Frequentist vs Bayesian

Responsi 1 – STA1312 Pengantar Statistika Bayes

Septian Rahardiantoro
Gerry Alfa Dito

Review Materi

Konteks	Frequentist	Bayesian
Peluang	Frekuensi dari kejadian ketika percobaan diulang beberapa kali	keyakinan
Data	random	fixed
Parameter	fixed	random
Selang kepercayaan	Berdasarkan contoh yang diambil secara acak berulang dari populasi	Berdasarkan informasi kontekstual khusus masalah dari sebaran prior

Latihan 1

B/S

1. Jika dilihat dari sudut pandang Bayesian, data diasumsikan bersifat acak yang memiliki sebaran.
2. Keunggulan dari metode Bayesian adalah dapat memperbaharui informasi yang diperoleh.
3. Posterior dalam konsep Bayesian merupakan sebaran awal sebelum data diperoleh.
4. Parameter memiliki sebaran adalah konsep dari sudut pandang Frequentist.

Latihan 2 – Peluang Bayes

Dari 200 orang pelamar terdapat 150 berasal dari SMA dan sisanya berasal dari SMK. Dari 150 pelamar SMA tersebut hanya 80 yang mempunyai sertifkat bahasa Inggris. Pelamar yang berasal dari SMK, hanya 20% yang mempunyai sertifikat bahasa Inggris. Jika seorang pelamar ditarik secara acak, berapa peluang pelamar tersebut berpendidikan SMK jika diketahui ybs mempunyai sertifikat bahasa Inggris?

Latihan 3

Selesaikan problem ini melalui Bayes' Box untuk memperoleh posterior: Secara manual;

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

Pembagian Kelompok

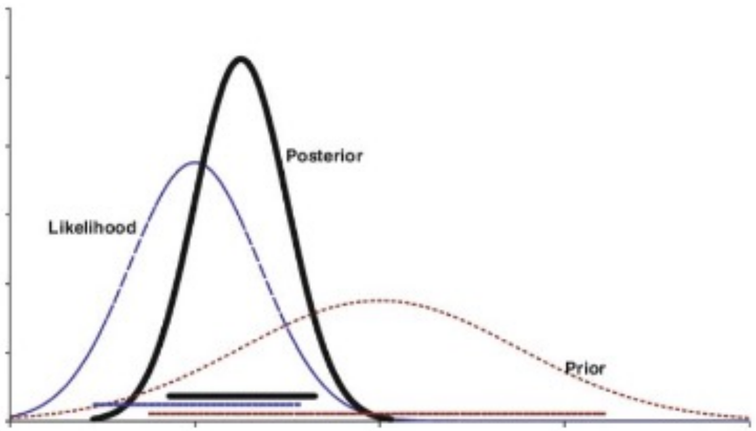
Terima kasih ☺

Konsep Inferensi Bayesian

Responsi 1 – STA1312 Pengantar Statistika Bayes

Septian Rahardiantoro

Review Materi:

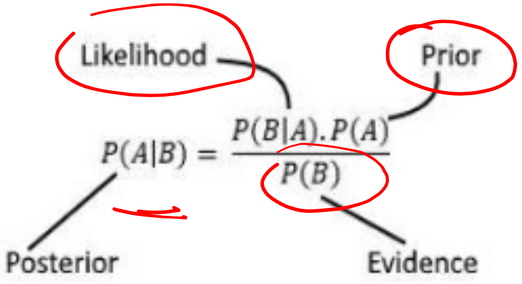


$Y_1, Y_2, \ldots, Y_n$ iid $Y \sim f_Y(y\,|\,\theta),\ \theta \in \Omega$

Fkp/fmp bersama dari Y dan θ adalah $f(y,\theta) = f_Y(y\,|\,\theta)\ \pi(\theta)$.

Fkp/fmp marginal bagi Y adalah $f(y) = \int_\theta f(y,\theta)\ d\theta$.

Fkp/fmp posterior bagi θ dengan syarat Y = y adalah $\pi(\theta\,|\,y) = \dfrac{f(y,\theta)}{f(y)}$.



Dengan menggunakan *loss function* kuadratik $L(\theta\ ,\ d) = (\theta - d)^2$ dimana d adalah fungsi keputusan, maka penduga bagi θ,

$$\hat{\theta}_B = E(\theta\,|\,y) = \int_\theta \theta\ \pi(\theta\,|\,y)\ d\theta.$$

name	notation	equal to
model, likelihood	$f(x\|\theta)$	
prior	$\pi(\theta)$	
joint	$h(x,\theta)$	$f(x\|\theta)\pi(\theta)$
marginal	$m(x)$	$\int_{\Theta} f(x\|\theta)\pi(\theta)d\theta$
posterior	$\pi(\theta\|x)$	$f(x\|\theta)\pi(\theta)/m(x)$
predictive	$f(y\|x)$	$\int_{\Theta} f(y\|\theta)\pi(\theta\|x)d\theta$

$$\underbrace{\pi(\theta|x)}_{\text{data}} = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

posterior $\propto$ prior $\times$ likelihood.

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."

K I   K II

0,01 D < T 0,95 ; $\bar{T}$ 0,05

0,99 $\bar{D}$ < T 0,1 ; $\bar{T}$ 0,9

$$\rightarrow P(D|T) = \frac{P(D) \cdot P(T|D)}{P(D) \cdot P(T|D) + P(\bar{D}) \cdot P(T|\bar{D})}$$

$= 8{,}75\%$

Latihan 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?

