

Algoritma untuk Membangkitkan *Missing Data*

ACENG KOMARUDIN MUTAQIN

Jurusan Statistika FMIPA - Universitas Islam Bandung
Email: shidiq_03@yahoo.com

ABSTRAK

Makalah ini membahas suatu algoritma untuk membangkitkan missing data (data hilang) untuk kasus univariat. Pendekatan yang digunakan untuk membangkitkan jenis data tersebut adalah teknik komposisi (composition technique). Untuk mengimplementasikan algoritma tersebut, diberikan contoh program komputer yang ditulis dalam makro lokal MINITAB.

Kata Kunci: incomplete data, missing data, composition technique

1. Pendahuluan

Missing data (data hilang) merupakan salah satu jenis data yang termasuk dalam *incomplete data* (data tak-lengkap). *Missing data* sering ditemukan dalam berbagai bidang seperti sosial, komputasi, biologi, kesehatan, dan fisika (Meng, 2000). *Missing data* terjadi dalam suatu kumpulan data jika beberapa entri dalam kumpulan data tersebut tidak terobservasi. Tidak terobservasinya data misalnya pada suatu survey, responden menolak untuk menjawab pertanyaan tertentu atau responden tidak mampu menyatakan pilihan dari beberapa pilihan yang ada. Pada studi pengamatan berulang, objek penelitian tidak ada pada satu atau beberapa waktu pengamatan yang telah ditetapkan sebelumnya. Tidak terobservasinya data juga dapat disebabkan oleh data tidak tercatat karena alat ukur yang digunakan rusak, atau datanya benar-benar hilang. *Missing data* juga sering ditemukan dalam data deret waktu.

Sepanjang abad terakhir ini, beberapa statistikawan dan ilmuwan telah menemukan beberapa metode untuk menangani *missing data*. Satu metode untuk menangani *missing data* adalah metode imputasi. Metode imputasi dapat berupa deterministik atau stokastik. Algoritma EM (*Expectation – Maximization*) termasuk yang sifatnya deterministik, sedangkan yang sifatnya stokastik biasanya menggunakan teknik resampling seperti bootstrap.

Karena banyaknya perhatian terhadap *missing data*, maka sangat penting mempunyai suatu metode untuk membangkitkan *missing data*. Makalah ini menawarkan suatu algoritma untuk membangkitkan *missing data* untuk kasus univariat. Algoritmanya didasarkan pada teknik komposisi – *composition technique* (Law & Kelton, 1991). Salah satu pendekatan yang sering digunakan untuk membangkitkan data dari suatu distribusi yang fungsi densitas (masa) peluangnya dapat dinyatakan sebagai kombinasi dari beberapa fungsi densitas (masa) peluang lainnya.

Makalah ini disusun sebagai berikut. Bagian 2 membahas algoritma untuk membangkitkan *missing data* menggunakan teknik komposisi untuk kasus univariat. Contoh program komputer untuk mengimplementasikan algoritma tersebut diberikan dalam Bagian 3.

2. Algoritma

Misalkan X merupakan peubah acak kontinu (diskrit) dari data lengkap (*complete data*) dengan fungsi densitas peluang (fungsi masa peluang) $f(x)$ pada $x \in S$. Misalkan juga bahwa p menyatakan proporsi data yang tidak terobservasi (data hilang), maka fungsi densitas peluang (fungsi masa peluang) dari peubah acak *missing data*, Y , dapat ditulis sebagai berikut:

$$g(y) = pI_{(*)}(y) + (1-p)I_{(A)}(y)f(y) \quad (1)$$

dimana * menyatakan data tidak terobservasi dan I_A menyatakan fungsi indikator dari himpunan A , yaitu:

$$I_A(y) = \begin{cases} 1 & ; \text{untuk } y \in A \\ 0 & ; \text{untuk } y \text{ lainnya} \end{cases} \quad (2)$$

22 Aceng Komarudin Mutaqin

Dengan menggunakan teknik komposisi – *composition technique* (Law & Kelton, 1991), algoritma untuk membangkitkan peubah acak Y adalah sebagai berikut:

Bangkitkan U dari distribusi uniform, $U(0,1)$.

Jika $U \leq p$, maka $Y = *$ (data tidak terobservasi).

