IPB University
Inspiring Innovation with Integrity

Pengantar Statistika Bayes - STA1312

Frequentist vs Bayesian

Dr. Kusman Sadik, S.Si, M.Si
Program Studi Statistika dan Sains Data IPB
Tahun Akademik 2022/2023

Pendekatan Utama dalam Statistika

Ada dua pendekatan filosofis utama dalam statistika:

- Pendekatan frequentist (*frequentist approach*) atau sering pula disebut pendekatan klasik (*classical approach*).
- Pendekatan Bayesian (*Bayesian approach*). Pendekatan ini berpotensi memberikan banyak keuntungan mendasar dibandingkan pendekatan frequentist.

What is statistics?

- The study about uncertainty
- Statistics is the exploration of a parameter (theta) in light of data X.

Perbedaan utama antara pendekatan classical dan Bayesian:

	Classical	Bayesian
(1). What's probability?		
(2). Data X		
(3). Parameter θ		
(4). Confidence Interval		

What is probability?

- **Classical**
  - frequency of a certain event when repeating a random experiment infinite times
- **Examples:**
  - 1. probability of heads in a coin toss experiment: defined
  - 2. probability that it will rain Jan 27, 2023: not defined.
  - 3. probability of me getting cancer by 70 if smoking 1 pack cig a day: not defined

- **Bayesian**
  - the quantification of uncertainty –de Finetti
  - this uncertainty can be equated with personal belief of a quantity
- **Examples**
  - 1. belief (can be updated by observation)
  - 2. belief (can be refined by modeling)
  - 3. belief (as above)

Data and Parameter

- **Classical:**
  - Parameter is fixed
  - Data is random
- **Bayesian**
  - Parameter has uncertainty (“random”)
  - Data is given (“fixed” in operational sense)

- Goal of both: draw conclusion/inference for θ

Data and Parameter (cont.)

- **Classical:**
  - Parameter θ is fixed
  - Data X is random

- **Bayesian**
  - Inference should be the same if data are the same: likelihood principle
  - We can talk about distribution of θ, to express uncertainty/belief about θ, whereas classical stats cannot.

Confidence/Credible Interval for θ

- **Classical**
  - Confidence Interval
  - If samples of the same size are drawn repeatedly from a population specified by θ, and a confidence interval is calculated from each sample for θ, then 95% of these intervals should contain θ.

- **Bayesian**
  - Credible Interval
  - Probability of 95% that the credible interval contains θ.

Pendekatan Bayesians Menggunakan Prior

- Prior: non-experimental knowledge.
- Examples from Berger's book:
  - A man tasting tea: tea poured into milk or vice versa.
  - A man predicts coin toss.
  - Each of the 10 times, they got it right.
  - Inference may differ due to different prior beliefs.

Criticisms of Bayesianism

- **Use of prior: not objective; different people come to different conclusions; this is not scientific.**
- **Answers:**
  - Absolute “objectivity” does not exist.
  - This provides a formal way of incorporating subjective belief.
  - To be objective: use noninformative priors.
  - Evidence-based: when data accumulate, all rational individuals will come to the same conclusion.

Bayesian Machinery

data

prior belief ----------------------------------→ posterior belief

likelihood

prior distribution -------------------------→ posterior

$L(\theta)=P(X|\theta)$

$p(\theta)$ --------------------------------------→ $p(\theta|X)$

<u>Ringkasan</u>: perbedaan utama antara pendekatan *classical* (*frequentist*) dan Bayesian

Approach	What does it assume?	What does it ask?	What do you need?	What does it give?	Main advantages	Main drawbacks	When to use it
Frequentist	Parameters that you estimate are fixed and are a single point	Is the hypothesis true or false?	A stopping criterion A predetermined experimental design	A point estimate (*p*-value)	Simple and easy to use Widely accepted	Significance depends on the sample size Only gives you a yes/no answer	When you have a large amount of data
Bayesian	There's a probability distribution around the parameters	What is the probability of the hypothesis given the data?	A prior Any data set	A probability for or against the hypothesis	You get strength of evidence Can update it with new information	Priors can be subjective or arbitrary **Requires advanced statistics**	When have limited data When you have priors **When you have computing power**

Ilustrasi Statistika Bayesian:

- Ini merupakan ilustrasi sederhana untuk menunjukkan dasar-dasar cara kerja statistika Bayesian.
- Dimulai dengan beberapa nilai peluang di awal (*prior*) dan bagaimana persisnya nilai ini diperbarui saat telah mendapatkan lebih banyak informasi.
- Nilai peluang yang diperbarui ini disebut sebagai peluang *posterior*.
- Untuk membantu memperjelas step-by-step, akan digunakan tabel yang biasanya disebut *Bayes' Box* untuk menghitung peluang posterior.

Ilustrasi Statistika Bayesian:

Suppose there are two balls in a bag. We know in advance that at least one of them is black, but we're not sure whether they're both black, or whether one is black and one is white. These are the only two possibilities we will consider. To keep things concise, we can label our two competing hypotheses. We could call them whatever we want, but I will call them `BB` and `BW`. So, at the beginning of the problem, we know that *one and only one* of the following statements/hypotheses is true:

`BB`: Both balls are black `BW`: One ball is black and the other is white.

Ilustrasi Statistika Bayesian:

BB: Both balls are black
BW: One ball is black and the other is white.

Suppose an experiment is performed to help us determine which of these two hypotheses is true. The experimenter reaches into the bag, pulls out one of the balls, and observes its colour. The result of this experiment is (drumroll please!):

D: The ball that was removed from the bag was black.

We will now do a Bayesian analysis of this result.

Ilustrasi Statistika Bayesian:

Bayes's Box

A Bayesian analysis starts by choosing some values for the prior probabilities. We have our two competing hypotheses BB and BW, and we need to choose some probability values to describe how sure we are that each of these is true. Since we are talking about two hypotheses, there will be two prior probabilities, one for BB and one for BW. For simplicity, we will assume that we don't have much of an idea which is true, and so we will use the following prior probabilities:

$$\begin{aligned} P(\texttt{BB}) &= 0.5 \\ P(\texttt{BW}) &= 0.5. \end{aligned}$$

The first column of a Bayes' Box is just the list of hypotheses we are considering.

Hypotheses	prior	likelihood	prior $\times$ likelihood	posterior
BB	0.5			
BW	0.5			
Totals:	1			

Ilustrasi Statistika Bayesian:

D: The ball that was removed from the bag was black.

The likelihood for a hypothesis is the probability that you would have observed the data, if that hypothesis were true. The values can be found by going through each hypothesis in turn, imagining it is true, and asking, "What is the probability of getting the data that I observed?".

Hypotheses	Possible Data	Probability
BB	Black Ball	1
	White Ball	0
BW	Black Ball	0.5
	White Ball	0.5

This table demonstrates a method for calculating the likelihood values, by considering not just the data that actually occurred, but all data that might have occurred. Ultimately, it is only the probability of the data which actually occurred that matters, so this is highlighted in blue.

Ilustrasi Statistika Bayesian:

To find the posterior probabilities, we take the `prior` $\times$ `likelihood` column and divide it by its sum, producing numbers that do sum to 1. This gives us the final posterior probabilities, which were the goal all along. The completed Bayes' Box is shown below:

Hypotheses	prior	likelihood	h = prior $\times$ likelihood	posterior
BB	0.5	1	0.5	0.667
BW	0.5	0.5	0.25	0.333
Totals:	1		0.75	1

We can see that the posterior probabilities are not the same as the prior probabilities, because we have more information now! The experimental result made BB a little bit more plausible than it was before. Its probability has increased from $1/2$ to $2/3$.

Materi Latihan

Selesaikan problem ini melalui Bayes' Box untuk memperoleh posterior: (a).Secara manual; (b).Menggunakan Program R.

You move into a new house which has a phone installed. You can't remember the phone number, but you suspect it might be 555-3226 (some of you may recognise this as being the phone number for Homer Simpson's "Mr Plow" business). To test this hypothesis, you carry out an experiment by picking up the phone and dialing 555-3226.

If you are correct about the phone number, you will definitely hear a busy signal because you are calling yourself. If you are incorrect, the probability of hearing a busy signal is 1/100. However, all of that is only true if you assume the phone is working, and it might be broken! If the phone is broken, it will always give a busy signal.

When you do the experiment, the outcome (the data) is that you do actually get the busy signal. The question asked us to consider the following four hypotheses, and to calculate their posterior probabilities:

Hypothesis	Description	Prior Probability
H_1	Phone is working and 555-3226 is correct	0.4
H_2	Phone is working and 555-3226 is incorrect	0.4
H_3	Phone is broken and 555-3226 is correct	0.1
H_4	Phone is broken and 555-3226 is incorrect	0.1

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 3th Edition*. John Wiley and Sons, Inc.

3. Lee, Peter M. (2012). *Bayesian Statistics, An Introduction 4th 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

IPB University
Inspiring Innovation with Integrity

Pengantar Statistika Bayes - STA1312

Konsep Inferensi Bayesian

(Part 1: Bayes' Rule dan Bayes' Box)

Dr. Kusman Sadik, S.Si, M.Si
Program Studi Statistika dan Sains Data IPB
Tahun Akademik 2022/2023

Teorema Bayes

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \begin{cases} P(A \cap B) = P(B) \times P(A|B) \\ P(A \cap \tilde{B}) = P(\tilde{B}) \times P(A|\tilde{B}) \\ P(A) = P(A \cap B) + P(A \cap \tilde{B}) \end{cases}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A \cap B) + P(A \cap \tilde{B})}$$

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|\tilde{B}) \times P(\tilde{B})}$$

Fakta hubungan ini dapat digunakan untuk menghitung P(B|A) berdasarkan P(A|B)

Teorema Bayes (*Cont.*)

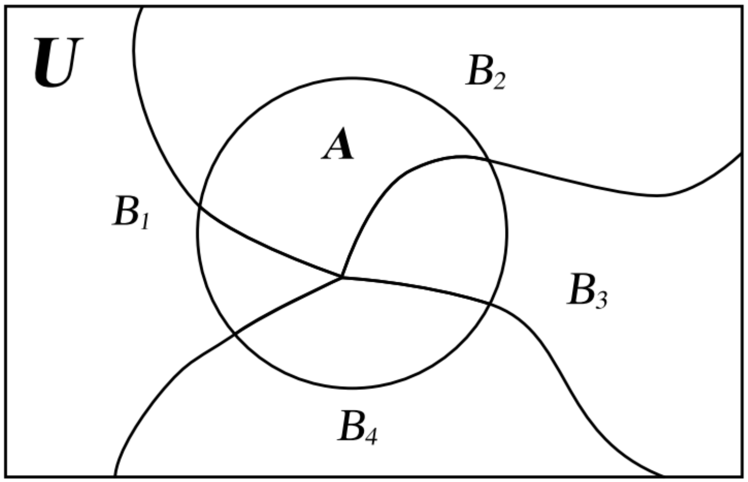
A set of events partitioning the universe. Often we have a set of more than two events that partition the universe. For example, suppose we have n events $B_1, \cdots, B_n$ such that:

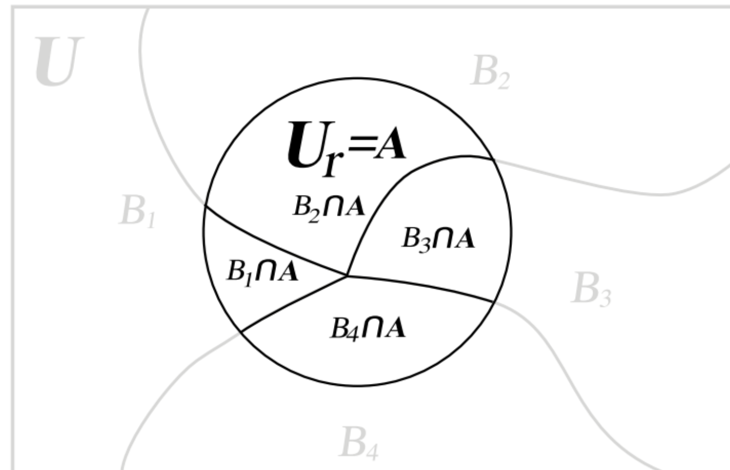
- The union $B_1 \cup B_2 \cup \cdots \cup B_n = U$, the universe, and
- Every distinct pair of the events are disjoint, $B_i \cap B_j = \phi$ for $i = 1, \ldots, n$, $j = 1, \ldots, n$, and $i \neq j$.

Then we say the set of events $B_1, \cdots, B_n$ *partitions* the universe. An observable event A will be partitioned into parts by the partition. $A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots (A \cap B_n)$. $(A \cap B_i)$ and $(A \cap B_j)$ are disjoint since B_i and B_j are disjoint. Hence

$$P(A) = \sum_{j=1}^{n} P(A \cap B_j)\,.$$

Teorema Bayes (*Cont.*)





$$P(A) = \sum_{j=1}^{n} P(A \cap B_j)$$

$$P(A) = \sum_{j=1}^{n} P(A|B_j) \times P(B_j)$$

Teorema Bayes (*Cont.*)

$$P(A) = \sum_{j=1}^{n} P(A \cap B_j)$$

$$P(A) = \sum_{j=1}^{n} P(A|B_j) \times P(B_j)$$

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)}$$

Bayes' Rule :

$$\boxed{P(B_i|A) = \frac{P(A|B_i) \times P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \times P(B_j)}}$$

Studi Kasus 1:

- The chance of a certain medical test being positive is 90%, if a patient has disease D. Known 1% of the population have the disease, and the test records a false positive 5% of the time. If you receive a positive test, what is your probability of having D?

- We are told: $P(+|D) = 0.90$, $P(D) = 0.01$, $P(+|\widetilde{D}) = 0.05$

- We want to know: $P(D|+)$

- Bayes' Theorem: $P(D|+) = \frac{P(+|D)\,P(D)}{P(+)} = \frac{P(+|D)\,P(D)}{[P(+|D)\,P(D)]+[P(+|\widetilde{D})\,P(\widetilde{D})]}$

- Substituting in the data: $P(D|+) = \frac{0.9\times0.01}{(0.9\times0.01)+(0.05\times0.99)} = 0.15$

- *Interpretation: although the test is correct 90% of the time, the probability of having D after a positive test is only 15%. This is because only a small fraction of the population have the disease.*

$$P(B_i|A) = \frac{P(A|B_i) \times P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \times P(B_j)}$$

Bayes' Theorem: The Key to Bayesian Statistics

To see how we can use Bayes' theorem to revise our beliefs on the basis of evidence, we need to look at each part. Let $B_1, \ldots, B_n$ be a set of unobservable events which partition the universe. We start with $P(B_i)$ for $i = 1, \ldots, n$, the *prior* probability for the events B_i, for $i = 1, \ldots, n$. This distribution gives the weight we attach to each of the B_i from our prior belief. Then we find that A has occurred.

The *likelihood* of the unobservable events $B_1, \ldots, B_n$ is the conditional probability that A has occurred given B_i for $i = 1, \ldots, n$. Thus the *likelihood* of event B_i is given by $P(A|B_i)$. We see the *likelihood* is a function defined on the events $B_1, \ldots, B_n$. The *likelihood* is the weight given to each of the B_i events given by the occurrence of A.

$P(B_i|A)$ for $i = 1, \ldots, n$ is the *posterior* probability of event B_i, given that event A has occurred. This distribution contains the weight we attach to each of the events B_i for $i = 1, \ldots n$ after we know event A has occurred. It combines our prior beliefs with the evidence given by the occurrence of event A.

