IIT KANPUR

CE462A PROJECT1

INSTRUCTOR
PROF. SHIVAM TRIPATHI

SUBMITTED BY

PRASHANT CHOUHAN
170492

Executive Summary

While going through this project the best-fitted distribution is identified out of given i.e. Gaussian, Log-Normal, Exponential, Pearson type III, Gamma on the basis of goodness-of-fit.

The given data set for annual daily maximum discharge for a hundred years is provided and discharge is assumed to be a random variable.

The probability distribution of various distribution is calculated by programming in Python language.

Chi-Squared, Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit at 5% significance level is performed to select the best distribution out of the given distribution. Preference is given to the Anderson-Darling test.

The corresponding p-value is tabulated on the basis of which the distribution is selected.

The graph of the best-fitted distribution is attached along with the distribution. Then the corresponding design discharge is calculated for best-fitted distribution for the given data set.

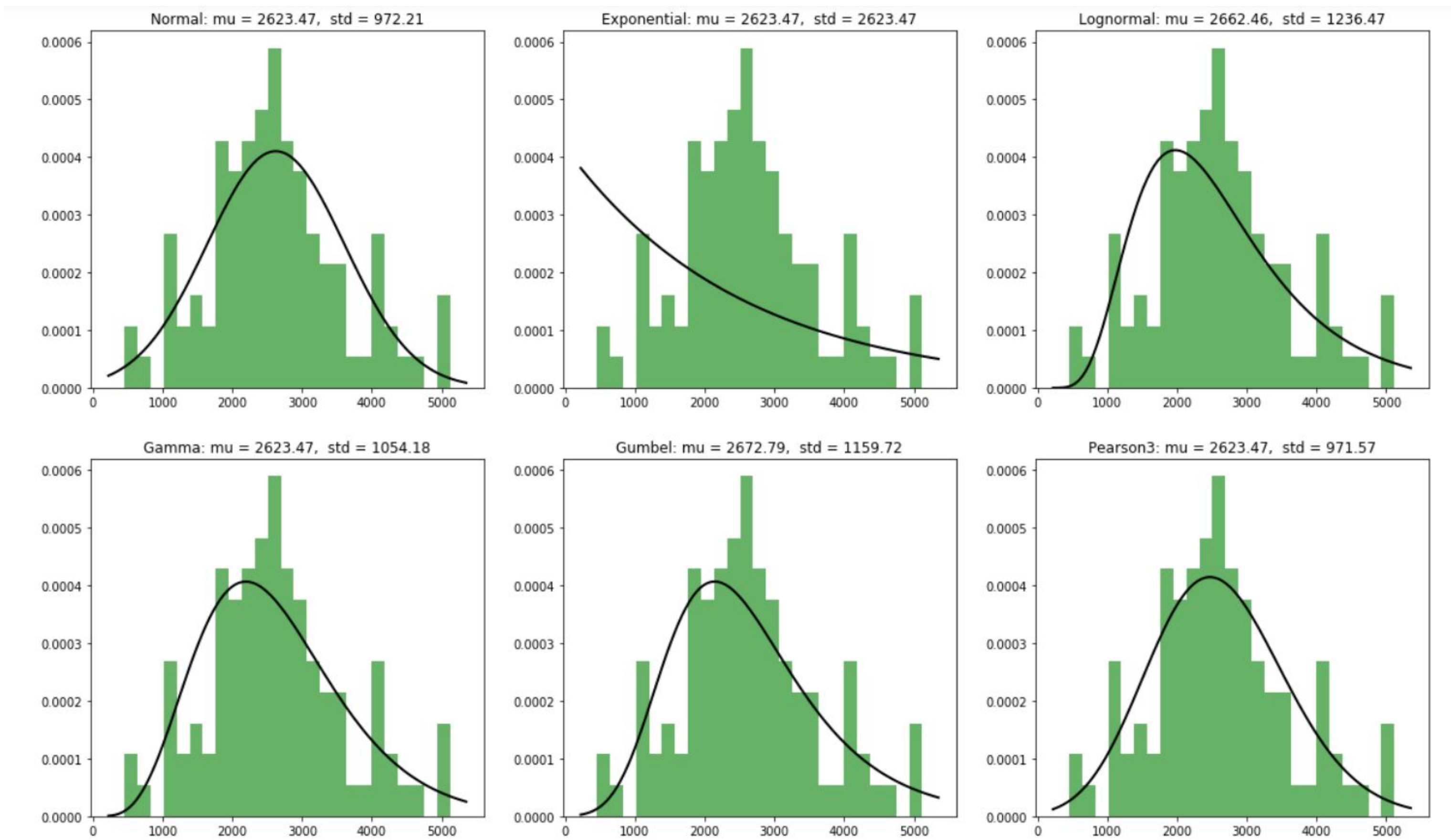
Methodology

Python language was used on Jupyter-notebook to define the distribution to calculate the probability distribution for these hundred data points. For the Chi-Squared test the data is divided into ten intervals of equal frequency. The three goodness of fit tests are defined then, and the value corresponding to 5% critical value for each distribution is calculated and compared with the statistical value of each distribution. The distributions which are within the 5% of critical value are accepted and the corresponding p-value is calculated using the critical value chart, this is done for each test. On the basis of these values, the best distribution is selected.

In this study, frequency Analysis was carried out using different distribution models such as Exponential, Gaussian, Lognormal, Gamma, Pearson Type-III and Gumbel distributions for each site. The annual peak flow series was used for this purpose and then Chi-squared, Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests were performed at 5% significance level and the best distribution was selected. Then using the distribution that fitted best was used to calculate the 50-year return period discharge at each site

Result:-

Site 1:



Distribution : norm
Mean = 2623.4709
Standard Deviation = 972.2060
Parameters = [2623.4709, 972.206]
Anderson Test: Accept
statistical Value (A2*) = 0.5532
Critical Value (α) = 0.7590
p Value = 0.1539
KS Test: Accept
P value = 0.7638
Stat value = 0.0678
Chi_square Test: Accept
P value = 0.5285
Stat value = 15.9341
AIC = 1663.7012
BIC = 1668.9116

Distribution : expon
Mean = 2623.4709
Standard Deviation = 2623.4709
Parameters = [0.0, 2623.4709]
Anderson Test: Reject
statistical Value (A2*) = 18.3595
Critical Value (α) = 1.3330
p Value = 0.0000
KS Test: Reject
P value = 0.0000
Stat value = 0.3514
Chi_square Test: Reject
P value = 0.0000
Stat value = 126.9073
AIC = 1778.4507
BIC = 1783.6610

Distribution : lognorm
Mean = 2662.4573
Standard Deviation = 1236.4680
Parameters = [0.4419, 0.0, 2414.7593]
Anderson Test: Reject
statistical Value (A2*) = 1.8496
Critical Value (α) = 0.7590
p Value = 0.0001
KS Test: Accept
P value = 0.2165
Stat value = 0.1038
Chi_square Test: Accept
P value = 0.1560
Stat value = 22.7955
AIC = 1684.3367
BIC = 1692.1522

Distribution : gamma
Mean = 2623.4709
Standard Deviation = 1054.1761
Parameters = [6.1934, 0.0, 423.5943]
Anderson Test: Accept
statistical Value (A2*) = 0.8494
Critical Value (α) = 2.4120
p Value = 0.0289
KS Test: Accept
P value = 0.6453
Stat value = 0.0736
Chi_square Test: Accept
P value = 0.4222
Stat value = 17.4821
AIC = 1670.6838
BIC = 1678.4993