K I   K II

0,001 B — A 0,9 ; $\bar{A}$ 0,1

0,999 $\bar{B}$ — A 0,05 ✓ ; $\bar{A}$ 0,95

a. $$P(B|A) = \frac{P(B)\cdot P(A|B)}{P(B)\cdot P(A|B) + P(\bar{B})\cdot P(A|\bar{B})}$$

$$= \frac{0{,}001 \cdot 0{,}9}{0{,}001 \cdot 0{,}9 + 0{,}999 \cdot 0{,}05} = 0{,}01769$$

Latihan 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.

X = byk bola merah dlm guci
$X = \{0, 1, 2, \ldots 9\}$

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.

1M 1H

1M 8H

X	*prior*	*likelihood*	*prior* × *likelihood*	*posterior*
0	$P(X=0)=\frac{1}{10}$	0		
1	$P(X=1)=\frac{1}{10}$	$\frac{\binom{1}{1}\binom{8}{1}}{\binom{9}{2}} = \frac{8}{36}$		
$\vdots$				
9	$P(X=9)=\frac{1}{10}$			

X	Prior	Likelihood $\to \boxed{MH}$	$h = P \cdot L$	Posterior
0	$1/10$	0 ?	0	0
1	$1/10$	$\binom{1}{1}\binom{8}{1}/\binom{9}{2} = 8/36$	$8/360$	$8/120$
2	$1/10$	$\binom{2}{1}\binom{7}{1}/\binom{9}{2} = 14/36$	$14/360$	$14/120$
3	$1/10$	$\binom{3}{1}\binom{6}{1}/\binom{9}{2} = 18/36$	$18/360$	$18/120$
$\boxed{4}$	$1/10$	$\binom{4}{1}\binom{5}{1}/\binom{9}{2} = 20/36$	$20/360$	$20/120$
$\boxed{5}$	$1/10$	$\binom{5}{1}\binom{4}{1}/\binom{9}{2} = 20/36$	$20/360$	$20/120$
6	$1/10$	$\binom{6}{1}\binom{3}{1}/\binom{9}{2} = 18/36$	$18/360$	$18/120$
7	$1/10$	$\binom{7}{1}\binom{2}{1}/\binom{9}{2} = 14/36$	$14/360$	$14/120$
8	$1/10$	$\binom{8}{1}\binom{1}{1}/\binom{9}{2} = 8/36$	$8/360$	$8/120$
9	$1/10$	0 ?	0	0

$$\text{Total} = \frac{120}{360}$$

Latihan 4

$Y_1 \sim \text{binomial}(n=10, \pi)$

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.

$Y_1 = 7 \rightarrow P(Y_1 = 7) = \binom{10}{7} \pi^7 (1-\pi)^3$

π	*prior*	*likelihood*	*prior × likelihood*	*posterior*
0,2	1/4			
0,4	1/4			
0,6	1/4			
0,8	1/4			

π	prior	likelihood $\rightarrow P(Y_1=7) = \binom{10}{7}\pi^7(1-\pi)^3$	$h = p \cdot L$	posterior
0,2	1/4	$\binom{10}{7} \cdot 0{,}2^7(0{,}8)^3 = 0{,}000786$	0,0001965	0,00171
0,4	1/4	$\binom{10}{7} 0{,}4^7(0{,}6)^3 = 0{,}042467$	0,010617	0,09241
$\boxed{0{,}6}$	1/4	$\binom{10}{7} 0{,}6^7(0{,}4)^3 = 0{,}214991$	0,053748	$\boxed{0{,}46781}$
0,8	1/4	$\binom{10}{7} 0{,}8^7(0{,}2)^3 = 0{,}201326$	0,050332	+ 0,938075

Total 0,1148935

$Y_1 \rightarrow$ Sukses $\rightarrow$ pendosan $Y_1 = 7$ ✓

$\pi = 0{,}6$

Latihan 5 (PR)

Let n be the unknown number of customers that visit a store on the day of a sale. The number of customers that make a purchase is $Y|n \sim \text{Binomial}(n, \theta)$ where θ is the known probability of making a purchase given the customer visited the store. The prior is $n \sim \text{Poisson}(5)$. Assuming θ is known and n is the unknown parameter, plot the posterior distribution of n for all combinations of $Y \in \{0, 5, 10\}$ and $\theta \in \{0.2, 0.5\}$ and comment on the effect of Y and θ on the posterior.

Dikumpulkan melalui newlms dalam format **pdf**, dengan batas waktu pengumpulan maksimal Kamis 9 Februari 16:00WIB

Terima kasih ☺

Konsep Inferensi Bayesian (2)

Responsi 3 - STA1312 Pengantar Statistika Bayes

Septian Rahardiantoro

Review

name	notation	equal to
model, likelihood	$f(x\|\theta)$	
prior	$\pi(\theta)$	
joint	$h(x,\theta)$	$f(x\|\theta)\pi(\theta)$
marginal	$m(x)$	$\int_\Theta f(x\|\theta)\pi(\theta)d\theta$
posterior	$\pi(\theta\|x)$	$f(x\|\theta)\pi(\theta)/m(x)$
predictive	$f(y\|x)$	$\int_\Theta f(y\|\theta)\pi(\theta\|x)d\theta$

