INFERENSI PARAMETER MEAN POPULASI NORMAL DENGAN METODE BAYESIAN OBYEKTIF

Adi Setiawan

Program Studi Matematika Industri dan Statistika, Fakultas Sains dan Matematika
Universitas Kristen Satya Wacana Jl. Diponegoro 52-60 Salatiga 50711, Indonesia

adi_setia_03@yahoo.com

ABSTRAK

Dalam statistika, seringkali dilakukan anggapan bahwa sampel diambil dari populasi yang berdistribusi normal. Dalam makalah ini akan dijelaskan tentang bagaimana menggunakan metode Bayesian obyektif untuk mengestimasi mean populasi dalam kasus parameter variansi populasi diketahui atau tidak diketahui. Studi simulasi dilakukan untuk memperjelas penggunaan metode tersebut.

Keywords: prior, posterior, deskrepansi intrinsik, statistik intrinsik

PENDAHULUAN

Dalam statistika, seringkali dilakukan anggapan bahwa sampel diambil dari populasi yang berdistribusi normal. Pada makalah-makalah terdahulu telah dijelaskan bagaimana menggunakan metode Bayesian obyektif dalam melakukan estimasi titik, estimasi interval dan pengujian hipotesis (Setiawan, 2009a; Setiawan, 2009b; Setiawan, 2010b dan Setiawan, 2011). Dalam makalah ini akan dijelaskan tentang bagaimana menggunakan metode Bayesian obyektif untuk mengestimasi mean populasi dalam kasus parameter variansi populasi diketahui atau tidak diketahui. Studi simulasi dilakukan untuk memperjelas penggunaan metode tersebut.

DASAR TEORI

Estimasi Titik

Dalam pandangan Bayesian, hasil dari sembarang masalah inferensi yang dinyatakan dalam distribusi posterior merupakan gabungan dari informasi yang disediakan oleh data dan informasi prior relevan yang tersedia. Akan tetapi apabila tidak tersedia informasi prior, akan dipilih fungsi prior yang relatif *uninformative* artinya fungsi prior yang memberikan pengaruh minimum pada inferensi fungsi posterior. Secara lebih formal, misalkan bahwa mekanisme probabilitas yang membangkitkan data yang tersedia $\boldsymbol{x}$ dianggap sebagai $p(\boldsymbol{x}|\,\theta)$ untuk suatu $\theta \in \Theta$ dan kuantitas yang menjadi perhatian adalah fungsi yang bernilai real $\phi(\theta)$ dari θ. Tanpa menghilangkan keumuman, hal itu juga dapat dijelaskan berikut ini. Misalkan model probabilitas yang digunakan berbentuk $\{\, p(x\,|\,\theta,\lambda)\}$ dengan λ adalah parameter *nuisance* yang dipilih. Dalam hal ini diperlukan untuk mengidentifikasi fungsi prior bersama $\pi(\phi,\lambda)$ yang akan mempunyai pengaruh minimal pada distribusi posterior marginal dengan kuantitas yang menjadi perhatian ϕ yaitu

$$\pi(\phi\,|\,x) \propto \int_{\Lambda} p(x\,|\,\phi,\lambda)\,\pi(\phi,\lambda)\,d\lambda\,.$$

Reference prior digunakan sebagai prior yang dapat memberikan pengaruh minimal pada distribusi posterior. Dalam kasus dimensi satu, *reference prior* merupakan prior Jeffry. Dengan menggunakan prior ini maka penyelesaian masalah estimasi hanya tergantung pada model anggapan dan data pengamatan sehingga estimasi titik yang menggunakan metode ini dinamakan sebagai estimasi titik Bayesian obyektif (Bernardo dan Juarez, 2003).

Diskrepansi intrinsik (*intrínsic discrepancy*) $\delta(p_1, p_2)$ antara dua fungsi densitas $p_1(\boldsymbol{x})$ dengan $\boldsymbol{x} \in X_1$ dan $p_2(\boldsymbol{x})$ dengan $\boldsymbol{x} \in X_2$ didefinisikan sebagai

$$\delta(p_1, p_2) = \min\left\{K(p_2(x) | p_1(x)),\ K(p_1(x) | p_2(x))\right\}$$

dengan

$$K(p_1(x) | p_2(x)) = \int_X p_1(x) \log \frac{p_1(x)}{p_2(x)} dx \,.$$

Untuk dua keluarga fungsi densitas

$$M_1 = \left\{p_1(x|\phi), x \in \mathrm{X}_1(\phi), \phi \in \Phi\right\}$$

dan

$$M_2 = \left\{p_2(x|\psi), x \in \mathrm{X}_2(\psi), \psi \in \Psi\right\}$$

dapat didefinisikan diskrepansi intrinsik

$$\delta^*(M_1, M_2) = \inf_{\phi \in \Phi, \psi \in \Psi} \delta\big(p_1(x|\phi),\ p_2(x|\psi)\big).$$

sx

Fungsi kerugian (*loss function*) dalam kasus ini adalah diskrepansi intrinsik. Misalkan bahwa deskripsi yang sesuai dari tingkah laku probabilistik dari kuantitas random $\boldsymbol{x}$ diberikan oleh model

$$\{\, p(x | \theta, \lambda), x \in \mathrm{X}, \theta \in \Theta, \lambda \in \Lambda\,\}.$$