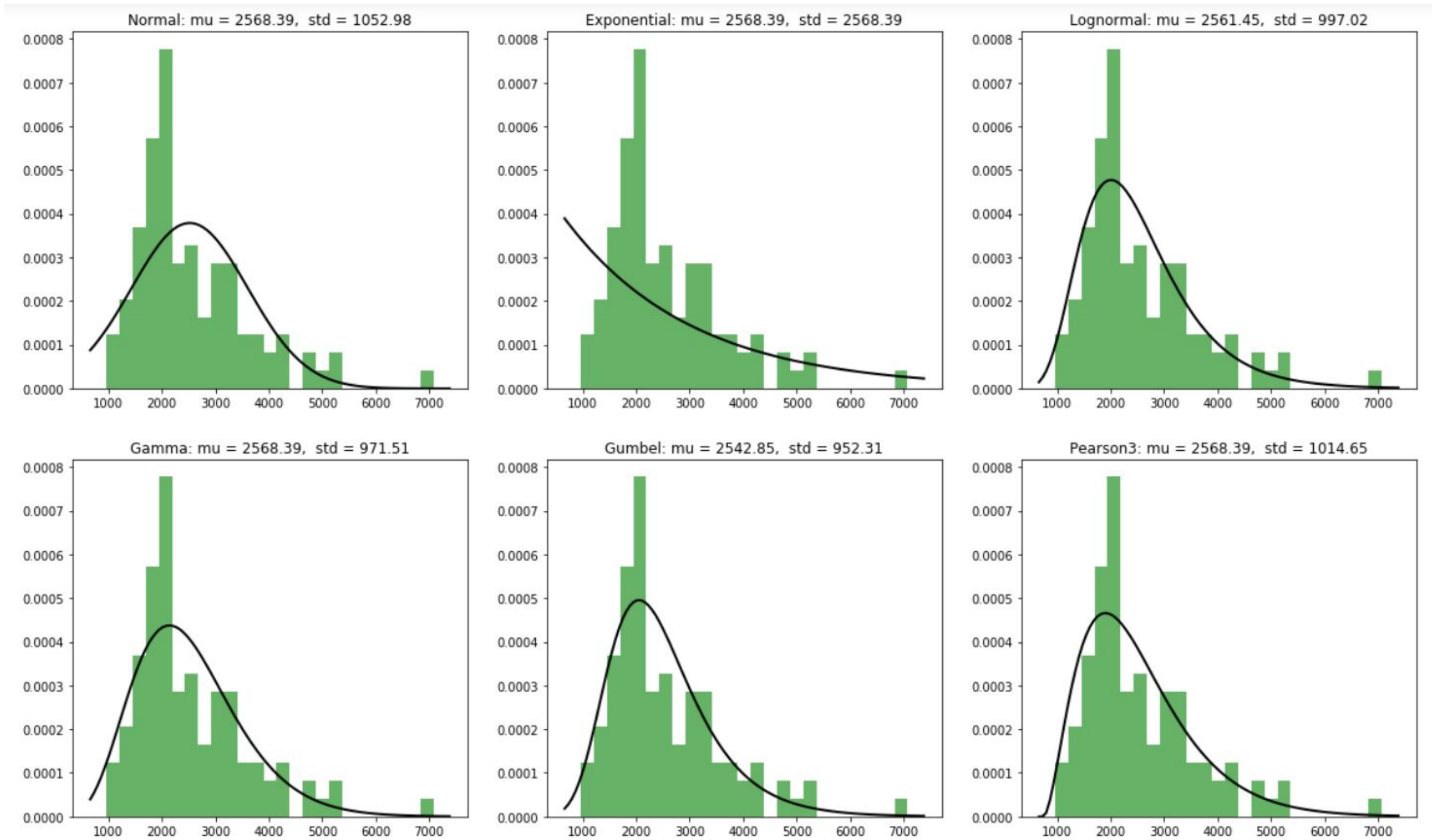
Distribution : gumbel_r
Mean = 2672.7943
Standard Deviation = 1159.7227
Parameters = [2150.8574, 904.2321]
Anderson Test: Reject
statistical Value (A2*) = 0.9948
Critical Value (α) = 0.7420
p Value = 0.0127
KS Test: Accept
P value = 0.6187
Stat value = 0.0750
Chi_square Test: Accept
P value = 0.3674
Stat value = 18.3443
AIC = 1669.9509
BIC = 1675.1612

Distribution : pearson3
Mean = 2623.4709
Standard Deviation = 971.5671
Parameters = [0.3031, 2623.4709, 971.5671]
Anderson Test: Accept
statistical Value (A2*) = 0.3580
p Value = 0.4533
KS Test: Accept
P value = 0.9682
Stat value = 0.0493
Chi_square Test: Accept
P value = 0.6073
Stat value = 14.8365
AIC = 1664.0300
BIC = 1671.8456

```
Best fitted Distribution:
According to AD test: pearson3
According to KS test: pearson3
According to Chi2 test: pearson3
According to AIC Value: norm
According to BIC Value: norm

Designed Discharge for 50 year return period = 4630.4628
```

Result:-

Site 2:



Distribution : norm
Mean = 2568.3929
Standard Deviation = 1052.9823
Parameters = [2568.3929, 1052.9823]
Anderson Test: Reject
statistical Value (A2*) = 3.0558
Critical Value (α) = 0.7590
p Value = 0.0000
KS Test: Reject
P value = 0.0288
Stat value = 0.1438
Chi_square Test: Reject
P value = 0.0001
Stat value = 49.4006
AIC = 1679.6641
BIC = 1684.8744

Distribution : expon
Mean = 2568.3929
Standard Deviation = 2568.3929
Parameters = [0.0, 2568.3929]
Anderson Test: Reject
statistical Value (A2*) = 18.7304
Critical Value (α) = 1.3330
p Value = 0.0000
KS Test: Reject
P value = 0.0000
Stat value = 0.3689
Chi_square Test: Reject
P value = 0.0000
Stat value = 130.9262
AIC = 1774.2071
BIC = 1779.4175

Distribution : lognorm
Mean = 2561.4545
Standard Deviation = 997.0182
Parameters = [0.3756, 0.0, 2387.0051]
Anderson Test: Accept
statistical Value (A2*) = 0.6344
Critical Value (α) = 0.7590
p Value = 0.0982
KS Test: Accept
P value = 0.3574
Stat value = 0.0912
Chi_square Test: Accept
P value = 0.1537
Stat value = 22.8651
AIC = 1649.4979
BIC = 1657.3134

Distribution : gamma
Mean = 2568.3929
Standard Deviation = 971.5075
Parameters = [6.9893, 0.0, 367.4776]
Anderson Test: Accept
statistical Value (A2*) = 1.1771
Critical Value (α) = 2.4120
p Value = 0.0045
KS Test: Accept
P value = 0.1536
Stat value = 0.1116
Chi_square Test: Reject
P value = 0.0373
Stat value = 28.7054
AIC = 1655.6721
BIC = 1663.4876

