LAPORAN PENELITIAN

PENAKSIRAN PARAMETER μ UNTUK DISTRIBUSI NORMAL DENGAN METODA MAXIMUM LIKELIHOOD DAN METODA BAYES

Oleh :

Drs. Herman, M.A
Dra. Nani Dianiyati
Albert Gamot Malau, S.Si

Universitas Terbuka
Lembaga Penelitian
Pusat Studi Indonesia
1999

Lembar Pengesahan
Laporan Penelitian PSI-UT

1. a. Judul Penelitian : Penaksiran Parameter μ untuk Distribusi Normal dengan Metoda Maximum Likelihood dan Metoda Bayes

   b. Bidang Penelitian : Statistika

2. Ketua Peneliti
   a. Nama lengkap dan gelar : Drs. Herman, M.A
   b. NIP : 131-628-379
   c. Golongan kepangkatan : III/c
   d. Jabatan Fungsional : Lektor Madya
   e. Fakultas/Unit Kerja : FMIPA/Statistika

3. Anggota tim peneliti
   a. Jumlah anggota : 2 orang
   b. Nama anggota/NIP/Gol. Kepangkatan :
      1. Dra. Nani Dianiyati/ 131-627-868/III/c
      2. Albert Gamot Malau, S.Si/ 132-174-680 /III/a

4. Lama Penelitian : 8 bulan

5. Biaya Penelitian : Rp.4.010.000,-
   (Empat juta sepuluh ribu rupiah).

Pondok Cabe, 11 Januari 1999

Mengetahui,
Dekan FMIPA

Dr. Djati Kerami
NIP 130 422 587

Ketua Peneliti

Drs. Herman, MA
NIP 131 628 379

Menyetujui,
Kepala PSI-UT

Dr. Tian Belawati
NIP 131 569 974

Menyetujui,
Ketua Lembaga
Penelitian UT

WBP Simanjuntak, MEd, PhD
NIP 130 212 017

Abstrak

Pada penelitian ini, parameter μ dari suatu distribusi normal ditaksir dengan cara Bayes. Di samping itu parameter yang sama juga ditaksir dengan cara konvensional menggunakan cara maximum likelihood. Ke dua penaksiran tersebut kemudian disimulasikan dengan menggunakan perangkat lunak Minitab. Hasil yang didapat menunjukkan bahwa untuk besar sampel di atas 30 ke dua penaksiran menunjukkan hasil yang hampir sama. Di samping itu, kalau distribusi prior diambil dalam bentuk normal maka distribusi posterior juga akan menghasilkan distribusi normal.

Daftar Isi

Penaksiran Parameter μ untuk Distribusi Normal dengan Metoda Maximum Likelihood dan Metoda Bayes

I. Pendahuluan

1.1. Latar Belakang

Statistik inferensi (pengambilan keputusan) digunakan untuk memprediksi "keadaan" dari suatu populasi berdasarkan data (sampel) yang diambil. Dalam statistika inferensi ini, seringkali diasumsikan bahwa distribusi populasi diketahui di mana parameter-parameternya bisa ditaksir dari sampel yang diambil. Asumsi seperti ini paling sering digunakan oleh peneliti. Teknik yang digunakan untuk keadaan seperti ini dikenal dengan nama Maximum Likelihood Estimation (MLE).

Akan tetapi ada ide lain tentang nilai dari parameter-parameter populasi. Menurut Bayes parameter-parameter berasal dari suatu distribusi. Kalau ini yang terjadi maka masalahnya tidaklah sesederhana asumsi yang diambil sebelumnya. Teknik penaksiran parameter cara Bayes dikenal dengan istilah Bayes Estimation (BE).

Sudah tentu ke dua pendekatan ini mempunyai kelebihan dan kekurangan masing-masing. Pada MLE, teknik penaksiran parameternya lebih mudah, sehingga orang banyak menggunakan teknik ini. Akan tetapi teknik ini hanya bisa digunakan bilamana distribusi populasi diketahui. Disamping itu MLE sangat sensitif terhadap data ekstrim. Pada BE karena parameter dianggap berasal dari suatu distribusi maka kesulitan pertama yang dijumpai adalah bentuk dari distribusi parameter tersebut. Kalau seorang peneliti yang sudah berpengalaman barangkali tidak terlampau sulit menentukan distribusi parameter. Tetapi bagaimana dengan peneliti pemula ?

Pada BE seorang peneliti harus menentukan distribusi awal (prior) dari parameter yang akan ditaksir. Penentuan distribusi prior ini menurut Hogg & Craig (1978) sangatlah subjektif. Semakin berpengalaman seseorang maka semakin mudahlah ia menentukan distribusi priornya. Sudah tentu penentuan distribusi prior ini harus berdasarkan alur berfikir yang logis (Bernardo & Smith, 1994).

Setelah informasi dari data (yang didapat dari sampling) digabungkan dengan informasi prior dari parameter maka akan didapat distribusi posterior dari paremeter. Dengan lebih rumitnya teknik ini maka peneliti masih jarang menggunakannya.

1.2. Perumusan Masalah

Dalam penelitian ini masalah dibatasi hanya untuk penaksiran parameter μ dari distribusi $N(\mu,\sigma^2)$. Pada pemakaian BE, bentuk distribusi prior untuk parameter ditentukan sama dengan bentuk distribusi populasi data. Dengan menggunakan penurunan matematika (setelah digabungkan dengan informasi data dari sampel), distribusi ini akan diturunkan ke distribusi posterior. Untuk menentukan taksiran parameter itu, dibutuhkan suatu fungsi keputusan yang bergantung dari fungsi kerugian.