Diskrepansi intrinsik antara $p(x | \theta, \lambda)$ dan keluarga densitas

$$\{\, p(x | \theta_0, \lambda),\ \lambda \in \Lambda\,\}$$

adalah

$$\delta^*(\theta, \lambda; \theta_0) = \inf_{\lambda_0 \in \Lambda} \delta(\theta, \lambda; \theta_0, \lambda_0)$$

dengan

$$\delta(\theta, \lambda; \theta_0, \lambda_0) = \min\left\{K(\theta_0, \lambda_0 | \theta, \lambda)\ ,\ K(\theta, \lambda | \theta_0, \lambda_0)\right\}.$$

Misalkan $\{\, p(x | \theta, \lambda), x \in \mathrm{X}, \theta \in \Theta, \lambda \in \Lambda\,\}$ adalah model parametrik yang dapat digunakan untuk menggambarkan tingkah laku kuantitas random $\boldsymbol{x}$. Didefinisikan intrinsik statistik (*intrinsic statistic*) sebagai

$$d(\theta_0 | x) = E_{\pi_{\delta^*}}[\delta^* | x] = \int_\Lambda \int_\Theta \delta^*(\theta, \lambda; \theta_0)\, \pi_{\delta^*}(\theta, \lambda | x)\, d\theta\, d\lambda \qquad (1)$$

dengan $\pi_{\delta^*}(\theta, \lambda | x)$ adalah posterior referensi untuk parameter dari model $p(x | \theta, \lambda)$ bila $\delta^*(\theta, \lambda; \theta_0)$ adalah parameter yang menjadi perhatian.

Estimator intrinsik (*intrinsic estimator*) atau estimasi titik Bayesian obyektif didefinisikan sebagai yaitu parameter θ yang meminimalkan statistik intrinsik

$$\theta^* = \theta^*(x) = \arg\min_{\tilde{\theta} \in \Theta} d(\tilde{\theta} | x)\,.$$

Estimasi interval kredibel

Interval kredibel intrinsik 100q% (*q-credible region intrinsic*) adalah himpunan bagian $R^*_q = R^*_q(x, \Theta) \subseteq \Theta$ dari ruang parameter Θ sehingga memenuhi

(i) $\displaystyle\int_{R^*_q} \pi(\theta, \theta_0 | x)\, d\theta = q$

(ii) Untuk setiap $\theta_i \in R^*q$, $\theta_j \notin R^*q$ dan untuk setiap berlaku $d(\theta_i \mid x) \leq d(\theta_j \mid x)$.

dengan $d(\theta_i \mid x)$ adalah harapan fungsi kerugian *reference posterior* sebagai *proxy* untuk nilai dari parameter yang diberikan pada persamaan (1).

Terlihat bahwa pernyataan pada persamaan (1) mempunyai bentuk yang sulit sehingga perhitungannya tidaklah mudah namun dengan menggunakan integrasi numerik, hal itu dengan mudah dapat dilakukan.

Pengujian Hipotesis

Apabila diinginkan untuk melakukan pengujian hipotesis $H_0 \equiv \{ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \}$ maka statistik intrinsik pada persamaan (1) merupakan ukuran dari kekuatan bukti melawan penggunaan model M_0 dengan

$$M_0 = \{ p(x|\theta_0, \lambda), \lambda \in \Lambda \}.$$

Hal itu berarti H_0 akan ditolak jika dan hanya jika $d(\boldsymbol{\theta}_0 \mid \boldsymbol{x})$ untuk suatu batas d^* (Juarez, 2004). Bernardo dan Rueda (2002) mengusulkan untuk menggunakan aturan sebagai berikut : jika $d^* \approx 1$ maka tidak ada bukti untuk menolak $H0$, jika $d^* \approx 2{,}5$ maka terdapat bukti lemah (*mild*) untuk menolak dan jika $d^* > 5$ maka terdapat bukti kuat (*strong*) untuk menolak H_0.

