

Tugas Praktikum - Pemodelan Risiko 2

Anggota kelompok:

- | | |
|--------------------------------|------------|
| • Andreas Jorgan | 2106725305 |
| • Laurensius Fabianus Steven | 1906352142 |
| • Priskila Lintang Marshaulina | 1906376003 |
| • Nazhira Ghaizani | 1906357805 |

Soal

Misalkan anda diberikan sampel dari 20 *ground-up loss* dalam satuan juta dollar berikut:

0,003; 0,012; 0,180; 0,253; 0,394; 0,430; 0,491; 0,743; 1,066; 1,126; 1,303; 1,508; 1,740; 4,757;
5,376; 5,557; 7,236; 7,465; 8,054; 14,938.

Kerjakan poin 1-3 di bawah ini menggunakan R serta buatlah video simulasi R tersebut.

1. Estimasi parameter distribusi uniform, eksponensial, dan lognormal dengan menggunakan metode MLE untuk sampel di atas.
2. Gambarkan plot cdf empiris serta cdf distribusi uniform, eksponensial, dan lognormal dari sampel di atas.
3. Ujilah kecocokan model distribusi uniform, eksponensial, dan lognormal terhadap sampel di atas menggunakan uji Kolmogorov-Smirnov untuk taraf signifikansi 5%.

Tautan video pembahasan

Pembahasan Jawaban:

Untuk menjawab soal di atas digunakan *library* EnvStats dan bbmle dengan sintaks berikut.

```
library(EnvStats)  
library(bbmle)
```

Setelah itu, data akan dimasukan dalam format vektor/tuple yang disimpan pada variabel `data`.

```
data <- c(0.003, 0.012, 0.180, 0.253, 0.394, |  
          0.430, 0.491, 0.743, 1.066, 1.126, |  
          1.303, 1.508, 1.740, 4.757, 5.376, |  
          5.557, 7.236, 7.465, 8.054, 14.938)
```

Nomor 1

Estimasi parameter menggunakan metode MLE.

a. Distribusi Uniform

Estimasi parameter dilakukan menggunakan fungsi `eunif` sesuai sintaks berikut:

```
parunif <- eunif(data, method = 'mle')
parunif$parameters
```

hasil penaksiran nilai parameter adalah sebagai berikut:

```
> parunif$parameters
      min      max
0.003 14.938
```

hasil taksiran ini akan disimpan sebagai variabel berikut untuk keperluan plot dan uji statistik.

```
unif_a <- parunif$parameters['min']
unif_b <- parunif$parameters['max']
```

b. Distribusi Eksponensial

Estimasi parameter dilakukan menggunakan fungsi `eexp` sesuai sintaks berikut:

```
parexp <- eexp(data, method = 'mle')
parexp$parameters
```

hasil penaksiran nilai parameter adalah sebagai berikut:

```
> parexp$parameters
      rate
0.3193256
```

hasil taksiran ini akan disimpan sebagai variabel berikut untuk keperluan plot dan uji statistik.

```
exp_lmd <- parexp$parameters['rate']
```

c. Distribusi Lognormal

Estimasi parameter dilakukan menggunakan fungsi `elnorm` sesuai sintaks berikut:

```
parlnorm <- elnorm(data, method = 'mle')
parlnorm$parameters
```

hasil penaksiran nilai parameter adalah sebagai berikut:

```
> parlnorm$parameters
      meanlog      sdlog
-0.04784929  2.10538383
```

hasil taksiran ini akan disimpan sebagai variabel berikut untuk keperluan plot dan uji statistik.

```
lnorm_mu <- parlnorm$parameters['meanlog']
lnorm_sg <- parlnorm$parameters['sdlog']
```

Nomor 2

Gambar plot CDF

a. CDF Empiris

Plot CDF empiris didapat menggunakan fungsi `ecdf` sesuai sintaks berikut.

