

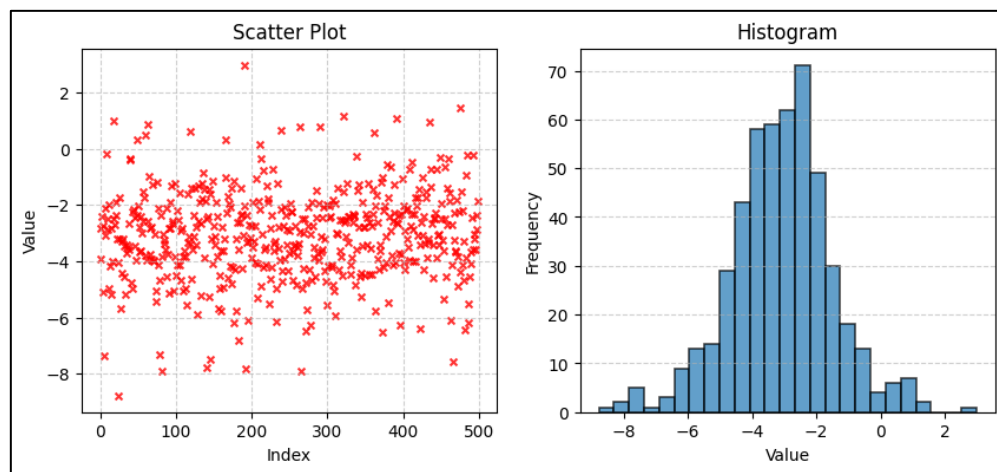
Nama : Mutiara Khairunnisa

NIM : 23/517062/PA/22149

## Tugas 2 Geostatistika Pra-UTS

### A. Analisis Data

Data yang digunakan dalam Tugas 2 Pra-UTS merupakan data yang berisi 500 angka yang dibuat secara *random* melalui *website* <https://octave-online.net/> dengan memasukkan 5digit terakhir NIM pada kode yang telah dibuat. Hasilnya diperoleh dataset dengan sebaran angka seperti gambar berikut:



Gambar 1. *Scatter Plot* dan Histogram dari Dataset

Seperti yang dapat diamati pada *scatter plot* di atas, data tersebar dari -9 hingga 3 dengan banyak angka yang tersebar di sekitar -5 hingga -1. Histogram dari dataset tersebut tampak seperti distribusi normal yang mana histogram tidak mengalami *skewness* ke kanan atau kiri.

### B. Analisis Pengolahan Data dan Hasilnya

Untuk pengolahan data, yaitu *fitting* fungsi probabilitas dengan manual menggunakan *python* dengan memanfaatkan *library* NumPy untuk pengolahan numerik dan Matplotlib untuk membuat plot. Untuk *fitting* distribusi probabilitas dengan histogram yang dimiliki, perlu menghitung PDF (*Probability Density Function*) dan CDF (*Cumulative Distribution Function*) terlebih dahulu sesuai dengan jenis parameter distribusi probabilitas yang digunakan. Sebelum itu, PDF merupakan fungsi yang menggambarkan kemungkinan nilai yang muncul dalam distribusi probabilitas. Sedangkan CDF merupakan fungsi yang menunjukkan total peluang (kumulatif) sampai titik tertentu yang dapat dihitung dengan mengintegrasikan PDF. Macam-macam fungsi distribusi probabilitas yang digunakan diantaranya sebagai berikut.

