BAB II

MATERI PENUNJANG

Dalam bab ini akan disajikan konsep mengenai estimasi parameter dan bias suatu estimator serta metode bootstrap yang akan mendukung untuk pembahasan selanjutnya mengenai estimasi bias dengan metode tersebut.

2.1. Estimasi Parameter dan Bias Estimator.

Inferensi statistik mempelajari cara pengambilan kesimpulan tentang parameter-parameter populasinya berdasarkan analisis data sampel. Dalam inferensi biasanya parameter populasi dipandang suatu konstanta yang tidak diketahui harga sebenarnya. Jika berdasarkan data sampel akan diduga atau diperkirakan harga parameter itu dengan keakuratan tertentu, maka estimasinya disebut estimasi parameter.

Definisi 2.1.1 Definisi Ekspektasi.

Suatu variabel random diskrit X dengan fungsi probabilitas f(x) didefinisikan harapannya atau ekspektasinya $[E(X)]$ sebagai $E(X)=\sum_i x_i f(x_i)$, asal saja jumlahannya ada dan terhingga.

Definisi 2.1.2 Definisi Statistik Fungsi.

Statistik adalah fungsi g(x) dari data observasi x = $(x_1,x_2,\ldots,x_n)$. Suatu statistik yang digunakan sebagai estimasi suatu parameter dinamakan estimator titik dari parameter itu. Estimator titik $\hat{\theta}$ adalah suatu variabel random yang mempunyai fungsi probabilitas tertentu yang dinamakan distribusi sampling $\hat{\theta}$. Metode yang digunakan untuk menentukan estimator umumnya adalah metode momen dan metode maksimum likelihood.

Definisi 2.1.3 Definisi Fungsi Probabilitas.

Fungsi f(x) adalah suatu fungsi probabilitas suatu variabel random diskrit X jika dan hanya jika memenuhi sifat-sifat berikut:

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

Definisi 2.1.4 Definisi Bias Estimator.

Misalkan X adalah variabel random independen dengan fungsi probabilitas f(x) dan θ merupakan suatu parameter. Bias dari $\hat{\theta}$ sebagai estimator untuk parameter θ didefinisikan sebagai selisih diantara ekspektasi $\hat{\theta}$ dengan nilai parameter θ,

$$\text{Bias} = \text{E}[\hat{\theta}] - \theta$$

Definisi 2.1.5 Definisi Mean dan Varian Sampel

Misalkan $X_1, X_2, \ldots, X_n$ merupakan sampel random berukuran n dari sebuah distribusi yang diberikan. Suatu statistik

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \text{ dan } S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n} = \sum_{i=1}^{n} \frac{X_i}{n} - \overline{X}^2$$

masing-masing disebut mean dan varian dari sampel random tersebut.

Pendefinisian persamaan untuk varian sampel di atas menunjukan bahwa sampel random diperoleh melalui resampling dengan pengembalian, yang mana peluang untuk setiap titik data terambil sebagai sampel random tersebut adalah 1/n.

Distribusi probabilitas yang menghasilkan sampel random tersebut dikenal sebagai distribusi probabilitas empiris. Jadi ada alasan kuat untuk menggunakan persamaan varian sampel seperti tersebut di atas dari pada rumus varian sampel $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ (Robert V. Hogg and AllanT. Craig, 1978).

Definisi 2.1.6 Definisi Vektor.

Suatu notasi $a = (a_1, a_2, ..., a_n)$ disebut vektor a pada ruang R^n dengan $a_1, a_2, ..., a_n$ merupakan komponen-komponen dari vektor tersebut.

2.2. Simulasi Monte Carlo dan Pembangkit Bilangan Acak.

Kemajuan komputasi telah menempatkan statistik dalam perspektif baru. Saat ini banyak masalah-masalah statistik dan analisa data yang dikerjakan atau diselesaikan dengan bantuan komputer, dimana masalah tersebut terlalu rumit untuk dikerjakan atau diselesaikan secara manual. Metode bootstrap ideal memerlukan penghitungan terhadap n^n kemungkinan sampel, sehingga jika menggunakan penghitungan manual maka akan menjadi sangat tidak efisien. Metode bootstrap yang merupakan suatu prosedur resampling akan lebih efisien jika menggunakan komputasi komputer, yaitu dengan mensimulasikannya. Simulasi yang digunakan dalam metode bootstrap ini adalah simulasi monte carlo, sebab dengan simulasi ini dapat dilakukan pendekatan tanpa harus melakukan penghitungan terhadap n^n buah sampel, tetapi cukup dengan mengambil sejumlah B sampel yang cukup besar tetapi akan tampak kecil jika dibandingkan dengan jumlah sampel bootstrap ideal sebenarnya.