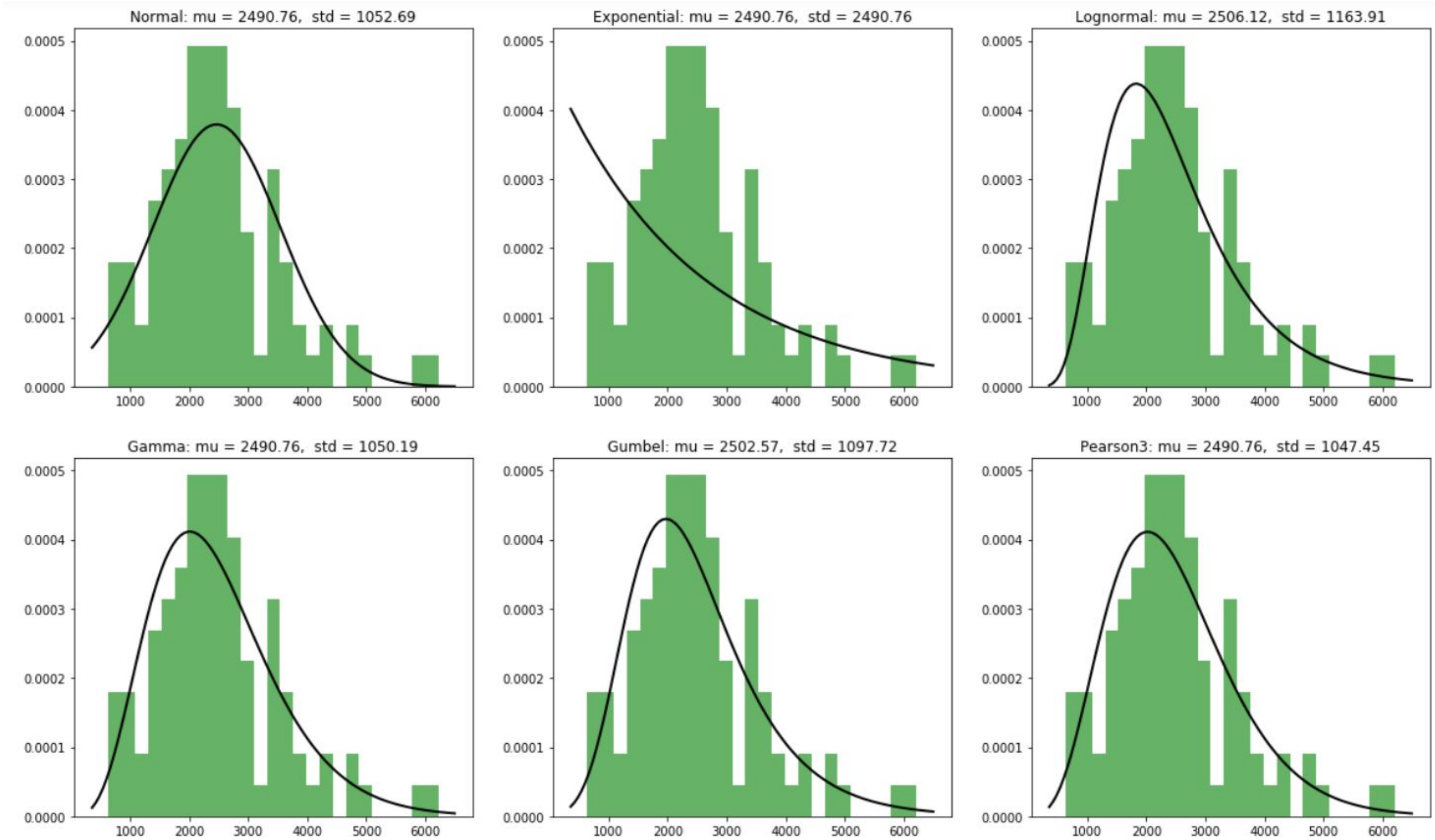
Distribution : gumbel_r
Mean = 2542.8527
Standard Deviation = 952.3104
Parameters = [2114.2623, 742.5133]
Anderson Test: Accept
statistical Value (A2*) = 0.7230
Critical Value (α) = 0.7420
p Value = 0.0594
KS Test: Accept
P value = 0.3072
Stat value = 0.0951
Chi_square Test: Accept
P value = 0.1436
Stat value = 23.1768
AIC = 1648.3307
BIC = 1653.5410

Distribution : pearson3
Mean = 2568.3929
Standard Deviation = 1014.6537
Parameters = [1.1785, 2568.3929, 1014.6537]
Anderson Test: Accept
statistical Value (A2*) = 0.5811
p Value = 0.1307
KS Test: Accept
P value = 0.5132
Stat value = 0.0808
Chi_square Test: Accept
P value = 0.1411
Stat value = 23.2548
AIC = 1649.0271
BIC = 1656.8427

```
Best fitted Distribution:
According to AD test: pearson3
According to KS test: pearson3
According to Chi2 test: lognorm
According to AIC Value: gumbel_r
According to BIC Value: gumbel_r

Designed Discharge for 50 year return period = 5037.4863
```

Result:-

Site 3:



Distribution : norm
Mean = 2490.7636
Standard Deviation = 1052.6854
Parameters = [2490.7636, 1052.6854]
Anderson Test: Reject
statistical Value (A2*) = 1.2250
Critical Value (α) = 0.7590
p Value = 0.0034
KS Test: Accept
P value = 0.1656
Stat value = 0.1099
Chi_square Test: Accept
P value = 0.1700
Stat value = 22.3948
AIC = 1679.6076
BIC = 1684.8180

Distribution : expon
Mean = 2490.7636
Standard Deviation = 2490.7636
Parameters = [0.0, 2490.7636]
Anderson Test: Reject
statistical Value (A2*) = 16.4447
Critical Value (α) = 1.3330
p Value = 0.0000
KS Test: Reject
P value = 0.0000
Stat value = 0.3234
Chi_square Test: Reject
P value = 0.0000
Stat value = 102.1503
AIC = 1768.0689
BIC = 1773.2793

Distribution : lognorm
Mean = 2506.1160
Standard Deviation = 1163.9111
Parameters = [0.4419, 0.0, 2272.9459]
Anderson Test: Reject
statistical Value (A2*) = 0.8051
Critical Value (α) = 0.7590
p Value = 0.0372
KS Test: Accept
P value = 0.3956
Stat value = 0.0884
Chi_square Test: Accept
P value = 0.1821
Stat value = 22.0686
AIC = 1672.2398
BIC = 1680.0554

Distribution : **gamma**

Mean = 2490.7636
Standard Deviation = 1050.1878
Parameters = [5.6251, 0.0, 442.7937]
Anderson Test: Accept
  statistical Value (A2*) = 0.4198
  Critical Value (α) = 2.4120
  p Value = 0.3257
KS Test: Accept
  P value = 0.8419
  Stat value = 0.0616
Chi_square Test: Accept
  P value = 0.3430
  Stat value = 18.7519
AIC = 1668.7435
BIC = 1676.5590

Distribution : **gumbel_r**

Mean = 2502.5680
Standard Deviation = 1097.7205
Parameters = [2008.5354, 855.8892]
Anderson Test: Accept
  statistical Value (A2*) = 0.4208
  Critical Value (α) = 0.7420
  p Value = 0.3239
KS Test: Accept
  P value = 0.7924
  Stat value = 0.0664
Chi_square Test: Accept
  P value = 0.3236
  Stat value = 19.0861
AIC = 1667.1129
BIC = 1672.3233

