IPB University
Inspiring Innovation with Integrity

Pengantar Statistika Bayes - STA1312

Konsep Inferensi Bayesian – Part 3

(*Decision Theory*: Pendugaan Parameter dan Pengujian Hipotesis)

Dr. Kusman Sadik, S.Si, M.Si
Program Studi Statistika dan Sains Data IPB
Tahun Akademik 2023/2024

Inferensi Parameter : Frequentist vs Bayesian

- Sebaran posterior dari parameter yang diberikan data memberikan **inferensi lengkap** dari sudut pandang Bayesian.
- Hal ini akan merangkum berbagai informasi tentang parameter berdasarkan data.
- Ada tiga inferensi penting terkait parameter tersebut:
  1. Pendugaan titik (*point estimation*).
  2. Pendugaan selang (*interval estimation*)
  3. Pengujian hipotesis (*hypothesis testing*).
- Pada ketiga inferensi tersebut **ada perbedaan** antara metode frequentist dengan Bayesian.
- Pada metode frequentist, inferensi tersebut didasarkan pada sebaran penarikan contoh data (***sampling distibution***).
- Sementara pada metode Bayesian, inferensi tersebut didasarkan pada ***sebaran posterior***.

Pendugaan Titik

(*Point Estimation*)

Teori Keputusan (*Decision Theory*) Bayesian

The outcome of a Bayesian analysis is the posterior distribution, which combines the prior information and the information from data. However, sometimes we may want to summarize the posterior information with a scalar, for example the mean, median or mode of the posterior distribution. In the following, we show how the use of scalar estimator can be justified using statistical decision theory.

Let $L(\theta, \hat{\theta})$ denote the loss function which gives the cost of using $\hat{\theta} = \hat{\theta}(y)$ as an estimate for θ. We define that $\hat{\theta}$ is a *Bayes estimate* of θ if it minimizes the *posterior expected loss*

$$\mathrm{E}[L(\theta, \hat{\theta})|y] = \int L(\theta, \hat{\theta}) p(\theta|y) d\theta.$$

Teori Keputusan (*Decision Theory*) Bayesian (cont.)

On the other hand, the expectation of the loss function over the sampling distribution of y is called *risk function*:

$$R_{\hat{\theta}}(\theta) = \mathsf{E}[L(\theta, \hat{\theta})|\theta] = \int L(\theta, \hat{\theta}) p(y|\theta) dy.$$

Further, the expectation of the risk function over the prior distribution of θ,

$$\mathsf{E}[R_{\hat{\theta}}(\theta)] = \int R_{\hat{\theta}}(\theta) p(\theta) d\theta,$$

is called *Bayes risk*.

Teori Keputusan (*Decision Theory*) Bayesian (cont.)

Absolute error loss:

Absolute error loss: $L(\theta, \hat{\theta}) = |\hat{\theta} - \theta|$. In general, if X is a random variable, then the expectation $\mathsf{E}(|X - d|)$ is minimized by choosing d to be the median of the distribution of X. Thus, the Bayes estimate of θ is the posterior median.

Quadratic loss function:

Quadratic loss function: $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$. In general, if X is a random variable, then the expectation $\mathsf{E}[(X - d)^2]$ is minimized by choosing d to be the mean of the distribution of X. Thus, the Bayes estimate of θ is the posterior mean.

Teori Keputusan (*Decision Theory*) Bayesian (cont.)

Quadratic loss function:

It turns out (we will prove this below) that *if the loss function is quadratic, the best estimate you can give is the posterior mean.* Here is the proof. The expected value of the loss is

$$\mathbb{E}\left[L(\theta,\hat{\theta})\right] = \int p(\theta|x)(\hat{\theta}-\theta)^2\,d\theta$$

Since we are summing (integrating) over all possible true θ values, the expected loss is only a function of our estimate $\hat{\theta}$. To minimise a function of one variable, you differentiate it and then set the derivative to zero. The derivative is

$$\begin{aligned}\frac{d}{d\hat{\theta}}\mathbb{E}\left[L(\theta,\hat{\theta})\right] &= \int p(\theta|x)\frac{d}{d\hat{\theta}}(\hat{\theta}-\theta)^2\,d\theta \\ &= \int p(\theta|x)2(\hat{\theta}-\theta)\,d\theta\end{aligned}$$

Teori Keputusan (*Decision Theory*) Bayesian (cont.)

