Pendahuluan

STK473 – Praktikum 1

Warm Up

Saya punya 27 buah bola besi yang masing-masing memiliki berat 1 kg. Ternyata salah satu bola besi *fake* memiliki berat 1.1 kg. Bagaimana caranya agar saya bisa menemukan bola besi yang berbeda tersebut jika saya hanya punya kesempatan menimbang menggunakan timbangan neraca **maksimal 3 kali**?

Peubah Acak

- Bernoulli
- Binomial
- Seragam Diskret
- Poisson
- Seragam
- Normal
- Eksponensial

dan lain-lain...

Peubah Acak Diskret

Bernoulli(p)

pmf	$P(X = x \mid p) = p^x(1-p)^{1-x}; \quad x = 0, 1; \quad 0 \le p \le 1$
mean and variance	$\mathrm{E}X = p, \quad \mathrm{Var}\, X = p(1-p)$
mgf	$M_X(t) = (1-p) + pe^t$

Binomial(n, p)

pmf	$P(X = x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, 2, \ldots, n; \quad 0 \le p \le 1$
mean and variance	$\mathrm{E}X = np, \quad \mathrm{Var}\, X = np(1-p)$
mgf	$M_X(t) = [pe^t + (1-p)]^n$

Peubah Acak Diskret

Discrete uniform

pmf	$P(X = x\|N) = \frac{1}{N}; \quad x = 1, 2, \ldots, N; \quad N = 1, 2, \ldots$
mean and variance	$\mathrm{E}X = \frac{N+1}{2}, \quad \mathrm{Var}\, X = \frac{(N+1)(N-1)}{12}$
mgf	$M_X(t) = \frac{1}{N} \sum_{i=1}^{N} e^{it}$

Poisson(λ)

pmf	$P(X = x\|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, \ldots; \quad 0 \leq \lambda < \infty$
mean and variance	$\mathrm{E}X = \lambda, \quad \mathrm{Var}\, X = \lambda$
mgf	$M_X(t) = e^{\lambda(e^t - 1)}$

Peubah Acak Kontinu

Uniform(a, b)

pdf	$f(x\|a,b) = \frac{1}{b-a}, \quad a \le x \le b$
mean and variance	$\mathrm{E}X = \frac{b+a}{2}, \quad \mathrm{Var}\, X = \frac{(b-a)^2}{12}$
mgf	$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$

Exponential(β)

pdf	$f(x\|\beta) = \frac{1}{\beta} e^{-x/\beta}, \quad 0 \le x < \infty, \quad \beta > 0$
mean and variance	$\mathrm{E}X = \beta, \quad \mathrm{Var}\, X = \beta^2$
mgf	$M_X(t) = \frac{1}{1-\beta t}, \quad t < \frac{1}{\beta}$

Peubah Acak Kontinu

$\boldsymbol{Normal}(\mu, \sigma^2)$

pdf	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty,$ $\sigma > 0$
mean and variance	$\mathrm{E}X = \mu, \quad \mathrm{Var}\,X = \sigma^2$
mgf	$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

Review R

```
#membuat matriks
a1<-matrix(1:9,3,3)
a1

#mengakses isi matriks
a1[3,c(1,3)]
a1[1:2,c(1,3)]

#menggunakan looping
sum<-0
for (i in 1:10) sum<-sum+i
sum
```

Review R

```
#membangkitkan bilangan yang menyebar seragam
seragam<-runif(1000,3,14)
hist(seragam)
plot(density(seragam))
plot.ecdf(seragam)

#membangkitkan bilangan yang menyebar binomial
binomial<-rbinom(1000,20,0.3)
barplot(table(binomial))
plot.ecdf(binomial)
```

Contoh Soal

Misalkan terdapat 12 nasabah asuransi di suatu tempat. Diketahui bahwa proporsi nasabah telat bayar polis ialah 1/6. Jika antar nasabah saling bebas, tentukanlah peluang bahwa terdapat 7 sampai 9 nasabah yang telat bayar polis!

- X = banyaknya nasabah asuransi yang telat membayar
- X ~ binomial(n = 12, p = 1/6)
- $P(7 \leq X \leq 9)$

```
> pbinom(9,size=12,p=1/6) - pbinom(6,size=12,p=1/6)
[1] 0.001291758
> dbinom(7,12,1/6) + dbinom(8,12,1/6) + dbinom(9,12,1/6)
[1] 0.001291758
```

Mengambil Contoh Acak

```
seragam1<-sample(seragam,100,replace=T)
hist(seragam1)

binomial1<-sample(binomial,100,replace=T)
barplot(table(binomial1))
```

Uniform Random Number

STK473 – Praktikum 2

Uniform Random Variable

PDF random variables `X ~ Uniform (a,b)`

$$f(x) = \begin{cases} \dfrac{1}{b-a} & , a \leq x \leq b \\ 0 & , x \ lainnya \end{cases}$$

We need to generate n random number $x_1, x_2, \dots, x_n$ that uniformly distributed.

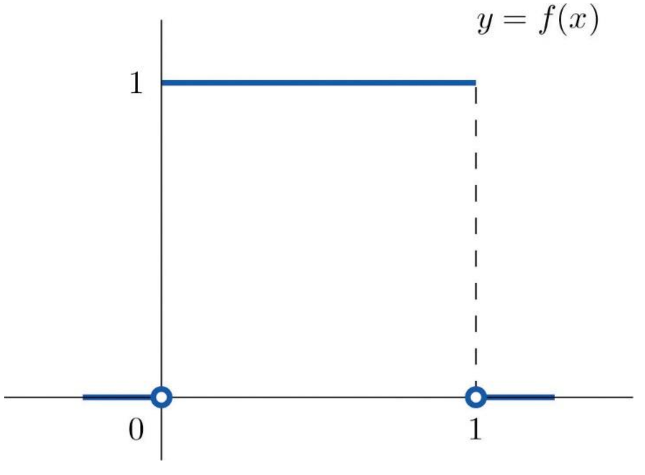
Why Uniform First?

1. The simplest of continuous distribution that described distribution of some number intervals
2. From Uniform, can be transformed to be others complicated distributions

So, how we generate random numbers that uniformly distributed ?

Start from the simplest Uniform distribution:

X ~ Uniform (0,1)



Congruential Generator

X ~ Uniform (0,1)

So we have intervals of x_i, $0 \leq x_i \leq 1$

Congruential Generator :

$$X_{n+1} = aX_n + b \ (\text{mod}\ m), n \geq 0$$

$$U_i = \frac{X_i}{m} \sim U(0,1)$$

Congruential Generator (2)

Maximum cycle

- b & m don't have same factors
- (a-1) (mod prime factor of m) = 0
- (a-1) (mod 4) = 0, if m (mod 4) = 0

So we got :

$$\boxed{m = 2^k, k \geq 2} \quad \boxed{a = 4c + 1}$$

$$\boxed{b > 0, \text{odd numbers}}$$

Independence of Observations

$$cov\left(X_i, X_j\right) \approx 0$$

$$\rho = \left[\frac{1}{a} - \frac{6b}{am}\left(1 - \frac{b}{m}\right)\right] \pm \frac{a}{m}$$

$$X_{n+1} = aX_n + b \pmod m, n \geq 0$$

$$U_i = \frac{X_i}{m} \sim U(0,1)$$

$a = 1598$

$b = 17$

$m = 1000$

$X_0 = 78$

i	$aX_{i-1} + b$	X_i	U_i
0		78	
1	124661	661	0.661
2	1056295	295	0.295
3	471427	427	0.427
4	682363	363	0.363
5	580091	91	0.091
6	145435	435	0.435
7	695147	147	0.147
8	234923	923	0.923
9	1474971	971	0.971
10	1551675	675	0.675
11	1078667	667	0.667
12	1065883	883	0.883
13	1411051	51	0.051

Application in R

```
x0<-78
n<-250
xi<-matrix(NA,n,3)
colnames(xi)<-c("aX(i-1)+b","Xi","Ui")
for (i in 1:n)
{
  xi[i,1]<-(1598*x0+17)
  xi[i,2]<-xi[i,1]%%1000
  xi[i,3]<-xi[i,2]/1000
  x0<-xi[i,2]
}
hist(xi[,3])
```

Other Generator Function

$$\boxed{U_{n+1} = (\pi + U_n)^5 (\text{mod } 1), n \geq 0}$$

```
n<-1000
x1<-0.9
for (i in 2:n) x1[i]<-(pi+x1[i-1])^5%%1
hist(x1)
```

```
> cbind("i"=1:20, "Xi"=xi[1:20,2], "i"=101:120,
  "Xi"=xi[101:120,2], "i"=201:220,
  "Xi"=xi[201:220,2])

       i  Xi   i  Xi   i  Xi
 [1,]  1 661 101 411 201 411
 [2,]  2 295 102 795 202 795
 [3,]  3 427 103 427 203 427
 [4,]  4 363 104 363 204 363
 [5,]  5  91 105  91 205  91
 [6,]  6 435 106 435 206 435
 [7,]  7 147 107 147 207 147
 [8,]  8 923 108 923 208 923
 [9,]  9 971 109 971 209 971
[10,] 10 675 110 675 210 675
[11,] 11 667 111 667 211 667
[12,] 12 883 112 883 212 883
[13,] 13  51 113  51 213  51
[14,] 14 515 114 515 214 515
[15,] 15 987 115 987 215 987
[16,] 16 243 116 243 216 243
[17,] 17 331 117 331 217 331
[18,] 18 955 118 955 218 955
[19,] 19 107 119 107 219 107
[20,] 20   3 120   3 220   3
```

Repeated Series

If we have $X \sim U(0,1)$, how to generate $Y \sim U(a,b)$?

$$X \sim \mathrm{Uniform}\ (0,1) \xrightarrow{?} Y \sim \mathrm{Uniform}\ (a,b)$$

X ~ Uniform (0,1)

$$f(x) = \begin{cases} 1, 0 \leq x \leq 0 \\ 0 \ , x\ lainnya \end{cases}$$

Y ~ Uniform (a,b)

$$f(x) = \begin{cases} \dfrac{1}{b-a}, a \leq x \leq b \\ 0 \ , x\ lainnya \end{cases}$$

Y = (b-a)X + a

Application in R

Y ~ Uniform (3,20)