Distribution : **pearson3**

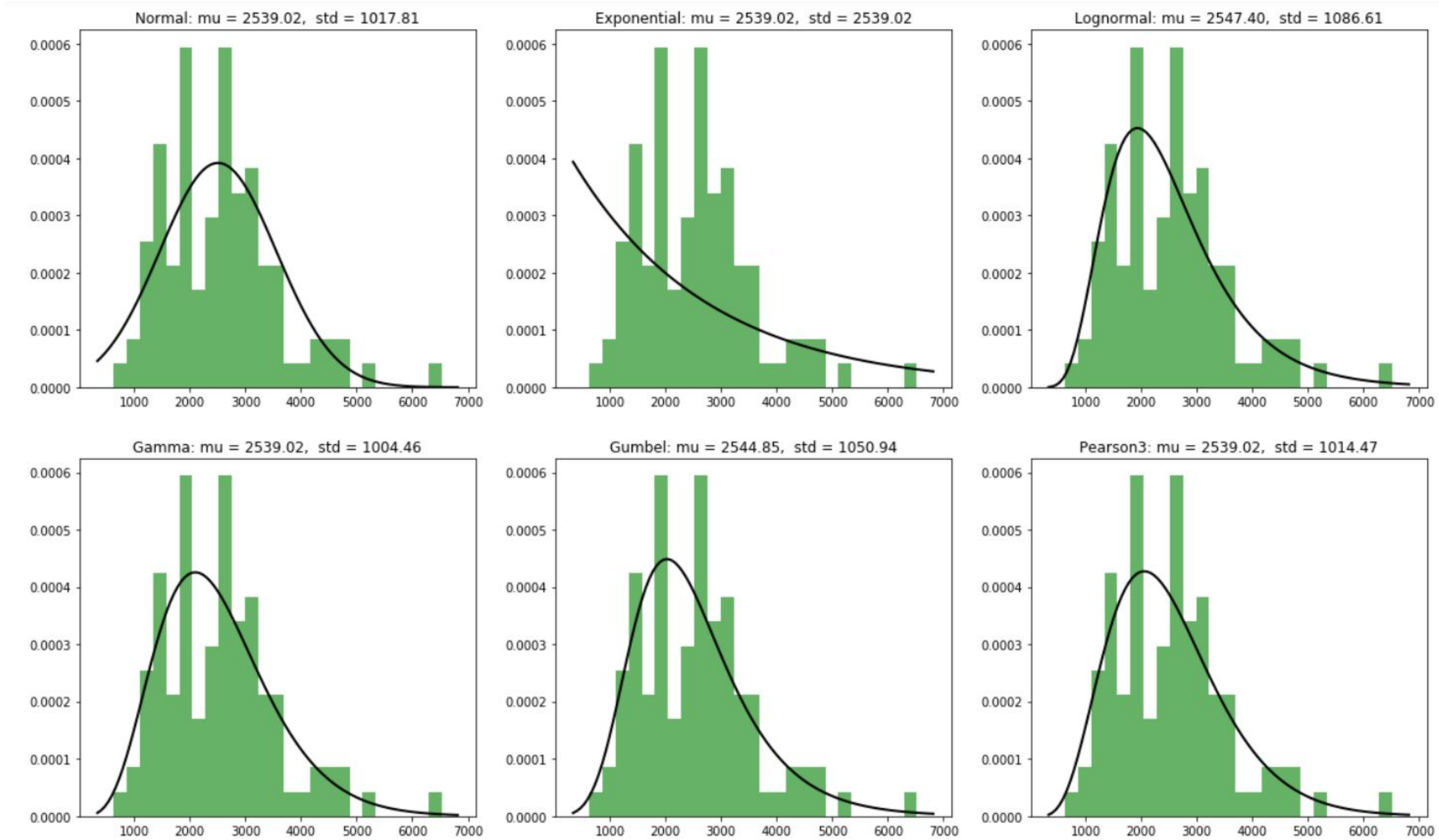
Mean = 2490.7636
Standard Deviation = 1047.4490
Parameters = [0.8263, 2490.7636, 1047.449]
Anderson Test: Accept
  statistical Value (A2*) = 0.4105
  p Value = 0.3426
KS Test: Accept
  P value = 0.8403
  Stat value = 0.0617
Chi_square Test: Accept
  P value = 0.3483
  Stat value = 18.6617
AIC = 1668.7367
BIC = 1676.5522

```
Best fitted Distribution:
According to AD test: pearson3
According to KS test: gamma
According to Chi2 test: pearson3
According to AIC Value: gumbel_r
According to BIC Value: gumbel_r
```

```
Designed Discharge for 50 year return period = 5483.9351
```

Result:-

Site 4:



Distribution : norm
Mean = 2539.0236
Standard Deviation = 1017.8083
Parameters = [2539.0236, 1017.8083]
Anderson Test: Reject
statistical Value (A2*) = 1.0233
Critical Value (α) = 0.7590
p Value = 0.0108
KS Test: Accept
P value = 0.4631
Stat value = 0.0839
Chi_square Test: Accept
P value = 0.0741
Stat value = 26.0187
AIC = 1672.8691
BIC = 1678.0794

Distribution : expon
Mean = 2539.0236
Standard Deviation = 2539.0236
Parameters = [0.0, 2539.0236]
Anderson Test: Reject
statistical Value (A2*) = 17.4858
Critical Value (α) = 1.3330
p Value = 0.0000
KS Test: Reject
P value = 0.0000
Stat value = 0.3324
Chi_square Test: Reject
P value = 0.0000
Stat value = 114.6693
AIC = 1771.9070
BIC = 1777.1173

Distribution : lognorm
Mean = 2547.4004
Standard Deviation = 1086.6112
Parameters = [0.4089, 0.0, 2343.1361]
Anderson Test: Accept
statistical Value (A2*) = 0.5320
Critical Value (α) = 0.7590
p Value = 0.1739
KS Test: Accept
P value = 0.4478
Stat value = 0.0849
Chi_square Test: Accept
P value = 0.2785
Stat value = 19.9179
AIC = 1662.7602
BIC = 1670.5757

Distribution : **gamma**
Mean = 2539.0236
Standard Deviation = 1004.4598
Parameters = [6.3895, 0.0, 397.373]
Anderson Test: Accept
statistical Value (A2*) = 0.3522
Critical Value (α) = 2.4120
p Value = 0.4673
KS Test: Accept
P value = 0.8710
Stat value = 0.0595
Chi_square Test: Accept
P value = 0.3187
Stat value = 19.1725
AIC = 1661.3793
BIC = 1669.1948