Nama Monte Carlo pertama kali digunakan oleh Jon Von Neumann dan Ulam yang bekerja pada pengembangan bom atom selama perang dunia kedua. Penggunaan nama Monte Carlo yang merupakan tempat judi di kota kerajaan Monaco dianggap tepat karena sifat acak dari permainan *russian roullete* serupa dengan metode pengambilan sampel dari suatu distribusi probabilitas. Metode monte carlo mulai dikenal pada tahun 1994, tetapi pengembangan sistematisnya merupakan hasil kerja Harris dan Herman Kahn pada tahun 1948. Pada umumnya metode ini digunakan pada proses simulasi, karena hampir semua proses simulasi menggunakan bilangan random dengan suatu distribusi tertentu.

Simulasi monte carlo merupakan prosedur komputasi numerik yang melibatkan pengambilan sampel eksperimen dengan bilangan random (acak), yaitu peubah acak U(0,1). Simulasi monte carlo didasarkan pada penggunaan sampling untuk memperkirakan hasil-hasil (output) dari suatu sistem yang dipelajari. Dari sampling data tersebut dapat disusun sebaran frekuensi dari parameter sistem yang dipelajari. Selanjutnya dari data frekuensi sebaran tersebut dapat disusun grafik sebaran kumulatifnya berdasarkan pada penggunaan bilangan-bilangan acak yaitu bilangan-bilangan yang tidak ada ketergantungan satu sama lain dan tidak mempunyai korelasi sama sekali untuk menyusun karakteristik dari parameter sistem yang dipelajari. Semakin banyak bilangan acak yang dipakai hasilnya akan semakin tepat. Perkembangan komputer digital telah membawa peningkatan untuk penggunaan simulasi, sebab dengan komputer digital memungkinkan pembangkitan bilangan acak dalam jumlah besar, sehingga ketepatan hasilnya pun semakin baik. Tahapan yang terpenting dalam simulasi monte carlo adalah dalam proses pembangkitan bilangan acak U(0,1).

Sebuah simulasi dari suatu sistem atau proses dimana terdapat komponen yang bersifat acak membutuhkan suatu metode pembangkitan atau perolehan bilangan acak. Selanjutnya akan dibahas pembangkitan bilangan acak dari sebaran uniform pada interval [0,1]. Sebaran ini dinyatakan dengan U(0,1). Distribusi uniform mempunyai daerah atau range dari suatu variabel kontinu X yang terdiri dari sebuah interval. Fungsi densitas uniform didefinisikan sebagai berikut :

$$f(x) = \frac{1}{b-a} \qquad a \le x \le b$$

$$= 0 \qquad \text{lainnya}$$

Prosedur pembangkitan bilangan acak yang dipakai pada umumnya menggunakan nilai awal sebagai bilangan awal untuk mencari bilangan kedua. Kemudian bilangan kedua ini dipakai untuk mencari bilangan ketiga, demikian seterusnya. Pembangkit bilangan acak yang banyak dipakai saat ini adalah Linear Congruential Generators (LCGs) yang diperkenalkan oleh Lehmer (1951). Suatu barisan intejer $z_1, z_2, \ldots$ didefinisikan dengan formula rekursif sebagai berikut :

$$Z_i = (aZ_{i-1} + c)(\text{mod } m), \quad i = 1,2,\ldots,n$$

dimana m = modulus a = faktor pengali

c = faktor penambah Z_0 = nilai awal

merupakan integer non-negatip.

