BAB II

ESTIMASI PARAMETER

2.1 Estimator Konsisten dan Tak Bias

Inferensi statistik mempelajari cara pengambilan kesimpulan (generalisasi) tentang parameter - parameter populasinya berdasarkan analisis data sampel . Sebelum data diambil dari populasi hal-hal yang perlu diperhatikan adalah : ukuran sampel meliputi cara pengambilan sampel , jenis inferensi yang diinginkan , kekuatan dan kecermatan penarikan kesimpulan. Dalam inferensi biasanya parameter populasi dipandang suatu konstan yang tidak diketahui harga sebenarnya. Jika berdasarkan data sampel akan diduga atau diperkirakan harga parameter itu dengan kecermatan tertentu, maka inferensinya disebut *Estimasi Parameter*.

Suatu keluarga fungsi probabilitas dikatakan parametrik jika ada sejumlah terhingga harga-harga parameter yang menentukan dengan tunggal suatu anggota keluarga ini. Dengan perkataan lain, bentuk distribusinya diketahui, tetapi harga-harga parameternya tidak diketahui. Suatu keluarga fungsi probabilitas dikatakan non parametrik apabila beberapa harga-harga parameternya ditentukan, tidak menentukan dengan tunggal suatu anggota keluarga itu.

<u>Definisi 2.1</u>

Statistik adalah fungsi g(x) dari data observasi $X' = (x_1, \ldots, x_n)$. Berarti harga g(x) dapat dihitung langsung dari harga $X' = (x_1, \ldots, x_n)$ karena fungsi ini tergantung harga parameter yang tidak diketahui.

Suatu statistik yang digunakan sebagai estimasi suatu parameter dinamakan estimator titik dari parameter itu. Estimator titik θ adalah suatu variabel random yang mempunyai fungsi probabilitas tertentu, yang dinamakan distribusi sampling θ. Fungsi probabilitas estimator θ seringkali dapat diturunkan apabila fungsi probabilitas populasinya diketahui.

<u>*Contoh 2.1*</u>:

Misalkan x_1 , ... , x_n adalah sampel random dari populasi berdistribusi normal dengan mean μ dan variansi σ^2. Maka estimator titik untuk μ adalah :

$$E(x) = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

<u>Definisi 2.2</u>

Fungsi f(x) adalah suatu fungsi probabilitas atau distribusi probabilitas suatu variabel random diskrit X jika dan hanya jika untuk setiap hasil x yang mungkin ,

1. $f(x) \geq 0$
2. $\sum f(x) = 1$
3. $P(X=x) = f(x)$

<u>Definisi 2.3</u>

Fungsi f(x) adalah fungsi kepadatan probabilitas variabel random kontinu X yang didefinisikan diatas himpunan semua bilangan riil $\Re$, bila memenuhi :

1. $f(x) \geq 0$ untuk semua $x \in \Re$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$

3. $P(a<x<b) = \int_a^b f(x)\, dx$

Dalam mengestimasi parameter digunakan metode titik estimasi untuk mendapatkan estimator yang terbaik. Sifat-sifat dari estimator terbaik adalah : konsisten dan tak bias.

Definisi 2.4

Misalkan $x_1 , \ldots , x_n$ variabel random independen berdistribusi identik. Baris estimator titik $g_1(x_1)$, $g_2(x_2)$, . . . dikatakan konsisten jika untuk setiap $\varepsilon>0$ berlaku,

$$\lim_{n\to\infty} P\{|g_n(x_1,\ldots,x_n) - \theta| \le \varepsilon\} = 1$$

Jadi konsistensi $g_1(x_1)$, $g_2(x_2)$, ... untuk θ adalah konvergen dalam probabilitas baris estimator itu ke variabel random konstan θ.

Contoh 2.2

Misalkan dilakukan ekperimen berulang-ulang dengan probabilitas sukses P. x_i menunjukkan banyaknya sukses dalam trial ke-i . Maka $\frac{1}{n}\sum_{i=1}^{n} x_i$ menunjukkan proporsi sukses dalam n trial yang pertama. Akan ditunjukkan bahwa $\frac{1}{n}\sum_{i=1}^{n} x_i$ n=1,2, ... adalah konsisten terhadap P.

