PENGUJIAN MODEL DISTRIBUSI

Oleh Sity Rachyany *)

RINGKASAN

Dalam melakukan analisa data dari suatu populasi, sering diasumsikan berdistribusi tertentu. Seperti yang akan dibicarakan disini adalah distribusi Poisson.

Untuk mengetahui apakah model distribusi yang diasumsikan betul-betul dapat dipenuhi, maka perlu diuji kebenarannya.

Pengujian ini dapat dilakukan dengan uji kecocokan chi-kuadrat yang bergantung kepada banyaknya parameter yang akan ditaksir.

1. PENDAHULUAN

Dalam melakukan analisa data mengenai suatu populasi, sering diasumsikan populasinya berbentuk distribusi tertentu. Jika asumsi tidak dipenuhi, maka semua kesimpulan yang dibuat berdasarkan asumsi ini tidak benar. Dari pernyataan ini timbul pertanyaan; bagaimana caranya untuk dapat menguji kebenaran asumsi tersebut, agar kesimpulan yang dibuat dapat dipercaya.

Berdasarkan keterangan diatas, maka dalam makalah ini akan dibahas mengenai pengujian suatu distribusi dengan menggunakan tes chi-kuadrat yang dinamakan dengan tes kesesuaian/tes kecocokan distribusi.

Untuk keperluan tes kecocokan ini, diperlukan frekuensi diharapkan atau frekuensi yang berdasarkan teoritis dan frekuensi observasi yang diperoleh dari hasil pengamatan.

*) Staf Kelompok Penelitian Ionosfer.

2. FREKUENSI OBSERVASI DAN FREKUENSI DIHARAPKAN

Seperti yang kerap kali kita temui bahwa hasil yang diperoleh dalam sampel tidak selalu tepat benar dengan hasil yang diharapkan secara teoritis sesuai dengan aturan peluang (probabilitas). Sebagai contoh meskipun secara teoritis, kita mengharapkan 50 muka G (Gambar) dan 50 muka H (Huruf) dalam pelemparan uang sebanyak 100 kali. Tetapi hasil ini jarang sekali diperoleh secara tepat.

Misalkan suatu eksperimen menghasilkan k buah peristiwa-peristiwa atau katagori-katagori E_1, E_2, ... ,E_k yang saling bebas; dengan masing-masing peluang adalah $p_1 = P(E_1)$, $p_2 = P(E_2)$, ... $p_k = P(E_k)$.

Dengan harga peluang p_i (i = 1, 2, ... ,k) yang diberikan, maka kita dapat menghitung masing-masing frekuensi yang diharapkan $e_1 = np_1$, $e_2 = np_2$, ... , $e_k = np_k$.

Untuk lebih mudahnya dapat dilihat dalam tabel (2.1) berikut ini.

TABEL 2.1

Katagori	E_1	E_2	......	E_k
Frekuensi observasi	O_1	O_2	......	O_k
Frekuensi diharapkan	e_1	e_2	....	e_k

dengan :

O_1, O_2, ... , O_k merupakan frekuensi-frekuensi yang terjadi berdasarkan hasil pengamatan, sedangkan frekuensi yang diperoleh berdasarkan teori yang berlaku untuk eksperimen itu; kita umumkan e_1, e_2, ... , e_k.

3. DEFINISI X^2

Suatu ukuran mengenai adanya perbedaan antara frekuensi observasi dan frekuensi yang diharapkan dapat diperlihatkan dengan statistik χ^2 (chi-kuadrat) yang ditentukan oleh :

$$\chi^2 = \frac{(O_1 - e_1)^2}{e_1} + \frac{(O_2 - e_2)^2}{e_2} + \ldots\ldots + \frac{O_k - e_k)^2}{e_k}$$

$$= \sum_{j=1}^{k} \frac{(O_j - e_j)^2}{e_j} \qquad (3\text{-}1)$$

Dengan jumlah frekuensi adalah N, maka :

$$\Sigma\, O_j = \Sigma\, e_j = N \qquad (3\text{-}2)$$

sehingga persamaan (3-1) dapat juga dituliskan sebagai :

$$\chi^2 = \sum_{j=1}^{k} \frac{O_j^{\,2}}{e_j} - N \qquad (3\text{-}3)$$

Nilai chi-kuadrat akan sama dengan nol jika $O_j = e_j$, yang berarti bahwa tidak ada perbedaan antara frekuensi observasi dengan frekuensi yang diharapkan. Makin besar perbedaan antara O_j & e_j makin besar pula nilai chi-kuadratnya.

Ternyata statistik χ^2 ini mendekati distribusi kuadrat dengan derajat kebebasan :

a. $\nu = k-1$; jika menghitung frekuensi diharapkan e_j tanpa menggunakan parameter yang ditaksir dari statistik sampel.

b. $\nu = k-m-1$; jika untuk memperoleh e_j dengan menggunakan m buah parameter populasi yang ditaksir berdasarkan statistik sampel.

4. UJI KECOCOKAN

Untuk melakukan uji kecocokan ini, akan dibandingkan antara data hasil pengamatan dengan frekuensi yang diharapkan berdasarkan model yang dihipotesakan; dan untuk ini, digunakan rumus (3-1). Nilai-nilai parameter populasi yang diharapkan atau teoritis, ditaksir berdasarkan nilai-nilai statistik sampel yang tak bias.

Misalkan untuk distribusi normal akan ditaksir rata-rata μ oleh $\overline{x}$ dan variasi σ^2 oleh S^2 dengan derajat kebebasan k-m-1;
di mana k adalah banyaknya katagori atau kelas interval dan
m adalah banyaknya parameter yang ditaksir.
Karena itu, maka derajat kebebasan untuk distribusi normal adalah (k-3).

Lebih lanjut, kita coba uji kecocokan ini terhadap data yang dihipotesakan mengikuti distribusi Poisson.

4.1 . Distribusi Poisson

Variabel acak x dikatakan mempunyai distribusi Poisson, jika fungsi peluang x berbentuk :

$$p(x) = \frac{e^{-\lambda}\, \lambda^{x}}{x!} \tag{4-1}$$

dengan x = 0, 1, 2, 3, ...

e adalah bilangan konstan ; e = 2,7183 dan

paramter λ adalah positip ; $\lambda > 0$

Sifat-sifat distribusi Poisson.

1. $E(x) = \lambda$

   Bukti :

   $$\text{Rata-rata } \mu = E(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\, \lambda^{x}}{x!}$$

   $$= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\, \lambda^{x}}{(x-1)!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!}$$

   $$= e^{-\lambda}\, \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda}\, \lambda\, e^{\lambda}$$

   $$= \lambda$$

   $$\therefore\ \mu = E(x) = \lambda$$

2. Varians untuk x = λ

   Bukti : Varians (x) = $E(x)^2 - \{E(x)\}^2$

   $$E\{x(x-1)\} = \sum_{x=1}^{\infty} x(x-1) \frac{e^{-\lambda}\, \lambda^{x}}{x!}$$

   $$= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\, \lambda^{x}}{(x-2)!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-2)!}$$

   $$= e^{-\lambda}\, \lambda^{2} \sum_{x=1}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}$$

   $$= e^{-\lambda}\, \lambda^{2}\, e^{-\lambda} = \lambda^{2}$$

   $E(x)^2 = \lambda^2 + \lambda$

   Sehingga varians (x) $= E(x^2) - \{E(x)\}^2$

   $$= \lambda^2 + \lambda - \lambda^2$$

   $$\therefore\ \text{Varians } (x) = \lambda$$

Misalkan x_1, x_2, ..., x_k merupakan sampel acak dari sebuah populasi. Selanjutnya akan diuji apakah populasi tersebut mengikuti distribusi Poisson atau tidak.

Untuk ini prosedur yang akan digunakan adalah :

- Rumuskan hipotesisnya dalam bentuk

  H_o : $F_x(\cdot)$ berdistribusi Poisson

  H_1 : $F_x(\cdot)$ tidak berdistribusi Poisson

- Susun deretan data menjadi k buah katagori/peristiwa yaitu :

  E_1, E_2, ... , E_k yang saling bebas

- Hitung n_i yang menyatakan frekuensi data sampel dalam katagori

  E_j ; $j = 1, 2, \ldots\ldots, k$

- Nyatakan peluang $P(E_j)$ oleh p_j

- Hitung statistik

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - e_j)^2}{e_j}$$

Dengan menggunakan distribusi chi-kuadrat yang bersifat pendekatan ini, maka kriteria pengujian adalah; Tolak hipotesa H_o jika χ^2 berdasarkan perhitungan > χ_t^2 dengan χ^2 tabel diperoleh dari daftar chi-kuadrat untuk taraf α yang dipilih, sedangkan dalam hal lain H_o diterima.

