ANALISIS REGRESI

DEPARTEMEN PENDIDIKAN DAN KEBUDAYAAN
DIREKTORAT JENDERAL PENDIDIKAN TINGGI
UNIVERSITAS AIRLANGGA

INFERENSI BAYESIAN PADA MODEL LINIER REGRESI

Ketua Peneliti :

Dra. Inna Kuswandari

Fakultas Matematika dan Ilmu Pengetahuan Alam

LEMBAGA PENELITIAN UNIVERSITAS AIRLANGGA
Dibiayai Oleh : DIP OPF Unair 1994/1995
SK.Rektor Nomor : 5655/PT03.H/N/1994
Nomor : 169

DEPARTEMEN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS AIRLANGGA
LEMBAGA PENELITIAN
Jl.Darmawangsa Dalam 2 Telp. (031) 42322 Surabaya 60286

IDENTITAS DAN PENGESAHAN
LAPORAN AKHIR HASIL PENELITIAN

1. a. Judul Penelitian : Inferensi Bayesian Pada Model Linier Regresi

   b. Macam Penelitian : (V) Fundamental, ( ) Terapan, ( ) Pengembangan

2. Kepala Proyek Penelitian
   a. Nama Lengkap Dengan Gelar : Dra. Inna Kuswandari
   b. Jenis Kelamin : W a n i t a
   c. Pangkat/Golongan dan NIP : Penata Muda/IIIa/131 933 022
   d. Jabatan Sekarang : Staf Pengajar
   e. Fakultas / Jurusan : FMIPA/Matematika
   f. Univ./Inst./Akademi : Universitas Airlangga
   g. Bidang Ilmu Yang Diteliti : Matematika

3. Jumlah Tim Peneliti : 5 (lima) orang

4. Lokasi Penelitian : Lab. Matematika FMIPA Unair

5. Kerjasama dengan Instansi Lain
   a. Nama Instansi : -
   b. A l a m a t : -

6. Jangka Waktu Penelitian : 5 (lima) bulan

7. Biaya Yang Diperlukan : Rp 1.500.000,00

8. Seminar Hasil Penelitian :
   a. Dilaksanakan Tanggal : 5 April 1995
   b. Hasil Penilaian : ~~( ) Baik Sekali~~ ~~( ) B a i k~~
   (V) S e d a n g ( ) K u r a n g

Surabaya, 19 April 1995

Mengetahui/ Mengesahkan :
a.n. Rektor
Ketua Lembaga Penelitian,

Prof. Dr. Noor Cholies Zaini
NIP. 130 355 372

DEPARTEMEN PENDIDIKAN DAN KEBUDAYAAN

DIREKTORAT JENDRAL PENDIDIKAN TINGGI

UNIVERSITAS AIRLANGGA

INFERENSI BAYESIAN PADA MODEL LINIER REGRESI

Peneliti

Dra. Inna Kuswandari

Drs. Kartono, Ms.

Dra. Rini S.

Dra. Utami Dyah Purwati

Drs. Eto Wuryanto

Lembaga Penelitian Universitas Airlangga

Dibiayai Oleh : DIP/OPF Unair 1994/1995

SK Rektor Nomor : 5655/PT.03.H/N/1994

Nomor Urut : 169

KATA PENGANTAR

Pertama-tama penulis panjatkan puji syukur kehadirat Allah SWT atas karunia-Nya, sehingga penelitian dengan judul Inferensi Bayesian Pada Model Linier Regresi dapat diselesaikan sesuai dengan waktunya.

Pada kesempatan ini penulis menyampaikan terima kasih kepada semua pihak yang telah membantu selesainya penulisan laporan ini.

Akhirnya penulis menyadari bahwa tulisan ini masih jauh dari sempurna, sehingga kritik dan saran dari para pembaca sangat diharapkan.

Surabaya, Februari 1995

Penulis

RINGKASAN PENELITIAN

Judul Penelitian	: Inferensi Bayesian Pada Model Linier Regresi
Ketua Peneliti	: Inna Kuswandari
Anggota peneliti	: Kartono Utami Dyah Purwati Rini Semiati Eto Wuryanto
Fakultas	: MIPA
Sumber biaya	: DIP Operasional Perawatan dan Fasilitas Universitas Airlangga tahun 1994/1995 SK Rektor no : 5655/PT03.H/N/1994 Tanggal : 20 Juli 1994

Penelitian ini bertujuan mengembangkan inferensi Bayesian pada model linier regresi. Pada metode statistik klasik inferensi hanya didasarkan pada informasi sampel saja, sedangkan metode Bayesian sudah memasukkan informasi awal (prior) selain informasi sampel untuk keperluan analisisnya. Metode klasik menganggap bahwa parameter populasi merupakan konstanta yang mempunyai nilai tertentu, sedangkan padametode Bayesian parameter merupakan variabel random yang mempunyai fungsi distribusi tertentu (distribusi prior). Distribusi sampel digabung dengan distribusi prior menghasilkan distribusi posterior.

Permasalahan yang dibahas dalam penelitian ini adalah bagaimanakah inferensi Bayesian pada model linier regresi kasus univariat jika dibandingkan dengan metode klasik.