Populasi Normal

Kasus 1

Misalkan dimiliki sampel x_1, x_2 , ..., x_n sampel dari distribusi normal N(μ, σ^2) dengan σ diketahui. Misalkan juga

$$\bar{x} = \frac{\sum_{j=1}^{n} x_j}{n}$$

adalah mean dari sampel. *Reference prior* untuk parameter yang menjadi perhatian yaitu μ adalah $\pi(\mu) = \sigma^{-1}$ sehingga *reference posterior* untuk parameter μ adalah

$$N(\mu \mid \bar{x}, \sigma/\sqrt{n}).$$

Dapat dibuktikan bahwa deskrepansi intrinsik antara distribusi normal N(μ_1, σ^2) dan distribusi normal N(μ_2, σ^2) adalah

$$\delta_x\{\mu_1, \mu_2 | \sigma^2\} = \frac{n}{2}\left[\frac{(\mu_1 - \mu_2)^2}{\sigma^2}\right] = \frac{1}{2}\left[\frac{\mu_1 - \mu_2}{\sigma/\sqrt{n}}\right]^2$$

sehingga diperoleh intrinsik statistik yaitu

$$d(\mu_0 \mid x_1, ..., x_n) = \int_{-\infty}^{\infty} \frac{1}{2}\left[\frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right]^2 N(\mu | \bar{x}, \sigma/\sqrt{n})\, d\mu. \qquad (2)$$

Hal itu berarti bahwa estimasi Bayesian obyektif untuk parameter μ adalah μ^* yang meminimalkan intrinsik statistik yaitu

$$\mu^* = \mu^*(x_1, ..., x_n) = \arg\min_{\tilde{\mu} \in \Omega} d(\tilde{\mu} | x_1, ..., x_n)$$

dengan $d(\tilde{\mu} | x_1, ..., x_n)$ merupakan persamaan (2). Estimasi interval kredibel untuk μ ditentukan sehingga syarat (i) dan (ii) di atas dipenuhi.

Kasus 2

Apabila dimiliki sampel x_1, x_2 , ..., x_n sampel dari distribusi normal N(μ, σ^2) dengan σ tidak diketahui. Misalkan

$$s^2 = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n}$$

adalah variansi dari sampel. *Reference prior* untuk parameter yang menjadi perhatian yaitu μ dengan σ sebagai parameter *nuisance* adalah

$$\pi(\mu, \sigma) = \pi(\mu)\, \pi(\sigma|\mu) = \sigma^{-1}$$

sehingga *reference posterior* yang bersesuaian adalah

$$\pi(\mu,\sigma | x_1, x_2, ..., x_n) = N(\mu \mid \bar{x}, \sigma/\sqrt{n})\, Gamma^{-1/2}[\,\sigma | (n-1)/2,\ ns^2/2\,]$$

$$\propto \sigma^{-(n+1)} \exp\left[-\frac{n}{\sigma^2}\left(s^2 + (\bar{x} - \mu)^2\right)\right].$$

Dapat dibuktikan bahwa deskrepansi intrinsik antara distribusi normal N(μ_1, σ^2) dan distribusi normal N(μ_2, σ^2) adalah

$$\delta_x\{(\mu_1, \sigma), (\mu_2, \sigma^2)\} = \frac{n}{2}\ln\left[1 + \frac{(\mu_1 - \mu_2)^2}{\sigma^2}\right]$$

sehingga diperoleh statistik intrinsik yaitu

$$d(\mu_0 \mid x_1, ..., x_n) = \int_{-\infty}^{\infty} \int_0^{\infty} \frac{n}{2} \ln\left[1 + \frac{(\mu - \mu_0)}{\sigma^2}\right] \pi(\mu, \sigma \mid x_1, ..., x_n)\, d\mu\, d\sigma\,. \qquad (3)$$

Fungsi mempunyai sumbu simetri $\mu_0 = \bar{x}$ sehingga estimasi titik dengan menggunakan metode Bayesian obyektif adalah $\hat{\mu} = \bar{x}$. Posterior referensi dari $\pi(\mu, \sigma | x_1, x_2, ..., x_n)$ adalah distribusi student *t* yaitu St(μ | $\mu_0 = \bar{x}$, s/$\sqrt{n-1}$, *n*-1). Hal itu berarti