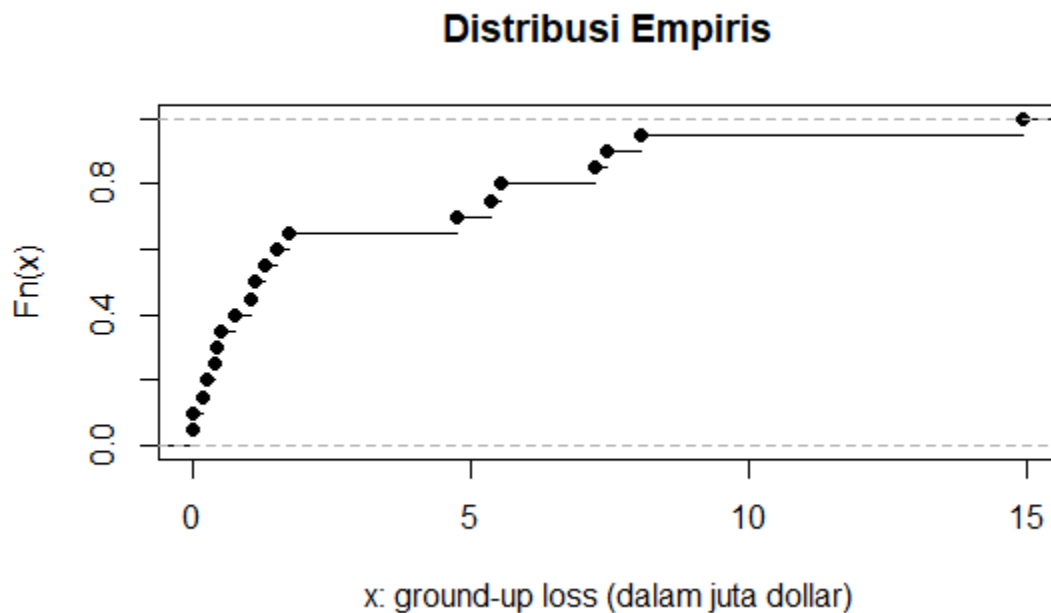
```
plot(ecdf(data), xlim = c(0,15), ylim = c(0,1),  
     xlab = 'x: ground-up loss (dalam juta dollar)',  
     main = '')  
plotemp <- recordPlot()  
title('Distribusi Empiris')
```

Fungsi plot memiliki beberapa parameter selain fungsi `ecdf`:

- `xlim` dan `ylim` untuk mengatur rentang nilai horizontal dan vertikal pada plot
- `xlab` dan `ylab` sebagai label sumbu vertikal. Dalam hal ini sumbu horizontal menyatakan ground-up loss dalam satuan juta dollar
- `main` sebagai judul plot

Hasil plot ini akan direkam dengan `recordPlot()` ke variabel `plotemp` untuk demonstrasi di bagian selanjutnya.

Hasil plot CDF distribusi empiris adalah sebagai berikut.