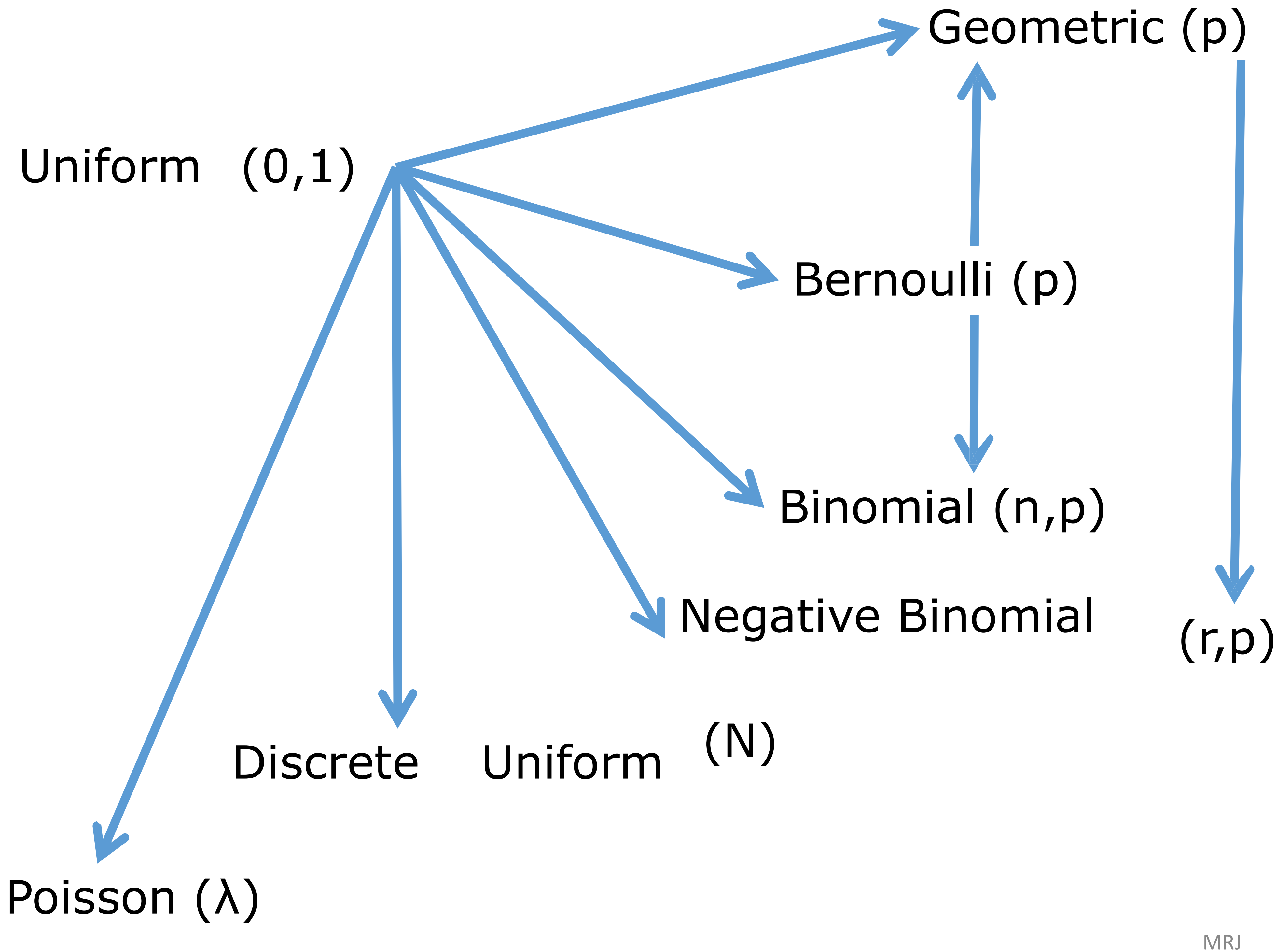
```
Y<-(20-3)*x1+3
hist(Y)
```

thank you!

Discrete & Continous Random Number

STK473 – Praktikum 3

Discrete Random Number



Geometric (p)
Uniform (0,1)
Bernoulli (p)
Binomial (n,p)
Negative Binomial
(r,p)
Discrete Uniform (N)
Poisson (λ)
MRJ

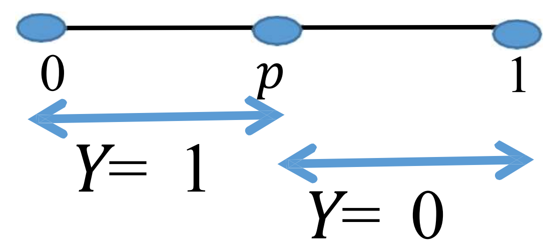
Bernoulli (p)

Uniform (0,1) $\xrightarrow{F_X(x)}$ Bernoulli (p)

→ X ~ Uniform (0,1)

$$0 \leq F_X(x) \leq 1$$

$$0 \leq X \leq 1$$



→ Y ~ Bernoulli (p)

***Bernoulli*(p)**

pmf	$P(X = x \mid p) = p^x(1-p)^{1-x}; \quad x = 0, 1; \quad 0 \leq p \leq 1$
mean and variance	$\mathrm{E}X = p, \quad \mathrm{Var}\, X = p(1-p)$
mgf	$M_X(t) = (1-p) + pe^t$

Application in R

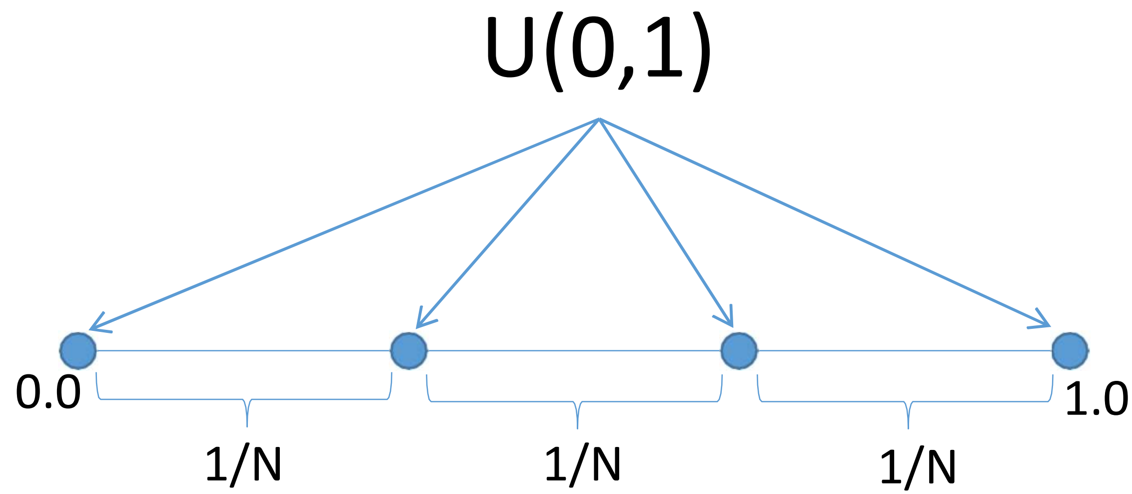
```
i<-1000
p<-.65
X<-runif(i)
Y<-NULL
for (z in 1:i) ifelse (X[z]<=p,Y[z]<-1,Y[z]<-0)
(tabel<-table(Y)/length(Y))
barplot(tabel,main="Bernoulli")

i<-1000
p<-.65
X<-runif(i)
Y<-(X<=p)+0
(tabel<-table(Y)/length(Y))
barplot(tabel,main="Bernoulli")
```

Discrete Uniform (N)

Uniform (0,1)

$\Big\downarrow F_X(x)$

Discrete Uniform (N)



Discrete uniform

pmf	$P(X=x\|N)=\frac{1}{N};\quad x=1,2,\ldots,N;\quad N=1,2,\ldots$
mean and variance	$\mathrm{E}X=\frac{N+1}{2},\quad \mathrm{Var}\,X=\frac{(N+1)(N-1)}{12}$
mgf	$M_X(t)=\frac{1}{N}\sum_{i=1}^{N}e^{it}$

Application in R

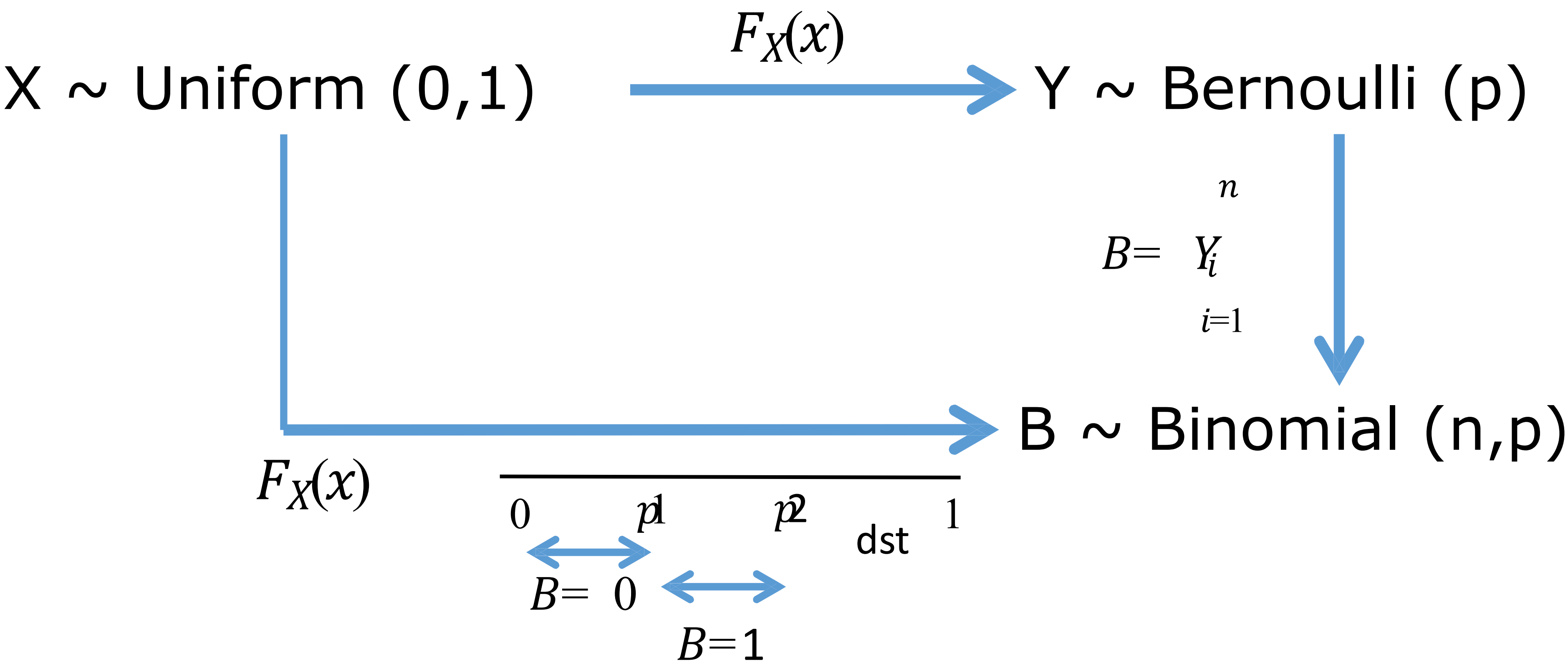
```
i<-1000
N<-4
X<-runif(i)
DU<-NULL
for (z in 1:i){
  if (X[z]<=1/N) DU[z]<-1
  else if (X[z]<=2/N) DU[z]<-2
  else if (X[z]<=3/N) DU[z]<-3
  else DU[z]<-4
}
(tabel<-table(DU)/length(DU))
barplot(tabel,main="Seragam Diskret")
```