X_i adalah independen dengan $E(x_i) = p$ dan $Var(x_i) = p(1-p)$, maka

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} x_i - p\right| \le \varepsilon\right\} = 1$$

jadi $\frac{1}{n}\sum_{i=1}^{n} x_i$ konsisten untuk p

Definisi 2.5

Misalkan x_1 , ... ,x_n adalah variabel random dengan distribusi f(x). θ adalah parameter yang ditentukan oleh distribusi f(x). Statistik g(x_1 , ... , x_n) adalah estimator tak bias untuk θ, jika dan hanya jika

$$E(g(x_1, \ldots, x_n)) = \theta$$

Contoh 2.3

Misalkan x_1, ... ,x_n sampel random dari suatu populasi dengan mean E(x) = μ dan variansi σ^2 , maka mean sampel X adalah estimator tak bias untuk μ yakni ,

$$E(\overline{X}) = E(\frac{1}{n}(X_1 + \ldots + X_n))$$

$$= \frac{1}{n}(E(x_1) + \ldots + E(x_n))$$

$$= \frac{1}{n}(\mu + \ldots + \mu) = \frac{1}{n} . n\mu = \mu$$

demikian juga variansi sampel

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

merupakan estimator tak bias untuk σ^2 , yakni

$$E(s^2) = E\left(\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i^2) - \frac{n}{n-1}E(\overline{X}^2)$$

sehingga $E(s^2) = \frac{1}{n-1}(n\sigma^2 + n\mu^2) - \frac{n}{n-1}(\frac{\sigma^2}{n} + \mu^2)$

$$= \frac{n}{n-1}\sigma^2 - \frac{1}{n-1}\sigma^2 = \frac{n-1}{n-1}\sigma^2 = \sigma^2$$

2.2 ESTIMATOR BAYES

Metode bayes merupakan metode estimasi secara parametrik untuk mengestimasi θ. Prinsip metode bayes adalah menggabungkan antara informasi sampel dengan informasi sebelumnya atau prior. Informasi prior adalah asumsi dari peneliti / pengamat mengenai bentuk distribusi yang diambil. Sehingga didefinisikan

Definisi 2.6

Diberikan $x_1, \dots, x_n$ adalah variabel random independen yang berdistribusi identik dengan distribusi sampel $f(x|\theta)$ untuk $\theta \in \Omega$. Estimator bayes $\delta(x_1, \dots, x_n)$ untuk θ yang mempunyai distribusi prior $\lambda(\theta)$ dirumuskan dengan :

$$\delta(x_1, \dots, x_n) = \frac{\int \theta f(x_1|\theta)\dots f(x_n|\theta)\lambda(\theta)d\theta}{\int f(x_1|\theta)\dots f(x_n|\theta)\lambda(\theta)d\theta}$$

estimator $\delta(x_1, \dots, x_n)$ adalah harga harapan θ yang berkaitan dengan fungsi distribusi prior $\lambda(\theta)$ pada ruang parameter Ω

Contoh 2.4 :

Misalkan $x_1, \dots, x_n$ variabel random independen berdistribusi identik dengan fungsi distribusinya :

$f(x|\theta) = \theta\, e^{-\theta x}$ dimana $\theta > 0$ serta $x>0$

diambil distribusi prior $\lambda(\theta) = e^{-\theta}$ untuk θ . Estimator bayes untuk θ adalah :

$$\delta(x_1, \ldots, x_n) = \frac{I_1}{I_2}$$

$$I_1 = \int \theta\, f(x_1|\theta) \ldots f(x_n|\theta)\, \lambda(\theta)\, d\theta$$