$$\tau = \sqrt{n-1}(\mu - \bar{x})/s$$

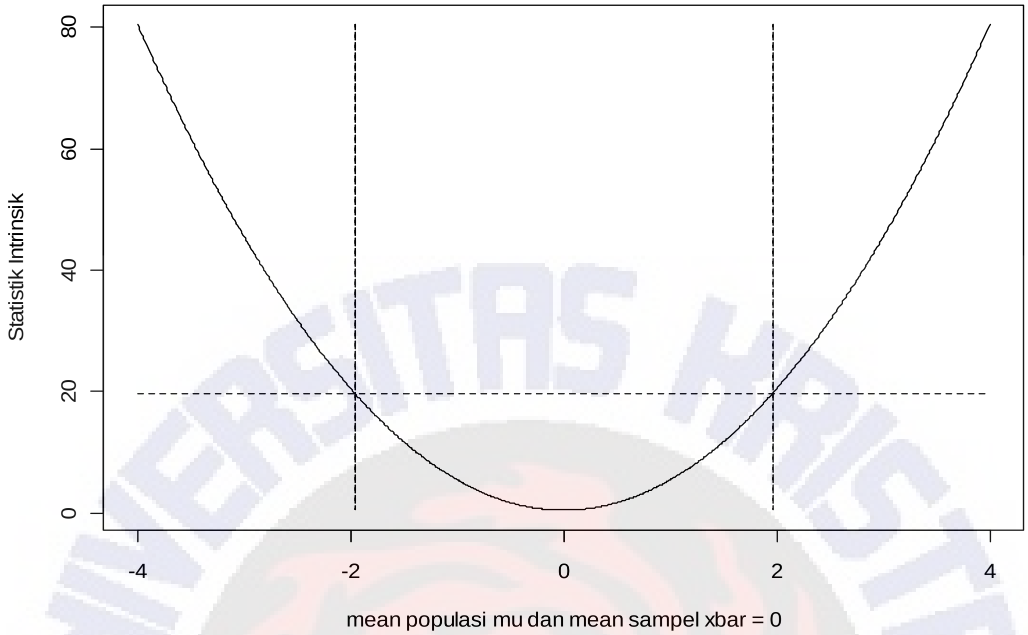
mempunyai distribusi *t* dengan derajat bebas *n*-1. Akibatnya interval kredibel (1-*q*)100% untuk mean μ adalah

$$\left(\bar{x} - t_{1-(q/2);n-1}\frac{s}{\sqrt{n-1}},\ \bar{x} + t_{1-(q/2);n-1}\frac{s}{\sqrt{n-1}}\right).$$

dengan $t_{1-(q/2);n-1}$ adalah kuantil ke-1-(*q*/2*)* dari distribusi *t* dengan derajat bebas *n*-1. Hipotesis H_0 : $\mu = \mu_0$ akan ditolak didasarkan pada statistik intrinsik pada persamaan (3) mempunyai nilai cukup besar. Berdasarkan saran Bernardo dan Rueda (2002) maka statistik intrinsik yang lebih besar dari 5 menunjukkan bahwa terdapat bukti yang kuat untuk menolak H_0.

