

Izvještaj druge laboratorijske vježbe

Zadatak 1:

Na temelju uputa sniman je promet generiran prilikom igranja online igra, točnije igre Agar.io. Promet je snimljen Wireshark-om, a ukupno je snimljeno 10436 paketa. Snimljeni promet je filtriran po IPv6 adresi poslužitelja sljedećim filterom:

```
ipv6.addr == 2a00:1bb8:11f:93b1:f1df:f0d0:9875:24c4
```

Filtriranjem je dobiveno 8889 paketa, uglavnom TLS ili TCP protokola.

Zadatak 2:

Osnovna zadaća programskog koda je identifikacije distribucije koja najbolje opisuje slučajno generirane podatke dobivene snimanjem prometa. To se obavlja u funkciji `best_fit_distribution`. U toj funkciji se na podatcima dobivenim snimanjem prometa provodi identifikacija parametara pomoću funkcije `.fit()`. Dobiveni parametri se koriste u izračunu funkcije gustoće razdiobe. Najbolja distribucija odabire se na temelju izračuna sume kvadratne greške, bira se ona s najmanjom sumom. Kod uz ovo generira i tri grafa:

- histogram dobivenih podataka
- histogram sa funkcijama gustoće vjerojatnosti za sve razmatrane razdiobe
- histogram s funkcijom gustoće vjerojatnosti razdiobe koja najbolje opisuje empirijske podatke

Zadatak 3:

U dodatnoj napisanoj skripti "data_separator.py" radi se separacija podataka na podatke o duljini paketa i međudolaznim vremenima. Ti podaci se spremaju u datoteke "length.csv" i "intervals.csv". U priloženoj slici ispod prikazano je čitanje tih datoteka u glavnom kodu te priprema za njihovo daljnje korištenje.

```
#Učitavanje podataka iz datoteke
data_array=[]
count = 1
with open(r'C:\Users\telav\OneDrive\Radna površina\8.semestar\MPP\LAB2\length.csv') as csvDataFile:
    csvReader = csv.reader(csvDataFile, delimiter=',')
    for row in csvReader:
        if count % 2 == 1:
            data_array.append(float(row[0]))
            count+=1
data = pd.Series(data=data_array)
```

Postupak je identičan i za datoteku "intervals.csv".

Zadatak 4:

S obzirom da na testovima provedenim za zadane 3 distribucije najbolje rezultate postiže Weibullova distribucija, odabrane su iduće 3 distribucije:

- Gamma distribucija
- Log-normalna distribucija
- Pareto distribucija

Zadatak 5:

Na slikama ispod prikazane su vrijednosti medijana, prosjeka, minimalne i maksimalne vrijednosti za empirijske podatke te podatke generirane po svakoj distribuciji koristeći parametre dobivene ranije spomenutom metodom `.fit()`. Na prvoj slici su podaci za međudolazno vrijeme, a na drugoj za veličinu paketa.

```
Medijan: 0.00717149999999922
Prosječna vrijednost: 0.008744238523852385
Maksimalna vrijednost0.4852170000000058
Minimalna vrijednost0.0
=====
expon
Medijan: 0.0060907810873676435
Prosječna vrijednost: 0.008684584137785516
Maksimalna vrijednost0.0708000315176042
Minimalna vrijednost1.1137509834109622e-06
=====
norm
Medijan: 0.008641815434310623
Prosječna vrijednost: 0.008800026592521795
Maksimalna vrijednost0.04686651864163148
Minimalna vrijednost-0.033298050506836326
=====
dweibull
Medijan: 0.007545273342166739
Prosječna vrijednost: 0.007492106614583411
Maksimalna vrijednost0.057806286290494
Minimalna vrijednost-0.032768554649280524
=====
gamma
Medijan: 6.426068317121437e-09
Prosječna vrijednost: 0.04587629219090357
Maksimalna vrijednost7.1514619191565805
Minimalna vrijednost-4.1159059092807274e-30
=====
lognorm
Medijan: 2.4756100034719366e-47
Prosječna vrijednost: inf
Maksimalna vrijednostinf
Minimalna vrijednost-5e-324
=====
pareto
Medijan: 0.0057344266788777395
Prosječna vrijednost: 0.008806127468736732
Maksimalna vrijednost0.10548691856657115
Minimalna vrijednost1.4712039434067181e-06
```