Quadratic loss function:

$$\frac{d}{d\hat{\theta}}\mathbb{E}\left[L(\theta,\hat{\theta})\right] = \int p(\theta|x)\frac{d}{d\hat{\theta}}(\hat{\theta}-\theta)^2\,d\theta$$
$$= \int p(\theta|x)2(\hat{\theta}-\theta)\,d\theta$$

Setting this equal to zero and then solving for $\hat{\theta}$ gives the final result:

$$\hat{\theta} = \int \theta p(\theta|x)\,d\theta.$$

which is the posterior mean. Some people call the posterior mean the "Bayes Estimate" for this reason.

Penduga Titik Bagi θ

- For a real valued θ, standard **Bayes estimates** are
  - Posterior mean, or
  - Posterior median.
- The **posterior mean** is the Bayes estimate corresponding with squared error loss.
- The **posterior median** is the Bayes estimate for absolute deviation loss.

Perhatikan pada contoh kasus materi sebelumnya:

(Binomial Bayes estimation) Let $X_1, \ldots, X_n$ be iid Bernoulli(p). Then $Y = \sum X_i$ is binomial(n, p). We assume the prior distribution on p is beta(α, β). The joint distribution of Y and p is

$$f(y,p) = \left[\binom{n}{y} p^y (1-p)^{n-y}\right]\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\right] \begin{pmatrix}\text{conditional} \times \text{marginal} \\ f(y|p) \times \pi(p)\end{pmatrix}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1}.$$

The marginal pdf of Y is

$$f(y) = \int_0^1 f(y,p)\,dp = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)},$$

The marginal pdf of Y is

$$f(y) = \int_0^1 f(y,p)\,dp = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)},$$

posterior distribution, the distribution of p given y, is

$$f(p|y) = \frac{f(y,p)}{f(y)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{y+\alpha-1}(1-p)^{n-y+\beta-1},$$

which is beta$(y+\alpha, n-y+\beta)$. (Remember that p is the variable and y is treated as fixed.) A natural estimate for p is the mean of the posterior distribution, which would give us as the Bayes estimator of p,

$$\boxed{\hat{p}_{\mathrm{B}} = \frac{y+\alpha}{\alpha+\beta+n}.}$$

Catatan: $X \sim beta(\alpha,\beta) \rightarrow E(X) = \frac{\alpha}{\alpha+\beta}$

Credible Intervals

(*Posterior Intervals*)

Credible Intervals

Credible intervals are another useful kind of summary. They are used to make statements like "There is a 95% probability the parameter is between 100 and 150". The basic idea is to use the posterior distribution to find an interval $[a, b]$ such that

$$\boxed{P(a \leq \theta \leq b|x) \quad = \quad \alpha}$$

where α is some pre-defined probability. 95% seems to be the most popular choice. An example of a 95% credible interval is given in Figure 1.

Note that the interval shown in Figure 1 is not the only possible interval that would contain 95% of the probability. However, to make the notion of a credible interval precise, we usually use a *central* credible interval. A central credible interval containing an amount of probability α will leave $(1-\alpha)/2$ of the probability to its left and the same amount $(1-\alpha)/2$ of the probability to its right.

Credible Intervals (Cont.)