## 1. Logistic Distribution.

Merupakan jenis distribusi probabilitas kontinu yang kerap digunakan untuk pemodelan statistik, seperti regresi logistic, pemodelan pertumbuhan populasi, maupun pemodelan data ekonomi dan sosial. Parameter yang digunakan dalam fungsi ini yaitu  $\mu$  (*mean*) dan  $s$  (skala). Nilai PDF dan CDF dapat diperoleh dengan persamaan berikut:

$$PDF = f(x; \mu, s) = \frac{e^{-\frac{x-\mu}{s}}}{s \left(1 + e^{-\frac{(x-\mu)}{s}}\right)^2}$$
$$CDF = F(x; \mu, s) = \frac{1}{1 + \frac{e^{-(x-\mu)}}{s}}$$

Dengan:

- $\mu$  = Parameter lokasi atau *mean*
- $s$  = Parameter skala yang diperoleh dari  $\frac{\sigma}{\pi/\sqrt{3}}$
- $e$  = Bilangan Euler

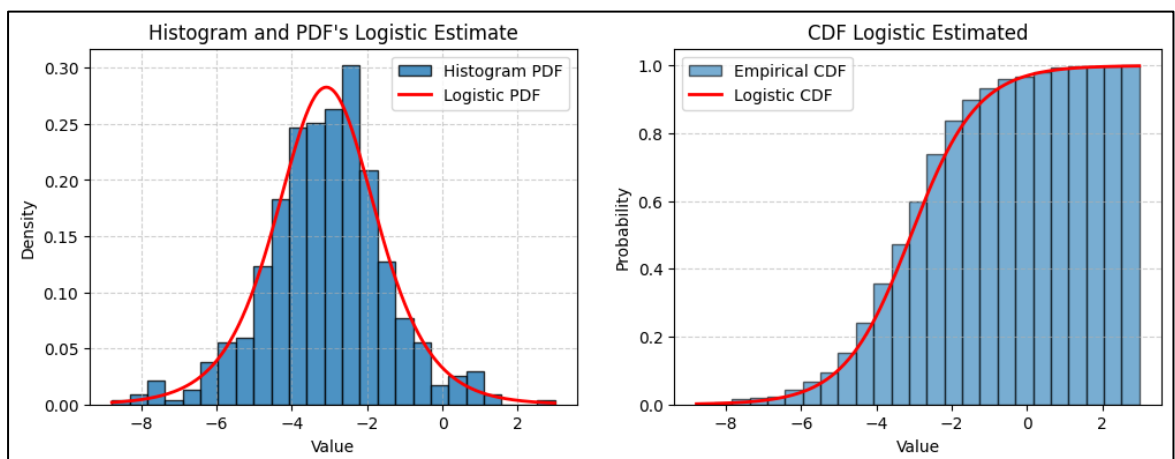
Sebelum perhitungan PDF, dilakukan perhitungan parameter terlebih dahulu, yaitu *mean* (rata-rata) dari dataset dan  $s$  yang memerlukan nilai standar deviasi. Setelah itu, persamaan dapat dimasukkan pada *Python* dengan kode di bawah akan menghasilkan grafik PDF dan CDF seperti gambar di bawah ini.

```
mu = mean
s = std_dev * np.sqrt(3) / np.pi
x = np.linspace(data.min(), data.max(), 500)

pdf_logistic = np.exp(-(x-mu) / s) / (s * (1 + np.exp(-(x - mu) / s)) ** 2)
cdf_logistic = 1 / (1 + np.exp(-(x - mu) / s))

# plot hist and scatter
fig, axes = plt.subplots(1, 2, figsize=(12, 4))
```

Gambar 2. Kode untuk Distribusi Probabilitas Logistik



Gambar 3. CDF dan PDF dari Distribusi Probabilitas Logistik

Dari hasil yang diperoleh tersebut, tampak bahwasanya PDF Logistik sangat sesuai dengan dataset yang dimiliki. Kemudian, pada CDF Logistik, juga berimpitan dengan CDF empiris, sehingga CDF Logistik sangat cocok untuk diterapkan pada dataset ini. Alhasil, distribusi probabilitas Logistik merupakan salah satu distribusi probabilitas yang dapat diterapkan pada dataset ini.