Dengan studi literatur melalui penelusuran pustaka untuk prior berdistribusi normal atau eksponensial didapatkan bahwa selang kepercayaan Bayesian adalah lebih pendek jika dibandingkan dengan metode klasik. Hal ini tentu saja akan memberikan hasil yang lebih teliti daripada metode klasik. Selanjutnya masih perlu dikaji inferensi Bayesian pada model linier regresi untuk kasus multivariat.

DAFTAR ISI

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam rangka memasuki pembangunan jangka panjang II yang menitikberatkan pada pengembangan ilmu pengetahuan dan teknologi, peranan ilmu-ilmu dasar khususnya statistika tidak dapat diabaikan. Secara mudah statistik sering diartikan sebagai data, baik berupa gambar, tabel, atau diagram. Pekerjaan statistik meliputi banyak hal, mulai dari pengambilan sampel, pengumpulan dan pengolahan data, dilanjutkan dengan analisis dan interpretasinya.

Terdapat beberapa macam pembagian statistik. Berdasarkan sifatnya ada statistik deskriptif dan inferensia, berdasarkan banyaknya variabel yang diteliti ada statistik univariat dan multivariat, berdasarkan ada tidaknya distribusi populasi ada statistik parametrik dan non parametrik. Terdapat pembagian lagi yang sangat penting, yaitu statistik klasik dan Bayesian.

Metode statistik klasik menganggap bahwa parameter (besaran populasi) merupakan suatu konstanta yang mempunyai nilai tertentu. Parameter diestimasi oleh statistik berdasarkan sampel yang diamati, dan sampel merupakan satu-satunya informasi untuk menduga populasi. Pendugaan parameter populasi dilakukan dengan berbagai cara, misalnya dengan metode kuadrat terkecil, metode kemungkinan maksimum (MLE), dan metode momen.

Dalam perkembangan selanjutnya parameter tidaklah dianggap sebagai konstanta, melainkan suatu variabel random yang mempunyai fungsi distribusi tertentu. Informasi dari sampel digabung dengan informasi awal (prior) menghasilkan distribusi posterior. Inilah yang dinamakan statistik Bayesian.

Persoalan analisis regresi merupakan masalah yang sering dijumpai dalam kehidupan sehari-hari berkenaan dengan beberapa faktor (variabel bebas) yang mempengaruhi variabel tak bebas/terikat. Inferensi terhadap parameter pada analisis regresi yang sering dijumpai masih didasarkan pada metode statistik klasik, yaitu tanpa mempertimbangkan prior dan hanya berdasarkan pada informasi sampel saja. Dengan demikian diharapkan metode Bayesian akan memberikan hasil yang lebih memuaskan (teliti), sebab sudah memasukkan prior dalam analisisnya.

Pada penelitian ini akan dibahas metode Bayes pada model linier regresi, khusus untuk kasus univariat beserta pengembangan inferensinya.

1.2 Permasalahan

Permasalahan yang akan dibahas pada penelitian ini adalah bagaimanakah bentuk estimasi parameter untuk model linier regresi, khususnya untuk kasus univariat jika prior berdistribusi normal atau eksponensial.

1.3 Tujuan

Tujuan penelitian ini adalah melakukan inferensi Bayesian terhadap parameter populasi pada model linier regresi, serta membandingkannya dengan metode klasik.

1.4 Manfaat

Penelitian ini bermanfaat untuk menambah wawasan dan pengetahuan tentang metode Bayesian disamping metode klasik yang sudah dikenal, sehingga dapat digunakan/diterapkan pada bidang ilmu lainnya.

BAB II

TINJAUAN PUSTAKA

2.1. Teorema Bayes

Telah dikemukakan bahwa statistik Bayes selalu mempertimbangkan informasi prior selain informasi sampel untuk keperluan analisisnya. Distribusi prior ditentukan dengan memilih distribusi sekawan/konjugat dari distribusi sampel, misalnya jika sampel berdistribusi Bernoulli priornya diambil berdistribusi beta, dan sebagainya.

Dengan demikian analisis Bayesian menggabungkan informasi awal (prior) $p(\theta)$ dan informasi sampel y ke dalam bentuk distribusi posterior dari θ. Distribusi posterior didefinisikan sebagai distribusi bersyarat θ jika diberikan nilai observasi y, sesuai dengan teorema berikut.

Teorema Bayes

Misalkan $y' = (y_1, y_2, \ldots, y_n)$ adalah vektor n observasi yang mempunyai fungsi kepadatan probabilitas. $p(y|\theta)$ tergantung pada k parameter $\theta' = (\theta_1, \theta_2, \ldots, \theta_k)$. Misalkan pula θ mempunyai distribusi probabilitas $p(\theta)$, maka

$$p(y|\theta).p(\theta) = p(y,\theta) = p(\theta|y).p(y)$$

Jika diberikan data observasi y, probabilitas bersyarat dari θ adalah

$$p(\theta|y) = \frac{p(y|\theta).p(\theta)}{p(y)} \tag{2.1}$$

dengan

$$p(y) = \mathbb{E}\ p(y|\theta) = C^{-1} = \begin{cases} \int p(y|\theta).p(\theta)\ d\theta, & \theta \text{ kontinu} \\ \Sigma\ p(y|\theta)\ p(\theta), & \theta \text{ diskrit} \end{cases}$$