Persamaan di atas menyatakan bahwa untuk memperoleh Z_i , bagilah $aZ_{i-1} + c$ dengan m dan Z_i merupakan sisa dari pembagian ini. Oleh karena itu $0 \le z_i \le m-1$ dan untuk memperoleh bilangan acak yang diinginkan, yaitu U_i

pada [0,1], diambil $U_i = Z_i/m$. Selain itu, pemilihan konstanta intejer m, a, c, dan Z_0 tidak boleh dilakukan secara sembarangan. Menurut Fuller (1998) ada beberapa syarat yang harus dipenuhi yaitu: modulus m > 0, faktor pengali $0 \le a \le m$, faktor penambah $0 \le c \le m$, dan nilai awal $Z_0 \ge 0$.

Pandang LCG yang didefinisikan dengan m = 17, a = 6, c = 1, Z_0 = 8. Tabel 2.1 memberikan Z_i dan U_i (tiga desimal dibelakang koma) untuk $i = 1,2,3,\ldots,40$. Dari tabel tersebut terlihat bahwa $Z_{17} = Z_1 = 15$, $Z_{18} = Z_2 = 6$ dan seterusnya.

Tabel 2.1. LCG di mana $Z_i = (6Z_{i-1} + 1)(\bmod 17)$ dengan $Z_0 = 8$.

i	Z_i	U_i	i	Z_i	U_i	i	Z_i	U_i	i	Z_i	U_i
1	15	0,882	11	0	0,000	21	13	0,765	31	4	0,235
2	6	0,353	12	1	0,059	22	11	0,647	32	8	0,471
3	3	0,176	13	7	0,412	23	16	0,941	33	15	0,882
4	2	0,118	14	9	0,529	24	12	0,706	34	6	0,353
5	13	0,765	15	4	0,235	25	5	0,294	35	3	0,176
6	11	0,647	16	8	0,471	26	14	0,824	36	2	0,118
7	16	0,941	17	15	0,882	27	0	0,000	37	13	0,765
8	12	0,706	18	6	0,353	28	1	0,059	38	11	0,647
9	5	0,294	19	3	0,176	29	7	0,412	39	16	0,941
10	14	0,824	20	2	0,118	30	9	0,529	40	12	0,706

Dari tabel di atas tampak adanya looping, sebab nilai Z_i untuk $i = 17,18,\ldots,32$ akan sama dengan nilai Z_i untuk $i = 1,2,3,\ldots,16$, demikian seterusnya jika perhitungan dilanjutkan maka akan diperoleh nilai yang disebut periode suatu pembangkit. Untuk LCGs, Z_i hanya bergantung pada intejer sebelumnya Z_{i-1} dan jika $0 \leq z_i \leq m$ maka periode paling besar yang mungkin adalah m; dan jika suatu pembangkit memiliki peiode m, maka LCG tersebut dikatakan memiliki periode penuh. Suatu LCG yang baik haruslah memiliki periode yang penuh atau sedikitnya panjang.

2.3 Metode Bootstrap

Metode bootstrap pertama kali diperkenalkan oleh Bradley Efron dalam Annals of Mathematical Statistics pada tahun 1979 sebagai suatu metode pendekatan non parametrik untuk menaksir berbagai kuantitas statistik seperti mean, standar error, dan bias suatu estimator atau untuk membentuk interval konfidensi. Metode bootstrap dapat juga digunakan untuk mengestimasi distribusi suatu statistik. Distribusi ini diperoleh dengan menggantikan distribusi populasi yang tidak diketahui dengan distribusi empiris berdasarkan data sampel, kemudian melakukan resampel kembali dari distribusi empiris untuk dipergunakan dalam mencari penaksir bootstrap. Dengan metode bootstrap tidak perlu melakukan asumsi distribusi dan asumsi-asumsi awal untuk menduga bentuk distribusi dan pengujian-pengujian statistiknya. Dengan demikian penarikan kesimpulan melalui metode ini memberikan hasil taksiran kuantitas statistik yang lebih baik apabila asumsi-asumsi yang diberikan tidak jelas bahkan mungkin kurang realistik untuk diterapkan terhadap suatu populasi.