Prior Distribution : Conjugate

Likelihood	Prior	Posterior
$X\|\theta \sim \mathcal{N}(\theta, \sigma^2)$	$\theta \sim \mathcal{N}(\mu, \tau^2)$	$\theta\|X \sim \mathcal{N}(\frac{\tau^2}{\sigma^2+\tau^2}X + \frac{\sigma^2}{\sigma^2+\tau^2}\mu, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$
$X\|\theta \sim \mathcal{B}(n, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X \sim \mathcal{B}e(\alpha + x, n - x + \beta)$
$X_1, \ldots, X_n\|\theta \sim \mathcal{P}(\theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{G}a(\sum_i X_i + \alpha, n + \beta)$.
$X_1, \ldots, X_n\|\theta \sim \mathcal{NB}(m, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{B}e(\alpha + mn, \beta + \sum_{i=1}^n x_i)$
$X \sim \mathcal{G}(n/2, 2\theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(n/2 + \alpha, (x/2 + \beta^{-1})^{-1})$
$X_1, \ldots, X_n\|\theta \sim \mathcal{U}(0, \theta)$	$\theta \sim \mathcal{P}a(\theta_0, \alpha)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{P}a(\max\{\theta_0, x_1, \ldots, x_n\}\alpha + n)$
$X\|\theta \sim \mathcal{N}(\mu, \theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(\alpha + 1/2, \beta + (\mu - X)^2/2)$
$X\|\theta \sim \mathcal{G}a(\nu, \theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X \sim \mathcal{G}a(\alpha + \nu, \beta + x)$

Non-informative Prior

Laplace's Prior

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

Jeffreys' Prior

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

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

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.

Latihan 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.

Latihan 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.

(d) Find a 95% Bayesian credible interval for μ.

Latihan 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.

(d) Find a 95% Bayesian credible interval for μ.

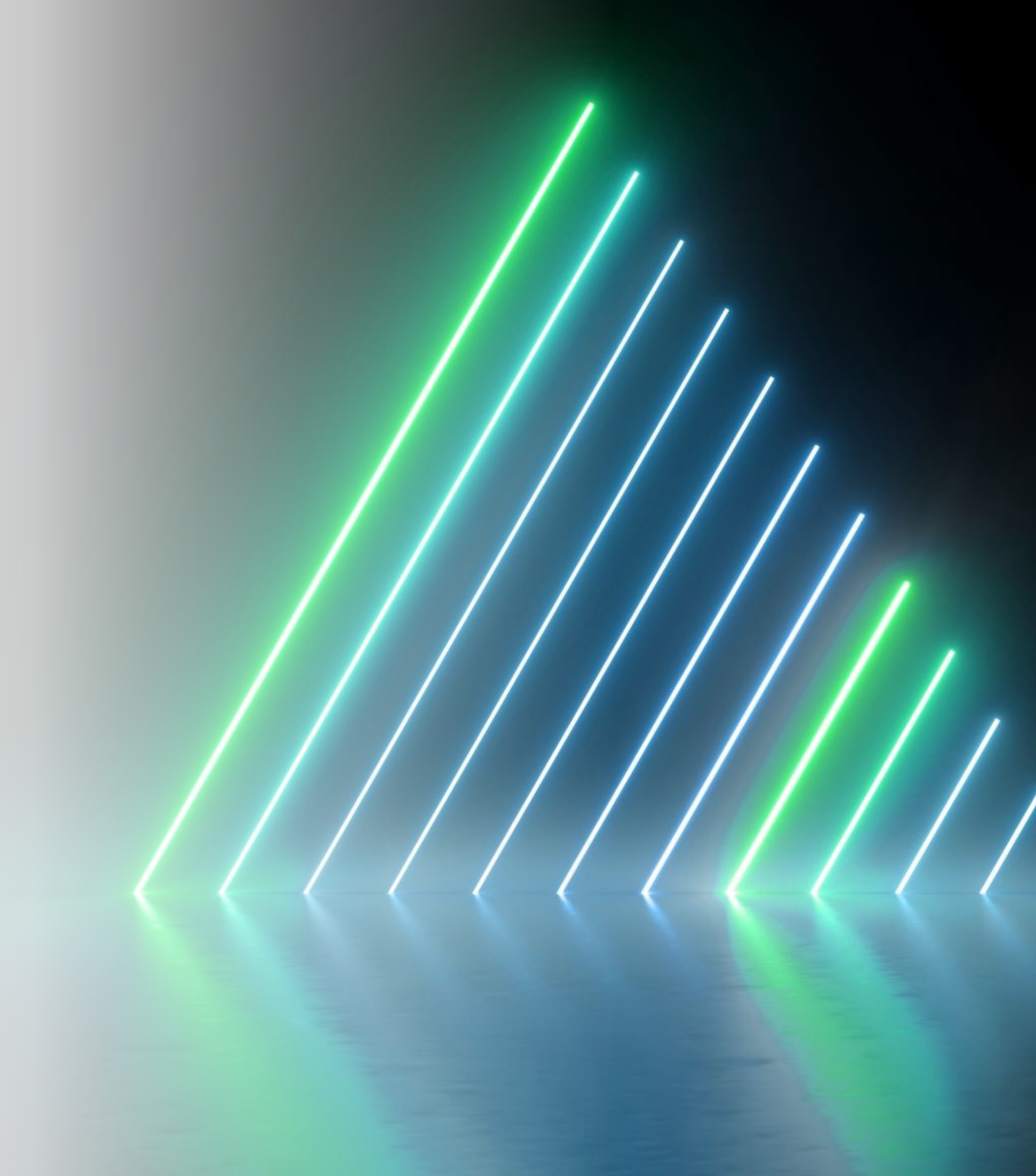
Jawaban Latihan 3 dan 4 dengan R

```
install.packages("Bolstad")
library(Bolstad)

##Latihan 3
poisgamp(c(3,4,3,0,1),11,5)
mean(poisgamp(c(3,4,3,0,1),11,5))
quantile(poisgamp(c(3,4,3,0,1),11,5), probs = 0.5)

##Latihan 4
poisgamp(c(3,4,3,0,1),11.5,5)
mean(poisgamp(c(3,4,3,0,1),11.5,5))
quantile(poisgamp(c(3,4,3,0,1),11.5,5), probs = 0.5)
```

Terima kasih ☺

Konsep Inferensi dan Komputasi Bayesian

Responsi 4 & 5 - STA1312 Pengantar Statistika Bayes



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.

