# MENAKSIR PARAMETER μ DARI N( μ,$\sigma^2$ ) DENGAN METODE BAYES

Hartayuni Saini[1]

[1]Jurusan Matematika, FMIPA-UNTAD. e-mail: yunh3_chendist@yahoo.co.id

**Abstrak**

Untuk menaksir nilai μ dari N(μ,$\sigma^2$) umumnya digunakan teknik Maximum Likelihood. Jarang sekali penggunaan Metode Bayes untuk menaksir nilai parameter μ tersebut. Cara Bayes ini memang lebih rumit dari cara Maximum Likelihood. Tulisan ini membahas penaksiran μ dari N(μ,$\sigma^2$) dengan cara Bayes. Solusi Bayes ini selanjutnya akan dibandingkan dengan solusi Maximum Likelihood dengan menggunakan hasil simulasi data dengan menggunakan perangkat lunak Manitab yang menunjukkan bahwa untuk ukuran data ≥ 30 ke dua pendekatan ini mempunyai nilai yang “sama”.

**Kata Kunci** : Maksimum likelihood estimasi, metode bayes, Bayes Estimation

## I. Pendahuluan

Statistik inferensi digunakan untuk memprediksi “keadaan” dari suatu populasi berdasarkan sampel yang diambil. Dalam statistika inferensi ini, seringkali diasumsikan bahwa distribusi populasi diketahui. Teknik yang digunakan untuk menaksir nilai parameter bila distribusi populasi diketahui adalah dengan Maximum Likelihood Estimation (MLE). Akan tetapi ada pendekatan lain yang diperkenalkan oleh Bayes untuk menaksir nilai parameter-parameter populasi. Menurut Bayes, parameter populasi berasal dari suatu distribusi sehingga nilainya tidaklah tunggal. Karena itu penaksiran nilai parameternya menjadi lebih rumit dari pada MLE. Teknik penaksiran parameter cara Bayes dikenal dengan istilah Bayes Estimation (BE).

Masing-masing pendekatan ini sudah tentu mempunyai kelebihan dan kekurangan masing-masing. Pada MLE, teknik penaksiran parameternya lebih mudah, sehingga orang banyak menggunakan teknik ini. Akan tetapi teknik ini hanya dapat digunakan bilamana distribusi populasi diketahui. Selain itu MLE sangat sensitif terhadap data ekstrim. Data ekstrim ini sangat berpengaruh terhadap nilai-nilai mean ataupun variansi. Pada BE, karena nilai parameter berasal dari suatu distribusi maka kesulitan pertama yang dijumpai adalah bagaimana bentuk 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 pengambilan sampel) digabungkan dengan informasi prior dari parameter, akan didapat distribusi posterior dari paremeter.

Tulisan ini bertujuan mencari solusi Bayes untuk $\mu$ dari $N(\mu,\sigma^2)$. Hasil solusi Bayes ini lalu dibandingkan dengan solusi Maximum Likelihood melalui simulasi data dengan menggunakan perangkat lunak Minitab.

## II. Teori Penaksiran Bayes Dan Maximum Likelihood

### II.1. Bayes Estimation (Penaksiran 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 $\theta \epsilon \Omega$) 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 dari Θ. Fungsi distribusi peluang dari Θ diberi notasi h(ω) dengan h(ω) = 0 , $\omega \,\epsilon\, \Omega$.

Misalkan $X_1, X_2, \ldots\ldots\ldots., X_n$ adalah sampel random yang distribusinya bergantung pada ω dan misalkan Y adalah fungsi dari $X_1, X_2, \ldots\ldots\ldots., X_n$ . Fungsi distribusi peluang Y diberikan Θ = ω diberi notasi g(y/ω). Distribusi gabungan dari Y dan Θ dinotasikan k(ω,y) dapat dinyatakan sebagai:

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

Fungsi distribusi peluang marginal y, dapat dicari dengan cara :

$$k(y) = \int_\Omega h(\omega) \cdot g\left(\frac{y}{\omega}\right) d\omega \quad \text{atau} \quad k(y) = \sum_\Omega h(\omega) \cdot g\left(\frac{y}{\omega}\right)$$

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

$$k\left(\frac{\omega}{y}\right) = \frac{k(y,\omega)}{k(y)} = \frac{h(\omega)\cdot g\left(\frac{y}{\omega}\right)}{k(y)},\ k(y) > 0$$

Bayes menyebut fungsi h(ω) sebagai distribusi peluang prior dari Θ dan fungsi k(ω /y) disebut fungsi distribusi peluang posterior dari Θ. Fungsi ini disebut prior karena h(ω) adalah fungsi distribusi peluang awal (mula-mula) dari Θ untuk pengamatan Y. Sedangkan k(ω/y) disebut posterior karena fungsi distribusi peluang dari Θ ini muncul setelah pengamatan Y dibuat. Pada umumnya h(ω) tidak diketahui. Karena itu pemilihan h(ω) akan mempengaruhi fungsi distribusi peluang k(ω/y). Sehingga untuk menentukan h(ω), semua pengetahuan awal dari pengamatan harus diperhitungkan. Sudah tentu pemilihan h(ω) 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(ω,f(y))*. Diharapkan nilai *f(y)* akan "mendekati" nilai ω. Untuk itu pilihlah *f(y)* yang meminimumkan ekspektasi fungsi kerugian. Dengan kata lain

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

### II. 2. Maximun Likelihood Estimation (Penaksiran Maximum Likelihood)

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 sedikit tentang penaksiran parameter ini.