## 2. Laplace Distribution

Distribusi probabilitas ini merupakan distribusi yang kerap digunakan untuk pemrosesan sinyal, statistik, dan model ekonomi. Distribusi ini juga kerap dikenal sebagai *double exponential distribution* karena memiliki bentuk dua distribusi eksponensial yang simetris. Distribusi ini memiliki parameter  $\mu$  (*pusat lokasi*) dan  $b$  (*skala*). Nilai CDF dan PDF dapat diperoleh dari persamaan berikut:

$$PDF = f(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right)$$

$$CDF = F(x) = \begin{cases} \frac{1}{2} \exp\left(\frac{x - \mu}{b}\right), & x < 0 \\ 1 - \frac{1}{2} \exp\left(-\frac{x - \mu}{b}\right), & x \geq 0 \end{cases}$$

Dengan:

$\mu$  = parameter lokasi pusat (*mean*)

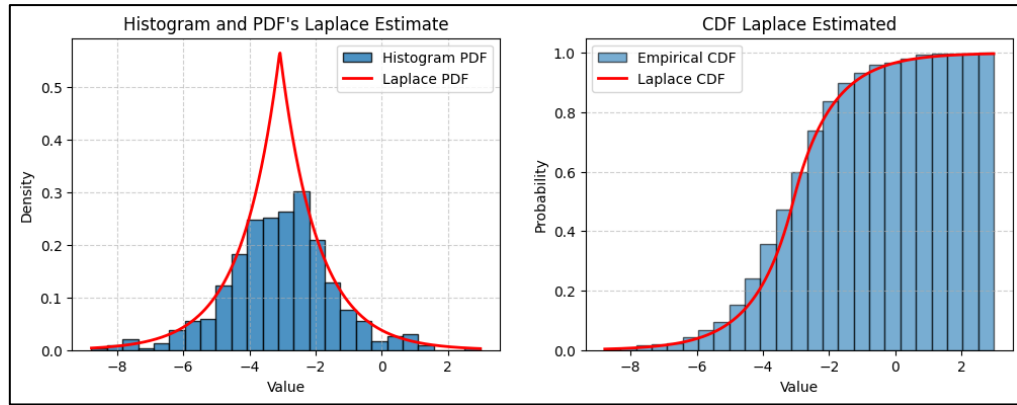
$b$  = parameter skala yang diperoleh dari  $\frac{\sigma}{\sqrt{2}}$ .

Sebelum perhitungan PDF dan CDF diperlukan perhitungan parameter tertentu, seperti *mean* (rata-rata) dari dataset dan  $b$  (skala) yang diperoleh dari standar deviasi dibagi  $\sqrt{2}$ . Setelah itu, perhitungan CDF dan PDF dapat dilakukan dengan memasukkan kode di bawah pada *Python* yang akan menghasilkan grafik PDF dan CDF seperti gambar di bawah ini.

```
mu = mean
b = std_dev / np.sqrt(2)
x = np.linspace(data.min(), data.max(), 500)

pdf_laplace = (1 / 2 * b) * np.exp(-np.abs(x - mu) / b)
cdf_laplace = np.where(x <= mu, 0.5 * np.exp((x - mu) / b),
| | | | | 1 - 0.5 * np.exp(-(x - mu) / b))
```

Gambar 4. Kode untuk Distribusi Probabilitas Laplace



Gambar 5. CDF dan PDF dari Distribusi Probabilitas Laplace

Pada gambar PDF di atas, tampak bahwasanya data *fitted* sepenuhnya namun masih terdapat puncak histogram yang kosong, menandakan bahwasanya distribusi ini tidak sepenuhnya cocok dengan dataset yang dimiliki. Distribusi ini tidak terlalu cocok untuk tipe data dengan banyak *noise*. Akan tetapi, untuk CDF, *Laplace CDF* memiliki kemiripan atau sangat dekat (berhimpitan) dengan *empirical CDF* yang dimiliki dataset, hal ini menunjukkan bahwasanya untuk CDF, distribusi Laplace cukup cocok untuk digunakan pada dataset ini.