Prior, Posterior, Likelihood, dan Marginal Likelihood

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$

In Bayesian statistics, most of the terms in Bayes' rule have special names. Some of them even have more than one name, with different scientific communities preferring different terminology. Here is a list of the various terms and the names we will use for them:

- $P(H|D)$ is the **posterior probability**. It describes how certain or confident we are that hypothesis H is true, given that we have observed data D. Calculating posterior probabilities is the main goal of Bayesian statistics!
- $P(H)$ is the **prior probability**, which describes how sure we were that H was true, before we observed the data D.
- $P(D|H)$ is the **likelihood**. If you were to assume that H is true, this is the probability that you would have observed data D.
- $P(D)$ is the **marginal likelihood**. This is the probability that you would have observed data D, *whether H is true or not.*

We often write Bayes' theorem in its proportional form as

$$posterior \propto prior \times likelihood$$

Studi Kasus 2:

(Teorema Bayes untuk Binomial dengan Prior Diskret)

Let $Y|\pi$ be $binomial(n = 4, \pi)$. Suppose we consider that there are only three possible values for π, .4,.5, and .6. We will assume they are equally likely. The joint probability distribution $f(\pi_i, y_j)$ is found by multiplying the conditional observation distribution $f(y_j|\pi_i)$ times the prior distribution $g(\pi_i)$. In this case, the conditional observation probabilities come from the $binomial(n = 4, \pi)$ distribution.

Suppose $Y = 3$ was observed.

Bayes' Box

π	*prior*	*likelihood*	*prior* $\times$ *likelihood*	*posterior*
.4	$\frac{1}{3}$	.1536	.0512	$\frac{.0512}{.2497} = .205$
.5	$\frac{1}{3}$	.2500	.0833	$\frac{.0833}{.2497} = .334$
.6	$\frac{1}{3}$	.3456	.1152	$\frac{.1152}{.2497} = .461$
marginal $P(Y = 3)$			.2497	1.000

Bayes' Box

π	*prior*	*likelihood*	*prior* × *likelihood*	*posterior*
.4	$\frac{1}{3}$	.1536	.0512	$\frac{.0512}{.2497} = .205$
.5	$\frac{1}{3}$	.2500	.0833	$\frac{.0833}{.2497} = .334$
.6	$\frac{1}{3}$	.3456	.1152	$\frac{.1152}{.2497} = .461$
marginal $P(Y = 3)$			.2497	1.000

- Put in the *parameter values*, the *prior*, and the *likelihood* in their respective columns. The *likelihood* values are *binomial(n,*π_i*)* evaluated at the observed value of y.

- Multiply each element in the *prior* column by the corresponding element in the *likelihood* column and put in the *prior* × *likelihood* column.

- Sum these *prior* × *likelihood.*

- Divide each element of *prior* × *likelihood* column by the sum of *prior* × *likelihood* column. (This rescales them to sum to 1.)

- Put these in the *posterior* column!

Studi Kasus 3:

(Teorema Bayes untuk Poisson dengan Prior Diskret)

Let $Y|\mu$ be Poisson(μ). Suppose that we believe there are only four possible values for μ, 1,1.5,2, and 2.5. Suppose we consider that the two middle values, 1.5 and 2, are twice as likely as the two end values 1 and 2.5. Suppose $y = 2$ was observed. Plug the value $y = 2$ into formula

$$f(y|\mu) = \frac{\mu^y e^{-\mu}}{y!}$$

to give the likelihood.

Bayes' Box

μ	*prior*	*likelihood*	*prior* $\times$ *likelihood*	*posterior*
1.0	$\frac{1}{6}$	$\frac{1.0^2 e^{-1.0}}{2!} = .1839$	.0307	$\frac{.0307}{.2473} = .124$
1.5	$\frac{1}{3}$	$\frac{1.5^2 e^{-1.5}}{2!} = .2510$	.0837	$\frac{.0837}{.2473} = .338$
2.0	$\frac{1}{3}$	$\frac{2.0^2 e^{-2.0}}{2!} = .2707$	.0902	$\frac{.0902}{.2473} = .365$
2.5	$\frac{1}{6}$	$\frac{2.5^2 e^{-2.5}}{2!} = .2565$	.0428	$\frac{.0428}{.2473} = .173$
marginal $P(Y=2)$			.2473	1.000

Materi Latihan

1

Suppose there is a medical diagnostic test for a disease. The *sensitivity* of the test is .95. This means that if a person has the disease, the probability that the test gives a positive response is .95. The *specificity* of the test is .90. This means that if a person does not have the disease, the probability that the test gives a negative response is .90, or that the *false positive* rate of the test is .10. In the population, 1% of the people have the disease. What is the probability that a person tested has the disease, given the results of the test is positive? Let D be the event "the person has the disease" and let T be the event "the test gives a positive result."

2

Suppose there is a medical screening procedure for a specific cancer that has *sensitivity* = .90, and *specificity* = .95. Suppose the underlying rate of the cancer in the population is .001. Let B be the event "the person has that specific cancer," and let A be the event "the screening procedure gives a positive result."

(a) What is the probability that a person has the disease given the results of the screening is positive?

(b) Does this show that screening is effective in detecting this cancer?

3

There is an urn containing 9 balls, which can be either green or red. The number of red balls in the urn is not known.

Let X be the number of red balls in the urn.

Suppose we look at the two draws from the urn (without replacement) as a single experiment. The results were first draw red, second draw green. Find the posterior distribution of X by filling in the simplified table.

X	*prior*	*likelihood*	*prior* $\times$ *likelihood*	*posterior*

4

Selesaikan problem ini melalui Bayes' Box untuk memperoleh posterior: (a).Secara manual; (b).Menggunakan Program R.

Let Y_1 be the number of successes in $n = 10$ independent trials where each trial results in a success or failure, and π, the probability of success, remains constant over all trials. Suppose the 4 possible values of π are .20, .40, .60, and .80. We do not wish to favor any value over the others so we make them equally likely. We observe $Y_1 = 7$. Find the posterior distribution by filling in the simplified table.

π	*prior*	*likelihood*	*prior* $\times$ *likelihood*	*posterior*

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 3th Edition*. John Wiley and Sons, Inc.

3. Lee, Peter M. (2012). *Bayesian Statistics, An Introduction 4th 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

IPB University
Inspiring Innovation with Integrity

Pengantar Statistika Bayes - STA1312

Konsep Inferensi Bayesian – Part 2

(Conjugate Prior dan Non-informative Prior)

Dr. Kusman Sadik, S.Si, M.Si
Program Studi Statistika dan Sains Data IPB
Tahun Akademik 2022/2023

Metode Bayesian Secara Analitik

- Sebelumnya telah dibahas penentuan peluang posterior berdasarkan Bayes' Box. Namun hal ini terbatas hanya untuk prior diskret.
- Penentuan peluang posterior yang bersifat general dapat dilakukan secara analitik, yakni melalui proses fungsi sebaran.
- Pada metode Bayes, parameter θ dianggap sebagai peubah acak dengan suatu sebaran peluang tertentu yang disebut sebagai sebaran prior bagi θ.
- Beradasarkan data Y_1, Y_2, …, Y_n, sebaran prior tersebut kemudian diperbaiki (di-*update*) menjadi sebaran posterior bagi θ.
- Selanjutnya inferensi (pendugaan parameter dan pengujian hipotesis) didasarkan pada sebaran posterior ini.

If we denote the prior distribution by $\pi(\theta)$ and the sampling distribution by $f(\mathbf{x}|\theta)$, then the posterior distribution, the conditional distribution of θ given the sample, $\mathbf{x}$, is

$$\pi(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)\pi(\theta)/m(\mathbf{x}), \qquad (f(\mathbf{x}|\theta)\pi(\theta) = f(\mathbf{x},\theta))$$

where $m(\mathbf{x})$ is the marginal distribution of $\mathbf{X}$, that is,

$$m(\mathbf{x}) = \int f(\mathbf{x}|\theta)\pi(\theta)d\theta.$$

Notice that the posterior distribution is a conditional distribution, conditional upon observing the sample. The posterior distribution is now used to make statements about θ, which is still considered a random quantity. For instance, the mean of the posterior distribution can be used as a point estimate of θ.

Komponen Metode Bayes Secara Analitik

1. Fungsi kepekatan peluang peubah acak X: $f(\boldsymbol{x}|\theta)$

2. Sebaran prior bagi θ: $\pi(\theta)$

3. Sebaran posterior bagi θ: $\pi(\theta|\boldsymbol{x})$

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x},\theta)}{m(\boldsymbol{x})} = \frac{\{f(\boldsymbol{x}|\theta)\}\{\pi(\theta)\}}{\int_{-\infty}^{\infty} f(\boldsymbol{x},\theta)d\theta} = \frac{\{f(\boldsymbol{x}|\theta)\}\{\pi(\theta)\}}{\int_{-\infty}^{\infty}\{f(\boldsymbol{x}|\theta)\}\{\pi(\theta)\}d\theta}$$

4. Inferensi didasarkan pada sebaran poterior $\pi(\theta|\boldsymbol{x})$ misalnya penduga Bayes bagi θ:

$$\hat{\theta}_B = E(\theta|\boldsymbol{x}) = \int_{-\infty}^{\infty} \theta\{\pi(\theta|\boldsymbol{x})\}\, d\theta$$

Komponen Metode Bayes Secara Analitik (Cont.)

When θ can get values continuosly on some interval, we can express our beliefs about it with a *prior density* $p(\theta)$. After we have obtained the data y, our beliefs about θ are contained in the conditional density,

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta}, \tag{1}$$

called *posterior density*.

Since θ is integrated out in the denominator, it can be considered as a constant with respect to θ. Therefore, the Bayes' formula in (1) is often written as

$$p(\theta|y) \propto p(\theta)p(y|\theta), \tag{2}$$

which denotes that $p(\theta|y)$ is proportional to $p(\theta)p(y|\theta)$.

Komponen Metode Bayes Secara Analitik (Cont.)

If we view $\boldsymbol{\theta}$ as a random variable, we can apply Bayes' tule to obtain the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\theta} \propto f(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

The posterior is proportional to the likelihood times the prior, and quantifies the uncertainty about the parameters that remain after accounting for prior knowledge and the new information in the observed data.

Prior density of $\boldsymbol{\theta}$:	$\pi(\boldsymbol{\theta})$
Likelihood function of $\mathbf{Y}$ given $\boldsymbol{\theta}$:	$f(\mathbf{Y}\|\boldsymbol{\theta})$
Marginal density of $\mathbf{Y}$:	$m(\mathbf{Y}) = \int f(\mathbf{Y}\|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$
Posterior density of $\boldsymbol{\theta}$ given $\mathbf{Y}$:	$p(\boldsymbol{\theta}\|\mathbf{Y}) = f(\mathbf{Y}\|\boldsymbol{\theta})\pi(\boldsymbol{\theta})/m(\mathbf{Y})$

Contoh Kasus (1):

A drug company would like to introduce a drug to reduce acid indigestion. It is desirable to estimate θ, the proportion of the market share that this drug will capture. The company interviews n people and Y of them say that they will buy the drug. In the non-Bayesian analysis $\theta \in [0, 1]$ and $Y \sim \text{Bin}(n, \theta)$.

We know that $\hat{\theta} = Y/n$ is a very good estimator of θ. It is unbiased, consistent and minimum variance unbiased. Moreover, it is also the maximum likelihood estimator (MLE), and thus asymptotically normal.

A Bayesian may look at the past performance of new drugs of this type. If in the past new drugs tend to capture a proportion between say .05 and .15 of the market, and if all values in between are assumed equally likely, then $\theta \sim \text{Unif}(.05, .15)$.

Thus, the prior distribution is given by

$$p(\theta) = \begin{cases} 1/(0.15 - 0.05) = 10, & 0.05 \le \theta \le 0.15 \\ 0, & \text{otherwise.} \end{cases}$$

and the likelihood function by

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

The posterior distribution is

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} = \begin{cases} \frac{\theta^y(1-\theta)^{n-y}}{\int_{0.05}^{0.15} \theta^y(1-\theta)^{n-y}d\theta} & 0.05 \le \theta \le 0.15 \\ 0, & \text{otherwise.} \end{cases}$$

Conjugate Prior

(1). Conjugate Prior: Beta-Binomial

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*.

Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$ (indexed by θ). A class $\prod$ of prior distributions is a *conjugate family* for $\mathcal{F}$ if the posterior distribution is in the class $\prod$ for all $f \in \mathcal{F}$, all priors in $\prod$, and all $x \in \mathcal{X}$.

The beta family is conjugate for the binomial family. Thus, if we start with a beta prior, we will end up with a beta posterior. The updating of the prior takes the form of updating its parameters. Mathematically, this is very convenient, for it usually makes calculation quite easy. Whether or not a conjugate family is a reasonable choice for a particular problem, however, is a question to be left to the experimenter.

Contoh Kasus (2):

(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

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

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,

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

Catatan:

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

(2). Conjugate Prior: Poisson-Gamma

The number of diseased trees per acre can be modeled by a Poisson distribution with mean θ. Since θ changes from area to area, the forester believes that $\theta \sim \text{Exp}(\lambda)$. Thus,

$$p(\theta) = (1/\lambda)e^{-\theta/\lambda}, \quad \text{if } \theta > 0, \text{and } 0 \text{ elsewhere}$$

The forester takes a random sample of size n from n different one-acre plots.

Note: Exp(λ) = Gamma(1, 1/ λ)

Note: Exp(λ) = Gamma(1, 1/ λ)

$$X \sim \Gamma(\alpha, \beta) \equiv \mathrm{Gamma}(\alpha, \beta)$$

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\beta x} \beta^{\alpha}}{\Gamma(\alpha)} \quad \text{for } x > 0 \quad \alpha, \beta > 0,$$

The number of diseased trees per acre can be modeled by a Poisson distribution with mean θ. Since θ changes from area to area, the forester believes that $\theta \sim \text{Exp}(\lambda)$. Thus,

Let $y = (y_1, ...y_n)$ be a sample from $\text{Poi}(\theta)$. Then the likelihood is

$$p(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \propto \theta^{\sum y_i} e^{-n\theta}.$$

prior $\longrightarrow$ $p(\theta) = (1/\lambda)e^{-\theta/\lambda}, \quad \text{if } \theta > 0, \text{and } 0 \text{ elsewhere}$

The likelihood function is

$$p(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!} e^{-\theta} = \frac{\theta^{\sum_{i=1}^{n} y_i}}{\prod y_i!} e^{-n\theta}.$$

Consequently, the posterior distribution is

posterior $\rightarrow$ $$p(\theta|y) = \frac{\theta^{\sum_{i=1}^{n} y_i} e^{-\theta(n+1/\lambda)}}{\int_0^\infty \theta^{\sum_{i=1}^{n} y_i} e^{-\theta(n+1/\lambda)}}.$$

The likelihood function is

$$p(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!} e^{-\theta} = \frac{\theta^{\sum_{i=1}^{n} y_i}}{\prod y_i!} e^{-n\theta}.$$

Consequently, the posterior distribution is

$$p(\theta|y) = \frac{\theta^{\sum_{i=1}^{n} y_i} e^{-\theta(n+1/\lambda)}}{\int_0^\infty \theta^{\sum_{i=1}^{n} y_i} e^{-\theta(n+1/\lambda)}}.$$

We see that this is a Gamma-distribution with parameters $\alpha = \sum_{i=1}^{n} y_i + 1$ and $\beta = n + 1/\lambda$. Thus,

posterior $\longrightarrow$ $$p(\theta|y) = \frac{(n+1/\lambda)^{\sum_{i=1}^{n} y_i + 1}}{\Gamma(\sum_{i=1}^{n} y_i + 1)} \theta^{\sum_{i=1}^{n} y_i} e^{-\theta(n+1/\lambda)}.$$

(3). Conjugate Prior: Normal-Normal

(Normal Bayes estimators) Let $X \sim \mathrm{n}(\theta, \sigma^2)$, and suppose that the prior distribution on θ is $\mathrm{n}(\mu, \tau^2)$. (Here we assume that σ^2, μ, and τ^2 are all known.) The posterior distribution of θ is also normal, with mean and variance given by

$$\mathrm{E}(\theta|x) = \frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\sigma^2+\tau^2}\mu,$$

$$\mathrm{Var}(\theta|x) = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}.$$

Penjabaran secara lengkap disediakan sebagai latihan !

