

KEUNGGULAN MODEL ι DALAM ANALISIS DATA DNA FCM

MOHD. KIDIN SHAHRAN

Jabatan Statistik
Fakulti Sains Matematik dan Komputer
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia.

Abstrak. Dalam kertas ini Model ι yang telah dibina oleh Mohd. Kidin [3, 4] dibandingkan dengan model seragam, dan model negatif eksponen. Perbandingan dilakukan melalui penganggaran silang data simulasi sebelum dilakukan ke atas data sebenar.

Abstract. In this paper Model ι which was developed by Mohd. Kidin [3, 4] is compared with other two models ie the uniform and negative exponential model. Comparison is made via cross estimation of simulated data before the models being applied to real data analysis.

1. PENGENALAN

Dalam pembinaan model kepada fenomena nyata tiga pendekatan mungkin digunakan, iaitu (a) pendekatan konsep (conceptual approach); (b) pendekatan empirik (emphirical approach); dan (c) pendekatan elektik (electical approach) (lihat Warren, [6]). Pendekatan pertama biasanya berasaskan teori; pendekatan kedua berasaskan kepada analisis data; dan pendekatan ketiga adalah

campuran atau gabungan terbaik antara dua pendekatan pertama. Setelah sesuatu dibina serta dikenalpasti kesesuaiannya dengan data, maka langkah-langkah seterusnya proses anggaran parameter anu yang mencirikan taburan data; diikuti dengan proses pengesahan; dan seterusnya aplikasi kepada data nyata.

Model ι telah dibina bagi taburan amaun DNA bagi sel pada fasa S dalam kitaran sel dan dikenal pasti oleh Mohd. Kidin [3]. Model adalah seperti berikut:

$$g(y) = \frac{K}{[y - A(1 - \alpha)]^{1+d} [2A(1 + \beta) - y]^{1-d}} \quad (1)$$

yang

A : amaun DNA bagi sel dalam fasa G1;

$2A$: amaun DNA bagi sel dalam fasa G2;

$y, \quad A \leq y \leq 2A$: amaun DNA bagi sel dalam fasa S;

K : pemalar lenoarmalan;

d, α, β : parameter bentuk, $0 < d, \alpha, \beta < 1$.

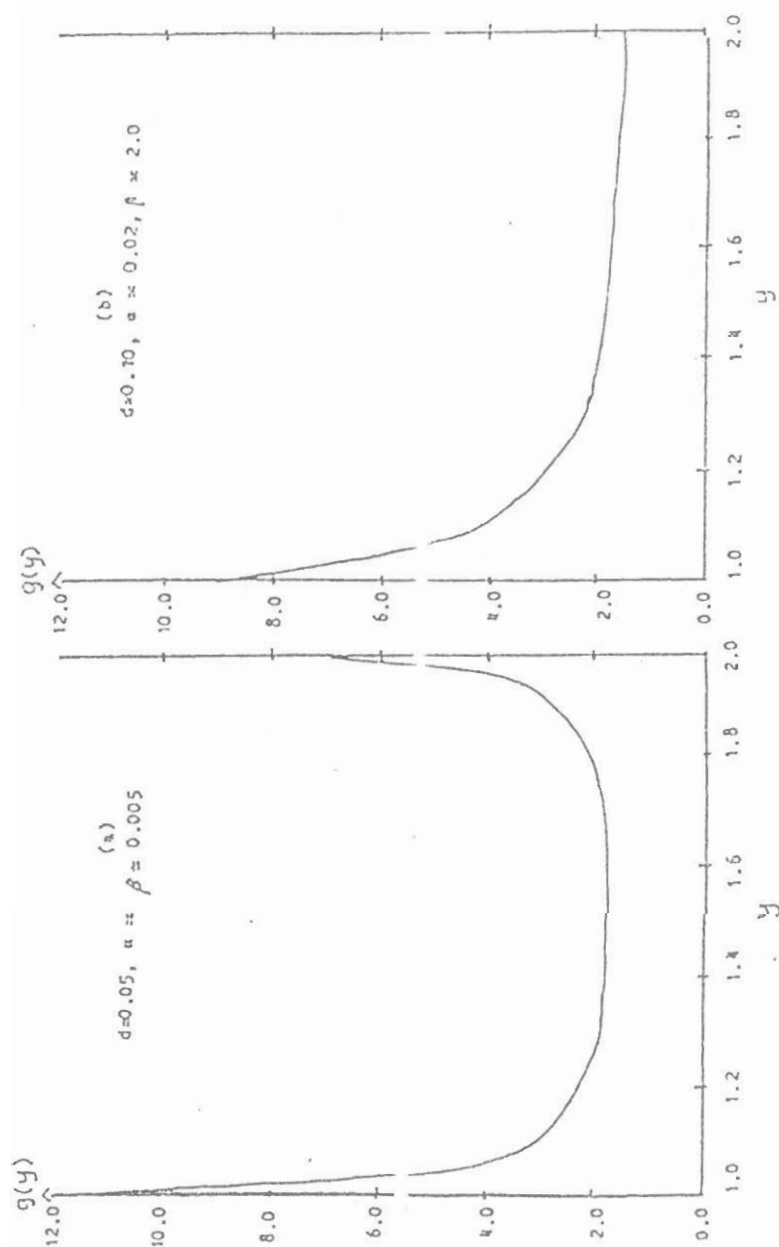
Dalam analisis, supaya model mempunyai bilangan parameter yang minimum, diandaikan $\alpha = \beta$, dan d diberikan nilai 0.05 (lihat Mohd. Kidin, [5])