Secara umum $\mathbb{E}\ f(\theta)$ adalah ekspektasi matematik dari $f(\theta)$ terhadap distribusi $p(\theta)$. Persamaan (2.1) juga dapat ditulis sebagai

$$p(\theta|y) = c\ p(y|\theta).p(\theta) \tag{2.2}$$

Pernyataan (2.1) dan (2.2) dinamakan teorema Bayes. Dalam hal ini $p(\theta)$ menyatakan segala sesuatu yang diketahui tentang θ tanpa melihat data, dinamakan distribusi prior dari θ. Sedangkan $p(\theta|y)$ menyatakan apa yang diketahui tentang θ jika diberikan data y, dinamakan distribusi posterior dari θ.

Jika diberikan data y, $p(\theta|y)$ menyatakan fungsi dari θ, bukan y, dan dapat dianggap sebagai fungsi likelihood dari θ untuk y yang diberikan, ditulis $l(\theta|y)$. Sehingga teorema Bayes dapat ditulis sebagai

$$p(\theta|y) = l(\theta|y)\ .p(\theta) \tag{2.3}$$

Dengan kata lain teorema Bayes menyatakan bahwa distribusi probabilitas posterior dari θ untuk data y sebanding dengan hasil kali distribusi prior θ dengan likelihood dari θ jika diberikan y, yaitu

distribusi posterior $\propto$ likelihood x distribusi prior

Tujuan statistik inferensi adalah melakukan inferensi tentang parameter populasi. Dalam metode statistik klasik inferensi hanya didasarkan pada informasi sampel. Dengan kata lain, informasi prior tidak dimasukkan dalam analisis statistik. Parameter bukanlah merupakan variabel random, melainkan suatu nilai tertentu.

Statistik klasik pada umumnya mengikuti interpretasi probabilitas secara frekuensi relatif dan tidak menggunakan interpretasi secara subyektif dalam analisis statistik fromal. Pada statistik klasik parameter populasi mempunyai nilai tertentu yang tidak diketahui, sehingga pernyataan probabilitas tentang parameter tidak mempunyai arti. Pada statistik Bayes terdapat ketidakpastian tentang parameter populasi, sehingga dapat dibuat pernyataan ketidakpastian tentang parameter tersebut, karena parameter dianggap sebagai variabel random.

Estimasi titik merupakan salah satu bentuk prosedur inferensi. Metode statistik klasik berdasarkan estimasi titik seluruhnya pada informasi sampel; jadi statistik digunakan untuk mengestimasi parameter yang tidak diketahui, misalkan $\bar{x}$ digunakan untuk mengestimasi rata-rata populasi, dan sebagainya.

Metode Bayes mendasarkan estimasi titik pada distribusi posterior dan membuat pernyataan tentang $\bar{\theta}$, sedangkan metode klasik membuat pernyataan tentang y (hasil sampel).

Bentuk kedua adalah estimasi interval. Dari distribusi posterior, probabilitas setiap interval yang memuat θ dapat ditentukan. Estimasi interval pada metode klasik dinamakan *confidence interval*, sedangkan pada metode Byaes dinamakan *credible interval*.

Kriteria estimator terbaik pada statistik klasik adalah estimator yang memenuhi sifat *UMVUE* (*Uniformly Minimum Variance Unbiased Estimator*), sehingga tidak semua estimator dapat ikut berkompetisi dalam penentuan estimator terbaik. Sedangkan kriteria estimator terbaik pada metode Bayes adalah estimator yang meminimumkan fungsi resiko $R(\theta,\delta)$ untuk setiap nilai θ. Fungsi resiko adalah ekspektasi dari fungsi kerugian (*loss function*). Jadi pada metode Bayesian semua estimator dapat ikut berkompetisi dalam penentuan estimator terbaik, asalkan fungsi resikonya minimum/sekecil mungkin.

Pendekatan Bayes untuk uji hipotesis dapat dinyatakan dalam bentuk ratio distribusi posterior yang besarnya sama dengan perkalian ratio prior dan ratio likelihood. Hasil numerik metode Bayes dengan ratio likelihood klasik akan sama jika ratio prior sama dengan 1.

Jadi metode Bayes dalam melakukan inferensi selalu mempertimbangkan adanya informasi prior dan fungsi kerugian, sehingga diperoleh keputusan yang memuaskan/lebih teliti. Dapatlah dikatakan bahwa metode Bayesian memasukkan lebih banyak informasi yang dapat digunakan untuk mengambil keputusan. Inilah mengapa metode Bayesian dikatakan lebih modern daripada metode klasik.

2.2. Model Linear Regresi

Misalkan terdapat model linier regresi

$$\underset{\sim}{Y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad (2.4)$$

Persamaan (2.4) dapat dinyatakan dalam bentuk matriks sebagai berikut

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \\ x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

dengan

$\underset{\sim}{Y}$ adalah vektor observasi berukuran (n x 1)

$\underset{\sim}{X}$ adalah matriks berukuran (n x p)

$\underset{\sim}{\beta}$ adalah vektor parameter berukuran (p x 1)

$\underset{\sim}{\varepsilon}$ adalah vektor galat berukuran (n x 1)

dan $E(\varepsilon) = 0$, $V(\varepsilon) = I\sigma^2$.

