Prosiding SPMIPA. pp. 153-156. 2006 *ISBN : 979.704.427.0*

PERBANDINGAN PEMBANGKITAN DATA ACAK DENGAN METODE TRANSFORMASI INVERS DAN METODE REJEKSI

Rukun Santoso
Program Studi Statistika Jurusan Matematika FMIPA UNDIP
Jl. Prof. Soedarto Kampus UNDIP Tembalang, Semarang

Abstrak: Seringkali penelitian statistika teoritis memerlukan data dengan distribusi tertentu. Pada kenyataannya data tersebut sulit diperoleh, sehingga diupayakan melalui pengambilan data acak dengan simulasi komputer. Tulisan ini membahas perbandingan dua metode pembangkitan data acak dengan simulasi komputer yaitu metode transformasi invers dan metode rejeksi. Dalam hal ini suatu metode dipandang lebih baik jika memberikan persentase penerimaan dalam uji kocхарактер distribusi yang lebih besar.

Kata Kunci: distribusi peluang kontinu, peluang kumulatif kontinu

PENDAHULUAN

Dalam penelitian statistika yang bersifat teoritik biasanya diperlukan data yang berasal dari distribusi peluang tertentu. Seringkali data yang sesuai dengan distribusi yang diasumsikan tidak dapat diperoleh. Dalam hal ini perlu dicari alternatif lain berupa data simulatif yang dibangkitkan secara acak dari ditribusi tertentu dengan bantuan komputer. Terdapat beberapa metode untuk melakukan pembangkitan data acak dari distribusi peluang kontinu yang bentuk fungsinya diberikan, seperti metode transformasi invers, metode rejeksi, dan metode nilai Hazard[2]

Beberapa perangkat lunak komputer seperti SPSS 11, Minitab, dan S-PLUS telah menyediakan fungsi pembangkit bilangan acak untuk distribusi peluang yang umum digunakan seperti distribusi normal, binomial, chi-squared, F, student t, dan uniform. Namun untuk distribusi peluang lain yang tidak dikenal secara umum perlu disusun algorithma dan kode program sendiri (*user defined).* Salah satu perangkat lunak komputer yang memadai untuk melakukan berbagai kebutuhan dalam riset statistika adalah S-PLUS. Disamping tersedia fungsi-fungsi tercadang, S-PLUS menyediakan fasilitas *programing* yang memungkinkan pengguna untuk membuat kode program sendiri sesuai kebutuhan.

METODE TRANSFORMASI INVERS

Metode transformasi invers digunakan untuk membangkitkan data acak dari distribusi peluang kontinu yang diketahui bentuk fungsinya. Algoritma metode ini diturunkan dari sifat 1.

Sifat 1. Diberikan U peubah acak uniform (0,1). Untuk sebarang fungsi distribusi peluang kumulatif kontinu F jika didefinisikan variabel acak $X=F^{-1}(U)$, maka X mempunyai distribusi peluang kumulatif F.

Bukti. Untuk sebarang quantil a berlaku

$$\begin{aligned} F_X(a) &= P\{X\leq a\} \\ &= P\{F^{-1}(U)\leq a\} \end{aligned}$$

Karena F fungsi monoton maka $F^{-1}(U)\leq a$ berlaku jika hanya jika $U\leq F(a)$. Sehingga diperoleh

$$\begin{aligned} P\{F^{-1}(U)\leq a\} &= P\{ U\leq F(a)\} \\ &= F(a) \quad \blacksquare \end{aligned}$$

Berdasarkan sifat 1 disusun algoritma untuk membangkitkan n bilangan acak dari distribusi F dan menguji tingkat kecocokan data tersebut dalam mengikuti distribusi F.

1. Bangkitkan sebuah bilangan acak u dari distribusi uniform U(0,1)
2. Hitung x dengan $x=F^{-1}(u)$. Dalam hal ini x adalah bilangan acak yang dicari
3. Ulangi langkah 1 dan 2 hingga mendapatkan sebanyak n bilangan acak

4. Lakukan uji kecocokan distribusi Kolmogorov-Smirnov untuk n bilangan acak yang diperoleh terhadap distribusi F

METODE REJEKSI

Misalkan telah diketahui suatu metode untuk membangkitkan bilangan acak dari densitas kontinu g(x). Pembangkitan bilangan acak dari densitas kontinu f(x) dapat dilakukan berdasarkan pembangkitan bilangan acak dari g(x) dengan cara menerima atau menolak hasil pembangkitan secara proporsional terhadap perandingan f(X)/g(X).

