ASSIGNMENT: **Question 1**

Normally you can cite as [H&M 2000, §1.1]

The number of meteorites falling on an ocean in a given year can be modeled by a geometric distribution counting the number of Bernoulli trials with p=0.44 until it succeeds. Give a graphic showing the probability of one, two, three … meteorites falling (until the probability remains provably less than 0.5% for any bigger number of meteorites). Calculate the expectation and median and show them graphically on this graphic.

Task 1: Basic Probabilities and Visualizations

In geometric distribution, we will consider performing independent Bernoulli trials where the probability of success, p, is fixed from one trial to the next, and the trials are performed until the first success is observed. In this case, the probability of success is 0.44 is the probability that a meteorite will land on an ocean in a given year. Let X count the failures until the first success occurs. The possible values of X are $0,1,2,3$ .... We may observe a success on the first trial, in which case $X = 0$ (no failures). We may get quite unlucky as there is no upper bound on the number of trials! However, we stop when the probability is provably less than 0.5% for any larger number of meteorites. In this setup, X is said to have a geometric distribution, with the probability of success p written as $\mathrm{X} \sim \mathrm{Geometric(p)}$. Remember that we can only observe outcomes with one success and this success occurs on the last trial. Therefore, possible outcomes look like *“S”*, *“FS”*, *“FFS”*,*“FFFS”*, *“FFFFS”*, etc. Using independence, values of the PMF, f, for X are:

- $f(0) = p(S) = p$
- $f(1) = p(FS) = (1-p)p$
- $f(2) = p(FFS) = (1-p)^2p$
- $f(3) = p(FFFS) = (1-p)^3p$

Continuing in this way, we see that the PMF of X is given by:

$$f(x) = P(X = x) = \begin{cases} (1-p)^x p & if\ x = 0,1,2,3,\dots, \\ 0 & Otherwise \end{cases} \qquad \text{Eq 1}$$

Similarly, If the probability of success (meteorite falling on the ocean) on each trial is *p*, the probability that the *k*th meteorite is the first success is given by Probability Mass Function (PMF) f:

$$f(x) = P(X = x) = \begin{cases} (1-p)^{x-1} p & if\ x = 1,2,3,\dots, \\ 0 & Otherwise \end{cases} \qquad \text{Eq 2}$$

The probability of that first meteorite falling on an ocean:

- $P(X = 1) = (1-0.44)^{1-1} * \ 0.44 = 0.44$
- $P(X = 2) = (1-0.44)^{2-1} * \ 0.44 = 0.2464$
- $P(X = 3) = (1-0.44)^{3-1} * \ 0.44 = \ 0.137984$

The calculation is performed iteratively until the cumulative probability exceeds 0.995 (1 - 0.005), ensuring that the probability of observing a larger number of meteorites is less than 0.5%.

The formula for the geometric distribution CDF is given as follows:

$$F(x) = P(X \leq x) = \begin{cases} 1 - (1-p)^x & if\ x > 0\,, \\ 0 & Otherwise \end{cases}$$

Using the source code, it is has been proved that beyond 10 the probability is less than 0.005

Expectation:

Let X Geometric p for 0 < p < 1, then its PMF is given by equation 2, The expected value is

$E[X] = \frac{1}{p} = \frac{1}{0.44} = 2.27$

The expectation represents the average number of trials (meteorites falling) until the first success.

The median is the middle value in a distribution, below which 50% of the observations fall. For the geometric distribution, the median can be approximated using the formula:

$Median[X] = \left\lceil \frac{-1}{\log_2(1-p)} \right\rceil = \left\lceil \frac{-1}{\log_2(1-0.44)} \right\rceil = 2$

Note that the ⌈ ⌉ notation represents ceiling function.

Question 2:

Let Y be the random variable with the time to hear an owl from your room's open window (in hours). Assume that the probability that you still need to wait to hear the owl after y hours is given by $0.53e^{-8y^2} + 0.46e^{-3y^8}$. Find the probability that you need to wait between 2 and 4 hours to hear the owl, compute and display the probability density function graph as well as a histogram by the minute. Compute and display in the graphics the mean, variance, and quartiles of the waiting times.

Solution:

The probability that you still need to wait to hear the owl after y hours is given by:

$$f(y) = \frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8}$$

The survival function of a random variable is defined as the probability that the random variable is greater than a certain value.