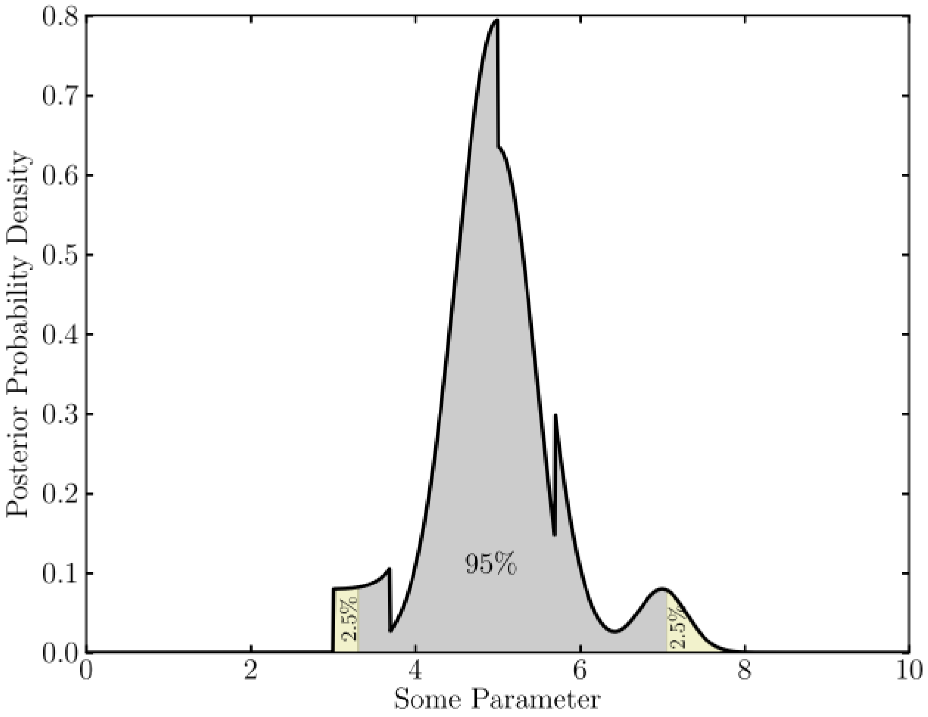
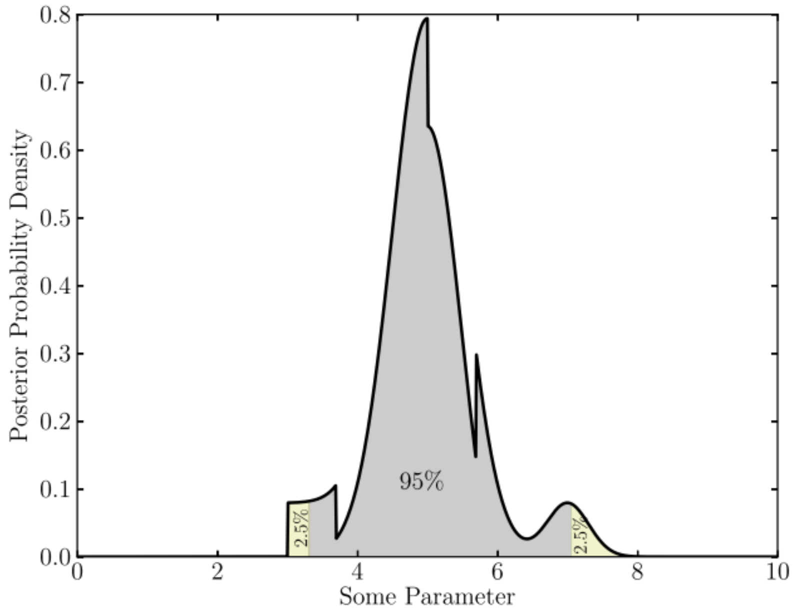


Figure 1. *A central 95% credible interval is defined as an interval that contains 95% of the posterior probability, while having 2.5% of the probability above the upper limit and 2.5% of the probability below the lower limit. The credible interval is formed by finding the edges of the grey region. In this case the credible interval is* [3.310, 7.056].