Untuk menjelaskan prosedur diatas, tentang pengujian suatu distribusi; akan saya sajikan hasil percobaan yang dilakukan oleh Rutherford dan Greiser [1].

Sejumlah partikel-partikel α yang dipancarkan oleh suatu zat radio aktif dalam $n = 2608$ perioda, yang masing-masing 7,5 detik mengikuti distribusi Poisson dengan parameter λ sebagai rata-ratanya.

Dalam tabel (4.1) diperlihatkan n_j menyatakan sejumlah perioda di mana sejumlah partikel-partikel yang dipancarkan setara dengan E_j.

Tabel 4.1 :

E_j	0	1	2	3	4	5	6	7	8	9	10
n_j	57	203	383	525	532	408	273	139	45	27	16

Rata-rata sejumlah partikel-partikel λ yang dipancarkan selama suatu perioda 7,5 detik adalah :

$$\lambda = \frac{57(0) + 203(1) + \ldots\ldots + 27(9) + 16(10)}{2608}$$

$$= \frac{10086}{2608} = 3{,}87$$

Untuk menghitung statistik χ^2, diperlukan tabel (4.2) berikut :

TABEL 4.2

E_j	$n_j = O_j$	$n\ p_j = e_j$	$\frac{(O_j - e_j)^2}{e_j}$
0	57	54,399	0,1244
1	203	210,523	0,2688
2	383	407,361	1,4568
3	525	525,496	$4{,}68.10^{-4}$
4	532	508,418	1,0938
5	408	393,515	0,5332
6	273	253,817	1,4498
7	139	140,325	0,0125
8	45	67,882	7,7132
9	27	29,189	0,1642
10	16	17,075	0,0677
-	2608	2608,00	12,8849

Dari daftar chi-kuadrat dengan derajat kebebasan sama dengan 9 dengan α = 0,05 ; maka diperoleh $\chi^2_{0,95(9)}$ sama dengan 16,9. Berdasarkan kriteria pengujian hipotesa, ternyata bahwa hipotesa model distribusi Poisson dapat diterima.

Contoh lain saya ambil data absorpsi dari ionosonde variabel. Pameungpeuk. Munculnya signal per noise ratio dalam setiap 2 jam, mulai jam 06.00 sampai dengan jam 20.00 untuk bulan Agustus 1983. Frekuensi yang diambil adalah frekuensi 2,37 MHz. Datanya seperti dalam Tabel (4.3) berikut ini :

Tabel 4.3 ....

TABEL 4.3

E_j	0	1	2	3	4	5	6	7	8
n_j	166	23	12	21	18	17	10	4	4

Rata-rata munculnya signal per noise ratio yang diamati selama 1 bulan (Agustus'83) adalah $\lambda = 1,41$. Dengan mensubstitusikan harga λ ke dalam rumus (4-1) dan rumus (3-1), maka diperoleh Tabel (4.4) berikut :

TABEL 4.4

E_j	$n_j = O_j$	$n\ p_j = e_j$	$\dfrac{(O_j - e_j)^2}{e_j}$
0	166	67,139	145,571
1	23	94.611	54,202
2	12	66,701	44,860
3	21	31,350	3,417
4	18	11,051	4,370
5	17	3,116	61,863
6	10	0,732	117,344
7	4	0,148	100,256
8	4	0,026	607,411
-	275	274,874	1139,294

Dengan mengambil $\alpha = 0,05$, ternyata model yang dihipotesakan berdistribusi Poisson ditolak.

5. PENUTUP

Dari contoh terakhir dinyatakan bahwa model distribusi Poisson ditolak. Ini berarti bahwa munculnya signal per noise ratio tidak berdistribusi Poisson. Dengan perkataan lain, data tersebut harus diuji kembali dengan model distribusi yang lain.

DAFTAR PUSTAKA

1. FISZ, MAREK : "Probability Theory and Mathematical Statistics", Third Edition, John Wiley & Sons, Inc New York, 1963.
2. SUDJANA : "Metoda Statistika", Penerbit Tarsito, Bandung, 1975.
3. SPIEGEL, M.R. : "Theory and Problems of Statistics", Schaum Publishing Co. New York, 1961.