Ringkasan Beberapa Conjugate Prior

$f(x\|\theta)$	$\pi(\theta)$	$\pi(\theta\|x)$
Normal $\mathcal{N}(\theta,\sigma^2)$	Normal $\mathcal{N}(\mu,\tau^2)$	$\mathcal{N}(\varrho(\sigma^2\mu+\tau^2 x),\varrho\sigma^2\tau^2)$ $\varrho^{-1}=\sigma^2+\tau^2$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha,\beta)$	$\mathcal{G}(\alpha+x,\beta+1)$
Gamma $\mathcal{G}(\nu,\theta)$	Gamma $\mathcal{G}(\alpha,\beta)$	$\mathcal{G}(\alpha+\nu,\beta+x)$
Binomial $\mathcal{B}(n,\theta)$	Beta $\mathcal{B}e(\alpha,\beta)$	$\mathcal{B}e(\alpha+x,\beta+n-x)$
Negative Binomial $\mathcal{N}eg(m,\theta)$	Beta $\mathcal{B}e(\alpha,\beta)$	$\mathcal{B}e(\alpha+m,\beta+x)$
Multinomial $\mathcal{M}_k(\theta_1,\ldots,\theta_k)$	Dirichlet $\mathcal{D}(\alpha_1,\ldots,\alpha_k)$	$\mathcal{D}(\alpha_1+x_1,\ldots,\alpha_k+x_k)$
Normal $\mathcal{N}(\mu,1/\theta)$	Gamma $\mathcal{G}a(\alpha,\beta)$	$\mathcal{G}(\alpha+0.5,\beta+(\mu-x)^2/2)$

Non-informative Prior

Non-informative Prior

- Sebelumnya telah dijabarkan bahwa *conjugate prior* berguna sebagai pendekatan bagi sebaran prior yang menghasilkan sebaran posterior yang bersifat *closed-form*.
- Namun apabila tidak ada pengetahuan sama sekali tentang nilai parameter θ, maka *non-informative prior* dapat digunakan.
- Pada beberapa literatur, *non-informative prior* ini seringkali disebut juga sebagai *objective prior*.
- Ada banyak metode pendekatan untuk memperoleh *non-informative prior.* Dua metode yang paling banyak digunakan adalah metode yang dikembangkan oleh Laplace dan Harold Jeffreys.
- *Non-informative prior* yang diperoleh melalui kedua metode tersebut kemudian dikenal dengan nama *Laplace's prior* dan *Jeffreys' prior*.

Laplace's Prior

- *Laplace's prior* dikenal juga sebagai sebaran eksponensial ganda (*double exponential distibution*), yaitu:

$$\pi(\theta) = \frac{1}{2\sigma} exp\left(-\frac{|\theta - \mu|}{\sigma}\right); \; -\infty < \mu < \infty; \;\; \sigma > 0$$

- Jadi *Laplace's prior* merupakan sebaran dengan dua parameter yang simetris di sekitar nilai tengahnya (μ) dan memiliki ekor sebaran yang bersifat eksponensial.

- Artinya nilai yang sangat besar atau sangat kecil berkurang secara eksponensial saat nilai menjauh dari nilai tengahnya (μ).

- *Laplace's prior* sering digunakan dalam model linear (regresi dll), karena memberikan prior yang fleksibel.

Jeffreys' Prior

A Jeffreys' prior (JP) is a method to construct a prior distribution that is invariant to reparameterization. This is a necessary condition for a method to be objective. For example, if one statistician places a prior on the standard deviation (σ) and another places a prior on the variance (σ^2) and the two get different results, then the method is subjective because it depends on the choice of parameterization.

The univariate JP for θ is

$$\pi(\theta) \propto \sqrt{I(\theta)}$$

where $I(\theta)$ is the expected Fisher information defined as

$$I(\theta) = -\mathrm{E}\left(\frac{d^2 \log f(\mathbf{Y}|\theta)}{d\theta^2}\right)$$

and the expectation is with respect to $\mathbf{Y}|\theta$.

Jeffreys' Prior (Multivariate)

JP can also be applied in multivariate problems. The multivariate JP for $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ is

$$\pi(\boldsymbol{\theta}) \propto \sqrt{|\mathbf{I}(\boldsymbol{\theta})|}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the $p \times p$ expected Fisher information matrix with (i, j) element

$$-\mathrm{E}\left(\frac{\partial^2 \log f(\mathbf{Y}|\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right).$$

Contoh: Jeffreys' Prior untuk Binomial

Binomial proportion: As a univariate example, consider the binomial model $Y \sim \text{Binomial}(n, \theta)$. The second derivative of the binomial log likelihood is

$$\begin{aligned}\frac{d^2 \log f(\mathbf{Y}|\theta)}{d\theta^2} &= \frac{d^2}{d\theta^2} \log \binom{n}{Y} + Y \log(\theta) + (n - Y) \log(1 - \theta) \\ &= \frac{d}{d\theta} \frac{Y}{\theta} - \frac{n - Y}{1 - \theta} \\ &= -\frac{Y}{\theta^2} - \frac{n - Y}{(1 - \theta)^2}.\end{aligned}$$

The information is the expected value of the negative second derivative. Under the binomial model, the expected value of Y is $n\theta$ and since the second derivative is linear in Y the expectation passes into each term as

$$I(\theta) = \frac{\mathrm{E}(Y)}{\theta^2} + \frac{n - \mathrm{E}(Y)}{(1 - \theta)^2} = \frac{n}{\theta} + \frac{n}{1 - \theta} = \frac{n}{\theta(1 - \theta)}.$$

Contoh: Jeffreys' Prior untuk Binomial

$$I(\theta) = \frac{\mathrm{E}(Y)}{\theta^2} + \frac{n - \mathrm{E}(Y)}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}.$$

The JP is then

$$\pi(\theta) \propto \sqrt{\frac{n}{\theta(1-\theta)}} \propto \theta^{1/2-1}(1-\theta)^{1/2-1},$$

which is the kernel of the Beta(1/2,1/2) PDF. Therefore, the JP for the binomial proportion is $\theta \sim \mathrm{Beta}(1/2, 1/2)$.

Materi Latihan

1

Sophie, the editor of the student newspaper, is going to conduct a survey of students to determine the level of support for the current president of the students' association. She needs to determine her prior distribution for π, the proportion of students who support the president. She decides her prior mean is .5, and her prior standard deviation is .15.

(a) Determine the $beta(a, b)$ prior that matches her prior belief.

(b) What is the equivalent sample size of her prior?

(c) Out of the 68 students that she polls, $y = 21$ support the current president. Determine her posterior distribution.

2

Say $Y|\lambda \sim \text{Poisson}(\lambda)$.

(a) Derive and plot the Jeffreys' prior for λ.

(b) Is this prior proper?

(c) Derive the posterior and give conditions on Y to ensure it is proper.

3

We will use the Minitab macro *PoisGamP*, or `poisgamp` function in R, to find the posterior distribution of the Poisson probability μ when we have a random sample of observations from a $Poisson(\mu)$ distribution and we have a $gamma(r, v)$ prior for μ. The *gamma* family of priors is the conjugate family for *Poisson* observations. That means that if we start with one member of the family as the prior distribution, we will get another member of the family as the posterior distribution. The simple updating rules are "add sum of observations to r" and "add sample size to v. When we start with a $gamma(r, v)$ prior, we get a $gamma(r', v')$ posterior where $r' = r + \sum(y)$ and $v' = v + n$.

Suppose we have a random sample of five observations from a $Poisson(\mu)$ distribution. They are:

3	4	3	0	1

(a) Suppose we start with a positive uniform prior for μ. What $gamma(r, v)$ prior will give this form?

(b) **[Minitab:]** Find the posterior distribution using the Minitab macro *PoisGamP* or the R function `poisgamp`.

[R:] Find the posterior distribution using the R function `poisgamp`.

(c) Find the posterior mean and median.

Catatan: Poin *b* dan *c* silakan gunakan **Program R** melalui metode **Monte Carlo** (Lihat buku Bolstad dan Curran (2017), hlm. 208).

4

Suppose we start with a Jeffreys' prior for the Poisson parameter μ.

$$g(\mu) = \mu^{-\frac{1}{2}}$$

(a) What $gamma(r, v)$ prior will give this form?

(b) Find the posterior distribution using the macro $PoisGamP$ in Minitab or or the function `poisgamp` in R.

(c) Find the posterior mean and median.

Catatan: Poin *b* dan *c* silakan gunakan **Program R** melalui metode **Monte Carlo** (Lihat buku Bolstad dan Curran (2017), hlm. 209).

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 3th Edition*. John Wiley and Sons, Inc.

3. Lee, Peter M. (2012). *Bayesian Statistics, An Introduction 4th 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

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 2022/2023

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*

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

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:

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

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

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

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

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

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

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.)

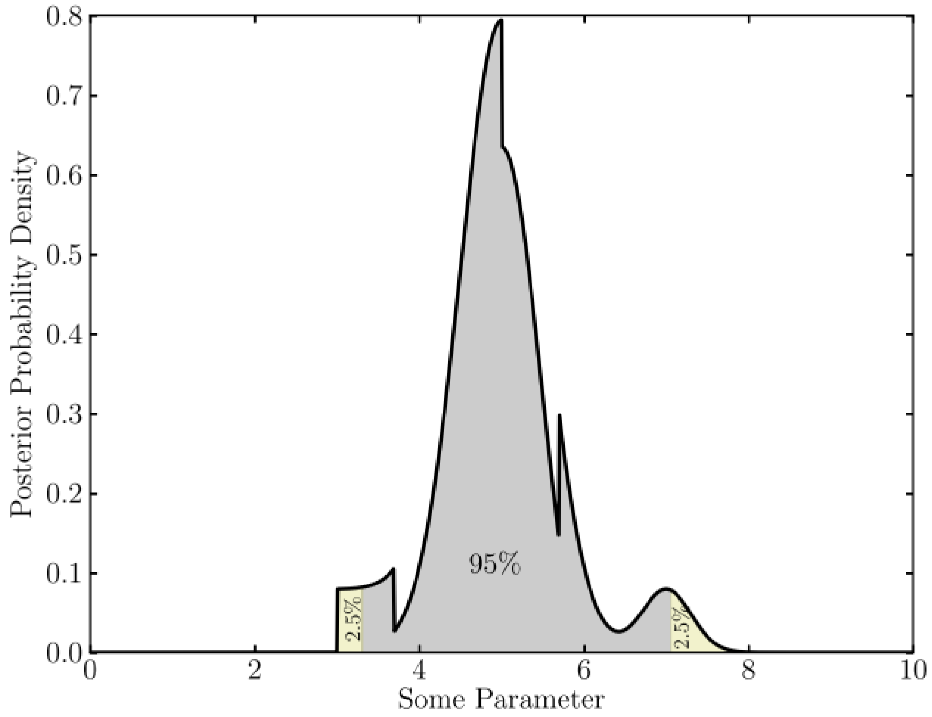
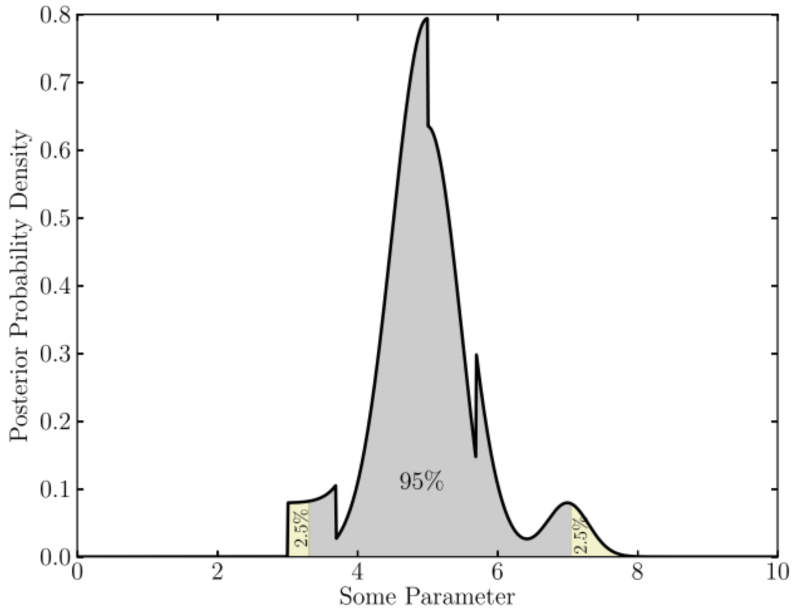


Figure 1. *A central 95% credible interval is defined as an interval that contains 95% of the posterior probability, while having 2.5% of the probability above the upper limit and 2.5% of the probability below the lower limit. The credible interval is formed by finding the edges of the grey region. In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

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

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

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

Equivalently, we choose H_0 if and only if

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

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 Latihan

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 3th Edition*. John Wiley and Sons, Inc.

3. Lee, Peter M. (2012). *Bayesian Statistics, An Introduction 4th 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

IPB University
Inspiring Innovation with Integrity

Pengantar Statistika Bayes - STA1312

Metode Komputasi Bayesian - Part 1

(MCMC, Algoritma Metropolis–Hastings, dan Gibbs Sampling)

Dr. Kusman Sadik, S.Si, M.Si
Program Studi Statistika dan Sains Data IPB
Tahun Akademik 2022/2023

Markov Chain Monte Carlo

Konsep Dasar Markov Chain Monte Carlo

- Pengembangan metode Markov Chain Monte Carlo (**MCMC**) telah menjadi hal yang sangat penting untuk statistika Bayes.
- Metode ini memungkinkan kita untuk **menarik sampel** dari sebaran posterior $g(\theta|y)$ yang eksak, berdasarkan bentuk proporsionalnya, yaitu $g(\theta).f(y|\theta)$ yang berupa perkalian prior dengan likelihood.
- Melalui MCMC, inferensi Bayesian didasarkan pada **sampel dari posterior** ini, bukan pada posterior yang eksak.
- Metode MCMC ini dapat pula digunakan untuk model yang rumit yang memiliki banyak parameter.
- Dua metode utama pada MCMC adalah: (i).Algoritma **Metropolis–Hastings**, dan (ii).Algoritma **Gibbs Sampling**.

Markov Chain

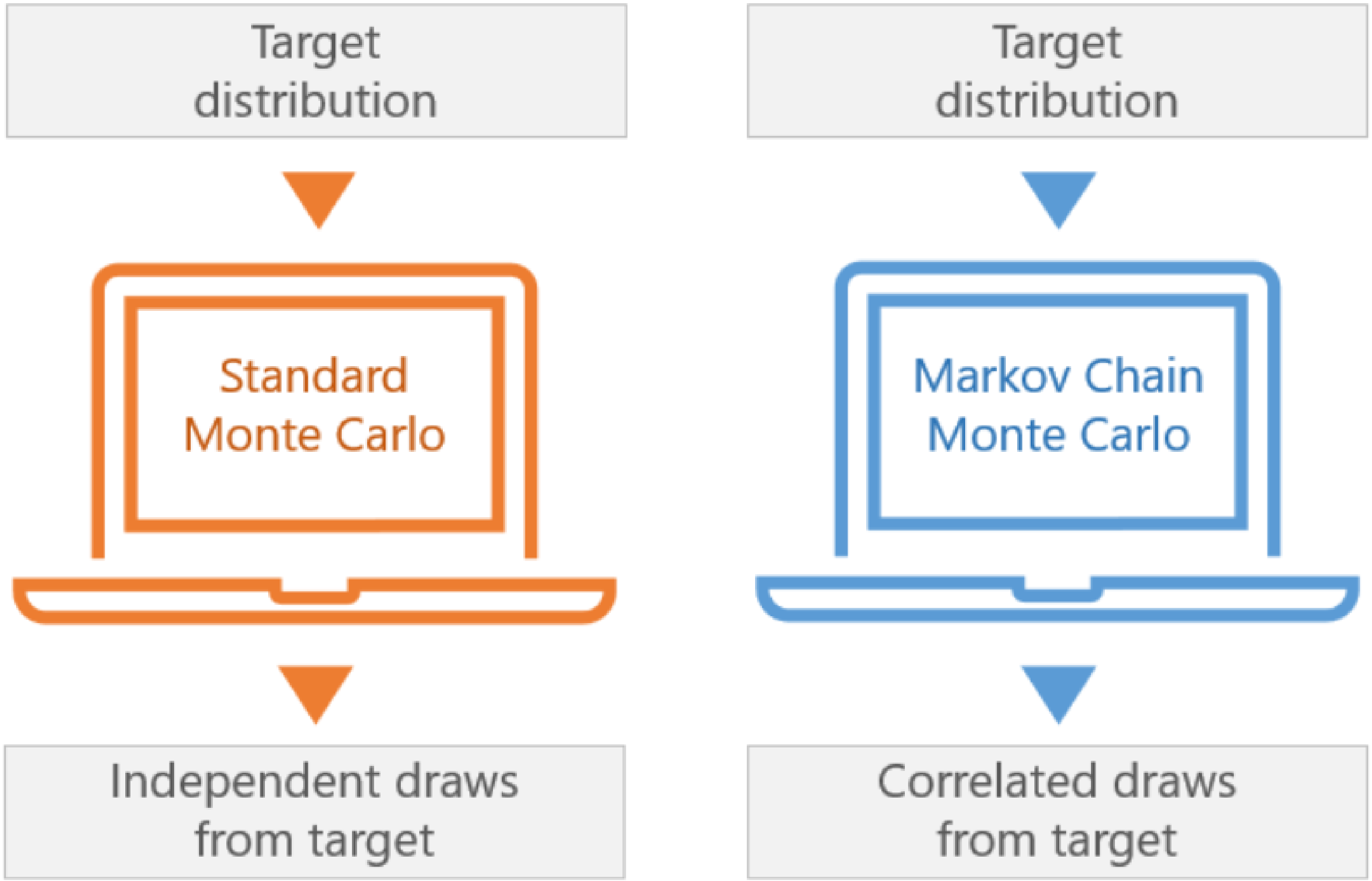
- Rantai Markov (Markov Chain) adalah model dari proses yang bergerak di sekitar serangkaian kemungkinan nilai yang disebut sebagai keadaan (***state***).
- Status keadaan masa depan (***future state***) akan dipilih secara acak menggunakan beberapa peluang transisi.
- Rantai Markov memiliki karakteristik "tanpa-memori“ (***memoryless***), artinya keadaan masa depan (*future state*) hanya bergantung pada keadaan saat ini (*current state*), tidak tergantung pada keadaan masa lalu (*past states*). Hal ini disebut sebagai **properti Markov**.
- Berdasarkan properti tersebut maka peluang transisi dari rantai Markov hanya akan bergantung pada keadaan saat ini, bukan pada keadaan sebelumnya.