Metode Menghitung Credible Intervals

The method for computing credible intervals is closely related to the method for computing the posterior median. With the median, we found the value of θ which has 50% of the posterior probability to its left and 50% to its right. To find the lower end of a 95% credible interval, we find the θ value that has 2.5% of the probability to its left. To find the upper end we find the value of θ that has 2.5% of the posterior probability to its right, or 97.5% to the left.



Pengujian Hipotesis

(Berbasis *Posterior*)

Pengujian Hipotesis

Suppose that we need to decide between two hypotheses H_0 and H_1. In the Bayesian setting, we assume that we know prior probabilities of H_0 and H_1. That is, we know $P(H_0) = p_0$ and $P(H_1) = p_1$, where $p_0 + p_1 = 1$. We observe the random variable (or the random vector) Y. We know the distribution of Y under the two hypotheses, i.e, we know

$$f_Y(y|H_0), \quad \text{and} \quad f_Y(y|H_1).$$

Using Bayes' rule, we can obtain the posterior probabilities of H_0 and H_1:

$$P(H_0|Y = y) = \frac{f_Y(y|H_0)P(H_0)}{f_Y(y)},$$
$$P(H_1|Y = y) = \frac{f_Y(y|H_1)P(H_1)}{f_Y(y)}.$$

Pengujian Hipotesis (Cont.)

One way to decide between H_0 and H_1 is to compare $P(H_0|Y=y)$ and $P(H_1|Y=y)$, and accept the hypothesis with the higher posterior probability. This is the idea behind the *maximum a posteriori (MAP) test*. Here, since we are choosing the hypothesis with the highest probability, it is relatively easy to show that the error probability is minimized.

To be more specific, according to the MAP test, we choose H_0 if and only if

$$P(H_0|Y=y) > P(H_1|Y=y).$$

In other words, we choose H_0 if and only if

$$f_Y(y|H_0)P(H_0) > f_Y(y|H_1)P(H_1).$$

Note that as always, we use the PMF instead of the PDF if Y is a discrete random variable. We can generalize the MAP test to the case where you have more than two hypotheses. In that case, again we choose the hypothesis with the highest posterior probability.

Langkah Pengujian Hipotesis (Bayesian)

- $\boldsymbol{X} \mid \theta \sim f(\boldsymbol{x} \mid \theta)$.
- To test: $H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$.
- Prior distribution:
  - Prior probabilities $\Pr(H_0)$ and $\Pr(H_1)$ of the hypotheses.
  - Proper prior densities $\pi_0(\theta)$ and $\pi_1(\theta)$ on Θ_0 and Θ_1.
- Posterior distribution $\pi(\theta \mid \boldsymbol{x})$:

$$\Pr(H_0 \mid \boldsymbol{x}) = 1 - \Pr(H_1 \mid \boldsymbol{x})$$

- Based on the posterior odds. By default, H_0 accepted if $\Pr(H_0 \mid \boldsymbol{x}) > \Pr(H_1 \mid \boldsymbol{x})$

Studi Kasus 2

Suppose that the random variable X is transmitted over a communication channel. Assume that the received signal is given by

$$Y = X + W,$$

where $W \sim N(0, \sigma^2)$ is independent of X. Suppose that $X = 1$ with probability p, and $X = -1$ with probability $1 - p$. The goal is to decide between $X = 1$ and $X = -1$ by observing the random variable Y. Find the MAP test for this problem.

Studi Kasus 2

H_0: $X = 1$,

H_1: $X = -1$.

Under H_0, $Y = 1 + W$, so $Y|H_0 \sim N(1, \sigma^2)$. Therefore,

$$f_Y(y|H_0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2\sigma^2}}.$$

Under H_1, $Y = -1 + W$, so $Y|H_1 \sim N(-1, \sigma^2)$. Therefore,

$$f_Y(y|H_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2\sigma^2}}.$$

Thus, we choose H_0 if and only if

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2\sigma^2}} P(H_0) > \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2\sigma^2}} P(H_1).$$

Studi Kasus 2

Thus, we choose H_0 if and only if

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2\sigma^2}} P(H_0) \gtrless \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y+1)^2}{2\sigma^2}} P(H_1).$$