Application in R

```
i<-1000
N<-4
X<-runif(i)
DU<-as.numeric(cut(X,breaks=c(0,1/N,2/N,3/N,1),
  include.lowest = T))
(tabel<-table(DU)/length(DU))
barplot(tabel,main="Seragam Diskret")
```

```
i<-1000
N<-4
X<-runif(i)
DU<-1+floor(N*X)
(tabel<-table(DU)/length(DU))
barplot(tabel,main="Seragam Diskret")
```

Binomial (n,p)



$F_X(x)$
X ~ Uniform (0,1)
Y ~ Bernoulli (p)
$B = \sum_{i=1}^{n} Y_i$
B ~ Binomial (n,p)
$F_X(x)$
0
p1
p2
dst
1
$B = 0$
$B = 1$

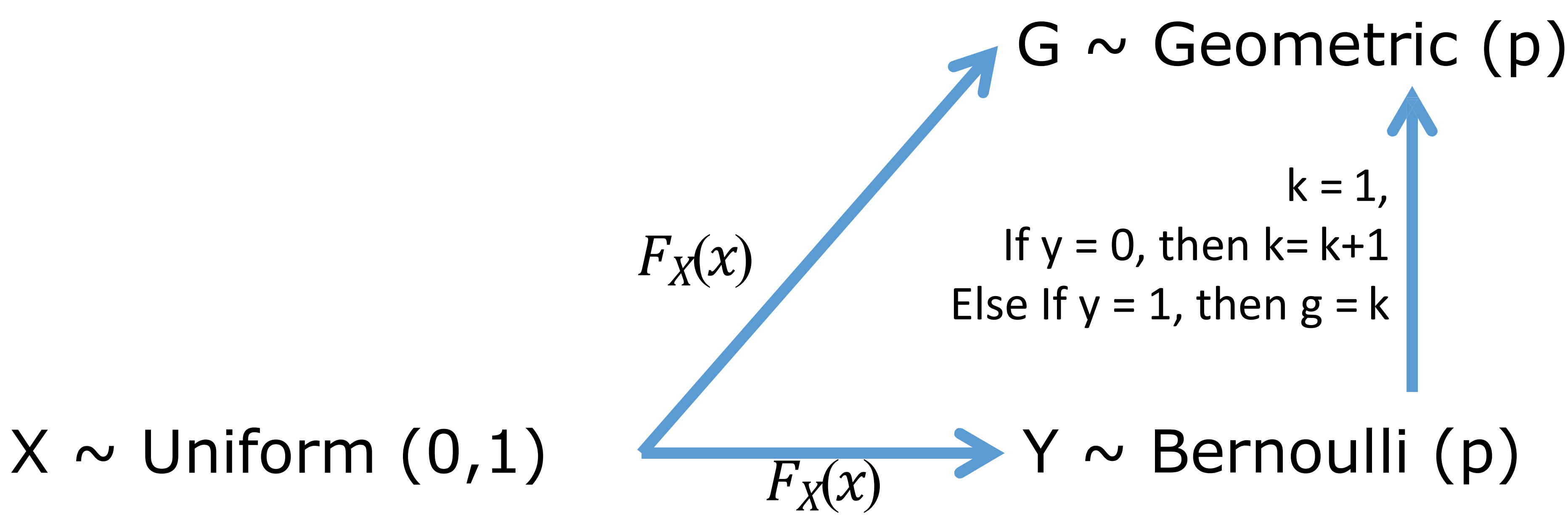
Application in R

```
#Binomial (5,0.65)
i<-1000
n<-5
p<-0.65
Binom<-NULL
for (z in 1:i){
  m<-0
  for (k in 1:n){
    y<-(runif(1)<=p)+0
    m<-m+y
  }
  Binom[z]<-m
}
(tabel<-table(Binom)/length(Binom))
barplot(tabel,main="Binomial")
```

Application in R

```
#Binomial (3,0.5)

i<-1000

X<-runif(i)

Binom<-as.numeric(cut(X,breaks=c(0,1/8,4/8,7/8,1),
   include.lowest = T))-1

(tabel<-table(Binom)/length(Binom))

barplot(tabel,main="Binomial")
```

Geometric (p)



Geometric(p)

pmf	$P(X = x \mid p) = p(1-p)^{x-1}; \quad x = 1, 2, \ldots; \quad 0 \leq p \leq 1$
mean and variance	$\mathrm{E}X = \frac{1}{p}, \quad \mathrm{Var}\, X = \frac{1-p}{p^2}$
mgf	$M_X(t) = \frac{pe^t}{1-(1-p)e^t}, \quad t < -\log(1-p)$
notes	$Y = X - 1$ is negative binomial$(1, p)$. The distribution is *memoryless*: $P(X > s \mid X > t) = P(X > s - t)$.

Negative Binomial (r,p)

$$
\begin{tikzcd}
X \sim \text{Uniform } (0,1) \arrow[r, "F_X(x)"] \arrow[dr, "F_X(x)"'] & G \sim \text{Geometric } (p) \arrow[d] \\
 & Y \sim \text{Neg Binomial } (r,p)
\end{tikzcd}
$$

Negative binomial(r, p)

pmf	$P(X = x\|r, p) = \binom{r+x-1}{x} p^r (1-p)^x; \quad x = 0, 1, \ldots; \quad 0 \le p \le 1$
mean and variance	$\mathrm{E}X = \frac{r(1-p)}{p}, \quad \mathrm{Var}\, X = \frac{r(1-p)}{p^2}$
mgf	$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r, \quad t < -\log(1-p)$
notes	An alternate form of the pmf is given by $P(Y = y\|r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$, $y = r, r+1, \ldots$. The random variable $Y = X + r$. The negative binomial can be derived as a gamma mixture of Poissons. (See Exercise 4.34.)

Tugas 1

Buatlah program R untuk membangkitkan 1000 bilangan acak yang menyebar:

- **Geometrik**
- **Binomial Negatif**

Dikumpulkan paling lambat hari .....
di GCR.

Format:

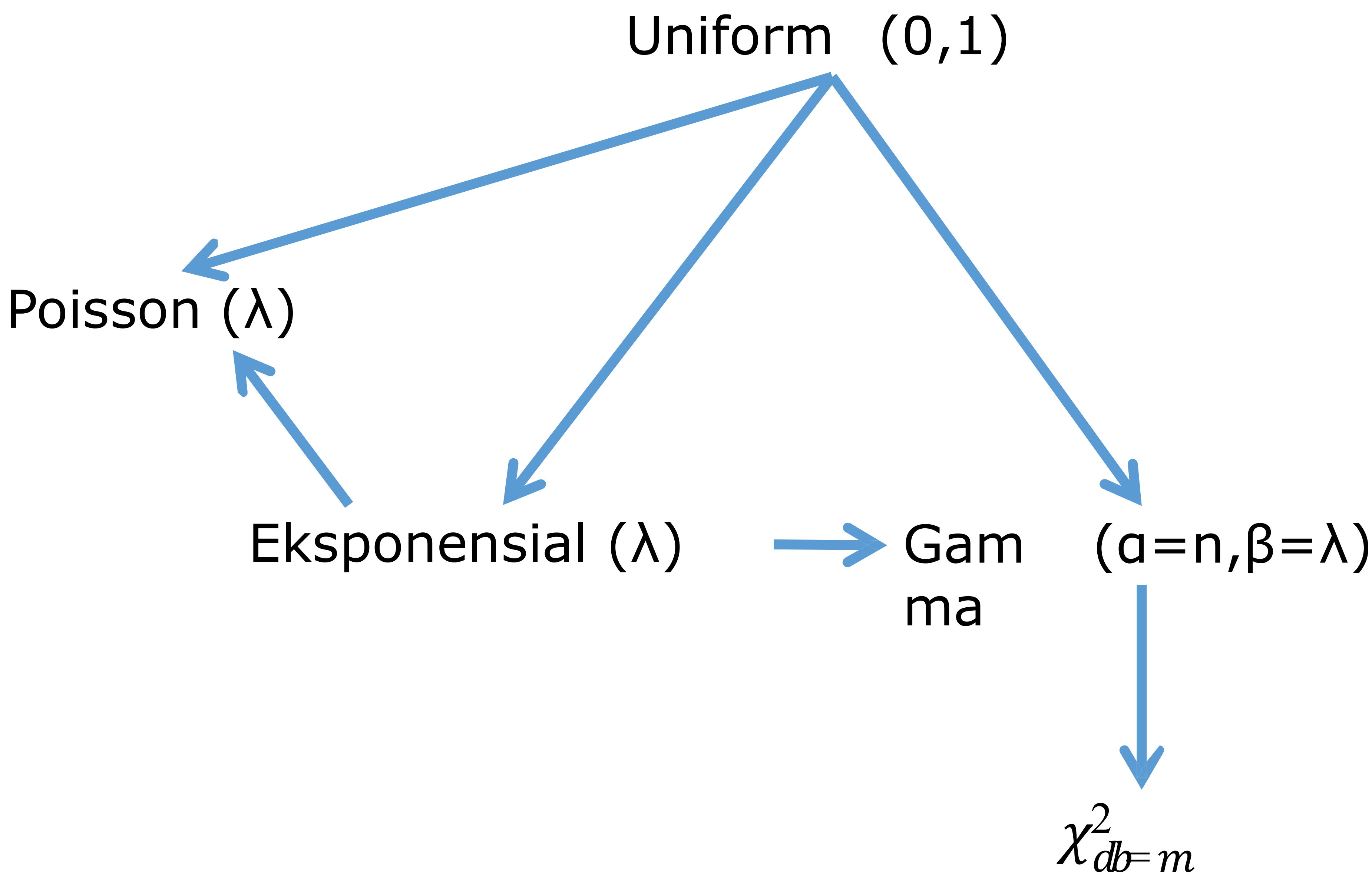
-Nama file: “Tugas 1 Kelompok [no kelompok]”