Masalah inferensi statistik seringkali melibatkan pengestimasian beberapa aspek dari suatu sebaran peluang F berdasarkan pada suatu sampel acak yang ditarik dari F. Fungsi sebaran empiris, $\hat{F}_n$, adalah suatu estimasi sederhana dari sebaran F. Suatu cara yang tepat untuk mengestimasi beberapa aspek yang diinginkan dari F, seperti mean, median, atau korelasi, adalah dengan menggunakan aspek dari $\hat{F}_n$. Inilah yang disebut prinsip penggantian (Plug-in). Metode bootstrap adalah suatu aplikasi langsung dari prinsip penggantian ini.

Pandang $x_1,x_2,\ldots,x_n$ sampel acak berukuran n dari populasi dengan sebaran F yang tidak diketahui, dimana $x_1,x_2,\ldots,x_n$ merupakan peubah acak bebas yang tersebar secara identik (var. iid), maka notasi :

$$\theta = t(F) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.1)$$

merupakan nilai sebenarnya dari parameter yang akan diestimasi. Parameter θ dapat dipandang sebagai nilai fungsi sebenarnya dari F. Suatu estimator $\hat{\theta}$ untuk θ adalah nilai fungsi dari data $x_1,x_2,\ldots,x_n$. Sebaran sebenarnya, F, tidak diketahui, akan tetapi berdasarkan sampel $x_1,x_2,\ldots,x_n$ maka F dapat diestimasi menggunakan fungsi sebaran empiris $\hat{F}_n$. Fungsi sebaran empiris $\hat{F}_n$ dari $x_1,x_2,\ldots,x_n$ didefinisikan sebagai :

$$\hat{F}_n(x) = \frac{1}{n}\{\#(x_i \le x), 1 \le i \le n\} \text{ untuk } -\infty < x < \infty \quad \ldots\ldots\ldots\ldots(2.2)$$

Prinsip penggantian adalah suatu metode sederhana untuk pengestimasian parameter dari sampel. Estimasi plug-in dari suatu parameter $\theta = t(F)$ didefinisikan sebagai

$$\hat{\theta} = t(\hat{F}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3)$$

Dengan kata lain, mengestimasi fungsi $\theta = t(F)$ dari sebaran peluang F sama saja dengan mengestimasi parameter dari sebaran empiris $\hat{F}_n$, yaitu $\hat{\theta} = t(\hat{F})$. Untuk selanjutnya istilah estimasi pengganti (plug-in) disebut estimator saja, dengan lambang "^". Jadi estimator untuk parameter θ adalah $\hat{\theta}$.

Untuk menjelaskan metode bootstrap dapat dibayangkan sebagai suatu masalah real (nyata) dan satu masalah buatan yang sangat mirip atau bisa dikatakan identik. Masalah buatan inilah yang disebut dengan masalah bootstrap. Gambar berikut ini dapat menjelaskan gambaran dari metode bootstrap tersebut.

DUNIA RIIL	DUNIA BOOTSTRAP
Distribusi peluang tak diketahui Data observasi	Distribusi Empiris Sampel bootstrap
$F \rightarrow x = (x_1, x_2, x_3, \ldots, x_n)$ $\Rightarrow$	$\hat{F} \rightarrow x^* = (x_1^*, x_2^*, x_3^*, \ldots, x_n^*)$
$\downarrow$	$\downarrow$
$\hat{\theta} = s(x)$	$\hat{\theta}^* = s(x^*)$
Statistik yang menjadi perhatian	Replikasi bootstrap

Gambar 2.1. *Diagram skema dari metode bootstrap untuk kasus satu sampel. Dalam dunia real distribusi peluang yang tidak diketahui F memberikan data* $x = (x_1, x_2, \ldots x_n)$ *melalui sampling random; dari x dihitung statistik yang menjadi perhatian* $\hat{\theta} = s(x)$. *Dalam dunia bootstrap,* $\hat{F}$ *membangkitkan* x^* *melalui sampling random, memberikan* $\hat{\theta}^* = s(x^*)$.