b. CDF Distribusi Uniform

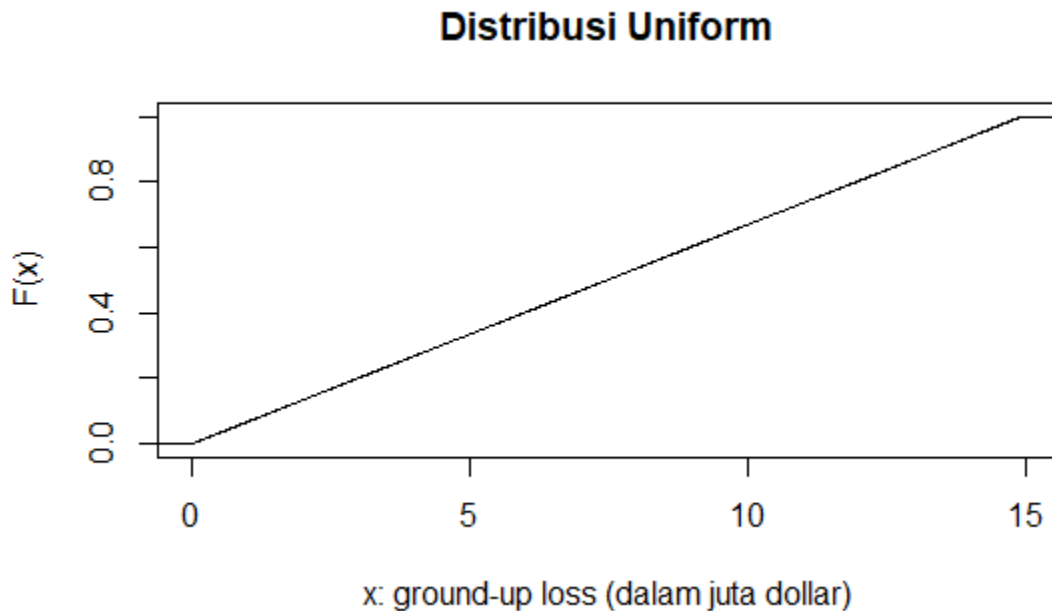
Plot CDF distribusi uniform didapat menggunakan fungsi **punif** sesuai sintaks berikut.

```
x <- seq(-1,16,0.02)
plot(x, punif(x, unif_a, unif_b, lower.tail = TRUE), type = 'l',
      xlim = c(0,15), ylim = c(0,1),
      xlab = 'x: ground-up loss (dalam juta dollar)',
      ylab = 'F(x)',
      main = 'Distribusi Uniform')
```

Fungsi **seq** digunakan untuk mendapatkan sampel berjarak seragam untuk mensimulasikan CDF distribusi parametrik yang disimpan pada variabel **x**.

Fungsi **plot** memiliki beberapa parameter selain **x** dan **punif** yang baru, yaitu **type** untuk menentukan jenis plot. Dalam hal ini plot yang diinginkan adalah plot garis dengan tipe **'l'** (*line*).

Hasil plot CDF distribusi uniform adalah sebagai berikut.

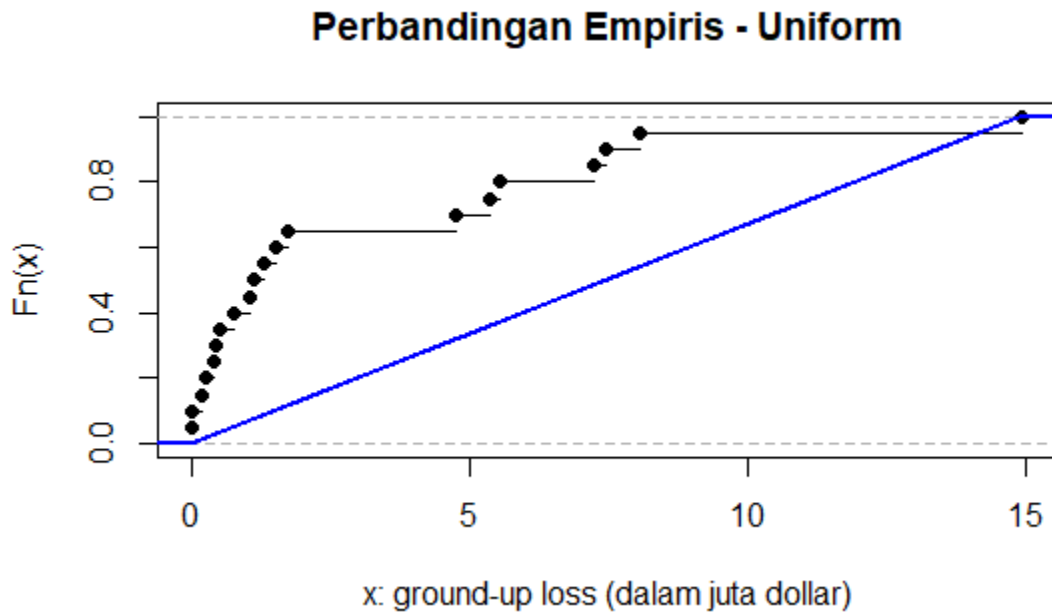


c. Perbandingan CDF Distribusi Uniform dengan CDF Empiris

Untuk membandingkan apakah distribusi uniform memodelkan data dengan baik, secara visual akan dibandingkan plot CDF distribusi uniform dengan CDF empiris. Plot empiris yang tersimpan pada `plotemp` dikombinasikan dengan plot uniform dengan fungsi `lines`.

```
plotemp  
lines(x, punif(x, unif_a, unif_b, lower.tail = TRUE),  
      col = 'blue', lwd = 2)  
title('Perbandingan Empiris - Uniform')
```

Berikut plot perbandingan CDF distribusi uniform dengan CDF empiris.