-Ekstensi file: “.r” atau “.txt”

Continous Random Numbers

Inverse Transform Method

- Metode Transformasi Kebalikan
- Dikenal juga sebagai Look-Up Table Method
- Didasari pada kenyataan bahwa
  - jika U adalah bilangan acak Seragam(0, 1)
  - dan didefinisikan $X = F^{-1}(U)$, dengan $F^{-1}(U)$ adalah fungsi kebalikan dari $F(X)$
  - maka X akan memiliki sebaran yang diinginkan
- Algoritma untuk mendapatkan bilangan acak X dengan sebaran tertentu
  - Tentukan bentuk dari fungsi sebaran kumulatif X yang diinginkan, misal $F(x)$
  - Cari fungsi kebalikan dari $F(x)$, yaitu $F^{-1}(x)$
  - Bangkitkan bilangan acak Seragam (0, 1), misal dilambangkan U
  - Hitung $X = F^{-1}(U)$


Uniform (0,1)
Poisson (λ)
Eksponensial (λ)
Gam
ma
(α=n,β=λ)
$\chi^2_{db=m}$

Inverse Transform Method

Seragam (a, b)

- Ilustrasi untuk membangkitkan sebaran Seragam(a, b)

- X ~ Seragam(a, b)
  - $F(x) = (x – a) / (b – a)$
  - $U = (x – a) / (b – a)$
  - $X = a + (b – a) U$

- Algoritma:
  - Bangkitkan U, bilangan acak Seragam(0, 1)
  - Hitung X = a + (b – a) * U
  - Ulangi berkali-kali sesuai dengan banyaknya bilangan yang diinginkan

Inverse Transform Method

Eksponensial (λ)

- Ilustrasi untuk membangkitkan sebaran Eksponensial(λ)

- X ~ Eksponensial(λ)
  - $f(x) = \lambda e^{-\lambda x}$, untuk $x \geq 0$
  - $F(x) = 1 - e^{-\lambda x}$, untuk $x \geq 0$
  - $U = 1 - e^{-\lambda x}$, untuk $x \geq 0$
  - $X = -\ln(1 - U) / \lambda$

- Algoritma:
  - Bangkitkan U, bilangan acak Seragam(0, 1)
  - Hitung $X = -\ln(1 - U) / \lambda$
  - Ulangi berkali-kali sesuai dengan banyaknya bilangan yang diinginkan

Application in R

```
#Eksponensial(λ=3)
i<-1000
lambda<-3
U<-runif(i)
X<--log(U)/lambda
hist(X)
```

Gamma (α,β)

$$
\begin{tikzcd}
U\sim\text{Uniform }(0,1) \arrow[d, "X=\frac{-\ln U}{\lambda}"'] \arrow[dr, "G=\frac{-\sum_{i=1}^{n}\ln U_i}{\lambda}"] & \\
X\sim\text{Eksponensial }(\lambda) \arrow[r, "G=\sum_{i=1}^{n} X_i"'] & G\sim\text{Gamma}(\alpha=n,\beta=\lambda)
\end{tikzcd}
$$

Application in R

```
#Gamma(α=5,β=3)
i<-1000
lambda<-3
alpha<-5
U<-log(runif(i*alpha))
Um<-matrix(U,i)
Y<-apply(Um,1,sum)
Gama<--Y/lambda
hist(Gama)
```

Chi-Square (m)

G~Gamma (α=n,β=λ)

↓ Gamma(2n/2,2)

$\chi^2_{db=m=2n}$

Gamma(v/2,2)→$\chi^2_{db=v}$

↓

n

Sehingga ketika $v$ganjil maka n= v/2 akan menghasilkan bilangan pecahan → tidak sesuai

Ingat:

- Jika Z~N(0,1), maka $Z^2 \sim \chi^2_{db=1}$
- Jika $X_1 \sim \chi^2_{db=m}$ dan $X_2 \sim \chi^2_{db=n}$, maka $X_3 = X_1 + X_2 \sim \chi^2_{db=(m+n)}$

Sehingga

$\chi^2{}_{db=v}$

v genap

$$Gamma(v/2,2)$$

v ganjil

$$Gamma\left(\frac{(v-1)}{2},2\right)+Z^2$$

$$Z^2 \sim N(0,1)$$

Application in R

```
#chi-square(10)
i<-1000
lambda<-2
alpha<-5
U<-log(runif(i*alpha))
Um<-matrix(U,i)
Y<-apply(Um,1,sum)
chi<--Y/lambda
hist(chi)
```

```
#chi-square(11)
i<-1000
lambda<-2
alpha<-5
U<-log(runif(i*alpha))
Um<-matrix(U,i)
Y<-apply(Um,1,sum)
chi<--Y/lambda
chi<-chi+(rnorm(i))^2
hist(chi)
```

Inverse Transform Method

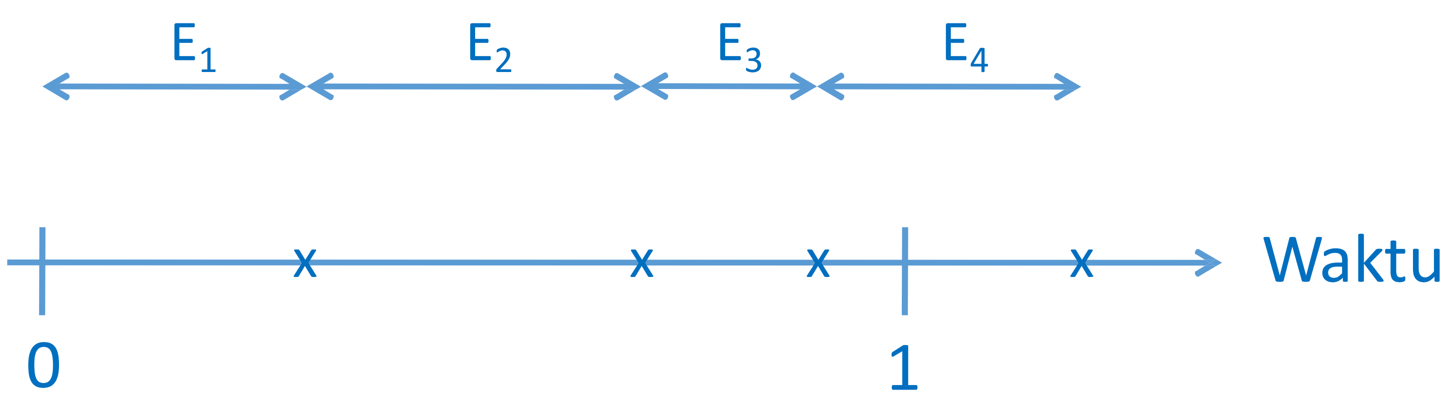
- Kesulitan utama: memperoleh kebalikan dari fungsi sebaran kumulatif
- Keunggulan: bisa digunakan untuk berbagai sebaran (termasuk sebaran diskret)

Memanfaatkan pengetahuan mengenai transformasi dan sifat sebaran peubah acak

- Y_i ~ Eksponensial(λ) $\rightarrow$ G = $\Sigma_i^n Y_i$ ~Gamma(n,λ)
- Gamma(m/2, 2) = Chi-square(m)
- Chi-square(db = 1) = Kuadrat dari Normal(0, 1)
- Jika X_1~Chi-square(m), X_2~Chi-square(n), maka X_3=X_1+X_2~Chi-square(m+n)
- dsb

Poisson(λ)

- Proses Poisson dengan laju sebesar λ
  - Waktu antar kejadian → saling bebas → Eksponensial (λ)
  - Banyaknya kejadian pada setiap selang waktu t menyebar Poisson(λt)



Poisson (λ)

U~Uniform (0,1)

$$X = \frac{-\ln U}{\lambda}$$

Y~Eksponensial (λ)

$$S_0 = 0$$

$$S_k = \sum_{i=1}^{k} Y_i$$

$$S_k \le 1 \le S_{k+1}$$

$$\prod_{i=1}^{k} U_i < e^{-\lambda}$$

K~Poisson (λ)

Application in R

```
#Poisson(1) melalui ekponensial
i<-1000
lambda<-1
K<-NULL
for (z in 1:i){
  sk<-0
  k<-0
  while (sk<=1){
    u<-runif(1)
    y<--log(u)/lambda
    sk<-y+sk
    k<-k+1
  }
  K[z]<-k-1
}
(tabel<-table(K)/length(K))
barplot(tabel)
```

Application in R

```
#Poisson(1) melalui seragam
i<-1000
lambda<-1
K<-NULL
for (z in 1:i){
  k<-0
  sk<-1
  while(sk>=exp(-lambda)){
    u<-runif(1)
    sk<-sk*u
    k<-k+1
  }
  K[z]<-k
}
(tabel<-table(K)/length(K))
barplot(tabel)
```

thank you!

Normal Random Number

STK473 – Praktikum 4

Normal Random Variable

PDF random variables `X ~ Normal (μ,σ²)`

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

We need to generate n random number $x_1, x_2, \ldots, x_n$ that normally distributed.

