ldots, y_n$, but *does not depends* on θ. (Definite integration with respect to θ produces a result that is free of θ.) Realizing that, with respect to θ, the denominator is a constant, we can write

$$g^*(\theta|y_1, y_2, \ldots, y_n) = c(y_1, y_2, \ldots, y_n)L(y_1, y_2, \ldots, y_n|\theta)g(\theta),$$

  where

$$c(y_1, y_2, \ldots, y_n) = \frac{1}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta)g(\theta)\,\mathrm{d}\,\theta}$$

  does not depend on θ.

- Further, notice that, because the posterior density is a *bona fide* density function, the quantity $c(y_1, y_2, \ldots, y_n)$ must be such that

$$\int_{-\infty}^{\infty} g^*(\theta|y_1, y_2, \ldots, y_n)\,\mathrm{d}\,\theta =$$

$$c(y_1, y_2, \ldots, y_n) \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta)g(\theta)\,\mathrm{d}\,\theta = 1$$

- Finally, we see that the posterior density is proportional to the product of the conditional likelihood of the data and the prior density for θ:

$$g^*(\theta|y_1, y_2, \ldots, y_n) \propto L(y_1, y_2, \ldots, y_n|\theta)g(\theta),$$

where the proportionally constant is chosen so that the integral of the posterior density function is 1.

- We illustrate by reconsidering Example 1.

Example 11. Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Bernoulli distribution where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$ and assume that the prior distribution for p is $beta(\alpha, \beta)$. Find the posterior distribution for p.

Solution

- As before,

$$\begin{aligned}
L(y_1, y_2, \ldots, y_n|p)g(p) &= p(y_1, y_2, \ldots, y_n|p)g(p) \\
&= p^{\sum y_i}(1-p)^{n-\sum y_i}\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \\
g^*(p|y_1, y_2, \ldots, y_n) &\propto p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}.
\end{aligned}$$

- From the above, we recognize that the resultant posterior for p must be beta with parameters $\alpha^* = \sum y_i + \alpha$ şi $\beta^* = n - \sum yi + \beta$.
- Advantage of proportionality method: less work.
- Disadvantage? We never exhibited the predictive mass function for the data and lost the opportunity to critique the Bayesian model.

8 References

References

References

[1] Bayes, T. 1764. "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans. Roy. Soc.* 53, 370–418.

[2] Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2d ed. New York: Springer-Verlag.

[3] Box, G. E. P. 1980. "Sampling and Bayes' Inference in Scientific Modeling and Robustness," *J. of the Royal Statistical Society*, Series A 143, 383–430.

[4] Box, G. E. P., and G. C. Tiao. 1992. *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics.

[5] Casella, G., and R. L. Berger. 2002. *Statistical Inference*, 2d ed. Pacific Grove, Calif.: Duxbury.

[6] Hogg, R. V., J. W. McKean, and A. T. Craig. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N.J.: Prentice Hall.

[7] Kepner, J., and D. Wackerly. 2002. "Observations on the Effect of the Prior Distribution on the Predictive Distribution in Bayesian Inferences," *Journal of Applied Statistics* 29(5): 761–769.

[8] Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill.

[9] Rice, J. A. 1995. *Mathematical Statistics and Data Analysis*, 2d ed. Belmont, Calif.: Duxbury.

[10] Wackerly D. D., and Mendenhall, W., and Scheaffer R. L. 2008. *Mathematical Statistics with Applications*, 7th ed., Thomson Brooks/Cole

8.1