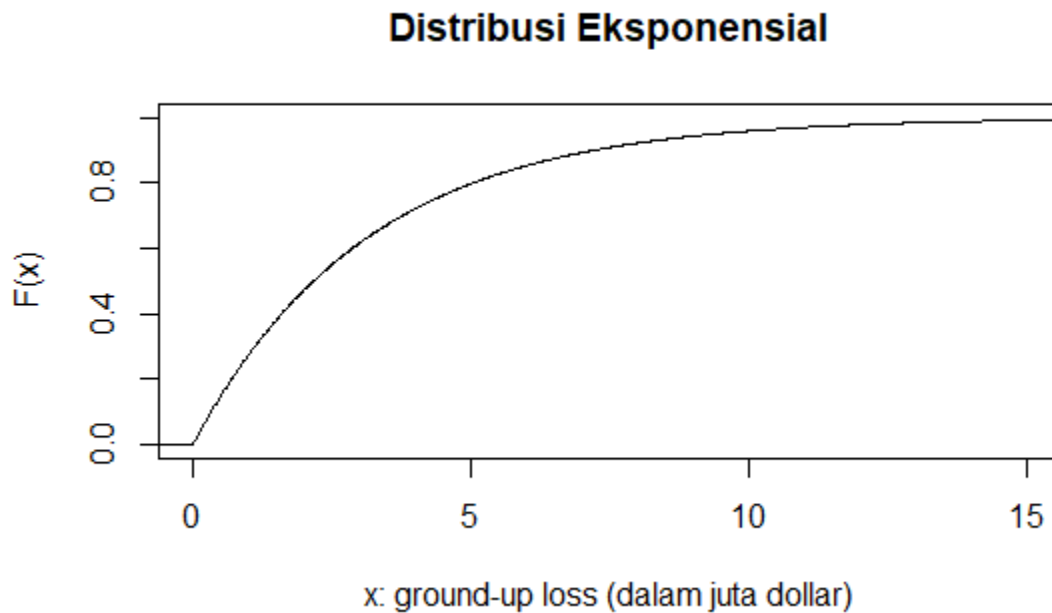


d. CDF Distribusi Eksponensial

Plot CDF distribusi eksponensial didapat menggunakan fungsi `pexp` sesuai sintaks berikut.

```
plot(x, pexp(x, exp_lmd, lower.tail = TRUE), type = 'l',  
     xlim = c(0,15), ylim = c(0,1),  
     xlab = 'x: ground-up loss (dalam juta dollar)',  
     ylab = 'F(x)',  
     main = 'Distribusi Eksponensial')
```

Hasil plot CDF distribusi eksponensial adalah sebagai berikut.