`Uniform (0,1)` $\longrightarrow$ `Normal (0,1)`

We have
3 ways

Method 1: Central Limit Theorem Approach

$$U_1, U_2, \dots, U_n \sim U(0,1)$$

$$N = \sum_{i=1}^{n} U_i \sim_{n\to\infty} N(\mu, \sigma^2)$$

- So how generate $N(0,1)$ from $U_1, U_2, \dots, U_n \sim U(0,1)$?
- How many n we need?

Uniform(a, b)

pdf	$f(x\|a,b) = \frac{1}{b-a}, \quad a \le x \le b$
mean and variance	$\mathrm{E}X = \frac{b+a}{2}, \quad \mathrm{Var}\,X = \frac{(b-a)^2}{12}$
mgf	$M_X(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$

Normal(μ, σ^2)

pdf	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty,$ $\sigma > 0$
mean and variance	$\mathrm{E}X = \mu, \quad \mathrm{Var}\,X = \sigma^2$
mgf	$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

$$E(U_i) = \frac{a+b}{2} = \frac{0+1}{2} = \frac{1}{2}$$

$$Var(U_i) = \frac{(b-a)^2}{12} = \frac{(1-0)^2}{12} = \frac{1}{12}$$

with

$E(N) = \mu$ and $Var(N) = \sigma^2$

So suppose $N = \sum_{i=1}^{n} U_i + k$

$$Var(N) = Var\left(\sum_{i=1}^{n} U_i + k\right) = \sum_{i=1}^{n} Var(U_i) = \frac{n}{12} = \sigma^2 \qquad n = 12\sigma^2$$

$$E(N) = E\left(\sum_{i=1}^{n} U_i + k\right) = \sum_{i=1}^{n} E(U_i) + k = \frac{n}{2} + k = \mu \qquad k = \mu - \frac{n}{2}$$

so

$N \sim N(0,1)$

$$N = \sum_{i=1}^{12} U_i - 6$$

$$Y = \sqrt{10}\,N + 5$$

$Y \sim N(5,10)$

$$Y = \sum_{i=1}^{120} U_i - 55$$

Application in R

```
i<-1000
mu<-0
sigma2<-1
n<-12*sigma2
k<-mu-n/2

U<-runif(n*i)
Um<-matrix(U,i)
N<-apply(Um,1,sum)+k

mean(N); var(N); hist(N)
```

```
i<-1000
mu<-5
sigma2<-10
n<-12*sigma2
k<-mu-n/2

U<-runif(n*i)
Um<-matrix(U,i)
N<-apply(Um,1,sum)+k

mean(N); var(N); hist(N)
```

Application in R

```
i<-1000
mu<-0
sigma2<-1
n<-12*sigma2
k<-mu-n/2

U<-runif(n*i)
Um<-matrix(U,i)
N<-apply(Um,1,sum)+k

Y<-sqrt(10)*N+5
mean(Y); var(Y); hist(Y)
```

$$N = a\left(\sum_{i=1}^{n} U_i - b\right)$$

```
i<-1000
mu<-40
var<-15
a<-3
n<-12*var/a^2
b<-n/2-mu/a

U<-runif(n*i)
Um<-matrix(U,i)
N<-a*(apply(Um,1,sum)-b)

mean(N); var(N); hist(N)
```

$$Var(N) = a^2\frac{n}{12} = \sigma^2 \qquad n = \frac{12\sigma^2}{a^2}$$

$$E(N) = a\left(\frac{n}{2} - b\right) = \mu \qquad b = \frac{n}{2} - \frac{\mu}{a}$$

Method 2:
Box-Muller Method

$$U_1, U_2 \sim^{id} U(0,1)$$

So

$$N_1 = (-2lnU_1)^{1/2}\cos(2\pi U_2)$$
$$N_2 = (-2lnU_1)^{1/2}\sin(2\pi U_2)$$

Finally we have N_1 & $N_2 \sim^{id} N(0,1)$

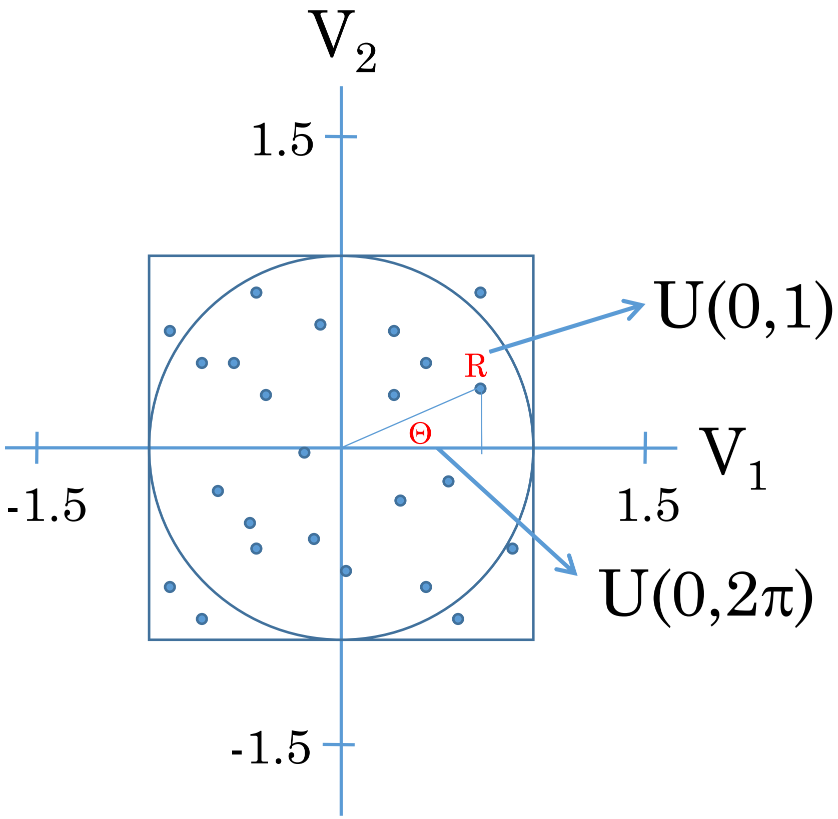
- It can be written:
  - $N_1 = R \cos \Theta$
  - $N_2 = R \sin \Theta$
- with
  - $R = (-2 \log_e U_1)^{1/2}$ and
  - $\Theta = 2\pi U_2 \sim U(0,2\pi)$

- $(N1, N2) \leftrightarrow (R, \Theta)$
- Cartesius Coordinate$\leftrightarrow$ polar coordinate

Application in R

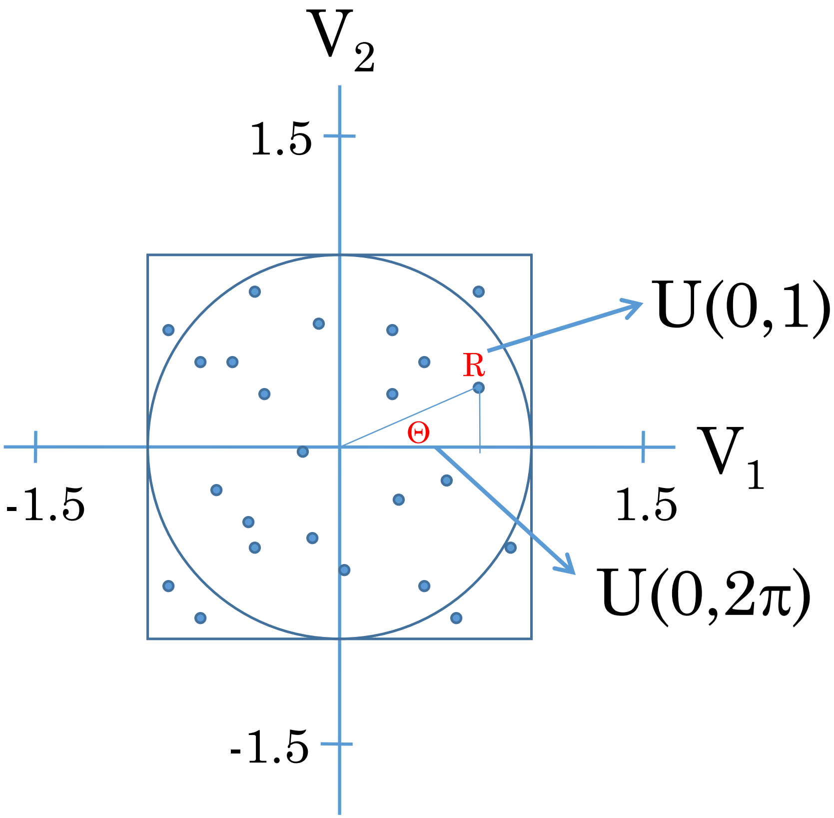
```
i<-1000
U1<-runif(i)
U2<-runif(i)
R<-sqrt(-2*log(U1))
Theta<-2*pi*U2
N1<-R*cos(Theta)
N2<-R*sin(Theta)
mean(N1); var(N1); hist(N1)
mean(N2); var(N2); hist(N2)
```

Method 3: Polar Marsaglia



- if $U \sim U(0,1)$
  - $2U \sim U(0,2)$
  - $V = (2U - 1) \sim U(-1,1)$

- $V_1, V_2 \sim U(-1,1)$
  - $R^2 = V_1^2 + V_2^2$
  - $\tan \Theta = V_2/V_1$



- from Box-Muller
  - $N_1 = (-2 \log_e U_1)^{1/2} \cos(2\pi U_2)$
  - $N_2 = (-2 \log_e U_1)^{1/2} \sin(2\pi U_2)$

R

$\sin \Theta = V_2/R$

$\cos \Theta = V_1/R$

$V_1(V_1^2+V_2^2)^{-1/2}$

$V_2(V_1^2+V_2^2)^{-1/2}$

- $N_1 = (-2 \log_e U_1)^{1/2} \cos (2\pi U_2)$
- $N_2 = (-2 \log_e U_1)^{1/2} \sin (2\pi U_2)$

- $N_1 = (-2 \log_e R^2)^{1/2} V_1(V_1^2+V_2^2)^{-1/2}$
- $N_2 = (-2 \log_e R^2)^{1/2} V_2(V_1^2+V_2^2)^{-1/2}$

- $N_1 = (-2 \log_e (V_1^2+V_2^2))^{1/2} V_1(V_1^2+V_2^2)^{-1/2}$
- $N_2 = (-2 \log_e (V_1^2+V_2^2))^{1/2} V_2(V_1^2+V_2^2)^{-1/2}$

- $N_1 = V_1\{(-2 \log_e W)/W\}^{1/2}$
- $N_2 = V_2\{(-2 \log_e W)/W\}^{1/2}$

- with $W = V_1^2+V_2^2$

$$V_1, V_2 \sim^{iid} U(-1,1)$$

- So

$$N_1 = V_1 \left[\frac{-2lnW}{W}\right]^{1/2}$$

$$N_2 = V_2 \left[\frac{-2lnW}{W}\right]^{1/2}$$

- With $W = V_1^2 + V_2^2$
- Finally we have N_1 & $N_2 \sim^{iid} N(0,1)$
- Simulation???