$$P(X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \dots, X_2, X_1) = P(X_{t+1} \mid X_t)$$

Markov Chain (*cont.*)

- Setiap keadaan (*state*) hanya terhubung langsung ke *state* sebelumnya dan tidak ke *state* lain yang lebih jauh ke belakang. Melalui proses ini maka hubungan antar *state* bersifat seperti **rantai**.
- Himpunan dari berbagai *state* disebut sebagai ruang *state* (***state-space***) yang dapat berupa diskrit maupun kontinu.
- Pada perkuliahan ini hanya menggunakan rantai Markov yang peluang transisinya tetap sama di setiap state yang disebut sebagai ***time-invariant***.
- Pada Markov Chain Monte Carlo (MCMC), kita perlu menemukan rantai Markov dengan sebaran jangka panjang (***long-run distribution***) yang sama dengan sebaran posterior $g(\theta|y)$.

Markov Chain (*cont.*)



Algoritma Metropolis–Hastings

Algoritma Metropolis–Hastings

- Alagoritma Metropolis–Hastings bertujuan untuk mengambil sampel dari beberapa **sebaran target** (*target density*) dengan memilih nilai dari sebaran kandidat (*candidate density*).
- Pemilihan untuk menerima suatu nilai kandidat, terkadang disebut **proposal**, hanya bergantung pada nilai yang diterima sebelumnya.
- Algoritma ini membutuhkan nilai awal (*initial value*) untuk memulai, peluang penerimaan (*acceptance*), dan peluang transisi.
- Apabila peluang transisi bersifat **simetris**, maka urutan nilai yang dihasilkan dari proses ini membentuk rantai Markov.
- Secara simetris artinya bahwa peluang bergerak dari state θ ke state θ' **sama** dengan peluang bergerak dari state θ' ke state θ.

Algoritma Metropolis–Hastings (*cont.*)

- Jika $g(\theta|y)$ adalah sebaran posterior (sebaran target) yang tidak diskalakan, dan $q(\theta, \theta')$ adalah sebaran kandidat, maka **peluang transisi** ditentukan sebagai berikut:

$$\alpha(\theta, \theta') = \min\left[1, \frac{g(\theta'|y)q(\theta', \theta)}{g(\theta|y)q(\theta, \theta')}\right]$$

Algoritma Metropolis–Hastings (*cont.*)

Langkah-langkah algoritma Metropolis–Hastings dapat diringkas sebagai berikut:

1. Start at an initial value $\theta^{(0)}$.
2. Do the following for $n = 1, \ldots, n$.
   (a) Draw θ' from $q(\theta^{(n-1)}, \theta')$.
   (b) Calculate the probability $\alpha(\theta^{(n-1)}, \theta')$.
   (c) Draw u from $U(0,1)$.
   (d) If $u < \alpha(\theta^{(n-1)}, \theta')$, then let $\theta^{(n)} = \theta'$, else let $\theta^{(n)} = \theta^{(n-1)}$

Catatan: $\alpha(\theta, \theta') = \min\left[1, \dfrac{g(\theta'|y)q(\theta', \theta)}{g(\theta|y)q(\theta, \theta')}\right]$

Algoritma Metropolis–Hastings (*cont.*)

1. Start at an initial value $\theta^{(0)}$.
2. Do the following for $n = 1, \ldots, n$.
   (a) Draw θ' from $q(\theta^{(n-1)}, \theta')$.
   (b) Calculate the probability $\alpha(\theta^{(n-1)}, \theta')$.
   (c) Draw u from $U(0, 1)$.
   (d) If $u < \alpha(\theta^{(n-1)}, \theta')$, then let $\theta^{(n)} = \theta'$, else let $\theta^{(n)} = \theta^{(n-1)}$

We should note that having the candidate density $q(\theta, \theta')$ close to the target $g(\theta|y)$ leads to more candidates being accepted. In fact, when the candidate density is exactly the same shape as the target

$$q(\theta, \theta') = k \times g(\theta'|y)$$

the acceptance probability is given by

$$\begin{aligned}
\alpha(\theta, \theta') &= \min\left[1, \frac{g(\theta'|y)\, q(\theta', \theta)}{g(\theta|y)\, q(\theta, \theta')}\right] \\
&= \min\left[1, \frac{g(\theta'|y) g(\theta|y)}{g(\theta|y) g(\theta'|y)}\right] \\
&= 1\,.
\end{aligned}$$

Algoritma Gibbs Sampling

Algoritma Gibbs Sampling

- Algoritma Gibbs-Sampling sangat relevan ketika terdapat **banyak parameter** dalam model.
- Algoritma **Metropolis-Hastings** dapat diperluas ke masalah dengan banyak parameter tersebut.
- Namun, seiring bertambahnya jumlah parameter, **performa** algoritma untuk *acceptance rate*-nya umumnya menurun.
- *Acceptance rate* dapat ditingkatkan di Metropolis–Hastings dengan hanya memperbarui satu blok parameter di setiap iterasi. Hal ini kemudian dikenal dengan algoritma ***blokwise***-Metropolis-Hastings.
- Algoritma **Gibbs-Sampling** merupakan **kasus khusus** dari algoritma *blokwise*-Metropolis-Hastings tersebut.

Algoritma Gibbs Sampling (*cont.*)

- Pada penerapannya, algoritma Gibbs-Sampling sangat berguna untuk model berhirarki (***hierarchical model***), karena ketergantungan antar parameter model terdefinisi dengan baik.
- Gibbs-Sampling merupakan algoritma simulasi rantai Markov dari sebaran posterior bersama (***joint posterior distribution***) dengan cara mensimulasikan tiap parameter secara individual (***univariate***) dari rangkaian sebaran bersyarat.
- Mensimulasikan satu nilai dari setiap parameter ini disebut sebagai satu siklus (***one cycle***) dari Gibbs-Sampling.
- Secara umum, performa algoritma simulasi Gibbs-Sampling akan **konvergen** pada sebaran target, yakni ***joint posterior distribution***.

Algoritma Gibbs Sampling (*cont*.)

Suppose we decide to use the true conditional density as the candidate density at each step for every parameter given all of the others. In that case

$$q(\theta_j, \theta_j' | \boldsymbol{\theta}_{-j}) = g(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y}) \,,$$

where $\boldsymbol{\theta}_{-j}$ is the set of all the parameters excluding the j^{th} parameters. Therefore, the acceptance probability for θ_j at the n^{th} step will be

$$\begin{aligned}\alpha\left(\theta_j^{(n-1)}, \theta_j' | \boldsymbol{\theta}_{-j}^{(n)}\right) &= \min\left[1, \frac{g(\theta_j' | \boldsymbol{\theta}_{-j}, \mathbf{y})\, q(\theta_j', \theta_j | \boldsymbol{\theta}_{-j})}{g(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y})\, q(\theta_j, \theta_{-j} | \boldsymbol{\theta}_{-j})}\right] \\ &= 1 \,.\end{aligned}$$

so the candidate will be accepted at each step. The case where we draw each candidate block from its true conditional density given all the other blocks at their most recently drawn values is known as Gibbs sampling.

Implementasi MCMC dalam Program R

Studi Kasus 1:

- Misalkan diketahui **sebaran posterior** bagi theta adalah g(θ|y) = $e^{-\theta}$, dengan θ > 0.
- Lakukan **sampling** bagi θ berdasarkan sebaran posterior tsb (N = 10000).
- Pada metode MCMC, sebaran posterior dinyatakan sebagai sebaran target.
- Secara **analitik** tentukan penduga Bayes bagi θ. Bandingkan dengan hasil yang diperoleh melalui MCMC. Apa kesimpulannya?

```
> #Metode MCMC dengan Algoritma Metropolis-Hastings
>
> par(mar = c(3, 3, 3, 3))
>
> target <- function(s){
+   if (s < 0) {
+     return(0)
+   } else {
+     return(exp(-s))
+   }
+ }
```

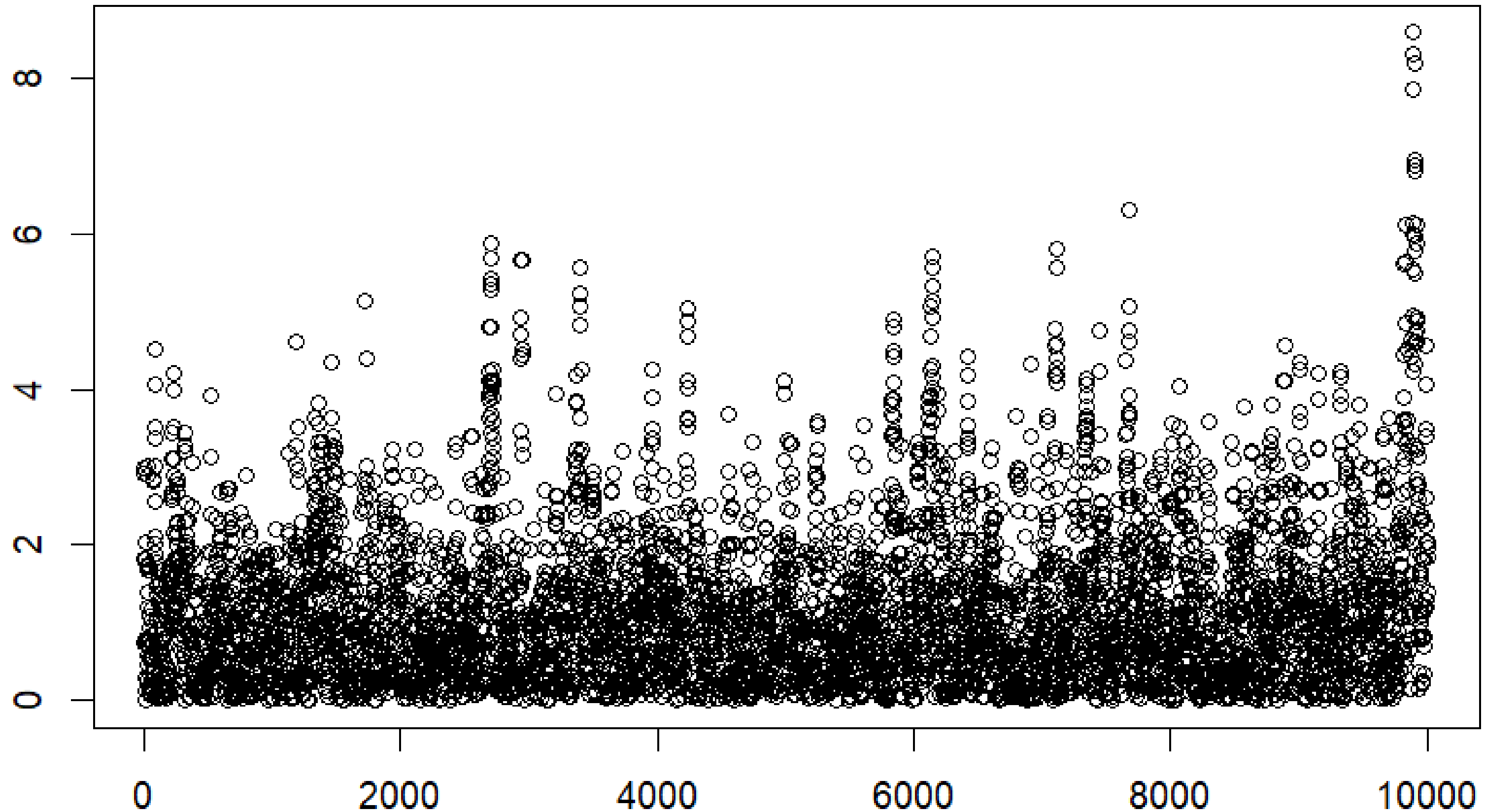
```
> # Algoritma Metropolis-Hastings untuk mendapatkan sampel yang
    sebarannya proporsional pada sebaran target (posterior)
>
> set.seed(1111)
> theta <- rep(0, 10000)
> theta[1] <- 3      # Sebagai nilai inisial bagi theta
> for (i in 2:10000) {
+   current.theta <- theta[i - 1]
> # Menggunakan Random-Walk
+   proposal.dtheta <- current.theta + rnorm(1, mean = 0, sd = 1)
+   A <- target(proposal.dtheta) / target(current.theta)
+   if (runif(1) < A) {
+     theta[i] <- proposal.dtheta
+   } else {
+     theta[i] <- current.theta
+   }
+ }
>
> # Nilai theta dari proses di atas merupakan realisasi dari Markov
    Chain
```

> # Nilai theta dari proses di atas merupakan realisasi dari Markov Chain (N = 10000)