Dengan metode kuadrat terkecil dan metode kemungkinan maksimum (MLE) akan dilakukan pendugaan terhadap parameter β. Selanjutnya akan dibahas prosedur inferensinya.

2.2.1. Metode Kuadrat Terkecil

Metode kuadrat terkecil adalah metode yang meminimumkan deviasi vertikal. Deviasi adalah selisih antara nilai observasi dengan nilai estimasi. Estimasi parameter dengan menggunakan metode ini tidak membutuhkan asumsi bahwa $\underset{\sim}{\varepsilon}$ harus berdistribusi normal.

Dari persamaan (2.4) maka

$$\underset{\sim}{\varepsilon} = \underset{\sim}{Y} - \underset{\sim}{X}\,\underset{\sim}{\beta}$$

$$\sum \varepsilon^2 = \underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} = (Y - X\beta)'(Y - X\beta)$$

$$= (Y' - (X\beta)')(Y - X\beta)$$

$$= Y'Y - Y'X\beta - (X\beta)'Y + (X\beta)'(X\beta)$$

$$= Y'Y - 2Y'X\beta + \beta'X'X\beta \qquad (2.5)$$

Keadaan $\varepsilon'\varepsilon$ minimum akan terjadi jika turunan parsial persamaan (2.5) terhadap β sama dengan nol, dan turunan keduanya definit positif. Bersamaan dengan itu β diganti dengan b (estimator dari β).

$$\frac{\partial\varepsilon'\varepsilon}{\partial\beta} = 0 \longrightarrow 0 - 2X'Y + 2X'Xb = 0$$

$$\frac{\partial^2\varepsilon'\varepsilon}{\partial\beta^2} = 2X'X > 0$$

Matriks X'X adalah matriks simetris, berarti elemen baris ke-i kolom ke-j sama dengan elemen baris ke-j kolom ke-i. Transpos dari matriks simetris adalah matriks itu sendiri. Jika X'X adalah matriks non singulir (mempunyai invers), maka

$$b = (X'X)^{-1}X'Y$$

Selanjutnya

$$\mathbb{E}(b) = \mathbb{E}((X'X)^{-1}X'Y)$$

$$= (X'X)^{-1}X'\ \mathbb{E}(Y)$$

$$= (X'X)^{-1}X'\ X\beta$$

$$= I\ \beta$$

$$= \beta$$

b adalah estimator linier dari β yang bersifat tak bias.

$$\text{var}(b) = \text{var}\ [(X'X)^{-1}X'Y\ ]$$

$$= (X'X)^{-1}\ \sigma^2$$

Jadi b berdistribusi normal $N(\beta, (X'X)^{-1}\ \sigma^2)$

2.2.2. Metode Kemungkinan Maksimum (MLE)

$$\underset{\sim}{Y} = \underset{\sim}{X}\ \underset{\sim}{\beta} + \underset{\sim}{\varepsilon},\ \text{dengan}\ \underset{\sim}{\varepsilon} \sim N(0, \sigma^2)$$

$$\underset{\sim}{Y} \overset{iid}{\sim} N(X\beta, \sigma^2)$$

Fungsi likelihood

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}}\ \exp\left[-\left\{ \frac{(Y-X\beta)'\ (Y-X\beta)}{2\ \sigma^2} \right\}\right]$$

Fungsi likelihood akan maksimum jika pangkat eksponen bernilai minimum yang pada dasarnya identik dengan metode kuadrat terkecil. Sehingga dengan metode MLE didapat :

$$b = (X'X)^{-1}X'Y$$

dan b berdistribusi normal $N(\beta, (X'X)^{-1}\sigma^2)$.

Dengan demikian metode kuadrat terkecil maupun metode MLE menghasilkan estimator b yang sama pada model regresi linier diatas, dan bersifat tak bias. Perbedaan utama antara kedua metode tersebut adalah mutlak adanya asumsi distribusi normal untuk $\underset{\sim}{\varepsilon}$ pada pendugaan terhadap estimator β dengan metode MLE. Asumsi distribusi normal untuk $\underset{\sim}{\varepsilon}$ pada metode kuadrat terkecil hanya digunakan untuk keperluan inferensi.

2.2.3. Inferensi b

Dari uraian dimuka di dapat bahwa b $\sim N(\beta, (X'X)^{-1}\sigma^2)$

- Untuk melakukan uji hipotesis $\underset{\sim}{\beta}$ secara keseluruhan, yaitu

  $H_0 = \beta_i = 0 \quad \forall\, i,\ i = 1,2,\ldots,p$

  $H_1 = \text{ada } \beta_i \neq 0$

  digunakan uji F dengan tabel analisis variansi sebagai berikut :

sumber	jumlah kuadrat	derajat bebas	kuadrat rata-rata
Regresi	b'X'Y(JKR)	p	KRR
Sisaan	Y'Y-bX'Y(JKS)	n-p	KRS
Total	Y'Y(JKT)	n	

JKR = Jumlah kuadrat regresi

JKS = Jumlah kuadrat sisaan

KRR = Kuadrat rata-rata regresi

$$= \frac{JKR}{p}$$

KRS = Kuadrat rata-rata sisaan

$$= \frac{JKS}{n-p}$$

Statistik uji yang digunakan ialah

$$F = \frac{KRR}{KRS}$$

Daerah kritis adalah $F > F(1-\alpha, p, n-p)$

$F(1-\alpha, p, n-p)$ diperoleh dari tabel F dengan nilai α tertentu, derajat bebas pembilang = p, dan derajat bebas penyebut = n-p.