We have $P(H_0) = p$, and $P(H_1) = 1 - p$. Therefore, we choose H_0 if and only if

$$\exp\left(\frac{2y}{\sigma^2}\right) \gtrless \frac{1-p}{p}.$$

Equivalently, we choose H_0 if and only if

$$y \gtrless \frac{\sigma^2}{2} \ln\left(\frac{1-p}{p}\right).$$

Pengujian Hipotesis Dapat Menggunakan *Credible Interval* (Khusus Uji Hipotesis Dua Arah)

$$H_0 : \pi = \pi_0 \qquad\qquad H_1 : \pi \neq \pi_0$$

Compute a $(1 - \alpha) \times 100\%$ credible interval for π. If π_0 lies inside the credible interval, accept (do not reject) the null hypothesis $H_0 : \pi = \pi_0$; and if π_0 lies outside the credible interval, then reject the null hypothesis.

Catatan:
Berdasarkan nilai α tersebut, maka pengujian hipotesisnya disebut mempunyai taraf nyata (*significance level*) sebesar α.

Materi Praktikum

1

Suppose that the random variable X is transmitted over a communication channel. Assume that the received signal is given by

$$Y = 2X + W,$$

where $W \sim N(0, \sigma^2)$ is independent of X. Suppose that $X = 1$ with probability p, and $X = -1$ with probability $1 - p$. The goal is to decide between $X = -1$ and $X = 1$ by observing the random variable Y. Find the MAP test for this problem.

2

In the study of water quality in New Zealand streams, documented in McBride et al. (2002), a high level of *Campylobacter* was defined as a level greater than 100 per 100 ml of stream water. $n = 116$ samples were taken from streams having a high environmental impact from birds. Out of these, $y = 11$ had a high *Campylobacter* level. Let π be the true probability that a sample of water from this type of stream has a high *Campylobacter* level.

(a) Find the frequentist estimator for π.

(b) Use a $beta(1, 10)$ prior for π. Calculate the posterior distribution $g(\pi|y)$.

(c) Find the posterior mean and variance. What is the Bayesian estimator for π?

(d) Find a 95% credible interval for π.

(e) Test the hypothesis

$$H_0 : \pi = .10 \quad \text{versus} \quad H_1 : \pi \neq .10$$

at the 5% level of significance.

Catatan: Penentuan nilai sebaran dan peluangnya yang dipakai untuk mecari *credible interval* dan pengujian hipotesis dapat menggunakan **Program R** melalui metode **Monte Carlo** (Lihat buku Bolstad dan Curran (2017), hlm. 189).

3

The number of defects per 10 meters of cloth produced by a weaving machine has the *Poisson* distribution with mean μ. You examine 100 meters of cloth produced by the machine and observe 71 defects.

(a) Your prior belief about μ is that it has mean 6 and standard deviation 2. Find a $gamma(r, v)$ prior that matches your prior belief.

(b) Find the posterior distribution of μ given that you observed 71 defects in 100 meters of cloth.

(c) Calculate a 95% Bayesian credible interval for μ.

Catatan: Penentuan *credible interval* dapat menggunakan **Program R** melalui metode **Monte Carlo** (Lihat buku Bolstad dan Curran (2017), hlm. 208).

Pustaka

1. Reich BJ dan Ghosh SK. (2019). *Bayesian Statistical Methods*. Taylor and Francis Group.

2. Bolstad WM dan Curran JM. (2017). *Introduction to Bayesian Statistics 3^th^ Edition*. John Wiley and Sons, Inc.

3. Lee, Peter M. (2012). *Bayesian Statistics, An Introduction 4^th^ Edition*. John Wiley and Sons, Inc.

4. Albert, Jim. (2009). *Bayesian Computation with R*. Springer Science and Business Media.

5. Ghosh JK, Delampady M, dan Samanta T. (2006). *An Introduction to Bayesian Analysis, Theory and Methods*. Springer Science and Business Media.

Terima Kasih