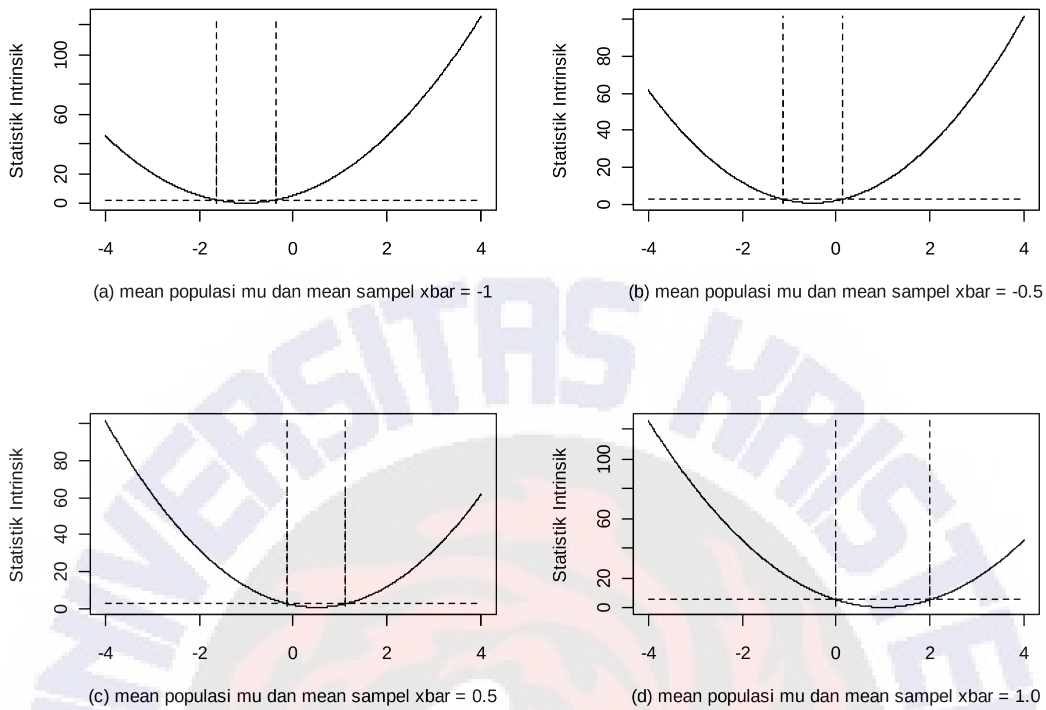
PERHITUNGAN STATISTIK INTRINSIK, STUDI SIMULASI DAN PEMBAHASAN

Estimasi titik untuk parameter mean populasi berdasarkan mean sampel ditentukan dengan cara memilih nilai μ yang meminimalkan nilai statistik intrinsik. Gambar 1 menunjukan nilai statistik intrinsik bila digunakan nilai μ antara -4 dan 4 dan *n* =10. Terlihat bahwa nilai statistik intrinsik akan mencapai minimum jika $\mu = \bar{x} = 0$ sehingga $\mu = \bar{x} = 0$ merupakan estimasi titik untuk mean populasi μ. Estimasi interval kredibel 95 % juga dapat ditentukan berdasarkan nilai statistik intrinsik. Batas bawah yaitu -1,96 dan batas atas 1,96 ditentukan sehingga nilai statistik intrinsik lebih kecil dari nilai statistik intrinsik 19.71.



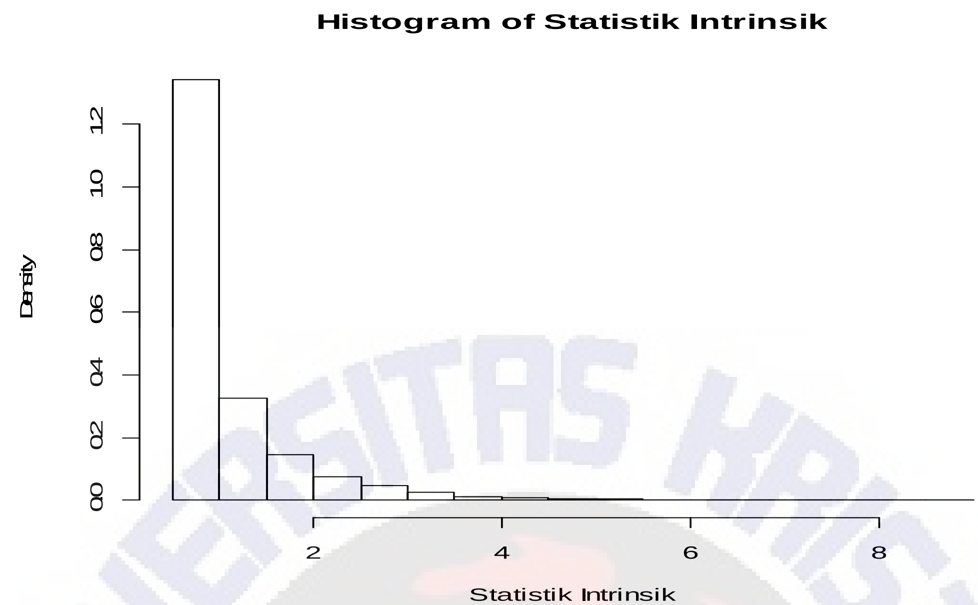
Gambar 1. Nilai statistik intrinsik untuk setiap nilai $\mu 0$ yang diberikan dengan menggunakan persamaan (1).

Gambar 2 menyatakan nilai-nilai statistik intrinsik jika dimiliki mean sampel berturut-turut yaitu (a) -1, (b) -0,5, (c) 0,5 dan (d) 1. Terlihat bahwa nilai statistik akan mencapai minimum pada mean sampelnya. Berdasarkan nilai statistik intrinsik ini juga dapat ditentukan interval kredibel untuk parameter μ. Interval kredibel 95 % untuk parameter μ bertutut-turut adalah (a) (-1,62 , -0,38) (b) (-1,12 , 0,12) (c) (-0,12 , 1,12) dan (d) (0, 2).