Pemilihan fungsi seragam dan negatif eksponen bukan sahaja kedua-duanya telah digunakan dalam analisis oleh penyelidik lain tetapi mereka boleh dimimikkan oleh model ι . Di sini berlaku dua keistimewaan iaitu

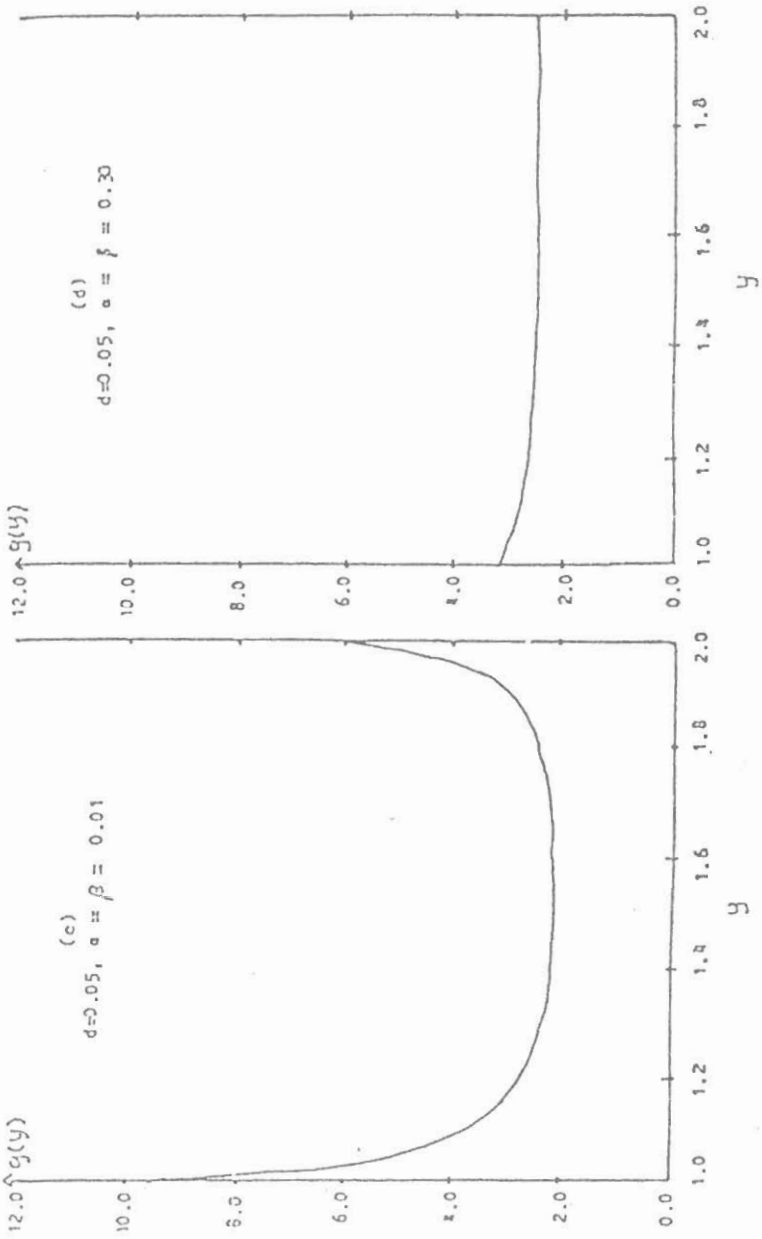
- (1) Pemilihan di atas bukan secara *ad hoc*;
- (2) kedua-dua fungsi itu merupakan sebahagian daripada model ι bagi nilai-nilai α dan β tertentu.