- Untuk melakukan uji hipotesis secara individual terhadap $\underset{\sim}{\beta}$ yaitu

  $H_0 : \beta_i = 0$ untuk satu i, i=1,2,...,p

  $H_1 : \beta_i \neq 0$

  digunakan statistik uji

$$t = \frac{b_i}{sd(b_i)}$$

  dengan $sd(b_i)$ adalah akar kuadrat dari elemen diagonal ke-i matriks $(X'X)^{-1} s^2$.

  Daerah kritis adalah $|t| > t(\alpha/2, n-2)$

  dengan $t(\alpha/2, n-2)$ dapat diperoleh dari tabel t dengan derajat bebas = n-2.

- Selang kepercayaan bersama $100(1-\alpha)\%$ untuk semua parameter $\underset{\sim}{\beta}$ adalah

$$(\beta - b)' X'X (\beta - b) \leq p.s^2 F(p, n-p, 1-\alpha)$$

dengan

$$s^2 = \frac{JKS}{n-p} = KRS$$

dan $F(p, n-p, 1-\alpha)$ diperoleh dari tabel F dengan nilai α tertentu, derajat bebas pembilang = p, derajat bebas penyebut = n-p.

- Selang kepercayaan 95% untuk $\underset{\sim}{\beta}$ secara individual adalah

$$\beta_i = b_i \pm 1{,}96.\ sd(b_i)$$

dengan $sd(b_i)$ adalah akar kuadrat dari elemen diagonal matriks $(X'X)^{-1}\ s^2$ atau matriks variansi-covariansi.

BAB III

METODE PENELITIAN

Penelitian ini dilakukan dengan metode studi literatur. Adapun langkah-langkah yang dilakukan adalah sebagai berikut :

1. Mengumpulkan/mempelajari literatur yang berhubungan dengan topik yang akan dibahas.
2. Memahami masalah analisis regresi secara umum.
3. Memahami metode Bayesian disamping metode klasik, dan membuat perbandingan diantara keduanya.
4. Mengembangkan inferensi Bayesian pada model linier regresi untuk kasus univariat dengan mengambil prior berdistribusi normal atau eksponensial.

BAB IV

PEMBAHASAN

Pada bab ini akan dibahas inferensi Bayesian pada model linier regresi untuk kasus univariat.

Diberikan model regresi

$$Y = \alpha + \beta X + \varepsilon \text{ dengan } \varepsilon \sim N(0,\sigma^2)$$

Dengan metode kuadrat terkecil (meminimumkan $\Sigma\, \varepsilon^2$) diperoleh $\hat{\alpha}$ dan $\hat{\beta}$ sebagai estimator dari α dan β.

$$\hat{\beta} = \frac{\Sigma\, X_i Y_i - (\Sigma\, X_i)(\Sigma\, Y_i)}{\Sigma\, (X_i - \bar{X})^2}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

Setelah didapat $\hat{\alpha}$ dan $\hat{\beta}$ selanjutnya persamaan regresi dapat ditulis sebagai

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X$$

dengan $\hat{\beta} \sim N\left(\beta, \dfrac{\sigma^2}{\Sigma\, (X_i - \bar{X})^2}\right)$

Karena $\sigma_x^2 = \frac{1}{n}\, \Sigma\, (X_i - \bar{X})^2$, maka

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2/\sigma_x^2}{n}\right)$$

Selanjutnya menentukan prior untuk β, kemudian dilakukan inferensi Bayesian terhadap β.

4.1. Prior Berdistribusi Normal

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X, \qquad \hat{\beta} \sim N\left(\beta, \frac{\sigma^2/\sigma_x^2}{n}\right)$$

Fungsi kepadatan probabilitas untuk $\hat{\beta}$ adalah

$$f(\hat{\beta}/\beta) = \frac{1}{\sqrt{2\pi}\,\dfrac{\sigma/\sigma_x}{\sqrt{n}}} \exp\left[-\left\{\frac{(\hat{\beta}-\beta)^2}{(2\,\sigma^2/\sigma_x^2)/n}\right\}\right]$$

Distribusi prior adalah normal $\beta \sim N(\beta_o, \sigma_o^2)$

$$p(\beta) = \frac{1}{\sqrt{2\pi}\,\sigma_o} \exp\left[-\left\{\frac{(\beta-\beta_o)^2}{2\,\sigma_o^2}\right\}\right]$$

