Post Graduate Project

Bayesian Framework of Geomteric Skew Normal Distribution

Submitted by:

Peshal Agarwal

Roll No: 13817472
MTH697A, MTH698A

Guide:

Prof. Debasis Kundu

Signature

1 Introduction

Distributions with better ability to represent data with minimal assumptions has always attracted statisticians. Over past few years, skew-normal distribution has been increasingly popular as we know that a lot of data is skewed these days. One of early developments was by Prof. Azzalini[2] who defined the skew normal distribution as follows

$$f(x;\lambda) = 2\phi(x)\Phi(\lambda x) \qquad x \in R, \quad \lambda \in R$$

Note that it is a proper density function because if we consider two random variables X and Y following standing normal distribution and any real number λ then,

$$P(Y \leq \lambda X) = \frac{1}{2} = \int_{-\infty}^{\infty} \phi(x)\Phi(\lambda x)dx$$

We can add the location and scale parameters to the distribution, and the density function changes as follows

$$f(x;\lambda,\mu,\sigma) = \frac{2}{\sigma}\phi\Big(\frac{x-\mu}{\sigma}\Big)\Phi\Big(\frac{\lambda(x-\mu)}{\sigma}\Big)$$

Here, λ is the skewness parameter and the pdf can take a variety of shapes, however, there are certain issues. The estimation of parameters is not easy, it cannot have a heavy tail and MLE estimates may not always exists. But it can be extended easily to multivariate setting.

1.1 Power Normal Distribution

The distrubtion of Power normal distribution looks like the following

$$F(x;\alpha) = [\Phi(x)]^{\alpha}$$

where $\alpha(> 0)$ is the skewness parameter.
Similar to skew normal, in this distribution also we can have the scale and location parameter, and the density function has the following form

$$f(x;\alpha,\mu\,\sigma) = \frac{\alpha}{\sigma}\Big[\Phi\Big(\frac{x-\mu}{\sigma}\Big)\Big]^{\alpha-1}\phi\Big(\frac{x-\mu}{\sigma}\Big)$$

Note that, for this distribution even though the MLE estimates of the parameters would always exists but are non-trivial to calculate. Moreover, it can't be used to analyze heavy tail data.

1.2 Geometric Skew Normal Distribution

Now we look at a three parameter skew distribution which does not have the issues present in the distributions discussed above.
Consider the random variabel N having Geometric distribution with parameter p and let $X_1, X_2, \ldots$ be *i.i.d* random variables following standard normal distribution. Now define a random variable X as follows

$$X = \sum_{i=0}^{N} X_i$$

We call X to follow geometric skew distribution (GSN). The pdf with scale and location parameter turns out to be the following

$$f(x;\mu,\sigma,p)=\sum_{k=1}^{\infty}\frac{p}{\sigma\sqrt{k}}\phi\Big(\frac{x-k\mu}{\sigma\sqrt{k}}\Big)(1-p)^{k-1}$$

With decrease in p the variance as well as the tail probabilities increase.

2 Characteristics

The expression of mean and variance are as follows

$$E(X)=\frac{\mu}{p} \qquad Var(X)=\frac{\sigma^2p+\mu^2(1-p)}{p^2}$$

$$E(X^m)=p\sum_{n=1}^{\infty}(1-p)^{n-1}c_m(n\mu,n\sigma^2)$$

$$\text{where} \quad c_m(n\mu,n\sigma^2)=E(Y^m), \quad \text{where} \quad Y{\sim}N(n\mu,n\sigma^2)$$

The distribution has the property of infinite divisibility. It is also geometric stable. Consider X_i's iid GSN(μ,σ,p) and $M\sim GE(q)$ both independent where $0<q<1$. Consider the random variable X s.t.

$$X=\sum_{i=1}^{M}X_i$$

Then it turns out that $X\sim GSN(\mu,\sigma,pq)$
Also, the $X\sim GSN(\mu,\sigma,p)$ can be decomposed in the following manner

$$X=Y+\sum_{i=1}^{Q}Y_i$$

where Q ~Poisson(λ) and $Y\sim N(\mu,\sigma^2)$. Moreover, $Y_i|Z_i\sim N(\mu Z_i,\sigma^2 Z_i)$ such that

$$P(Z_i=k)=\frac{(1-p)^k}{\lambda k}$$

$$\text{where} \quad k\in \quad \text{and} \quad \lambda=-lnp$$

3 Bayesian Inference

3.1 Objective

We want to incorporate prior information in order to get better estimates of the parameters, thereby leading to a better fit to the data. We use joint distribution of x and N instead of marginal of X, because the calculations become intractable in case of marginal distribution of X. We take suitable priors on each the parameters simultaneously and find their MAP estimates in an iterative manner. Earlier research work has also assumed priors, but none contains priors on all the three-parameters at the same time. We would use statistical tests to measure the goodness of fit.

3.2 The Likelihood and the priors

The joint PDF of X and N is where $X \sim GSN(x;\mu,\sigma,p)$ and $N \sim GE(p)$ is given by

$$f_{X,N}(x,n) = \frac{1}{\sigma\sqrt{2\pi n}} exp\Big(-\frac{1}{2n\sigma^2}(x-n\mu)^2\Big)p(1-p)^{n-1}$$

We will take beta prior on p, Gaussian prior on μ and inverse gamma prior on σ^2, all independent of each other

$$\begin{aligned} P(p) &= \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1} \\ P(\mu) &= \frac{1}{\sqrt{2\pi b^2}} exp\Big(-\frac{1}{2b^2}(\mu-a)^2\Big) \\ P(\sigma^2) &= \frac{d^c}{\Gamma(c)}(\sigma^2)^{(-c-1)}e^{\frac{-d}{\sigma^2}} \end{aligned}$$

3.3 Posterior

$$\Pi(\mu,\sigma,p|\{x_i\},\{m_i\}) \propto \prod_{i=1}^{n}\Big(\frac{1}{\sigma}exp\Big(-\frac{1}{2m_i\sigma^2}(x_i-m_i\mu)^2\Big)p(1-p)^{m_i-1}\Big)p^{\alpha-1}(1-p)^{\beta-1}$$
$$exp\Big(-\frac{1}{2b^2}(\mu-a)^2\Big)(\sigma^2)^{(-c-1)}exp\Big(\frac{-d}{\sigma^2}\Big)$$