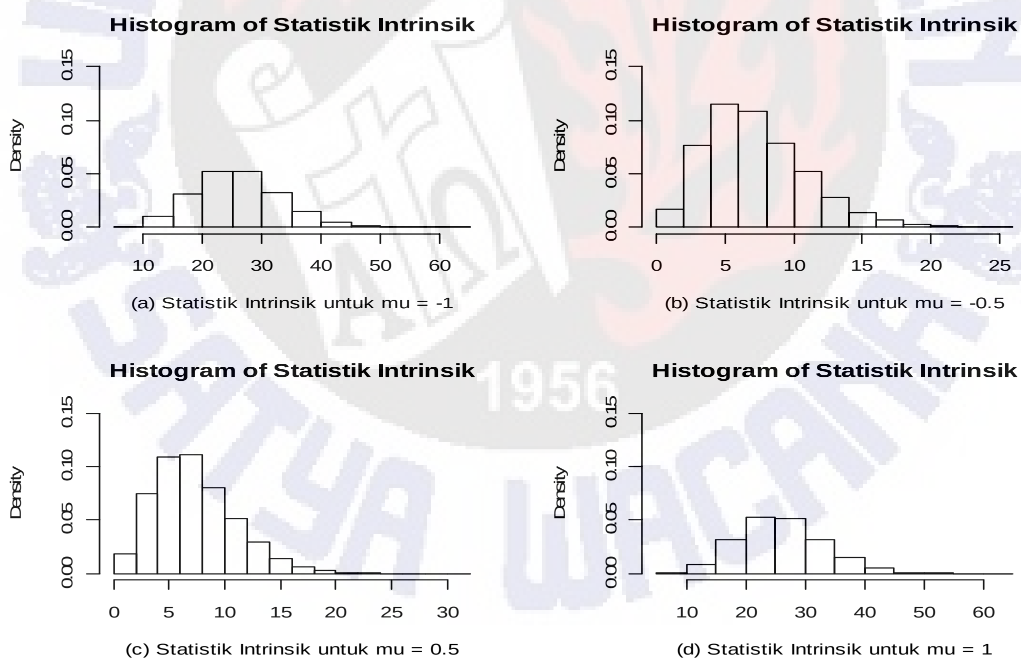


Gambar 2. Nilai-nilai statistik intrinsik jika dimiliki mean sampel berturut-turut yaitu (a) -1, (b) -0,5, (c) 0,5 dan (d) 1.

Misalkan dimiliki sampel $x_1, x_2, ..., x_n$ ukuran n = 50 dari populasi berdistribusi normal dengan mean μ tidak diketahui dan variansi populasi σ^2 diketahui yaitu dipilih $\sigma = 1$. Apabila diambil sampel dari distribusi normal dengan mean 0 maka nilai-nilai statistik intrinsik yang merupakan ukuran kekuatan untuk menolak hipotesis $H_0 : \mu = 0$ dan bila hal tersebut diulang sebanyak B = 10000 kali maka hasilnya dinyatakan pada Gambar 3. Terlihat bahwa nilai-nilai statistik intrinsik cenderung kecil dengan rata-rata 1,0287 dan hanya 0,46 % yang mempunyai nilai lebih dari 5. Apabila sampel diambil dari populasi yang mempunyai mean berturut-turut (a) -1 (b) -0,5 (c) 0,5 dan (d) 1 maka nilai-nilai statistik intrinsik dinyatakan pada Gambar 4. Terlihat bahwa nilai-nilai statistik intrinsik cenderung makin membesar jika mean populasi yang menjadi asal dari sampel jauh dari 0.



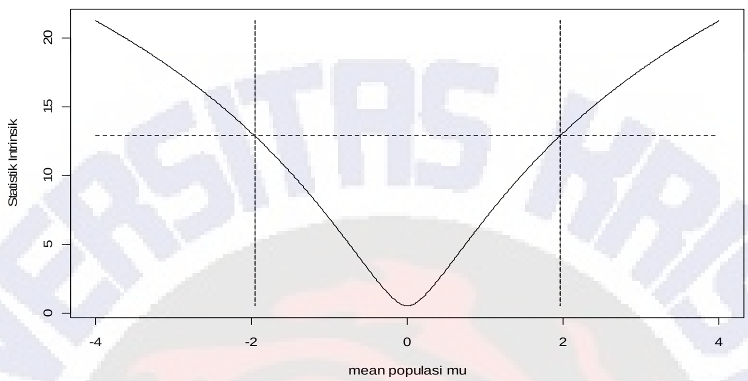
Gambar 3. Histogram dari nilai-nilai statistik intrinsik jika sampel diambil dari distribusi normal dengan mean 0 dan variansi 1 serta digunakan untuk pengujian hipotesis H0 : $\mu = 0$. Dalam hal ini digunakan pengulangan B = 10000.



Gambar 4. Histogram dari nilai-nilai statistik intrinsik jika sampel diambil dari populasi yang mempunyai mean berturut-turut (a) -1 (b) -0,5 (c) 0,5 dan (d) 1. Dalam hal ini digunakan pengulangan B = 10000.