2. KEUNGGULAN MODEL ι

Model ι sebagai fungsi pendasar fasa S mempunyai beberapa bentuk tipikal seperti dalam Rajah 1. Gabungan parameter bentuk α dan β boleh mewujudkan kepelbagaian bentuk model dan



RAJAH 1. Kepelbagaian bentuk tipikal fungsi pendaras $g(y)$.



RAJAH 1(SAMBUNGAN). Kepelbagaian bentuk tipikal fungsi pendasar $g(y)$.

seterusnya menunjukkan kemampuan model memimik model-model lain yang telah digunakan oleh penyelidik terdahulu dalam analisis data yang sama. Umpamanya panel (b) dalam rajah itu adalah milik model negatif eksponen iaitu model yang telah digunakan oleh Baisch dan Gohde [2]; panel (d) pula adalah mimik model seragam iaitu model yang telah digunakan oleh Baisch drk [1]. Dengan demikian perbandingan analisis akan dibuat antara tiga model berikut:

Model I (seragam) : $g(y) = 1/A$;

Model II (negatif eksponen) : $g(y) = \lambda \exp(-\lambda y)$;

Model III (model ι) : $g(y)$, persamaan (1).

Oleh sebab adanya perbezaan bilangan parameter anu dalam model-model tersebut, maka perbandingan analisis tertumpu kepada anggaran x_1 (keamatan pendarflor bagi DNA dalam sel pada fasa G1), anggaran x_2 (keamatan pendarflor bagi DNA dalam sel pada fasa G2+M), ρ_1 (perkadaran sel dalam fasa G1), ρ_3 (perkadaran sel dalam fasa S), dan γ (pekali variasi, parameter taburan sukatan). Sementara ρ_2 (perkadaran sel dalam fasa G2+M) diperolehi dengan penolakan ρ_3 daripada 1.

3. PENGANGGARAN SILANG

Penganggaran ini bermaksud menyesuaikan satu model kepada data yang disimulasikan daripada model lain. Tujuan penganggaran silang ini ialah untuk menunjukkan betapa pentingnya mempunyai model yang betul, sesuai dan tepat untuk menganalisis set data yang diberi. Hal ini semakin bertambah penting apabila set data yang diberi itu dianggap daripada taburan campuran seperti dalam kes yang sedang dibincangkan. Oleh sebab model III atau Model ι dibina berasaskan fenomena fizik replikasi DNA maka ia dianggap lebih jitu kesesuaiannya dibandingkan dengan Model I dan Model II. Dengan demikian data simulasi daripada

Jadual 1

Anggaran Silang Data Simulasi Kumpulan 1;
 Input: $x_1 = 40.0$, $x_2 = 80.0$, $\alpha = 0.05$,
 $\rho_1 = 0.55$, $\rho_3 = 0.35$,
 $\gamma = 0.025$, 0.075 , 0.10

Anggaran; (sisihan ralat); dan [sisihan relatif,%]