$\Pi(\mu,\sigma,p|\{x_i\},\{m_i\}) = \Pi(\mu,\sigma,\mid\{x_i\},\{m_i\})\Pi(p\mid\{x_i\},\{m_i\})$

$$\begin{aligned} \Pi(\mu\mid\sigma,\{x_i\},\{m_i\}) &\sim Normal(\hat{a},\hat{b}^2) \\ \Pi(\sigma^2\mid\mu,\{x_i\},\{m_i\}) &\sim InvGamma(c+n/2,d+K/2) \\ \Pi(p\mid\{x_i\},\{m_i\}) &\sim Beta(n+\alpha,\sum_{i=1}^{n}m_i-n+\beta) \end{aligned}$$

where

$$\begin{aligned} \widehat{a} &= \Big(\frac{\sum_{i=1}^n x_i}{\sigma^2}+\frac{a}{b^2}\Big)\Big/\Big(\frac{\sum_{i=1}^n m_i}{\sigma^2}+\frac{1}{b^2}\Big) \\ \widehat{b}^2 &= \Big(\frac{\sum_{i=1}^n m_i}{\sigma^2}+\frac{1}{b^2}\Big)^{-1} \\ K &= \sum_{i=1}^{n}\frac{(x_i-m_i\mu)^2}{m_i} \end{aligned}$$

We will have to choose appropriate values of all hyper-parameters and sample μ,σ,p using the above distributions in an iterative manner.
We will make use of the following equation

$$\begin{aligned} E(N\mid X=x,\mu,\sigma,p) &= \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}exp(-\frac{1}{2\sigma^2 n}(x-n\mu)^2)\sqrt{n}}{\sum_{k=1}^{\infty}(1-p)^{k-1}exp(-\frac{1}{2\sigma^2 k}(x-k\mu)^2)/\sqrt{k}} \\ E(N^{-1}\mid X=x,\mu,\sigma,p) &= \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}exp(-\frac{1}{2\sigma^2 n}(x-n\mu)^2)/n^{3/2}}{\sum_{k=1}^{\infty}(1-p)^{k-1}exp(-\frac{1}{2\sigma^2 k}(x-k\mu)^2)/\sqrt{k}} \end{aligned}$$

The iterative algorithm to find m_i's by sampling from the posterior

1. Initialize μ^0, σ^0, p^0
2. For $t = 0, 1, 2 \dots$
   (a) $m_{i,t} = E(N \mid X = x_i, \mu^{(t)}, \sigma^{(t)}, p^{(t)})$ for each $i \in [n]$
   (b) $l_{i,t} = E(N^{-1} \mid X = x, \mu, \sigma, p)$
   (c) $\mu^{(t+1)} \sim \Pi(\mu \mid \sigma^{(t)}, \{x_i\}, \{m_{i,t}\})$
   (d) $(\sigma^{(t+1)})^2 \sim \Pi(\sigma^2 \mid \mu^{(t)}, \{x_i\}, \{m_{i,t}\}, \{l_{i,t}\})$
   (e) $p^{(t+1)} \sim \Pi(p \mid \{x_i\}, \{m_{i,t}\})$
3. Repeat until convergence

The iterative algorithm to find m_i's by taking MAP estimates of parameters

1. Initialize μ^0, σ^0, p^0
2. For $t = 0, 1, 2 \dots$
   (a) $m_{i,t} = E(N \mid X = x_i, \mu^{(t)}, \sigma^{(t)}, p^{(t)})$ for each $i \in [n]$
   (b) $l_{i,t} = E(N^{-1} \mid X = x, \mu, \sigma, p)$
   (c) $\mu^{(t+1)} = \hat{a}^t$
   (d) $(\sigma^{(t+1)})^2 = \frac{d+K^t/2}{c+n/2+1}$
   (e) $p^{(t+1)} = \frac{n+\alpha-1}{\sum_{i=1}^n m_{i,t}+\alpha+\beta-2}$
3. Repeat until convergence

where

$$\hat{a}^t = \Big(\frac{\sum_{i=1}^n x_i}{(\sigma^{(t)})^2} + \frac{a}{b^2}\Big) \Big/ \Big(\frac{\sum_{i=1}^n m_{i,t}}{(\sigma^{(t)})^2} + \frac{1}{b^2}\Big)$$

$$K^t = \sum_{i=1}^n x_i^2 l_{i,t} - 2x_i\mu^{(t)} + m_{i,t}(\mu^{(t)})^2$$

If we take try to get the MAP estimate for m_i instead of taking the expected value, then following happens:

$$P(N = n \mid X = x) = \frac{P(X = x, N = n)}{P(X = x)} \propto P(X = x, N = n) = f_{X,N}(x, n)$$

$$f_{X,N}(x, n) = \frac{1}{\sigma\sqrt{2\pi n}} exp\Big(- \frac{1}{2n\sigma^2}(x - n\mu)^2\Big) p(1-p)^{n-1}$$

$$log\ f = -\frac{1}{2} ln\ n - \frac{1}{2\sigma^2}\big(\frac{x^2}{n} + n\mu^2\big) + n\ ln(1-p) + k$$

$$\frac{\partial log\ f}{\partial n} = \frac{1}{\sigma^2}\Big(- \frac{\sigma^2}{n} + \frac{x^2}{n^2} - \mu^2 + 2\sigma^2 ln(1-p)\Big)$$

$$\frac{\partial^2 log\ f}{\partial n^2} = \frac{1}{2n^2} - \frac{x^2}{\sigma^2 n^3}$$

Applying the first order optimality,

$$\frac{\partial log\ f}{\partial n} = 0 \implies cn^2 + \sigma^2 - x^2 = 0$$

where $c = \mu^2 - 2\sigma^2 ln(1-p)$, solving the above quadratic in n we get

$$n = \frac{-\sigma^2 \pm \sqrt{\sigma^4 + 4cx^2}}{2c}$$

since, n cannot take negative values, we take

$$n = \frac{-\sigma^2 + \sqrt{\sigma^4 + 4cx^2}}{2c} > 0$$

But, in order for above n to be a maxima, we must have $\frac{\partial^2 log\ f}{\partial n^2} < 0$ at this value of n.

$$\frac{\partial^2 log\ f}{\partial n^2} < 0 \implies \frac{1}{2} - \frac{x^2}{\sigma^2 n} < 0 \implies n < \frac{2x^2}{\sigma^2}$$

The above condition need not be always satisfied, so the root of quadratic can also be a minima instead of a maxima depending on the values of x, σ, μ, and p. Moreover, if we perform linear search over some finite values (say 1, 2, . . . , 8) and choose the one with highest probability, even then the estimates are farther to the true value compared to the estimates obtained using expected value. Hence using any of these techniques, we don't get good estimates, thus we stick to expected value estimate.