Diberikan c bernilai konstat sedemikian hingga berlaku

$$\frac{f(x)}{g(x)} \leq c$$

untuk semua $x \in R$, maka pembangkitan n bilangan acak dari densitas kontinu f(x) dapat mengikuti langkah berikut

1. Bangkitkan bilangan acak X dari densitas g dan bangkitkan bilangan acak U dari U(0,1)
2. Jika $\dfrac{f(X)}{c.g(X)} \geq U$ maka diambil Y=X dengan Y bilangan acak dari densitas f, jika tidak ulangi langkah 1.
3. Ulangi langkah 1 dan 2 sampai mendapatkan $Y_1, Y_2, \ldots, Y_n$

Sifat 2. Variabel acak Y yang dibangkitkan dengan metode rejeksi mempunyai distribusi peluang f

Bukti. Ambil Y hasil yang diperoleh dan N banyaknya iterasi yang diperlukan

$P\{Y \leq y\} = P\{X_N \leq y\}$

$$= P\{X \leq y | \frac{f(X)}{c.g(X)} \geq U\}$$

$$= \frac{P\{X \leq y, \frac{f(X)}{c.g(X)} \geq U\}}{K} \quad \text{dengan } K = P\{\frac{f(X)}{c.g(X)} \geq U\}$$

$$= \frac{\int P\{X \leq y, \frac{f(X)}{c.g(X)} \geq U \mid X = x\} g(x)dx}{K}$$

$$= \frac{\int_{-\infty}^{y} \frac{f(x)}{c.g(x)} g(x)dx}{K}$$

$$= \int_{-\infty}^{y} \frac{f(x)}{cK} dx$$

$$= \int_{-\infty}^{y} f(x)dx \quad \text{untuk } x \to \infty \quad \blacksquare$$

PEMBAHASAN

Untuk membandingkan metode invers dan metode rejeksi mula-mula kedua metode digunakan untuk membangkitkan data acak ukuran n tertentu, dalam tulisan ini diambil n=20, dari densitas peluang yang diberikan. Dalam tulisan ini diberikan densitas

$$f(x) = \alpha(\tfrac{1}{x})^{\alpha+1}, x > 1$$

dengan parameter α=1.

Selanjutnya data acak yang diperoleh dari masing-masing metode diuji kecocokan distribusinya denga uji kecocokan distribusi Kolmogorov-Smirnov. Proses pembangkitan data dan pengujian kecocokan distribusi diulangi sebanyak K kali, dalam tulisan in diambil K=20. Hasil yang diperoleh dapat dilihat pada tabel 1. Dari hasil tersebut uji kecocokan distribusi Kolmogorov-Smirnov memberikan kesimpulan yang cenderung sama untuk kedua metode. Kode program S-PLUS untuk pembangkitan data masing-masing metode serta uji perbandingannya ada pada lampiran.

Tabel 1. Statistik untuk Uji Kolmogorov-Smirnov

No	Metode Invers	Metode Rejeksi
(1)	(2)	(3)
1	0.19113647	0.2222693
2	0.15561656	0.1233807
3	0.19731645	0.1113558
4	0.19996062	0.2659975
5	0.14296094	0.1303280
6	0.17874166	0.2560272
7	0.19324084	0.1320149
8	0.32879235	0.2527861
9	0.11693419	0.1764089
10	0.26141846	0.3371369
11	0.23621021	0.1433832
12	0.18270650	0.2297219
13	0.19414563	0.2715748
14	0.17174603	0.2548396
15	0.16738349	0.1119099
16	0.20982712	0.1835328
17	0.16501631	0.1793180
18	0.16604243	0.2660100
19	0.18292247	0.2896157
20	0.08471651	0.3230420

DAFTAR PUSTAKA

[1]. Conover, W. J., *Practical Nonparametric Statistics*, New York: John Wiley and Sons, 1980.

[2]. Ross, S.M *Introduction to Probability Models* Academic Press, 1997.

[3]. S-Plus User Guide, 1995

LAMPIRAN

```
rinv1<-function(a)
{
u<-runif(1)
x<-(1/(1-u))^(1/a)
return(x)
}

rinv<-function(n,a)
{
an<-rep(a,n)
x<-sapply(an,rinv1)
return(x)
}
```

```
rrejek1<-function (a)
{
repeat
{
  u1<-runif(1,1,1000)
  u2<-runif(1)
  if (u2<=(1/u1)^(a+1))
     {
       x<-u1
       break
     }
  }

return(x)
}

rrejek<-function(n,a)
{
an<-rep(a,n)
x<-sapply(an,rrejek1)
return(x)
}

empirik.p<-function(y)
{
        n <- length(y)
        p <- rep(0, n)
        y <- sort(y)
        for(i in 1:n) {
              p.i <- length(y[y <= y[i]])/n
              p[i] <- p.i
        }
        return(unique(y), unique(p))
}

tes.ks<-function(n,a)
{
  x1<-rinv(n,a)
  x2<-rrejek(n,a)
  p1<-empirik.p(x1)
  p2<-empirik.p(x2)
  xp1<- p1[[1]]
  yp1<- p1[[2]]
  xp2<- p2[[1]]
  yp2<- p2[[2]]
  f1<-1-(1/xp1)^a
  f2<-1-(1/xp2)^a
  k1<-length(yp1)-1
  k2<-length(yp2)-1
  yp11<-c(0,yp1[1:k1])
  yp21<-c(0,yp2[1:k2])
  m1<-max(abs(yp1-f1),abs(f1-yp11))
  m2<-max(abs(yp2-f2),abs(f2-yp21))
  return(c(m1,m2))
}

utama<-function(n,a,k,sig)
{
  m<-NULL
  for(i in 1:k)
  {
     ksi<-tes.ks(n,a)
     m<-c(m, i, ksi)
  }
   m<-matrix(m,nrow=k)
   return(m)
}
```