Karena pada distribusi normal bentuk eksponensialnya mengandung bentuk kuadrat maka fungsi kerugiannya diambil dalam bentuk $L(\omega,f(y)).=(\omega - f(y))^2$. Solusi Bayes untuk bentuk ini adalah $f(y) = E(\Theta/y)$, yaitu mean dari distribusi bersyarat Θ, bila diketahui Y=y.

Masalah pertama yang muncul adalah "bagaimana bentuk distribusi posteriornya ?". Selain itu, masalah lainnya adalah "apakah nilai taksiran parameter μ dengan MLE dan dengan BE akan mempunyai nilai yang sama?". "Berapa besar data yang dibutuhkan agar ke dua pendekatan ini mempunyai nilai yang "sama" ?" Penelitian ini mencoba menjawab masalah-masalah tersebut.

1.3. Tujuan Penelitian

Tujuan pertama dari penelitian ini adalah mempelajari bagaimana parameter μ dari populasi yang berdistribusi normal ditaksir dengan metode Bayes. Penelitian ini juga bertujuan membandingkan hasil penaksiran parameter μ yang diperoleh dari cara MLE dan cara BE. Selain itu tujuan lainnya adalah mencari ukuran besarnya sampel agar ke dua taksiran yang diteliti mempunyai nilai yang "sama".

1.4. Manfaat Penelitian

Dalam prakteknya penaksiran parameter biasanya menggunakan cara MLE. Penaksiran dengan cara ini seringkali tidak memuaskan peneliti. Karena itu pada penelitian ini dikembangkan metode lain untuk menaksir parameter, yaitu dengan menggunakan cara Bayes. Diharapkan metoda ini bisa menambah variasi atau ragam untuk penaksiran parameter-parameter populasi dalam bidang-bidang yang menggunakan statistik inferensi.

II. Metodologi Penelitian

Ada beberapa hal yang harus dilakukan untuk memperoleh hasil yang ingin dicapai dalam penelitian ini. Langkah pertama adalah menentukan distribusi prior untuk menaksir nilai μ dari suatu distribusi $N(\mu,\sigma^2)$, dengan $\sigma^2 < \sim$. Karena distribusi populasinya adalah normal maka adalah sangat beralasan kalau distribusi prior untuk μ juga diambil dalam bentuk distribusi normal. Misalkan distribusi priornya memiliki fungsi kepadatan $f(\theta)$.

Dengan menggunakan sedikit teknik kalkulus maka distribusi posterior diturunkan dari distribusi prior dan distribusi data. Misalkan diambil sampel random dari populasi sebanyak *n* data. Lalu misalkan $Y = \overline{X}$ adalah mean dari sampel random tersebut. Distribusi dari sampel adalah $N(\theta,\sigma^2/n)$. Misalkan fungsi kepadatannya adalah $g(y|\theta)$. Distribusi gabungan dari distribusi prior dan y adalah :

$$k(\theta,y) = f(\theta).g(y|\,\theta)$$

Dengan menggunakan teknik integral maka akan didapat distribusi posterior $k(\theta|y)$. Lalu diambil bentuk fungsi kerugian yang akan menghasilkan taksiran untuk nilai θ menurut cara Bayes.

Penaksiran nilai mean populasi yang berdistribusi normal dengan Maximum Likelihood Estimation adalah seperti berikut ini. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random dari populasi yang berdistribusi $N(\theta,\sigma^2)$ dengan $-\infty < \theta < \infty$.

Fungsi likelihoodnya adalah $L(\theta; x_1, x_2, \ldots, x_n) = \left(\dfrac{1}{\sigma\sqrt{2\pi}}\right)^n exp\left[-\sum_{i=1}^{n}\dfrac{(x_i-\theta)^2}{2\sigma^2}\right]$.

Untuk memaksimumkan L turunkanlah L terhadap θ, dengan $\dfrac{dL}{d\theta} = 0$. Dari sini akan didapatkan nilai taksiran θ.

Hasil taksiran dari ke dua pendekatan ini lalu dibandingkan. Untuk perbandingan tersebut ambil sampel random dari suatu distribusi normal, dengan ukuran sampel yang berbeda-beda. Pengambilan sampel dilakukan dengan menggunakan perangkat lunak Minitab Release 8. Kemudian taksiran untuk masing-masing pendekatan dihitung dan dibandingkan.

III. Latar Belakang Teori

Menurut Bernardo dan Smith (1994), di dalam kehidupan sehari hari banyak sekali hal-hal yang mengandung ketidak pastian. Ketidak pastian ini mungkin berkaitan dengan masa lalu di mana pengetahuan ataupun bukti yang berkaitan dengan hal tersebut hilang ataupun tidak tersedia. Ketidak pastian dapat juga berkaitan dengan belum selesainya suatu pengembangan pada saat ini ataupun pada masa datang. Sehingga apapun keadaannya, selalu ada ketidak pastian dengan alasan ketidak yakinan akan keabsahan informasi yang dimiliki. Walaupun tingkat ketidak pastian untuk setiap orang belum tentu sama.

Banyak perasaan tentang ketidak pastian adalah tidak mendasar. Karena itu untuk hal-hal yang tidak mendasar tidak perlu mengubah urutan pikiran logis yang sudah dimiliki sebelumnya. Hal seperti ini dapat terjadi bila tidak adanya keterlibatan secara langsung antara seseorang dengan situasi yang sedang dipermasalahkannya. Dengan kata lain hal ini dapat terjadi bila seseorang merasa bahwa dia tidak mempunyai kapasitas untuk mempengaruhi keadaan, atau hasil yang akan terjadi tidak akan sesuai dengan yang diharapkan. Untuk hal-hal seperti ini tidaklah perlu berfikir hati-hati tentang ketidak pastian itu karena usaha yang diperlukan tidak akan sesuai dengan hasil yang diharapkan.