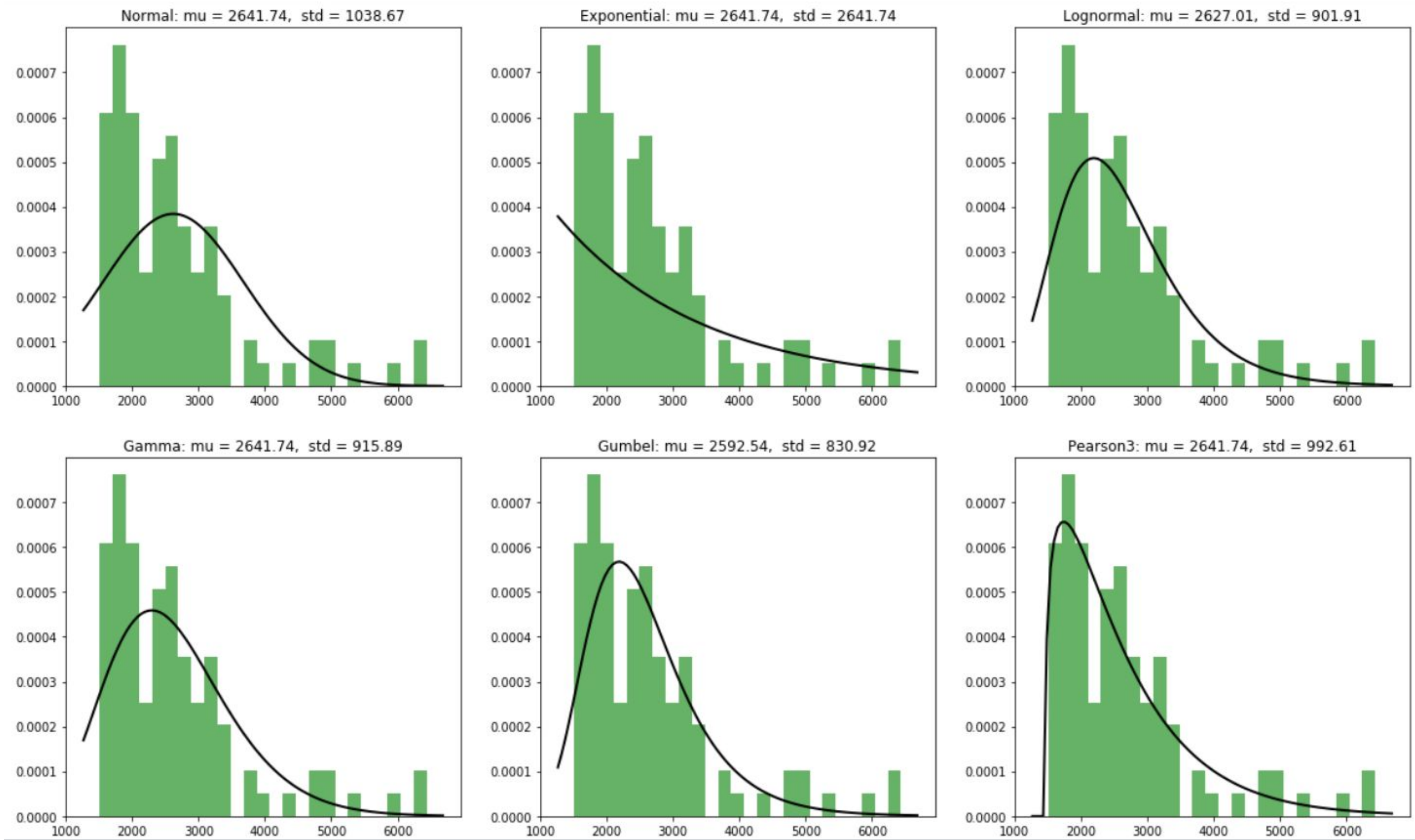
Distribution : **gumbel_r**
Mean = 2544.8519
Standard Deviation = 1050.9387
Parameters = [2071.8736, 819.4135]
Anderson Test: Accept
statistical Value (A2*) = 0.4344
Critical Value (α) = 0.7420
p Value = 0.3007
KS Test: Accept
P value = 0.6195
Stat value = 0.0749
Chi_square Test: Accept
P value = 0.3137
Stat value = 19.2627
AIC = 1659.7383
BIC = 1664.9487

Distribution : **pearson3**
Mean = 2539.0237
Standard Deviation = 1014.4682
Parameters = [0.8623, 2539.0237, 1014.4682]
Anderson Test: Accept
statistical Value (A2*) = 0.3589
p Value = 0.4512
KS Test: Accept
P value = 0.8035
Stat value = 0.0643
Chi_square Test: Accept
P value = 0.3191
Stat value = 19.1652
AIC = 1661.2315
BIC = 1669.0470

```
Best fitted Distribution:
According to AD test: gamma
According to KS test: gamma
According to Chi2 test: pearson3
According to AIC Value: gumbel_r
According to BIC Value: gumbel_r

Designed Discharge for 50 year return period = 5380.5926
```

Result:-

Site 5:



Distribution : norm

Mean = 2641.7391
Standard Deviation = 1038.6654
Parameters = [2641.7391, 1038.6654]
Anderson Test: Reject
statistical Value (A2*) = 5.0738
Critical Value (α) = 0.7590
p Value = 0.0000
KS Test: Reject
P value = 0.0219
Stat value = 0.1484
Chi_square Test: Reject
P value = 0.0000
Stat value = 92.3763
AIC = 1676.9261
BIC = 1682.1364

Distribution : expon

Mean = 2641.7391
Standard Deviation = 2641.7391
Parameters = [0.0, 2641.7391]
Anderson Test: Reject
statistical Value (A2*) = 21.2733
Critical Value (α) = 1.3330
p Value = 0.0000
KS Test: Reject
P value = 0.0000
Stat value = 0.4447
Chi_square Test: Reject
P value = 0.0000
Stat value = 180.7174
AIC = 1779.8385
BIC = 1785.0489

Distribution : lognorm

Mean = 2627.0113
Standard Deviation = 901.9077
Parameters = [0.3338, 0.0, 2484.6571]
Anderson Test: Reject
statistical Value (A2*) = 1.7095
Critical Value (α) = 0.7590
p Value = 0.0002
KS Test: Accept
P value = 0.4318
Stat value = 0.0859
Chi_square Test: Reject
P value = 0.0001
Stat value = 47.3597
AIC = 1633.9248
BIC = 1641.7403

Distribution : gamma
Mean = 2641.7391
Standard Deviation = 915.8914
Parameters = [8.3194, 0.0, 317.5397]
Anderson Test: Reject
statistical Value (A2*) = 2.5890
Critical Value (α) = 2.4120
p Value = 0.0000
KS Test: Accept
P value = 0.2253
Stat value = 0.1028
Chi_square Test: Reject
P value = 0.0000
Stat value = 58.1250
AIC = 1645.5096
BIC = 1653.3251

Distribution : gumbel_r
Mean = 2592.5437
Standard Deviation = 830.9238
Parameters = [2218.5838, 647.8686]
Anderson Test: Reject
statistical Value (A2*) = 1.6175
Critical Value (α) = 0.7420
p Value = 0.0004
KS Test: Accept
P value = 0.2655
Stat value = 0.0988
Chi_square Test: Reject
P value = 0.0000
Stat value = 49.9773
AIC = 1629.3676
BIC = 1634.5779

Distribution : pearson3
Mean = 2641.7391
Standard Deviation = 992.6116
Parameters = [1.7484, 2641.7391, 992.6116]
Anderson Test: Accept
statistical Value (A2*) = 0.4386
p Value = 0.2940
KS Test: Accept
P value = 0.9715
Stat value = 0.0487
Chi_square Test: Accept
P value = 0.1433
Stat value = 23.1860
AIC = 1608.8119
BIC = 1616.6274

```
Best fitted Distribution:
According to AD test: pearson3
According to KS test: pearson3
According to Chi2 test: pearson3
According to AIC Value: pearson3
According to BIC Value: pearson3
```

```
Designed Discharge for 50 year return period = 5485.2158
```

Summary

On the basis of analyzing the given distributions are best fitted for given datasets:

Site	Best Fitted Distribution	Design Discharge for 50 years return Period
1	Gaussian (Normal)	4630.4628
2	Gumbel	5037.4863
3	Gumbel	5483.9351
4	Gumbel	5380.5926
5	Pearson type III	5485.2158

In [796]:
```
import csv
import scipy.stats as stats
import numpy as np
from matplotlib import pyplot as plt
import pandas as pd
from pylab import *
```

In [797]:
```
data = np.loadtxt('170470.txt', skiprows=1)
data1 = data[:,0]
N = len(data1)
KS = {}
AD = {}
Chi2 = {}
AIC = {}
BIC = {}
para = {}
```

