Uji Dipendensi Serial Pada Model Runtun Waktu Frekuensi Dengan Menggunakan *Simple Runs Test*

Herni Utami
Jurusan Matematika FMIPA UGM
herni_utami@ugm.ac.id

Intisari.

Di dalam analisis runtun waktu $\{X_t,\ t = 0,\ \pm 1,\ \pm 2,\ ...\}$ dengan X_t integer positif atau nol (runtunwaktu frekuensi), kebutuhan untuk menguji adanya dipendensi adalah suatu yang rutin dilakukan. Salah satu cara untuk uji tersebut adalah dengan uji nonparametrik yaitu *run test*.

Kata kunci: runtunwaktu frekuensi, run test, INARMA, INAR, INMA

1. Pendahuluan

Proses runtun waktu frekuensi $\{X_t,\ t = 0,\ \pm 1,\ \pm 2,\ ...\}$dengan X_t integer bernilai kecil, muncul di berbagai bidang statistika, diantaranya: runtun waktu banyaknya pelanggan yang menunggu dilayani di suatu konter yang dicatat dengan waktu diskret, banyaknya karyawan yang absen di suatu perusahaan, dan banyaknya kasus per bulan tentang suatu penyakit langka yang cepat menular. Nilai-nilai variabel random ke-*t* untuk kasus di atas bernilai positif atau nol dengan mean sampel barangkali kurang dari 10. Beberapa model terkait dengan kasus-kasus seperti di atas, telah dikembangkan dibeberapa literatur. Pada makalah ini, akan difokuskan pada model *integer-valued autoregressive-moving average* (INARMA)

2. Proses INAR(1)

Misalkan proses $\{X_t,\ t = 0,\ \pm 1,\ \pm 2,\ ...\}$ mengikuti model INAR(1), maka proses tersebut akan memenuhi persamaan:

Dipresentasikan dalam Seminar Nasional Matematika dan Pendidikan Matematika 2006 dengan tema " **Trend Penelitian dan Pembelajaran Matematika di Era ICT "** yang diselenggarakan pada tanggal 24 Nopember 2006

$$X_t = a \circ X_{t-1} + W_t \qquad t = 0, \pm 1, \pm 2, \ldots$$

Dengan state space proses adalah bilangan cacah, dan diasumsikan bahwa $a \in [0,1)$ dan W_t adalah barisan variabel random diskrit identik independen (iid) dengan mean μ_W dan variansi σ_W^2 yang masing-masing berhingga. Untuk sembarang *t*, variabel random W_t dan X_{t-1} independen.

Proses INAR(1) diasumsikan stasioner. Hal tersebut analog dengan proses AR(1) yang sudah umum dikenal, tetapi model INAR(1) adalah nonlinear jika dikaitkan dengan o-operator yang didefinisikan sebagai berikut:

$$a \circ X_{t-1} \equiv \sum_{i=1}^{X_{t-1}} Y_{i,t-1}$$

dimana $Y_{i,t-1}$ diasumsikan variabel random bernoulli identik independen, dengan $P(Y_{i,t-1} = 1) = a$ dan $P(Y_{i,t-1} = 0) = 1 - a$. Fungsi autokorelasi (ACF) proses INAR(1) adalah $\rho(k) = a^k$ untuk $k = 0,1,2,\ldots$ Hal ini identik dengan fungsi autokorelasi proses AR(1), hanya saja $\rho(k)$ untuk proses INAR(1) selalu positif sedanga $\rho(k)$ untuk proses AR(1) tidak selalu positif.

Secara umum, proses INARMA ditandai dengan adanya struktur dependensi dan sejauh ini tidak asumsi tentang distribusi marginal dari W_t. Al-Osh dan Alzaid (1987) mengasumsikan $W_t \sim \mathrm{Poi}(\lambda)$ dengan $\lambda > 0$ sehingga $X_t \sim \mathrm{Poi}(\lambda/(1-a))$, selanjutnya proses disebut PoINAR(1).

3. Proses INMA(1)

Type struktur dipendensi yang lain dinyatakan dengan *first-order integer-valued moving average* atau INMA(1). Model proses INMA(1) dari $\{X_t,\ t = 0,\ \pm 1,\ \pm 2,\ \ldots\}$ dinyatakan sebagai berikut:

$$X_t = b \circ W_{t-1} + W_t \qquad t = 0,\ \pm 1,\ \pm 2, \ldots$$

dimana $b \in [0,1]$ dan W_t adalah iid variabel random diskret. Mean dari W_t adalah μ_W dan variansi σ_W^2 yang masing-masing berhingga. o-operator didefinisikan sebagai berikut:

$$b \circ W_{t-1} = \sum_{i=1}^{W_{t-1}} Y_{i,t-1}$$

dimana $Y_{i,t-1}$ adalah iid variabel random bernoulli dengan $P(Y_{i,t-1} = 1) = b$. Sedang struktur dipendensi dari proses INMA(1) dinyatakan dengan ACF, yaitu:

$$\rho(k) = \begin{cases} \dfrac{b\sigma_W^2}{[b(1-b)\mu_W + (1+b^2)\sigma_W^2}, & k = 1 \\ 0 & k > 1 \end{cases}.$$

Dari ACF di atas diperoleh $0 \le \rho(1) \le 0{,}5$.