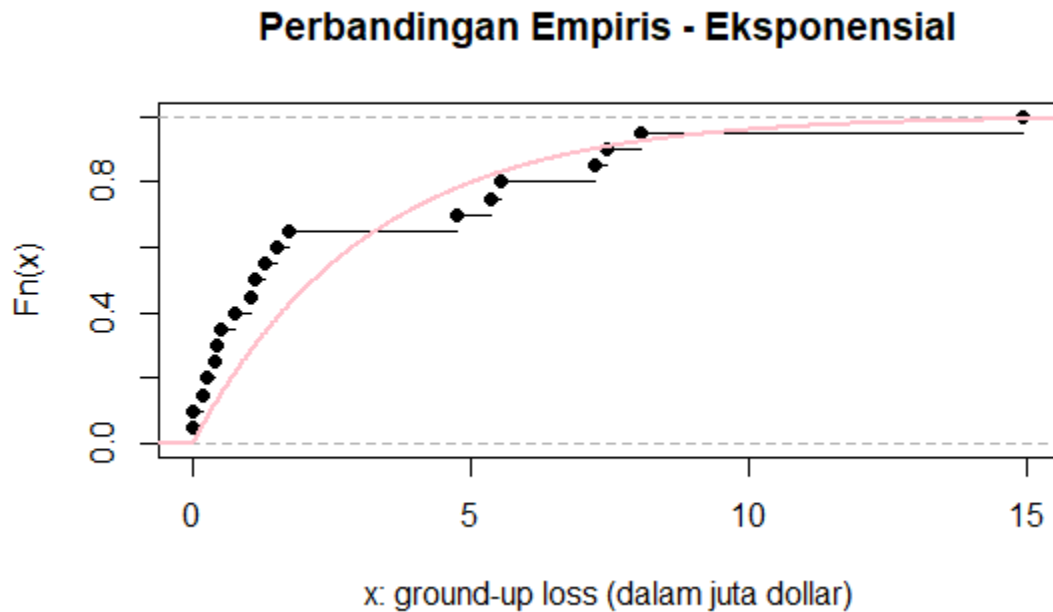


e. Perbandingan CDF Distribusi Eksponensial dengan CDF Empiris

Untuk membandingkan apakah distribusi eksponensial memodelkan data dengan baik, secara visual akan dibandingkan plot CDF distribusi eksponensial dengan CDF empiris. Plot empiris yang tersimpan pada `plotemp` dikombinasikan dengan plot eksponensial dengan fungsi `lines`.

```
plotemp  
lines(x, pexp(x, exp_lmd, lower.tail = TRUE), col = 'pink', lwd = 2)  
title('Perbandingan Empiris - Eksponensial')
```

Berikut plot perbandingan CDF distribusi eksponensial dengan CDF empiris.

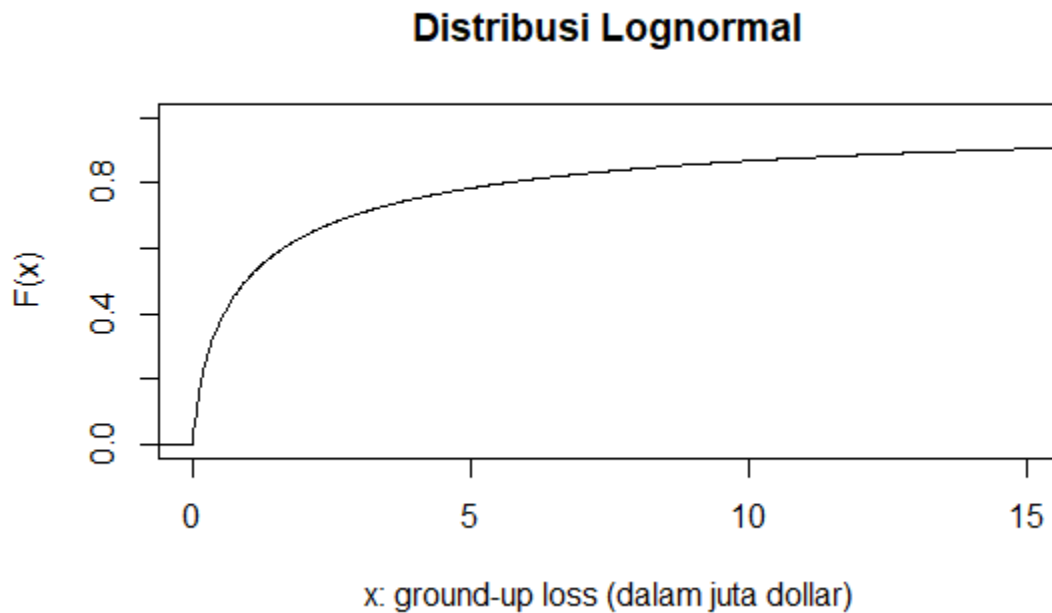


f. CDF Distribusi Lognormal

Plot CDF distribusi lognormal didapat menggunakan fungsi `plnorm` sesuai sintaks berikut.

```
plot(x, plnorm(x, lnorm_mu, lnorm_sg, lower.tail = TRUE), type = 'l',  
      xlim = c(0,15), ylim = c(0,1),  
      xlab = 'x: ground-up loss (dalam juta dollar)',  
      ylab = 'F(x)',  
      main = 'Distribusi Lognormal')
```

Hasil plot CDF distribusi lognormal adalah sebagai berikut.