Dalam masalah real terdapat data $x_1, x_2, \ldots, x_n$ dan suatu model menyatakan bahwa data tersebut merupakan hasil dari suatu pengamatan bebas dari suatu distribusi F(x). Masalah bootstrap menyerupai masalah real, tetapi dengan mengganti distribusi F(x) dengan distribusi empiris $\hat{F}_n(x)$. Selanjutnya data $x_1, x_2, \ldots, x_n$ disimulasikan sehingga diperoleh sampel bootstrap. Sampel bootstrap didefinisikan sebagai sampel random berukuran n yang diambil dari $\hat{F}_n(x)$ dengan pengembalian, ditulis sebagai :

$$x_1^*, x_2^*, x_3^*, \ldots, x_n^* \text{ dimana } x_i^* \in \hat{F}_n(x) \qquad (2.4)$$

Sehingga ada n^n kombinasi yang mungkin untuk sampel bootstrap, bisa saja didapatkan $x_i^* = x_j^*$, untuk $i \neq j$. Pada sampel bootstrap, x_i bisa muncul nol kali, bisa satu kali dan maksimal n kali.

Setelah sampel bootstrap diperoleh, kemudian menghitung estimasi parameter

$$\hat{\theta}^* = \hat{\theta}(x_1^*, x_2^*, \ldots, x_n^*) \qquad (2.5)$$

dengan estimator yang sama dengan masalah real. Misal jika terdapat sampel bootstrap $x_1^*, x_2^*, \ldots, x_n^*$ merupakan sampel random independen berukuran n yang ditarik dari suatu distribusi empiris, maka menurut Robert V. Hogg dan Allen T. Craig taksiran atau estimasi parameter untuk rata-rata (mean) sampel dan varian sampel bootstrap adalah sebagai berikut :

Untuk mean sampel,

$$E(X^*) = \sum_{i=1}^{n} x_i^* P_i$$

karena fungsi distribusi empiris memberikan peluang 1/n untuk setiap titik data $x_1^*, x_2^*, \ldots, x_n^*$ maka $P_i = \frac{1}{n}$, untuk semua $i = 1,2,\ldots,n$, sehingga

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

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

$$= \overline{X}^*$$

Untuk varian sampel,

$$\mathrm{var}(X^*) = E\left[\left(X^* - E(X^*)\right)^2\right]$$

karena $E(X^*) = \overline{X}^*$ dan diketahui setiap data sampel mempunyai peluang 1/n untuk terambil sebagai sampel random (berdasarkan prinsip resampling dengan pengembalian) maka

$$\mathrm{var}(X^*) = \sum_{i=1}^{n} (x_i^* - \overline{x}^*)^2 P_i$$

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

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

$$= S^{*2}$$

Varian sampel random tersebut merupakan suatu estimator yang tidak unbias. Penekanan pada hasil pembootstrapan adalah sebaran frekuensi relatif dari $\hat{\theta}^*$ yang dihitung dari resampel merupakan dugaan dari sebaran sampling $\hat{\theta}$.

Selanjutnya didefinisikan suatu statistik, yaitu :

$$Q = \sqrt{n}(\hat{\theta} - \theta) \quad \text{...(2.6)}$$

Gagasan pendefinisian statistik ini muncul dari teorema limit pusat yang menyatakan bahwa jumlah dari peubah acak dikurangi ekspektasinya mempunyai distribusi normal simetris disekitar nol untuk n yang cukup besar. Maka pendekatan bootstrap berdasarkan statistik tersebut adalah :

$$Q^* = \sqrt{n}(\hat{\theta}^* - \hat{\theta}) \quad \text{...(2.7)}$$

Fungsi distribusi dari Q adalah :

$$G_n(x) = P(Q(x_1,...,x_n;F) \leq x)$$

$$= P((\sqrt{n}(\hat{\theta} - \theta)) \leq x) \quad \text{...(2.8)}$$

dan fungsi distribusi dari Q^* adalah ;

$$G_n(x^*) = P^*(Q^*(x_1^*,...,x_n^*;\hat{F}) \leq x)$$

$$= P^*((\sqrt{n}(\hat{\theta}^* - \hat{\theta})) \leq x) \quad \text{...(2.9)}$$

Oleh karena $x_1^*, x_2^*, x_3^*,..., x_n^*$ merupakan sampel bootstrap yang diambil dari $x_1, x_2, x_3,..., x_n$ berdasarkan $\hat{F}_n$, distribusi empiris dari $x_1, x_2, x_3,..., x_n$, maka dengan sendirinya G_n^* dapat diketahui karena didefinisikan berdasarkan $\hat{F}_n$. Jadi G_n^* merupakan estimator bootstrap dari G_n. G_n^* dapat dipandang sebagai estimasi plug-in dari G_n dengan meletakan $\hat{F}_n$ ke F dan x_i^* ke x_i. Perhitngan selanjutnya secara eksak G_n^* berdasarkan semua kemungkinan sampel bootstrap inilah yang disebut dengan "bootstrap ideal". Dengan simulasi monte carlo dapat diperoleh pengestimasian G_n^* yang cukup baik, artinya cukup dekat ke G_n.

Langkah-langkah dasar dalam metode bootstrap adalah sebagai berikut :

1. Menentukan distribusi empiris dari F, yaitu $\hat{F}_n(x)$ dengan memberi peluang 1/n untuk masing-masing titik $x_1, x_2, ..., x_n$.
2. Menentukan sampel bootstrap $x_1^*, x_2^*, ..., x_n^*$.

Dari $\hat{F}_n(x)$ yang telah ditentukan, diambil sampel bootstrap ukuran n secara random dari $x_1,x_2,...,x_n$ dengan pengembalian, sebut nilai sampelnya $x_1^*,x_2^*,...,x_n^*$.

3. Menentukan statistik bootstrap $\hat{\theta}^*$.

   Dari $x_1^*,x_2^*,...,x_n^*$ yang diperoleh pada langkah ke-2, selanjutnya hitung statistik bootstrap untuk menghasilkan $\hat{\theta}^*$.

4. Menentukan $Q^* = Q^*(x_1^*,x_2^*,...,x_n^*;\ \hat{F}_n)$.

   Yaitu dengan menghitung $Q^* = \sqrt{n}(\hat{\theta}^* - \hat{\theta})$

5. Menentukan distribusi Q^*.

   Kita pelajari sifat-sifat dari Q^*. Misalnya menentukan rata-rata dari Q^*.

6. Ulangi langkah 2, 3, 4, dan 5 sebanyak B kali, untuk B yang cukup besar.

7. Berikan sebaran peluang dari B $\hat{\theta}^*$ dengan menempatkan peluang bagi masing-masing $\hat{\theta}^{*1},\hat{\theta}^{*2},...,\hat{\theta}^{*B}$. Sebaran ini adalah estimasi bootstrap untuk sebaran sampling $\hat{\theta}$.

Sebagai contoh, misalkan $x = (x_1,x_2,x_3)$ sampel random berukuran n = 3 yang diambil dari suatu distribusi probabilitas F yang tidak diketahui dan $x = (x_1,x_2,x_3) = (3, 6, 9)$ adalah hasil pengamatan. Selanjutnya akan ditaksir distribusi sampling dari $Q = Q(x_1,x_2,...,x_n, F) = \sqrt{n}(\hat{\theta} - \theta)$, maka yang harus dilakukan adalah :

1. $\hat{F}_n(x)$ memberi peluang 1/3 untuk setiap (3, 6, 9).

2. Menurut ketentuan dari $\hat{F}_n(x)$ diambil sampel bootstrap berukuran n = 3, maka x^* yang mungkin adalah :

   {(3,6,9) , (3,9,6) , (9,6,3) , (9,3,6) , (6,3,9) , (6,9,3) , (3,3,6) , (3,3,9) , (6,6,9) , (6,6,3) , (9,9,3) , (9,9,6) , (6,3,3) , (9,3,3) , (9,6,6) , (3,6,6) , (3,9,9) , (6,9,9) , (3,6,3) , (3,9,3) , (6,9,6) , (6,3,6) , (9,3,9) , (9,6,9) , (9,9,9) , (3,3,3) , (6,6,6)}.

3. Tentukan $\hat{\theta} = t(\hat{F})$ dari $\bar{x} = \sum_{i=1}^{3} \frac{x_i}{3}$ yaitu $\hat{\theta} = \frac{3+6+9}{3} = 6$.