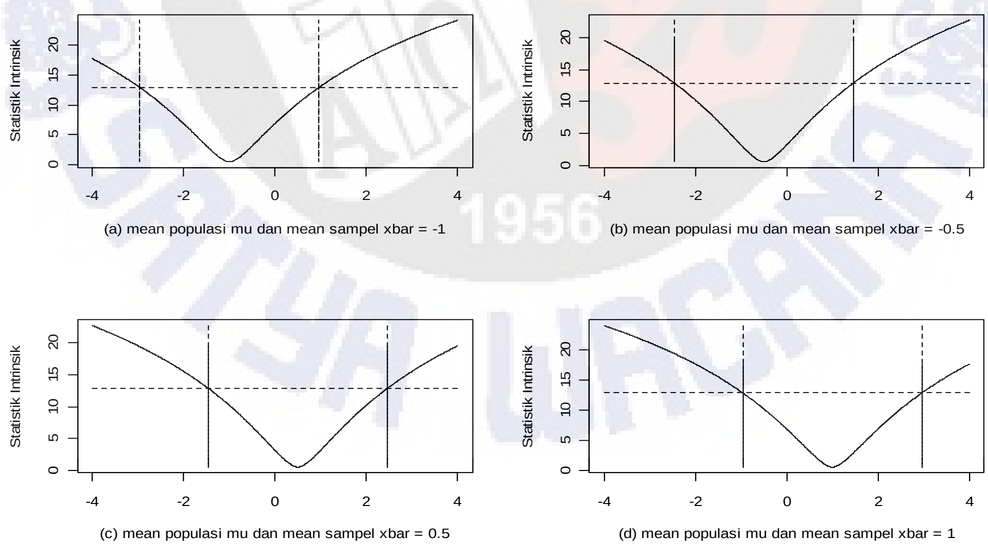
Kasus 2

Pada kasus σ tidak diketahui, estimasi titik untuk parameter mean populasi μ berdasarkan mean sampel ditentukan dengan cara memilih nilai μ yang meminimalkan nilai statistik intrinsik. Gambar 5 menunjukan nilai statistik intrinsik bila digunakan nilai μ antara -4 dan 4 jika diberikan mean sampel $\bar{x} = 0$ dan n=10. Terlihat bahwa nilai statistik intrinsik akan mencapai minimum jika $\mu = \bar{x} = 0$ sehingga $\mu = \bar{x} = 0$ merupakan estimasi titik untuk mean populasi μ.



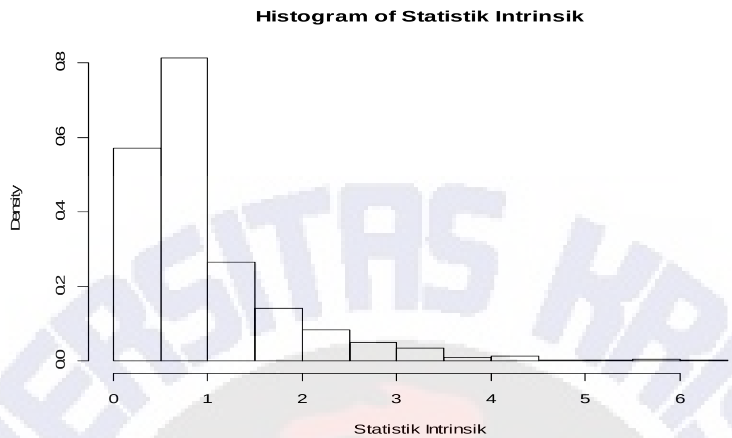
Gambar 5. Nilai statistik intrinsik jika diberikan mean μ dan mean sampel $\bar{x} = 0$.

Gambar 6 menyatakan nilai-nilai statistik intrinsik untuk μ jika diberikan mean sampel berturut-turut yaitu (a) -1, (b) -0,5, (c) 0,5 dan (d) 1. Dalam hal ini juga diberikan variansi sampel adalah 1. Terlihat bahwa nilai statistik akan mencapai minimum pada mean sampelnya. Berdasarkan nilai statistik intrinsik ini juga dapat ditentukan interval kredibel untuk parameter μ.