4. **Proses INAR(2)**

Struktur dipendensi proses $\{X_t,\ t = 0,\ \pm 1,\ \pm 2,\ ...\}$dengan order lebih tinggi dapat digambarkan dengan model INAR(2), yaitu

$$X_t = a_1 \circ X_{t-1} + a_2 \circ X_{t-2} + W_t \qquad t = 0, \pm 1, \pm 2, ...$$

o-operator analog dengan definisi sebelumnya. Untuk menjamin kestasioneran proses maka haruslah $a_1 + a_2 < 1$. Fungsi autokorelasi dari proses INAR(2) adalah

$$\rho(k) = \begin{cases} a_1, & k = 1 \\ a_1\rho(k-1) + a_2\rho(k-2), & k \ge 2 \end{cases}$$

5. **Uji Independensi Serial**

Dari uraian di atas, bisa dikatakan bahwa dalam proses INAR(1), INMA(1), dan INAR(2) terdapat adanya dependensi. Hal ini terlihat dari fungsi autokorelasi ketiga proses tersebut. Dengan demikian model-model INAR(1), INMA(1), dan INAR(2) bisa digunakan jika terdapat dependensi dalam proses $\{X_t,\ t = 0,\ \pm 1,\ \pm 2,\ \ldots\}$ sehingga uji independensi menjadi suatu hal yang penting untuk dilakukan dalam pemodelan runtun waktu frekuensi. Ada beberapa tes yang digunakan untuk menguji adanya dipendensi pada suatu runtun waktu frekuensi. Diantaranya adalah *simple runs test*..

Metode pertama yang dibahas untuk uji indipendensi adalah *simple runs test*. Pada metode ini data runtun waktu original didefinisikan sedemikian hingga ke dalam dua kategori. Misalkan mendefinisikan data yang lebih atau kurang dari nilai tengah sampel, sehingga setiap data akan masuk kesalah satu kategori lebih dari atau kurang dari nilai tengah sampel dan membuang data yang sama dengan nilai tengah yang digunakan. Umumnya nilai tengah yang digunakan adalah median. Dalam kasus dimana data runtun waktu mengenai frekuensi selalu bernilai kecil untuk setiap *t*, maka kemungkinan besar akan terdapat banyak data bernilai sama dengan median sampel sehingga banyak data yang akan dibuang. Akibatnya kekuatan uji akan melemah. Gibbons dan chakraborti (1992) menggunakan mean sampl dengan pertimbangan bahwa mean sample kemungkinan besar bukan integer sedang data runtun waktu yang ada integer, sehingga data yang sama dengan mean sedikit atau tidak ada.

Hipotesis null yang digunakan adalah tidak ada dipendensi serial versus hipotesis alternatif adalah ada dipendensi serial. Runs test didasarkan pada urutan pengambilan sampel. Setiap data observasi dinyatakan dalam dua kategori. Run didefinisikan sebagai suatu urutan terdiri dari satu atau lebih data berkategori sama. Misalkan setelah data dikategorikan diperoleh: L L K L L L K K K K L, maka dari data tersebut terdapat 5 run, yang pertama terdiri dari dua L, yang kedua satu K, yang ketiga tiga L, yang keempat tiga K, dan yang kelima satu L. Statistik uji yang digunakan adalah *u*: banyaknya run. Untuk *u* terlalu kecil, dicurigai adanya pengelompokan atau kemungkinan lain adanya tren. Tetapi jika u terlalu besar dicurigai adanya pola selang seling. Jadi kalau u terlalu besar atau terlalu kecil mengindikasikan adanya dipendensi.

Untuk menentukan daerah kritis, perlu dicari distribusi dari u. Untuk menentukan probabilitas u, dimisalkan n: banyaknya data masuk kategori 1 dan m: banyaknya data masuk kategori 2. Total susunan terbentuk ada $\binom{n+m}{n}$. Jika u genap, maka bisa dinyatakan u=2k dan k adalah integer positif. Pada kasus ini terdapat k run dari kategori 1 dan k run dari kategori 2. Banyaknya cara membentuk k run dari n data adalah $\binom{n-1}{k-1}$. Begitu juga banyaknya cara membentuk k run dari m data ketegori 2 adalah $\binom{m-1}{k-1}$. Jadi banyaknya cara membentuk 2k run dari n+m data yang terdiri dari n data kategori 1 dan m data kategori 2 adalah $2\binom{n-1}{k-1}\binom{m-1}{k-1}$. Dengan cara yang sama diperoleh banyaknya cara untuk membentuk 2k+1 run (u ganjil) adalah $\binom{n-1}{k}\binom{m-1}{k-1}+\binom{n-1}{k-1}\binom{m-1}{k}$. Sehingga diperoleh distribusi probabilitas u adalah;

$$f(u)=\begin{cases} 2\binom{n-1}{k-1}\binom{m-1}{k-1}, & u=2k \\ \binom{n-1}{k}\binom{m-1}{k-1}+\binom{n-1}{k-1}\binom{m-1}{k}, & u=2k+1 \end{cases}.$$

Apabila n dan m besar, maka $Z=\frac{u-\mu_u}{sd(u)}\xrightarrow{d}N(0,1)$ dengan $\mu_u=\frac{2nm}{n+m}+1$ dan $sd(u)=\sqrt{\frac{2nm(2nm-n-m)}{(n+m)^2(n+m-1)}}$. Hipotesis null ditolak jika $Z\le -Z_{\alpha/2}$ atau $Z\ge Z_{\alpha/2}$.

Daftar Pustaka

Brännäs, K dan Quoreshi, S, (2004), *Integer-Valued Moving Average Modelling of the Number of Transactions in Stocks*, Department of Econometrics & USBE, Umeå University, Sweden

Freend's, J, (2002), *Mathematical Statistics*, edisi ke-enam, Irwin Miller and Marylees Miller.

Jung, R dan Tremayne, A.R., (2003), *Testing for Serial Dependence in Time Series Models of Counts*, Journal of Time Series Analysis, vol. 24, No.1, p65-84