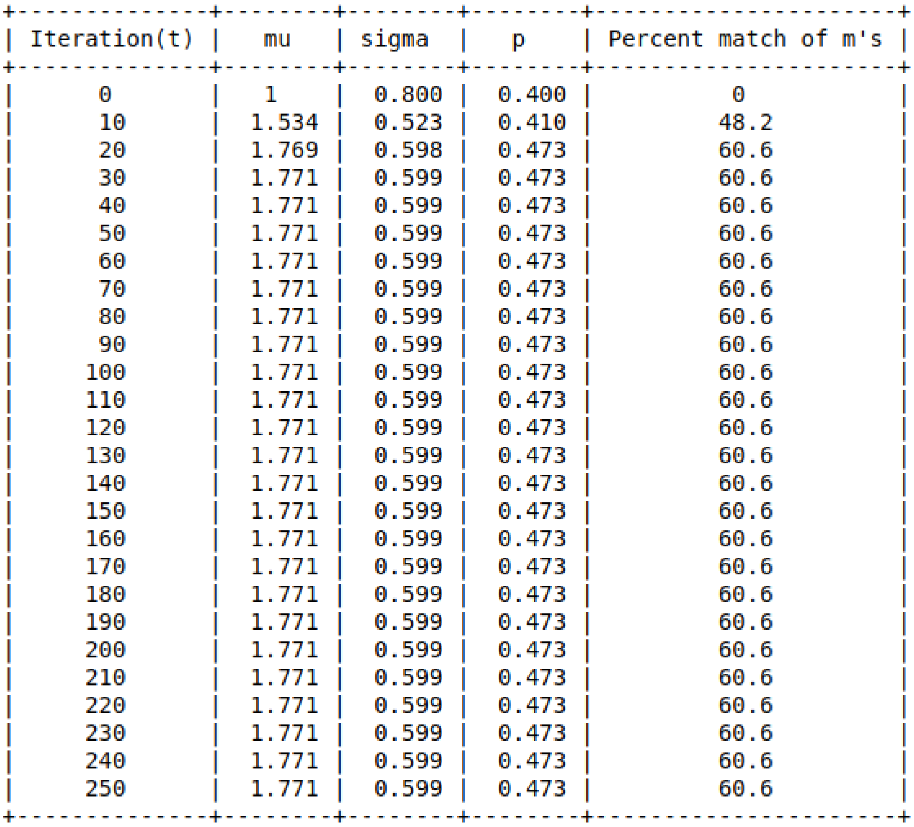
3.4 Simulation

We generate 500 data points generated with the following parameters

$$\mu = 2$$
$$\sigma = 1$$
$$p = 0.5$$

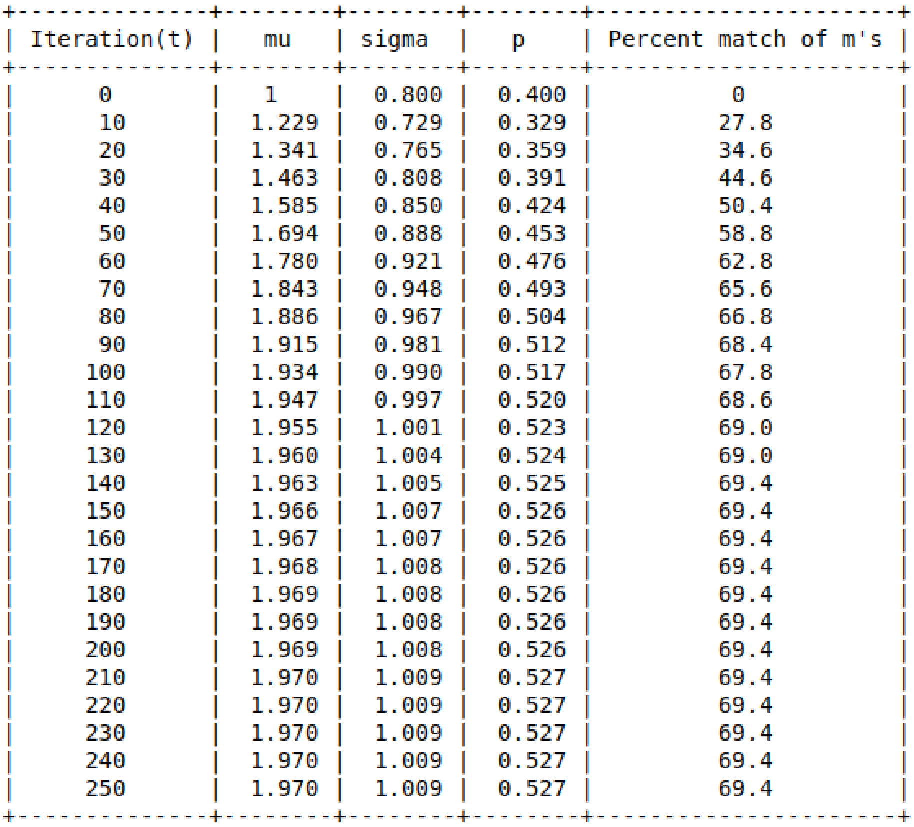
We denote the hyperparameters as follows

α := shape parameter of Beta prior on p
β := scale parameter of Beta prior on p
a := mean of Gaussian prior on μ
b := standard deviation of Gaussain prior on μ
c := shape parameter of Inverse Gamma prior on σ^2
d := scale parameter of Inverse Gamma prior on σ^2