- - - o0o - - -

D I S K U S I

1. WILSON SINAMBELA

Tanya : Dalam ringkasan model distribusi adalah "Distribusi Poisson" adalah distribusi diskrit. Kebenaran diuji dengan chi-kuadrat yang fungsinya kontinu
- Apakah fungsi diskrit dapat diuji dengan fungsi kontinu ?
- Kalau dapat, sampai di mana tingkat kebenarannya ?

Jawab : Uji kecocokan ini berlaku umum baik fungsi diskrit maupun kontinu.

2. MASPUL AINI

Tanya : Distribusi Poisson $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ dengan $x = 0, 1, 2, \ldots\ldots$ sudah dikatakan bahwa λ positip, apakah boleh λ negatip atau nol ($\lambda = 0$) ?

Jawab : Seperti yang sudah disebutkan bahwa λ positip; ini berarti bahwa λ negatip tidak mungkin sedangkan $\lambda = 0$ sudah jelas boleh.

3. M. PARDEDE

Tanya : Apakah uji kecocokan hanya dengan chi-kuadrat atau ada metoda lain ?

Jawab : Tidak hanya chi-kuadrat, ada metoda lain yaitu metoda non parametrik.

4. TATANG T. SOELAEMAN

Tanya : Untuk data apa saja, metoda ini telah dipakai di Pusrigan ?

Jawab : Metoda ini belum pernah dipakai di Pusrigan.

5. SURATNO

Tanya : Mohon penjelasan tentang banyaknya parameter yang akan ditaksir !

Jawab : Misalkan kita ambil distribusi normal dengan fungsi :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Kita tahu bahwa rata-rata μ dan variansi σ^2 tidak diketahui tapi harus dicari dengan jalan menaksir μ oleh $\bar{x}$ dan σ^2 oleh s^2 . μ & σ^2 diumpamakan parameter. Jadi dalam distribusi normal ada 2 parameter sedangkan untuk distribusi Poisson, parameternya hanya satu yaitu λ .

Tanya : Apakah pengujian ini hanya berlaku untuk distribusi Poisson ?

Jawab : Tidak, distribusi Poisson hanya sebagai contoh saja. Pengujian ini berlaku untuk umum.

6. S.L. MANURUNG

Tanya : Persamaan $-X^2 = \sum_{j=1}^{k} \frac{O_j^{\,2}}{e_j} - N$

$-X^2 = 0$, jika $O_j = e_j$; mohon dijelaskan !

Jawab : Berdasarkan persamaan (3-2), kalau persamaan (3-1) diuraikan akan didapat persamaan (3-3)

$$X^2 = \sum_{j=1}^{k} \frac{(O_j - e_j)^2}{e_j} = \Sigma \frac{(O_j^{\,2} - 2\,O_j e_j + e_j^{\,2})^2}{e_j}$$

$$= \Sigma \frac{O_j^{\,2}}{e_j} - 2\ \Sigma O_j + \Sigma e_j = \Sigma \frac{O_j^{\,2}}{e_j} - 2N + N$$

$$= \frac{O_j^{\,2}}{e_j} - N \quad \text{(terbukti)}$$

Telah dijelaskan bahwa X^2 merupakan suatu ukuran, mengenai perbedaan antara frekuensi observasi dan frekuensi yang diharapkan. Jadi sudah jelas jika $O_j = e_j$, maka $X^2 = 0$, karena antara frekuensi hasil pengamatan O_j dengan frekuensi yang diharapkan adalah sama.

Tanya : Mohon diberikan contoh sehingga pengertian akan dk(ν) dan penggunaan teori ini !

Jawab : Dapat dilihat dalam contoh.

7. BINSAR GULTOM

Tanya : Pemilihan distribusi Poisson alasannya apa, karena di bidang Atmosfer atas jarang dipakai. Apakah tidak lebih baik distribusi Weibull yang diuji ?

Jawab : Distribusi Poisson hanya sebagai contoh. Tetapi jika kita ingin menguji distribusi Weibull juga bisa.

Tanya : Tadi ada d.k = k-m-1 . Dari mana asalnya ?

Jawab : Ini sudah ketentuan bahwa chi-kuadrat mempunyai d.k = k-m-1; di mana k adalah banyaknya kejadian atau katagori sedangkan m adalah banyaknya parameter yang akan ditaksir.