BAYESIAN STATISTICS

Chapter 6

Instructor: Seokho Lee (이석호)

Hankuk University of Foreign Studies

6. Estimating Proportion

6.1. Dichotomous response variable

The response variable taking either of 'success'/'failure' or '0'/'1' is called the **dichotomous response variable**. For the dichotomous response variable, our interest is on the **proportion** such as 'the success probability' or 'recovery rate.'

Such dichotomous nominal variable can be coded quantitatively.

- Product: 0 if good, 1 if bad
- Treatment: 0 if ineffective, 1 if effective
- etc.

The inference on the proportion is usually conducted under the normal theory in the introductory statistics, or logistic regression model and loglinear model in the advanced statistics.

6.2. Prior, likelihood, and posterior

Consider a sport game. The prior distribution for the winning probability, p, of the team A may be obtained from some prior information: winning rate for the recent games, current physical and mental condition, etc.

Suppose the prior distribution of p is given below:

p	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
prior probability	.00	.05	.10	.20	.25	.15	.10	.08	.05	.02	.00

Thus, the probability of A wins is

$$\begin{aligned}
\Pr(\text{A wins}) &= \Pr(\text{A wins}|p=.0)\Pr(p=.0) \\
&\quad + \Pr(\text{A wins}|p=.1)\Pr(p=.1) \\
&\quad \ldots \\
&\quad + \Pr(\text{A wins}|p=1.0)\Pr(p=1.0) \\
&= (.0)(0)+(.1)(.05)+(.2)(.1)+\cdots+(1.0)(.0) \\
&= .334.
\end{aligned}$$

33% is the average of the prior probabilities. This is called **predicted probability**.

This predicted probability 33% that the team A wins is solely obtained from the prior distribution representing the prior information.

Suppose we observe that A wins the first 2 games of the recent 3 games. Then, is still the probability .334? That's not a sensible conclusion. The probability should be adjusted by the new information that the first 2 wins of 3 games. Probably, the updated winning probability must be larger than the previous predicted probability .334 as well as the adjusted probability must be updated.

The updated prior distribution is the posterior distribution. The posterior distribution is obtained using the prior and the likelihood.

The likelihood is the probability that the observed data occur when a model is given. If the probability of 'A wins' is .4, the likelihood (the probability that A wins the first 2 games of 3 games) is $.4 \times .4 \times .6 = .096$. The likelihood depends on the assumed models (i.e. which p is used, in this example).

p (winning probability)	prior	likelihood of the first 2 wins of 3 games
0.0	.00	$(.0)(.0)(1.0) = .000$
0.1	.05	$(.1)(.1)(.9) = .009$
0.2	.10	$(.2)(.2)(.8) = .032$
0.3	.20	$(.3)(.3)(.7) = .063$
0.4	.24	$(.4)(.4)(.6) = .096$
0.5	.15	$(.5)(.5)(.5) = .125$
0.6	.10	$(.6)(.6)(.4) = .144$
0.7	.08	$(.7)(.7)(.3) = .147$
0.8	.05	$(.8)(.8)(.2) = .128$
0.9	.02	$(.9)(.9)(.1) = .081$
1.0	.00	$(1.0)(1.0)(.0) = .000$
	1.00	

Now, let E be the event that A wins the first 2 games of the 3 games. Then,

$$\begin{aligned}
\Pr(E) &= \Pr(E|p=.0)\Pr(p=.0) + \Pr(E|p=.1)\Pr(p=.1) + \cdots \\
&\quad + \Pr(E|p=1.0)\Pr(p=1.0) \\
&= \sum_x \underbrace{\Pr(E|p=x)}_{\text{likelihood}}\underbrace{\Pr(p=x)}_{\text{prior}}
\end{aligned}$$

p (winning probability)	prior	likelihood	prior $\times$ likelihood
0.0	.00	.000	.000000
0.1	.05	.009	.000045
0.2	.10	.032	.003200
0.3	.20	.063	.012600
0.4	.24	.096	.024000
0.5	.15	.125	.018750
0.6	.10	.144	.014400
0.7	.08	.147	.011760
0.8	.05	.128	.006400
0.9	.02	.081	.001620
1.0	.00	.000	.000000
	1.00		$\Pr(E) = .092775$

The posterior is, then, obtained by following formula:

$$\Pr(p = x|E) = \frac{\Pr(p = x)\Pr(E|p = x)}{\Pr(E)} = \frac{\text{prior} \times \text{likelihood}}{.092775}.$$