Fungsi kepadatan probabilitas bersama

$$f(\hat{\beta},\beta) = f(\hat{\beta}/\beta).p(\beta)$$

Fungsi kepadatan probabilitas marginal untuk $\hat{\beta}$ adalah

$$m(\hat{\beta}) = \int_{-\infty}^{\infty} f(\hat{\beta}/\beta).p(\beta)\, d\beta$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\,\dfrac{\sigma\sigma_o/\sigma_x}{\sqrt{n}}} \exp\left[-\left\{\frac{(\hat{\beta}-\beta)^2}{(2\,\sigma^2/\sigma_x^2)/n} + \frac{(\hat{\beta}-\beta)^2}{2\,\sigma_o^2}\right\}\right] d\beta$$

Pangkat dari eksponen =

$$= \frac{1}{(2\,\sigma^2/\sigma_x^2)/n}\left(\hat{\beta}^2 - 2\beta\hat{\beta} + \beta^2\right) + \frac{1}{2\,\sigma_o^2}\left(\beta^2 - 2\beta\beta_o + \beta_o^2\right)$$

$$= \frac{\sigma_0^2\,\hat{\beta}^2 - 2\beta\hat{\beta}\sigma_0^2 + \beta^2\sigma_0^2 + \dfrac{\sigma^2/\sigma_x^2}{n}\left[\beta^2 - 2\beta\beta_0 + \beta_0^2\right]}{\dfrac{2\,\sigma^2\sigma_0^2\,/\,\sigma_x^2}{n}}$$

$$= \frac{\beta^2\left[\sigma_0^2 + \dfrac{\sigma^2/\sigma_x^2}{n}\right] - 2\beta\left[\hat{\beta}\sigma_0^2 + \beta_0\,\dfrac{\sigma^2/\sigma_x^2}{n}\right] + \sigma_0^2\,\hat{\beta}^2 + \beta_0^2\,\dfrac{\sigma^2/\sigma_x^2}{n}}{\dfrac{2\,\sigma^2\sigma_0^2\,/\,\sigma_x^2}{n}}$$

$$= \frac{\left[\sigma_0^2 + \dfrac{\sigma^2/\sigma_x^2}{n}\right]\left[\beta^2 - 2\beta\left[\dfrac{\hat{\beta}\sigma_0^2 + \beta_0(\sigma^2/\sigma_x^2)/n}{\sigma_0^2 + (\sigma^2/\sigma_x^2)/n}\right] + \dfrac{\sigma_0^2\,\hat{\beta}^2 + \beta_0^2(\sigma^2/\sigma_x^2)/n}{\sigma_0^2 + (\sigma^2/\sigma_x^2)/n}\right]}{\dfrac{2\,\sigma^2\sigma_0^2\,/\,\sigma_x^2}{n}}$$

$$= \frac{\left[n\sigma_0^2 + \sigma^2/\sigma_x^2\right]\left[\beta^2 - 2\beta\left[\dfrac{n\hat{\beta}\sigma_0^2 + \beta_0\,\sigma^2/\sigma_x^2}{n\sigma_0^2 + \sigma^2/\sigma_x^2}\right] + \dfrac{n\sigma_0^2\,\hat{\beta}^2 + \beta_0^2\,\sigma^2/\sigma_x^2}{n\sigma_0^2 + \sigma^2/\sigma_x^2}\right]}{\dfrac{2\,\sigma^2\sigma_0^2\,/\,\sigma_x^2}{n}}$$

$$= \frac{\left(\beta - \dfrac{n\hat{\beta}\sigma_0^2 + \beta_0\,\sigma^2/\sigma_x^2}{n\sigma_0^2 + \sigma^2/\sigma_x^2}\right)^2 - \left(\dfrac{n\hat{\beta}\sigma_0^2 + \beta_0\,\sigma^2/\sigma_x^2}{n\sigma_0^2 + \sigma^2/\sigma_x^2}\right)^2 + \dfrac{n\sigma_0^2\,\hat{\beta}^2 + \beta_0^2\,\sigma^2/\sigma_x^2}{n\sigma_0^2 + \sigma^2/\sigma_x^2}}{\dfrac{2\,\sigma^2(\sigma_0^2\,/\,\sigma_x^2)/n}{n\sigma_0^2 + \sigma^2/\sigma_x^2}}$$

$$= \frac{\left(\beta - \dfrac{n\hat{\beta}\sigma_o^2\sigma_x^2 + \beta_o\sigma^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}\right)^2 - \left(\dfrac{n\hat{\beta}\sigma_o^2\sigma_x^2 + \beta_o\sigma^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}\right)^2 + \dfrac{n\sigma_o^2\hat{\beta}^2\sigma_x^2 + \beta_o^2\sigma^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}}{\dfrac{2\sigma^2\sigma_o^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}}$$

Dengan mengingat sifat bahwa

Jika $X \sim N(\mu,\sigma^2)$ maka

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}\right] dx = 1$$

sehingga

$$\therefore\ m(\hat{\beta}) = \frac{\dfrac{\sigma\sigma_o}{\sqrt{n\sigma_o^2\sigma_x^2 + \sigma^2}}}{\sqrt{2\pi}\,\dfrac{\sigma\sigma_o/\sigma_x}{n}} \exp\left[-\left\{\frac{\dfrac{n\hat{\beta}\sigma_o^2\sigma_x^2 + \beta_o\sigma^2}{n\sigma_o^2\sigma_x^2 + \sigma^2} - \left(\dfrac{n\hat{\beta}\sigma_o^2\sigma_x^2 + \beta_o\sigma^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}\right)^2}{\dfrac{2\sigma^2\sigma_o^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}}\right\}\right]$$