Funkcija `best_fit_distribution` nam kaže kako je najbolja distribucija za ove podatke Paretova. Na temelju usporedbe ovih podataka s empirijskim podacima možemo vidjeti kako je ta procjena korektna s obzirom da se Pareto uz eksponencijalnu distribucija najbolje poklapa s empirijskim podacima.

```

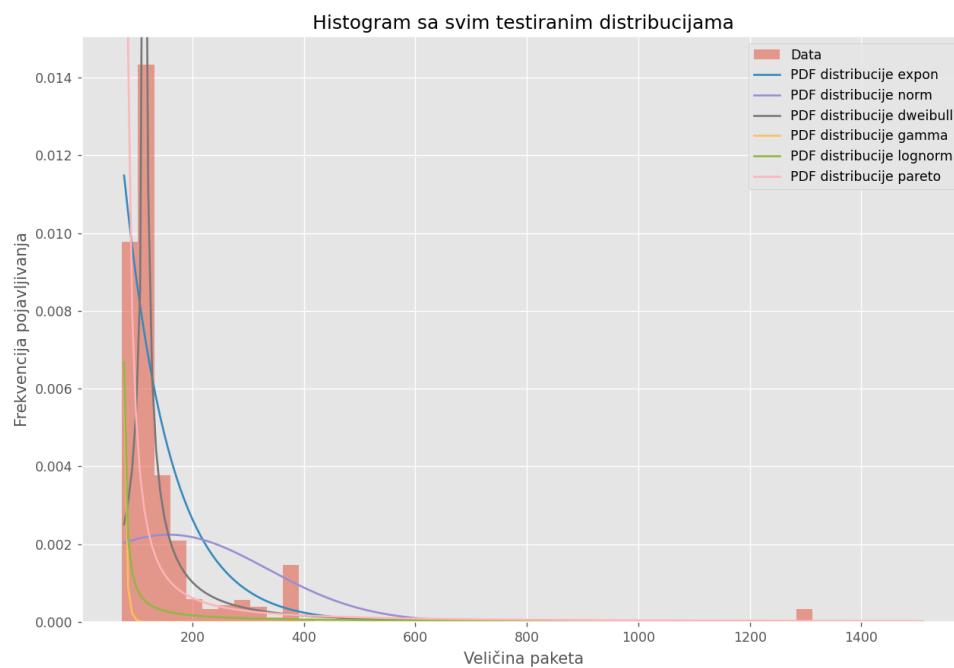
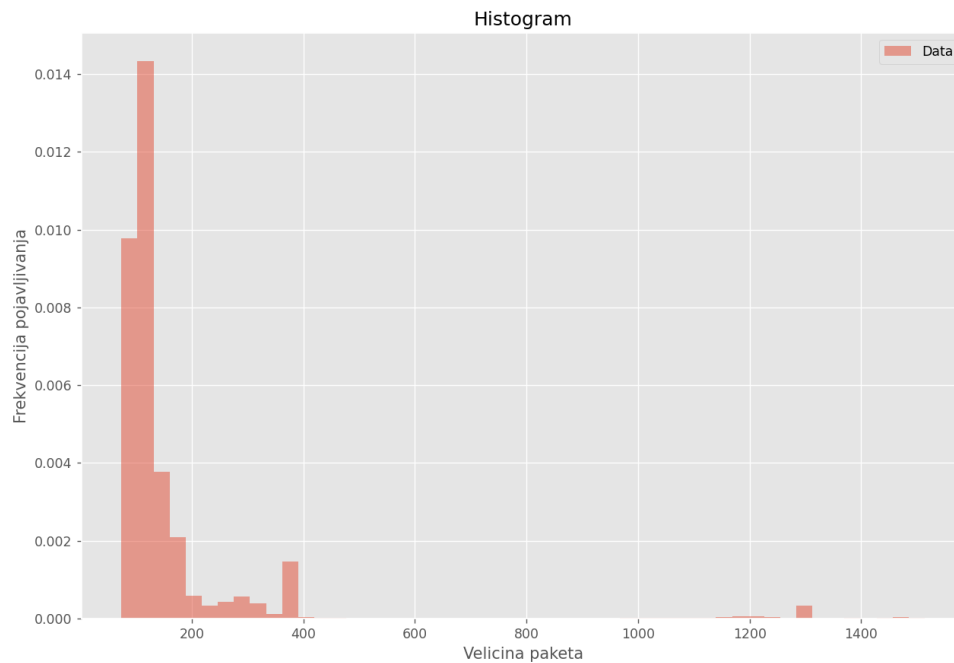
Medijan: 115.0
Prosječna vrijednost: 157.42186972662842
Maksimalna vrijednost1514.0
Minimalna vrijednost74.0
=====
expon
Medijan: 130.96283683319294
Prosječna vrijednost: 157.01877144500602
Maksimalna vrijednost806.59804356444
Minimalna vrijednost74.01198728292165
=====
norm
Medijan: 155.31720259060734
Prosječna vrijednost: 156.0187135625186
Maksimalna vrijednost813.9495535019024
Minimalna vrijednost-485.86603051216866
=====
dweibull
Medijan: 114.9960872061912
Prosječna vrijednost: 114.9407778030729
Maksimalna vrijednost2857.2118911371667
Minimalna vrijednost-2372.439890173327
=====
gamma
Medijan: 74.00000004416034
Prosječna vrijednost: 74.29598358306292
Maksimalna vrijednost109.03540621778882
Minimalna vrijednost73.99999999999999
=====
lognorm
Medijan: 74.01358242768235
Prosječna vrijednost: 2.08366847793913e+18
Maksimalna vrijednost9.257662831311198e+21
Minimalna vrijednost73.99999999999999
=====
pareto
Medijan: 85.54277431772243
Prosječna vrijednost: 121557584.387372
Maksimalna vrijednost1054204588536.442
Minimalna vrijednost74.00087376734825

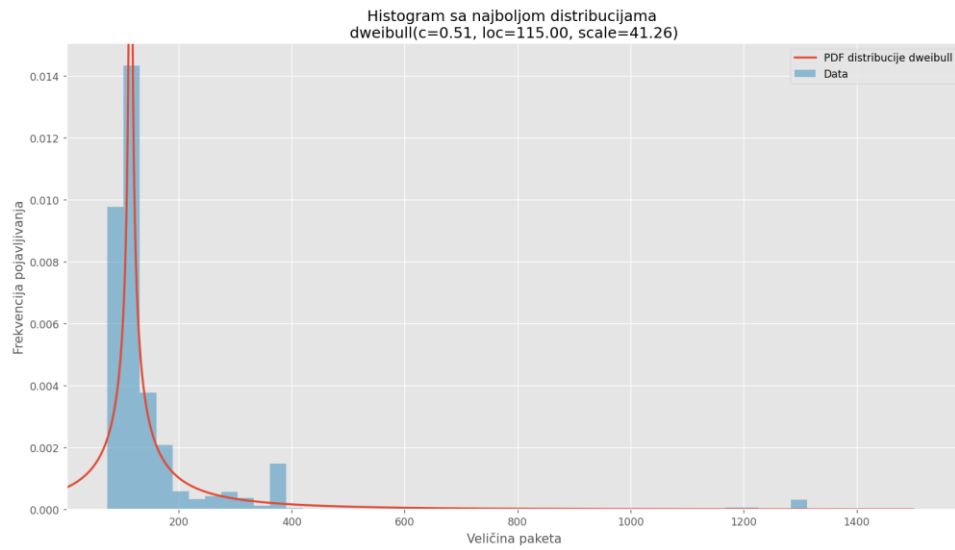
```