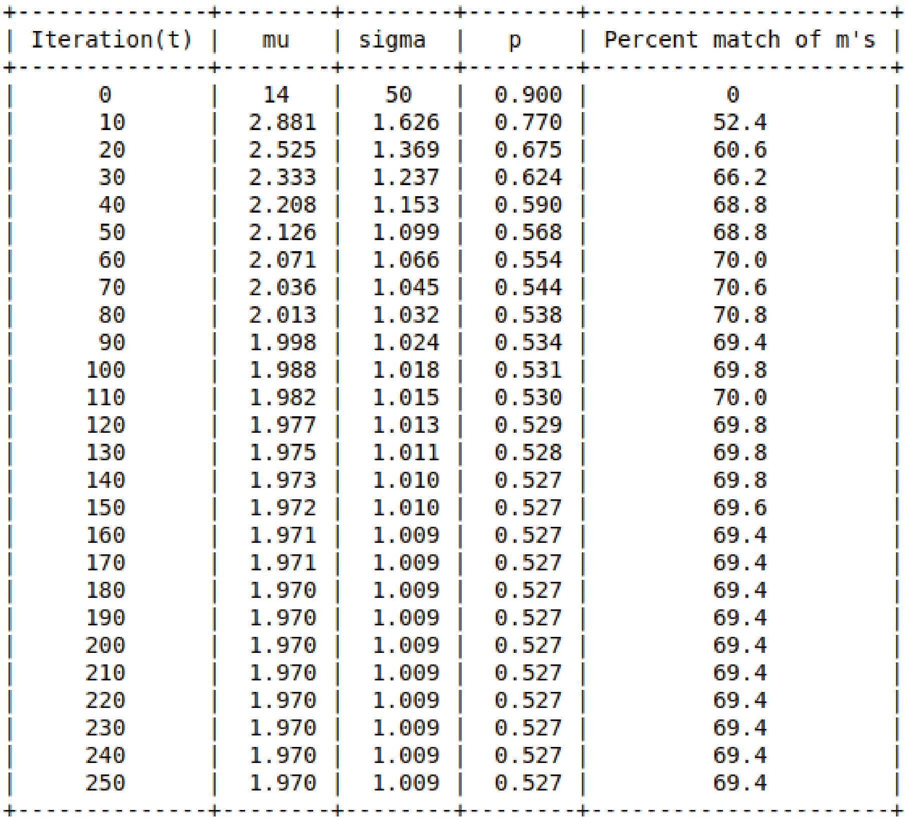
```
+--------------+--------+--------+--------+----------------------+
| Iteration(t) |   mu   | sigma  |   p    | Percent match of m's |
+--------------+--------+--------+--------+----------------------+
|      0       |   1    | 0.800  | 0.400  |          0           |
|      10      | 1.534  | 0.523  | 0.410  |         48.2         |
|      20      | 1.769  | 0.598  | 0.473  |         60.6         |
|      30      | 1.771  | 0.599  | 0.473  |         60.6         |
|      40      | 1.771  | 0.599  | 0.473  |         60.6         |
|      50      | 1.771  | 0.599  | 0.473  |         60.6         |
|      60      | 1.771  | 0.599  | 0.473  |         60.6         |
|      70      | 1.771  | 0.599  | 0.473  |         60.6         |
|      80      | 1.771  | 0.599  | 0.473  |         60.6         |
|      90      | 1.771  | 0.599  | 0.473  |         60.6         |
|     100      | 1.771  | 0.599  | 0.473  |         60.6         |
|     110      | 1.771  | 0.599  | 0.473  |         60.6         |
|     120      | 1.771  | 0.599  | 0.473  |         60.6         |
|     130      | 1.771  | 0.599  | 0.473  |         60.6         |
|     140      | 1.771  | 0.599  | 0.473  |         60.6         |
|     150      | 1.771  | 0.599  | 0.473  |         60.6         |
|     160      | 1.771  | 0.599  | 0.473  |         60.6         |
|     170      | 1.771  | 0.599  | 0.473  |         60.6         |
|     180      | 1.771  | 0.599  | 0.473  |         60.6         |
|     190      | 1.771  | 0.599  | 0.473  |         60.6         |
|     200      | 1.771  | 0.599  | 0.473  |         60.6         |
|     210      | 1.771  | 0.599  | 0.473  |         60.6         |
|     220      | 1.771  | 0.599  | 0.473  |         60.6         |
|     230      | 1.771  | 0.599  | 0.473  |         60.6         |
|     240      | 1.771  | 0.599  | 0.473  |         60.6         |
|     250      | 1.771  | 0.599  | 0.473  |         60.6         |
+--------------+--------+--------+--------+----------------------+
```

Figure 1: $\alpha = 1$, $\beta = 1$, a = 0, b = 100, c = 1, d = 1 when using linear search for optimal m's instead of expected value



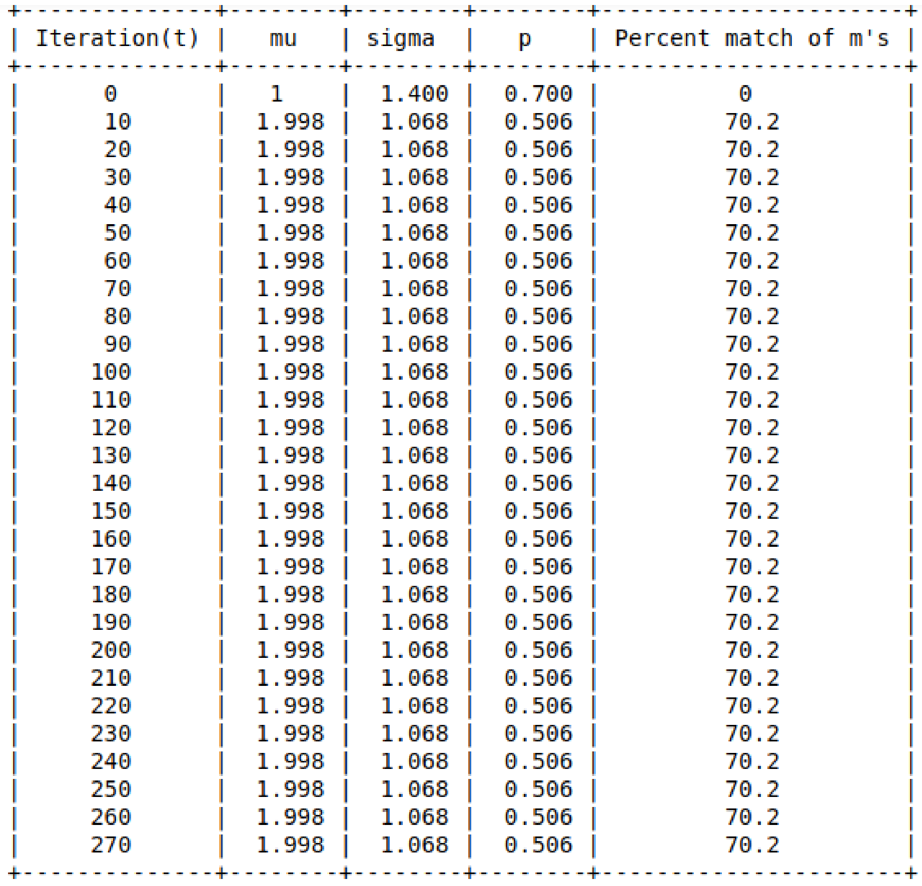
```
+--------------+--------+--------+--------+---------------------+
| Iteration(t) |   mu   | sigma  |   p    | Percent match of m's |
+--------------+--------+--------+--------+---------------------+
|      0       |   1    | 0.800  | 0.400  |         0           |
|      10      | 1.229  | 0.729  | 0.329  |        27.8         |
|      20      | 1.341  | 0.765  | 0.359  |        34.6         |
|      30      | 1.463  | 0.808  | 0.391  |        44.6         |
|      40      | 1.585  | 0.850  | 0.424  |        50.4         |
|      50      | 1.694  | 0.888  | 0.453  |        58.8         |
|      60      | 1.780  | 0.921  | 0.476  |        62.8         |
|      70      | 1.843  | 0.948  | 0.493  |        65.6         |
|      80      | 1.886  | 0.967  | 0.504  |        66.8         |
|      90      | 1.915  | 0.981  | 0.512  |        68.4         |
|     100      | 1.934  | 0.990  | 0.517  |        67.8         |
|     110      | 1.947  | 0.997  | 0.520  |        68.6         |
|     120      | 1.955  | 1.001  | 0.523  |        69.0         |
|     130      | 1.960  | 1.004  | 0.524  |        69.0         |
|     140      | 1.963  | 1.005  | 0.525  |        69.4         |
|     150      | 1.966  | 1.007  | 0.526  |        69.4         |
|     160      | 1.967  | 1.007  | 0.526  |        69.4         |
|     170      | 1.968  | 1.008  | 0.526  |        69.4         |
|     180      | 1.969  | 1.008  | 0.526  |        69.4         |
|     190      | 1.969  | 1.008  | 0.526  |        69.4         |
|     200      | 1.969  | 1.008  | 0.526  |        69.4         |
|     210      | 1.970  | 1.009  | 0.527  |        69.4         |
|     220      | 1.970  | 1.009  | 0.527  |        69.4         |
|     230      | 1.970  | 1.009  | 0.527  |        69.4         |
|     240      | 1.970  | 1.009  | 0.527  |        69.4         |
|     250      | 1.970  | 1.009  | 0.527  |        69.4         |
+--------------+--------+--------+--------+---------------------+
```

Figure 2: $\alpha = 1$, $\beta = 1$, a = 0, b = 100, c = 1, d = 1



```
+--------------+--------+--------+--------+----------------------+
| Iteration(t) |   mu   | sigma  |   p    | Percent match of m's |
+--------------+--------+--------+--------+----------------------+
|      0       |   14   |   50   | 0.900  |          0           |
|      10      | 2.881  | 1.626  | 0.770  |         52.4         |
|      20      | 2.525  | 1.369  | 0.675  |         60.6         |
|      30      | 2.333  | 1.237  | 0.624  |         66.2         |
|      40      | 2.208  | 1.153  | 0.590  |         68.8         |
|      50      | 2.126  | 1.099  | 0.568  |         68.8         |
|      60      | 2.071  | 1.066  | 0.554  |         70.0         |
|      70      | 2.036  | 1.045  | 0.544  |         70.6         |
|      80      | 2.013  | 1.032  | 0.538  |         70.8         |
|      90      | 1.998  | 1.024  | 0.534  |         69.4         |
|     100      | 1.988  | 1.018  | 0.531  |         69.8         |
|     110      | 1.982  | 1.015  | 0.530  |         70.0         |
|     120      | 1.977  | 1.013  | 0.529  |         69.8         |
|     130      | 1.975  | 1.011  | 0.528  |         69.8         |
|     140      | 1.973  | 1.010  | 0.527  |         69.8         |
|     150      | 1.972  | 1.010  | 0.527  |         69.6         |
|     160      | 1.971  | 1.009  | 0.527  |         69.4         |
|     170      | 1.971  | 1.009  | 0.527  |         69.4         |
|     180      | 1.970  | 1.009  | 0.527  |         69.4         |
|     190      | 1.970  | 1.009  | 0.527  |         69.4         |
|     200      | 1.970  | 1.009  | 0.527  |         69.4         |
|     210      | 1.970  | 1.009  | 0.527  |         69.4         |
|     220      | 1.970  | 1.009  | 0.527  |         69.4         |
|     230      | 1.970  | 1.009  | 0.527  |         69.4         |
|     240      | 1.970  | 1.009  | 0.527  |         69.4         |
|     250      | 1.970  | 1.009  | 0.527  |         69.4         |
+--------------+--------+--------+--------+----------------------+
```

Figure 3: $\alpha = 1$, $\beta = 1$, a = 0, b = 100, c = 1, d = 1, when starting with a wild guess

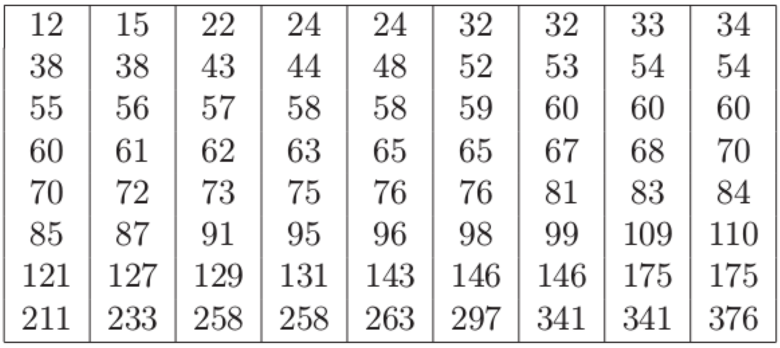


```
+--------------+--------+--------+--------+----------------------+
| Iteration(t) |   mu   | sigma  |   p    | Percent match of m's |
+--------------+--------+--------+--------+----------------------+
|      0       |   1    | 1.400  | 0.700  |          0           |
|      10      | 1.998  | 1.068  | 0.506  |         70.2         |
|      20      | 1.998  | 1.068  | 0.506  |         70.2         |
|      30      | 1.998  | 1.068  | 0.506  |         70.2         |
|      40      | 1.998  | 1.068  | 0.506  |         70.2         |
|      50      | 1.998  | 1.068  | 0.506  |         70.2         |
|      60      | 1.998  | 1.068  | 0.506  |         70.2         |
|      70      | 1.998  | 1.068  | 0.506  |         70.2         |
|      80      | 1.998  | 1.068  | 0.506  |         70.2         |
|      90      | 1.998  | 1.068  | 0.506  |         70.2         |
|     100      | 1.998  | 1.068  | 0.506  |         70.2         |
|     110      | 1.998  | 1.068  | 0.506  |         70.2         |
|     120      | 1.998  | 1.068  | 0.506  |         70.2         |
|     130      | 1.998  | 1.068  | 0.506  |         70.2         |
|     140      | 1.998  | 1.068  | 0.506  |         70.2         |
|     150      | 1.998  | 1.068  | 0.506  |         70.2         |
|     160      | 1.998  | 1.068  | 0.506  |         70.2         |
|     170      | 1.998  | 1.068  | 0.506  |         70.2         |
|     180      | 1.998  | 1.068  | 0.506  |         70.2         |
|     190      | 1.998  | 1.068  | 0.506  |         70.2         |
|     200      | 1.998  | 1.068  | 0.506  |         70.2         |
|     210      | 1.998  | 1.068  | 0.506  |         70.2         |
|     220      | 1.998  | 1.068  | 0.506  |         70.2         |
|     230      | 1.998  | 1.068  | 0.506  |         70.2         |
|     240      | 1.998  | 1.068  | 0.506  |         70.2         |
|     250      | 1.998  | 1.068  | 0.506  |         70.2         |
|     260      | 1.998  | 1.068  | 0.506  |         70.2         |
|     270      | 1.998  | 1.068  | 0.506  |         70.2         |
+--------------+--------+--------+--------+----------------------+
```

Figure 4: $\alpha = 100$, $\beta = 100$, a = 2, b = 0.01, c = 100, d = 101, when starting with an informative prior

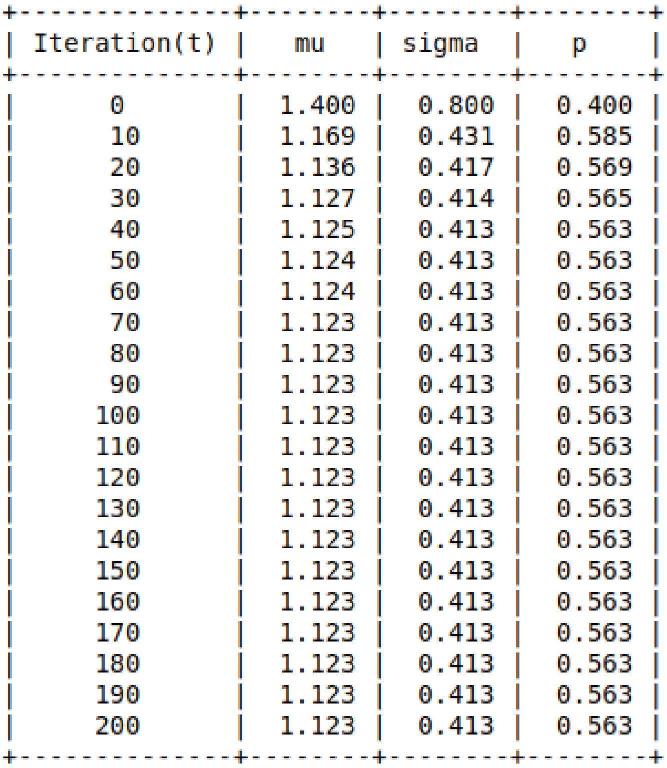
4 Data Analysis

We use the following dataset of 72 observations to test the effectiveness of the model. Without loss of generality, we divide all observations by 50 to ease estimation purposes.



12	15	22	24	24	32	32	33	34
38	38	43	44	48	52	53	54	54
55	56	57	58	58	59	60	60	60
60	61	62	63	65	65	67	68	70
70	72	73	75	76	76	81	83	84
85	87	91	95	96	98	99	109	110
121	127	129	131	143	146	146	175	175
211	233	258	258	263	297	341	341	376

Figure 5: Guinea pig data set



```
+--------------+--------+--------+--------+
| Iteration(t) |   mu   | sigma  |   p    |
+--------------+--------+--------+--------+
|      0       | 1.400 | 0.800 | 0.400 |
|      10      | 1.169 | 0.431 | 0.585 |
|      20      | 1.136 | 0.417 | 0.569 |
|      30      | 1.127 | 0.414 | 0.565 |
|      40      | 1.125 | 0.413 | 0.563 |
|      50      | 1.124 | 0.413 | 0.563 |
|      60      | 1.124 | 0.413 | 0.563 |
|      70      | 1.123 | 0.413 | 0.563 |
|      80      | 1.123 | 0.413 | 0.563 |
|      90      | 1.123 | 0.413 | 0.563 |
|     100      | 1.123 | 0.413 | 0.563 |
|     110      | 1.123 | 0.413 | 0.563 |
|     120      | 1.123 | 0.413 | 0.563 |
|     130      | 1.123 | 0.413 | 0.563 |
|     140      | 1.123 | 0.413 | 0.563 |
|     150      | 1.123 | 0.413 | 0.563 |
|     160      | 1.123 | 0.413 | 0.563 |
|     170      | 1.123 | 0.413 | 0.563 |
|     180      | 1.123 | 0.413 | 0.563 |
|     190      | 1.123 | 0.413 | 0.563 |
|     200      | 1.123 | 0.413 | 0.563 |
+--------------+--------+--------+--------+
```

Figure 6: $\alpha = 1$, $\beta = 1$, a = 0, b = 100, c = 0.1, d = 0.1

As evident from the table above, we get the following estimates

$$\hat{\mu} = 1.123$$
$$\hat{\sigma} = 0.413$$
$$\hat{p} = 0.563$$

The log-likelihood for these estimates turns out to be -107.13
Comparing it with the MLE estimates [1]

$$\hat{\mu} = 1.131$$
$$\hat{\sigma} = 0.589$$
$$\hat{p} = 0.566$$

The log-likelihood for these ML estimates turns out to be -109.92

Estimates	**Log Likelihood**	**KS Test Statistic**	**p-value**
MLE	-109.25	0.112	0.3
MAP	**-107.13**	**0.089**	**0.6**