Application in R

```
i<-1000
U1<-runif(i)
U2<-runif(i)
V1<-2*U1-1
V2<-2*U2-1
W<-V1^2+V2^2

W1<-W[W<1]
V1<-V1[W<1]
V2<-V2[W<1]

i_1<-i-length(W1)
while (i_1>0) {
  U_1<-runif(i_1)
  U_2<-runif(i_1)
  V_1<-2*U_1-1
  V_2<-2*U_2-1
  W_1<-V_1^2+V_2^2
  W1<-c(W1,W_1[W_1<1])
  V1<-c(V1,V_1[W_1<1])
  V2<-c(V2,V_2[W_1<1])
  i_1<-i-length(W1)
}

N1<-V1*sqrt(-2*log(W1)/W1)
N2<-V2*sqrt(-2*log(W1)/W1)

mean(N1);var(N1);hist(N1)
mean(N2);var(N2);hist(N2)
```

thank you!

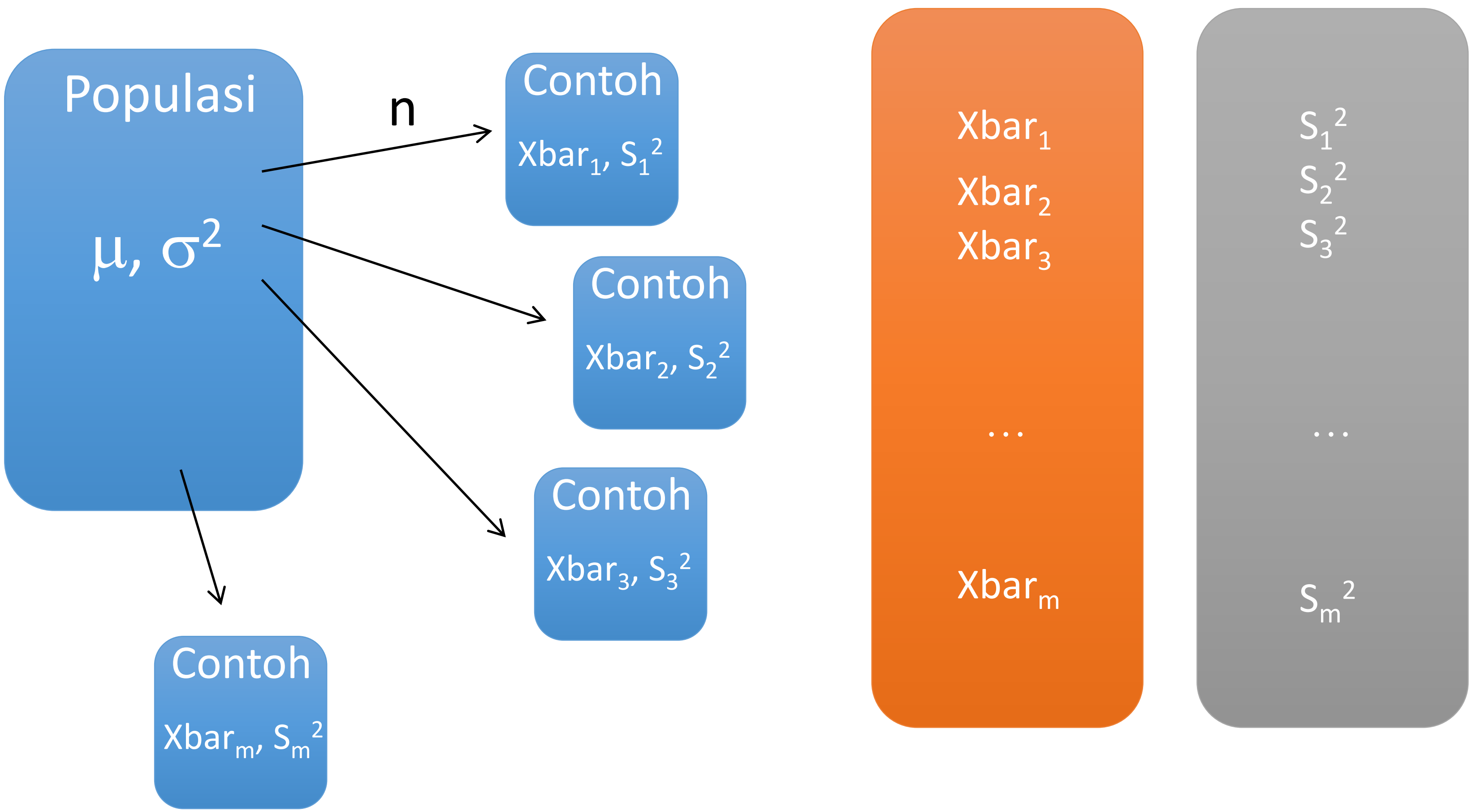
MRJ

Pembuktian Teorema Statistika

Sebaran Percontohan dan Dalil Limit Pusat

STK473 – Praktikum 5

Sebaran Percontohan



→ Sebaran peluang dari suatu statistik tertentu ←

Teorema Limit Pusat

Misalkan $X_1, X_2, \ldots, X_n$ adalah contoh acak dari populasi dengan nilai tengah μ dan ragam σ^2 (sebarannya tidak harus normal). Jika $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ dan n besar (secara matematis n→∞) maka $\bar{X}$ akan menyebar NORMAL dengan nilai tengah μ dan ragam σ^2/n

- Dalil limit pusat sangat berguna sebagai dasar atau alasan mengapa kita sering menggunakan sebaran NORMAL dalam inferensi statistika walaupun sebaran datanya TIDAK NORMAL.

Teorema Limit Pusat

- Algoritme
  - Tentukan ukuran contoh (n)
  - Tentukan sebaran data
  - Ulang k kali
    - Ambil n contoh acak dari sebaran data yang sudah ditentukan
    - Hitung rataannya lalu simpan
  - Periksa sebaran dari k rataan

Simulasi

- Kondisi:
  - Populasi tak terhingga
  - Populasi terhingga
- Faktor:
  - Banyaknya contoh (10, 30, 100)
  - Jenis distribusi (Normal, Exponential, Uniform)

Simulasi

- Populasi Tak Terhingga

	n=10	n=30	n=100
Normal	k<-1000 n<-10 x11<-matrix(rnorm(n*k),k) x11<-apply(x11,1,mean) hist(x11); mean(x11); var(x11)	k<-1000 n<-30 x12<-matrix(rnorm(n*k),k) x12<-apply(x12,1,mean) hist(x12); mean(x12); var(x12)	k<-1000 n<-100 x13<-matrix(rnorm(n*k),k) x13<-apply(x13,1,mean) hist(x13); mean(x13); var(x13)
Eksponensial	k<-1000 n<-10 x21<-matrix(rexp(n*k),k) x21<-apply(x21,1,mean) hist(x21); mean(x21); var(x21)	k<-1000 n<-30 x22<-matrix(rexp(n*k),k) x22<-apply(x22,1,mean) hist(x22) ; mean(x22); var(x22)	k<-1000 n<-100 x23<-matrix(rexp(n*k),k) x23<-apply(x23,1,mean) hist(x23); mean(x23); var(x23)
Seragam	k<-1000 n<-10 x31<-matrix(runif(n*k),k) x31<-apply(x31,1,mean) hist(x31); mean(x31); var(x31)	k<-1000 n<-30 x32<-matrix(runif(n*k),k) x32<-apply(x32,1,mean) hist(x32); mean(x32); var(x32)	k<-1000 n<-100 x33<-matrix(runif(n*k),k) x33<-apply(x33,1,mean) hist(x33); mean(x33); var(x33)

Simulasi

```
windows();par(mfrow=c(3,3))
hist(x11);hist(x12);hist(x13);
hist(x21);hist(x22);hist(x23);
hist(x31);hist(x32);hist(x33);
```

Simulasi

- Populasi Terhingga

```
y1<-rnorm(10000)
y2<-rexp(10000)
y3<-runif(10000)
hist(y1); mean(y1); var(y1)
hist(y2); mean(y2); var(y2)
hist(y3); mean(y3); var(y3)
```