Distribusi posterior

$$p(\beta/\hat{\beta}) = \frac{f(\hat{\beta},\beta)}{m(\hat{\beta})}$$

$$= \frac{1}{\sqrt{2\pi}\,\dfrac{\sigma\sigma_o}{\sqrt{n\sigma_o^2\sigma_x^2 + \sigma^2}}} \exp\left[-\left\{\frac{\left(\beta - \dfrac{n\hat{\beta}\sigma_o^2\sigma_x^2 + \beta_o\sigma^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}\right)^2}{\dfrac{2\sigma^2\sigma_o^2}{n\sigma_o^2\sigma_x^2 + \sigma^2}}\right\}\right]$$

Mean dan variansi posterior adalah

$$E(\beta/\hat{\beta}) = \frac{n\hat{\beta}\sigma_0^2\sigma_x^2 + \beta_0\sigma^2}{n\sigma_0^2\sigma_x^2 + \sigma^2}$$

$$var(\beta/\hat{\beta}) = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2\sigma_x^2 + \sigma^2}$$

Jika dimisalkan $n_0 = \dfrac{\sigma^2/\sigma_x^2}{\sigma_0^2}$, maka

$$E(\beta/\hat{\beta}) = \frac{n\hat{\beta}\left(\dfrac{\sigma^2/\sigma_x^2}{\sigma^2}\right)\sigma_0^2 + \beta_0\sigma^2}{n\left(\dfrac{\sigma^2/\sigma_x^2}{n_0}\right)\sigma_x^2 + \sigma^2}$$

$$= \frac{n\hat{\beta} + n_0\beta_0}{n + n_0}$$

$$var(\beta/\hat{\beta}) = \frac{\dfrac{\sigma^2/\sigma_x^2}{n_0}\sigma^2}{n\left(\dfrac{\sigma^2/\sigma_x^2}{n_0}\right)\sigma_x^2 + \sigma^2} = \frac{\sigma^2/\sigma_x^2}{n + n_0}$$

Distribusi posterior β adalah distribusi normal dengan mean dan variansi seperti tersebut diatas.

Jadi $\beta \sim N\left(\dfrac{n\hat{\beta} + n_o\beta_o}{n + n_o}, \dfrac{\sigma^2/\sigma_x^2}{n + n_o}\right)$

Dengan demikian selang kepercayaan Bayesian 95 % bagi β adalah

$$\beta = \frac{n\hat{\beta} + n_o\beta_o}{n + n_o} \pm 1{,}96 \sqrt{\frac{\sigma^2/\sigma_x^2}{n + n_o}} \qquad (*)$$

sedangkan selang kepercayaan klasik sebesar 95 % bagi β adalah

$$\beta = \hat{\beta} \pm 1{,}96 \sqrt{\frac{\sigma^2/\sigma_x^2}{n}} \qquad (**)$$

Jika dibandingkan antara (*) dan (**) terlihat bahwa panjang interval (*) lebih pendek, karena adanya faktor pembagi sebesar $n + n_o$ yang pasti bernilai positif. Dengan kata lain nilai β lebih memusat jika dibandingkan dengan metode klasik.

4.2. Prior Berdistribusi Eksponensial

Jika $\beta \sim \exp(\lambda)$ maka distribusi priornya adalah

$$p(\beta) = \lambda\, e^{-\lambda\beta}$$

Fungsi kepadatan probabilitas bersama antara $\hat{\beta}$ dan β adalah

$f(\hat{\beta},\beta) = f(\hat{\beta}/\beta).p(\beta)$

$$= \frac{\lambda}{\sqrt{2\pi}\ \frac{\sigma/\sigma_x}{\sqrt{n}}} \exp\left[-\left\{\frac{(\hat{\beta}-\beta)^2}{(2\ \sigma^2/\sigma_x^2)/n} + \lambda\beta\right\}\right]$$

Fungsi kepadatan probabilitas marginal untuk $\hat{\beta}$

$$m(\hat{\beta}) = \int_{-\infty}^{\infty} f(\hat{\beta},\beta).p(\beta)\ d\beta$$

$$= \int_{-\infty}^{\infty} \frac{\lambda}{\sqrt{2\pi}\ \frac{\sigma/\sigma_x}{\sqrt{n}}} \exp\left[-\left\{\frac{(\hat{\beta}-\beta)^2}{(2\ \sigma^2/\sigma_x^2)/n} + \lambda\beta\right\}\right] d\beta$$