Functions For Testing

In [798]:
```
def ks_test(data,distribution,args):
    stat_value , p_value = stats.kstest(data,distribution,args)
    significance = 0.05
    if(p_value < significance):
#         print("KS_test : H0 is rejected : P_value(=","%.3f" % p_value,") < α(=",significance, ")")
        return p_value ,stat_value
    else:
#         print("KS_test : H0 is accepted : P_value(=","%.3f" % p_value,") > α(=",significance, ")")
        return p_value , stat_value

def anderson_test(data, distribution):
    significance = 0.05
    data = sorted(data)
    if(distribution=='pearson3'):
        A_star = stats.anderson(data,distribution)
        A_star,temp1,temp2 =A_star
    else:
        A_star , C_alpha , alpha_val = stats.anderson(data,distribution)

    if ((A_star) >= 0.00 and (A_star) < 0.200):
            P = 1 - np.exp(-13.436 + 101.14*(A_star) - 223.73*((A_star)**2))
    if ((A_star) >= 0.200 and (A_star) < 0.340):
        P = 1 - np.exp(-8.318 + 42.796*(A_star) - 59.938*((A_star)**2))
    if ((A_star) >= 0.340 and (A_star) < 0.600):
        P = np.exp(0.9177 - 4.279*(A_star) - 1.38*((A_star)**2))
    if ((A_star) >= 0.600):
        P = np.exp(1.2937 - 5.709*(A_star) + 0.0186*((A_star)**2))
    if(distribution =='pearson3'):
        if(P  > significance):
#             print("Anderson_test:  H0 is accepted : P(=","%.3f" % P,") > α(=",significance, ")")
            return P, A_star
        else:
#             print("Anderson_test: H0 is rejected : P(=","%.3f" % P,") < α(=",significance, ")")
            return P, A_star

    else:
        if(A_star  > C_alpha[2]):
#             print("Anderson_test:  H0 is rejected : A*(=","%.3f" % A_star,") > Cα(=",C_alpha[2], ") &  P_value
=","%.3f" % P  )
            return P, A_star , C_alpha[2]
        else:
#             print("Anderson_test: H0 is accepted : A*(=","%.3f" % A_star,") < Cα(=",C_alpha[2], ") &  P_value =",
"%.3f" % P )
            return P, A_star , C_alpha[2]


def chi2(dataset,dist):
    lower_val = []
    upper_val = []
    p = []
    e = []
    qc = pd.qcut(dataset,q = 20,precision=0)
    n = qc.value_counts().to_list()
    ppp = qc.categories
    for k in range(len(qc.categories)):
        lower_val.append(ppp[k].left)
        upper_val.append(ppp[k].right)
        if(dist=='norm'):
            loc,scale = stats.norm.fit(dataset)
            p.append( stats.norm.cdf(upper_val[k],loc,scale) - stats.norm.cdf(lower_val[k],loc,scale ))
        if(dist=='expon'):
            loc,scale = stats.expon.fit(dataset,floc=0)
            p.append(stats.expon.cdf(upper_val[k],loc,scale) - stats.expon.cdf(lower_val[k],loc,scale ))
        if(dist=='lognorm'):
            shape,loc,scale = stats.lognorm.fit(dataset,floc=0)
            p.append( stats.lognorm.cdf(float(upper_val[k]),shape,loc,scale) - stats.lognorm.cdf(lower_val[k],shap
e,loc,scale ))
        if(dist=='gamma'):
            shape,loc,scale = stats.gamma.fit(dataset,floc=0)
            p.append(stats.gamma.cdf(upper_val[k],shape,loc,scale) - stats.gamma.cdf(lower_val[k],shape,loc,scale
))
        if(dist=='gumbel_r'):
            loc,scale = stats.gumbel_r.fit(dataset)
            p.append( stats.gumbel_r.cdf(upper_val[k],loc,scale) - stats.gumbel_r.cdf(lower_val[k],loc,scale ))
        if(dist=='pearson3'):
            skew,loc,scale = stats.pearson3.fit(dataset)
```

```
            skew,loc,scale = stats.pearson3.fit(dataset)
            p.append(stats.pearson3.cdf(upper_val[k],skew,loc,scale) - stats.pearson3.cdf(lower_val[k],skew,loc,sca
le ))

        e.append(p[k]*N)
    stat_value , p_value = stats.chisquare(n,e,ddof = 2)
    significance = 0.05
    if(p_value < significance):
#         print("Chi2_test : H0 is rejected : P_value(=","%.3f" % p_value,") < α(=",significance, ")")
        return p_value ,stat_value
    else:
#         print("Chi2_test : H0 is accepted : P_value(=","%.3f" % p_value,") > α(=",significance, ")")
        return p_value,stat_value
```

Distribution Functions

In [799]:

```
def norm(data1):
    mu, std = stats.norm.fit(data1)
    args = mu , std
    logLik = np.sum(stats.norm.logpdf(data1,mu,std) )
    xmax, xmin = np.max(data1) + 200 , np.min(data1) - 200
    x = np.linspace(xmin, xmax, 100)
    p = stats.norm.pdf(x, mu, std)
    para["norm"] = args
    AD["norm"] = anderson_test(data1,'norm')
    KS["norm"] = ks_test(data1,'norm',args)
    Chi2["norm"] = chi2(data1,'norm')
    AIC["norm"] = -2*logLik + 2*len(args)
    BIC["norm"] = -2*logLik + np.log(N)*len(args)
    return p , args, mu , std
def expon(data1):
    loc , scale = stats.expon.fit(data1,floc=0)
    args = loc , scale
    mean, var, skew, kurt = stats.expon.stats(loc,scale,moments='mvsk')
    xmax, xmin = np.max(data1) + 200 , np.min(data1) - 200
    x = np.linspace(0, xmax, 100)
    p = stats.expon.pdf(x,loc,scale)
    logLik = np.sum(stats.expon.logpdf(data1,loc,scale) )
    para["expon"] = args
    AD["expon"] = anderson_test(data1,'expon')
    KS["expon"] = ks_test(data1,'expon',args)
    Chi2["expon"] = chi2(data1,'expon')
    AIC["expon"] = -2*logLik + 2*len(args)
    BIC["expon"] = -2*logLik + np.log(N)*len(args)
    return p, args, mean , var**0.5
def lognorm(data1):
    xmax, xmin = np.max(data1) + 200 , np.min(data1) - 200
    x = np.linspace(xmin, xmax, 100)
    args = stats.lognorm.fit(data1,floc=0)
    shape, loc, scale = args
    mean, var, skew, kurt = stats.lognorm.stats(shape,loc,scale,moments='mvsk')
    p = stats.lognorm.pdf(x,shape,loc,scale)
    logLik = np.sum(stats.lognorm.logpdf(data1,shape,loc,scale) )
    para["lognorm"] = args
    AD["lognorm"] = anderson_test(data1,'lognorm')
    KS["lognorm"] = ks_test(data1,'lognorm',args)
    Chi2["lognorm"] = chi2(data1,'lognorm')
    AIC["lognorm"] = -2*logLik + 2*len(args)
    BIC["lognorm"] = -2*logLik + np.log(N)*len(args)
    return p, args, mean , var**0.5
def gamma(data1):
    xmax, xmin = np.max(data1) + 200 , np.min(data1) - 200
    x = np.linspace(xmin, xmax, 100)
    args = stats.gamma.fit(data1,floc=0)
    shp, locs, scle = args
    mean, var, skew, kurt = stats.gamma.stats(shp,locs,scle,moments='mvsk')
    p = stats.gamma.pdf(x,shp,locs,scle)
    logLik = np.sum(stats.gamma.logpdf(data1,shp,locs,scle) )
    para["gamma"] = args
    AD["gamma"] = anderson_test((data1),'gamma')
    KS["gamma"] = ks_test(data1,'gamma',args)
    Chi2["gamma"] = chi2(data1,'gamma')
    AIC["gamma"] = -2*logLik + 2*len(args)
    BIC["gamma"] = -2*logLik + np.log(N)*len(args)
    return p, args, mean , var**0.5
def gumbel_r(data1):
    xmax, xmin = np.max(data1) + 200 , np.min(data1) - 200
    x = np.linspace(xmin, xmax, 100)
    args = stats.gumbel_r.fit(data1)
    loc, scale = args
    mean, var, skew, kurt = stats.gumbel_r.stats(loc,scale,moments='mvsk')
    p = stats.gumbel_r.pdf(x,loc,scale)
    logLik = np.sum(stats.gumbel_r.logpdf(data1,loc,scale))
    para["gumbel_r"] = args
    AD["gumbel_r"] = anderson_test(data1,'gumbel_r')
    KS["gumbel_r"] = ks_test(data1,'gumbel_r',args)
    Chi2["gumbel_r"] = chi2(data1,'gumbel_r')
    AIC["gumbel_r"] = -2*logLik + 2*len(args)
    BIC["gumbel_r"] = -2*logLik + np.log(N)*len(args)
    return p, args, mean , var**0.5
def pearson3(data1):
    xmax, xmin = np.max(data1) + 200 , np.min(data1) - 200
    x = np.linspace(xmin, xmax, 100)
    args = stats.pearson3.fit(data1)
    skew ,loc, scale = args
    mean, var, skew, kurt = stats.pearson3.stats(skew,loc,scale,moments='mvsk')
    p = stats.pearson3.pdf(x,skew,loc,scale)
    logLik = np.sum(stats.pearson3.logpdf(data1,skew,loc,scale) )
    para["pearson3"] = args
```

```
    AD["pearson3"] = anderson_test(data1,'pearson3')
    KS["pearson3"] = ks_test(data1,'pearson3',args)
    Chi2["pearson3"] = chi2(data1,'pearson3')
    AIC["pearson3"] = -2*logLik + 2*len(args)
    BIC["pearson3"] = -2*logLik + np.log(N)*len(args)
    return p, args , mean , var**0.5
```