$H_0 = X = 1$

$H_1 = X = -1$

MAP $\rightarrow$ Dipilih H_0 $\leftrightarrow$ $\underbrace{f_Y(y|H_0)}\cdot P(H_0) > \underbrace{f_Y(y|H_1)}\cdot P(H_1)$

$W \sim N(0, \sigma^2)$

- dibawah kondisi H_0 $\rightarrow$ $Y = 2 + W \rightarrow E(Y) = E(2+W) = 2$

$Var(Y) = Var(2+W) = Var(W) = \sigma^2$

$Y | H_0 \sim N(2, \sigma^2)$

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

- dibawah kondisi H_1 $\rightarrow$ $Y = -2 + W$

$Y | H_1 \sim N(-2, \sigma^2)$

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

$\Rightarrow$ Dpilih H_0 $\leftrightarrow$ $f_Y(y|H_0) P(H_0) > f_Y(y|H_1) . P(H_1)$

$$\cancel{\frac{1}{\sigma\sqrt{2\pi}}} e^{-\frac{(y-2)^2}{2\sigma^2}} . p > \cancel{\frac{1}{\sigma\sqrt{2\pi}}} e^{-\frac{(y+2)^2}{2\sigma^2}} . (1-p)$$

$$e^{-\frac{(y-2)^2}{2\sigma^2}} . e^{\frac{(y+2)^2}{2\sigma^2}} > \frac{1-p}{p}$$

$$\exp\left(\frac{1}{2\sigma^2}\left(-\cancel{y^2} + 4y - \cancel{4} + \cancel{y^2} + 4y + \cancel{4}\right)\right) > \frac{1-p}{p}$$

$$\rightarrow \exp\left(\frac{8y}{2\sigma^2}\right) > \frac{1-p}{p}$$

$$\frac{4y}{\sigma^2} > \ln\left(\frac{1-p}{p}\right)$$

$$y > \frac{\sigma^2}{4} \ln\left(\frac{1-p}{p}\right)$$

Latihan 2

9.3. 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. $= \frac{y}{n} = \frac{11}{116}$

(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.

$\to y|\pi \sim \text{binomial}(n, \pi)$ ✓

$\pi \sim \text{beta}(1, 10)$

(b) posterior :

$\pi|y \sim \text{beta}(1+y, n-y+10)$

$\pi|y \sim \text{beta}(12, 115)$

(c) $E(\pi|y) = \frac{\alpha}{\alpha+\beta} = \frac{12}{12+115} = \frac{12}{127}$

$Var(\pi|y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$= \frac{12 \cdot 115}{127^2(128)}$

Prior Distribution : Conjugate

Likelihood	Prior	Posterior
$X\|\theta \sim \mathcal{N}(\theta, \sigma^2)$	$\theta \sim \mathcal{N}(\mu, \tau^2)$	$\theta\|X \sim \mathcal{N}(\frac{\tau^2}{\sigma^2+\tau^2}X + \frac{\sigma^2}{\sigma^2+\tau^2}\mu, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$
$X\|\theta \sim \mathcal{B}(n, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X \sim \mathcal{B}e(\alpha + x, n - x + \beta)$
$X_1, \ldots, X_n\|\theta \sim \mathcal{P}(\theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{G}a(\sum_i X_i + \alpha, n + \beta)$.
$X_1, \ldots, X_n\|\theta \sim \mathcal{NB}(m, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{B}e(\alpha + mn, \beta + \sum_{i=1}^n x_i)$
$X \sim \mathcal{G}(n/2, 2\theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(n/2 + \alpha, (x/2 + \beta^{-1})^{-1})$
$X_1, \ldots, X_n\|\theta \sim \mathcal{U}(0, \theta)$	$\theta \sim \mathcal{P}a(\theta_0, \alpha)$	$\theta\|X_1, \ldots, X_n \sim \mathcal{P}a(\max\{\theta_0, x_1, \ldots, x_n\}\alpha + n)$
$X\|\theta \sim \mathcal{N}(\mu, \theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta\|X \sim \mathcal{IG}(\alpha + 1/2, \beta + (\mu - X)^2/2)$
$X\|\theta \sim \mathcal{G}a(\nu, \theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta\|X \sim \mathcal{G}a(\alpha + \nu, \beta + x)$

```
###Latihan 2
pi.duga <- 11/116 #[1] 0.09482759 (a)
m1 <- 12/(12+115) #[1] 0.09448819 (c)
v1 <- (12*115)/((12+115)^2*(12+115+1)) #[1] 0.0006684388 (c)
ba <- m1+1.96*sqrt(v1) #(d)
bb <- m1-1.96*sqrt(v1) #(d)
cat(paste("95% Credible Interval: [", round(bb,5)," ", round(ba,5), "]\n", sep="")) #(d)
#95% Credible Interval: [0.04381 0.14516]
#pi=0.1 masuk selang --> tidak tolak H0 (e)
```

Latihan 3

$y|\mu \sim poisson(\mu)$

10.4. 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.

$\hat{\mu} = \frac{71}{10} = 7,1$

$E(\mu) = \frac{r}{v}$

$Var(\mu) = r/v^2$

$\mu \sim gamma(r, v)$

(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.