$$= \int_0^{\infty} \theta\theta^n e^{-\theta\sum x} e^{-\theta} d\theta$$

$$= \int_0^{\infty} \theta^{n+1} e^{-\theta(\sum x+1)} d\theta$$

$$= \frac{\Gamma(n+2)}{(\sum x+1)^{n+2}}$$

$$I_2 = \int f(x_1|\theta) \ldots f(x_n|\theta)\, \lambda(\theta)\, d\theta$$

$$= \int_0^{\infty} \theta^n e^{-\theta\sum x} e^{-\theta} d\theta$$

$$= \int_0^{\infty} \theta^n e^{-\theta(\sum x+1)} d\theta$$

$$= \frac{\Gamma(n+1)}{(\sum x+1)^{n+1}}$$

Sehingga diperoleh estimator bayes sebagai berikut :

$$\delta(x_1, \ldots, x_n) = \frac{I_1}{I_2} = \frac{\Gamma(n+2)}{(\sum x+1)^{n+2}} \cdot \frac{(\sum x+1)^{n+1}}{\Gamma(n+1)}$$

$$= \frac{n+1}{\sum x+1}$$

2.3 METODE BOOTSTRAP

Metode Bootstrap adalah metode yang dikembangkan pertama kali oleh Efron,B pada tahun 1979 yang menitik beratkan analogi antara sampel dan populasi dari mana sampel tersebut diambil untuk mengestimasi sebaran sampling dari θ yang meliputi mean dan variansi. Untuk memperoleh suatu dugaan empirik dari sebaran statistik sampel atau sampel tersebut mengikuti suatu distribusi apa dilakukan dengan cara pembootstrapan yang meliputi proses resampling dari data dengan pengembalian beberapa kali. Jika diketahui dari suatu observasi X_1 , X_2 , ... ,Xn maka sampel buatan yang diambil dengan pengembalian dari X_1 , X_2 , ... ,Xn, akan mempunyai titik sampel dengan peluang yang sama untuk terpilih yaitu $\frac{1}{n}$ dari setiap X_i . Dari sini tehnik pembootstrapannya meliputi pendugaan secara empirik semua sebaran sampling dari $\hat{\theta}$. Sehingga kadang-kadang pembootstrapan untuk penarikan kesimpulan akan memberikan hasil yang lebih baik apabila asumsi yang ada tidak jelas atau mungkin kurang realistik untuk diterapkan.

Suatu sampel berukuran n diambil berulang-ulang secara acak dengan pengembalian , dari sampel asli (seperti pada Diagram 2.1 struktur komponen ragam sampling dan resampling) dapat menghasilkan suatu resampel yang dinotasikan dengan tanda asterik (*). Penekanan dari hasil pembootstrapan ini adalah bahwa sebaran frekuensi relatif dari $\hat{\theta}^*$ yang dihitung dari resampel adalah merupakan dugaan dari sebaran sampling $\hat{\theta}$.

Dengan langkah-langkah dasar pembootstrapannya sebagai berikut :

1. Diberikan suatu sebaran peluang empirik $\hat{F}(x)$ bagi sampel dengan peluang $\frac{1}{n}$ untuk masing-masing titik $x_1, x_2, \ldots, x_n$. $\hat{F}(x)$ adalah fungsi sebaran empirik dari x yang juga merupakan dugaan maksimum likelihood non parametrik bagi sebaran populasi F(x) .