### 3. *Gumbel Distribution*

Distribusi ini merupakan distribusi yang digunakan untuk memodelkan nilai ekstrem, seringkali digunakan dalam analisis risiko, hidrologi, meteorologi, dan ilmu geofisika. Parameter pada distribusi ini diantaranya  $\mu$  dan  $\beta$ . Nilai PDF dan CDF dapat diperoleh dari persamaan berikut:

$$PDF = f(x; \mu, \beta) = \frac{1}{\beta} e^{-(z + e^{-z})}$$

$$CDF = F(x; \mu, \beta) = e^{-e^{-z}}$$

$$\text{dengan } z = \frac{x - \mu}{\beta}$$

Di mana:

- $x$  = variabel acak
- $\mu$  = parameter pusat lokasi (*mean*)
- $\beta$  = parameter skala yang diperoleh dengan  $\frac{\sigma\sqrt{6}}{\pi}$

Persamaan tersebut apabila dimasukkan pada *Python* dengan kode di bawah akan menghasilkan grafik PDF dan CDF seperti gambar di bawah ini.

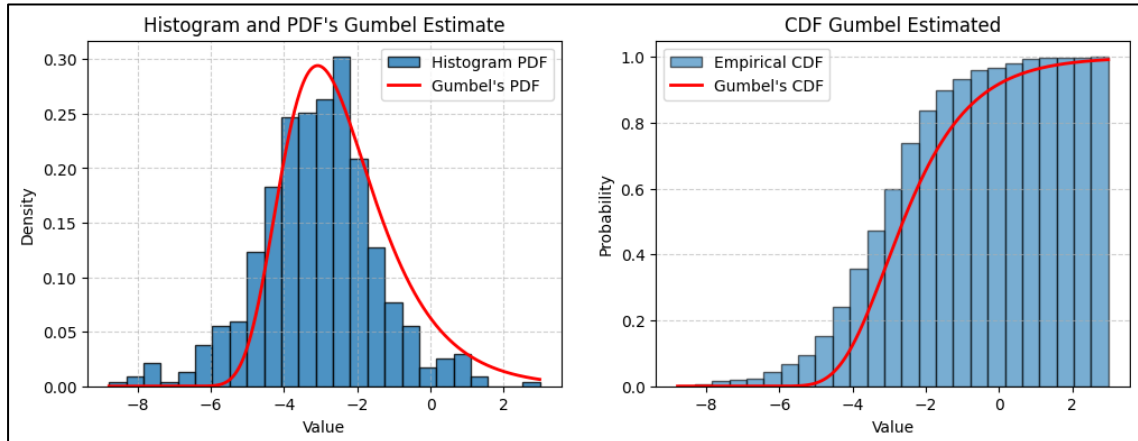
```

b = std_dev * np.sqrt(6) / (np.pi)
mu = mean
x = np.linspace(min(data), max(data), 500)
Z = (x - mu) / b # parameter tambahan untuk mempersingkat rumus

pdf_gumbel = 1 / b * np.exp(-(Z + np.exp(-Z)))
cdf_gumbel = np.exp(-np.exp(-Z))

```

Gambar 6. Kode untuk Distribusi Probabilitas Gumbel



Gambar 7. PDF dan CDF dari Distribusi Gumbel

Dari hasil pada gambar 7 di atas, dapat diamati bahwasanya PDF dari distribusi Gumbel memiliki cukup kemiripan dengan histogram meskipun pada PDF Gumbel memiliki ekor yang lebih panjang dan lebar ke arah kanan karena PDF Gumbel menurun secara eksponensial ke kanan sehingga data pada bagian kiri tidak *fitted* sepenuhnya. Kemudian, pada CDF Gumbel, CDF empiris tidak *fitted* sepenuhnya karena CDF Gumbel kurang menangkap kejadian ekstrem di bagian kiri bawah ( $x < -5$ ). Selain itu, pada ekor kanan, CDF empiris mencapai 1 lebih cepat dibandingkan CDF Gumbel. Oleh karena itu, model distribusi probabilitas ini kurang tepat untuk di-*fitting* dengan model empiris.

