SDS 383C - Statistical Modeling 1: Homework 2

Rahul Nandakumar

Graduate Student, ORIE Program (E-ID: rn9355)

September 30, 2022
Resubmitted on December 12, 2022

1. For $y_1, \ldots y_n \overset{iid}{\sim} Normal(\mu, \sigma^2)$, show that $\mathbf{s} = (\bar{y}, s^2)$ are sufficient. Also, find the observed and the expected information matrices.

Solution: We know that the probability distribution of a Normal random variable is given by,

$$f(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}$$

The likelihood function and the log likelihood function is given as,

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}$$

$$\mathcal{L}(\mu, \sigma^2) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu)^2}\right)$$

$$= -n\log(\sqrt{2\pi}\sigma) + \sum_{i=1}^{n} \frac{-1}{2\sigma^2}(y_i - \mu)^2$$

Differentiating the log likelihood function with respect to μ and σ, we get the following,

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{i=1}^{n} \frac{1}{\sigma^2}(y_i - \mu)$$

$$= \frac{n}{\sigma^2}(\bar{y} - \mu)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \sum_{i=1}^{n} (y_i - \mu)^2 \frac{1}{\sigma^3} - \frac{n}{\sigma}$$

$$= n\left(\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^3} - \frac{1}{\sigma}\right)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \mu^2} = \frac{-n}{\sigma^2}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} (y_i - \mu)^2$$

$$\frac{\partial^2 \mathcal{L}}{\partial \mu \partial \sigma} = \frac{\partial^2 \mathcal{L}}{\partial \sigma \partial \mu} = \frac{-n(\bar{y} - \mu)}{\sigma^2}$$

The observed information matrix is given by

$$J(\theta) = -\begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \mu^2} & \frac{\partial^2 \mathcal{L}}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \mathcal{L}}{\partial \sigma \partial \mu} & \frac{\partial^2 \mathcal{L}}{\partial \sigma^2} \end{bmatrix}$$

$$J(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n(\bar{y}-\mu)}{\sigma^2} \\ \frac{n(\bar{y}-\mu)}{\sigma^2} & \frac{-n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2 \end{bmatrix}$$

The expected information matrix is given by

$$\mathbf{I}(\theta) = \mathbb{E}\left\{ -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta^T} \right\}$$

Now, let us calculate

$$\mathbb{E}\left[\frac{n(\bar{y}-\mu)}{\sigma^2}\right] = \mathbb{E}\left[\frac{n\bar{y}}{\sigma^2}\right] - \mathbb{E}[\frac{n\mu}{\sigma^2}]$$

$$= \mathbb{E}\left[\frac{n\mu}{\sigma^2}\right] - \mathbb{E}\left[\frac{n\mu}{\sigma^2}\right] = 0$$

$$\mathbb{E}[\frac{-n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2] = \mathbb{E}\left[\frac{-n}{2\sigma^4}\right] + \mathbb{E}\left[\frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$= \frac{-n}{2\sigma^4} + 0 = \frac{-n}{2\sigma^4}$$

Therefore, applying the expectation to the matrix $J(\theta)$, we get

$$\mathbf{I}(\theta) = \begin{bmatrix} \mathbb{E}[\frac{n}{\sigma^2}] & \mathbb{E}[\frac{n(\bar{y}-\mu)}{\sigma^2}] \\ \mathbb{E}[\frac{n(\bar{y}-\mu)}{\sigma^2}] & \mathbb{E}[\frac{-n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix};$$

A statistic $s(y_{1:n})$ is said to be sufficient for $y_{1:n}$ iff

$$f(y_{1:n}|\theta) = g(s|\theta)h(y_{1:n}); \text{ By factorization theorem}$$

Considering this approach, for the given Normal(μ, σ^2) distribution, we get

$$f(y_1 \ldots y_n|\mu, \sigma^2) = f(y_1|\mu, \sigma^2) \times \cdots \times f(y_n|\mu, \sigma^2); \text{ Since they are iid}$$

$$= \left(\frac{1}{\sqrt{(2\pi)}\sigma}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i^2+\mu^2-2y_i\mu+\bar{y}^2-\bar{y}^2)}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{\frac{-1}{2\sigma^2}[\sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2]}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{\frac{-n(\bar{y}-\mu)^2}{2\sigma^2}} e^{\frac{-1}{2\sigma^2}[\sum_{i=1}^{n}(y_i-\bar{y})^2]}$$

$$= (2\pi)^{\frac{n}{2}}\sigma^{-n} e^{\frac{-(n-1)s^2}{2\sigma^2}} e^{\frac{-n(\bar{y}-\mu)^2}{2\sigma^2}}$$

Since this final expression is of the form $f(y_{1:n}|s, \theta) = g(s|\theta)h(y_{1:n})$, where,

$$g(s|\theta) = g(s|\mu, \sigma^2) = \sigma^{-n} e^{\frac{-(n-1)s^2}{2\sigma^2}} e^{\frac{-n(\bar{y}-\mu)^2}{2\sigma^2}}$$

$$h(y_{1:n}) = (2\pi)^{\frac{n}{2}}$$

we can conclude that $\mathbf{s} = (\bar{y}, s^2)$ is a sufficient statistic for the given distribution.

2. From AC Davison's Statistical Models, do problem 4.4.6.

Solution: We have the pdf of the $N(\mu, c\mu^2)$ distribution given as,

$$f(y_i|\mu, c\mu^2) = \frac{1}{\sqrt{2\pi c}} e^{\frac{-1}{2c\mu^2}(y_i - \mu)^2}$$

Since it is given that $y_1, \ldots y_n$ are iid, we can write the joint distribution as,

$$
\begin{aligned}
f(y_1, \ldots, y_2|\mu, c\mu^2) &= \left(\frac{1}{\sqrt{2\pi c\mu}}\right)^n e^{\frac{-1}{2c\mu^2} \sum_{i=1}^n [(y_i - \mu)^2]} \\
&= \left(\frac{1}{\sqrt{2\pi c\mu}}\right)^n e^{\frac{-1}{2c\mu^2} \sum_{i=1}^n [y_i^2 + \mu^2 - 2y_i\mu + \bar{y}^2 - \bar{y}^2]} \\
&= \left(\frac{1}{\sqrt{2\pi c\mu}}\right)^n e^{\frac{-1}{2c\mu^2} [\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2]} \\
&= \frac{e^{n(\bar{y}-\mu)^2}}{(\sqrt{2\pi c\mu})^n} e^{\sum_{i=1}^n (y_i - \bar{y})^2}
\end{aligned}
$$

The final result is of the form, $f(s|\theta, y_{1:n}) = g(s|\theta)h(y_{1:n})$. For a regular $N(\mu, \sigma^2)$ distribution,

$$
\begin{aligned}
f(y_1, \ldots, y_2|\mu, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi \sigma}}\right)^n e^{\frac{-1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2} \\
&= \frac{e^{n(\bar{y}-\mu)^2}}{(\sqrt{2\pi \sigma})^n e^{2\sigma^2}} e^{\sum_{i=1}^n (y_i - \bar{y})^2}
\end{aligned}
$$

From both these results, we can conclude that the minimal sufficient statistic is the same for both these distributions. To find the maximum likelihood function of μ, we consider the log likelihood function.

$$
\begin{aligned}
\mathcal{L}(\mu) &= \sum_{i=1}^n \log\left(\frac{1}{\sqrt{2\pi c\mu}} e^{\frac{-1}{2c\mu^2}(y_i - \mu)^2}\right) \\
&= -n\log(\sqrt{2\pi c}\mu) - \frac{1}{2c\mu^2} \sum_{i=1}^n (y_i - \mu)^2 \\
\frac{d\mathcal{L}(\mu)}{d\mu} &= \frac{-n}{\mu} - \frac{1}{2c}\left[\frac{\mu^2 \sum_{i=1}^n -2(y_i - \mu) - \sum_{i=1}^n (y_i - \mu)^2 2\mu}{\mu^4}\right] = 0 \\
&= \frac{-1}{\mu} - \frac{1}{2c}\frac{\mu^2(-2n\bar{y} + 2n\mu) - 2n\sum_{i=1}^n (y_i - \bar{y})^2 + (\bar{y} - \mu)^2}{\mu^4} = 0 \\
&= c\mu^2 + 3\bar{y}\mu - \frac{1}{n}\sum_{i=1}^n y_i^2 = 0
\end{aligned}
$$

We solve this using the quadratic formula to get $\hat{\mu}_{MLE}$

$$\hat{\mu}_{MLE} = \frac{-3\bar{y} \pm \sqrt{9\bar{y}^2 + \frac{4c}{n}\sum_{i=1}^n y_i^2}}{2c}$$

3

Now, let us find the minimal sufficient statistic for our model. We start with the minimal sufficient statistic for the known model $\text{Normal}(\mu, \sigma^2)$. From AC Davison, we are given the statement that if the likelihood of two sets of data y and z are same upto a constant, then $\frac{L(\theta;y)}{L(\theta;z)}$ doesn't depend on θ. Further, the partition that this equivalence relationship generates is minimally sufficient.

Let $Z_1, \ldots, Z_n \sim \text{Normal}(\mu, \sigma^2)$, then we derive the minimal sufficient statistic of $Y_1, \ldots Y_N \sim \text{Normal}(\mu, \sigma^2)$ as,

$$\frac{L(\mu, \sigma^2 | Z_{1:n})}{L(\mu, \sigma^2 | Y_{1:n})} = \frac{f(Z_1, \ldots, Z_n | \mu, \sigma^2)}{f(Y_1, \ldots, Y_n | \mu, \sigma^2)}$$

$$= \exp\left\{ \frac{\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 - \sum_{i=1}^n z_i^2 + 2\mu \sum_{i=1}^n z_i - n\mu^2}{2\sigma^2} \right\}$$

This expression is constant w.r.t μ iff $\sum_{i=1}^n z_i = \sum_{i=1}^n y_i$. Furthermore, the expression is constant w.r.t σ^2 iff $\sum_{i=1}^n z_i = \sum_{i=1}^n y_i$ and $\sum_{i=1}^n z_i^2 = \sum_{i=1}^n y_i^2$. Therefore, we can conclude that the minimal sufficient statistic for $Y_1, \ldots Y_N \sim \text{Normal}(\mu, \sigma^2)$ is $s = (\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2)$, with $\sum_{i=1}^n y_i$ being the sufficient statistic for μ.

Now, let us consider the case of finding the minimal sufficient statistic for $Y_1, \ldots Y_N \sim \text{Normal}(\mu, c\mu^2)$. Let $Z_1, \ldots Z_N \sim \text{Normal}(\mu, c\mu^2)$. Then,

$$\frac{L(\mu, c\mu^2 | Z_{1:n})}{L(\mu, c\mu^2 | Y_{1:n})} = \frac{f(Z_1, \ldots, Z_n | \mu, c\mu^2)}{f(Y_1, \ldots, Y_n | \mu, c\mu^2)}$$

$$= \exp\left\{ \frac{1}{2c\mu^2} \left(\sum_{i=1}^n y_i^2 - z_i^2 \right) + \frac{1}{c\mu} \left(\sum_{i=1}^n z_i - \sum_{i=1}^n y_i \right) \right\}$$

Again, this expression is constant w.r.t μ iff $\sum_{i=1}^n z_i = \sum_{i=1}^n y_i$. and $\sum_{i=1}^n z_i^2 = \sum_{i=1}^n y_i^2$. Therefore, we can conclude that the minimal sufficient statistic for $(\mu, c\mu^2)$ is $s = (\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2)$. Hence we can conclude that minimal sufficient statistic for $\text{Normal}(\mu, c\mu^2)$ is the same as for $\text{Normal}(\mu, \sigma^2)$.

Now let us work on showing the large sample standard error of the calculated $\hat{\mu}_{MLE}$. We know that, for large n,

$$(\hat{\theta}_{MLE} - \theta) \approx \text{MVN}(0, V)$$

where, $V = \left(\mathbb{E}\left(\frac{-\partial^2 \mathcal{L}(\hat{\theta})}{\partial \hat{\theta} \partial \hat{\theta}^T} \right) \right)$ Taking the second derivative of the log-likelihood, we get

$$\frac{d\mathcal{L}(\mu)}{d\mu} = \frac{-n}{\mu} - \frac{1}{2c}\left[\frac{\mu^2 \sum_{i=1}^n -2(y_i - \mu) - \sum_{i=1}^n (y_i - \mu)^2 2\mu}{\mu^4} \right]$$

$$\frac{d^2\mathcal{L}(\mu)}{d\mu^2} = -n\frac{d}{d\mu}\left(\frac{1}{\mu} \right) + \frac{1}{c}\sum_{i=1}^n \frac{d}{d\mu}\left(\frac{(y_i - \mu^2)}{\mu^3} \right) - \frac{1}{c}\sum_{i=1}^n y_i \frac{d}{d\mu}\left(\frac{1}{\mu^2} \right) + \frac{n}{c}\frac{d}{d\mu}\left(\frac{1}{\mu} \right)$$

$$= \frac{n}{\mu^2} + \frac{1}{c}\sum_{i=1}^n \left(\frac{-2y_i + 2\mu}{\mu^3} - \frac{3(y_i - \mu)^2}{\mu^4} \right) + \frac{1}{c}\sum_{i=1}^n y_i \frac{2}{\mu^3} - \frac{n}{c\mu^2}$$

$$= \frac{n}{\mu^2} - \frac{2}{c\mu^3}\sum_{i=1}^n y_i + \frac{2n}{c\mu^2} - \frac{3}{c\mu^4}\sum_{i=1}^n y_i^2 + \frac{6}{c\mu^3}\sum_{i=1}^n y_i - \frac{3n}{c\mu^2} + \frac{2}{c\mu^3}\sum_{i=1}^n y_i - \frac{n}{c\mu^2}$$

$$= \frac{-3}{c\mu^4}\sum_{i=1}^n y_i^2 + \frac{6}{c\mu^3}y_i - \frac{2n}{c\mu^2} + \frac{n}{\mu^2}$$

Further, we calculate the expected value of the negative of this double derivate as,

$$\mathbb{E}\left(-\frac{d^2\mathbb{L}(\mu)}{d\mu^2} \right) = \mathbb{E}\left(\frac{3}{c\mu^4}\sum_{i=1}^n y_i^2 - \frac{6}{c\mu^3}y_i + \frac{2n}{c\mu^2} - \frac{n}{\mu^2} \right)$$

$$= \frac{3}{6\mu^4}(c\mu^2 + \mu^2) - \frac{6}{c\mu^3}n\mu + \frac{2n}{c\mu^2} - \frac{n}{\mu^2}$$

$$= \frac{c + 1 - 2n}{2\mu^2} - \frac{4n}{c\mu^2}$$

$$\mathbb{E}\left(-\frac{d^2\mathbb{L}(\mu)}{d\mu^2} \right)\Big|_{\mu=\hat{\mu}_{MLE}} = \frac{c + 1 - 2n}{2\hat{\mu}_{MLE}^2} - \frac{4n}{c\hat{\mu}_{MLE}^2}$$

Thus,

$$SE(\hat{\mu}_{MLE}) = \sqrt{\frac{c + 1 - 2n}{2\hat{\mu}_{MLE}^2} - \frac{4n}{c\hat{\mu}_{MLE}^2}}$$

3.

Solution(a): We draw a random sample of size = 20 from a Cauchy(0,1) distribution. This is done in Python using the *scipy.stats* package where we can call the *cauchy.pdf()* function with *loc = 0, scale = 1*. The plot of the Cauchy(0,1) distribution is given below.
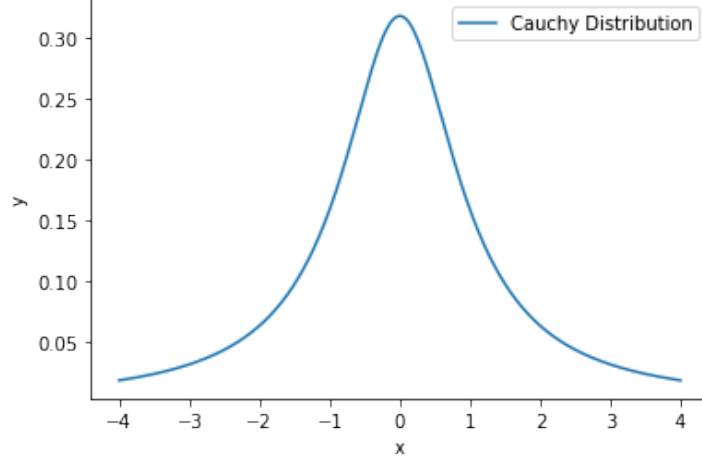
Figure 1: Cauchy(0,1) distribution

From here, we repeat this process for $B = 1000$ times and create a list containing all the sample means. Then, we plot a histogram of all the sample means, and superimpose a Standard Cauchy distribution over it. The histogram of means also follows a Cauchy(0,1) distribution. This is illustrated in the plot below. The code for generating this plot is also attached with the assignment.
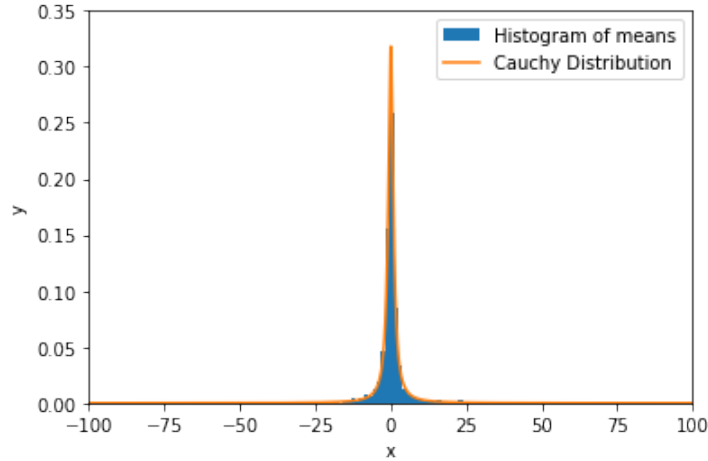

Figure 2: Cauchy(0,1) distribution superimposed over Histogram of means

Now, for the second part of the problem, we can use the Likelihood function to estimate θ. Since there is no closed form solution, we turn to numerical approximation methods implemented in Python. These methods are iterative, and convergence is acheived after set number of iterations. The methods are as follows

a. Step-wise gradient ascent: Here, we have γ^m as the learning rate.

$$\theta^{m+1} = \theta^m + \gamma^m \frac{\partial \mathcal{L}(\theta^m)}{\partial \theta}$$

b. Newton-Raphson: Similar to gradient ascent, except our learning rate is computed as a function of the inverse of second gradient of the log likelihood function.

$$\theta^{m+1} = \theta^m + J(\theta^m)^{-1} \frac{\partial \mathcal{L}(\theta^m)}{\partial \theta}$$

c. stochastic gradient ascent. Here, we take a batch of values from our sample, and compute the gradients accordingly. Taking random batches from sample for each iteration helps us avoid saddle points, and helps us achieve convergence quickly.

$$\theta^{m+1} = \theta^m + \gamma^m \frac{\partial \mathcal{L}_{i^m}(\theta^m)}{\partial \theta}$$

To estimate the first and the second gradient required in the implementations of these methods, we arrive at the following expression for the Log Likelihood function. We know that,

$$f(y_i|\theta, 1) = \frac{1}{\pi} \frac{1}{1 + (y_i - \theta)^2}$$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\pi} \frac{1}{1 + (y_i - \theta)^2}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\pi} \frac{1}{1 + (y_i - \theta)^2}\right)$$

$$= -\left[n\log(\pi) + \sum_{i=1}^{n} \log(1 + (y_i - \theta)^2)\right]$$

Now, the first and second derivative can be calculated as,

$$\frac{\partial \mathcal{L}}{\partial \theta} = -2 \sum_{i=1}^{n} \frac{(y_i - \theta)}{1 + (y_i - \theta)^2}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta^2} = -2 \sum_{i=1}^{n} \frac{[1 + (y_i - \theta)^2](-1) - (y_i - \theta)[2(y_i - \theta)]}{(1 + (y_i - \theta)^2)^2}$$

$$= 2 \sum_{i=1}^{n} \frac{1 + 3(y_i - \theta)^2}{(1 + (y_i - \theta)^2)^2}$$

Using these derivate formulas, we can implement the methods given above. Our convergence criteria is met, when the total number of iterations reach 100000. For the stochastic gradient ascent, we consider a batch size of 12. We also consider a learning rate of 0.5 for the stochastic and stepwise gradient ascent, and we consider an initial guess value for θ as 5 in all cases from where we iteratively try to find actual θ. We arrive at the following results for the estimated values of θ. We can verify the integrity of these results by fitting a cauchy distribution using the *cauchy.fit()* method in *scipy.stats*. Using this, we get a value for θ as 9.12286148071289. Therefore, we can also calculate the rms error for each method as follows. The plots have also been attached.

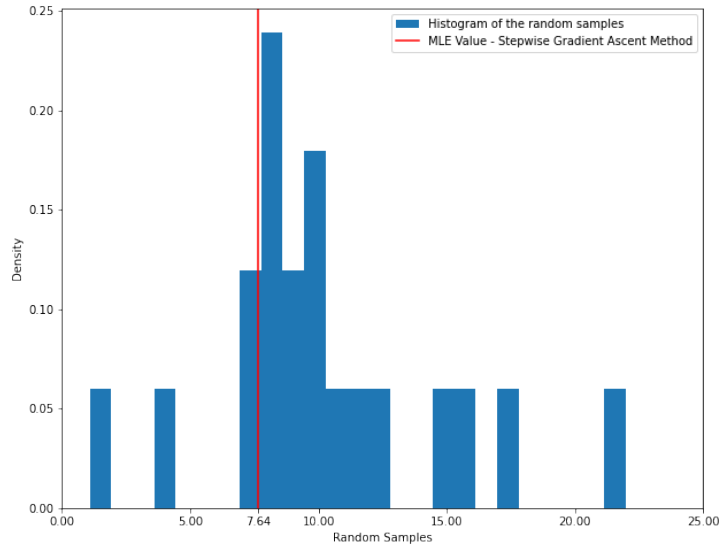a. step-wise gradient ascent: $\theta = 7.637799268662628$, rms error: 1.485062212050262

Figure 3: Stepwise Gradient Ascent method calculated theta

b. Newton-Raphson method: $\theta = 9.122853472425337$, rms error: $8.008287553096238 \times 10^{-6}$
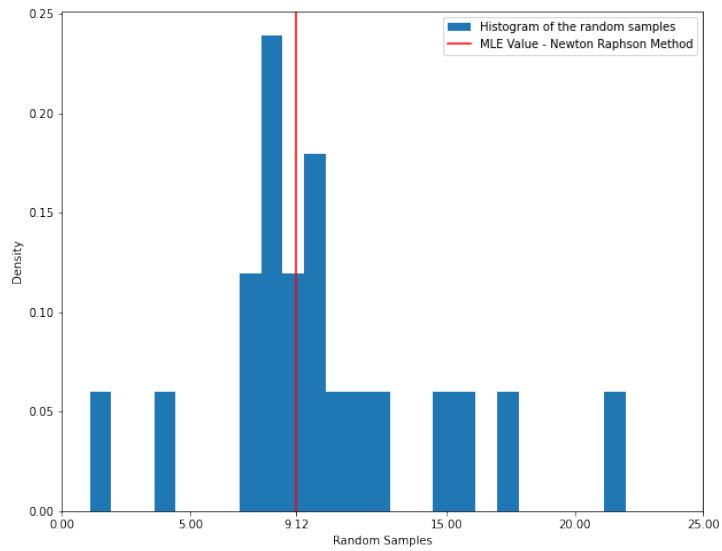

Figure 4: Newton-Raphson method calculated theta

c. Stochastic gradient ascent: $\theta = 9.472881177927484$, rms error: 0.35001969721459325
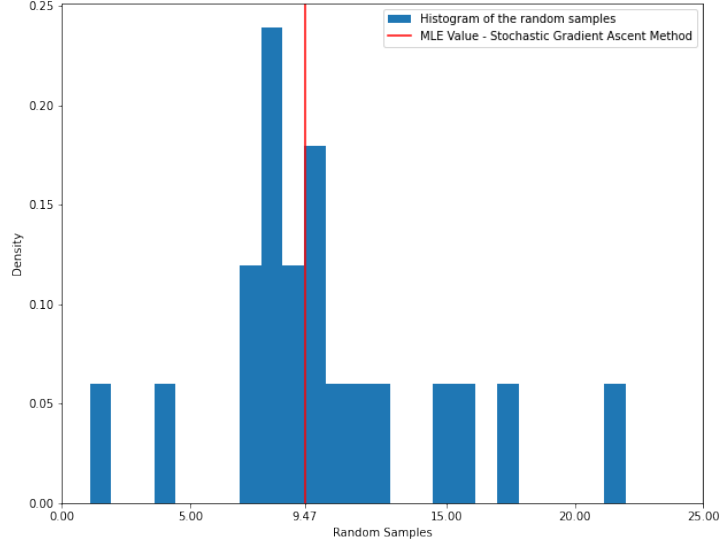
Figure 5: Stochastic gradient ascent method calculated theta

The code accomanying these implementations is attached along with this assignment.

4. Let $y_{ij} \overset{ind}{\sim} Normal(\mu_i, \sigma^2)$, $i = 1 \ldots n$ and $j = 1 \ldots m$ with $n \to \infty$ but m fixed. Find out MLEs of μ_i and σ^2. Show that $\hat{\sigma}^2_{MLE}$ is inconsistent for σ^2. Adjust this estimator to come up with a consistent estimator of σ^2.

Solution: It is given that

$$f(y_{ij}|\mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu_i)^2}; \forall j = 1, \ldots, m$$

Now, the likelihood and the log likelihood function can be calculated as,

$$L(y_{ij}|\mu_i, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu_i)^2}; \forall j = 1 \ldots m$$

$$\mathcal{L}(y_{ij}|\mu_i, \sigma^2) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y_i - \mu_i)^2}\right); \forall j = 1 \ldots m$$

$$= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \sum_{i=1}^{n} \log e^{\frac{-1}{2\sigma^2}(y_i - \mu_i)^2}$$

$$= -n\log(\sqrt{2\pi}\sigma) + \sum_{i=1}^{n} \frac{-1}{2\sigma^2}(y_i - \mu_i)^2$$

Now, taking the derivative of the log likelihood function with respect to μ_i and σ, and

9

solving them, we get the following results.

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - \mu_i)^2 = 0$$

$$= -n\sigma^2 + \sum_{i=1}^{n} (y_i - \mu_i)^2 = 0$$

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu_i)^2$$

$$\frac{\partial \mathcal{L}}{\partial \mu_i} = 0 + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (2)(y_i - \mu_i) = 0$$

$$\hat{\mu}_{i,MLE} = y_i$$

Therefore, we have found the MLEs for μ_i and σ^2

5. For $y_1, \ldots, y_n \overset{iid}{\sim} Uniform(0, \theta)$, show that $z_n = \frac{n(\theta - \hat{\theta}_{MLE})}{\theta} \overset{D}{\to} E$, where E is exponential. Design a simulation study illustrating this nice result.

Solution: The given distribution is Uniform(0, θ), and the pdf is given by

$$f(y_i|\theta) = \frac{1}{\theta}$$

Let us construct the likelihood and log likelihood function for this distribution.

$$L(\theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta}\right)$$

$$= \left(\frac{1}{\theta}\right)^n$$

$$\mathcal{L}(\theta) = -n\log(\theta)$$

Therefore, since the function is monotonous, the MLE for θ is given by $y_{(n)}$, which is the nth order statistic. In simple words, $\hat{\theta}_{MLE} = \max\{y_1, \ldots, y_n\}$. Now, we consider the given problem, $z_n = \frac{n(\theta - \hat{\theta}_{MLE})}{\theta}$. The convergence in distribution is defined as when

$$\lim_{n \to \infty} \Pr(x_n \le a) = \Pr(x \le a)$$

In our problem, as $n \to \infty$,

$$Pr(z_n \leq t) = Pr\left(\frac{n(\theta - \hat{\theta}_{MLE})}{\theta} \leq t\right)$$

$$= Pr\left(n(\theta - \hat{\theta}_{MLE}) \leq \theta t\right)$$

$$= Pr\left(\hat{\theta}_{MLE} \geq \theta - \frac{\theta t}{n}\right)$$

$$= 1 - Pr\left(\hat{\theta}_{MLE} \leq \theta - \frac{\theta t}{n}\right)$$

$$= 1 - \left(\frac{\theta - \frac{\theta t}{n}}{\theta}\right)^n$$

$$= 1 - e^{-t}; \text{ Since } \lim_{n \to \infty}\left(1 + \frac{x}{n}\right)^n = e^x$$

This is the pdf of a exponential distribution with parameter $\lambda = 1$. Therefore we can conclude that for large n, z_n converges in distribution to E. This is also illustrated in the graph below. The procudure to obtain the graph is also illustrated. We create a sorted array of a randomly generated uniform distribution with $\theta = 10$. For $B = 100000$ times, we generate random samples of size $= 1000$ from this population. The MLE of this sample is the maximum value of the samples, i.e., $\hat{\theta}_{MLE} = max\{x_1, x_2 \ldots, x_n\} = x_{(n)}$, as established above. By calculating the value of z_n as asked in the question, and by taking the plot of z_n overlaid with an exponential, we can see that the distribution converges. The code for generating this plot is also attached along with this assignment.
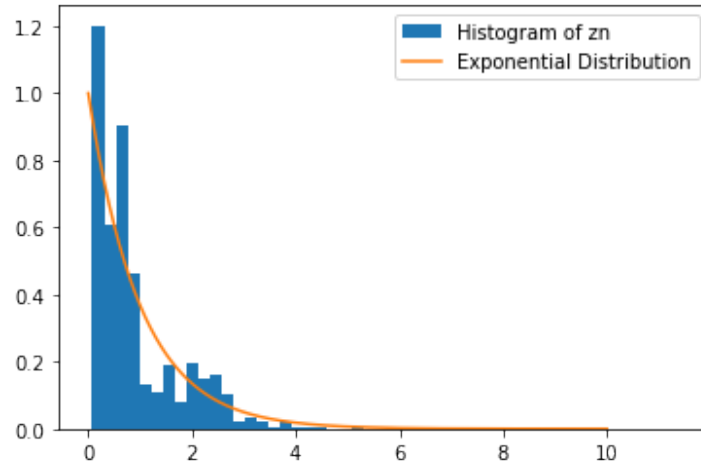

Figure 6: Showing convergence in distribution

6. Consider the spring failure data from Davison, providing the failure times at stress 950 N/mm^2 in 10^3 cycles:

$$225, 171, 198, 189, 189, 135, 162, 135, 117, 162$$

Using the method of maximum likelihood, fit a Weibull distribution to these data [use the parametrization considered in class]. Also find an approximate 90% joint confidence region for the parameters (k, λ). Supplement your analysis with appropriate plots.

Solution: The pdf of a Weibull distribution is given by

$$f(y_i|k, \lambda) = \frac{k}{\lambda}\left(\frac{y_i}{\lambda}\right)^{k-1} e^{-\left(\frac{y_i}{\lambda}\right)^k}$$

We can find the likelihood and log likelihood function as follows.

$$L(\theta) = \prod_{i=1}^{n} \frac{k}{\lambda}\left(\frac{y_i}{\lambda}\right)^{k-1} e^{-\left(\frac{y_i}{\lambda}\right)^k}$$

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log\left(\frac{k}{\lambda}\left(\frac{y_i}{\lambda}\right)^{k-1} e^{-\left(\frac{y_i}{\lambda}\right)^k}\right)$$

$$= n\log(k) - nk\log(\lambda) - \sum_{i=1}^{n}\left(\frac{y_i}{\lambda}\right)^k + (k-1)\sum_{i=1}^{n}\log(y_i)$$

Taking the partial derivate wrt. k, λ, and equating to zero, we get

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -\frac{nk}{\lambda} + \frac{k}{\lambda^{k+1}} + \frac{k}{\lambda^{k+1}}\sum_{i=1}^{n}(y_i)^k \qquad = 0$$

$$\frac{\partial \mathcal{L}}{\partial k} = \frac{n}{k} - n\log(\lambda) - \frac{\sum_{i=1}^{n} y_i^k - \log(\lambda)\sum_{i=1}^{n} y_i^k}{\lambda^k} \qquad = 0$$

From the first equation, we can arrive at a value for $\hat{\lambda}$ as

$$\hat{\lambda} = \left(\sum_{i=1}^{n} y_i^k\right)^{\frac{1}{k}}$$

We cannot solve the second equation directly to find $\hat{k}$, therefore we use the Newton Raphson method to find an iterative solution for $\hat{k}$ in the second equation. We update $\hat{k}_{new}$ as

$$\hat{k}_{new} = \hat{k}_{old} - \frac{f(\hat{k}_{old})}{f'(\hat{k}_{old})}$$

The derivative can be calculated as

$$\frac{\partial^2 \mathcal{L}}{\partial k^2} = \left(\frac{\sum_{i=1}^{n} y_i^k \log y_i}{\sum_{i=1}^{n} y_i^k}\right)^2 - 2\left(\frac{\sum_{i=1}^{n} y_i^k \log y_i}{\sum_{i=1}^{n} y_i^k}\right) - \frac{1}{k^2}$$

Our stopping criterion is reached, until the first derivative value hits 0.0001. This is also the point where the increment in $\hat{k}_{new}$ is negligible compared to $\hat{k}_{old}$. We use the value of k_{new} in our original expression to estimate $\hat{\lambda}$. Using this procudure, for the given data, we get the following parameter values

$$(\hat{k}, \hat{\lambda}) = (5.974797076433545, 181.40553454869354)$$

We can verify the integrity of these results by using the *weibull_min.fit()* function in the *scipy.stats* package and estimating the parameters. The RMS value of the errors are

$$e_k = (\hat{k} - k)^2 = 0.00000447717$$

$$e_\lambda = (\hat{\lambda} - \lambda)^2 = 1.15940309 \times 10^{-9}$$

Therefore, we can conclude that our estimates are pretty accurate and our method works. Now, using this estimated values of $(\hat{k}, \hat{\lambda})$, we can calculate the joint confidence region as follows.

7. For a Weibull(k, λ) distribution, design a simulation study to show that $(\hat{\theta}-\theta)^T(\mathbf{J}(\hat{\theta}))(\hat{\theta}-\theta) \xrightarrow{D} \chi_2^2$ and $\mathbf{W}(\theta) \xrightarrow{D} \chi_2^2$

Solution: The pdf of a Weibull distribution is given by

$$f(x_i|k, \lambda) = \frac{k}{\lambda}\left(\frac{x_i}{\lambda}\right)^{k-1} e^{-\left(\frac{x_i}{\lambda}\right)^k}$$

For this question, let us consider the following.

$$\hat{\theta} = [\hat{k}, \hat{\lambda}]$$
$$\theta = [k, \lambda]$$

The likelihood function is given by,

$$\mathcal{L}(\theta) = -kn\log(\lambda) + n\log(k) + (k-1)\sum_{i=1}^{n}\log(x_i) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k$$

We find the first and second derivatives of this log-likelihood function as follows.

$$\frac{\partial \mathcal{L}(\theta)}{\partial k} = -n\log(\lambda) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k\log\left(\frac{x_i}{\lambda}\right) + \sum_{i=1}^{n}\log(x_i) + \frac{n}{k}$$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \lambda} = \frac{-kn}{\lambda} + \sum_{i=1}^{n}\frac{k\left(\frac{x_i}{\lambda}\right)^k}{\lambda}$$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial k^2} = \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k\log\left(\frac{x_i}{\lambda}\right)^2 - \frac{n}{k^2}$$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \lambda^2} = \frac{kn}{\lambda^2} - \sum_{i=1}^{n}\left(\frac{k^2\left(\frac{x_i}{\lambda}\right)^k}{\lambda^2} + \frac{k\left(\frac{x_i}{\lambda}\right)^k}{\lambda^2}\right)$$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \lambda \partial k} = -\sum_{i=1}^{n}\left(-\frac{k\left(\frac{x_i}{\lambda}\right)^k\log\left(\frac{x_i}{\lambda}\right)}{\lambda} - \frac{\left(\frac{x_i}{\lambda}\right)^k}{\lambda}\right) - \frac{n}{\lambda} = \frac{\partial^2 \mathcal{L}(\theta)}{\partial k \partial \lambda}$$

Using the second derivate values, we can calculate $\mathbf{J}(\theta)$ as,

$$\mathbf{J}(\theta) = \begin{bmatrix} -\frac{\partial^2 \mathcal{L}(\theta)}{\partial k^2} & -\frac{\partial \mathcal{L}(\theta)}{\partial k \partial \lambda} \\ -\frac{\partial \mathcal{L}(\theta)}{\partial \lambda \partial k} & -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \lambda^2} \end{bmatrix}$$

$$\mathbf{J}(\hat{\theta}) = \begin{bmatrix} -\frac{\partial^2 \mathcal{L}(\theta)}{\partial k^2}\big|_{\hat{k},\hat{\lambda}} & -\frac{\partial \mathcal{L}(\theta)}{\partial k \partial \lambda}\big|_{\hat{k},\hat{\lambda}} \\ -\frac{\partial \mathcal{L}(\theta)}{\partial \lambda \partial k}\big|_{\hat{k},\hat{\lambda}} & -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \lambda^2}\big|_{\hat{k},\hat{\lambda}} \end{bmatrix}$$

Also, we have the following vectors,

$$(\hat{\theta} - \theta)^T = \begin{bmatrix} \hat{k} - k & \hat{\lambda} - \lambda \end{bmatrix}$$

13

$$\hat{\theta} - \theta = \begin{bmatrix} \hat{k} - k \\ \hat{\lambda} - \lambda \end{bmatrix}$$

By combining the expressions, we arrive at the following result.

$$(\hat{\theta} - \theta)^T \mathbf{J}(\hat{\theta})(\hat{\theta} - \theta) = \begin{bmatrix} \hat{k} - k & \hat{\lambda} - \lambda \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 \mathcal{L}(\theta)}{\partial k^2}|_{\hat{k},\hat{\lambda}} & -\frac{\partial \mathcal{L}(\theta)}{\partial k \partial \lambda}|_{\hat{k},\hat{\lambda}} \\ -\frac{\partial \mathcal{L}(\theta)}{\partial \lambda \partial k}|_{\hat{k},\hat{\lambda}} & -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \lambda^2}|_{\hat{k},\hat{\lambda}} \end{bmatrix} \begin{bmatrix} \hat{k} - k \\ \hat{\lambda} - \lambda \end{bmatrix}$$

$$= \Bigg[(-k + \hat{k}) \Bigg((-k + \hat{k}) \Bigg(\sum_{i=1}^{n} \Big(\frac{x_i}{\hat{\lambda}} \Big)^{\hat{k}} \log \Big(\frac{x_i}{\hat{\lambda}} \Big)^2 + \frac{n}{\hat{k}^2} \Bigg) +$$

$$(\hat{\lambda} - \lambda) \Bigg(\sum_{i=1}^{n} \Big(-\frac{\hat{k}(\frac{x_i}{\hat{\lambda}})^{\hat{k}} \log(\frac{x_i}{\hat{\lambda}})}{\hat{\lambda}} - \frac{(\frac{x_i}{\hat{\lambda}})^{\hat{k}}}{\hat{\lambda}} \Big) + \frac{n}{\hat{\lambda}} \Bigg) \Bigg)$$

$$+ (\hat{\lambda} - \lambda) \Bigg((-k + \hat{k}) \Bigg(\sum_{i=1}^{n} \Big(-\frac{\hat{k}(\frac{x_i}{\hat{\lambda}})^{\hat{k}} \log(\frac{x_i}{\hat{\lambda}})}{\hat{\lambda}} - \frac{(\frac{x_i}{\hat{\lambda}})^{\hat{k}}}{\hat{\lambda}} \Big) + \frac{n}{\hat{\lambda}} \Bigg) +$$

$$(\hat{\lambda} - \lambda) \Bigg(-\frac{\hat{k}n}{\hat{\lambda}^2} + \sum_{i=1}^{n} \Big(\frac{\hat{k}^2(\frac{x_i}{\hat{\lambda}})^{\hat{k}}}{\hat{\lambda}^2} + \frac{\hat{k}(\frac{x_i}{\hat{\lambda}})^{\hat{k}}}{\hat{\lambda}^2} \Big) \Bigg) \Bigg) \Bigg]$$

We have to show that this expression $\xrightarrow{D} \chi_2^2$. Similarly, we can get the expression for $\mathbf{W}(\theta)$. It is given by,

$$\mathbf{W}(\theta) = 2\log \Bigg\{ \frac{L(\hat{\theta}_{MLE})}{L(\theta)} \Bigg\} = 2 \Bigg\{ \mathcal{L}(\hat{\theta}_{MLE}) - \mathcal{L}(\theta) \Bigg\}$$

We also get an expression for this $\mathbf{W}(\theta)$ and show that $\mathbf{W}(\theta) \xrightarrow{D} \chi_2^2$