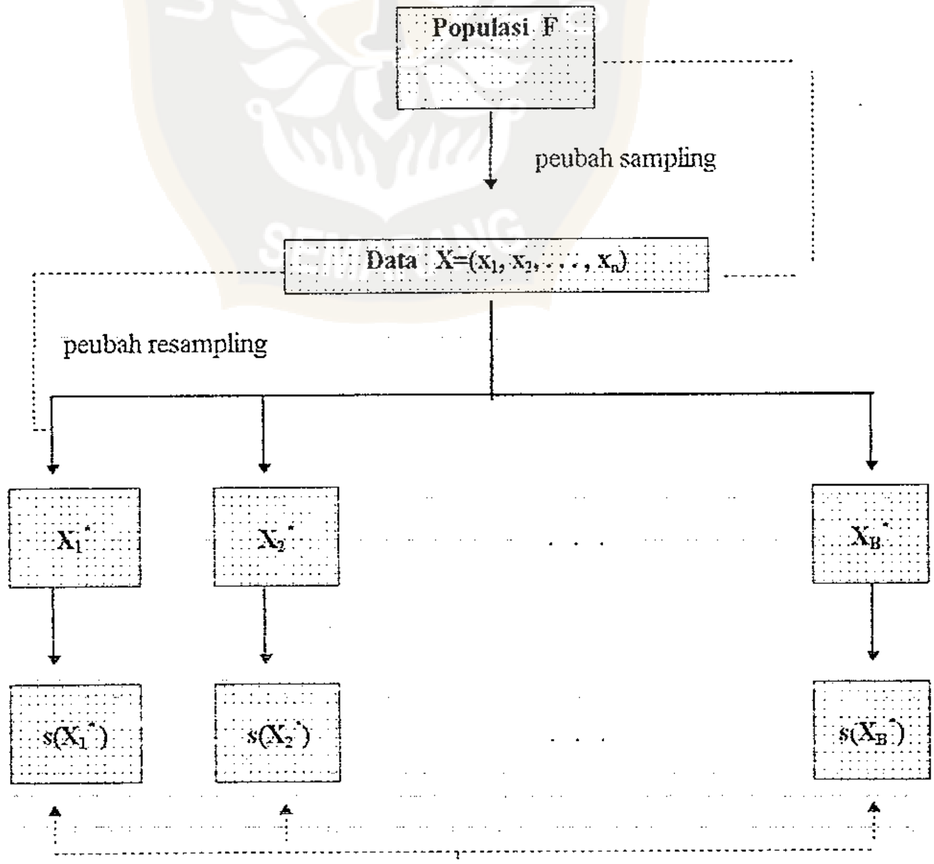
2. Dari $\hat{F}(x)$ ditarik sampel acak sederhana berukuran n dengan pengembalian maka dapat menghasilkan sampel sebanyak n^n buah sampel yang berbeda. Perolehan sampel acak ini disebut sebagai *resampel* dari x_b^*

3. Setiap resampel yang dihasilkan dihitung rata-ratanya yang dinotasikan $\hat{\theta}_b^*$ dan ini merupakan sampel baru yang ke-b. Langkah ke-2 dan ke-3 ini diulang sebesar B kali sampai mencapai B yang cukup besar, karena untuk B yang cukup besar juga akan menghasilkan sampel baru yang cukup besar pula.

4. Setelah proses resampel dirasa cukup besar maka dapat diberikan sebaran peluang dari B yaitu $\hat{\theta}_b^*$ dengan menempatkan peluang $\frac{1}{B}$ bagi masing-masing $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ yang mana sebaran peluang ini adalah dugaan bootstrap sebaran sampling $\hat{\theta}$ dari $\hat{F}^*(\hat{\theta}^*)$.

5. Demikian juga untuk dugaan bootstrap galat baku statistik dari $\sigma_F(\hat{\theta})$ adalah dugaan "*plug-in*" yang dipakai untuk menggantikan sebaran F yang tidak diketahui dengan suatu fungsi sebaran empirik $\hat{F}$. Sehingga dugaan bootstrap dari $\sigma_F(\hat{\theta})$ adalah $\sigma_{\hat{F}}(\hat{\theta}^*)$.

Maka untuk nilai B yang cukup besar ($B \rightarrow \sim$) maka σ_B adalah mendekati $\sigma_{\hat{F}}(\hat{\theta}^*)$.



Sampel Bootstrap
(Diagram 2.1)
Struktur komponen ragam sampling dan resampling

Contoh 2.4

Diambil suatu data dari pengamatan n=3 yakni $(X_1,X_2,X_3) = (x_1,x_2,x_3) = (4,1,10)$. Rata - rata sampel $\overline{X} = \frac{(4+1+10)}{3} = 5$. Akan diestimasi parameter μ dan σ^2 secara non parametrik dengan metode bootstrap.