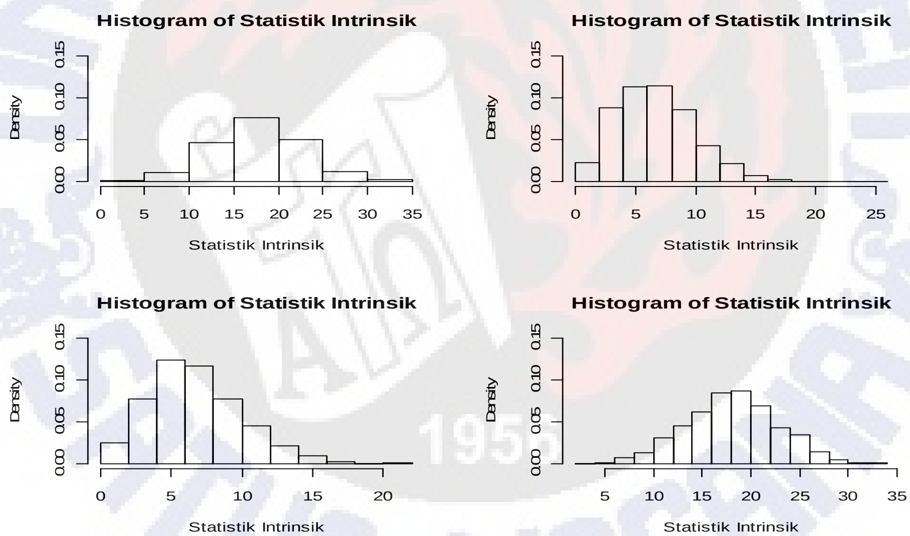


Gambar 6. Nilai-nilai statistik intrinsik untuk μ jika diberikan variansi sampel 1 dan mean sampel berturut-turut yaitu (a) -1, (b) -0,5, (c) 0,5 dan (d) 1.

Misalkan dimiliki sampel $x_1, x_2, ..., x_n$ berukuran n = 50 dari popolasi berdistribusi normal dengan mean μ dan variansi populasi σ^2 tidak diketahui. Apabila diambil sampel dari distribusi normal dengan mean 0 dan variansi 1 maka nilai-nilai statistik intrinsik yang merupakan ukuran kekuatan untuk menolak hipotesis nol H0 : $\mu = 0$ dinyatakan pada Gambar 7. Terlihat bahwa nilai-nilai statistik intrinsik cenderung kecil dengan rata-ratanya 0,99 dan hanya 0,6 % yang mempunyai nilai lebih dari 5. Apabila sampel diambil dari populasi yang mempunyai mean berturut-turut (a) -1 (b) -0,5 (c) 0,5 dan (d) 1 maka nilai-nilai statistik intrinsik dinyatakan pada Gambar 8. Terlihat bahwa nilai-nilai statistik intrinsik cenderung makin membesar jika mean populasi yang menjadi asal dari sampel jauh dari 0.



Gambar 7. Histogram dari B = 1000 nilai-nilai statistik intrinsik yang merupakan ukuran kekuatan untuk menolak H0 : $\mu = 0$ jika diberikan sampel yang diambil dari populasi normal dengan mean 0 dan variansi 1.



Gambar 8. Histogram dari B = 1000 nilai-nilai statistik intrinsik yang merupakan ukuran kekuatan untuk menolak H0 : $\mu = 0$ jika diberikan sampel yang diambil dari populasi normal dengan variansi 1 dan mean populasi berturut-turut (a) – 1 (b) -0,5 (c) 0,5 dan (d) 1.

KESIMPULAN

Dalam makalah di atas telah dijelaskan bagaimana parameter mean populasi diestimasi dan dilakukan uji hipotesis dengan menggunakan metode Bayesian obyektif jika dianggap sampel diambil dari distribusi normal. Metode tersebut dapat juga diperluas penggunaannya untuk parameter variansi populasi.

DAFTAR PUSTAKA

[1] Bernardo, J. dan R. Rueda (2002) Bayesian Hypotesis Testing : A Reference Approach, *International Statistical Review* 70, 351-372.

[2] Juarez, M. A. (2004) *Objective Bayesian Methods for Estimation and Hypothesis Testing*, Valencia : University of Valencia.

[3] Setiawan, A. (2009a) Estimasi Titik Bayesian Obyektif, *Prosiding Seminar Sains dan Pendidikan Sains* IV FSM UKSW, Salatiga.

[4] Setiawan, A. (2009b) *Credible Interval* Bayesian Obyektif, *Prosiding Seminar Nasional Matematika*, Universitas Katolik Parahyangan, Bandung.

[5] Setiawan, A. (2010b) Pengujian Hipotesis dengan Metode Bayesian Obyektif, *disampaikan dalam Konferensi Nasional Matematika XV 30 Juni – 3 Juli 2010*, UNIMA, Tondano.

[6] Setiawan, A. (2011) Pengujian Hipotesis tentang Parameter Populasi Berdistribusi Eksponensial dengan Metode Bayesian Obyektif, *Prosiding Seminar Nasional Statistika Universitas Diponegoro 2011*, ISBN : 987-979-097-142-4.