Jika $U > p$, maka bangkitkan Y dari distribusi dengan fungsi densitas peluang (fungsi masa peluang) $f(y)$ (data terobservasi).

3. Program Komputer

Dalam bagian ini akan diberikan dua contoh membangkitkan *missing data*. contoh pertama untuk kasus kontinu (distribusi Poisson) dan contoh kedua untuk kasus diskrit (distribusi normal). Kedua contoh tersebut ditulis dalam makro lokal MINITAB, dan MATLAB 6.5.

Contoh 1. (Distribusi Poisson)

Gambar 1 menyajikan program dan output membangkitkan *missing data* berukuran n dari distribusi Poisson yang ditulis dalam makro lokal MINITAB.

Contoh 2. (Distribusi Normal)

Gambar 2 menyajikan program dan output membangkitkan *missing data* berukuran n dari distribusi normal yang ditulis dalam makro lokal MINITAB.

MACRO	6	*
MD1 n p lambda Y	7	6
mconstant i n p lambda	4	3
mcolumn U X Y	*	3
base 100	5	7
random n U;	8	2
uniform 0 1.	7	9
do i=1:n	3	4
if U(i) <= p	1	7
let Y(i) = '*'	4	8
else	5	*
random 1 X;	8	3
poisson lambda.	4	5
let Y(i) = X(1)	7	6
endif	6	5
enddo		
ENDMACRO		

(a)

(b)

Gambar 1. Program dan Output Makro Lokal MINITAB untuk Membangkitkan *Missing Data* Berukuran n dari Distribusi Poisson. (a) Program Makro Lokal MINITAB
(b) Output *Missing Data* dari Distribusi Poisson $\lambda = 5$, $n = 30$, dan $p = 0.2$.

<pre> MACRO MD1 n p mu sigma Y mconstant i n p mu sigma mcolumn U X Y base 100 random n U; uniform 0 1. do i=1:n if U(i)<=p let Y(i)='*' else random 1 X; normal mu sigma. let Y(i)=X(1) endif enddo ENDMACRO </pre>	<table border="1"> <tr><td>26.3972</td><td>*</td></tr> <tr><td>19.5075</td><td>23.1291</td></tr> <tr><td>19.6377</td><td>25.6626</td></tr> <tr><td>*</td><td>22.4312</td></tr> <tr><td>26.6461</td><td>18.6874</td></tr> <tr><td>28.6761</td><td>25.4476</td></tr> <tr><td>26.3220</td><td>25.1252</td></tr> <tr><td>23.9615</td><td>27.8171</td></tr> <tr><td>24.1144</td><td>*</td></tr> <tr><td>26.5960</td><td>20.2657</td></tr> <tr><td>24.3620</td><td>*</td></tr> <tr><td>24.6988</td><td>19.0478</td></tr> <tr><td>24.4138</td><td>26.5067</td></tr> <tr><td>*</td><td>28.8656</td></tr> <tr><td>30.3261</td><td>27.8915</td></tr> </table>	26.3972	*	19.5075	23.1291	19.6377	25.6626	*	22.4312	26.6461	18.6874	28.6761	25.4476	26.3220	25.1252	23.9615	27.8171	24.1144	*	26.5960	20.2657	24.3620	*	24.6988	19.0478	24.4138	26.5067	*	28.8656	30.3261	27.8915
26.3972	*																														
19.5075	23.1291																														
19.6377	25.6626																														
*	22.4312																														
26.6461	18.6874																														
28.6761	25.4476																														
26.3220	25.1252																														
23.9615	27.8171																														
24.1144	*																														
26.5960	20.2657																														
24.3620	*																														
24.6988	19.0478																														
24.4138	26.5067																														
*	28.8656																														
30.3261	27.8915																														

(a)

(b)

Gambar 1. Program dan Output Makro Lokal MINITAB untuk Membangkitkan *Missing Data* Berukuran n dari Distribusi Normal. (a) Program Makro Lokal MINITAB (b) Output *Missing Data* dari Distribusi Normal $\mu = 25$, $\sigma = 3$, $n = 30$, dan $p = 0.3$.

4. Daftar Pustaka

Law, A. M. and Kelton, W. D. (1991). *Simulation Modeling and Analysis*. Third Edition, McGraw Hill, Inc., New York.

Meng, X. L. (2000). Missing Data: Dial M for ??? *Journal of American Statistical Association* 95, 1325-1330.