- Populasi Terhingga

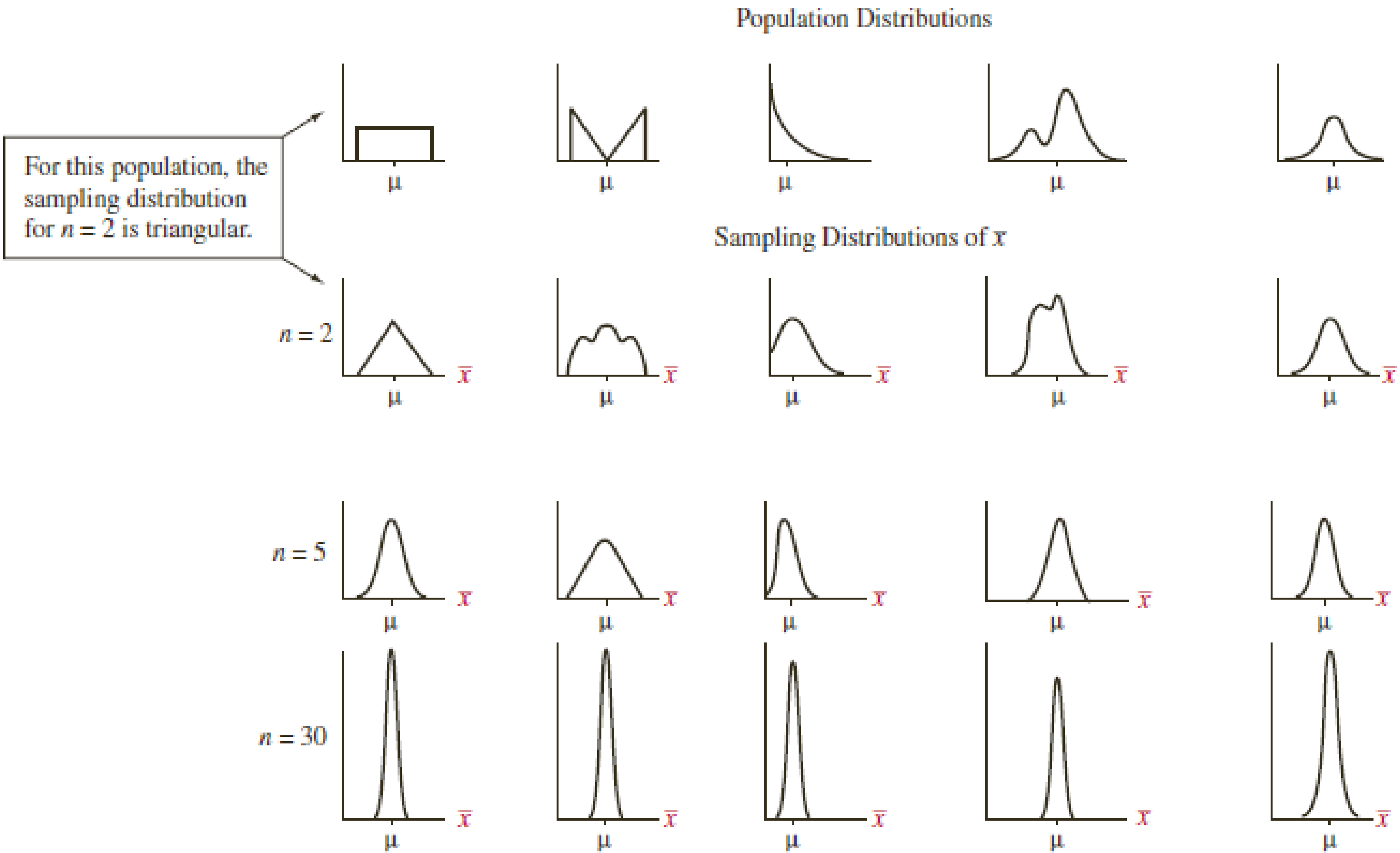
```
k<-1000
n<-10
z11<-matrix(sample(y1,
   n*k),k)
z21<-matrix(sample(y2,
   n*k),k)
z31<-matrix(sample(y3,
   n*k),k)
z11<-apply(z11,1,mean)
z21<-apply(z21,1,mean)
z31<-apply(z31,1,mean)
hist(z11)
mean(z11)
var(z11)
hist(z21)
mean(z21)
var(z21)
hist(z31)
mean(z31)
var(z31)
```

```
k<-1000
n<-30
z12<-matrix(sample(y1,
   n*k),k)
z22<-matrix(sample(y2,
   n*k),k)
z32<-matrix(sample(y3,
   n*k),k)
z12<-apply(z12,1,mean)
z22<-apply(z22,1,mean)
z32<-apply(z32,1,mean)
hist(z12)
mean(z12)
var(z12)
hist(z22)
mean(z22)
var(z22)
hist(z32)
mean(z32)
var(z32)
```

```
k<-1000
n<-100
z13<-matrix(sample(y1,
   n*k),k)
z23<-matrix(sample(y2,
   n*k),k)
z33<-matrix(sample(y3,
   n*k),k)
z13<-apply(z13,1,mean)
z23<-apply(z23,1,mean)
z33<-apply(z33,1,mean)
hist(z13)
mean(z13)
var(z13)
hist(z23)
mean(z23)
var(z23)
hist(z33)
mean(z33)
var(z33)
```

Simulasi

```
windows();par(mfrow=c(3,3))
hist(z11);hist(z12);hist(z13);
hist(z21);hist(z22);hist(z23);
hist(z31);hist(z32);hist(z33);
```


Population Distributions
For this population, the sampling distribution for n = 2 is triangular.
Sampling Distributions of x̄
n = 2
n = 5
n = 30
μ
x̄

thank you!

MRJ

Pembuktian Teorema Statistika

Ketakbiasan dan Selang Kepercayaan

STK473 – Praktikum 6

Ketakbiasan $\overline{x}$ dan s^2 sebagai Penduga μ dan σ^2

- $\bar{x}$ adalah penduga tak bias bagi μ → E($\bar{x}$) = μ
- s^2 adalah penduga tak bias bagi σ^2 → E(s^2) = σ^2

$$s^2 = \sum \frac{(x - \bar{x})^2}{n - 1}$$

Ketakbiasan $\overline{x}$ dan s^2 sebagai Penduga μ dan σ^2

- Algoritme
  - Tentukan sebaran yang akan digunakan, termasuk μ dan σ^2
  - Ulangi k kali
    - Bangkitkan n buah data dari sebaran yang ditentukan
    - Hitung $\bar{x}$ dan s^2
  - Hitung rata-rata dari $\bar{x}$ dan s^2, bandingkan dengan μ dan σ^2

Simulasi

```
n<-10
k<-1000 #ulangan
pop<-rnorm(100,10,sqrt(5)) #populasi terhingga
mean(pop)
var(pop)*(100-1)/100 #fungsi var di R adalah ragam contoh
cth<-matrix(NA,k,n)
for (i in 1:k) cth[i,]<-sample(pop,n)

xbar<-apply(cth,1,mean)
dev<-apply(cth,1,var)
mean(xbar)
mean(dev)
```

Ketakbiasan Penduga Least-Square dalam Regresi Linier

- $b = (X'X)^{-1}X'y$ dapat ditunjukkan sebagai penduga tak bias bagi β jika $E(\varepsilon) = 0$

Ketakbiasan Penduga Least-Square dalam Regresi Linier

- Algoritme
  - Tentukan β_0 dan β_1
  - Ulangi k kali
    - Bangkitkan n buah data X dari sebaran yang tertentu (misal: seragam)
    - Bangkitkan ε yang menyebar normal dengan nilai tengah 0 dan ragam σ_ε^2
    - Hitung b_0 dan b_1
  - Hitung rata-rata dari b_0 dan b_1, bandingkan dengan β_0 dan β_1

Simulasi

```
n<-20 #ukuran contoh
ulangan<-100
beta<-c(8,20) #beta0=8 dan beta1=20
sigmaerror<-3

betaduga<-matrix(NA,ulangan,2)
for (i in 1:ulangan){
  x<-runif(n)*10
  epsilon<-rnorm(n,0,sqrt(sigmaerror))
  X<-cbind(1,x)
  y<-X %*% beta + epsilon
  betaduga[I,]<-solve(t(X) %*% X) %*% (t(X) %*% y)
}
rataanbetaduga<-apply(betaduga,2,mean)
names(rataanbetaduga)<-c("b0","b1")
rataanbetaduga
```

Selang Kepercayaan

- Apa arti dari SK 95%?
  - **SK 95% bagi θ**: Kita percaya 95% bahwa selang a sampai b memuat nilai parameter θ yang sebenarnya
  - **SK 95%**: Jika kita melakukan 100 kali percontohan acak dan setiap percontohan acak dibuat SK-nya, maka dari 100 SK yang terbentuk, ada 95 SK yang mencakup parameter, sisanya sebanyak 5 SK meleset tidak mencakup parameter

Selang Kepercayaan

- Algoritme
  - Tentukan sebaran yang akan digunakan, termasuk μ dan σ^2
  - Ulangi k kali
    - Bangkitkan n buah data dari sebaran yang ditentukan
    - Hitung $\bar{x}$ dan s^2
    - Hitung $\sigma_{\bar{x}}^2$ dan buat selang kepercayaan (1-α)%
  - Hitung proporsi banyaknya selang kepercayaan yang memuat μ, bandingkan dengan (1-α)

Simulasi

```
n<-50
ulangan<-50
alpha<-0.05
mu<-50; std<-10
sampel<-matrix(rnorm(n*ulangan,mu,std),ulangan)
xbar<-apply(sampel,1,mean)
dev<-apply(sampel,1,sd)
SE<-dev/sqrt(n)
z<-qnorm(1-alpha/2)
sk<-(xbar-z*SE<mu & mu<xbar+z*SE)
sum(sk)/ulangan #proporsi banyaknya SK yang memuat mu
matplot(rbind(xbar-z*SE,xbar+z*SE), rbind(1:ulangan,1:ulangan),
        col=ifelse(sk,"blue","red"), type="l",lty=1)
abline(v=mu)
```

thank you!

MRJ

Resampling

STK473 – Praktikum 7

Simulasi Monte Carlo

- Simulasi yang memanfaatkan informasi mengenai sebaran data yang **diketahui (dihipotesiskan, dianggap tahu)** dengan pasti.

- Pendugaan Peluang

$$P(Y > a) \; atau \; P(Y < a)$$

dengan $Y = g(x),\ X \sim f(x)$

  - Menghitung peluang dari suatu peubah acak
  - Pengujian hipotesis

Simulasi Monte Carlo

- Pengujian Hipotesis

  Pada kasus contoh acak yang diambil berasal dari populasi yang menyebar normal, nilai statistik uji t akan menyebar $t_{db=(n-1)}$, sementara untuk populasi dengan sebaran lainnya, perlu dikaji ulang sebaran dari nilai statistik uji t

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

  Simulasi dilakukan untuk melihat hasil dari sebaran hipotetik populasi ketika menghitung nilai statistik uji t

Ilustrasi 1

Misal X~N(0,1), jika diketahui Y=2X, berapa P(Y>2)?