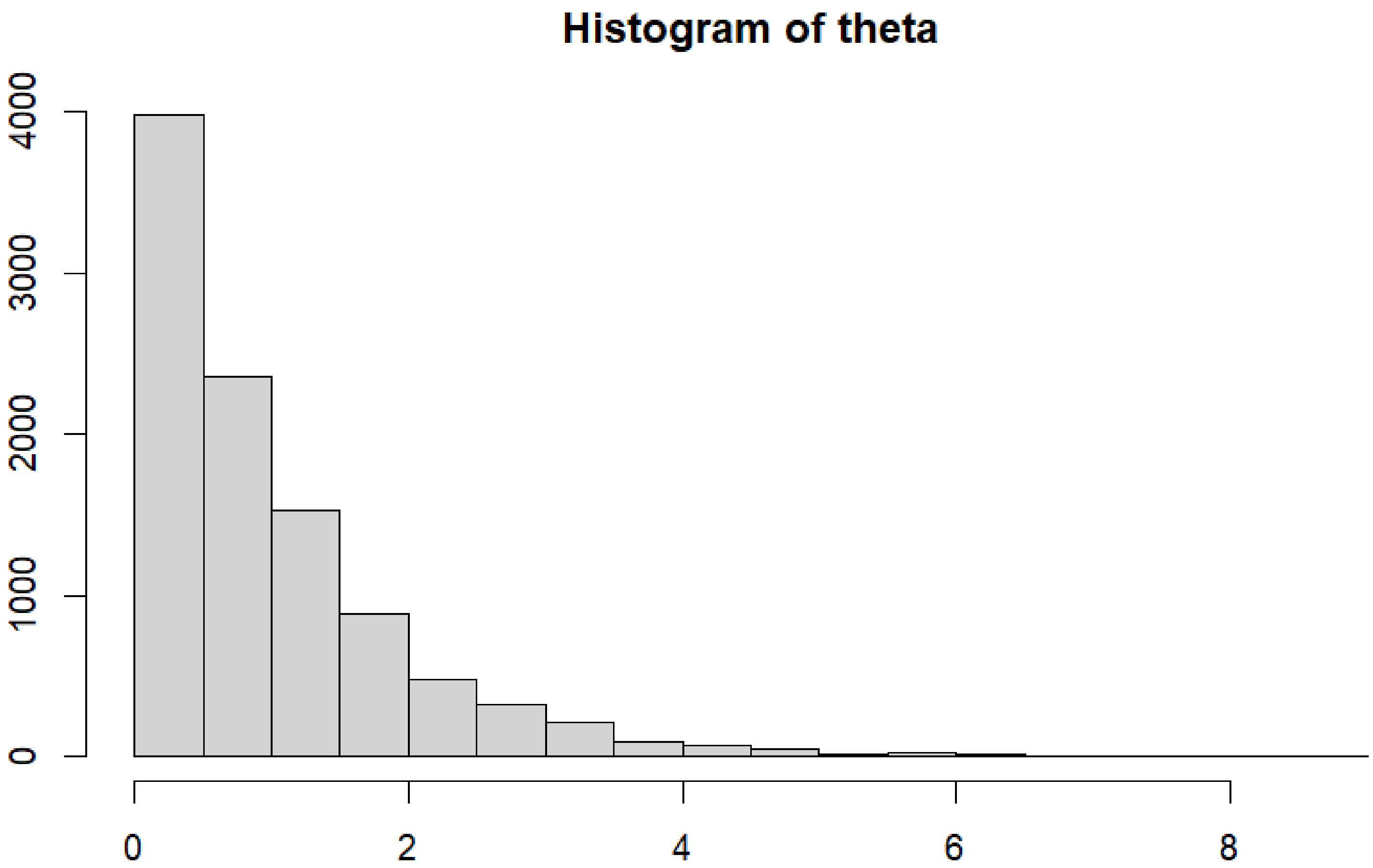
```
> theta
  [1] 3.0000000000 2.9134198887 1.8200030456 1.8200030456 0.7402045871 0.7402045871
  [7] 0.7402045871 2.0241454426 1.8253681405 0.2910865746 0.0001625694 0.0001625694
 [13] 0.0001625694 0.1124735338 0.1124735338 0.7900738330 0.7900738330 0.7900738330
 [19] 0.7900738330 0.7900738330 0.4626884062 0.4626884062 1.3838190122 1.2074620342
 [25] 1.2074620342 1.7179799001 0.2600910150 0.5680476567 0.7776767306 3.0059679173
 [31] 2.8881451089 1.7518898929 1.6694306328 1.6694306328 1.6694306328 1.0783515638
 [37] 0.1838411383 0.7689496223 0.6973162403 0.5216867201 0.5216867201 0.0841627928
 [43] 0.1700531043 0.5578498143 0.1531323266 0.1531323266 0.9450452004 0.9450452004
 [49] 0.9450452004 0.6200579158 0.6201201571 0.1489274163 0.8295695106 1.0301174526
 [55] 0.2632931578 0.0666508234 0.0666508234 0.5333585788 0.5333585788 1.1655024278
 .
 .
 .
[967] 0.0100827629 0.0100827629 0.2344831061 0.2344831061 0.5066017372 1.0802644561
[973] 1.0802644561 0.5790052709 0.3076113715 0.7081145189 0.7081145189 0.3796614005
[979] 1.1317112828 1.1317112828 1.4653975502 1.4653975502 1.4653975502 1.4653975502
[985] 1.8659347420 1.4170129586 1.2667857262 0.3088388461 0.3088388461 0.3088388461
[991] 0.3088388461 0.3088388461 1.1143153294 0.9167967531 1.1041458576 1.1041458576
[997] 1.1041458576 1.1979666189 1.1979666189 1.1979666189
[ reached getOption("max.print") -- omitted 9000 entries ]
```

> # Deskripsi theta yang diperoleh

```
> plot(theta)
```



```
> hist(theta)
```



```
> # Inferensi Bayesian: penduga bagi theta
>
> mean(theta)
[1] 0.9785864

> var(theta)
[1] 0.9494195
>
> # Inferensi Bayesian: menentukan 95% credible interval
>
> nilai.kuantil <- quantile(theta, probs = c(0.025, 0.975))
> cat(paste("Pendekatan 95% credible interval : ["
+           , round(nilai.kuantil[1], 4), " ",
round(nilai.kuantil[2], 4), "]\n", sep = ""))
>
Pendekatan 95% credible interval : [0.0223 3.5293]
```

```
> # Inferensi Bayesian: penduga bagi theta
>
> mean(theta)
  0.9785864
```

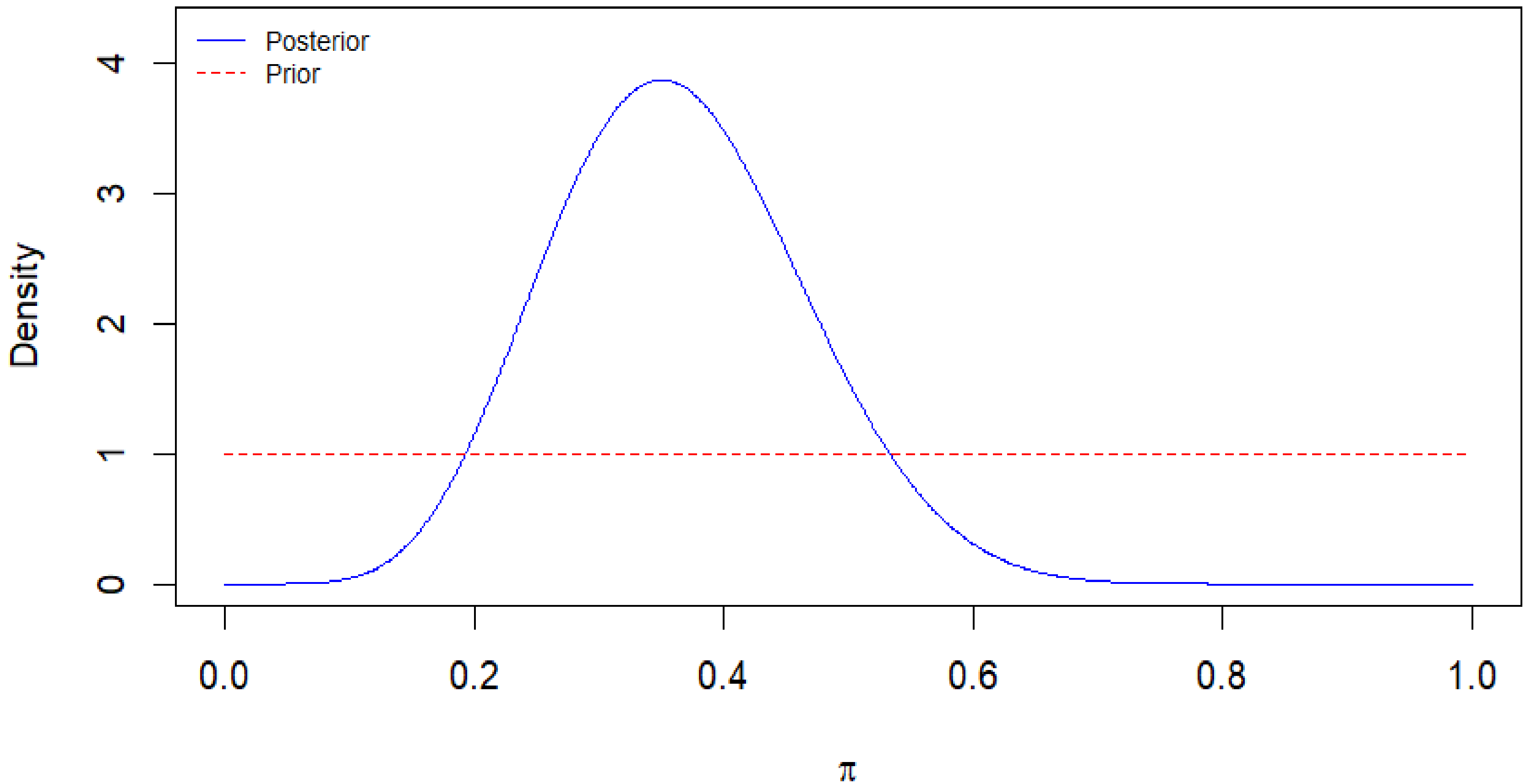
Secara analitik tentukan penduga Bayes bagi θ. Bandingkan dengan hasil yang diperoleh melalui MCMC. Apa kesimpulannya?

Studi Kasus 2:

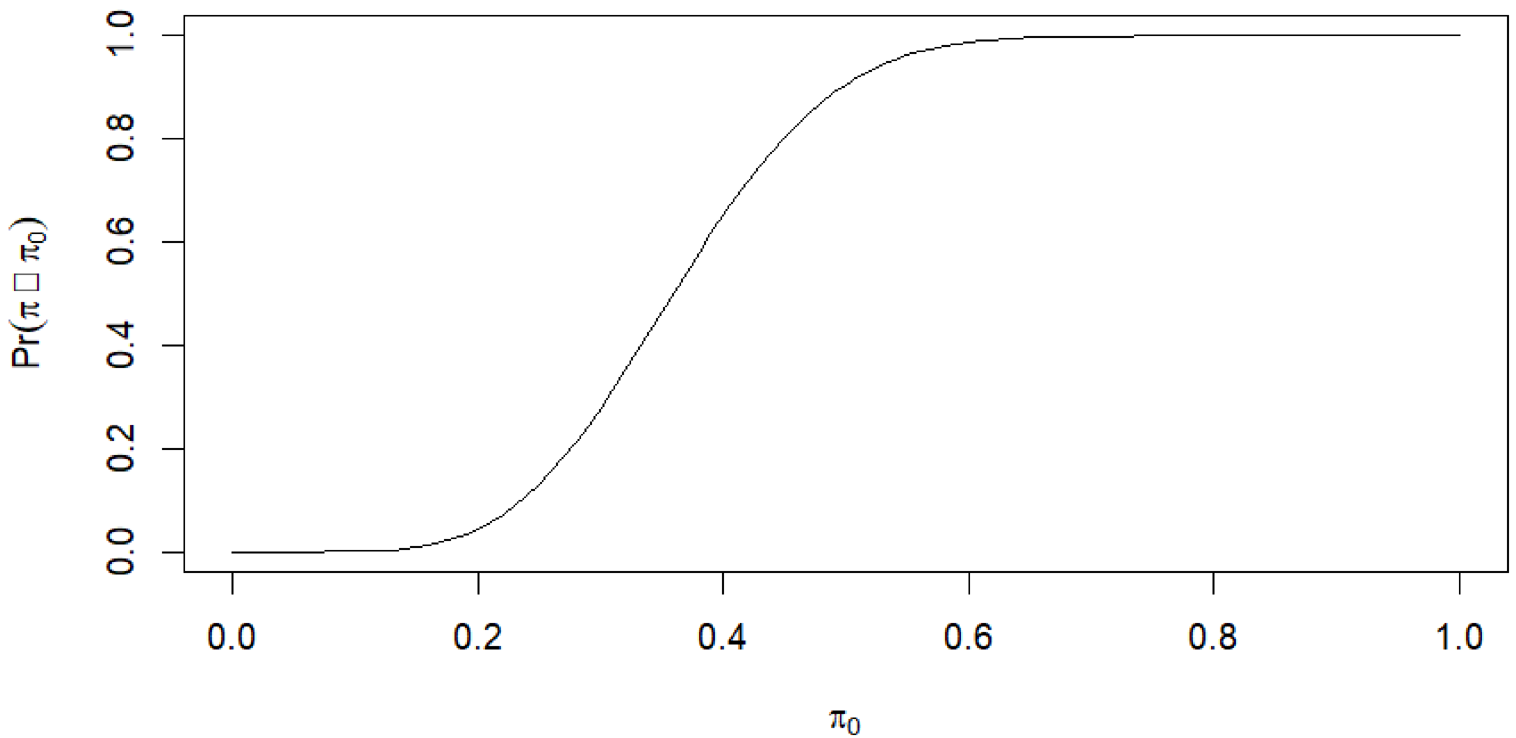
- Misalkan diketahui data menyebar Binomial(y=7, n=20), y adalah banyaknya sukses dari n percobaan.
- Sebaran prior bagi π adalah Uniform(0, 1), sedangkan π adalah peluang terjadinya sukses.
- Sebaran posterior akan dicari melalui MCMC menggunakan *function* **`binogcp()`** pada *package* “**`Bolstad`**” dalam R.

```
> ## melakukan simulasi untuk memperoleh sebaran
     posterior bagi pi melalui MCMC
>
> library(Bolstad)
>
> simulasi.post.pi <- binogcp(7, 20, density = "uniform",
                              params = c(0, 1))
```

```
> ## hasil simulasi berdasarkan MCMC
> simulasi.post.pi
```



```
> ## mencari sebaran kumulatif posterior (CDF)
     menggunakan Simpson's rule
> cdf.post = cdf(simulasi.post.pi)
> plot(cdf.post, type = "l", xlab = expression(pi[0]),
       ylab = expression(Pr(pi <= pi[0])))
```



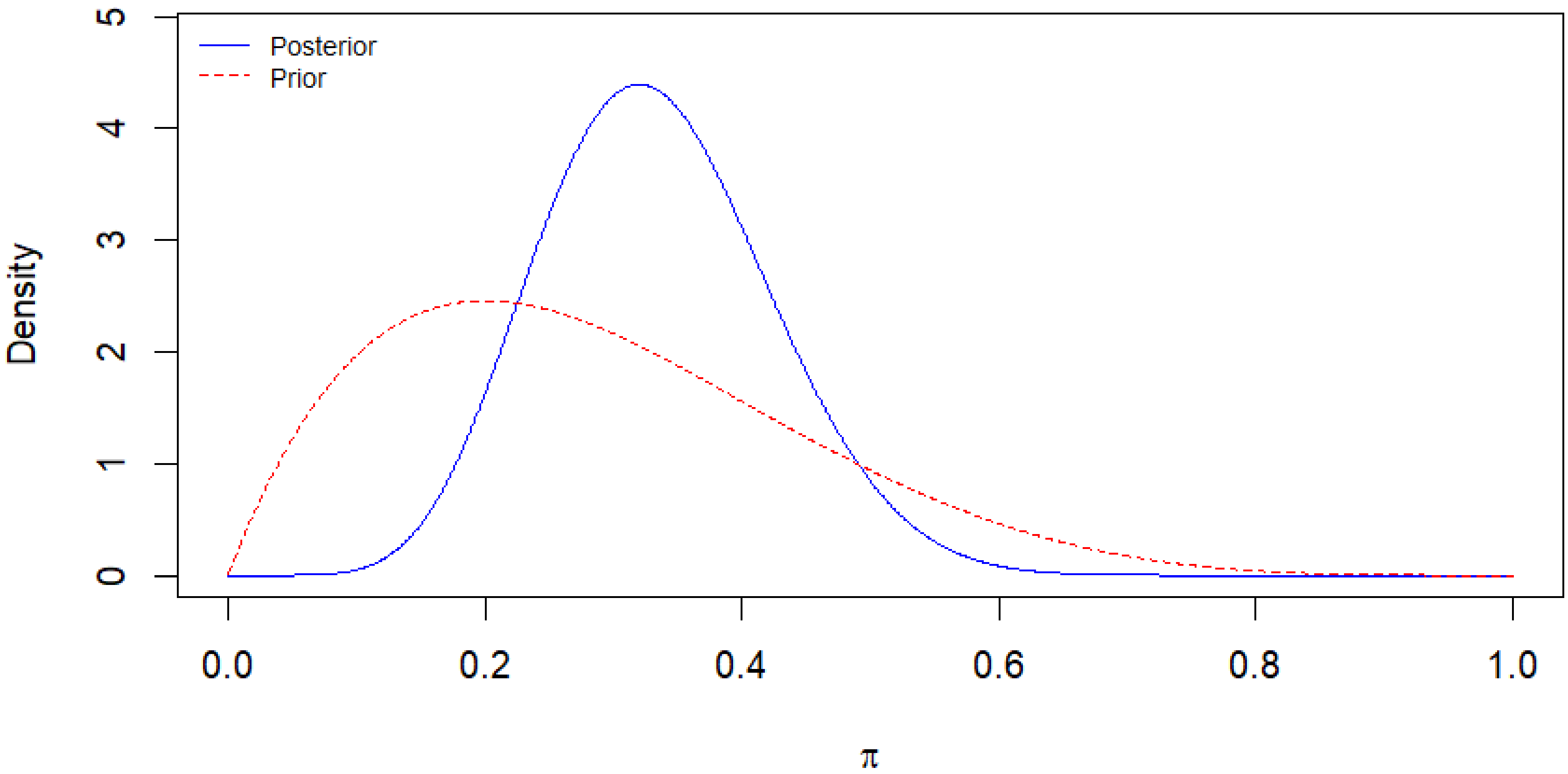
```
> ## mencari inferensi Bayesian dari posterior: nilai
     harapan dan ragam

> round(mean(simulasi.post.pi),5)
[1] 0.36364

> round(var(simulasi.post.pi),5)
[1] 0.01006

> ## menggunakan fungsi quantile untuk mencari 95%
     credible interval

> nilai.kuantil <- quantile(simulasi.post.pi, probs =
                     c(0.025, 0.975))

> cat(paste("Approximate 95% credible interval : ["
            , round(nilai.kuantil[1], 5), " ",
            round(nilai.kuantil[2], 5), "]\n", sep = ""))

Approximate 95% credible interval : [0.18057 0.56918]
```