#### 4. Weibull Distribution

Distribusi ini merupakan distribusi probabilitas yang kerap digunakan dalam menganalisis data yang berhubungan dengan *reliability*. Distribusi ini cukup fleksibel karena dapat menyesuaikan dengan bentuk histogram karena terdapat parameter yang dapat diatur, seperti  $k$ , dan  $\lambda$ . Akan tetapi, parameter ini hanya berlaku untuk nilai positif, sehingga data perlu di-*shifting* terlebih dahulu. Nilai PDF dan CDF *Weibull Distribution* dapat diperoleh dengan persamaan berikut:

$$PDF = f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, x \geq 0$$

$$CDF = F(x; k, \lambda) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}, x \geq 0$$

Dimana:

- $k = \text{shape parameter}$ , yang diperoleh dari hasil percobaan *trial and error*
- $\lambda = \text{scale parameter}$ , yang diperoleh dari  $\frac{\text{median}}{\log_2 \frac{1}{k}}$

Karena distribusi Weibull hanya berlaku untuk data positif, maka perlu dilakukan *shifting* agar data dapat diolah dengan baik. Proses *shifting* dilakukan dengan kode berikut agar dapat diperoleh hasil *mean*, *bin width*, dan jumlah *bin* baru sesuai data yang *dishifting*.

```
# before we start weibull n dagum, data perlu dishifting agar hasilnya lebih sesuai.

data_shifted = data + 9 # agar posisi dimulai dari 0
mean_shifted = np.mean(data_shifted)
Q1_shifted = np.percentile(data_shifted, 25)
Q3_shifted = np.percentile(data_shifted, 75)
bin_width_shifted = 2 * (Q3_shifted - Q1_shifted) / (3 * np.sqrt(mean_shifted))
num_bin_shifted = int(np.ceil(((max(data_shifted) - min(data_shifted)) / bin_width_shifted)))
```

Gambar 8. Kode untuk *Shifting* Dataset

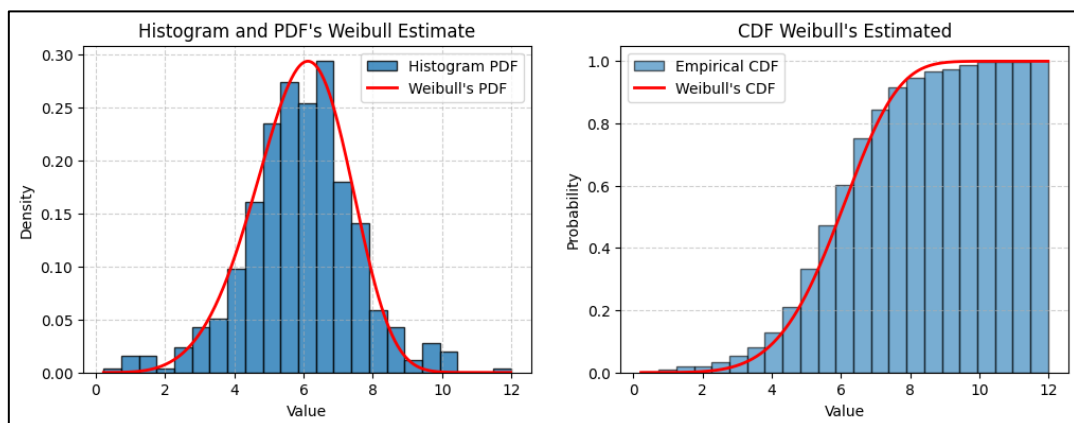
Setelah dilakukan *shifting*, hasil data baru dapat langsung diolah untuk memperoleh PDF dan CDF sebagaimana kode di bawah. Karena distribusi ini memiliki parameter yang dapat diatur sendiri, yaitu  $k$ , maka dilakukan percobaan *trial and error* untuk menemukan bentuk PDF dan CDF yang tepat sesuai dengan data. Akhirnya, diperoleh hasil seperti gambar berikut.

```
k = 5
lamda = np.median(data_shifted) / (np.log(2) ** (1/k))
x = np.linspace(data_shifted.min(), data_shifted.max(), 300)

# PDF weibull
pdf_weibull = np.where(x < 0, 0, (k / lamda) * ((x / lamda) ** (k - 1)) * np.exp(-(x / lamda) ** k))

# CDF weibull
cdf_weibull = np.where(x < 0, 0, 1 - np.exp(-(x / lamda) ** k))
```