Using this, we get the posterior:

p (winning probability)	prior	likelihood	prior$\times$likelihood	posterior
0.0	.00	.000	.000000	.0000000
0.1	.05	.009	.000045	.0004850
0.2	.10	.032	.003200	.0344921
0.3	.20	.063	.012600	.1358124
0.4	.24	.096	.024000	.2486904
0.5	.15	.125	.018750	.2021019
0.6	.10	.144	.014400	.1552142
0.7	.08	.147	.011760	.1267583
0.8	.05	.128	.006400	.0689841
0.9	.02	.081	.001620	.0174616
1.0	.00	.000	.000000	.0000000
	1.00		$\Pr(E) = .092775$	1.0000000

The expected probability using the posterior distribution is

$$\begin{aligned}
\Pr(\text{A wins}) &= \Pr(\text{A wins}|p=.0)\Pr(p=.0) + \Pr(\text{A wins}|p=.1)\Pr(p=.1) \\
&\quad + \cdots + \Pr(\text{A wins}|p=1.0)\Pr(p=1.0) \\
&= (.0)(0) + (.1)(.0000458) + (.2)(.0344921) + \cdots + (1.0)(.0) \\
&= .505.
\end{aligned}$$

This is the Bayes estimates for the proportion p.

The estimate of p, .334, from the prior information is updated by .505 by adding the new information!!

6.5. Conjugate prior: beta distribution

In the previous example, the prior distribution for p is given specifically at a set of discrete values. (You may wonder why $p = .15$ or other values are not considered.) This specific prior is available and acceptable only when the past information validates such prior distribution. However, most of the real problems are not in the case. Instead we better assume the distribution family for the prior distribution.

For example, since p is the probability between 0 and 1, we may consider beta distribution $\mathrm{B}(\alpha, \beta)$. At this moment, the **hyperparameters** α and β are not specified.

Suppose we observe the new information that the team A got x wins of the n games. Then

- prior $= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \propto p^{\alpha-1}(1-p)^{\beta-1}$
- likelihood $= \binom{n}{x} p^x (1-p)^{n-x} \propto p^x(1-p)^{n-x}$
- posterior $=$ likelihood $\times$ prior $\propto p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \sim \mathrm{Beta}(\alpha+x, \beta+n-x)$

The posterior mean is $\frac{\alpha+x}{\alpha+\beta+n}$, which is the Bayes estimator under the square-error loss function.

Example (6-1)

The prior distribution of p, the probability that A wins, is $p \sim Beta(2,3)$. Suppose we got data that A wins 2 games of 3 games. Give the posterior and the Bayes estimator of p.

(solution) The posterior is $\text{Beta}(2+2, 3+3-2)$ or $\text{Beta}(4,4)$. And the Bayes estimate is $4/(4+4) = .5$.

Note that the Bayes estimator can be expressed as

$$\hat{p}_B = \frac{\alpha+x}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n} \cdot \frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n} \cdot \frac{x}{n}.$$

Thus, the Bayes esitmator is the weighted average between the prior mean and the maximum likelihood estimator (mle).

- If n is large, the Bayes estimator is close to the mle and the prior mean does not affect very much.
- If n is small, the prior mean considerably affect the Bayes estimator.

The beta prior may not make sense in the case that the prior distribution is evidently bimodal. Moreover, we must specify the hyperparameters, α and β, in advance. The **subjective Bayesian approach** will specify α and β based on the subjective belief on p. The **empirical Bayesian approach** tries to estimate α and β from the data, which is out of scope in this lecture.

6.4. Interval estimation for the proportion

In the frequentist statistics (traditional statistics), the confidence interval is a measure used for the interval estimation for the unknown parameter.

When $X \sim \mathrm{B}(n, p)$, the $100(1-\alpha)\%$ confidence interval for the proportion, p, is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

with $\hat{p} = X/n$.

For example, if $n = 25$ and $X = 5$, then 95% confidence interval is

$$.2 \pm 1.96\sqrt{(.2)(.8)/25} = .2 \pm (1.96)(.4)/5 = .2 \pm .1568 = (.0432, .3568).$$

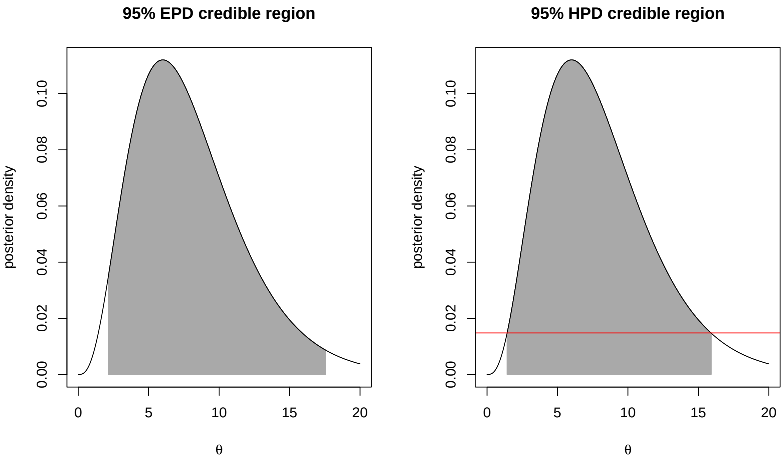
Think about its interpretation:

- We often say it is wrong that "the probability of $p \in (.0432, .3568)$ is .95."
  It is because p either inside or outside the interval. This view comes from that the proportion, p, is not a random variable but a unknown constant. Thus, the probability that p is inside the interval must be 1 (if p is inside the interval) or 0 (if p is not in the interval).
- The classical interpretation is this: "This particular interval $(.0432, .3568)$ may or may not include the true value p. However, we can say that if we conduct many experiments in the same way and this way of getting the confidence interval can be computed, then 95% of such confidence intervals include the true p."
  This interpretation is not probabilistic nor, thus, comfortable.

Contrast to the frequentist approach (confidence interval), Bayesian approaches view the parameter, p, as a random variable. Thus, interval estimate of p is interpreted straightforwardly in the probabilistic way.

HPD and EPD credible regions (or intervals) are used for the interval estimation for the parameter in Bayesian inference.

- **HPD**: This gives the shortest credible regions, which is good in the precision sense. However, it is hard to compute, and its interpretation is difficult.
- **EPD**: Easy to inteprete and compute. For example, when $p \sim \text{Beta}(a, b)$, $100(1-\alpha)\%$ EPD credible interval is (l, u) where l is the $100(\alpha/2)\%$ and u is the $100(1-\alpha/2)\%$ percentile, respectively.



95% EPD credible region
95% HPD credible region
posterior density
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
θ

Example (6-2)

If the posterior is $p \sim Beta(6, 3)$, give the 90% EPD credible interval for p.

(solution) (.40, .89)

```
> qbeta(0.05,6,3)
[1] 0.4003106
> qbeta(0.95,6,3)
[1] 0.8888873
```

Example (6-3)

19 persons having cold medication were cured and 8 were not. If the prior for p (probability of recovery) is $Beta(1, 1)$, what is the 95% EPD credible interval for p?

(solution) The posterior distribution for p is $\text{Beta}(1 + 19, 1 + 8)$=$\text{Beta}(20, 9)$. Thus, 95% EPD credible interval is $(.51, .84)$.

```
> qbeta(0.025,20,9)
[1] 0.5133317
> qbeta(0.975,20,9)
[1] 0.841224
```

Example (6-4)

None of 20 samples turns out to be defectives. If the prior distribution for the defective rate p is Beta$(1, 4)$, provide the 90% EPD credible interval for p.

(solution) The posterior distribution is $\text{Beta}(1 + 0, 4 + 20) = \text{Beta}(1, 24)$. Thus, 90% EPD credible interval is $(.002, .117)$.

```
> qbeta(0.05,1,24)
[1] 0.002134938
> qbeta(0.95,1,24)
[1] 0.1173462
```

6.5. Ignorance on the prior distribution

Sometimes there is no information to guess the prior distribution on p. This is called **complete prior ignorance** or **total ignorance**. For such case, we use $\text{Beta}(0, 0)$ for the prior (meaning that no success, no failure, thus, no trials at all). Suppose we get the new information that s successes and f failures. Then

- **prior**: $\propto p^{-1}(1-p)^{-1}$
- **likelihood**: $\propto p^s(1-p)^f$
- **posterior**: $\propto p^{s-1}(1-p)^{f-1} \sim \text{Beta}(s, f)$

Note that the prior is not integrable, thus not a proper density. This kind of prior is called **improper prior**. Even though the prior is not properly defined as a distribution, the corresponding posterior is a proper distribution. Moreover, Bayes estimator is the same as the maximum likelihood estimator, $s/(s+f)$.

Total ignorance of p can be interpreted as "no preference" on any value on $(0, 1)$. This can be represented by $\text{Uniform}(0, 1) = \text{Beta}(1, 1)$ as a "vague" prior. Then

- **prior**: $\propto p^0(1-p)^0$
- **likelihood**: $\propto p^s(1-p)^f$
- **posterior**: $\propto p^s(1-p)^f \sim \text{Beta}(s+1, f+1)$

Example (6-5)

1 defective is observed among 25 samples. Suppose the prior for p (defective rate) is Beta$(0, 0)$. Find 90% EPD credible interval for p.

(solution) Since the posterior is Beta$(1, 25)$, the 95% EPD credible interval is $(.002, .113)$.

```
> qbeta(0.05,1,25)
[1] 0.002049628
> qbeta(0.95,1,25)
[1] 0.1129281
```