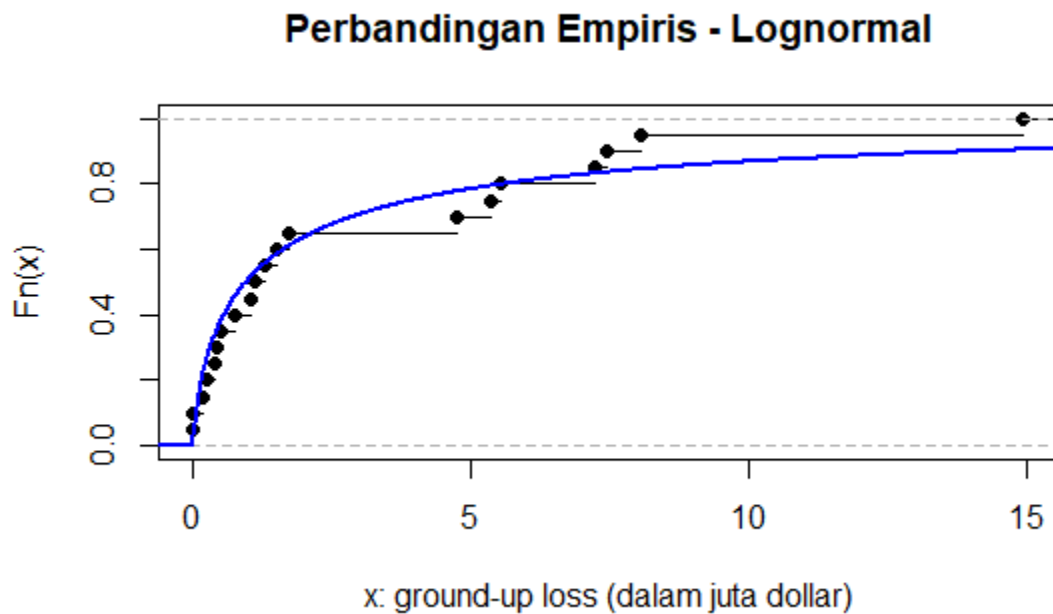


g. Perbandingan CDF Distribusi Lognormal dengan CDF Empiris

Untuk membandingkan apakah distribusi lognormal memodelkan data dengan baik, secara visual akan dibandingkan plot CDF distribusi lognormal dengan CDF empiris. Plot empiris yang tersimpan pada `plotemp` dikombinasikan dengan plot lognormal dengan fungsi `lines`.

```
plotemp  
lines(x, plnorm(x, lnorm_mu, lnorm_sg, lower.tail = TRUE),  
      col = 'blue', lwd = 2)  
title('Perbandingan Empiris - Lognormal')
```

Berikut plot perbandingan CDF distribusi eksponensial dengan CDF empiris.



Nomor 3

Uji Kolmogorov-Smirnov untuk menentukan kelayakan distribusi parametrik untuk memodelkan sampel *ground-up loss* dengan tingkat signifikansi $\alpha = 5\%$.

a. Distribusi Uniform

Uji Kolmogorov-Smirnov dilakukan dengan fungsi `ks.test` sesuai sintaks berikut.

```
ks.test(data, 'punif', unif_a, unif_b)
```

Fungsi `ks.test` mengambil input data sampel bersama dengan distribusi parametrik yang akan dibandingkan. Dalam hal ini distribusi tersebut adalah distribusi uniform yang diwakilkan oleh `punif` dengan parameter yang didapat pada soal nomor 1.