$\mu \sim gamma(9, \frac{3}{2})$

$E(\mu) = 6$

$\sqrt{Var(\mu)} = 2$

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

$\mu|y \sim gamma(\Sigma y_i + r, n + v) \rightarrow \mu|y \sim gamma(80, 11,5)$

$n = 10$

$\frac{r}{v} = 6 \rightarrow r = 6v$

$r = 9$

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

```
###Latihan 3
#c
m2 <- 80/11.5
v2 <- 80/(11.5^2)
ba2 <- m2+1.96*sqrt(v2)
bb2 <- m2-1.96*sqrt(v2)
cat(paste("95% Credible Interval: [", round(bb2,5)," ", round(ba2,5), "]\n", sep=""))
#95% Credible Interval: [5.43211 8.48094]
```

$\sqrt{\frac{r}{v^2}} = 2$

$\sqrt{\frac{6}{v}} = 2$

$\frac{6}{v} = 4 \rightarrow v = \frac{3}{2}$

$E(\mu|y) = \frac{r}{v}$

$Var(\mu|y) = \frac{r}{v^2}$

Komputasi Bayesian

- MCMC → membangun Rantai Markov dengan memulai dari keadaan tertentu dan membuat perubahan acak pada keadaan selama setiap iterasi. Menghasilkan Rantai Markov yang menyatu dengan target distribusi.
  - Algoritma Metropolis-Hastings → sederhana untuk diterapkan, tidak perlu menentukan distribusi bersyarat. Dapat digunakan untuk pengambilan sampel dari distribusi dimensi tinggi yang berkorelasi.
  - Algoritma Gibbs Sampling → distribusi bersama mungkin rumit untuk diambil secara langsung, tetapi mungkin dapat diambil sampel langsung dari distribusi bersyarat yang tidak terlalu rumit. Distribusi bersyarat memiliki dimensi yang lebih rendah daripada distribusi bersama dan mungkin lebih cocok untuk menerapkan teknik pengambilan sampel lainnya.

Latihan 4

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.

```r
###Latihan 4
target <- function(s){
    if(s<0 & s>1){
      return(0)
    } else {
      return(6*s*(1-s))
    }
}

set.seed(1234)
theta <- rep(0,1000)
theta[1] <- 0.2 #inisiasi
for(i in 2:1000) {
  current.theta <- theta[i-1]
  p.dtheta <- current.theta+rnorm(1,mean=0,sd=1)
  A <- target(p.dtheta)/target(current.theta)
  if(runif(1)<A){
    theta[i]<-p.dtheta
  } else {
    theta[i]<-current.theta
  }
}
```

```r
theta
plot(theta)
hist(theta)

#inferensi
mean(theta) #[1] 0.4879163
var(theta) #[1] 0.05286215

#90% credible interval
kuan <- quantile(theta,probs=c(0.05,0.95))
cat(paste("90% Credible Interval: [",
round(kuan[1],5)," ", round(kuan[2],5), "]\n",
sep=""))
#90% Credible Interval: [0.11252 0.87171]
```

Latihan 5

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.

```r
#Latihan 5
library(Bolstad)
sim1 <- poisgcp(c(12,17,10,8,23),density="gamma",params=c(3,6))
mean(sim1) #[1] 1.520086
var(sim1) #[1] 0.0007225749

#90% credible interval
kuan2 <- quantile(sim1,probs=c(0.05,0.95))
cat(paste("Approximate 90% Credible Interval: [", round(kuan2[1],5)," ",
round(kuan2[2],5), "]\n", sep=""))
#Approximate 90% Credible Interval: [1.46297 1.5412]
```

Latihan 6

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:

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

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.

```
#Latihan 6
y6 <- c(12,17,10,8,23)
lambda <- seq(0,8,by=0.001)
lap1 <- function(xx){
  return((1/4)*exp(-abs(xx-3)/2))
}
lp <- lap1(lambda)
sim2 <- poisgcp(y6,density="user",mu=lambda,mu.prior = lp)

mean(sim2) #[1] 7.739274
var(sim2) #[1] 0.05922999

#90% credible interval
kuan3 <- quantile(sim2,probs=c(0.05,0.95))
cat(paste("Approximate 90% Credible Interval: [", round(kuan3[1],5)," ",
round(kuan3[2],5), "]\n", sep=""))
#Approximate 90% Credible Interval: [7.24851 7.9851]
```

Latihan 7

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?