The CDF of a random variable Y is the probability that Y is less than or equal to a certain value y. In this case, the CDF of Y is given by:

$$F(y) = 1 - \left(\frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8}\right)$$

To find the probability that you need to wait between 2 and 4 hours to hear the owl, we use:

$$F(4) - F(2) = \left(1 - \left(\frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8}\right)\right) - \left(1 - \left(\frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8}\right)\right)$$

$$= \left(1 - \left(\frac{53}{99}e^{-128} + \frac{46}{99}e^{-196608}\right)\right) - \left(1 - \left(\frac{53}{99}e^{-32} + \frac{46}{99}e^{-768}\right)\right)$$

The **probability density function (PDF)** of Y is given as:

$$f(y) = \frac{dF(y)}{d(y)} = \frac{d(1 - (\frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8})}{d(y)} = 0 - \frac{d\left(\frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8}\right)}{d(y)}$$

Simplifying each term, we get:

$$f(y) = \frac{848y}{99}e^{-8y^2} + \frac{1104}{99}e^{-3y^8}$$

The mean is given by:

$E(Y) = \int y\, f(y)\, dy$

The variance is given by:

Var(Y) = $\int (y - E(Y))$^2 $f(y)\, dy$

The quartiles can be calculated by finding the values of y for which the cumulative distribution function (CDF) reaches 0.25, 0.5

$$1 - \int_y^{\infty} \left(\frac{53}{99}e^{-8y^2} + \frac{46}{99}e^{-3y^8}\right) dy$$

Question 3:`

A type of network router has a bandwidth total to first hardware failure called S expressed in terabytes. The random variable S is modelled by a distribution whose density is given by the function:

$$F_s(s) = \frac{1}{\theta}\ for\ S\ \in [0, \theta]$$

with a single parameter θ. Consider the bandwidth total to failure T of the sequence of the two routers of the same type (one being brought up automatically when the first is broken). Express T in terms of the bandwidth total to failure of single routers $S1$ and $S2$. Formulate realistic assumptions about these random variables. Calculate the density function of the variable T.

Given an experiment with the dual-router-system yielding a sample $T1, T2, \ldots, Tn$, calculate the likelihood function for θ. Propose a transformation of this likelihood function whose maximum is the same and can be computed easily.

An actual experiment is performed, the infrastructure team has obtained the bandwidth totals to failure given by the sequence 77, 2, 20, 32, 14 of numbers. Estimate the model-parameter with the maximum likelihood and compute the expectation of the bandwidth total to failure of the dual-router-system.

Solution:

We assume that:

1. S_1 and S_2 are independent random variables: We assume that the failures of the two routers are independent events. The failure of one router does not affect the failure probability of the other.
2. S_1 and S_2 follow the same distribution with parameter θ. This assumption implies that the two routers have the same failure characteristics and are subject to the same environmental factors.
3. The distribution of S1 and S2 is uniform with parameter θ.

Since the dual-router system consists of two routers of the same type, the total bandwidth to failure T can be expressed as the sum of the bandwidths to failure of the individual routers, S_1 and S_2 with density functions $f_1(s)$ and $f_2(s)$ defined for all $S \in [0, \theta]$. Then the sum, $T = S_1 + S_2$ is a random variable with density function $f_T(t)$, where f_T is the convolution of S_1 and S_2 .

Since $f_1(s) = f_2(s) = \begin{cases} \frac{1}{\theta} & if\ S \in [0, \theta] \\ 0 & Otherwise \end{cases}$

Then, the density function $f_T(t) = (f_1 * f_1)(t) = \int_0^s f_1(s) f_2(t-s) ds$

This a corollary of the transformation theorem states that the density function of

the sum of two continuous random variables is the convolution of the density functions.

If the two random variables are uniformly distributed on [a,b], then the sum is uniformly distributed on [2a,2b]. hence,

$f_1(s) = \frac{1}{\theta}\ \ if\ 0 \leq S \leq \theta,$

Hence $f_T(t) = \frac{1}{\theta} \int_0^t f_2(t-s) ds$

Now the integrand is 0 unless $0 \leq t-s \leq \theta$ (i.e unless $t-\theta \leq s \leq t)$ and then it is $\frac{1}{\theta}$ So if $0 \leq t \leq \theta$, we have

$f_T(t) = \frac{1}{\theta}\int_0^t \frac{1}{\theta} ds = \frac{t}{\theta^2}$

Then for $\theta \le t \le 2\theta$, we have

$f_T(t) = \frac{1}{\theta}\int_{t-\theta}^{\theta} \frac{1}{\theta} ds = \frac{2\theta - t}{\theta^2}$

$$f_T(t) = \begin{cases} \frac{t}{\theta^2} & if\ 0 \le t \le \theta \\ \frac{2\theta - t}{\theta^2} & if\ \theta \le t \le 2\theta \\ 0 & Otherwise \end{cases}$$