Studi Kasus 3:

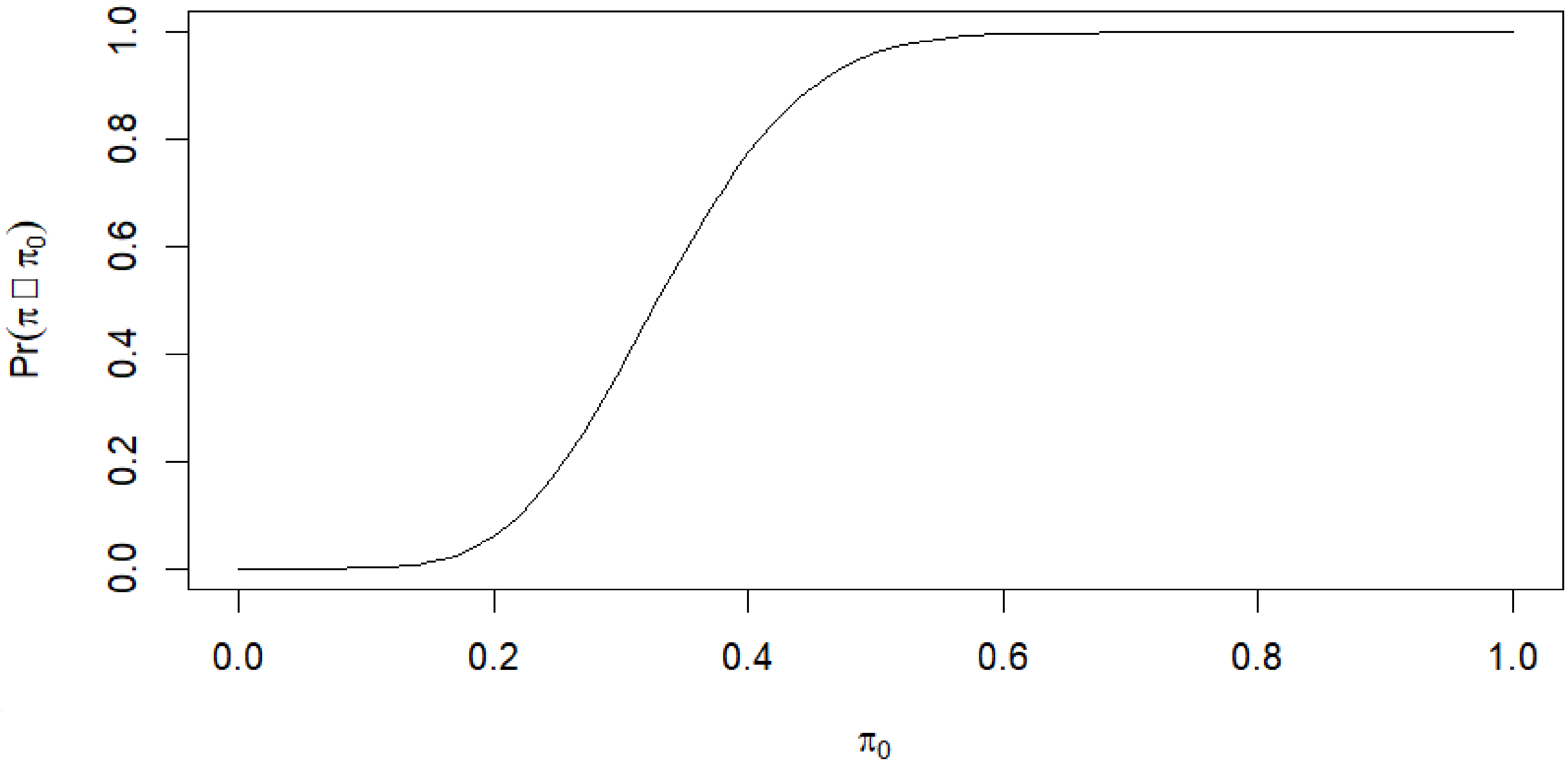
- Misalkan diketahui data menyebar Binomial(y=7, n=20), y adalah banyaknya sukses dari n percobaan.
- Sebaran prior bagi π adalah Beta(2, 5), sedangkan π adalah peluang terjadinya sukses.
- Sebaran posterior akan dicari melalui MCMC menggunakan *function* **`binogcp()`** pada *package* "**`Bolstad`**" dalam R.

```
> ## melakukan simulasi untuk memperoleh sebaran
     posterior bagi pi melalui MCMC
>
> library(Bolstad)
>
> simulasi.post.pi <- binogcp(7, 20, density = "beta",
                              params = c(2, 5))
```

```
> ## hasil simulasi berdasarkan MCMC
> simulasi.post.pi
```



```
> ## mencari sebaran kumulatif posterior (CDF)
     menggunakan Simpson's rule
> cdf.post = cdf(simulasi.post.pi)
> plot(cdf.post, type = "l", xlab = expression(pi[0]),
          ylab = expression(Pr(pi <= pi[0])))
```



```
> ## mencari inferensi Bayesian dari posterior: nilai
     harapan dan ragam

> round(mean(simulasi.post.pi),5)
[1] 0.33333

> round(var(simulasi.post.pi),5)
[1] 0.00794

> ## menggunakan fungsi quantile untuk mencari 95%
     credible interval

> nilai.kuantil <- quantile(simulasi.post.pi, probs =
                   c(0.025, 0.975))

> cat(paste("Approximate 95% credible interval : ["
            , round(nilai.kuantil[1], 5), " ",
            round(nilai.kuantil[2], 5), "]\n", sep = ""))

Approximate 95% credible interval : [0.17164 0.5174]
```

Materi Latihan

1

a. Misalkan diketahui sebaran posterior bagi θ adalah $g(\theta|y) = 6\theta(1-\theta)$ dengan $0 < \theta < 1$.

b. Gunakan metode MCMC pada Program *R* untuk melakukan sampling bagi θ berdasarkan sebaran posterior tersebut.

c. Berdasarkan poin (b) di atas tentukan penduga Bayes bagi θ.

d. Tentukan pula *credible interval* 90% bagi θ.

e. Secara analitik tentukan penduga Bayes bagi θ. Bandingkan hasilnya dengan poin (c) di atas. Apa kesimpulan Anda.

2

a. Misalkan diketahui *Y* menyebar Poisson(λ).

b. Diketahui 5 nilai pengamatan bagi *Y* yaitu 12, 17, 10, 8, dan 23.

c. Misalkan sebaran prior yang digunakan bagi λ adalah Gamma(3, 6).

d. Melalui metode MCMC pada Program *R* lakukan sampling bagi λ. Gunakan *function* poisgcp() pada *package* "Bolstad".

e. Berdasarkan poin (d) di atas tentukan penduga Bayes bagi λ serta *credible interval* 90% bagi λ.

3

a. Misalkan diketahui *Y* menyebar Poisson(λ).

b. Diketahui 5 nilai pengamatan bagi *Y* yaitu 12, 17, 10, 8, dan 23.

c. Misalkan sebaran prior yang digunakan bagi λ adalah *Laplace's prior*, yaitu:

$$\pi(\lambda) = \frac{1}{4}\exp\left(-\frac{|\lambda-3|}{2}\right), \; -\infty < \lambda < \infty$$

d. Melalui metode MCMC pada Program *R* lakukan sampling bagi λ. Gunakan *function* poisgcp() pada *package* "Bolstad". Sebaran prior pada poin (c) di atas harus didefinisikan terlebih dahulu di *option* density = "user".

e. Berdasarkan poin (d) di atas tentukan penduga Bayes bagi λ serta *credible interval* 90% bagi λ.

4

a. Bangkitkan 100 data peubah acak *X*, yang mana diketahui *X* menyebar Normal(μ=5, σ^2=7).

b. Misalkan sebaran prior yang digunakan bagi μ adalah *Laplace's prior*, yaitu:

$$\pi(\mu) = \frac{1}{6}\exp\left(-\frac{|\mu-2|}{3}\right),\ -\infty < \mu < \infty$$

c. Melalui metode MCMC pada Program *R* lakukan sampling bagi μ. Gunakan *function* normgcp() pada *package* "Bolstad". Sebaran prior pada poin (b) di atas harus didefinisikan terlebih dahulu di *option* density = "user".

d. Berdasarkan poin (c) di atas tentukan penduga Bayes bagi μ serta *credible interval* 90% bagi μ. Bandingkan hasilnya dengan nilai μ yang sebenarnya yaitu μ = 5. Apa kesimpulan Anda?

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 4th 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

IPB University
Inspiring Innovation with Integrity

Pengantar Statistika Bayes - STA1312

Metode Komputasi Bayesian - Part 2

(Contoh Penerapan Komputasi Bayesian pada Regresi Linear)

Dr. Kusman Sadik, S.Si, M.Si
Program Studi Statistika dan Sains Data IPB
Tahun Akademik 2022/2023

Regresi Linear dalam Bayesian

- Perhatikan persamaan regresi linear sederhana berikut ini:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

- Asumsi yang digunakan adalah:

$$\varepsilon_i \sim N(0, \sigma^2) \quad \text{dan} \quad y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

- Metode Bayesian digunakan untuk membuat inferensi mengenai parameter model, yaitu β_0 dan β_1.

- Sehingga diperlukan sebaran prior bagi β_0 dan β_1.

- Selanjutnya inferensi didasarkan pada sebaran posterior bagi β_0 dan β_1.

Regresi Linear dalam Bayesian (*cont.*)

$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$

Bayes' rule (parameter estimation version) tells us how to calculate the posterior distribution:

$$p(\theta|x) \propto p(\theta)p(x|\theta)$$

This is the generic form for parameters θ and data x. In our particular case, the unknown parameters are β_0 and β_1, and the data are the y values of the data points. The data also consist of a number N of points and the x-values, but we shall assume that these on their own provide no information about the slope and intercept (it would be a bit strange if they did). So the x-values and the number of points N act like prior information that lurks "in the background" of this entire analysis. The y-values are our data in the sense that we will obtain our likelihood by writing down a probability distribution for the y-values given the parameters.

Sebaran Prior bagi $\boldsymbol{\beta_0}$ dan $\boldsymbol{\beta_1}$

Therefore, Bayes' rule *for this problem* (i.e. with the actual names of our parameters and data, rather than generic names) reads:

$$\boxed{p(\beta_0, \beta_1 | y_1, y_2, ..., y_N) \propto p(\beta_0, \beta_1) p(y_1, y_2, ..., y_N | \beta_0, \beta_1)}$$

We can now say some things about Bayesian linear regression by working analytically. For starters, let's assume uniform priors for both β_0 and β_1, and that the prior for these two parameters are independent. The probability density for a uniform prior distribution can be written simply as:

$$p(\beta_0, \beta_1) \propto 1.$$

Fungsi Likelihood

Now, on to the likelihood. There are N data points and so there are N y-values in the dataset, called $\{y_1, y_2, ..., y_N\}$. We can obtain the likelihood by writing down a probability distribution for the data given the parameters, sometimes called a "sampling distribution". This describes our beliefs about the connection between the data and the parameters, without which it would be impossible to learn anything from data. If we knew the true values of β_0 and β_1, then we would predict the y-values to be scattered around the straight line. Specifically we will assume that each point departs from the straight line by an amount ϵ_i which has a $\mathcal{N}(0, \sigma^2)$ probability distribution. For now, we will assume σ, the standard deviation of the scatter, is known. In "$\sim$" notation, this can be written as:

$$\boxed{y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2).}$$

It is implied that all of the data values are independent (given the parameters). Therefore the likelihood can be written as a product of N normal densities, one for each data point:

$$p(\{y_1, y_2, ..., y_N\} | \beta_0, \beta_1) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right].$$

Sebaran Posterior bagi β_0 dan β_1

$$p(\{y_1, y_2, ..., y_N\}|\beta_0, \beta_1) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right].$$

Remember, when we combine the likelihood with the prior using Bayes' rule, we can usually ignore any constant factors which do not depend on the parameters. This allows us to ignore the first part of the product, outside the exponential (since we are assuming σ is known).

$$\begin{aligned}
p(\beta_0, \beta_1|y_1, y_2, ..., y_N) &\propto p(\beta_0, \beta_1)p(y_1, y_2, ..., y_N|\beta_0, \beta_1) \\
&\propto 1 \times \prod_{i=1}^{N} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right].
\end{aligned}$$

We have just found the expression for the posterior distribution for β_0 and β_1. This is a distribution for two parameters (i.e. it is bivariate).

Penggunaan MCMC

The above analytical results made the unrealistic assumption that the standard deviation σ, of the scatter, was known. In practice, σ usually needs to be estimated from the data as well. Therefore, in the Bayesian framework, we should include it as an extra unknown parameter. Now we have three unknown parameters instead of two. Our parameters are now β_0, β_1, and σ. One major advantage of MCMC is that we can increase the number of unknown parameters without having to worry about the fact that the posterior distribution might be hard to interpret or plot.

The data is the same as before, $\{y_1, y_2, ..., y_N\}$. The likelihood is also the same as before:

$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2).$$

Our three parameters will need priors.

Pengujian Hipotesis bagi β_0 dan β_1

- Pengujian hipotesis bagi β_0 dan β_1 dapat menggunakan pendekatan sebaran ***t-student***, yaitu:

$$\frac{\hat{\beta}_j - \beta_j}{st.dev(\hat{\beta}_j)} \sim t_{(n-2)}$$

dengan *j* = 0 atau 1.

- $\hat{\beta}_j$ merupakan penduga Bayes (***posterior mean***) sedangkan $st.dev(\hat{\beta}_j)$ merupakan galat baku bagi $\hat{\beta}_j$ (***posterior standard deviation***).

- Nilai $\hat{\beta}_j$dan $st.dev(\hat{\beta}_j)$ dapat diperoleh melalui metode **MCMC**.