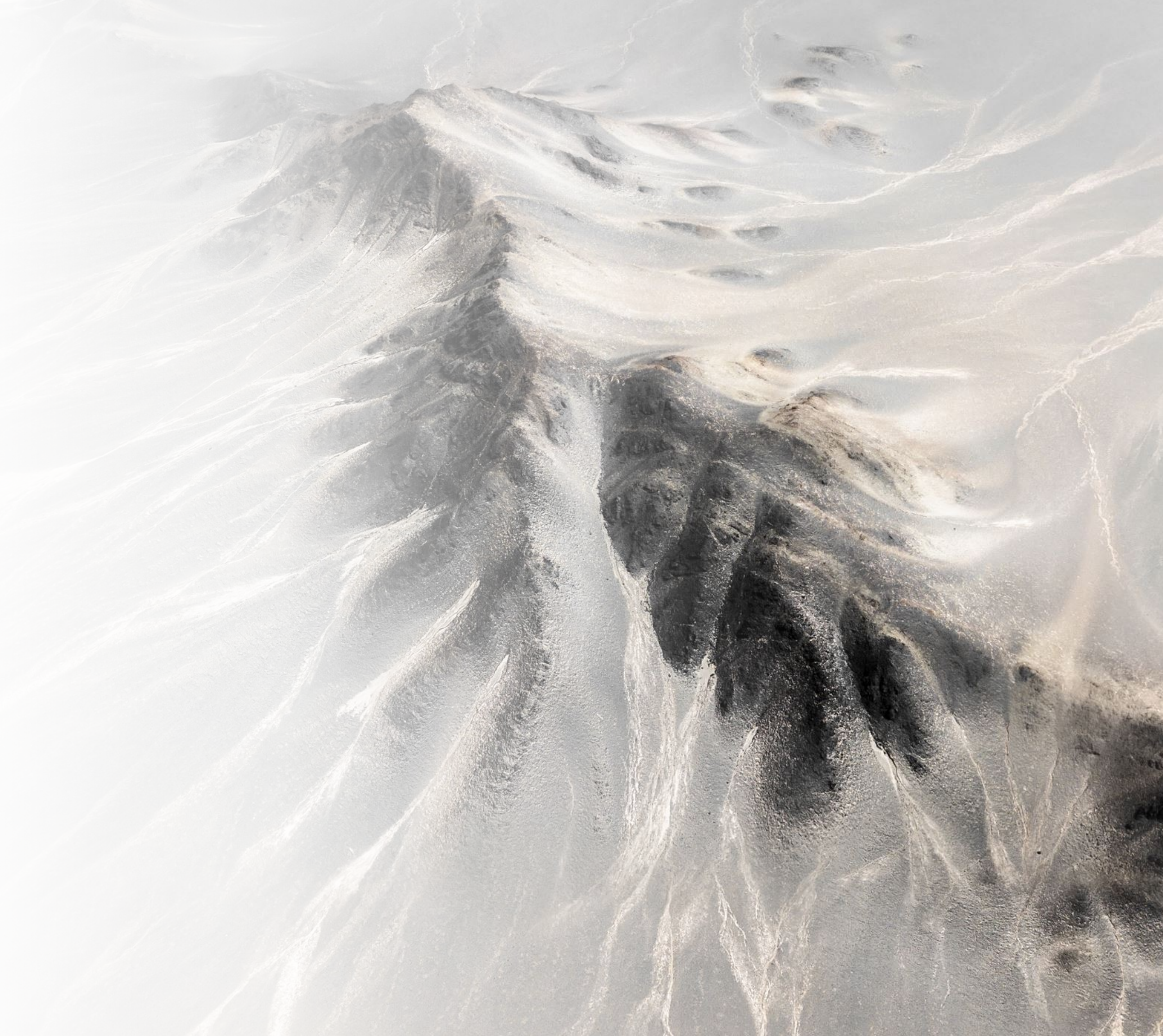
```
#Latihan 7
set.seed(12345)
y7 <- rnorm(100,5,sqrt(7))
lambda <- seq(0,8,by=0.001)
lap2 <- function(xx){
  return((1/6)*exp(-abs(xx-2)/3))
}
lp2 <- lap2(lambda)
sim3 <- normgcp(y7,density="user",mu=lambda,mu.prior = lp2)

mean(sim3) #[1] 5.619735
var(sim3) #[1] 0.08698377

#90% credible interval
kuan4 <- quantile(sim3,probs=c(0.05,0.95))
cat(paste("Approximate 90% Credible Interval: [", round(kuan4[1],5)," ",
round(kuan4[2],5), "]\n", sep=""))
#Approximate 90% Credible Interval: [5.13412 6.10435]
```

Terima kasih ☺

Metode Komputasi Bayesian (2)

Responsi 6 - STA1312 Pengantar Statistika Bayes

Septian Rahardiantoro



Review: Regresi Linier dalam Bayesian

- Regresi linier sederhana:

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

KMKT : min $\sum_{i=1}^{n} \varepsilon_i^2$

untuk $i = 1,2, \dots, n$.

- Asumsi: $\varepsilon_i \sim N(0, \sigma^2)$ dan $y_i \sim N(\underbrace{\beta_0 + \beta_1 x_i}_{\mu}, \sigma^2)$ $\rightarrow f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right)^2\right)$

- Pendekatan Bayesian:
  - β_0 dan β_1 memiliki sebaran prior, yang kemudian nilainya diduga dari sebaran posteriornya
  - Prior β_0 dan β_1 biasanya sebaran uniform(0,1) $\rightarrow p(\beta_0, \beta_1) = 1$
  - Dengan sebaran fungsi likelihood $y_1, y_2, \dots, y_n | \beta_0, \beta_1 = \prod_i f(y_i)$ $\rightarrow p(y_1 \dots y_n | \beta_0, \beta_1)$
  - Maka sebaran posterior β_0 dan β_1 dapat dicari

$$p(\beta_0, \beta_1 | y_1 \dots y_n) \propto p(\beta_0, \beta_1) \cdot p(y_1 \dots y_n | \beta_0, \beta_1)$$

Sebaran posterior β_0 dan β_1

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

Pendekatan pendugaan β_0 dan β_1 dengan MCMC untuk membangkitkan sebaran posteriornya

Latihan 1

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

$y|x \sim$ emphnormal $(\alpha_0 + \beta x, \sigma^2)$

$\sigma^2 = 3^2$

prior

$\beta \sim N(2, 2^2)$

posterior β ?