In [800]:
```
def Plot():
    plt.figure(figsize=(20,12))
    plt.subplot(2,3,1)
    plt.hist(data1, bins=25, density=True, alpha=0.6, color='g')
    xmin, xmax = plt.xlim()
    x = np.linspace(xmin, xmax, 100)
    plt.plot(x, norm(data1)[0], 'k', linewidth=2)
    title = "Normal: mu = %.2f,  std = %.2f" % (norm(data1)[2],norm(data1)[3])
    plt.title(title)

    plt.subplot(2,3,2)
    plt.hist(data1, bins=25, density=True, alpha=0.6, color='g')
    plt.plot(x, expon(data1)[0], 'k', linewidth=2)
    title = "Exponential: mu = %.2f,  std = %.2f" % (expon(data1)[2],expon(data1)[3])
    plt.title(title)

    plt.subplot(2,3,3)
    plt.hist(data1, bins=25, density=True, alpha=0.6, color='g')
    plt.plot(x, lognorm(data1)[0], 'k', linewidth=2)
    title = "Lognormal: mu = %.2f,  std = %.2f" % (lognorm(data1)[2],lognorm(data1)[3])
    plt.title(title)

    plt.subplot(2,3,4)
    plt.hist(data1, bins=25, density=True, alpha=0.6, color='g')
    plt.plot(x, gamma(data1)[0], 'k', linewidth=2)
    title = "Gamma: mu = %.2f,  std = %.2f" % (gamma(data1)[2],gamma(data1)[3])
    plt.title(title)

    plt.subplot(2,3,5)
    plt.hist(data1, bins=25, density=True, alpha=0.6, color='g')
    plt.plot(x, gumbel_r(data1)[0], 'k', linewidth=2)
    title = "Gumbel: mu = %.2f,  std = %.2f" % (gumbel_r(data1)[2],gumbel_r(data1)[3])
    plt.title(title)

    plt.subplot(2,3,6)
    plt.hist(data1, bins=25, density=True, alpha=0.6, color='g')
    xmin,xmax = plt.xlim()
    x = np.linspace(xmin, xmax, 100)
    plt.plot(x, pearson3(data1)[0], 'k', linewidth=2)
    title = "Pearson3: mu = %.2f,  std = %.2f" % (pearson3(data1)[2],pearson3(data1)[3])
    plt.title(title)
    plt.show()
```

Finding design discharge for return period

In [801]:
```
def cdf2x(data1,dist):
    data2 =  sorted(data1)

    if(dist=='norm'):
        mu, std = stats.norm.fit(data1)
        temp = stats.norm.cdf(data2, mu, std)
    if(dist=='expon'):
        mu , std = stats.expon.fit(data1,floc=0)
        temp = stats.expon.cdf(data2, mu, std)
    if(dist=='lognorm'):
        args = stats.lognorm.fit(data1,floc=0)
        shape, loc, scale = args
        temp = stats.lognorm.cdf(data2,shape, loc, scale)
    if(dist=='gamma'):
        args = stats.gamma.fit(data1,floc=0)
        shp, locs, scle = args
        temp = stats.gamma.cdf(data2,shp, locs, scle)
    if(dist=='gumbel_r'):
        args = stats.gumbel_r.fit(data1)
        loc, scale = args
        temp = stats.gumbel_r.cdf(data2, loc, scale)
    if(dist=='pearson3'):
        args = stats.pearson3.fit(data1)
        skew ,loc, scale = args
        temp = stats.pearson3.cdf(data2, skew ,loc, scale)
    cnt = 0
    for i in temp:
        cnt = cnt+1
        if(i>49/50):
            break
    cnt = cnt-1
    temp_cdf = data2[cnt] - (data2[cnt]-data2[cnt-1])*(temp[cnt]-0.98)/(temp[cnt]-temp[cnt-1])
    return temp_cdf
```