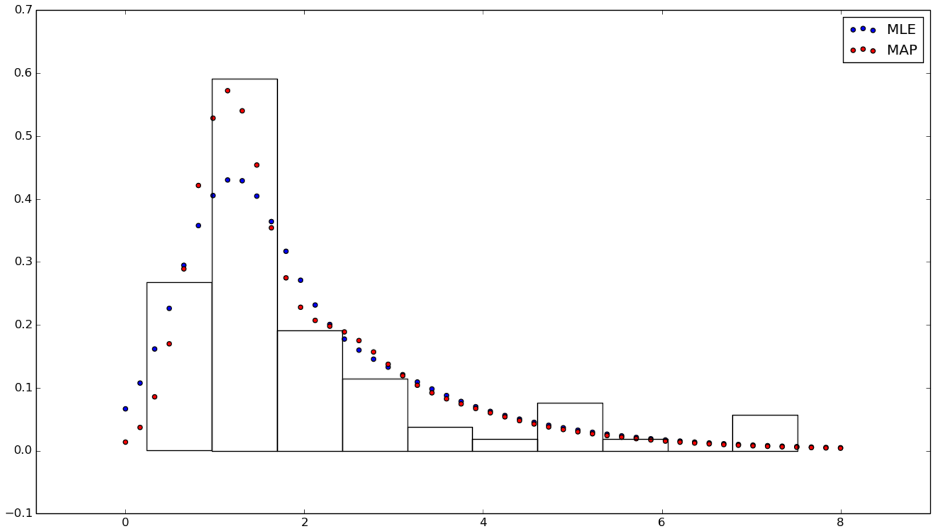


Figure 7: A comparison of probability density function along with histogram, between MLE and MAP estimates

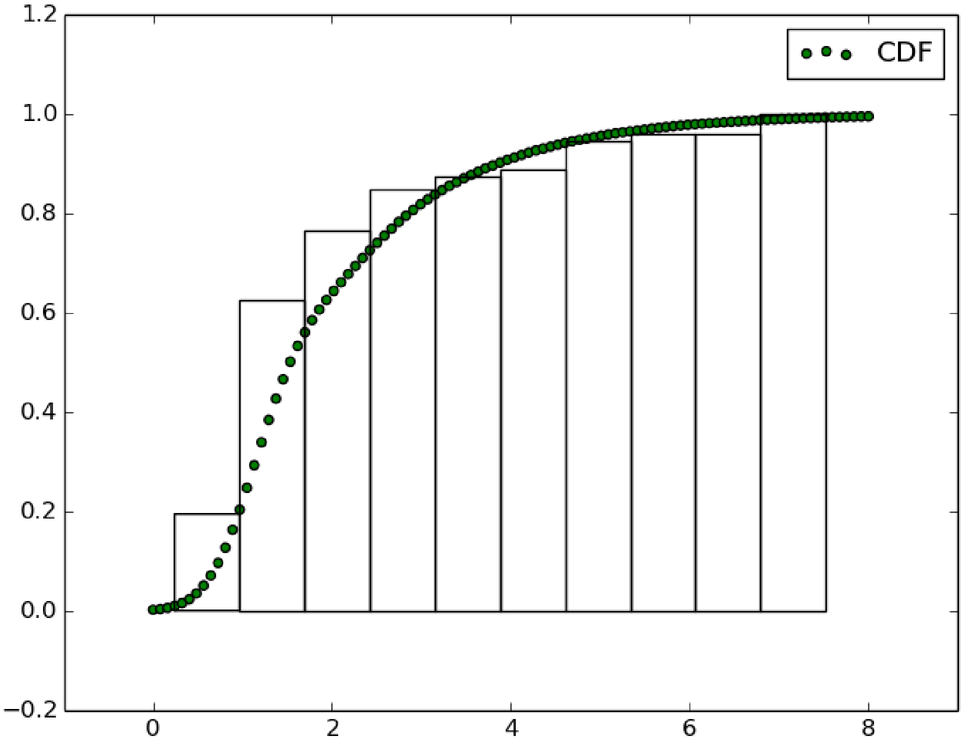
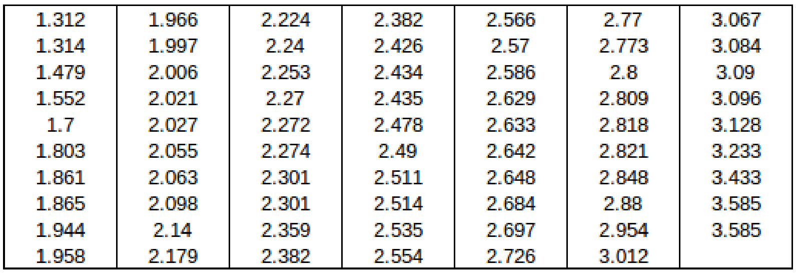


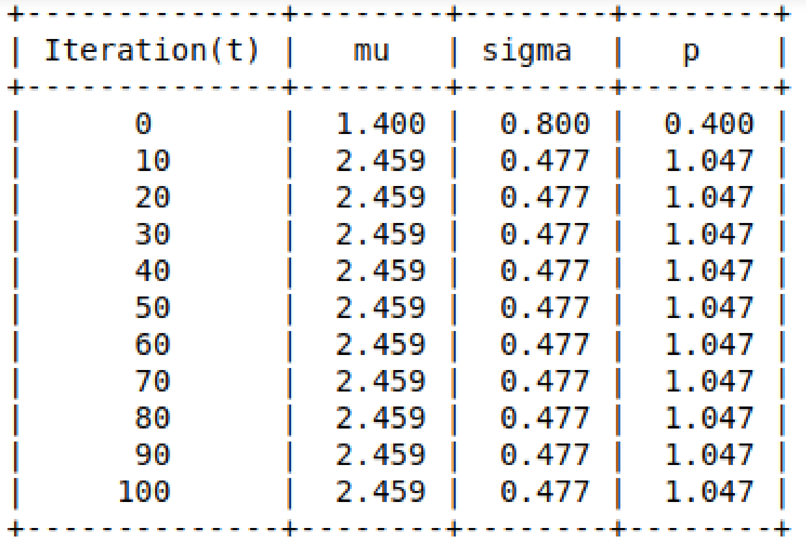
Figure 8: A comparison of probability distribution function with histogram

We also did a similar analysis for another dataset containing 69 observations.