Funkcija `best_fit_distribution` nam kaže kako je najbolja distribucija za ove podatke Weibullova. Ova procjena je korektna, iako bi se čak po podacima moglo reći da je bolja eksponencijalna distribucija s obzirom da su njene prosječne vrijednosti, maksimum i minimum bliži empirijskim podacima, dok Weibullova distribucija prednost ima samo kod medijana koji je iznimno blizu empirijskim podacima.

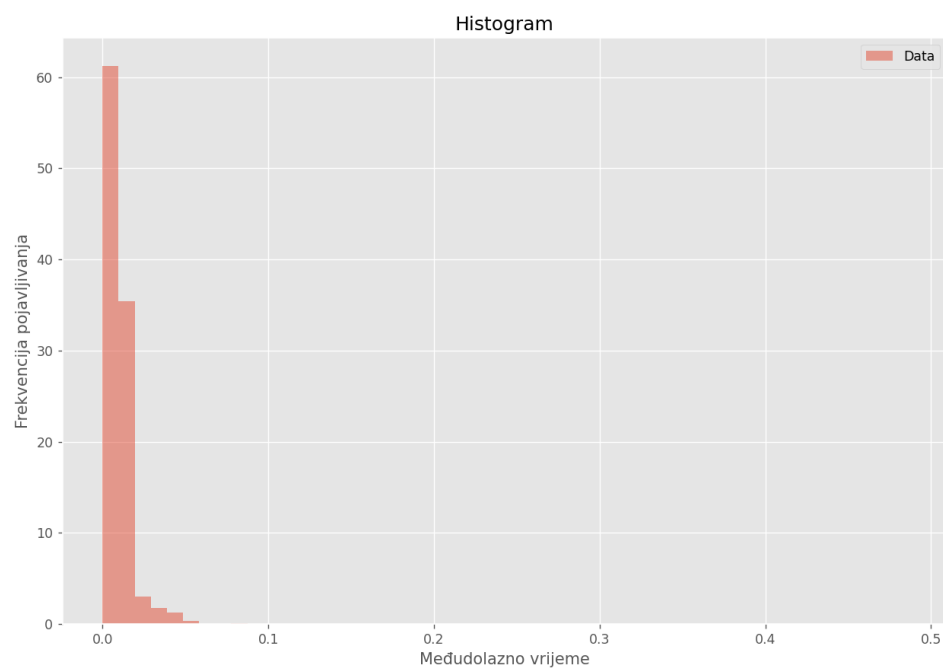
Zadatak 6:

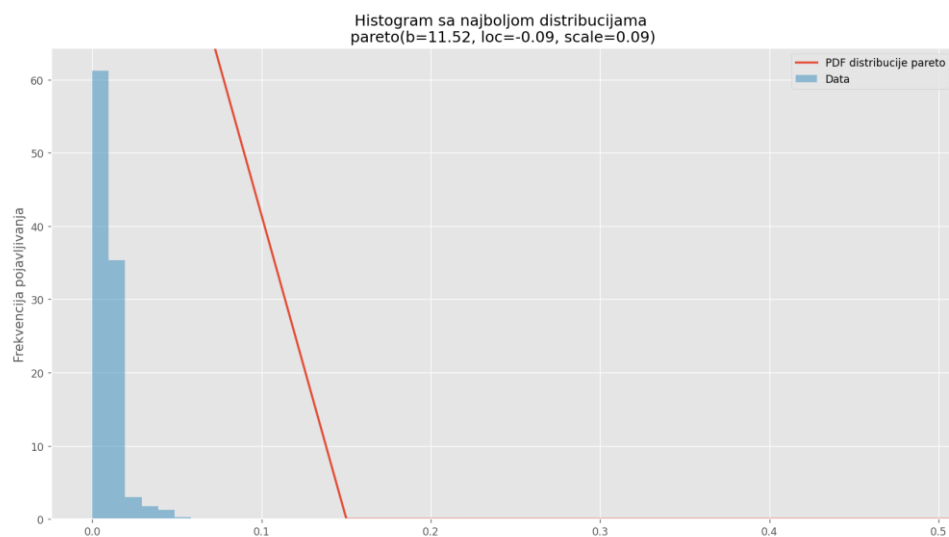
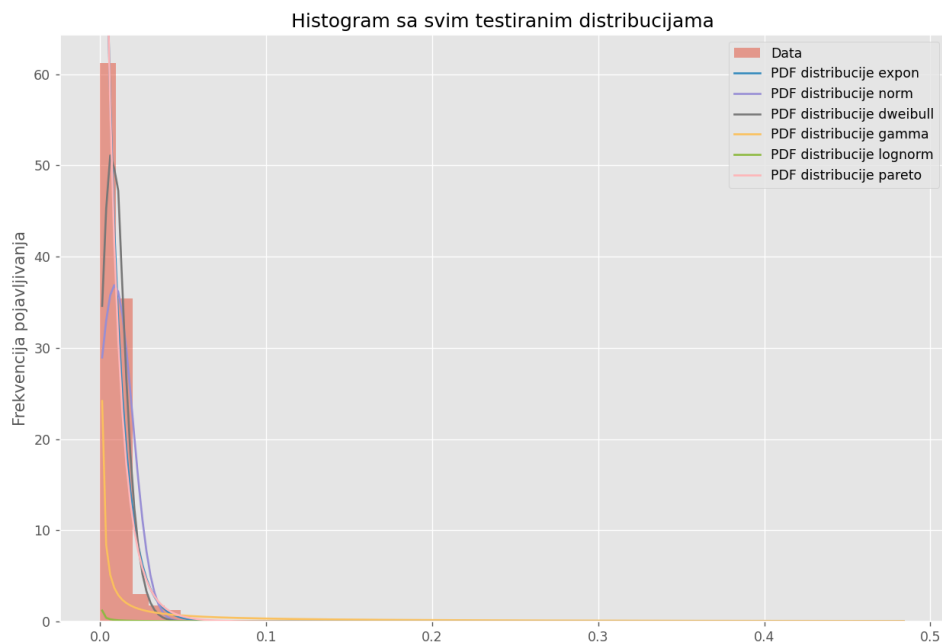
Iduća tri grafa se odnose na podatke o veličini paketa. Na njima se vidi kako je Weibullova distribucija najbolja prema izračunima funkcije `best_fit_distribution`.





Iduća tri grafa se odnose na podatke o veličini paketa. Na njima se vidi kako je Paretova distribucija najbolja prema izračunima funkcije `best_fit_distribution`.





Zadatak 7:

U priloženim datotekama se uz već spomenute “length.csv” i “intervals.csv” datoteke nalazi i “lab2_data.csv” u kojoj se nalaze svi podatci izvedeni iz Wiresharka. Uz to je priloženi i originalni zapis snimljenih paketa koji se može pregledati u Wireshark-u.