$\hat{x}_1$			$\hat{\rho}_1$			$\hat{\rho}_3$		
<u>Model</u>			<u>Model</u>			<u>Model</u>		
III	II	I	III	II	I	III	II	I
$\gamma = 0.025$			$\gamma = 0.025$			$\gamma = 0.025$		
40.00	40.09	40.11	0.549	0.592	0.601	0.347	0.274	0.276
(0.04)	(0.01)	(0.01)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)
[0.0]	[0.2]	[0.3]	[-0.2]	[7.6]	[9.3]	[-0.8]	[-21.7]	[-21.1]
$\gamma = 0.075$			$\gamma = 0.075$			$\gamma = 0.075$		
40.00	40.25	40.34	0.530	0.605	0.622	0.377	0.246	0.249
(0.07)	(0.05)	(0.04)	(0.023)	(0.007)	(0.006)	(0.04)	(0.007)	(0.008)
[0.0]	[0.6]	[0.8]	[-3.6]	[10.0]	[13.1]	[7.7]	[-29.7]	[-21.9]
$\gamma = 0.10$			$\gamma = 0.10$			$\gamma = 0.10$		
40.00	40.20	40.29	0.518	0.603	0.617	0.397	0.253	0.255
(0.13)	(0.08)	(0.06)	(0.048)	(0.010)	(0.006)	(0.008)	(0.010)	(0.01)
[0.0]	[0.5]	[0.7]	[-5.8]	[9.6]	[12.2]	[13.4]	[-27.7]	[-27.1]

Jadual 2

Anggaran Silang Data Simulasi Kumpulan 1;		
Input: $x_1 = 40.0$, $x_2 = 80.0$, $\alpha = 0.05$,		
$\rho_1 = 0.55$, $\rho_3 = 0.35$,		
$\gamma = 0.025$, 0.075 , 0.10		

Anggaran; (sisihan ralat); dan [sisihan relatif, %]

$\hat{x}_1$			$\hat{\rho}_1$			$\hat{\rho}_3$		
<u>Model</u>			<u>Model</u>			<u>Model</u>		
III	II	I	III	II	I	III	II	I
$\gamma = 0.05$			$\gamma = 0.05$			$\gamma = 0.05$		
39.96	40.08	40.10	0.427	0.484	0.493	0.387	0.290	0.292
(0.02)	(0.02)	(0.02)	(0.011)	(0.005)	(0.005)	(0.018)	(0.005)	(0.005)
[-0.1]	[0.2]	[0.2]	[-5.1]	[7.5]	[9.5]	[10.6]	[-17.1]	[-16.6]
$\gamma = 0.075$			$\gamma = 0.075$			$\gamma = 0.075$		
39.86	40.12	40.21	0.426	0.509	0.526	0.386	0.238	0.243
(0.09)	(0.06)	(0.05)	(0.035)	(0.008)	(0.006)	(0.058)	(0.008)	(0.008)
[-0.3]	[0.3]	[0.5]	[-5.3]	[13.1]	[16.9]	[10.3]	[-32.0]	[-30.6]
$\gamma = 0.10$			$\gamma = 0.10$			$\gamma = 0.10$		
39.97	40.29	40.38	0.427	0.509	0.524	0.371	0.239	0.245
(0.10)	(0.08)	(0.07)	(0.020)	(0.009)	(0.006)	(0.033)	(0.01)	(0.01)
[-0.1]	[0.7]	[0.9]	[-5.1]	[13.1]	[16.4]	[6.0]	[-31.7]	[-30.0]

Jadual 3

Anggaran Silang Data Simulasi Kumpulan 1;		
Input: $x_1 = 40.0$, $x_2 = 80.0$, $\alpha = 0.05$,		
$\rho_1 = 0.55$, $\rho_3 = 0.35$,		
$\gamma = 0.025$, 0.075 , 0.10		

Anggaran; (sisihan ralat); dan [sisihan relatif, %]