Gambar 9. Kode untuk Distribusi Probabilitas Weibull



Gambar 10. PDF dan CDF dari Distribusi Probabilitas Weibull

Dari hasil yang diperoleh di atas pada gambar 10, pada histogram dengan PDF Weibull cukup sesuai dengan bentuk histogram, hal ini terjadi karena penyesuaian besar  $k$  sehingga bentuk PDF Weibull dapat *fitted* dengan histogram. Selain itu, besar  $k > 1$ , akan menghasilkan bentuk PDF Weibull berbentuk seperti lonceng (normal distribution), sehingga dapat sesuai dengan bentuk histogram yang dimiliki. Kemudian, pada kurva CDF, dapat diamati bahwasanya model CDF Weibull memiliki kemiripan yang berarti estimasi model Weibull sesuai dengan data. Dengan kata lain, setelah menganalisis bentuk PDF dan CDF Weibull, model ini memiliki kesesuaian dengan baik dan cocok untuk digunakan dalam dataset ini.

### 5. Dagum's Distribution

Distribusi ini merupakan salah satu distribusi probabilitas yang sering digunakan dalam analisis ekonomi. Layaknya distribusi Weibull, distribusi ini juga hanya berlaku untuk nilai positif, sehingga nantinya dataset perlu di-*shifting* agar dapat dilakukan *fitting* data dengan PDF dan CDF Dagum. Pada distribusi ini, terdapat beberapa parameter yang dapat diatur secara fleksibel, diantaranya  $a, b$ , dan  $p$ . Nilai PDF dan CDF Dagum dapat diperoleh dengan persamaan berikut:

$$PDF = f(x; a, b, p) = \frac{ap}{b} \left(\frac{x}{b}\right)^{ap-1} \left(1 + \left(\frac{x}{b}\right)^a\right)^{-(p+1)}$$

$$CDF = F(x; a, b, p) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p}$$

Untuk  $x > 0$  dengan parameter:

- $a > 0$  = menentukan bentuk distribusi
- $b > 0$  = parameter skala
- $c > 0$  = mengontrol ekor distribusi