Tetapi ada juga perasaan ketidak pastian yang harus dipikirkan. Hal ini terjadi karena seseorang menghadapi problem praktis secara langsung. Ia harus memilih satu dari sekumpulan aksi-aksi yang mungkin, dimana masing-masing aksi melibatkan sekumpulan akibat yang tidak pasti dan orang ini berharap agar dapat menghindari pilihan-pilihan yang tidak logis. Sehingga orang itu perlu membuat daftar tentang keyakinannya mengenai aspek-aspek ketidak pastian tersebut (daftar ini mungkin dapat juga digunakan oleh orang lain). Karena itu, daftar ini sebaiknya dalam suatu bentuk yang dapat memungkinkan orang mengambil pilihan secara rasional.

Konsep rasional ini akan terpakai dalam menyajikan "keyakinan" atau dalam memilih aksi dalam situasi ketidak pastian. Dalam memilih yang terbaik dari sekumpulan aksi-aksi akanlah sangat menolong bilamana terdapat informasi akurat tentang akibat-akibat dari aksi-aksi yang tersedia. Tetapi dalam pengambilan keputusan yang sangat menarik adalah informasi yang akurat tidak ada sehingga orang harus meletakkan ketidak pastian sebagai suatu problem. Untuk itu haruslah ada *proses yang logis untuk mengambil keputusan dalam situasi ketidak pastian*.

Penentuan distribusi prior haruslah juga berdasarkan keputusan yang logis. Mengambil suatu bentuk distribusi prior haruslah ada dasarnya. Informasi untuk membuat distribusi prior harus ada. Kalau informasi tersebut belum ada paling tidak harus ada kerangka teori yang mendukung penentuan suatu distribusi prior. Memang distribusi yang diambil belum tentu benar, tetapi alasan pengambilannya haruslah logis. Mungkin saja pada prakteknya nanti akan terjadi penentuan bentuk distribusi prior lebih dari satu kali. Tapi seperti yang diuraikan sebelumnya, penentuan tersebut haruslah mempunyai dasar.

3.1 Teorema Bayes

Bayes menganggap bahwa parameter -parameter dari suatu distribusi merupakan suatu variabel random. Hogg dan Craig (1978) menjelaskan teorema Bayes seperti berikut ini. Misalkan X suatu variabel random yang distribusinya bergantung pada parameter θ (di mana $\theta \in \Omega$) yang tidak diketahui. Sebagai contoh, kalau θ adalah mean dari suatu distribusi normal, maka Ω adalah bilangan real. Sekarang pandang variabel random Θ, yang mempunyai distribusi peluang di Ω. Misalkan distribusi X bergantung pada ω, di mana ω diperoleh dari variabel random Θ. Fungsi distribusi peluang dari Θ diberi notasi $h(\omega)$ dengan $h(\omega) = 0$, $\omega \notin \Omega$.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random yang distribusinya bergantung pada ω dan misalkan Y adalah fungsi dari $X_1, X_2, \ldots, X_n$. Fungsi distribusi peluang Y diberikan $\Theta = \omega$ diberi notasi $g(y/\omega)$.

Distribusi gabungan dari Y dan Θ dinotasikan $k(\omega,y)$ dapat dinyatakan sebagai:

$$k(\omega,y) = h(\omega) \,.\, g(y/\omega).$$

Penulisan ini diturunkan dari sifat peluang bersyarat.

Jika Θ adalah variabel random dengan tipe kontinu, fungsi distribusi peluang marginal Y adalah :

$$k(y) = \int_{\Omega} h(\omega).g(y/\omega)\, d\omega$$

Bilamana Θ adalah variabel random tipe diskrit, fungsi distribusi peluang marginal dari Y adalah :

$$k(y) = \sum_{\Omega} h(\omega).g(y/\omega)$$

Dari ke tiga persamaan di atas, bisa diperoleh fungsi distribusi peluang bersyarat dari Θ diberikan Y = y sebagai :

$$k_1(\omega / y) = \frac{k(y,\omega)}{k_1(y)} = \frac{h(\omega).g(y/\omega)}{k_1(y)}, \qquad k_1(y) > 0$$

Hubungan di atas adalah salah satu bentuk dari formula Bayes.

Bayes menyebut fungsi $h(\omega)$ sebagai distribusi peluang prior dari Θ dan fungsi $k(\omega/y)$ disebut fungsi distribusi peluang posterior dari Θ. Fungsi ini disebut prior karena $h(\omega)$ adalah fungsi distribusi peluang awal (mula-mula) dari Θ untuk pengamatan Y. Sedangkan $k(\omega/y)$ disebut posterior karena fungsi distribusi peluang dari Θ ini muncul setelah pengamatan Y dibuat. Pada umumnya $h(\omega)$ tidak diketahui. Karena itu pemilihan $h(\omega)$ akan mempengaruhi fungsi distribusi peluang $k(\omega/y)$. Sehingga untuk menentukan $h(\omega)$, semua pengetahuan awal dari pengamatan harus diperhitungkan. Sudah tentu pemilihan $h(\omega)$ di sini sangat subjektif sekali.

Andaikata parameter ω ingin ditaksir dengan taksiran titik, maka dengan cara Bayes haruslah dipilih f sebagai fungsi keputusan sedemikian hingga $f(y)$ adalah penaksir ω. Pemilihan fungsi keputusan tersebut bergantung pada fungsi kerugian $L(\omega, f(y))$. Diharapkan nilai $f(y)$ akan "mendekati" nilai ω. Untuk itu pilihlah $f(y)$ yang meminimumkan ekspektasi fungsi kerugian. Dengan kata lain

$E(L(\Theta, f(y))/Y=y) = \int L(\omega, f(y)) \,.\, k(\omega/y)\, d\omega$ harus minimum.