$\hat{x}_1$			$\hat{\rho}_1$			$\hat{\rho}_3$		
<u>Model</u>			<u>Model</u>			<u>Model</u>		
III	II	I	III	II	I	III	II	I
$\gamma = 0.025$			$\gamma = 0.025$			$\gamma = 0.025$		
40.00	40.18	40.21	0.341	0.412	0.421	0.466	0.345	0.348
(0.04)	(0.02)	(0.02)	(0.021)	(0.005)	(0.005)	(0.036)	(0.006)	(0.006)
[0.0]	[0.4]	[0.5]	[-2.6]	[17.7]	[20.3]	[3.5]	[-23.3]	[-22.7]
$\gamma = 0.075$			$\gamma = 0.075$			$\gamma = 0.075$		
39.90	40.30	40.43	0.318	0.418	0.439	0.497	0.319	0.325
(0.09)	(0.07)	(0.06)	(0.012)	(0.008)	(0.006)	(0.018)	(0.009)	(0.009)
[-0.2]	[0.7]	[1.1]	[-9.1]	[19.4]	[25.4]	[10.4]	[-29.1]	[-27.8]
$\gamma = 0.10$			$\gamma = 0.10$			$\gamma = 0.10$		
39.95	40.29	40.49	0.333	0.418	0.446	0.468	0.314	0.319
(0.17)	(0.11)	(0.08)	(0.038)	(0.012)	(0.007)	(0.064)	(0.012)	(0.012)
[-0.1]	[0.7]	[1.2]	[-4.8]	[19.4]	[27.4]	[4.1]	[-30.2]	[-29.1]

Model III boleh dikatakan lebih mirip kepada data sebenar dalam kontek populasi sel bersinkroni. Sementara Model I dan Model II digunakan untuk menganggar parameter anu. Dalam penganggaran ini hanya tiga parameter terpenting ditimbangkan iaitu

- x_1 : Keamatan pendarflor bagi A unit amaun DNA bagi sel dalam fasa G1, ia dikenal juga sebagai min fasa G1;
- ρ_1 : perkadaran sel dalam fasa G1;
- ρ_3 : perkadaran sel dalam fasa S.

Dengan α dan d ditetapkan pada nilai 0.05, dan γ pada nilai-nilai 0.025, 0.075 dan 0.10; maka sebanyak tiga kumpulan data telah disimulasi dengan saiz sampel 10000.

Data Kumpulan 1 : $\rho_1 = 0.55, \rho_3 = 0.35$;

Data Kumpulan 1 : $\rho_1 = 0.45, \rho_3 = 0.35$;

Data Kumpulan 1 : $\rho_1 = 0.35, \rho_3 = 0.45$;

Dengan menggunakan data simulasi ada kebaikannya kerana hasil anggaran boleh dibandingkan dengan nilai input parameter. Dalam hal ini kejituan anggaran diukur dengan menggunakan sishan relatif, SR, dalam peratusan yang diberikan oleh

$$SR = \frac{\text{anggaran} - \text{input}}{\text{input}} \times 100\%.$$

Ukuran ini bermaksud, nilai mutlak SR yang besar menunjukkan anggaran diperolehi daripada jauh tersisih daripada sasaran input.

Jika ditinjau hasil anggaran silang bagi ketiga-tiga kumpulan data, didapati

- (1) $|SR|$ bagi anggaran hampir kesemua parameter dengan menggunakan Model III sentiasa lebih kecil daripada $|SR|$ diperolehi dengan menggunakan model I dan II.

- (2) Bagi kebanyakan kes, $|SR|$ bagi Model I dan II biasanya dalam pangkat yang sama.
- (3) Sebahagian besar anggaran daripada Model III mempunyai $|SR| < 10$, khususnya bagi parameter $\hat{\rho}_1$ dan $\hat{\rho}_3$.
- (4) Bagaimana pun untuk parameter x_1 , ketiga-tiga model menghasilkan anggaran yang setanding, dengan $|SR| < 1$.
- (5) (Samb) Melainkan bagi kumpulan 3, Model I menghasilkan anggaran dengan $|SR| > 1$ tetapi masih kurang daripada 2%.

Jadual 4
 Penyuaian Model I, II, III
 Ke atas Set data Nyata 1

Parameter	Anggaran Watson	Anggaran Mohd. Kidin		
		I	II	III
x_1	63.62	63.47 (0.13)	63.39 (0.14)	63.38 (0.13)
ρ_1	0.477	0.499 (0.009)	0.490 (0.011)	0.488 (0.009)
ρ_3	0.336	0.327 (0.013)	0.323 (0.013)	0.328 (0.013)
γ	0.074	0.0803 (0.001)	0.0801 (0.001)	0.080 (0.001)
λ	—	—	0.006 (0.004)	—
α	—	—	—	0.971 (0.186)

4. APLIKASI KE ATAS DATA SEBENAR

Sebagai lanjutan kepada proses penganggaran silang, perbandingan ketiga-tiga model boleh dilaksanakan melalui aplikasi kepada data sebenar. Dalam hal ini dua set data sebenar telah digunakan. Dalam set data 1, perkadaran sel dalam fasa G1 kelihatan lebih besar daripada perkadaran sel dalam fasa S. Sementara dalam set data nyata 2 pula, kes sebaliknya berlaku. Setiap set data disertakan dengan anggaran parameter yang telah diperoleh oleh Watson [7]. Anggaran ini juga boleh dijadikan kayu pengukur dalam perbandingan model-model.