1.312	1.966	2.224	2.382	2.566	2.77	3.067
1.314	1.997	2.24	2.426	2.57	2.773	3.084
1.479	2.006	2.253	2.434	2.586	2.8	3.09
1.552	2.021	2.27	2.435	2.629	2.809	3.096
1.7	2.027	2.272	2.478	2.633	2.818	3.128
1.803	2.055	2.274	2.49	2.642	2.821	3.233
1.861	2.063	2.301	2.511	2.648	2.848	3.433
1.865	2.098	2.301	2.514	2.684	2.88	3.585
1.944	2.14	2.359	2.535	2.697	2.954	3.585
1.958	2.179	2.382	2.554	2.726	3.012	

Figure 9: DataSet



```
+--------------+--------+--------+--------+
| Iteration(t) |   mu   | sigma  |   p    |
+--------------+--------+--------+--------+
|      0       | 1.400 | 0.800 | 0.400 |
|      10      | 2.459 | 0.477 | 1.047 |
|      20      | 2.459 | 0.477 | 1.047 |
|      30      | 2.459 | 0.477 | 1.047 |
|      40      | 2.459 | 0.477 | 1.047 |
|      50      | 2.459 | 0.477 | 1.047 |
|      60      | 2.459 | 0.477 | 1.047 |
|      70      | 2.459 | 0.477 | 1.047 |
|      80      | 2.459 | 0.477 | 1.047 |
|      90      | 2.459 | 0.477 | 1.047 |
|     100      | 2.459 | 0.477 | 1.047 |
+--------------+--------+--------+--------+
```

Figure 10: Estimates of μ, σ, p after convergence

As evident from the table above, the estimates are

$$\begin{aligned}\hat{\mu} &= 2.459\\ \hat{\sigma} &= 0.477\\ \hat{p} &\approx 1\end{aligned}$$

Note that this data turns out to follow a normal distribution (not skewed, $p = 1$). As stated earlier, GSN incorporates the usual normal distribution as a special case.

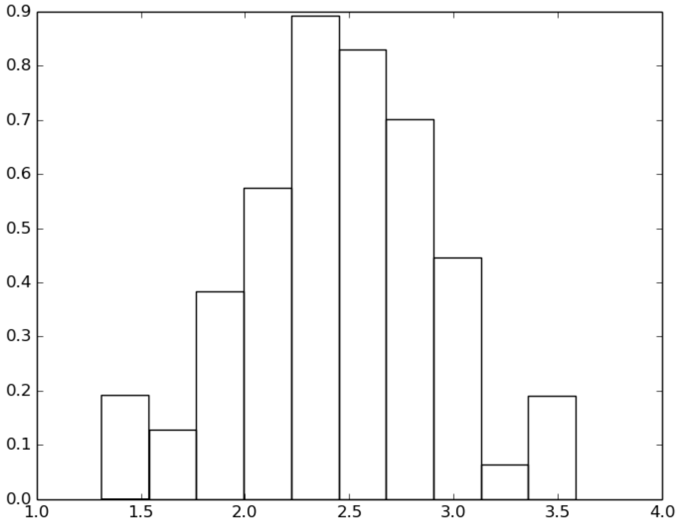


Figure 11: Histogram of data

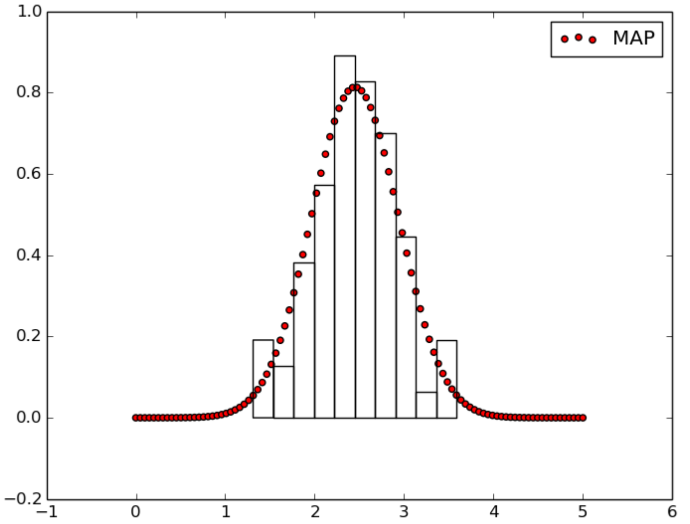


Figure 12: Normed histogram of the data and PDF

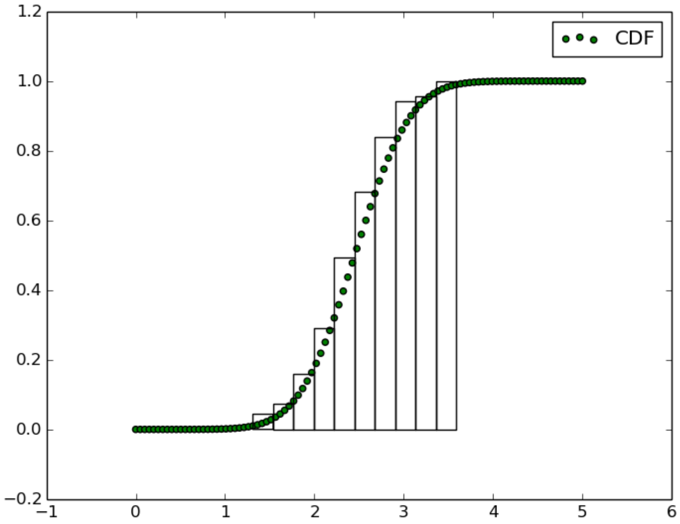


Figure 13: Cumulative histogram of the data and CDF

With these estimates, the Kolmogorov-Smirnov (KS) test statistic and the associated p values are 0.037 and 0.9, respectively.

5 Conclusion

In this project, we developed a Bayesian framework for GSN and estimate its parameters using prior, data and an iterative procedure, since estimate of one parameter depends on rest of the parameters. We found that these estimates are better compared to MLE[1] as well as Azzalini's [2] skew normal distribution in terms of good fit measured by KS test statistic and its associated p value.

6 Future Work

We would like to extend this analysis to a mixture of GSNs, where we expect the data to come from more than one GSN with different weights to each. We can also develop better ways to set the prior parameters.

$$f(x \mid \{\mu_k\},\ \{\sigma_k\},\ \{p_k\},\ \{\pi_k\}) = \sum_{k=1}^{K} \pi_k f(x; \mu_k, \sigma_k, p_k)$$

$$s.t. \qquad \sum_{k=1}^{K} \pi_k = 1$$

We futher plan to extend each of these scenarios to a multivariate setting.