Algoritme :

1. Bangkitkan X~N(0,1)
2. Hitung Y=2X
3. Ulangi langkah 1 dan 2 sebanyak 100000 kali
4. Hitung presentasi nilai Y>2

Simulasi

```
n<-100000
x<-rnorm(n)
y<-2*x
y1<-ifelse(y>2,1,0)
mean(y1) #P(Y>2) secara empirik
```

$$X \sim N(0,1)\,;Y = 2X$$
$$E(Y) = 2E(X) = 0\,;\,V(Y) = 2^2V(X) = 4$$
$$Y \sim N(0,4)$$

```
pnorm(2,0,2,lower.tail=F) #P(Y>2) dari sebaran hipotetik
```

Ilustrasi 2

- Misal peubah acak $X \sim \exp(\lambda)$, kita ingin menguji hipotesis

$$H_0 : \lambda = 2$$
$$H_1 : \lambda = 4$$

Jika kita menolak H_0 ketika $x > 1$.

Hitung α dan kuasa ujinya !

Jawaban

- $\alpha = P(tolak\ H_0 |\ H_0\ benar)$
  $$= P(x > 1 | \lambda = 2)$$
  $$= \int_1^{\infty} \frac{1}{2} e^{-x/2}\, dx = 0.135$$
- Kuasa uji $= 1 - \beta$
  $$= P(tolak\ H_0 |\ H_0\ salah)$$
  $$= P(x > 1 | \lambda = 4)$$
  $$= \int_1^{\infty} \frac{1}{4} e^{-x/4}\, dx = 0.018$$

Simulasi

```
lambda<-2
N<-100000
k<-1
x<-rexp(N,lambda)
y<-ifelse(x>k,1,0)
hasil<-mean(y) #alpha

lambda<-4
N<-100000
k<-1
x<-rexp(N,lambda)
y<-ifelse(x>k,1,0)
hasil<-mean(y) #kuasa uji
```

Ilustrasi 3

- Berikut tersedia data dari sebaran eksponensial:

1.68760	0.03120
0.03037	0.61068
0.91673	0.63169
1.34939	2.99986
0.08164	2.70955

- $H_0: \lambda = 1$
- Apa kesimpulan yang bisa diambil?

Ilustrasi 3

- Algoritme :

1. Hitung $t_{data} = \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|$ dari data
2. Bangkitkan 10 contoh acak~$\exp(\lambda = 1)$
3. Hitung $t_{dist} = \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right|$ dari sebaran hipotetik
4. Ulangi langkah 2 dan 3 sebanyak 10000 kali
5. Hitung nilai-p=$P(t_{dist}>t_{data})$

Simulasi

```
lambda<-1
k<-10000
data1<-c(1.6876,0.03037,0.91673,1.34939,0.08164,
         0.0312,0.61068,0.63169,2.99986,2.70955)
n<-length(data1)
m<-mean(data1)
s<-sd(data1)
tdata<-abs((m-1/lambda)/(s/sqrt(n)))

data2<-matrix(rexp(n*k,lambda),k)
m1<-apply(data2,1,mean)
s1<-apply(data2,1,sd)
tdist<-abs((m1-1/lambda)/(s1/sqrt(n)))

y<-ifelse(tdist>tdata,1,0)
pvalue<-mean(y)
kesimpulan<-ifelse(pvalue<0.05,"Tolak H0","Tak Tolak H0")
pvalue;kesimpulan
```

Simulasi

- Cara lain menghitung nilai-p. Ulangi 9999 kali. Gabungkan nilai t dari data asli dengan nilai t dari simulasi. Urutkan nilai t, dan perhatikan pada persentil keberapa nilai t hitung yang kita miliki.

```
y1<-c(tdata,tdist[-k])
y1<-sort(y1)
pvalue1<-1-which(y1==tdata)/k
kesimpulan1<-ifelse(pvalue1<0.05,"Tolak H0",
    "Tak Tolak H0")
pvalue1;kesimpulan1
```

Ilustrasi 3

- Bagaimana jika contoh acak tersebut diasumsikan berasal dari sebaran normal, lalu hipotesis yang ingin diuji adalah:

- $H_0: \mu = 2$

- Kesimpulan apa yang bisa diambil? Dari program sebelumnya, apa saja yang perlu diubah?
- Hint: untuk kesederhanaan gunakan $\sigma^2 = 1$

Case

Suppose we interest to estimate the ratio of Y/X, or Y/(X+Y), etc

So we decide to take a sample of X and Y

The question is "What is the estimated value of the ratio" and also "How we can estimate the SE of estimated ratio" ?

Case

The sample

$X = \{x_1, x_2, \ldots, x_n\}$

$Y = \{y_1, y_2, \ldots, y_n\}$

Ratio Y/X

Estimated by : $r = \frac{\bar{y}}{\bar{x}}$

Estimated SE : $\sqrt{\hat{V}(r)} = \left(1 - \frac{n}{N}\right)\left(\frac{1}{{\mu_X}^2}\right)\frac{{s_r}^2}{n}$

Ratio Y/(X+Y)

?

Case

Ratio Y/(X+Y)

Estimator : $r = \frac{\bar{y}}{\bar{x}+\bar{y}}$

Estimator of SE ?

→ **Resampling** For estimating SE

Jackknife

- We have a sample $y = (y_1, \dots, y_n)$ to estimate θ with the estimator $\hat{\theta} = f(y)$
- Target : estimate standard error of $\hat{\theta}$
- The leave-one-out observation samples
$$y_{(i)} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$$
for $i = 1, \dots, n$ are called **jacknife samples**
- Jacknife statistics are $\hat{\theta}_{(i)} = f(y_{(i)})$

Jacknife estimators

$$\hat{\theta}_{jack} = n\hat{\theta} - (n-1)\hat{\theta}_{(.)} \qquad \hat{\theta}_{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}$$

$$V_{jack}(\hat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(.)}\right)^2$$

Example 1

Data

Branch store	Income	
	Item1	Item2
A	1363	1087
B	670	571
C	761	518
D	746	612
E	991	770
F	798	655

Estimate the ratio of Item1/total item using jacknife!

Simulation

Algorithm:

1. Compute $\hat{\theta} = r = \frac{\bar{y}}{\bar{x}+\bar{y}}$
2. Take jacknife sample $x_{(i)}$ & $y_{(i)}$; $i = 1, \dots, n$
3. Compute jacknife estimator

$$\hat{\theta}_{jack} = n\hat{\theta} - (n-1)\hat{\theta}_{(.)} = n\frac{\bar{y}}{\bar{x}+\bar{y}} - \frac{(n-1)}{n}\sum_{i=1}^{n}\frac{\bar{y}_{(i)}}{\bar{x}_{(i)}+\bar{y}_{(i)}}$$

$$V_{jack}(\hat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{\bar{y}_{(i)}}{\bar{x}_{(i)}+\bar{y}_{(i)}} - \frac{1}{n}\sum_{i=1}^{n}\frac{\bar{y}_{(i)}}{\bar{x}_{(i)}+\bar{y}_{(i)}}\right)^2$$

Simulation

```
store<-LETTERS[1:6]
item1<-c(1363,670,761,746,991,798)
item2<-c(1087,571,518,612,770,655)
data1<-data.frame(store,item1,item2)
n<-nrow(data1)
teta_hat<-mean(data1$item1)/
     (mean(data1$item1)+mean(data1$item2))
y<-matrix(NA,n,n-1)
x<-matrix(NA,n,n-1)
for (i in 1:n) {
  y[i,]<-data1$item1[-i]
  x[i,]<-data1$item2[-i]
}
ybar<-apply(y,1,mean)
xbar<-apply(x,1,mean)
teta_i<-ybar/(xbar+ybar)
teta_jk<-n*teta_hat-(n-1)*mean(teta_i)
var_jk<-(n-1)/n*(sum(teta_i^2)-(n*mean(teta_i)^2))
se_jk<-sqrt(var_jk)
```

Bootstrap

- We have a sample $y = (y_1, \dots, y_n)$ to estimate θ with the estimator $\hat{\theta} = f(y)$
- Steps
  - Repeatedly simulate sample of size n from $y \rightarrow {y^b}_{(i)}$
  - Compute statistic of interest ${\hat{\theta}^b}_{(i)} = f({y^b}_{(i)})$
  - Study behavior of statistic over N repetitions

Bootstrap estimators

$$\hat{\theta}_b = \frac{1}{N} \sum_{i=1}^{N} {\hat{\theta}^b}_{(i)} \qquad V_b(\hat{\theta}) = \frac{1}{N-1} \sum_{i=1}^{N} \left({\hat{\theta}^b}_{(i)} - \hat{\theta}_b\right)^2$$

Example 2

Data

Branch store	Income	
	Item1	Item2
A	1363	1087
B	670	571
C	761	518
D	746	612
E	991	770
F	798	655

Estimate the ratio of Item1/total item using bootstrap!

Simulation

Algorithm:

1. Repeatedly simulate sample of size n from $x \rightarrow x^{b}{}_{(i)}$ & $y \rightarrow y^{b}{}_{(i)}$
2. Compute statistic $\hat{\theta}^{b}{}_{(i)} = \frac{\bar{y}^{b}{}_{(i)}}{\bar{x}^{b}{}_{(i)} + \bar{y}^{b}{}_{(i)}}$
3. Compute bootstrap estimator

$$\hat{\theta}_b = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}^{b}{}_{(i)} = \frac{1}{N}\sum_{i=1}^{N} \frac{\bar{y}^{b}{}_{(i)}}{\bar{x}^{b}{}_{(i)} + \bar{y}^{b}{}_{(i)}}$$

$$V_b(\hat{\theta}) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\theta}^{b}{}_{(i)} - \hat{\theta}_b\right)^2$$

Simulation

```
store<-LETTERS[1:6]
item1<-c(1363,670,761,746,991,798)
item2<-c(1087,571,518,612,770,655)
data1<-data.frame(store,item1,item2)
n<-nrow(data1)
b<-1000
y<-matrix(sample(data1$item1,n*b,replace=T),b)
x<-matrix(sample(data1$item2,n*b,replace=T),b)
ybar<-apply(y,1,mean)
xbar<-apply(x,1,mean)
teta_i<-ybar/(xbar+ybar)
teta_b<-mean(teta_i)
var_b<-(sum(teta_i^2)-(b*teta_b^2))/(b-1)
se_b<-sqrt(var_b)
```

thank you!

MRJ