Andaikan variabel random X mempunyai nilai-nilai terbilang $x_1, x_2,$..........., dengan $P_\theta(x)$ = $P_\theta\{X=x\}$. Seseorang ingin menaksir nilai yang sebenarnya dari θ tersebut dari nilai-nilai observasi $x_1, x_2$........... 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(θ) > 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(θ).$L_x(\theta)$ 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(θ) 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.

## III. Pembahasan

Pada tulisan ini distribusi populasi data diambil berbentuk normal. Hal ini adalah untuk memenuhi syarat dalam penaksiran dengan MLE dan juga untuk memudahkan penurunan formula-formula matematiknya. Akan tetapi pada situasi yang sebenarnya, kalau distribusi data tidak diketahui maka bentuk distribusi ini haruslah ditaksir.

### III.1. Bayes Estimation

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 :

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

Andaikan $x_1$, $x_2$, $x_3$, ...... , $x_n$ adalah sampel berukuran n yang diambil dari populasi normal di atas dan misalkan

$Y = \bar{X} = \sum_{i=1}^{n} x_i$ . Fungsi kepadatan dari Y adalah:

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

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

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

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

Dari $k(\theta, y)$ dapat diturunkan $(\theta|y)$ . Menurut Hogg & Craig (1978) distribusi posteriornya adalah :

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

Dengan mengeliminasi semua faktor konstanta, maka didapat

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

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

Fungsi kepadatan dari distribusi posterior adalah juga normal dengan mean $\frac{y\sigma_0^2+\frac{\theta_0\sigma^2}{n}}{\sigma_0^2+\frac{\sigma^2}{n}}$ dan variansi $\frac{(\sigma^2/n)\sigma_0^2}{\sigma_0^2+\frac{\sigma^2}{n}}$ . Bila fungsi kerugian yang digunakan berbentuk kuadrat maka nilai mean merupakan solusi Bayesnya. Karena pada distribusi normal, fungsi eksponensialnya berbentuk kuadrat maka untuk tulisan ini fungsi kerugiannya juga diambil dalam bentuk kuadrat, yaitu $\left(\omega, f(y)\right) = (\omega - f(y))^2$ . Solusi Bayesnya adalah f(y) = E(Θ/y), yaitu mean dari distribusi bersyarat Θ, bila diketahui Y=y. Dengan demikian nilai mean di atas adalah solusi Bayesnya.

### III.2. Maximum Likelihood Estimation

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

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

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

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

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

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

### III. 3. Simulasi Data dengan Minitab

Untuk membandingkan nilai dari ke dua penaksiran di atas maka perlu dilakukan simulasi dengan data. Untuk tulisan ini simulasi data menggunakan perangkat lunak Minitab. Perangkat lunak ini dapat membuat data dengan berbagai macam distribusi. Pada Minitab, pemakai juga dapat membuat program untuk penghitungan nilai-nilai variabel yang dinginkan.

Untuk simulasi data dengan minitab ini, ada beberapa langkah yang harus dilakukan. Langkah-langkah tersebut adalah sebagai berikut:

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;
$\left(Let\ K6\ =\ \frac{(K5\ *K3)+\ (K2*K4)/K1}{K3+\ (K4/K1)}\right)$
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; $\left(Let\ K\ =\ \frac{(K4/K1)*K3}{K3+\ (K4/K1)}\right)$

Hasil dari simulasi ini dapat dilihat pada Tabel 1 dan Tabel 2.

**Tabel 1 Nilai Taksiran Bayes dan Taksiran Maximum Likelihood**

$\boldsymbol{\mu = 1,\quad \sigma_0^2\ =\ \sigma^2\ =\ 4,\quad \theta_o = 1}$

| n | $\bar{X}$ | $S^2$ | MPost | VPost | $\bar{X}\ -\ MP$ |
|---|---|---|---|---|---|
| 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: $\bar{X}$ 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. Memang 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 2 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 sebesar 16.

**Tabel 2. Nilai Taksiran Bayes dan Taksiran Maximum Likelihood**

$\boldsymbol{\mu = 3}, \quad \boldsymbol{\sigma_0^2 = \sigma^2 = 16}, \quad \boldsymbol{\theta_o = 3}$

| n | $\bar{X}$ | $S^2$ | MPost | VPost | $\bar{X} - MP$ |
|---|---|---|---|---|---|
| 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: $\bar{X}$ 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 ≥ 30 (perhatikan kolom : $\bar{X} - MP$ ). Untuk n < 30 fluktuasi perbedaan nilai selisih ke dua taksiran masih cukup tinggi.

## IV. Kesimpulan

1. Untuk populasi yang distribusinya diketahui, penaksiran parameter akan lebih mudah bila menggunakan MLE.
2. Kalau populasi dianggap berdistribusi normal dan distribusi prior dari parameter juga berdistribusi normal, maka distribusi posteriornya juga akan berdistribusi normal.
3. Simulasi menunjukkan bahwa untuk ukuran sampel n ≥ 30, bila populasi data distribusinya diketahui, maka MLE dan BE akan mempunyai nilai yang hampir sama besarnya.
4. 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.

## V. Daftar Pustaka


1. Bernardo, J.M & Smith, A.F.M (1994). *Bayesian Theory*. John Willey & Sons: Biddles Ltd, Guildford and King's Lynn:England.
2. Hogg, R.V & Craig, A.T (1978). *Introduction to Mathematical Statistics*. Macmillan Publishing C0., Inc.: New York.
3. Lehmann.E.L (1986). Testing Statistical Hypotheses. 2nd. Ed. : John Willey & Sons,Inc: New York.
4. Minitab Reference Manual, Release 8, PC Version.
5. Ryan, B.F, Joiner, B.L & Ryan, T.A (1985). *Minitab Handbook*. PWS-KENT Publishing Company : Boston