Hasil uji adalah sebagai berikut.

```
> ks.test(data, 'punif', unif_a, unif_b)

      One-sample Kolmogorov-Smirnov test

data:  data
D = 0.5337, p-value = 7.548e-06
alternative hypothesis: two-sided
```

Nilai *p-value* uji adalah 7.548×10^{-6} yang lebih kecil dari tingkat signifikansi α . Oleh karena itu, H_0 ditolak, yang menyimpulkan bahwa distribusi uniform tidak dapat memodelkan data dengan baik.

b. Distribusi Eksponensial

Uji Kolmogorov-Smirnov dilakukan dengan fungsi `ks.test` sesuai sintaks berikut.

```
ks.test(data, 'pexp', exp_lmd)
```

Distribusi eksponensial yang diwakilkan oleh `pexp` dengan parameter yang didapat pada soal nomor 1.

Hasil uji adalah sebagai berikut.

```
> ks.test(data, 'pexp', exp_lmd)

      One-sample Kolmogorov-Smirnov test

data:  data
D = 0.22371, p-value = 0.2321
alternative hypothesis: two-sided
```

Nilai *p-value* uji adalah 0.231 yang lebih besar dari tingkat signifikansi α . Oleh karena itu, H_0 tidak ditolak, yang menyimpulkan bahwa distribusi eksponensial dapat memodelkan data dengan baik.

c. Distribusi Lognormal

Uji Kolmogorov-Smirnov dilakukan dengan fungsi `ks.test` sesuai sintaks berikut.

```
ks.test(data, 'plnorm', lnorm_mu, lnorm_sg)
```

Distribusi eksponensial yang diwakilkan oleh `plnorm` dengan parameter yang didapat pada soal nomor 1.

Hasil uji adalah sebagai berikut.

```
> ks.test(data, 'plnorm', lnorm_mu, lnorm_sg)
```

```
One-sample Kolmogorov-Smirnov test
```

```
data: data  
D = 0.13737, p-value = 0.7965  
alternative hypothesis: two-sided
```

Nilai *p-value* uji adalah 0.7965 yang lebih besar dari tingkat signifikansi α . Oleh karena itu, H_0 tidak ditolak, yang menyimpulkan bahwa distribusi lognormal dapat memodelkan data dengan baik.

Bonus

Hasil dari nomor 3 menyimpulkan bahwa distribusi eksponensial dan distribusi lognormal adalah kandidat yang baik untuk memodelkan sampel *ground-up loss*. Untuk menentukan model mana yang lebih baik dibutuhkan perbandingan kriteria informasi. Dalam hal ini akan digunakan kriteria informasi Akaike (AIC) untuk membandingkan kedua model.

Kriteria informasi Akaike (AIC) dapat digunakan pada R dengan definisi berikut:

$$AIC = -2\ln L + 2p$$

dengan L menyatakan *likelihood* dari model dan p menyatakan banyaknya parameter. AIC dapat dipandang sebagai penalti model yang akan semakin baik untuk nilai AIC yang semakin rendah.

Perhitungan AIC akan menggunakan dua *library* berikut.

```
library(survival)
library(flexsurv)
```

Nilai AIC untuk distribusi eksponensial dihitung dengan bantuan fungsi `flexsurvreg` sebagai berikut.

```
fitexp <- flexsurvreg(Surv(data)~1,
                     dist = 'exponential')
fitexp$AIC
> fitexp$AIC
[1] 87.66176
```

Hasil nilai AIC distribusi eksponensial adalah 87.66176

Dengan cara serupa AIC distribusi lognormal dihitung sebagai berikut.

```
fitlnorm <- flexsurvreg(Surv(data)~1,
                       dist = 'lognormal')
fitlnorm$AIC
> fitlnorm$AIC
[1] 88.62348
```

Hasil nilai AIC distribusi lognormal adalah 88.62348

Dapat dilihat bahwa AIC distribusi eksponensial lebih rendah, sehingga distribusi yang paling tepat untuk memodelkan sampel *ground-up loss* adalah distribusi eksponensial.