4. Dari x^* ditentukan $\hat{\theta}^*$.

   Untuk setiap pengambilan sampel bootstrap akan dihitung statistik bootstrap $\hat{\theta}_b^* = \sum_{i=1}^{3} \frac{x_i^*}{3}$ untuk b = 1,2,...,B, yaitu :

   $$\hat{\theta}_1^* = \frac{3+6+9}{3} = 6$$

   $$\hat{\theta}_2^* = \frac{3+9+6}{3} = 6$$

   $$\vdots$$

   $$\hat{\theta}_{27}^* = \frac{6+6+6}{3} = 6$$

5. Menentukan Q^*, yaitu $Q^* = Q(x_1^*, x_2^*, \ldots, x_n^*;\ \hat{F}_n) = \sqrt{n}(\hat{\theta}^* - \hat{\theta})$, dimana $\hat{\theta}^*$ adalah rataan dari salah satu 27 kemungkinan x^* yang mungkin.

   $$Q_1^* = \sqrt{3}(3-6) = -5{,}196152$$

   $$Q_2^* = \sqrt{3}(4-6) = -3{,}4641016$$

   $$Q_3^* = \sqrt{3}(5-6) = 1{,}7320508$$

   $$\vdots$$

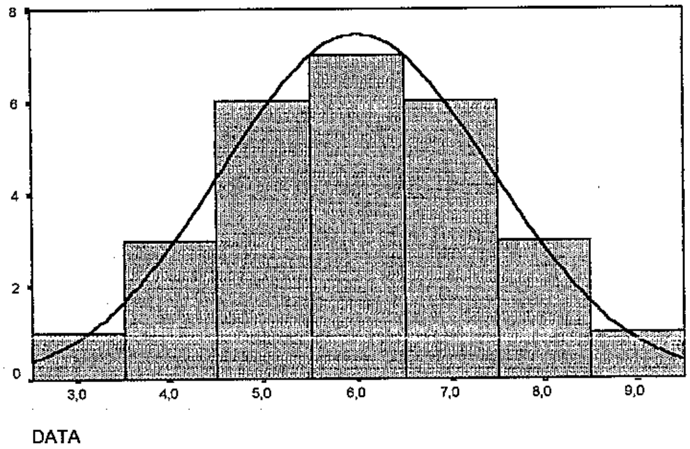
$Q_{27}{}^{*} = \sqrt{3}(7-6) = 1{,}7320508$

6. Hitung distribusi bootstrap dari $Q^* = Q(x_1{}^*, x_2{}^*, \ldots, x_n{}^*;\ \hat{F}_n)$ yang merupakan estimasi dari distribusi sampling $\hat{\theta}$.

Tabel 2.2.Tabel sebaran frekuensi rataan resampling untuk $(x_1,x_2,x_3) = (3, 6, 9)$.

i	$\hat{\theta}_{3i}{}^{*}$	$\sqrt{n}(\hat{\theta}_{3i}{}^{*} - \hat{\theta})$	Frekuensi(f_i)	Frekensi Relatif(P_i)	P_i Kumulatif
1	3	-5,1965	1	1/27	1/27
2	4	-3,4641	3	3/27	4/27
3	5	-1,732	6	6/27	10/27
4	6	0	7	7/27	17/27
5	7	1,7320	6	6/27	23/27
6	8	3,46410	3	3/27	26/27
7	9	5,1961	1	1/27	1

Resampling dilakukan dengan pengembalian untuk semua kemungkinan sampel. Rataan setiap sampel bootstrap ($\hat{\theta}_{3i}{}^{*}$) menyebar seperti pada Tabel 2.2 di atas. Sedangkan grafik dari sebaran frekuensi rata-rata resampling dapat dilihat pada Gambar 2.2 di bawah ini.



Gambar 2.2. *Grafik histogram dari rataan booststrap untuk data x = (3,6,9).*

Penaksir bootstrap $G_n^*(x) = P^*(\sqrt{3}(\hat{\theta}_{3i}^* - \hat{\theta}) \le x)$ untuk setiap x bagi $\sqrt{3}(\hat{\theta} - \theta)$ adalah $G_n^*(x) = \sum_{w_i \le x}^{9} P_i$ dengan $w_i = \sqrt{3}\left(\hat{\theta}_{3i}^* - \hat{\theta}\right)$, $i = 1,2,\ldots,9$. Semakin besar n semakin baik taksirannya (hal ini dikaitkan apabila grafik histogramnya mendekati bentuk grafik distribusi normal).