3.2 Maximun Likelihood Estimation

Maximum likelihood adalah teknik yang sangat luas dipakai dalam penaksiran suatu parameter distribusi data dan tetap dominan dipakai dalam pengembangan uji -uji yang baru (Lehmann, 1986). Berikut ini akan disinggung edikit tentang penaksiran parameter ini.

Andaikan variabel random X mempunyai nilai-nilai terbilang $x_1, x_2, \ldots\ldots\ldots,$ dengan $P_\theta(x) = P_\theta\{X=x\}$. Seseorang ingin menaksir nilai yang sebenarnya dari θ tersebut dari nilai-nilai observasi $x_1, x_2, \ldots\ldots\ldots$. Sehingga untuk setiap nilai θ yang mungkin perlu dipertimbangkan probabiliti nilai x diketahui bahwa nilai θ benar. Semakin tinggi peluangnya, maka seseorang akan semakin ingin menjelaskan bahwa nilai θ dapat dijelaskan dengan x, dan θ akan semakin sering muncul. Karena itu ekspresi $P_\theta(x)$ sebagai fungsi θ untuk *x fixed* disebut *likelihood* dari θ. Simbol lain untuk likehood θ adalah $L_x(\theta)$.

Misalkan ada terbilang banyaknya keputusan-keputusan yang diformulasikan dengan fungsi keuntungan (lawan dari fungsi kerugian) dimana fungsi tersebut bernilai 0 kalau keputusannya salah dan $a(\theta) > 0$ bilamana keputusannya benar dengan nilai θ benar. Likelihood $L_x(\theta)$ diberi bobot tertentu (yang dihasilkan bilamana nilai θ benar), untuk menaksir nilai θ yang memaksimumkan $a(\theta)$, $L_x(\theta)$ dan memilih keputusan yang benar. Kemudian juga akan dipilih fungsi keputusan yang benar dengan asumsi θ benar. Penjelasan akan sama juga untuk $P_\theta(x)$ sebagai fungsi kepadatan (data kontinu).

Pada penaksiran parameter biasanya $a(\theta)$ adalah bebas dari θ. Sehingga hal ini akan menggiring orang untuk menaksir θ dengan memaksimumkan nilai $L_x(\theta)$ yang dikenal dengan *maximum likelihood estimate* dari θ.

IV. Pembahasan

4.1 Distribusi Normal

Pandang integral

$$I = \int_{-\infty}^{\infty} \exp(-y^2/2)dy$$

Menurut Hogg dan Craig (1978) integral ini ada karena integrannya adalah suatu fungsi kontinu positif yang terbatas oleh suatu fungsi yang terintegral, yaitu

$0 < \exp(-y^2/2) < \exp(-|y|+1), \ -\infty < y < \infty$ dengan

$$\int_{-\infty}^{\infty} \exp(-|y|+1)\, dy = 2e$$

Bukti

$$\int_{-\infty}^{\infty} e^{(-|y|+1)}\, dy = \int_{-\infty}^{0} e^{y+1}\, dy + \int_{0}^{\infty} e^{-y+1}\, dy$$

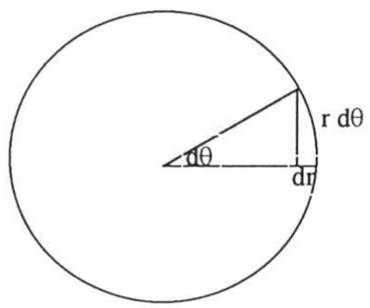
$$= \int_{-\infty}^{0} e^{y+1}\, d(y+1) - \int_{0}^{\infty} e^{-y+1}\, d(-y+1) = [e^{y+1} \Big/_{-\infty}^{0}] - [e^{-y+1} \Big/_{0}^{\infty}]$$

$$= e^1 - e^{-\infty} - e^{-\infty} + e^1 = 2e$$

Untuk mengevaluasi integral I (ingat $I > 0$), tuliskan I^2 sebagai :

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{y^2+z^2}{2}\right) dy\, dz$$

Integral ganda ini dapat diselesaikan dengan mengubahnya menjadi koordinat polar. Misalkan $y = r \cos \theta$ dan $x = r \sin \theta$.



Dengan menggunakan koordinat polar bentuk integral di atas berubah menjadi

$$I^2 = \int_0^{2\pi}\int_0^{\infty} \exp(-r^2/2)\, r\, d\theta dr = \int_0^{2\pi}\int_0^{\infty} \exp(-r^2/2)\, r\, dr\, d\theta = \int_0^{2\pi}\int_0^{\infty} \exp(-r^2/2)\, \frac{dr^2}{2}\, d\theta$$

$$I^2 = \int_0^{2\pi}\left[-\exp(-r^2/2) \Big/_{0}^{\infty}\right] d\theta = 2\pi$$

$$I = \sqrt{2\pi}$$

Karena $I = \sqrt{2\pi}$ maka $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp(-y^2/2)dy = 1$.

Definisikan suatu variabel baru dari integrasi, sebut sebagai x, sebagai berikut :

$y = \frac{x-a}{b}$, *dengan* $b > 0$, sehingga integral sebelum ini bentuknya menjadi

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

Karena $b > 0$, maka akibatnya $f(x) = \frac{1}{b\sqrt{2\pi}} exp\left[-\frac{(x-a)^2}{2b^2}\right],\ -\infty < x < \infty$

yang memenuhi persyaratan sebagai suatu fungsi kepadatan untuk suatu variabel random yang kontinu. Bentuk variabel random kontinu yang mempunyai fungsi kepadatan seperti di atas disebut mempunyai *distribusi normal* dan setiap fungsi yang bentuknya seperti fungsi di atas disebut *fungsi kepadatan normal*.

Fungsi pembangkit momen (*moment generating function*) untuk distribusi normal dapat diturunkan seperti di bawah ini.

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2b^2}\right] dx = \int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{2b^2tx + x^2 - 2ax + a^2}{2b^2}\right) dx$$

$$M(t) = \exp\left[-\frac{a^2 - (a+b^2t)^2}{2b^2}\right] \underbrace{\int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{(x-a-b^2t)^2}{2b^2}\right] dx}_{=1}$$

$$M(t) = \exp\left(at + \frac{b^2t^2}{2}\right)$$

Mean μ dan variansi σ^2 dari distribusi normal akan diturunkan dari $M(t)$.

$$M'(t) = M(t)(a + b^2 t)$$

$$M''(t) = M(t)(b^2) + M(t)(a + b^2 t)^2$$

$$\mu = M'(0) = a$$

$$\sigma^2 = M''(0) - \mu^2 = b^2 + a^2 - a^2 = b^2$$

Dengan demikian penulisan fungsi kepadatan normal dapat ditulis sebagai :

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

Bentuk ini langsung memperlihatkan secara eksplisit nilai mean dan variansinya. Demikian juga fungsi pembangkit momen $M(t)$ ditulis sebagai :

$$M(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

4.2 Taksiran Bayes

Misalkan X adalah variabel random yang berdistribusi normal dengan mean θ yang tidak diketahui, dengan variansi $\sigma^2 < \sim$. Bentuk fungsi kepadatan dari distribusi ini adalah :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right], dengan -\infty < x < \infty$$

Andaikan $x_1, x_2, x_3, \ldots\ldots, x_n$ adalah sampel berukuran n yang diambil dari populasi normal di atas dan misalkan $Y = \overline{X} = \sum_{i=1}^{n} x_i$. Fungsi kepadatan dari Y adalah:

$$f(y|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma\ /\sqrt{n}} exp\left[-\frac{(y-\theta)^2}{2\sigma^2/n}\right], dengan -\infty < y < \infty$$

Karena dalam hal ini distribusi yang diselidiki adalah distribusi normal, maka parameter θ yang akan ditaksir dianggap juga mempunyai distribusi normal. Distribusi prior ini bentuknya adalah:

$$g(\theta) \approx \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right], -\infty < \theta < \infty, \textit{ dengan } \sigma_0 \textit{ diketahui}.$$

Distribusi gabungan antara $f(y)$ dan $g(\theta)$ adalah $k(\theta, y) \approx g(\theta).f(y|\theta)$. Sehingga

$$k(\theta,y) \approx \frac{1}{\sqrt{2\pi}\sigma_0} exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] \cdot \frac{1}{\sqrt{2\pi}\sigma \ / \sqrt{n}} exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n}\right]$$

Dari $k(\theta, y)$ dapat diturunkan $k(\theta|y)$ dan menurut Hogg & Craig (1978)

$$k(\theta|y) \approx \frac{1}{\sqrt{2\pi}\sigma \ / \sqrt{n}} \frac{1}{\sqrt{2\pi}\sigma_0} exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n} - \frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right]$$

<u>Bukti</u>

$$k(\theta, y) = h(\theta).g(y|\theta)$$

$$k_1(y) = \int_{-\infty}^{\infty} k(\theta, y)\, d\theta$$

$$k(\theta|y) = \frac{k(\theta, y)}{k_1(y)}, \ dengan \ k_1(y) > 0$$

$$h(\theta) \approx \frac{1}{\sqrt{2\pi}\,\sigma_0} exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right]$$

$$g(y|\theta) = \frac{1}{\sqrt{2\pi}\,(\sigma \ / \sqrt{n})} exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n}\right]$$

$$k(\theta, y) = h(\theta).g(y|\theta)$$

$$k(\theta,y) \approx \frac{1}{\sqrt{2\pi}\,\sigma_0} \frac{1}{\sqrt{2\pi}\,(\sigma \ / \sqrt{n})} exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n}\right]$$

$$k_1(y) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_0\sigma \ / \sqrt{n}} exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n}\right]$$

$$k_1(y) = C \ (konstan)$$

$$k(\theta|y) = \frac{k(\theta,y)}{k_1(y)}$$

$$k(\theta|y) \approx \frac{\frac{1}{\sqrt{2\pi}\,\sigma_0} \frac{1}{\sqrt{2\pi}\,(\sigma \ / \sqrt{n})} exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n}\right]}{C}$$

$$k(\theta|y) \approx \frac{1}{\sqrt{2\pi}\,\sigma_0} \frac{1}{\sqrt{2\pi}\,(\sigma \ / \sqrt{n})} exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma_0^2}\right] exp\left[-\frac{(y-\theta)^2}{2\sigma^2 / n}\right]$$

Dengan mengeliminasi semua faktor konstanta, maka didapat

$$k(\theta|y) \approx exp\left[-\frac{(\sigma_0^2+\sigma^2/n)\theta^2 - 2(y\sigma_0^2+\theta_0\sigma^2/n)\theta}{2(\sigma^2/n)\sigma_0^2}\right]$$

$$k(\theta|y) \approx exp\left[-\frac{\left(\theta - \dfrac{y\sigma_0^2+\theta_0\sigma^2/n}{\sigma_0^2+\sigma^2/n}\right)^2}{\dfrac{2(\sigma^2/n)\sigma_0^2}{\sigma_0^2+\sigma^2/n}}\right]$$

Fungsi kepadatan dari distribusi posterior adalah juga normal dengan mean

$$\frac{y\sigma_0^2+\theta_0\sigma^2/n}{\sigma_0^2+\sigma^2/n} \text{ dan variansi } \frac{(\sigma^2/n)\sigma_0^2}{\sigma_0^2+\sigma^2/n}.$$

Bila fungsi kerugian yang dipakai adalah fungsi kerugian yang berbentuk kuadrat maka nilai mean tersebut merupakan solusi Bayes. Untuk penelitian ini diambil fungsi kerugian yang berbentuk kuadrat, yaitu $L(\omega,f(y)).= (\omega - f(y))^2$ dengan solusi Bayesnya adalah $f(y) = E(\Theta/y)$, yaitu mean dari distribusi bersyarat Θ, bila diketahui Y=y. Dengan demikian nilai mean di atas adalah solusi Bayesnya.

4.3 Maximum Likelihood Estimation

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random dari distribusi $N(\theta,\sigma^2)$ dengan $-\infty < \theta < \infty$ dan $\sigma^2 < \infty$.

$$L(\theta, x_1, x_2, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)\exp\left[-\sum_{i=1}^{n}(x_i-\theta)^2/2\sigma^2\right]$$

$$\ln L(\theta, x_1, x_2, \ldots, x_n) = n \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) + \left[-\sum_{i=1}^{n}(x_i-\theta)^2/2\sigma^2\right]$$

$$\frac{\partial \ln L(\theta, x_1, x_2, \ldots, x_n)}{\partial\theta} = 0 + 2\sum_{i=1}^{n}(x_1-\theta)$$

$$\sum_{i=1}^{n} x_i - n\theta = 0$$

$$\sum_{i=1}^{n} x_i = n\theta$$

$$\theta = \frac{\sum_{i=1}^{n} x_i}{n}$$

Dengan demikian $\sum_{i=1}^{n} x_i / n$ adalah nilai θ yang memaksimumkan L(θ). Statistik $\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$ disebut *maximum likelihood estimator* untuk θ.

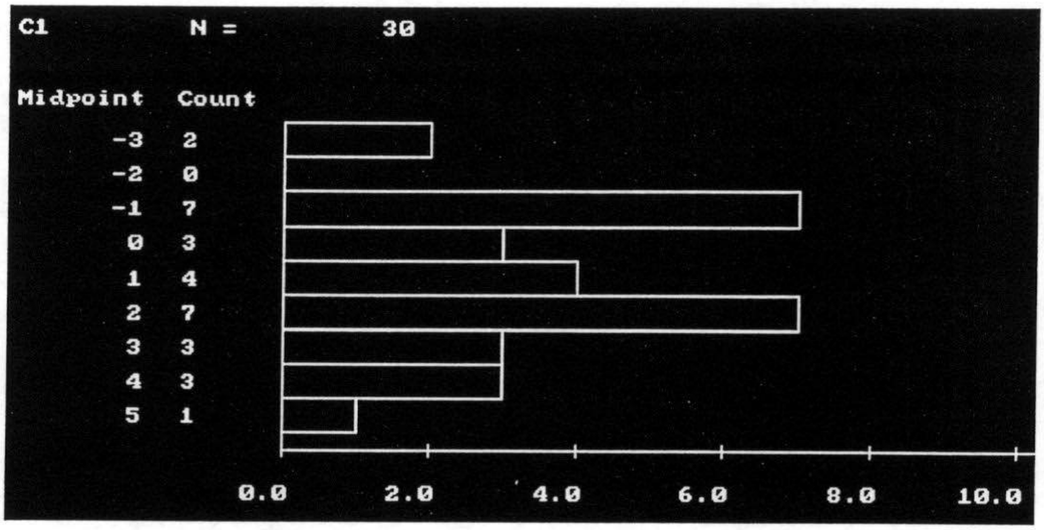
V. Simulasi Data dengan Minitab

Pada simulasi data dengan minitab ada beberapa langkah yang harus dilakukan. Langkah-langkah tersebut adalah :

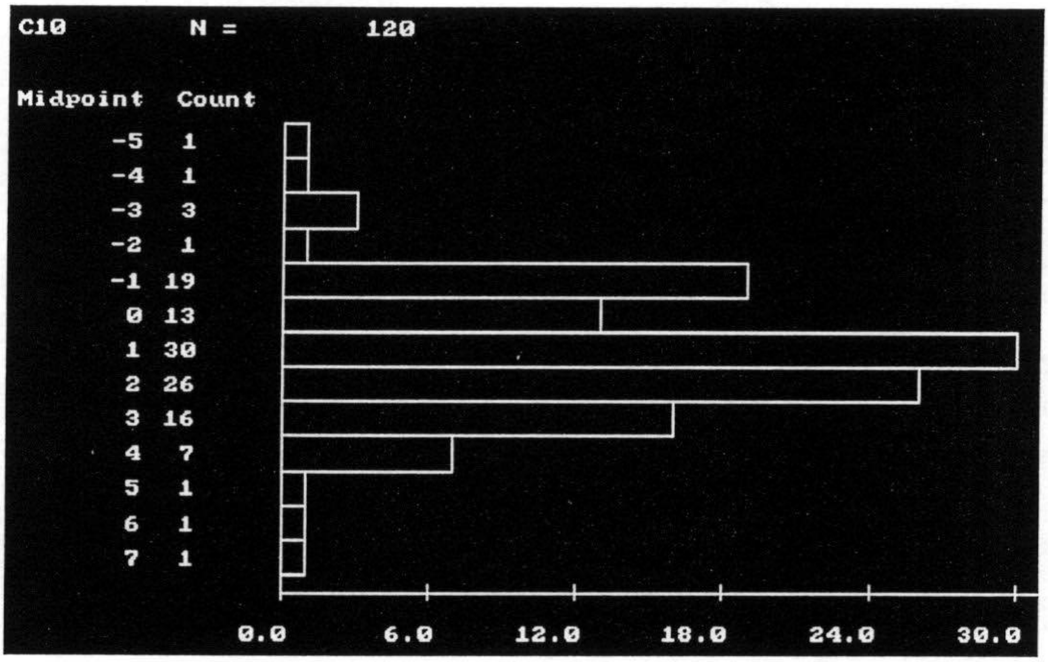
1. Menetapkan ukuran sampel; (Let K1 = N_1).
2. Menetapkan nilai mean populasi (K2), nilai mean distribusi prior diambil sama dengan nilai mean populasi; (Let K2 = N_2).
3. Menetapkan nilai variansi populasi; (Let K3 = N_3).
4. Menetapkan nilai variansi distribusi prior; (Let K4 = N_4).
5. Memilih sampel secara random dari distribusi N(K2,K3) yang besarnya K1. (Random K1,C1; Normal K2, K3).
6. Menghitung nilai mean sampel; (Let K5 = mean (C1).
7. Menghitung nilai mean distribusi posterior; (Let K6 = $\frac{(K5*K3)+(K2*K4)/K1}{K3+(K4/K1)}$).
8. Menghitung selisih antara K6 dengan K5; (Let K7 = K5 - K6).
9. Menghitung nilai standar deviasi distribusi posterior; (Let K8 = stdv (C1)).
10. Menghitung nilai variansi sampel; (Let K9 = K8 * K8).
11. Menghitung nilai variansi distribusi posterior; (Let K10 = $\frac{(K4/K1)*K3}{K3+(K4/K1)}$)

