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 2023/2024

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

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

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

The marginal pdf of Y is

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

The marginal pdf of Y is

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

posterior distribution, the distribution of p given y, is

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

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

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

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 $\theta \sim \mathrm{Exp}(\lambda)$. 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)

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

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

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 $y = (y_1, \ldots y_n)$ 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 $\rightarrow$ $p(\theta) = (1/\lambda)e^{-\theta/\lambda}, \quad \text{if } \theta > 0, \text{and } 0 \text{ elsewhere}$

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 Praktikum

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