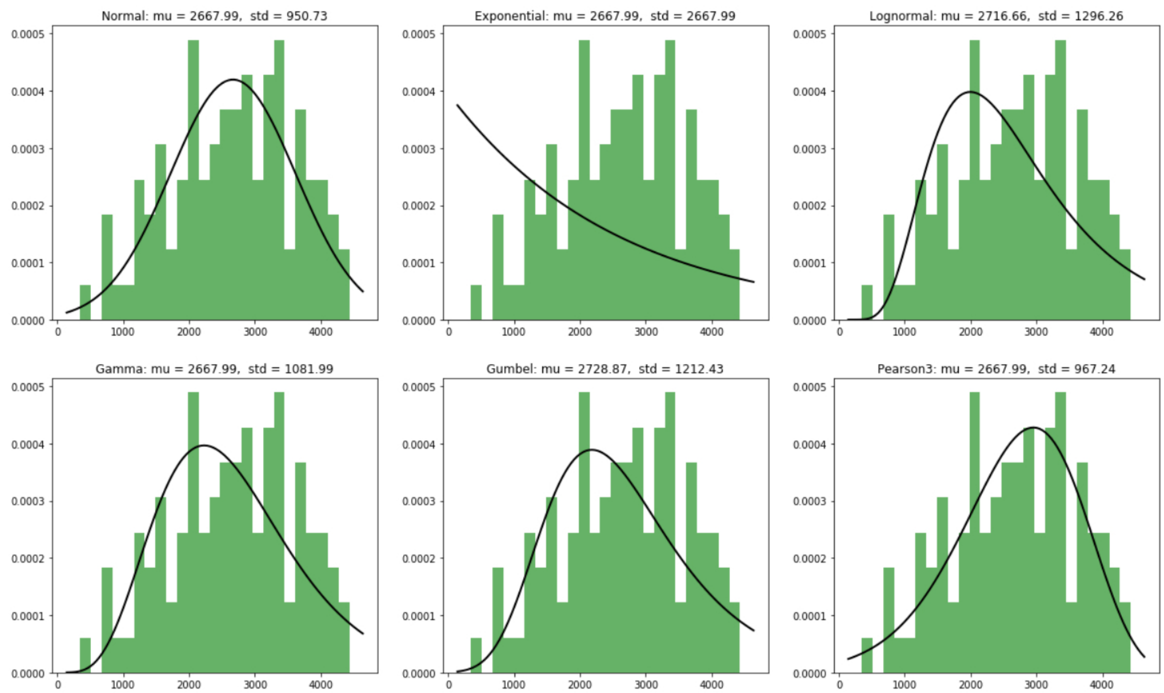
Print Functions

In [802]:
```
def Print(dist,distribution):
    print("Distribution :",'\x1b[1;31m'+distribution+'\x1b[0m')
    print("Mean = ","%.4f" % dist(data1)[2])
    print("Standard Deviation = ", "%.4f" %dist(data1)[3])
    print("Parameters = ", [float("{0:.4f}".format(n)) for n in dist(data1)[1]])
    if(dist == pearson3):
        print("Anderson Test:","Accept" if AD[distribution][0]>0.05 else "Reject", "\n    statistical Value (A2*) =
","%.4f" %AD[distribution][1],"\n    p Value = ","%.4f" %AD[distribution][0])
```

```
    else:
        print("Anderson Test:","Accept" if AD[distribution][2]>AD[distribution][1] else "Reject", "\n    statistica
l Value (A2*) = ","%.4f" %AD[distribution][1],"\n    Critical Value (α) = ","%.4f" %AD[distribution][2],"\n    p Va
lue = ","%.4f" %AD[distribution][0])
    print("KS Test:","Accept" if KS[distribution][0]>0.05 else "Reject", "\n    P value = ","%.4f" %KS[distributio
n][0],"\n    Stat value = ","%.4f" %KS[distribution][1])
    print("Chi_square Test:","Accept" if Chi2[distribution][0]>0.05 else "Reject", " \n    P value = ","%.4f" %Chi2
[distribution][0]," \n    Stat value = ","%.4f" %Chi2[distribution][1])
    print("AIC = ","%.4f" %AIC[distribution])
    print("BIC = ","%.4f" %BIC[distribution])
    print("\n\n")

def Best_print():
    print("Best fitted Distribution: ")
    print( "According to AD test:",AD_accept[0][0])
    print( "According to KS test:",KS_accept[0][0])
    print( "According to Chi2 test:",Chi2_accept[0][0])
    print( "According to AIC Value:",AIC[0][0])
    print( "According to BIC Value:",BIC[0][0])
```

In [803]:

```
Plot()
Print(norm,'norm')
Print(expon,'expon')
Print(lognorm,'lognorm')
Print(gamma,'gamma')
Print(gumbel_r,'gumbel_r')
Print(pearson3,'pearson3')
AD_accept = dict((k, v) for k, v in AD.items() if (v[0] >= 0.05))
AD_accept = sorted(AD_accept.items(),reverse=True,  key=lambda x: x[1][0])

KS_accept = dict((k, v) for k, v in KS.items() if v[0] >= 0.05)
KS_accept = sorted(KS_accept.items(),reverse=True,  key=lambda x: x[1][0])
KS_best = dict((k, v) for k, v in KS.items() if v==max(KS.values()))

Chi2_accept = dict((k, v) for k, v in Chi2.items() if v[0] >= 0.05)
Chi2_accept = sorted(Chi2_accept.items(),reverse=True,  key=lambda x: x[1][0])
Chi2_best = dict((k, v) for k, v in Chi2.items() if v==max(Chi2.values()))
AIC = sorted(AIC.items(),reverse=False,  key=lambda x: x[1])
BIC = sorted(BIC.items(),reverse=False,  key=lambda x: x[1])
Best_print()
print("\n\nDesigned Discharge for 50 year return period =","%.4f" %cdf2x(data1,AIC[0][0]))
```




```
Distribution : norm
Mean =  2667.9877
Standard Deviation =  950.7308
Parameters =  [2667.9877, 950.7308]
Anderson Test: Accept
    statistical Value (A2*) =  0.4650
    Critical Value (α) =  0.7590
    p Value =  0.2541
KS Test: Accept
    P value =  0.6424
    Stat value =  0.0737
Chi_square Test: Accept
    P value =  0.5993
    Stat value =  14.9467
AIC =  1659.2339
BIC =  1664.4442



Distribution : expon
Mean =  2667.9877
Standard Deviation =  2667.9877
Parameters =  [0.0, 2667.9877]
```

```
Anderson Test: Reject
    statistical Value (A2*) =  18.0862
    Critical Value (α) =  1.3330
    p Value =  0.0000
KS Test: Reject
    P value =  0.0000
    Stat value =  0.3103
Chi_square Test: Reject
    P value =  0.0000
    Stat value =  119.3418
AIC =  1781.8160
BIC =  1787.0263



Distribution : lognorm
Mean =  2716.6622
Standard Deviation =  1296.2593
Parameters =  [0.4529, 0.0, 2451.8517]
Anderson Test: Reject
    statistical Value (A2*) =  2.5645
    Critical Value (α) =  0.7590
    p Value =  0.0000
KS Test: Accept
    P value =  0.1049
    Stat value =  0.1197
Chi_square Test: Reject
    P value =  0.0346
    Stat value =  28.9920
AIC =  1692.2919
BIC =  1700.1074



Distribution : gamma
Mean =  2667.9877
Standard Deviation =  1081.9860
Parameters =  [6.0803, 0.0, 438.7928]
Anderson Test: Accept
    statistical Value (A2*) =  1.5178
    Critical Value (α) =  2.4120
    p Value =  0.0007
KS Test: Accept
    P value =  0.2368
    Stat value =  0.1016
Chi_square Test: Accept
    P value =  0.2265
    Stat value =  20.9950
AIC =  1675.6746
BIC =  1683.4901



Distribution : gumbel_r
Mean =  2728.8698
Standard Deviation =  1212.4304
Parameters =  [2183.2117, 945.3281]
Anderson Test: Reject
    statistical Value (A2*) =  1.6235
    Critical Value (α) =  0.7420
    p Value =  0.0004
KS Test: Accept
    P value =  0.2929
    Stat value =  0.0964
Chi_square Test: Accept
    P value =  0.1746
    Stat value =  22.2676
AIC =  1676.8689
BIC =  1682.0793



Distribution : pearson3
Mean =  2667.9876
Standard Deviation =  967.2408
Parameters =  [-0.5881, 2667.9876, 967.2408]
Anderson Test: Reject
    statistical Value (A2*) =  200.4932
    p Value =  0.0000
KS Test: Reject
    P value =  0.0000
    Stat value =  0.9867
Chi_square Test: Accept
    P value =  1.0000
    Stat value =  -414.7810
AIC =  1658.7865
BIC =  1666.6020



Best fitted Distribution:
According to AD test: norm
According to KS test: norm
According to Chi2 test: pearson3
According to AIC Value: pearson3
According to BIC Value: norm


Designed Discharge for 50 year return period = 353.1010
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: