Replikasi Signal dengan Menggunakan Metode Bootstrap

Siana Halim
Jurusan Teknik Industri, Fakultas Teknologi Industri
Universitas Kristen Petra, Surabaya
E-mail: halim@petra.ac.id

ABSTRAK

Signal dapat dimodelkan sebagai proses stokastik yang berperiode ataupun tidak berperiode. Untuk itu dalam mereplikasi sebuah signal, kita harus tetap menjaga karakter asli dari signal dan juga sifat keacakannya. Salah satu metode yang mungkin untuk dilakukan adalah bootstrap. Namun demikian, kita harus memodifikasi metode bootrap ini untuk mengakomodasi sifat ketergantungan dari series beserta periodisitasnya. Sebagai langkah awal dalam bootstrap ini diperlukan uji ada tidaknya periodisitas dalam signal. Diberikan dua metode untuk mendeteksi periodisitas, yaitu Fisher statistik dan Chiu statistik dan sebuah ilustrasi dengan menggunakan data simulasi untuk menguji dan mereplikasi sebuah signal.

Kata kunci: bootstrap, periodogram, fisher statistic

ABSTRACT

Signal can be modeled as a periodic or a nonperiodic stochastic process. Therefore to replicate a signal, we should keep the original character of the signal as well as the random character in it. One of plausible methods for doing such kind of job is bootstrap. However, we should modify the boostrap to accomodate the dependency in the series and their periodicities. As the pre bootraping we need to detect the existence of periodicities in the series. Two methods are given for detecting the existence of periodicities, i.e. the Fisher classical statistic, and the Chiu statistic. At the end we give an illustration. We used simulated data for testing and replicating a signal.

Keywords: *bootstrap, periodogram, fisher statistics*

PENDAHULUAN

Bootstrap merupakan alat bantu umum (*general tool*) yang biasa digunakan untuk mencari pendekatan dalam distribusi statistik yang dikehendaki. Bootstrap mengantikan atau bahkan seringkali memperbaiki hasil yang diperoleh berdasarkan analisa asymptotic secara klasik, terutama untuk data sampel yang berukuran kecil sampai menengah. Beberapa aplikasi yang menggunakan teknik bootstrap ini adalah untuk membandingkan dua random signal ataupun dua buah gambar random yang tidak bersih (*noisy image*) satu terhadap yang lain. Fanke dan Halim [1], menerapkan aplikasi ini untuk mendeteksi kerusakan pada texture. Pada makalah ini akan digunakan teknik Bootstrap untuk mereplikasi signal yang memiliki sifat periodisitas. Halim [2] menggunakan teknik ini untuk mensintesa texture yang memiliki sifat semiregular, yaitu texture yang memiliki sifat random dan regularitas didalamnya.

Untuk mereplikasi signal yang memiliki sifat periodisitas ini diperlukan tiga langkah, yaitu mendeteksi posisi periodisitas dari signal tersebut, mengujinya serta melakukan bootstrap sampling untuk data yang berperiodik. Langkah-langkah tersebut akan dijelaskan pada subbab-subbab berikut.

Catatan: Diskusi untuk makalah ini diterima sebelum tanggal 1 Desember 2007. Diskusi yang layak muat akan diterbitkan pada Jurnal Teknik Elektro volume 8, nomor 1, Maret 2008.

MENDETEKSI PERIODISITAS YANG TERSEMBUNYI

Pada dasarnya sebuah signal dapat kita modelkan sebagai fungsi periodik dari random variable yang mengandung error yang stasioner [3], yaitu

$$X_t = f(t) + \varepsilon_t \qquad (1)$$

dimana $X_1,\ldots,X_N$ adalah variable random dan $f(t+T) = f(t)$, $t = 1,\ldots,N$ dan T adalah periodisitas, ε_t adalah proses stasioner. Sebuah proses stokastik dikatakan stasioner (lemah) bila

- $E(\varepsilon_t) = \mu$
- $Var(\varepsilon_t) = \sigma^2$, $\sigma < \infty$
- $Cov(\varepsilon_{t+s}, \varepsilon_t) = \gamma_s$, $s \neq t$

Untuk mendeteksi periodisitas Chiu [3] mendekati persamaan (1) melalui fungsi regresi harmonik seperti di bawah ini

$$X_t = \mu + S(t) + \varepsilon_t, \quad t = 1,\ldots N \qquad (2)$$
$$\varepsilon_t \sim N(0, \sigma^2)$$

$$S(t) = \sum_{k=1}^{K} (A_k \, Cos\, \omega_k \, t + \phi_t) \qquad (3)$$

Terdiri dari K gelombang sinusoidal pada frekuensi $\omega_k \neq 0$ dengan amplitude $A_k \neq 0$, phase ϕ_k. K dapat saja sama dengan nol, untuk mengantisipasi bias signal tidak mengandung periodisitas. Tanpa menghilangkan sifat keumuman, asumsikan $\mu = 0$.

Untuk memprediksi letak dari periodisitas yang tersembunyi, seringkali kita menggunakan periodogram. Periodogram pula yang biasa digunakan sebagai alat untuk menguji keberadaan periodik komponen.

Definisi 1. Periodogram dari deret X_t didefinisikan sebagai

$$I_x(\omega) = d_x^{(N)}(\omega) d_x^{(N)}(-\omega) / 2\pi N \quad (4)$$

dimana

$$d_x^{(N)}(\omega) = \sum_{t=0}^{N-1} X_t \exp(-i\omega t)$$

adalah finit Fourier transform dari deret X_t □

Bila $I_x(\omega_j)$ adalah periodogram dari X_t pada frekuensi Fourier

$$\omega_j = \left(\frac{2\pi j}{N}\right), \ j = 0,1,\ldots,[N] \quad (5)$$

Dimana [N] = N/2 + 1 jika N adalah bilangan genap dan [N] = (N+1)/2 jika N adalah bilangan ganjil. Dengan mengurutkan $I_x(\omega)$ secara menaik, yaitu

$$I_1 < I_2 < \ldots < I_{[N]}$$

Maka Fisher – statistik dapat dirumuskan sebagai berikut

$$F = \frac{I_{[N]}}{\sum_{j=1}^{[N]} I_x(\omega_j)} \quad (6)$$

Fisher statistik inilah yang digunakan untuk menguji hipotesa keberadaan periodisitas dalam sebuah signal. Namun demikian Fisher statistik bukanlah statistik yang tangguh (*robust statistics*). Hal ini dapat dilihat bahwa Fisher test statistik (6) proportional terhadap nilai maksimum dari periodogram dibandingkan nilai sample mean dari periodogram. Seperti kita tahu bahwa sample mean sangatlah sensitive terhadap nilai pencilan (*outlier*).

Untuk mengatasi hal tersebut di atas Chiu [3] mengusulkan dua statistik untuk menguji hipotesa

H_o : X_t bukan signal harmonik (*Zero harmonics*)

H_1 : X_t memiliki h harmonik $(h \geq 1)$

Kedua statistik yang diusulkan oleh Chiu adalah

$$U(h) = \frac{I_{[N]-h+1}}{\sum_{i=1}^{\{N\}} I_i} \quad (7)$$

$$V(h) = \frac{I_{[N]-h+1}}{\sum_{i=1}^{\{N\}-h} I_i} \quad (8)$$

Persamaan (7) telah digunakan oleh Shimshoni [4] dan Lewis and Fieller [5] untuk mendeteksi k pencilan. Tetapi Chiu, membuang r buah nilai periodogram terbesar dari penyebut pada(7) dan sebagai gantinya menggunakan persamaan (8). Distribusi asimptotik dari $U(.)$ dan $V(.)$ untuk menguji H_0 dapat diturunkan seperti di bawah ini

$$\begin{aligned} Z_1(h) &= [N]U(h) - \ln([N]-h+1) \\ Z_2(h) &= c([N]-k)V(k) - \ln([N]-h+1) \end{aligned} \quad (9)$$

dimana

$$c = 1 + \frac{h\left(\ln\left(\frac{h}{[N]}\right)\right)}{[N]-h} \quad (10)$$

Misalkan

$$P_i(h) = \exp\{-\exp(-Z_i(h))\}\sum_{j=0}^{h-1}\exp\{-jZ_i(h)/j!\}, i = 1,2 \quad (11)$$

Jika tingkat kepercayaan yang kita pilih adalah α, maka H_0 ditolak apabila

$$P_i(h) > 1 - \alpha, \ i = 1, 2 \quad (12)$$

Melalui statistik di atas, maka keberadaan periodisitas dapat dideteksi, dan nilai periodenya dapat ditentukan sebagai berikut. Misalkan $\omega^* = \arg_\omega \max I_x(\omega_j)$ dimana i = 1,…, [N] dan ω adalah frekuensi angular. Maka $\hat{T} = \frac{2\pi}{\omega^*}$ adalah nilai estimasi dari T.

Catatan 1 Untuk menghitung nilai factorial, pada persamaan (11), yang sangat besar maka gunakanlah nilai logaritma seperti contoh di bawah ini

50! = exp(log(50!)) = exp (log1 + …+log50)

BOOTSTRAP UNTUK MEREPLIKASI SIGNAL

Telah kita ketahui bahwa metode bootstrap yang diperkenalkan oleh Efron [6] dapat bekerja dengan baik untuk data independen. Sampling untuk random variable independent yang didasarkan pada bootstrap $\mathbf{x}^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$ diperoleh dengan melakukan random sampling sebanyak n kali, dengan *replacement*, dari data aslinya. Misalkan, n = 7, bootstrap sampling yang kita peroleh bisa saja $\mathbf{x}^* = \left(x_5, x_7, x_5, x_1, x_4, x_1, x_7\right)$. Algoritma bootstrap untuk data independent dilakukan dengan membangkitkan bilangan random secara diskret uniform dengan *replacement*, misalnya, I ~ dunif(N,n), dimana N adalah jumlah data original dan n adalah jumlah data yang akan kita bangkitkan secara Bootstrap. Data pada index bangkitan inilah yang akan menjadi data bootstrap.

Namun demikian untuk melakukan resampling pada data yang memiliki ketergantungan dan periodisitas seperti pada signal, metode bootstrap ini haruslah dimodifikasi. Jika kita menggunakan bootstrap untuk data independen pada data periodik maka data bootstrap ini akan kehilangan sifat aslinya. Hal ini tidak dikehendaki dalam resampling data. Salah satu metode yang akan diadaptasi pada makalah ini adalah metode bootstrap untuk data time series musiman [7].

Replikasi signal dengan metode bootstrap dapat dilakukan dengan langkah-langkah sebagai berikut. Pertama-tama partisilah sample dari signal menjadi blok dengan panjang T, sedemikian hingga kita mendapakan $n = \lfloor N/T \rfloor$. T adalah nilai periodisitas yang telah kita dapatkan dan kita uji dengan menggunakan metode di atas, dan $\lfloor x \rfloor$ adalah nilai integer dari x. Algoritma untuk mereplikasi signal diberikan sebagai berikut :

Algoritma 1: Bootstrap untuk mereplikasi signal

- Misalkan k adalah jumlah blok bootstrap yang hendak dibangun
- Bangkitkan $b_0,\ldots,b_{k-1}$ random uniform pada himpunan $\{0,\ldots,n-1\}$
- Data bootstrap $X_1^*, X_2^*, \ldots, X_{kn}^*$ dapat dibangun dari sample mula–mula (*original sample*) sedemikian hingga

$$X^*_{mT+j} = X^*_{bmT+j} \tag{13}$$

dimana $m = \{0,\ldots,k\text{-}1\}$ dan $j = \{1,\ldots,nk\}$

ILUSTRASI

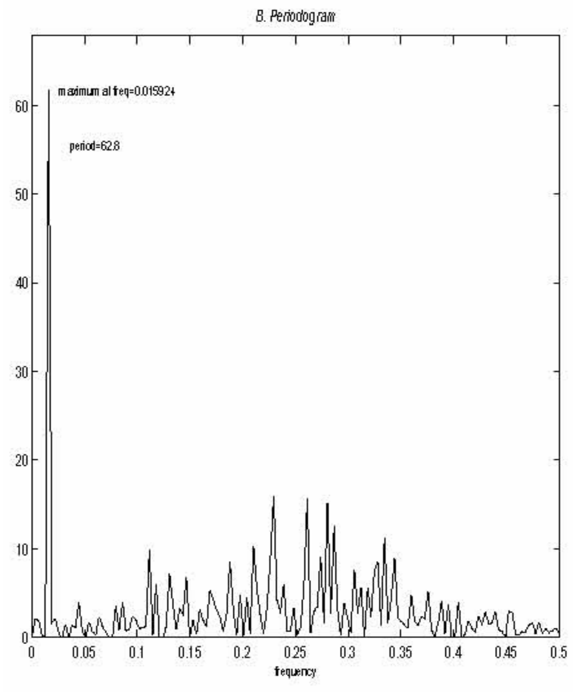
Sebagai ilustrasi dari metode ini, kita bangun random variable dari fungsi sinus dengan t = 1,…,310 dan periodisitas T = 62. Kita ganggu signal ini dengan menambahkan error yang berkorelasi. Error dibangun dari moving average dari random variable yang berdistribusi Gaussian dengan mean nol dan varians sama dengan dua.

$$X_t = \sin\left(\frac{2\pi t}{T}\right) + \varepsilon_t \tag{14}$$

Dimana ε_t dibangun sebagai berikut

$$\eta \sim N(0,2);$$
$$\varepsilon_t = 0.25\eta_t - 0.5\eta_{t-1}$$

Periodogram dari signal ini akan maximum pada frekuensi 0.015924. Hal ini berarti bahwa nilai periodisitas dari signal adalah 62.8 (atau 1/0.015924). Kita menggunakan versi pembulatan dari nilai periodisitas ini untuk mereplikasi signal dengan menggunakan metode bootstrap. Pada Gambar 1 di bawah ini dapat dilihat periodogram peserta nilai puncaknya.



Gambar 1. Periodogram dari signal X_t

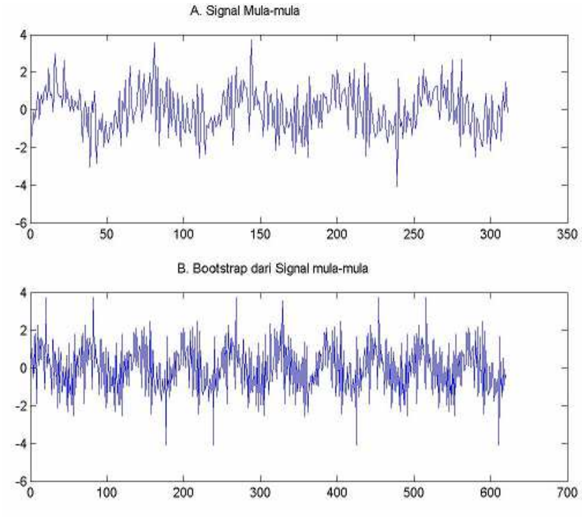
Fisher dan Chiu statistic untuk signal yang disimulasikan berdasarkan persamaan (14) adalah : F-statistik = 0.1529, dengan nilai p-value mendekati nol. Nilai Z_1 dan Z_2 dengan h=1 dan k=1 pada persamaan (9) adalah 33.7383 dan 32.8042 secara berturutan. Dengan demikian nilai P pada persamaan (11) dapat dihitung dan nilai keduanya adalah 1. Dengan mengambil nilai $\alpha = 0.05$, dapat disimpulkan bahwa untuk signal ilustrasi ini, baik dengan menggunakan Fisher statistic maupun Chiu statistik keduanya menolak H_0. Signal pada persamaan (14) memiliki satu nilai harmonik.

Selanjutnya dengan menggunakan Algoritma 1, kita dapat mereplikasi signal sinus di atas. Pada ilustrasi ini digunakan 10 blok, sehingga signal tiruan akan terdiri dari 620 series. Signal asli beserta tiruannya dapat dilihat pada Gambar 2.

Periodisitas pada signal bootstrap ini adalah 62, Fisher statistic = 0.1378 dengan p-value=0. Untuk Chiu statistik dengan h = 10, nilai Z_1 = -0.5180 dan Z_2 = -0.9734 dengan nilai P_1 = 2.0098 dan P_2 = 0.9618. Dengan mengambil nilai $\alpha = 0.05$, signal tiruan ini merupakan signal yang harmonik.

Catatan 2

Untuk menguji kesamaan karakteristik pada signal tiruan terhadap aslinya, gunakan statistik yang dipaparkan oleh Franke dan Halim [1].



Gambar 2. Signal original (atas) dan signal dari hasil Bootstrap (bawah)

KESIMPULAN

Pada makalah ini telah dipaparkan statistik untuk mendeteksi keberadaan periodisitas pada signal. Nilai periodisitas yang diperoleh pada langkah awal tadi akan digunakan untuk membangun blok pada signal. Blok inilah yang akan digunakan dalam membangun bootstrap untuk membentuk signal replikasi dari signal aslinya. Signal tiruan ini tetap memiliki karakteristik dari signal asli serta memiliki karakter keacakan.

DAFTAR PUSTAKA

[1] J. Franke, S. Halim: "Wild bootstrap tests for signals and images". IEEE Signal Processing Magazine 24, no. 4, 2007.

[2] S. Halim: Spatially adaptive detection of local disturbances in time series and stochastic processes on the integer lattice Z2. PhD Thesis. University of Kaiserlautern , 2005.

[3] S. T. Chiu. "Detecting periodic components in a white Gaussian time series". *Journal of Royal Statistics Society series B,* 51,2:249–259, 1989.

[4] M. Shimshoni. "On fisher's test of significance in harmonic analysis". *Geophys. J.R.Astronom. Society*, 23:373–377, 1971.

[5] E. Lewis and N. Fieller. "A recursive algorithm for null distributions for outliers: I. gamma samples". *Technometrics*, 21:371–376, 1979.

[6] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

[7] D. N. Politis. Resampling time series with seasonal components. Website, 2001. http://www.galaxy.gmu.edu/interface/I01/I2001Proceedings/DPolitis/DPolitis.pdf.