Hasil analisis diberikan dalam jadual 4 dan 5. Angka dalam kurungan adalah ralat anggaran.

Jadual 5

Penyuaian Model I, II, III
Ke atas Set data Nyata 2

Parameter	Anggaran Watson	Anggaran Mohd. Kidin		
		I	II	III
x_1	64.02	64.37 (0.12)	64.10 (0.14)	63.76 (0.24)
ρ_1	0.346	0.392 (0.008)	0.490 (0.010)	0.329 (0.025)
ρ_3	0.469	0.422 (0.012)	0.367 (0.010)	0.490 (0.038)
γ	0.063	0.065 (0.001)	0.418 (0.012)	0.0618 (0.001)
λ	—	—	0.0642 (0.001)	—
α	—	—	—	0.176 (0.09)

5. KESIMPULAN

Merujuk kepada hasil analisis seperti dalam Jadual 4, nampaknya ketiga-tiga model memberikan anggaran yang seimbang bagi ke semua parameter penting. Hal ini memang dijangka bagi kes data di mana ρ_1 tidak terlalu besar jika dibandingkan dengan ρ_3 , iaitu kes biasa bagi populasi sel berasinkroni. Kes ini adalah asas kepada pembinaan model yang dibincangkan.

Bagi kes data set kedua di mana $\rho_1 < \rho_3$, wujud variasi dalam anggaran yang diperolehi. Sebagai fungsi seragam, Model I sewajarnya menghasilkan anggaran yang lebih untuk ρ_1 dan ρ_2 ($1 - \rho_3$) sedangkan Model II dengan wajar terlalu kecil memberi yang lebih kepada ρ_1 berbanding kepada ρ_3 . Untuk Model III pula, dengan α yang besar akan memberikan anggaran lebih untuk ρ_2 ($1 - \rho_3$) dan kurang untuk ρ_3 . Walau bagaimanapun ketiga-tiga model memberikan anggaran yang seimbang bagi perkadaran sel dalam fasa S iaitu ρ_3 .

Dalam perbandingan ini, Model III masih lagi lebih baik dalam konteks ia dibina berasaskan kepada fenomena biofizik replikasi DNA (Mohd. Kidin, [3]).

RUJUKAN

- [1]. Baisch H., Gohde W. & Linden W.A., *Analisis of PCP data to determine the fraction of cells in the various phases of cell-cycle*, Radiation and environmental Biophysics **12** (1975), 31-39.
- [2]. Baisch H. & Gohde W., *Second International Symposium Pulse Cytophotometry* (eds. W. Gohde, J. Schumann et al.) (1976 Mathematical analysis DNA protein two parameter pulse cytophotometric data), European Press, Ghent: 71.
- [3]. Mohd. Kidin Shahrhan, *Prosiding Simposium Kebangsaan Sains Matematik Ke-2 UKM Bangi, Selangor* (1986 Pemodelan fungsi pendaras taburan fasa S dengan variasi kadar sintesis DNA), 351-363.
- [4]. Mohd Kidin Shahrhan, *Prosiding Simposium Kebangsaan Sains Matematik Ke-V; UTM, Sekudai, Johor* (1992 Taburan pendaras mewakili taburan DNA sito-metrik alir), 175-182.
- [5]. Mohd. Kidin Shahrhan (1993 Tinjauan model matematik untuk data DNA FCM FSMK/Sep 1993/LT 30 8).
- [6]. Warren G., *Statistical Modelling*, John Wiley & Sons., New York, 1984.
- [7]. Watson J. V., *A pragmatic approach to the analysis of DNA histograms with a definable peak*, Cytometry **4** (1986), 120-130.