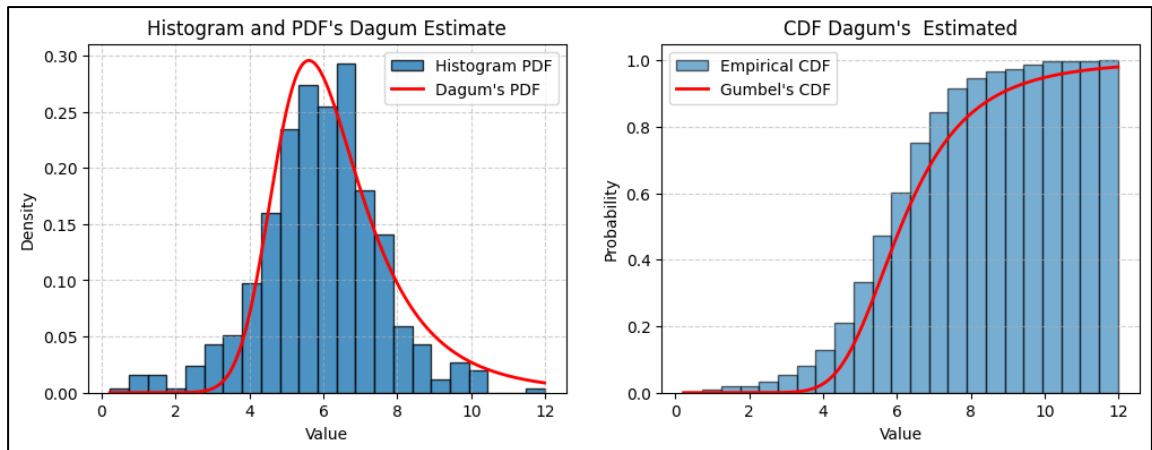
Karena data perlu dilakukan *shifting*, maka digunakan data hasil pengolahan gambar 8. Dan ditentukan nilai  $a, b$ , dan  $p$  dengan *trial and error* untuk memperoleh kurva PDF dan CDF Dagum yang *fit* dengan PDF histogram dan CDF empiris. Setelah *trial and error*, diperoleh nilai parameter yang sesuai, yaitu  $a = 5.5$ ,  $b = Q1$ , dan  $p = 2.5$ . Dari nilai tersebut, dataset dapat diolah dan diperoleh nilai PDF dan CDF Dagum sebagaimana gambar di bawah.

```
a = 5.5
b = np.percentile(data_shifted, 25)
c = 2.5
x = np.linspace(data_shifted.min(), data_shifted.max(), 500)

pdf_dagum = np.where(x < 0, 0, (a*c / x) * ((x/b)**(a*c) / (((x/b)**a) + 1) ** (c + 1)))

# CDF dagum
cdf_dagum = np.where(x < 0, 0, (1 + (x / b) ** -a) ** -c)
```

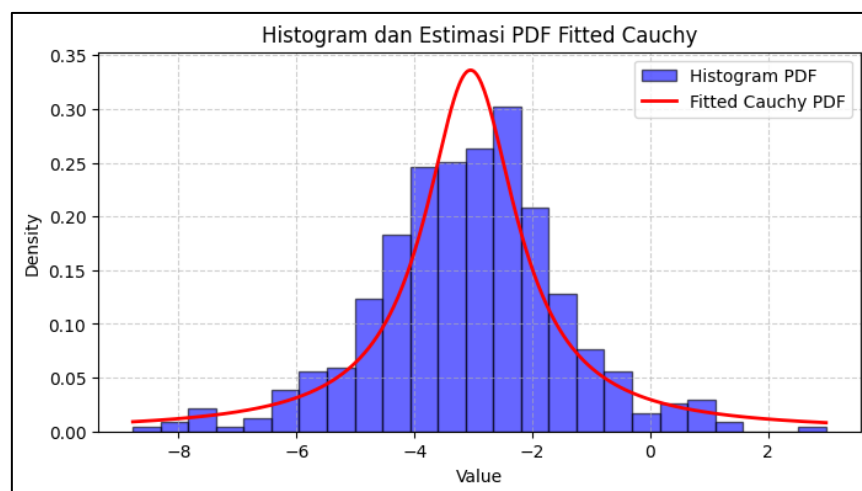
Gambar 11. Kode untuk Distribusi Probabilitas Dagum



Gambar 12. PDF dan CDF dari Distribusi Probabilitas Dagum

Dari hasil yang diperoleh sebagaimana gambar di atas, tampak bahwasanya PDF Dagum cukup sesuai dengan data yang dimiliki. Namun, bentuk asimetris dengan *right skewed* khas dari PDF Dagum menyebabkan data pada sebelah kiri tidak *fit* sepenuhnya. Kemudian, pada kurva CDF Dagum, dapat diamati bahwasanya kurva tersebut cukup sesuai dengan data empirisnya. Akan tetapi, adanya deviasi di ekor distribusi menandakan model distribusi ini kurang cocok untuk diterapkan dengan dataset yang digunakan.