Untuk satu kali simulasi, waktu yang dibutuhkan tidak lebih dari 2 detik. Hasil dari simulasi ini dapat dilihat pada Tabel I dan Tabel II.

Histogram dari data (n = 30) yang diambil secara random dari N(1,4).



Histogram dari data (n = 120) yang diambil secara random dari N(1,4).



Walau ke dua histogram di atas berasal dari distribusi normal, akan tetapi tampak bahwa untuk ukuran sampel yang kecil bentuknya tidak terlalu simetris. Sedangkan untuk ukuran sampel yang besar bentuk simetrisnya lebih tegas.

Tabel I
Nilai Taksiran Bayes dan Taksiran Maximum Likelihood
$\mu = 1, \sigma_0^2 = \sigma^2 = 4, \theta_0 = 1$

n	$\overline{X}$	S^2	MPost	VPost	$\overline{X}$ – MP
5	2,593	1,541	2,32736	0,66667	0,26547
10	1,656	8,520	1,59640	0,36364	-0,05960
20	1,464	3,359	1,44190	0,19048	-0,02210
30	0,966	4,482	0,96710	0,12903	0,00110
40	1,261	4,113	1,25460	0,09756	-0,00640
50	1,337	3,197	1,33040	0,07843	-0,00660
60	1,041	5,574	1,04030	0,06557	-0,00070
70	1,005	3,729	1,00490	0,05634	-0,00010
80	0,755	3,595	0,75802	0,04938	0,00302
90	0,801	3,378	0,80319	0,04396	0,00219
100	0,546	4,048	0,55055	0,03960	0,00450
110	1,105	4,813	1,10410	0,03604	-0,00090
120	1,110	3,531	1,10910	0,03306	-0,00090

Catatan: $\overline{X}$ adalah nilai rata-rata dari sampel yang diambil.
S^2 adalah nilai variansi dari sampel yang diambil.
MPost adalah nilai rata-rata dari distribusi posterior.
VPost adalah variansi distribusi posterior.

Tampak bahwa mulai dari n = 30 perbedaan antara taksiran Bayes dan taksiran Maximum Likelihood tidak banyak berubah. Variansi dari distribusi posterior akan semakin kecil dengan bertambahnya nilai n. Hal ini adalah akibat dari formula variansi yang berbanding terbalik dengan nilai n.

Bagaimana kalau untuk satu ukuran sampel dicoba beberapa kali. Untuk itu perhatikan tabel berikut ini di mana pada ukuran sampel n = 5, 10, 30, 50, 100 dicoba lebih dari satu kali. Pada percobaan ini mean populasi diambil sama besarnya dengan mean distribusi prior yaitu 3. Sedangkan variansi populasi diambil sama besarnya dengan variansi distribusi prior yaitu sebesar 16.

Tabel II
Nilai Taksiran Bayes dan Taksiran Maximum Likelihood
$\mu = 3$, $\sigma_0^2 = \sigma^2 = 16$, $\theta_0 = 3$

n	$\overline{X}$	S^2	MPost	VPost	$\overline{X}$ – MP
5	3,5021	38,2949	3,4185	2,6667	0,083695
5	3,5598	38,4450	3,4665	2,6667	0,093297
5	-1,6696	1,8485	-0,8913	2,6667	-0,778266
10	3,6078	17,7516	3,5526	1,4545	0,055259
10	2,2782	7,8959	2,3407	1,4545	-0,065925
10	2,8277	9,0861	2,8434	1,4545	-0,015662
20	2,4117	20,6686	2,4397	0,7619	-0,028016
20	2,6730	19,5036	2,6885	0,7619	-0,015572
20	1,2336	16,1426	1,3177	0,7619	-0,084114
30	3,0217	19,5193	3,0211	0,5161	0,000702
30	4,3785	17,0183	4,3340	0,5161	0,044468
40	3,7599	13,5954	3,7414	0,3902	0,018536
40	3,2983	14,2059	3,2911	0,3902	0,007277
40	3,4312	13,6803	3,4206	0,3902	0,010516
50	2,1592	23,6914	2,1756	0,3137	-0,016487
50	2,3694	13,8028	2,3818	0,3137	-0,012364
50	3,4266	18,1557	3,4182	0,3137	0,008365
100	3,5739	16,4282	3,5683	0,1584	0,005682
100	2,7608	19,2006	2,7632	0,1584	-0,002368

Catatan: $\overline{X}$ adalah nilai rata-rata dari sampel yang diambil.
S^2 adalah nilai variansi dari sampel yang diambil.
MPost adalah nilai rata-rata dari distribusi posterior.
VPost adalah variansi distribusi posterior.

