BAB II

MATERI PENUNJANG

Pada bab ini akan dibahas beberapa materi penunjang pada simulasi untuk estimasi parameter dalam metode bootstrap, diantaranya adalah tentang dasar-dasar simulasi dan langkah-langkah dasar untuk pembootstrapan.

2.1 Dasar-Dasar Simulasi

Simulasi adalah suatu metode yang digunakan untuk pengamatan sistem. Sistem adalah sekumpulan obyek yang tergabung dalam suatu interaksi atau kesalingtergantungan yang teratur. Untuk mempelajari sistem secara ilmiah, harus dibuat sekelompok asumsi tentang bagaimana cara kerja sistem tersebut. Asumsi-asumsi ini , yang biasanya menggunakan hubungan secara matematika atau logika, merupakan suatu model untuk mendapatkan beberapa pengertian tentang bagaimana kerja sistem yang bersangkutan. Dengan demikian simulasi dapat didefinisikan sebagai suatu proses pembuatan model dari suatu sistem nyata dan menghubungkan eksperimen dengan model ini untuk tujuan pengertian kerja dari sistem atau mengevaluasi berbagai macam strategi untuk pengoperasian sistem.

Jika hubungan yang menyusun model tersebut cukup sederhana, maka dapat digunakan metode matematika (seperti aljabar, kalkulus, atau teori peluang) untuk memperoleh informasi yang tepat pada masalah tersebut; ini disebut solusi analitik. Akan tetapi, sebagian besar sistem (dunia nyata) terlalu kompleks jika menggunakan model realistik untuk dievaluasi secara analitik, dan model ini harus dipelajari dengan menggunakan simulasi. Dalam suatu simulasi , komputer digunakan untuk mengevaluasi suatu model secara numerik, dan data dikumpulkan untuk mengestimasi karakteristik model yang diinginkan.

Setiap pelaksanaan simulasi pada dasarnya merupakan suatu eksperimen pada sistem. Keuntungan simulasi adalah bahwa eksperimen tersebut dapat dikontrol dan diamati secara lengkap.

Jika model sistem dapat dianalisa menggunakan teknik matematika, berarti metode tersebut lebih unggul daripada metode simulasi. Sebenarnya solusi analitik lebih akurat, lebih banyak memberikan informasi , dan biasanya lebih mudah untuk memperoleh hasilnya daripada hasil simulasi.

Alasan utama dari penggunaan simulasi adalah banyak model yang tidak dapat dianalisa secara tepat dengan teknik matematika standar . Pengalaman menunjukkan bahwa sering terdapat kesukaran dalam memecahkan suatu persoalan karena karakteristik dari permasalahan tidak dapat diselesaikan dengan rumus-rumus yang ada, atau persamaan dari permasalahan yang terlalu rumit untuk dipecahkan secara analitik.

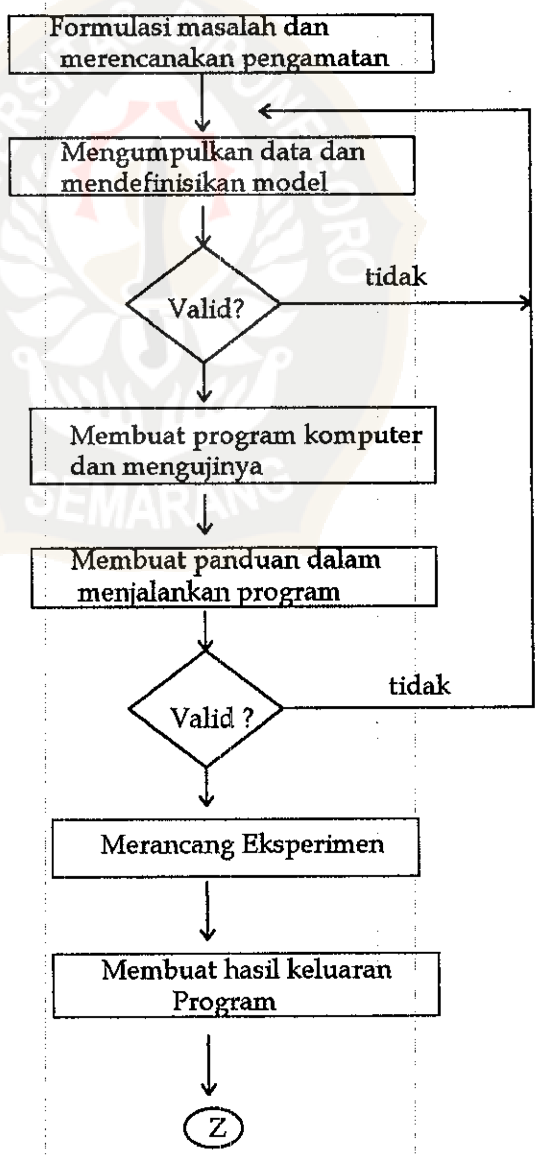
Langkah-langkah dalam proses simulasi adalah sebagai berikut:

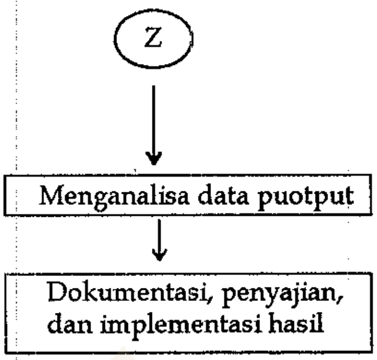
1. *Memformulasikan masalah dan merencanakan pengamatan.* Harus dilakukan identifikasi situasi masalah yang membuat perkembangan dari model dan sasaran dari pengamatan harus dinyatakan dengan jelas.
2. *Mengumpulkan data dan mendefinisikan sebuah model.* Informasi dan data harus dikumpulkan pada sistem yang diinginkan dan digunakan untuk menspesifikasi pengoperasian prosedur dan sebaran peluang untuk peubah acak yang digunakan dalam model.
3. *Test dan validasi model.* Ini merupakan langkah penting dalam simulasi, yaitu untuk meyakinkan bahwa model yang telah dikembangkan secara realistik menyajikan masalah yang sedang dipelajari, sehingga menghasilkan penyelesaian yang benar-benar tepat dan dapat dipercaya.
4. *Membuat suatu program komputer dan mengujinya.* Program komputer dapat dibuat dengan bahasa pemrograman yang bersifat serba guna (*general purpose language*) atau dalam suatu bahasa yang khusus dirancang untuk simulasi.
5. *Membuat panduan dalam menjalankan program komputer (pilot runs).* Panduan dalam menjalankan komputer dapat digunakan untuk

menguji sensitivitas output model terhadap perubahan kecil dalam suatu parameter input.

6. *Valid ?* Jika suatu sistem yang mirip dengan suatu sistem yang diinginkan telah ada , data output dari panduan dalam menjalankan komputer untuk suatu model dari sistem yang sudah ada dapat dibandingkan dengan output data dari sistem aktual yang ada. Jika perjanjiannya baik , model yang "divalidasi" dimodifikasi sedemikian sehingga mewakili sistem yang diinginkan.
7. *Merancang Eksperimen.* Untuk setiap rancangan sistem yang akan disimulasi, keputusan harus dibuat dalam hal seperti kondisi awal untuk menjalankan simulasi, panjang dari suatu periode, panjang dari pelaksanaan simulasi, dan jumlah replikasi bebas yang dibuat untuk setiap alternativ.
8. *Menghasilkan keluaran suatu pelaksanaan program (production runs).* Pada langkah ini diberikan data pada rancangan sistem yang digunakan.
9. *Menganalisa data output.* Teknik statistik digunakan untuk menganalisa data output dari hasil keluaran pelaksanaan program. Tujuan khususnya adalah untuk menentukan sistem simulasi mana yang terbaik .

10. *Dokumentasi, penyajian , dan implementasi hasil.* Karena model simulasi sering digunakan untuk lebih dari satu aplikasi, maka penting untuk mendokumentasi asumsi-asumsi yang mengarah pada model sebagaimana program komputer itu sendiri.





Gambar 2.1. Langkah-langkah dalam proses simulasi

2.2 Estimasi Parameter

Estimasi parameter merupakan salah satu bagian besar dari teknik statistik induktif. Statistik induktif merupakan proses memperoleh informasi dari data sampel yang digunakan untuk menarik kesimpulan tentang populasi dari sampel yang dipilih.

Suatu perkiraan tunggal pada sebuah parameter populasi adalah nilai tunggal numerik pada sebuah statistik yang berhubungan dengan parameter tersebut. Perkiraan tunggal adalah sebuah pemilihan yang unik untuk sebuah nilai parameter populasi yang tidak diketahui. Lebih jelasnya , jika X sebuah peubah acak dengan sebaran peluang $f(x)$, mempunyai parameter θ yang tidak diketahui , dan jika X_1 , X_2 , ... , X_n sebuah sampel acak yang besarnya n dari X , maka statistik

$\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ yang berhubungan dengan θ disebut *estimator* θ. Perhatikan bahwa estimator θ adalah sebuah peubah acak, karena estimator tersebut merupakan sebuah fungsi data sampel. Setelah sampel dipilih, $\hat{\theta}$ diperoleh berdasarkan nilai tertentu yang disebut perkiraan tunggal θ ..

Sebagai contoh, misalkan sebuah peubah acak X yang memiliki sebaran normal dengan rataan μ yang tidak diketahui dan varian σ^2 diketahui. Rataan sampel $\overline{X}$ adalah suatu estimator tunggal dari rataan populasi μ yang tidak diketahui. maka $\hat{\mu} = \overline{X}$. Sesudah sampel dipilih, nilai numerik $\bar{x}$ adalah perkiraan tunggal μ. Demikian juga jika varian populasi σ^2 adalah varian sampel S^2, dan nilai numerik s^2 dihitung berdasarkan data sampel yang merupakan perkiraan tunggal σ^2.

Jika rataan sebuah peubah acak akan diestimasi, harus diperhatikan salah satu dari rataan sampel, median sampel, atau mungkin observasi yang paling kecil dalam sampel sebagai estimator tunggal.

Sebuah sifat yang diinginkan estimator yaitu menjadi "tertutup" dalam beberapa pengertian pada nilai sebenarnya dari parameter yang tidak diketahui. Singkatnya dapat dikatakan bahwa $\hat{\theta}$ merupakan sebuah estimator yang *unbias* dari parameter θ jika

$$E(\hat{\theta}) = \theta \quad (2.1)$$

Yaitu , $\hat{\theta}$ merupakan sebuah estimator yang unbias dari θ , jika "pada rataan" nilainya adalah sama untuk θ . Dengan kata lain, rataan sebaran sampling $\hat{\theta}$ sama dengan θ .

Dalam beberapa keadaan, sebuah perkiraan tunggal tidak memberikan cukup penjelasan tentang parameter yang menarik. Sebagai contoh dalam memperkirakan kekuatan tekanan beton, sebuah bilangan tunggal mungkin tidak mempunyai arti. Sebuah perkiraan interval dengan bentuk $L \le \mu \le U$ mungkin lebih berguna.

Pada umumnya untuk membuat suatu estimator interval parameter tidak diketahui θ, harus ditentukan dua statistik L dan U sebagai berikut :

$$P\{L \le \theta \le U\} = 1 - 2\alpha \quad (2.2)$$

Interval yang dihasilkan adalah

$$L \le \theta \le U \quad (2.3)$$

disebut *interval keyakinan* $100(1 - 2\alpha)$ persen untuk parameter θ tidak diketahui. L dan U disebut batas keyakinan atas dan bawah, dan $1 - 2\alpha$ disebut koefisien keyakinan. Interpretasi sebuah interval keyakinan yaitu jika beberapa sampel acak dikumpulkan dan sebuah interval keyakinan

$100(1-2\alpha)$ persen pada θ dihitung dari setiap sampel, maka $100(1-2\alpha)$ persen interval ini akan berisi nilai θ sebenarnya.

Panjang interval keyakinan yang diobservasi adalah sebuah ukuran penting kualitas informasi yang diperoleh dari sampel. Luas setengah interval $\theta - L$ atau U-θ disebut ketepatan estimator. Lebih panjang interval keyakinan, lebih meyakinkan bahwa interval tersebut sebenarnya berisi nilai θ sebenarnya. Di lain pihak , lebih panjang interval, lebih kecil informasi yang diperoleh mengenai θ sebenarnya. Pada suatu keadaan yang ideal, dapat diperoleh suatu interval yang pendek dengan tingkat keyakinan yang tinggi.

2.3 Fungsi Sebaran Empiris dan Prinsip Penggantian (*Plug-in*)

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 , yaitu menggunakan aspek dari $\hat{F}_n$. Inilah yang disebut prinsip penggantian (*plug-in*). Metode bootstrap adalah suatu aplikasi langsung dari prinsip *plug-in* ini.

Pandang X_1 , X_2 , ... ,X_n sampel acak berukuran n dari populasi dengan sebaran F yang tidak diketahui ,di mana X_1 , X_2 , ... ,X_n merupakan peubah acak bebas yang tersebar secara identik (variabel i.i.d), maka notasi :

$$\theta = \theta(F) \qquad (2.4)$$

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 ,... ,x_n . Sebaran sebenarnya ,F, tidak diketahui, akan tetapi berdasarkan sampel X_1 , X_2 , ... ,X_n maka F dapat diestimasi menggunakan fungsi sebaran empiris $\hat{F}_n$.

Fungsi sebaran empiris $\hat{F}_n$ dari x_1 ,x_2 , ... ,x_n didefinisikan sebagai :

$$\hat{F}_n(x) = \frac{1}{n}\{\#\ \{x_i \leq x\},\ 1 \leq i \leq n\}, \text{untuk } -\infty < x < \infty \qquad (2.5)$$

Prinsip penggantian adalah suatu metoda sederhana untuk pengestimasian parameter dari sampel. Estimasi penggantian dari suatu parameter $\theta = t(F)$ didefinisikan sebagai :

$$\hat{\theta} = t\,(\hat{F}) \qquad (2.6)$$

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 estimator pengganti disebut estimator saja, dengan lambang "^". Jadi estimator dari parameter θ adalah $\hat{\theta}$.

2.4 Ukuran Klasik dan Metode Bootstrap

Misal $x_1, x_2, \ldots, x_n$ adalah n observasi dari suatu objek, yang merupakan peubah acak bebas dengan sebaran F(x). Misal θ bernilai real, adalah parameter skalar yang akan diestimasi dengan suatu fungsi dari observasi, $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$. Nilai harapan dari estimator diberikan dengan integral :

$$E_F[\hat{\theta}] = \int \ldots \int \hat{\theta}(x_1, x_2, \ldots, x_n) f(x_1) \ldots f(x_n) dx_1 \ldots dx_n \qquad (2.7)$$

Notasi E_F menandakan bahwa nilai harapan dihitung dengan peluang F. Bias estimator didefinisikan sebagai :

$$\text{Bias} = E_F\left[\hat{\theta} - \theta\right] \qquad (2.8)$$

dan dapat berupa sebuah fungsi dari parameter θ (dan parameter lain yang mungkin dalam F). Estimator $\hat{\theta}$ unbias jika $E_F\left[\hat{\theta} - \theta\right] = 0$ untuk semua θ atau jika ada parameter lain untuk semua vektor parameter dalam model. Variansi $\hat{\theta}$ yaitu :

$$\mathrm{Var}(\hat{\theta}) = E_F\left[\left(\hat{\theta} - E_F[\hat{\theta}]\right)^2\right] \tag{2.9}$$

dan standar deviasi $\sigma_{\hat{\theta}}$ adalah akar pangkat dari variansi.

Ketika parameter diestimasi, sebaran F tidak diketahui (asing). Sebaran estimator dan khususnya nilai harapan pada persamaan (2.8) dan (2.9), didefinisikan dengan sebaran asing ini. Oleh karena itu, mustahil untuk menghitung dengan tepat. Metode inferensi statistik mencoba berbagai cara untuk mengatasi masalah ini, dan akhirnya ditemukan suatu gagasan dengan mengganti sebaran asing F(x) dengan sebaran empiris $F_n(x)$ untuk menghitung persamaan (2.8) dan (2.9) dan untuk penghitungan - penghitungan lainnya. Sebaran empiris adalah ukuran peluang yang mengarah pada suatu kelompok dari suatu ukuran yang sama dengan proporsi nilai sampel yang terdapat dalam kelompok tersebut. Gagasan inilah yang selanjutnya disebut dengan metode bootstrap.

Bootstrap merupakan suatu metode berbasis komputer untuk penentuan ukuran ketepatan estimasi statistik. Pembootstrapan bertitik tolak atas dasar analogi antara sampel dan populasi dari mana sampel tersebut diambil. Kadang-kadang pembootstrapan penarikan kesimpulan akan memberikan hasil yang lebih baik apabila asumsi yang ada tidak jelas dan mungkin kurang realistik untuk diterapkan pada populasi ini. Selain itu sampel sudah di tangan. Untuk memperoleh suatu dugaan empiris dari sebaran statistik sampling dilakukan pembootstrapan yang meliputi "resampling" data dengan pengembalian beberapa kali. Jika diketahui observasi $x_1, x_2, \ldots, x_n$ maka sampel buatan yang diambil dengan pengembalian dari $x_1, x_2, \ldots, x_n$ akan mempunyai titik sampel dengan peluang yang sama untuk terpilih yaitu 1/n dari setiap x_i, $i=1,2,\ldots,n$. Dari setiap sampel bootstrap ini dihitung $\hat{\theta}^*$ (lihat gambar 2.1 dan 2.2). Penekanan pada hasil pembootstrapan adalah bahwa sebaran frekuensi relatif dari $\hat{\theta}^*$ yang dihitung dari resampel merupakan dugaan dari sebaran sampling $\hat{\theta}$.

Langkah-langkah dasar untuk pembootstrapan adalah :

a) Berikan sebaran peluang empiris, $\hat{F}_n(x)$ bagi sampel dengan peluang 1/n untuk masing-masing titik $x_1, x_2, \ldots, x_n$. $\hat{F}_n(x)$ adalah fungsi sebaran empiris dari x.

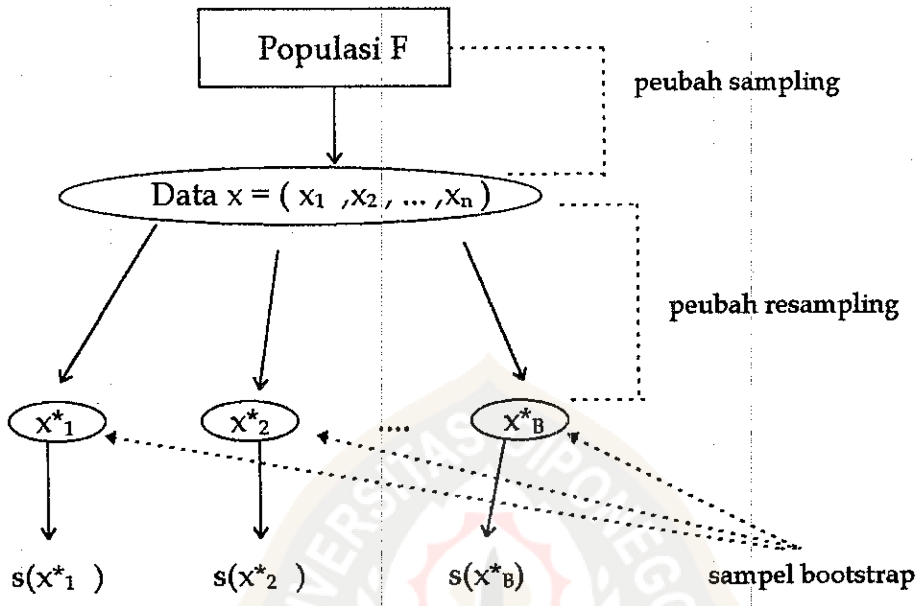
b) Dari $\hat{F}_n(x)$ tarik sampel acak sederhana berukuran n dengan pengembalian . Ini merupakan "resampel" ; x^*_b.

c) Hitung statistik yang diinginkan $\hat{\theta}$, dari resampel, dan menghasilkan $\hat{\theta}^*_b$.

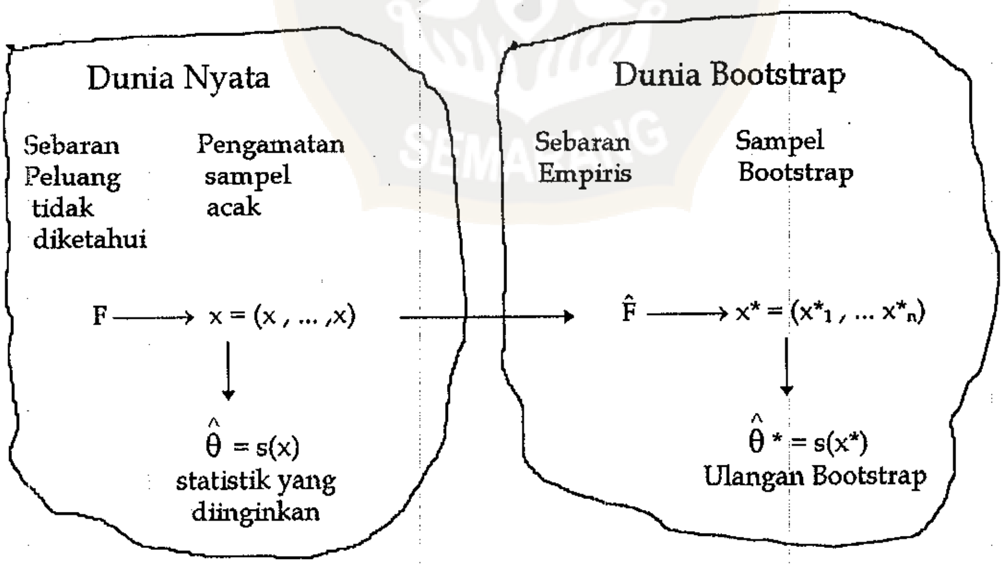
d) Ulangi langkah b dan c sebanyak B kali, untuk B cukup besar.

e) Berikan sebaran peluang dari B $\hat{\theta}^*_b$ dengan menempatkan peluang bagi masing-masing $\hat{\theta}^*_1$, $\hat{\theta}^*_2$, ... , $\hat{\theta}^*_B$. Sebaran ini adalah dugaan bootstrap bagi sebaran sampling $\hat{\theta}$; $F^*(\hat{\theta}^*)$.

Tanda "*" menunjukkan bahwa x* bukan data sesungguhnya x, melainkan bentuk acak dari x atau hasil resampling dari data x secara acak dengan pengembalian. Dengan kata lain bahwa titik-titik data bootstrap x^*_1 ,x^*_2 , ... ,x^*_n adalah sampel acak berukuran n yang diambil secara acak dengan pengembalian dari populasi n obyek (x_1 ,x_2 , ... ,x_n) . Suatu set data bootstrap (x^*_1 ,x^*_2 , ... ,x^*_n) berisi anggota data sesungguhnya (x_1 ,x_2 , ... ,x_n) yang mungkin muncul nol kali, satu kali, atau lebih dari satu kali. Jadi ada n cara untuk memperoleh sampel bootstrap atau jumlah sampel bootstrap seluruhnya adalah n.n.n...n $=n^n$ buah sampel bootstrap. Metode bootstrap yang ideal adalah dengan memperlakukan seluruh kemungkinan sampel yang ada. Namun dapat dibayangkan , misalnya saja untuk n=20 maka akan terdapat 20^{20} kemungkinan sampel. Sehingga dapat pula dibayangkan berapa lamanya komputer bekerja. Ini menjadi tidak efektif lagi, sehingga akhirnya digunakan pendekatan simulasi Monte-Carlo, di mana semua kemungkinan sampel bootstrap dibuat menjadi sejumlah B yang cukup besar (tetapi cukup kecil jika dibandingkan dengan jumlah sampel bootstrap ideal).



Gambar 2.2 Diagram untuk komponen ragam dari sampling dan resampling



Gambar 2.3 Diagram dari Bootstrap, penerapan masalah satu sampel. Dalam dunia nyata sebaran peluang yang tidak diketahui F memberikan data $x=(x_1,x_2,\ldots,x_n)$ dengan pengambilan sampel secara acak; dari x dihitung statistik yang diinginkan $\hat{\theta}^* = s(x^*)$. Hanya ada satu nilai pengamatan $\hat{\theta}$, namun demikian replikasi bootstrap $\hat{\theta}^*$ dapat dibangkitkan sebanyak yang diinginkan.