(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

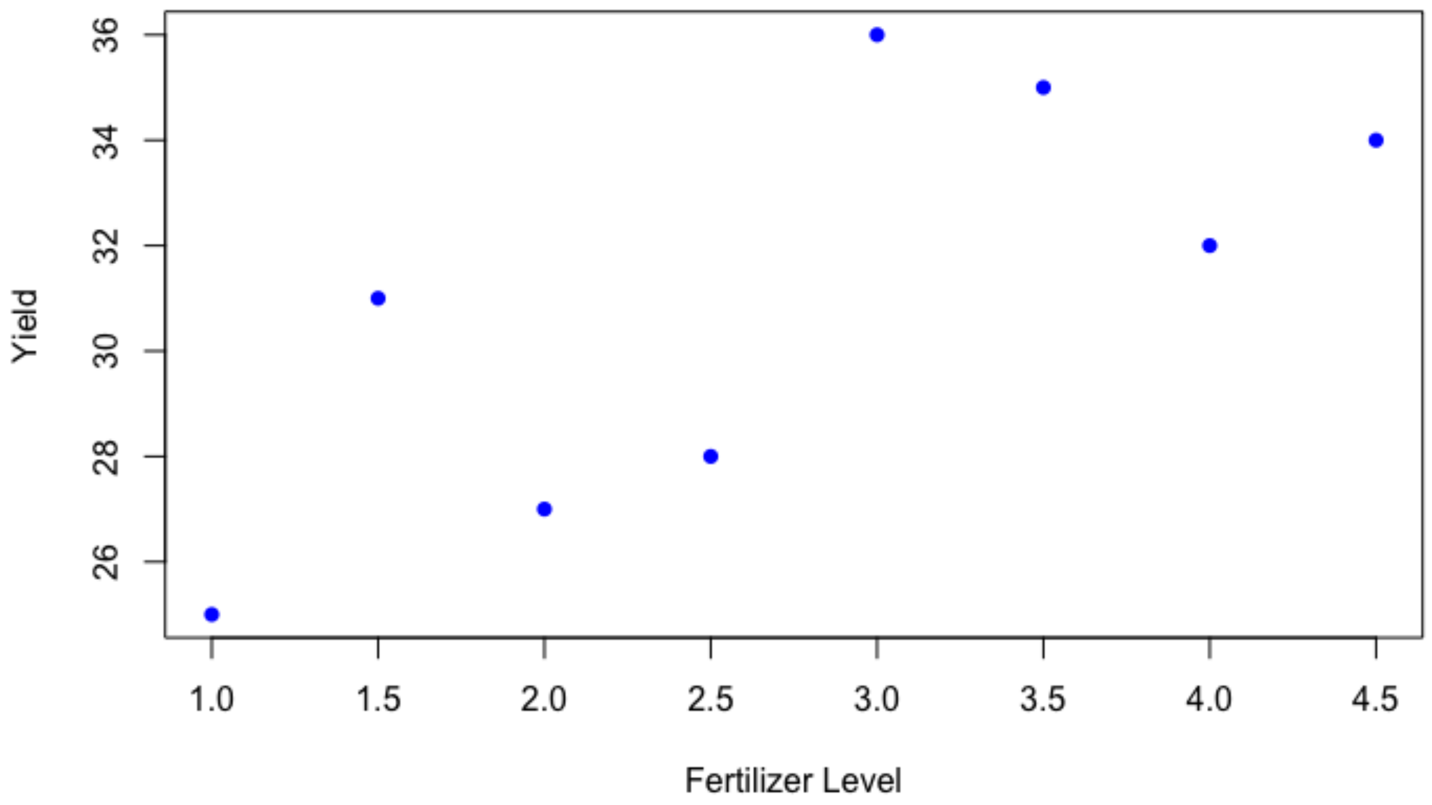
$$H_0 : \beta \le 0 \quad \text{versus} \quad H_1 : \beta > 0$$

at the 5% level of significance.