Tampak juga bahwa untuk satu ukuran sampel yang dicoba beberapa kali, ternyata nilai selisih antara taksiran MLE dan taksiran Bayes akan lebih stabil untuk $n \geq 30$. Untuk $n < 30$ fluktuasi perbedaan nilai selisih ke dua taksiran masih cukup tinggi.

VI. Kesimpulan

- Walaupun untuk menentukan distribusi prior dari parameter adalah sulit, tetapi penaksiran parameter dengan metoda Bayes tampaknya lebih menjanjikan karena peneliti tidak perlu tahu tentang distribusi awal dari populasi.
- Untuk populasi yang distribusinya diketahui, penaksiran parameter akan lebih mudah bila menggunakan MLE.
- Kalau populasi dianggap berdistribusi normal dan distribusi prior dari parameter juga berdistribusi normal, maka distribusi posteriornya juga akan berdistribusi normal.
- Simulasi menunjukkan bahwa untuk ukuran sampel $n \geq 30$, bila populasi data distribusinya diketahui, maka MLE dan BE akan mempunyai nilai yang hampir sama besarnya.
- Simulasi data (dengan perangkat lunak Minitab) untuk melihat hasil perbandingan MLE dan BE pada satu ukuran sampel ternyata hanya membutuhkan waktu tidak lebih dari 2 detik.

VII. Saran

Bila distribusi populasi adalah normal dan distribusi prior untuk parameter diambil normal juga, maka hasil yang didapat untuk distribusi posterior adalah juga normal. Bagaimana kalau distribusi prior untuk parameter diambil bukan berbentuk normal ? Apakah bentuk dari distribusi posteriornya ? Dengan demikian seberapa dekatkah nilai taksiran MLE dan BE ? Untuk menjawab pertanyaan-pertanyaan di atas disarankan untuk membuat penelitian lain yang mengambil distribusi prior yang bentuknya bukan distribusi normal.

VIII. Literatur yang akan dipakai

Bernardo, J.M & Smith, A.F.M (1994). *Bayesian Theory*. John Willey & Sons: Biddles Ltd, Guildford and King's Lynn: England.

Hogg, R.V & Craig, A.T (1978). *Introduction to Mathematical Statistics*. Macmillan Publishing C0., Inc.: New York.

Lehmann.E.L (1986). *Testing Statistical Hypotheses*. 2nd. Ed. : John Willey & Sons,Inc: New York.

Minitab Reference Manual, Release 8, PC Version.

Ryan, B.F, Joiner, B.L & Ryan, T.A (1985). *Minitab Handbook*. PWS-KENT Publishing Company : Boston.

Riwayat Hidup Peneliti

1.	Nama	: Drs. Herman, M.A
	Unit	: Jurusan Statistika FMIPA - UT
	Tempat / Tgl. Lahir	: Palembang / 25 Mei 1956
	Pendidikan	: S1, Matematika, ITB, 1984 S2, Educational Psychology, University of Victoria, 1993
	Pengalaman Penelitian	: - A Study of Relationship between Achievement in Prerequisite Course in Applied Statistics and Economics Study Programs at Universitas Terbuka, 1993. - Uji Umur dan Penaksiran Keandalan Alat-alat yang Berdistribusi Eksponensial dengan Pendekatan Bayes, 1997. - Item Analisis, Statistika dan Penurunan Rumus Item Analisis, 1997. - Penaksiran Parameter μ untuk Distribusi Normal dengan Metoda Maximum Likelihood dan Metoda Bayes, 1999.
2.	Nama	: Dra. Nani Dianiyati
	Unit	: Pusat Komputer/ Statistika FMIPA - UT
	Tempat / Tgl. Lahir	: Ciamis / 25 September 1959
	Pendidikan	: S1, Matematika, Universitas Pajajaran, 1984
	Pengalaman Penelitian	: - Item Analisis, Statistika dan Penurunan Rumus Item Analisis, 1997. - Penaksiran Parameter μ untuk Distribusi Normal dengan Metoda Maximum Likelihood dan Metoda Bayes, 1999.

3.	Nama	: Albert Gamot Malau, S.Si
	Unit	: Statistika FMIPA - UT
	Tempat / Tgl. Lahir	: Pematang Siantar / 30 April 1970
	Pendidikan	: S1 Matematika, Universitas Sumatra Utara, 1996
	Pengalaman Penelitian	: - Penaksiran Parameter μ untuk Distribusi Normal dengan Metoda Maximum Likelihood dan Metoda Bayes, 1999.

Ucapan Terima Kasih

Sehubungan dengan kegiatan penelitian ini , penulis mengucapkan banyak terima kasih kepada pihak-pihak yang sudah membantu kelancaran penyelesaian penelitian ini.

Pertama-tama penulis mengucapkan terima kasih kepada Dekan FMIPA Universitas Terbuka, Dr. Djati Kerami yang sudah meluangkan waktunya dalam memberikan kritik dan saran untuk penelitian ini.

Tak lupa pula ucapan terimakasih penulis tujukan kepada Dr Ujiana Pasaribu, dosen di jurusan Matematika ITB Bandung yang telah memberikan banyak masukan dan kritik untuk penelitian ini.

Selain itu juga penulis mengucapkan banyak terima kasih kepada Ketua Lembaga Penelitian, WBP Simanjuntak, M.Ed, PhD dan Kepala Pusat Studi Indonesia, Dr. Tian Belawati yang telah memberi kesempatan kepada penulis untuk mengikuti seleksi pembiayaan penelitian di Universitas Terbuka ini.

Juga penulis tak lupa mengucapkan terima kasih kepada rekan-rekan yang sudah membantu penelitian ini, baik berupa kritik dan saran ataupun peminjaman buku-buku yang diperlukan.