Resampling yang dilakukan dengan pengembalian untuk semua kemungkinan sampel akan diperoleh sebanyak 3^3 sampel yang berbeda , apabila dikelompokkan berdasarkan besarnya rata - rata dari setiap sampel yang terambil , didapatkan sebanyak 9 kelompok sampel seperti yang ditunjukkan pada tabel 2.1

Estimasi parameter yang diperoleh adalah :

$$\mu = E(x_i^*)$$

$$= \tfrac{1}{9}\sum_{i=1}^{9} x_i^*$$

$$= 5{,}11$$

$$\sigma^2 = E(s_i^*) = \tfrac{1}{9-1}\sum_{i=1}^{9} s_i^*$$

$$= \frac{68{,}8889}{8}$$

$$= 8{,}61$$

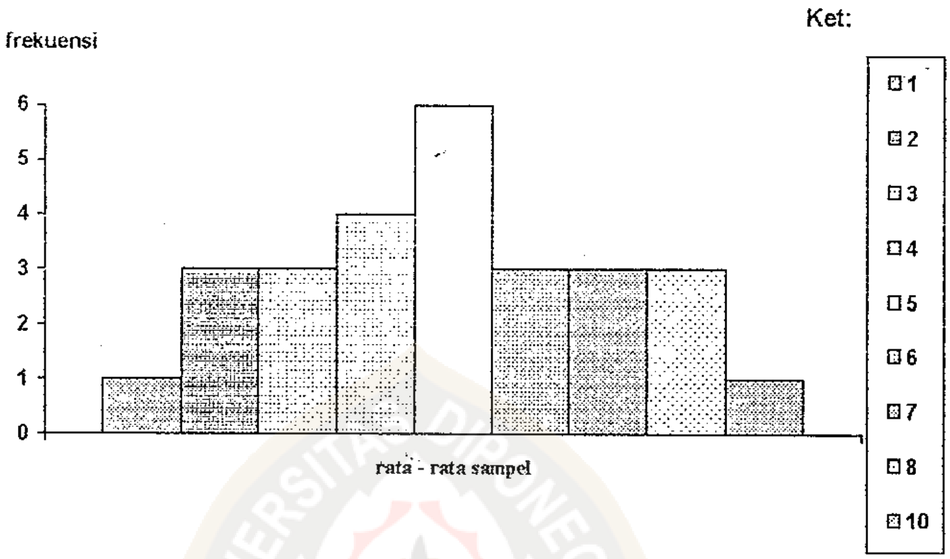
Tabel 2.1

Sebaran frekuensi rata-rata resampling

	X_i^*	$\sqrt{3}(X_i^* - \overline{X})$ (w_i)	frekuensi (f_i)	frekuensi relatif (p_i)	frekuensi relatif kumulatif
1	1	-6,9282	1	1/27	1/27
2	2	-5,1962	3	3/27	4/27
3	3	-3,4641	3	3/27	7/27
4	4	-1,7320	4	4/27	11/27
5	5	0	6	6/27	17/27
6	6	1,7320	3	3/27	20/27
7	7	3,4641	3	3/27	23/27
8	8	5,1962	3	3/27	26/27
9	10	8,6603	1	1/27	1
		Jumlah	$27 = 3^3$	1	

(sumber : Helmer,R. ;(1995))

Gambar 2.1 memperlihatkan bahwa sebaran dari resampling bootstrap yang dihasilkan terlihat tidak normal. Untuk menghasilkan penggambaran yang normal dipengaruhi oleh semakin besarnya sampel yang akan diteliti dengan semakin besarnya sampel maka akan semakin besar juga sampel buatan dari resampling. Jika penggambaran histogram terlihat normal, maka estimasi parameternya merupakan estimator terbaik dalam observasi tersebut.



Gambar 2.1

Bentuk sebaran dari sampel bootstrap