- ***Credible-interval*** (1 – α)100% juga dapat menggunakan sebaran *t-student* tersebut, yaitu:

$$\hat{\beta}_j \pm \left(t_{(\alpha/2;\ n-2)}\right)\left(st.dev(\hat{\beta}_j)\right)$$

Implementasi dalam Program R

Studi Kasus 1:

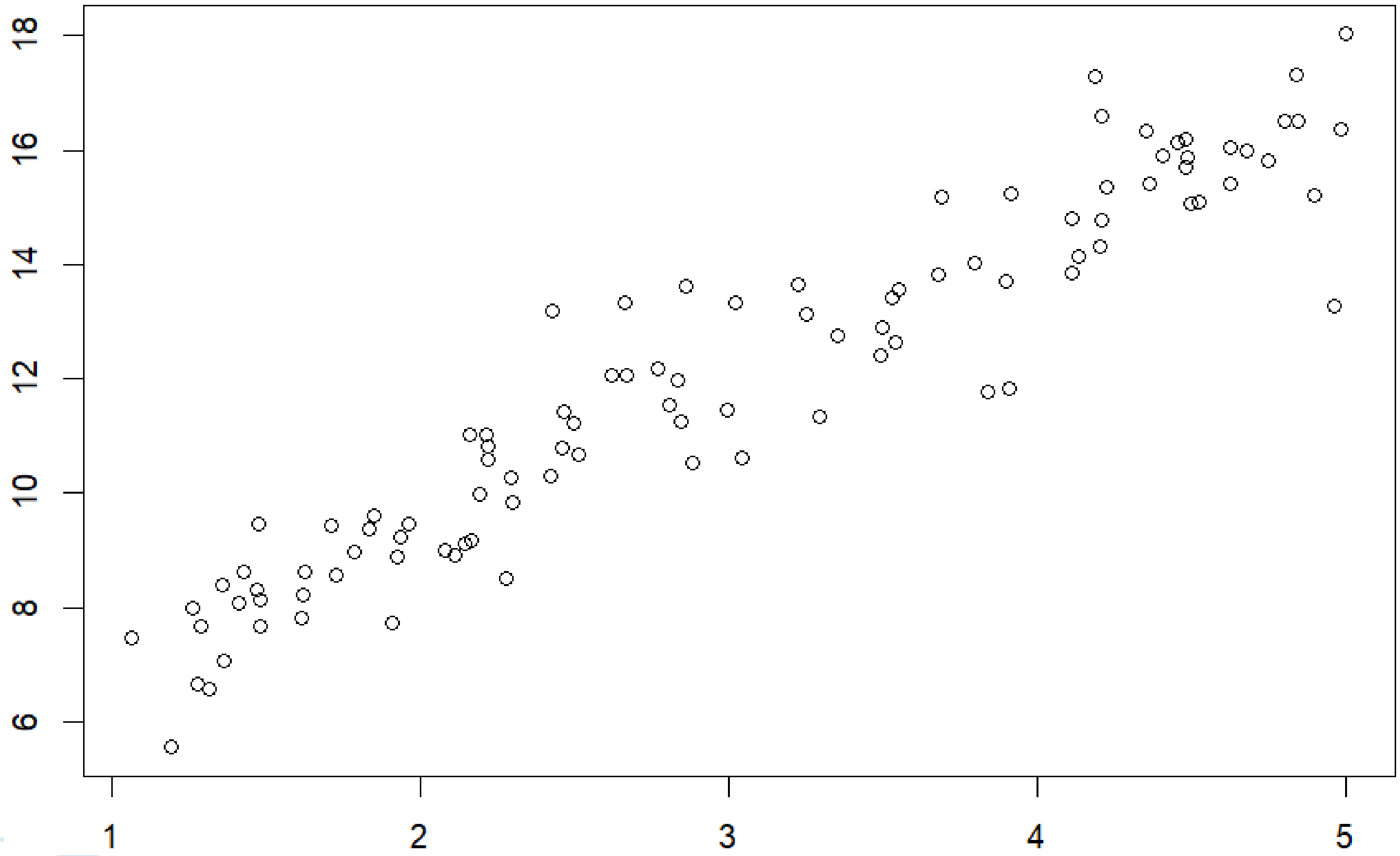
- Misalkan model yang digunakan $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, dengan $\varepsilon_i \sim N(0, \sigma^2)$.
- Data y_i dibangkitkan dengan x_i ~ Uniform(1, 5), $\beta_0 = 4.5$, $\beta_1 = 2.5$, dan $\varepsilon_i \sim N(0,1)$ serta banyak data *n* = 100.
- Sebaran prior yang digunakan bagi β_0 dan β_1 adalah "*flat*" yakni Uniform(0,1).
- Implementasi MCMC Bayesian pada Program R menggunakan *function* `bayes.lin.reg()` dalam *package* "`Bolstad`".
- Berikan penjelasan terkait inferensi Bayes dari *output* `bayes.lin.reg()` tersebut.
- Lakukan pengujian hipotesis untuk H_0: β_1 = 0 vs H_1: β_1 ≠ 0 pada taraf uji α = 0.05.
- Berdasarkan model yang diperoleh tentukan nilai prediksi bagi *y* apabila diketahui nilai *x* adalah 4, 1, dan 2.

Studi Kasus 1: *(cont.)*

```
> library(Bolstad)
>
> ## adjust plot margins
> par(mar = c(2, 2, 2, 2))
>
> ## Pembangkitan data x dan y
> set.seed(1312)
> x = runif(100, min=1, max=5)
> y = 4.5 + 2.5*x + rnorm(100)
>
> data.frame(x,y)
      x    y
1   2.3 10.3
2   2.2  9.2
3   5.0 18.0
.
.
99  2.8 12.2
100 1.3  7.7
```

Studi Kasus 1: *(cont.)*

```
> plot(x,y,xlab ="Nilai X",ylab="Nilai Y")
```



Studi Kasus 1: *(cont.)*

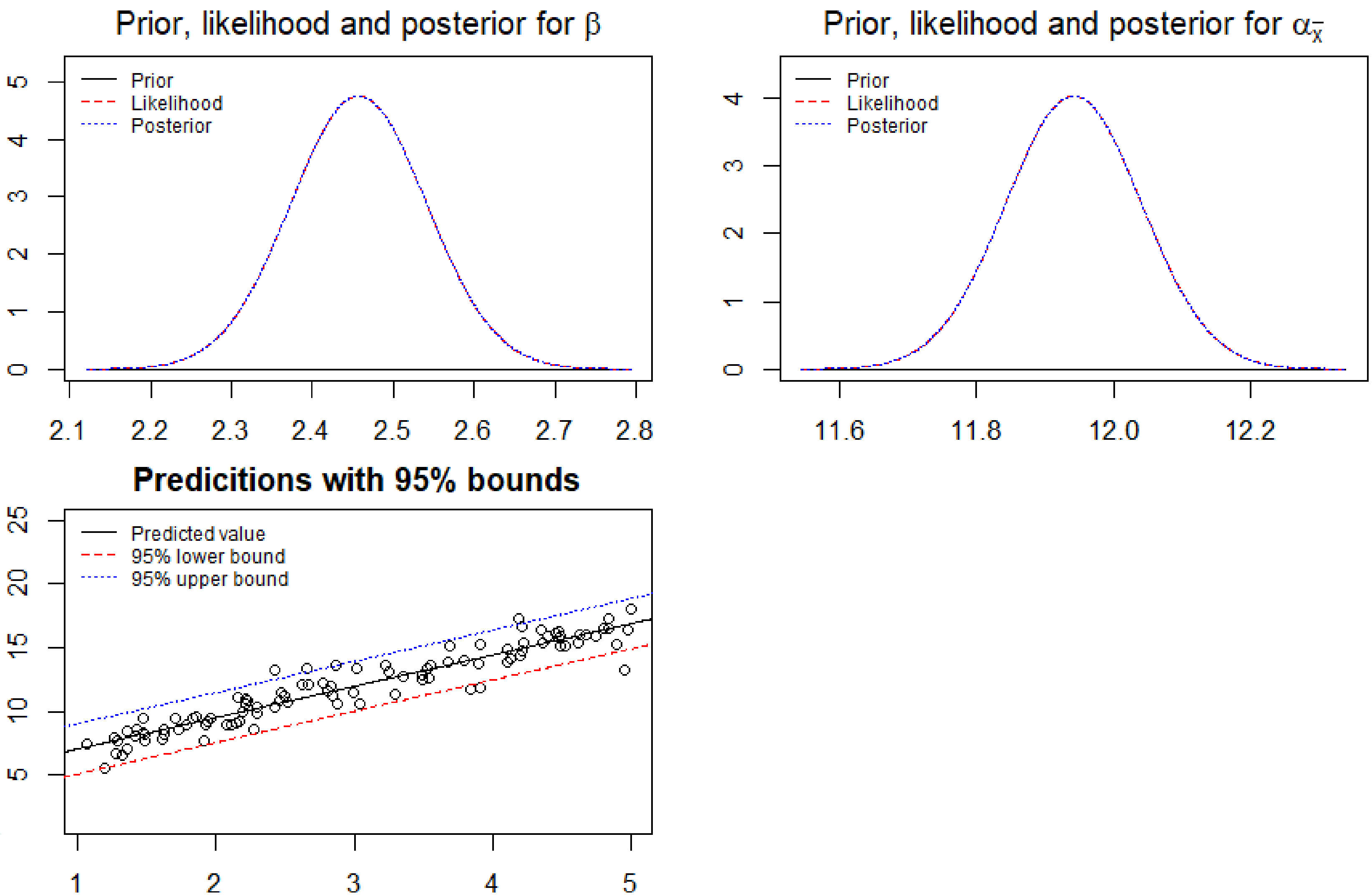
```
> ## Menggunakan prior flat untuk slope.prior (beta1)
> ## Menggunakan prior flat untuk intcpt.prior (beta0)
>
> bayes.lin.reg(y,x, slope.prior = "flat",
                intcpt.prior = "flat",
                alpha = 0.05, plot.data = TRUE)

Standard deviation of residuals:  0.99

            Posterior Mean Posterior Std. Deviation
            -------------- ------------------------
Intercept:  11.94          0.099036
Slope:       2.46          0.084019
```

Studi Kasus 1: *(cont.)*

```
            Posterior Mean Posterior Std. Deviation
            -------------- ------------------------
Intercept:  11.94          0.099036
Slope:       2.46          0.084019
```

Lakukan pengujian hipotesis untuk H_0: $\beta_1 = 0$ vs H_1: $\beta_1 \neq 0$ pada taraf uji $\alpha = 0.05$.

Studi Kasus 1: *(cont.)*



Studi Kasus 1: *(cont.)*

```
> ## Memprediksi y berdasarkan data x yang baru: x = 4, 1, 2
>
> bayes.lin.reg(y,x, slope.prior = "flat",
                intcpt.prior = "flat",
                pred.x=c(4, 1, 2))

x         Predicted y     SE
------    -----------   -----------
4           14.41         0.9989
1            7.04         1.0093
2            9.50         0.9988
```

Studi Kasus 2:

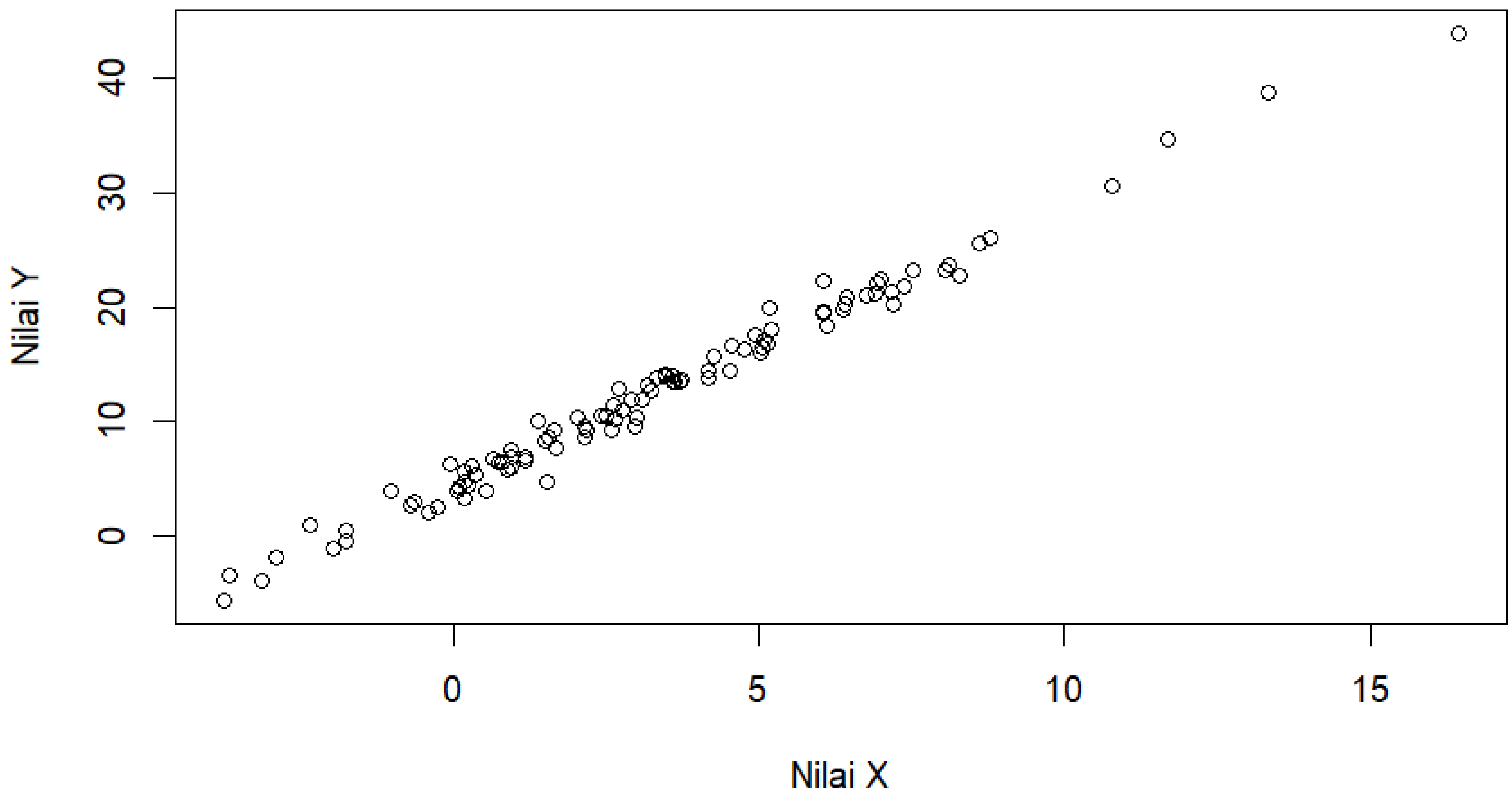
- Misalkan model yang digunakan $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, dengan $\varepsilon_i \sim N(0, \sigma^2)$.
- Data y_i dibangkitkan yang mana $x_i \sim$ Normal(μ = 3, σ = 4), $\beta_0 = 4.5$, $\beta_1 = 2.5$, dan $\varepsilon_i \sim N(0, 1)$.
- Sebaran prior yang digunakan bagi β_0 adalah Normal(μ = 0, σ = 2) dan bagi β_1 adalah Normal(μ = 1, σ = 3).
- Implementasi MCMC Bayesian pada Program R menggunakan *function* **`bayes.lin.reg()`** dalam *package* "**`Bolstad`**".
- Berikan penjelasan terkait inferensi Bayes dari *output* **`bayes.lin.reg()`** tersebut.
- Berdasarkan model yang diperoleh tentukan nilai prediksi bagi *y* apabila diketahui nilai *x* adalah 3, -1, dan 6.

Studi Kasus 2: *(cont.)*

```
> library(Bolstad)
>
> ## adjust plot margins
> par(mar = c(2, 2, 2, 2))
>
> ## Pembangkitan data x dan y
> set.seed(1312)
> x = rnorm(100, mean = 3, sd = 4)
> y = 4.5 + 2.5*x + rnorm(100)
>
> data.frame(x,y)
          x        y
1     1.176     6.68
2    16.438    44.05
3     2.707    12.84
.
.
99    2.186     9.35
100   8.796    26.03
```

Studi Kasus 2: *(cont.)*

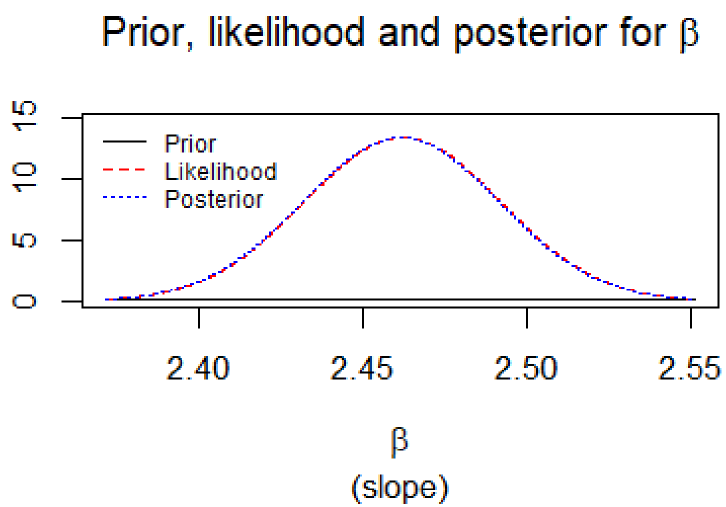
```
> plot(x,y,xlab ="Nilai X",ylab="Nilai Y")
```

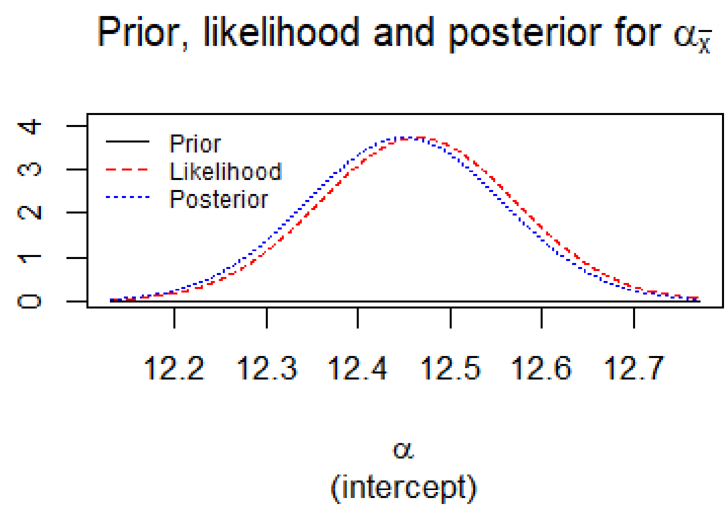


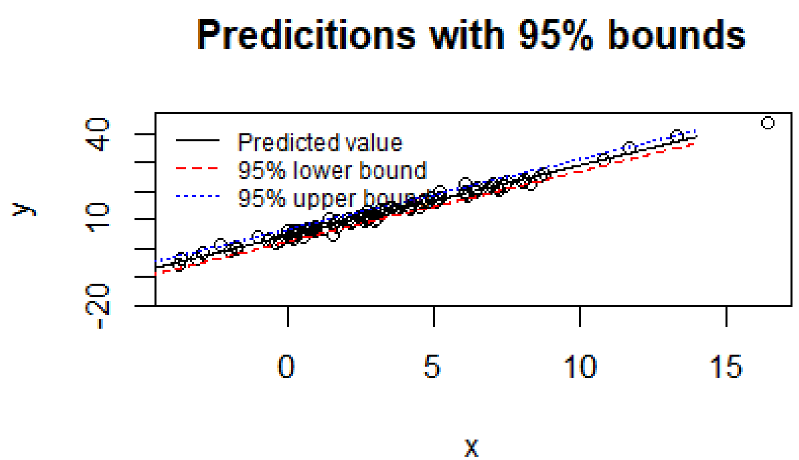
Studi Kasus 2: *(cont.)*

```
> ## Menggunakan slope.prior Normal(mean=0, sd=2)
> ## Menggunakan intcpt.prior Normal(mean=1,sd=3)>

> bayes.lin.reg(y,x, slope.prior = "normal",
               intcpt.prior = "normal",
               0, 2, 1, 3, alpha = 0.05,
               plot.data = TRUE)

Standard deviation of residuals:  1.07


            Posterior Mean Posterior Std. Deviation
            -------------- ------------------------
Intercept:  12.45          0.10690
Slope:       2.46          0.02995
```

Studi Kasus 2: *(cont.)*







Studi Kasus 2: *(cont.)*

```
> ## Memprediksi y berdasarkan data x yang baru: x = 3, -1, 6
>
> bayes.lin.reg(y,x, slope.prior = "normal",
                intcpt.prior = "normal",
                0, 2, 1, 3, pred.x=c(3, -1, 6))

 x      Predicted y     SE
------  -----------  -----------
 3         11.81        1.0750
-1          1.97        1.0825
 6         19.20        1.0781
```

Studi Kasus 3:

Bolstad dan Curran (2017), hlm. 303

14.1. A researcher measured heart rate (x) and oxygen uptake (y) for one person under varying exercise conditions. He wishes to determine if heart rate, which is easier to measure, can be used to predict oxygen uptake. If so, then the estimated oxygen uptake based on the measured heart rate can be used in place of the measured oxygen uptake for later experiments on the individual:

Suppose that we know that oxygen uptake given the heart rate is emphnormal$(\alpha_0 + \beta \times x, \sigma^2)$, where $\sigma^2 = .13^2$ is known. Use a *normal*$(0, 1^2)$ prior for β. What is the posterior distribution of β?

Find a 95% credible interval for β.

Perform a Bayesian test of $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$

at the 5% level of significance.

Studi Kasus 3:

Bolstad dan Curran (2017), hlm. 303

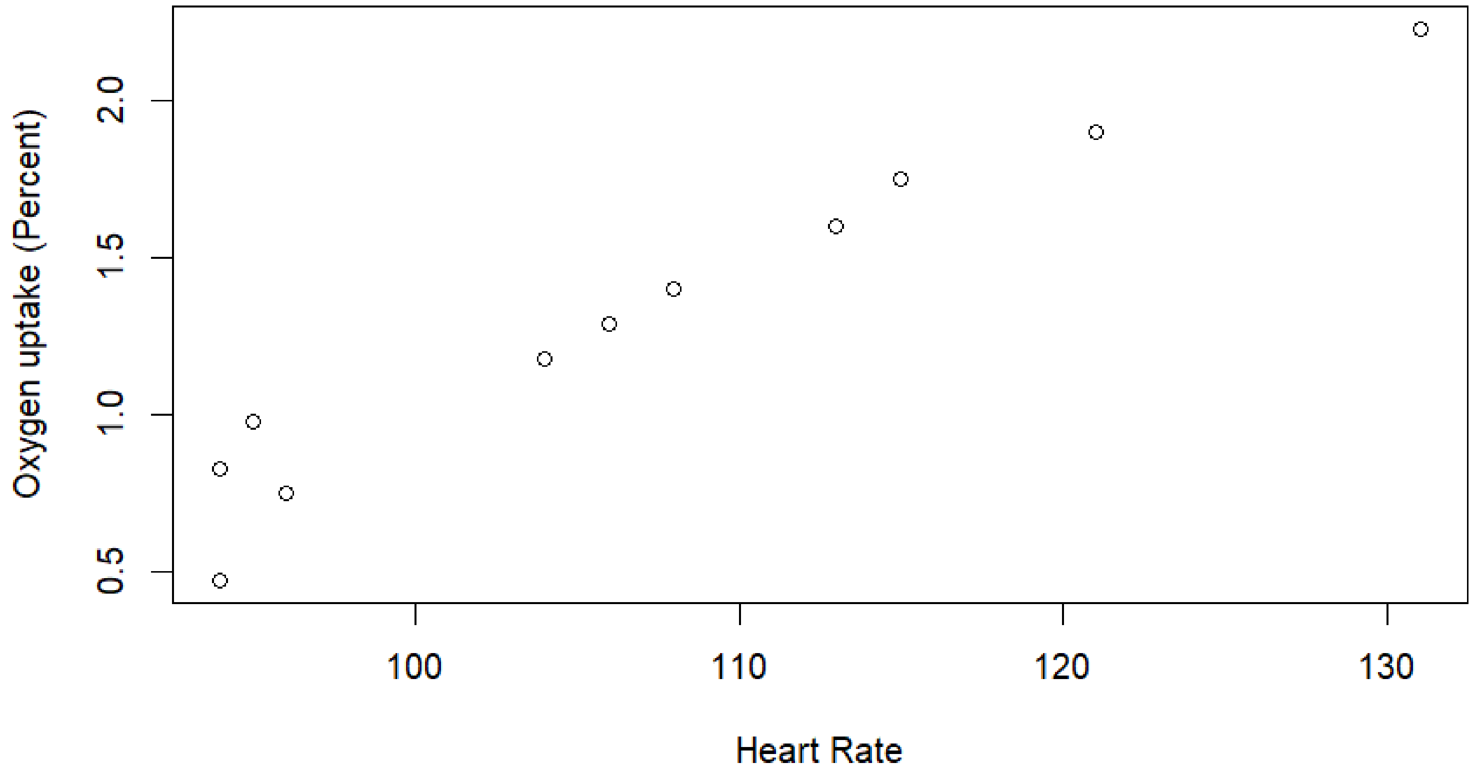
Heart Rate x	Oxygen Uptake y
94	.47
96	.75
94	.83
95	.98
104	1.18
106	1.29
108	1.40
113	1.60
115	1.75
121	1.90
131	2.23

Studi Kasus 3: *(cont.)*

```
> library(Bolstad)
>
> ## adjust plot margins
> par(mar = c(2, 2, 2, 2))
>
> ## data x dan y
> OU = c(0.47,0.75,0.83,0.98,1.18,1.29,1.40,
         1.60,1.75,1.90,2.23)
> HR = c(94,96,94,95,104,106,108,113,
         115,121,131)
```

Studi Kasus 3: *(cont.)*

```
> plot(HR,OU,xlab="Heart Rate",ylab="Oxygen uptake
      (Percent)")
```



Studi Kasus 3: *(cont.)*

```
> bayes.lin.reg(OU,HR,slope.prior = "normal",
                intcpt.prior = "flat",0,1,sigma=0.13)


Known standard deviation:  0.13


              Posterior Mean Posterior Std. Deviation
              -------------- ------------------------
Intercept:    1.307          0.039196
Slope:        0.043          0.003372
```

Studi Kasus 3: *(cont.)*

	Posterior Mean	Posterior Std. Deviation
Intercept:	1.307	0.039196
Slope:	0.043	0.003372

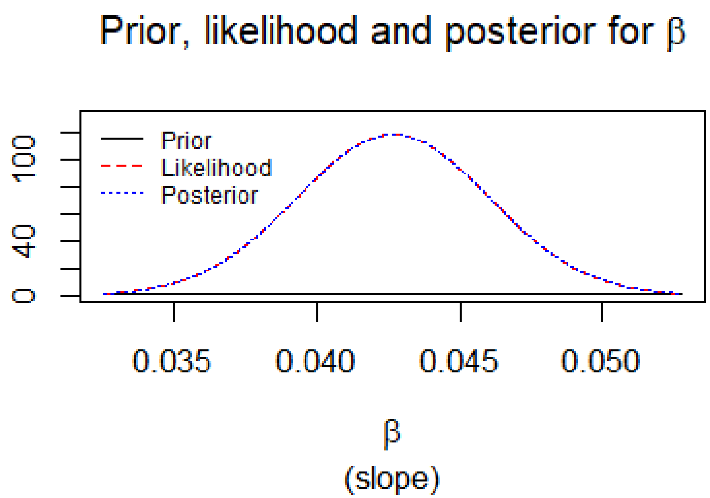
Find a 95% credible interval for β.

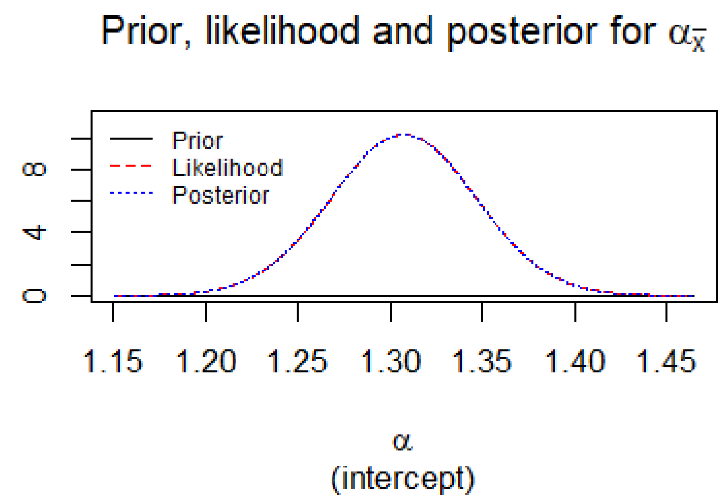
Perform a Bayesian test of

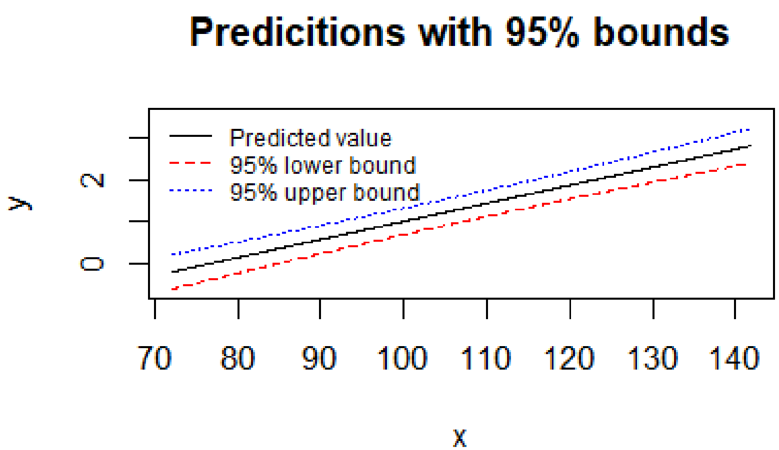
$$H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0$$

at the 5% level of significance.

Studi Kasus 3: *(cont.)*







Materi Latihan

1 Bolstad dan Curran (2017), hlm. 305 → Gunakan Program R

14.2. A researcher is investigating the relationship between yield of potatoes (y) and level of fertilizer (x.) She divides a field into eight plots of equal size and applied fertilizer at a different level to each plot. The level of fertilizer and yield for each plot is recorded below:

Fertilizer Level x	Yield y
1	25
1.5	31
2	27
2.5	28
3	36
3.5	35
4	32
4.5	34

1 Bolstad dan Curran (2017), hlm. 305

(a) Plot a scatterplot of yield versus fertilizer level.

(b) Calculate the parameters of the least squares line.

(c) Graph the least squares line on your scatterplot.

(d) Calculate the estimated variance about the least squares line.

(e) Suppose that we know that yield given the fertilizer level is emphnormal$(\alpha_0 + \beta \times x, \sigma^2)$, where $\sigma^2 = 3.0^2$ is known. Use a *normal*$(2, 2^2)$ prior for β. What is the posterior distribution of β?

(f) Find a 95% credible interval for β.

(g) Perform a Bayesian test of

$$H_0 : \beta \leq 0 \quad \text{versus} \quad H_1 : \beta > 0$$

at the 5% level of significance.

2 Bolstad dan Curran (2017), hlm. 307 → Gunakan Program R

14.5. A textile manufacturer is concerned about the strength of cotton yarn. In order to find out whether fiber length is an important factor in determining the strength of yarn, the quality control manager checked the fiber length (x) and strength (y) for a sample of 10 segments of yarn. The results are:

Fiber Length x	Strength y
85	99
82	93
75	103
73	97
76	91
73	94
96	135
92	120
70	88
74	92

2 Bolstad dan Curran (2017), hlm. 307

(a) Plot a scatterplot of strength versus fiber length.

(b) Calculate the parameters of the least squares line.

(c) Graph the least squares line on your scatterplot.

(d) Calculate the estimated variance about the least squares line.

(e) Suppose we know that the strength given the fiber length is emphnormal$(\alpha_0 + \beta \times x, \sigma^2)$, where $\sigma^2 = 7.7^2$ is known. Use a *normal*$(0, 10^2)$ prior for β. What is the posterior distribution of β.

(f) Find a 95% credible interval for β.

(g) Perform a Bayesian test of

$$H_0 : \beta \leq 0 \quad \text{versus} \quad H_1 : \beta > 0$$

at the 5% level of significance.

(h) Find the predictive distribution for y_{11}, the strength of the next piece of yarn which has fiber length $x_{11} = 90$.

(i) Find a 95% credible interval for the prediction.

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