$P\left(Z \le \frac{0 - \hat{\beta}}{S_\beta}\right) = \ldots < \alpha$

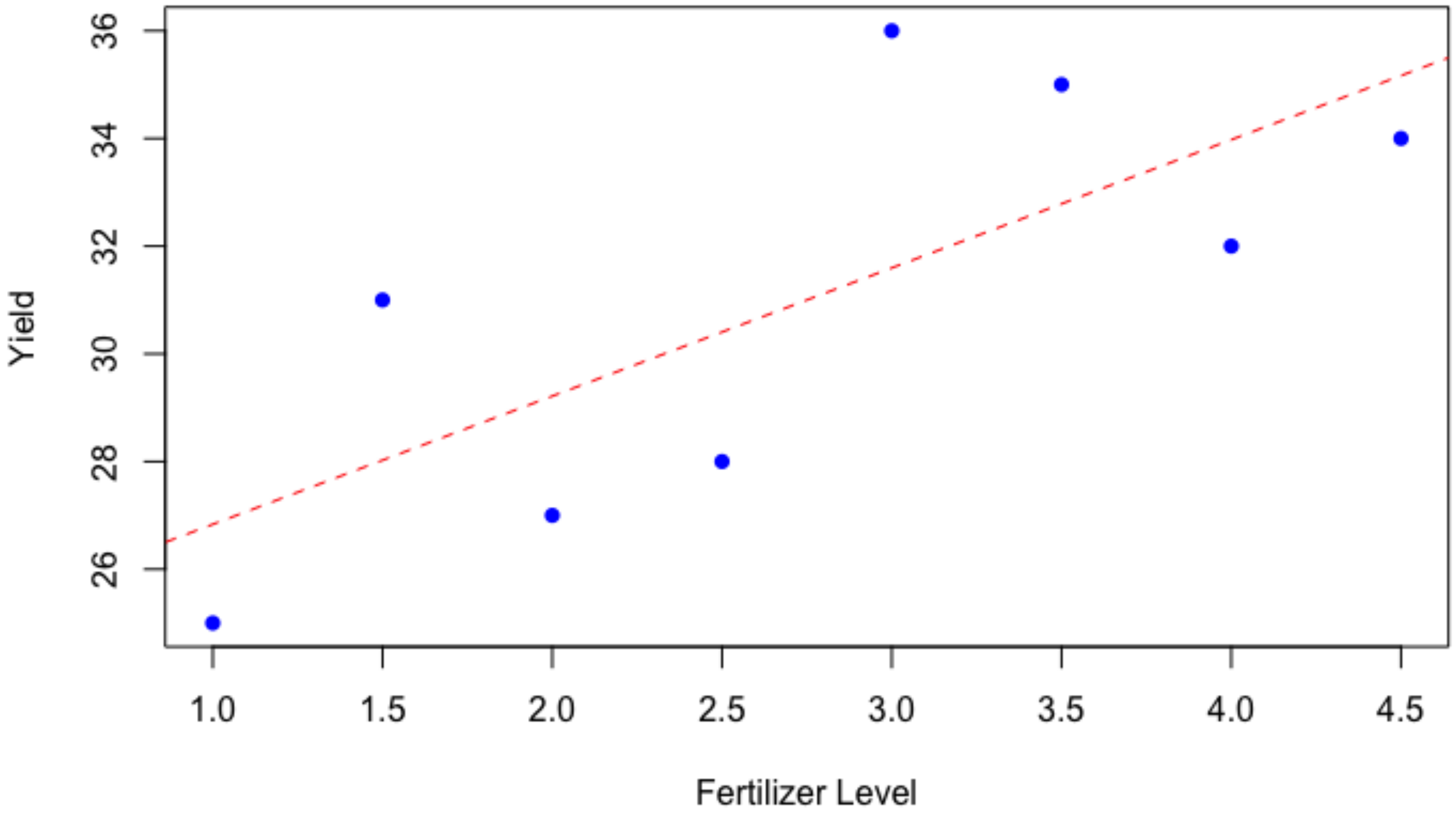
Tabel tho

```
#Latihan 1
##a
x <- c(1,1.5,2,2.5,3,3.5,4,4.5)
y <- c(25,31,27,28,36,35,32,34)
plot(x,y,xlab="Fertilizer Level",ylab="Yield",pch=16,col="blue")
```



```
##b
mod1 <- lm(y~x)
summary(mod1)

> summary(mod1)

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max 
-2.405 -2.036 -1.500  2.405  4.405 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  24.4524     2.7189   8.993 0.000106 ***
x             2.3810     0.9127   2.609 0.040185 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.957 on 6 degrees of freedom
Multiple R-squared:  0.5315,  Adjusted R-squared:  0.4534 
F-statistic: 6.806 on 1 and 6 DF,  p-value: 0.04019
```

```
##c
plot(x,y,xlab="Fertilizer Level",ylab="Yield",pch=16,col="blue")
abline(reg=mod1,col="red",lty=2)
```




```
##d
2.957^2
anova(mod1) #mean squared residuals
```

```
#e
library(Bolstad)
bayes.lin.reg(y,x,slope.prior = "normal",
              intcpt.prior = "flat",2,2,sigma=3)

 Known standard deviation:  3
           Posterior Mean Posterior Std. Deviation
           -------------- ------------------------
 Intercept:  31           1.0607
 Slope:      2.314        0.84017

##f
###credible interval
c(2.314-qnorm(0.975)*0.84017,2.314+qnorm(0.975)*0.84017) #slope
#[1] 0.6672971 3.9607029

##g
pnorm(q=(0-2.314)/0.84017)
###[1] 0.00294175 < 0.05 -> tolak H0
```

Latihan 2

A textile manufacturer is concerned about the strength of cotton yarn. In order to find out whether fiber length is an important factor in de- termining 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

$\alpha \rightarrow$ intercept

$\beta \rightarrow$ slope

$y = .$ $\quad P\left(Z \le \dfrac{0-\hat{\beta}}{S_{\hat{\beta}}}\right) < \alpha \rightarrow$ Tolak H_0

(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

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

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 $\boxed{x_{11} = 90.}$

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

```
#Latihan 2
##a
FL <- c(85,82,75,73,76,73,96,92,70,74)
St <- c(99,93,103,97,91,94,135,120,88,92)
plot(FL,St,xlab="Fiber
Length",ylab="Strength",pch=16,col="purple")

##b
mod2 <- lm(St~FL)
summary(mod2)

##c
plot(FL,St,xlab="Fiber
Length",ylab="Strength",pch=16,col="purple")
abline(reg=mod2,col="red",lty=2)

##d
7.667^2
anova(mod2) #mean squared residuals

##e
library(Bolstad)
bay2 <- bayes.lin.reg(St,FL,slope.prior = "normal",
              intcpt.prior = "flat",0,10,sigma=7.7)
```

```
##f
###credible interval
c(1.476-qnorm(0.975)*0.29041,1.476+qnorm(0.975)*0.29041)
#slope
##[1] 0.9068069 2.0451931

##g
pnorm(q=(0-1.476)/0.29041)
###[1] 1.862795e-07 < 0.05 -> tolak H0

##h
###prediction distribution
bay3 <- bayes.lin.reg(St,FL,slope.prior = "normal",
                      intcpt.prior =
"flat",0,10,sigma=7.7,
                      pred.x = 90)

##i
###prediction interval
c(116.6-qnorm(0.975)*8.6221,116.6+qnorm(0.975)*8.6221)
##[1]  99.70099 133.4990
```

Terima kasih ☺