Pangkat eksponen

$$= \frac{\hat{\beta}^2 - 2\hat{\beta}\beta + \beta^2 + 2\lambda\beta\ \frac{\sigma^2/\sigma_x^2}{n}}{2\ \frac{\sigma^2/\sigma_x^2}{n}}$$

$$= \frac{n\sigma_x^2\ \hat{\beta}^2 - 2n\hat{\beta}\beta\sigma_x^2 + n\beta^2\sigma_x^2 + 2\lambda\beta\sigma^2}{2\ \sigma^2}$$

$$= \frac{\beta^2(n\sigma_x^2) - 2\beta(n\hat{\beta}\sigma_x^2 - \lambda\sigma^2) + n\hat{\beta}^2\sigma_x^2}{2\ \sigma^2}$$

$$= \frac{n\sigma_x^2\left\{\beta^2 - 2\beta\left(\frac{n\hat{\beta}\sigma_x^2 - \lambda\sigma^2}{n\ \sigma_x^2}\right) + \frac{n\hat{\beta}^2\sigma_x^2}{n\ \sigma_x^2}\right\}}{2\ \sigma^2}$$

$$= \frac{\left[\beta - \dfrac{n\hat{\beta}\sigma_x^2 - \lambda\sigma^2}{n\,\sigma_x^2}\right]^2 - \left(\dfrac{n\hat{\beta}\sigma_x^2 - \lambda\sigma^2}{n\,\sigma_x^2}\right)^2 + \hat{\beta}^2}{\dfrac{\sigma^2}{n\,\sigma_x^2}}$$

Jadi

$$m(\hat{\beta}) = \lambda\ \exp\left[-\left\{\frac{\hat{\beta}^2 - \left(\dfrac{n\hat{\beta}\sigma_x^2 - \lambda\sigma^2}{n\,\sigma_x^2}\right)^2}{\dfrac{2\ \sigma^2/\sigma_x^2}{n}}\right\}\right]$$

Distribusi posterior

$$p(\beta/\hat{\beta}) = \frac{f(\hat{\beta},\beta)}{m(\hat{\beta})}$$

$$= \frac{1}{\sqrt{2\pi}\ \dfrac{\sigma/\sigma_x}{\sqrt{n}}}\ \exp\left[-\left\{\frac{\left(\beta - \dfrac{n\hat{\beta}\sigma_x^2 - \lambda\sigma^2}{n\,\sigma_x^2}\right)^2}{\dfrac{2\ \sigma^2/\sigma_x^2}{n}}\right\}\right]$$

Distribusi posterior adalah distribusi normal dengan mean dan variansi sebagai berikut

$$E(\beta/\hat{\beta}) = \frac{n\hat{\beta}\sigma_x^2 - \lambda\sigma^2}{n\,\sigma_x^2} = \hat{\beta} - \frac{\lambda\sigma^2/\sigma_x^2}{n}$$

$$\mathrm{var}(\beta/\hat{\beta}) = \frac{\sigma^2/\sigma_x^2}{n}$$

Jadi $\beta \sim N\left(\hat{\beta} - \frac{\lambda\sigma^2/\sigma_x^2}{n}, \frac{\sigma^2/\sigma_x^2}{n}\right)$

Jika dimisalkan $k = \frac{\sigma^2/\sigma_x^2}{n}$, maka

$$\beta \sim N(\beta - \lambda k,\ k)$$

Dengan demikian selang kepercayaan Bayesian 95 % bagi β adalah

$$\beta = \hat{\beta} - \frac{\lambda\sigma^2/\sigma_x^2}{n} \pm 1,96 \sqrt{\frac{\sigma^2/\sigma_x^2}{n}} \qquad (*)$$

sedang selang kepercayaan klasik 95 % bagi β adalah

$$\beta = \hat{\beta} \pm 1,96 \sqrt{\frac{\sigma^2/\sigma_x^2}{n}} \qquad (**)$$

Jika dibandingkan antara (*) dan (**) terlihat bahwa panjang interval (*) lebih pendek, sebab ada faktor pengurang sebesar $\frac{\lambda\sigma^2/\sigma_x^2}{n}$ yang tidak pernah negatif, sehingga

$$\hat{\beta} - \frac{\lambda\sigma^2/\sigma_x^2}{n} \leq \hat{\beta}, \text{ untuk } \hat{\beta} > 0.$$

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari pembahasan dimuka dapat diambil kesimpulan bahwa selang kepercayaan Bayesian pada model linier regresi adalah lebih pendek jika dibandingkan dengan metode klasik, baik untuk prior berdistribusi normal maupun eksponensial. Hal itu berarti bahwa metode Bayesian memberikan hasil yang lebih teliti daripada metode klasik.

5.2. Saran

Masih perlu dikaji lebih lanjut inferensi Bayesian pada model linier regresi, khususnya untuk kasus multivariat.

DAFTAR PUSTAKA

[1]. Box, G.E.P., and Tiao, George C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Massachusetts.

[2]. Draper,N., and Smith,H. (1981), *Applied Regression Analysis*, second edition, John-Wiley, New York.

[3]. Judge, George G., et.al., (1982), *Introduction to The Theory and Practice of Econometrics*, John-Wiley, New York.

[4]. Seber, G.A.F., (1977), *Linear Regression Analysis*, John-Wiley, New York.

[5]. Wonnacott, Thomas H., and Wonnacott, Ronald J., (1981), *Regression : A Second Course in Statistics*, John-Wiley, New York.