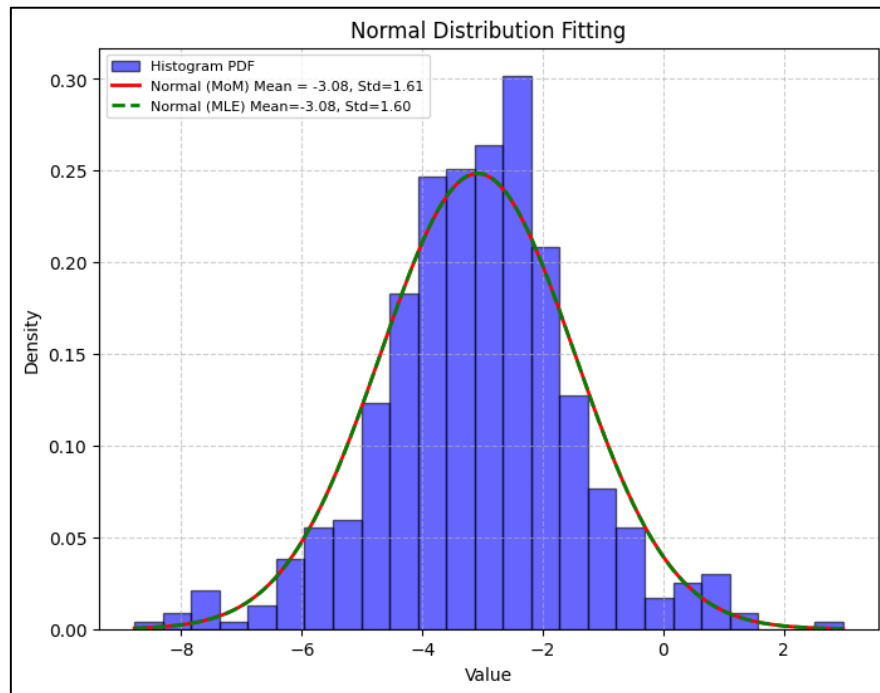
Setelah *fitting* distribusi probabilitas pada histogram PDF dan CDF, dilakukan transformasi data agar menyerupai distribusi normal dengan metode tertentu. Kali ini, digunakan distribusi probabilitas Cauchy karena memiliki bentuk PDF yang *fitted* dengan histogram data dan memiliki kemiripan dengan bentuk *normal distribution* hanya saja memiliki ekor yang panjang. Hasil dari *fitting* Cauchy dapat diamati pada gambar berikut:



Gambar 13. Histogram dan PDF dari Distribusi Cauchy



Kemudian, dengan estimator dari *normal distribution*, seperti MoM (*Methods of Moments*) dan MLE (*Methods Likelihood Estimation*). Diperoleh histogram dengan *fitted* PDF yang diperoleh melalui MoM dan MLE sebagaimana gambar berikut:



Gambar 14. *Normal Distribution Fitting*

### C. Interpretasi Hasil Pengolahan Data

Dari pengolahan yang telah dilakukan sebelumnya dengan berbagai macam distribusi probabilitas yang disesuaikan dengan data yang dimiliki, diperoleh bahwasanya tidak semua distribusi probabilitas dapat diterapkan pada suatu dataset. Hal ini dapat dipengaruhi dari sifat distribusi probabilitas tertentu yang menyebabkan tidak semua data *fitted* pada PDF dan CDF dari distribusi probabilitas. Pada dataset yang dimiliki, distribusi probabilitas yang sesuai dengan dataset tersebut diantaranya *Logistic Distribution*, *Weibull Distribution*, *Cauchy Distribution*, dan tentunya *Normal Distribution*.