Given an experiment with the dual-router-system yielding a sample $T1, T2, \ldots, Tn$ are independent with common pdf $f_T(t_i|\theta)$, the likelihood function for θ is the product of the probability density functions of the individual observations of T:

$$L(\theta) = \prod_{i=1}^{N} f_T(t_i|\theta)$$

$$L(\theta) = \begin{cases} \prod_{i=1}^{N} \frac{t}{\theta^2} & if\ 0 \le t \le \theta \\ \prod_{i=1}^{N} \frac{2\theta - t}{\theta^2} & if\ \theta \le t \le 2\theta \\ 0 & Otherwise \end{cases}$$

$$L(\theta) = \begin{cases} \frac{1}{\theta^{2n}}\left(\prod_{i=1}^{N} t_i\right) & if\ 0 \le t \le \theta \\ \frac{1}{\theta^{2n}}\left(\prod_{i=1}^{N}(2\theta - t)\right) & if\ \theta \le t \le 2\theta \end{cases}$$

The density function of T is not continuously derivable, so using a derivative to calculate the maximum likelihood estimate of θ does not work.

Since the likelihood function is a decreasing function of θ, $L(\theta)$ is maximized when $\theta = \max(T_i)$ for $0 \le t \le \theta$ and also $L(\theta)$ is maximized when $\theta = \min(T_i)$ for $\theta \le t \le 2\theta$.

That is $\hat{\theta} = \max(T_i)$.

So, for the sequence of numbers representing the bandwidth totals to failure as 77, 2, 20, 32, 14,

$\hat{\theta} = \max(T_i) = \max(77, 2, 20, 32, 14) = 77$

Remember to plot theta against L(theta)

The expected value, denoted as E(T), is calculated as follows:

$E(T) = \int_0^{\theta} t * f_T(t)\, dt = \int_0^{\theta} t * \frac{t}{\theta^2}\, dt = \frac{\theta}{3} = 77/3$

https://stats.stackexchange.com/questions/387826/maximum-likelihood-estimator-mle-for-2-theta2-x-3
https://cs.du.edu/~paulhorn/362/362assn6-solns.pdf

https://math.hawaii.edu/~grw/Classes/2018-2019/2019Spring/Math472_1/Assignment10.nb.html

https://mediaspace.baylor.edu/media/Finding+the+maximum+likelihood+estimator+of+the+upper+bound+of+a+uniform%280%2C+B%29+distribution/1_9ku3alr3

Task 4: Hypothesis Test

Over a long period of time, the production of 1000 high-quality hammers in a factory seems to have reached a weight with an average of 816 (in g) and standard deviation of 62.9 (in g). Propose a model for the weight of the hammers including a probability distribution for the weight. Provide all the assumptions needed for this model to hold (even the uncertain ones)? What parameters does this model have?
One aims at answering one of the following questions about a new production system: Does the new system make *less constant* weights?

To answer this question a random sample of newly produced hammers is evaluated yielding the weights in 803, 818, 793, 795, 807, 794, 823, 784, 786, 849.
What hypotheses can you propose to test the question? What test and decision rule can you make to estimate if the new system answers the given question? Express the decision rules as logical statements involving critical values. What error probabilities can you suggest and why? Perform the test and draw the conclusion to answer the question.

SOLUTION

Since the sample size is **n ≥ 30.** To model the weight of the hammers, we can assume that the weights follow a normal distribution. This assumption is based on the Central Limit Theorem, which states that for a large enough sample size, the distribution of the sample mean approaches a normal distribution regardless of the shape of the population distribution.

Assumptions for this model to hold:

1. The weights of the hammers are independent and identically distributed (i.i.d.).
2. The distribution of weights is approximately normal.

The model has two parameters:

- Mean (μ): The average weight of the hammers.
- Standard deviation (σ): The spread of the weights around the mean.

To confirm if the new system makes *less constant* weights, we need to state the hypothesis:

Null Hypothesis (H0): $\sigma_{new}^2 \ = \sigma_{old}^2 = \ 62.9^2$ (The variance of the new system weights is equal to the variance of the old system weights.)
Alternative Hypothesis (HA): $\sigma_{new}^2 \ < \ \sigma_{old}^2$ (The variance of the new system weights is less than the variance of the old system weights.)

[Wolfram inc 2023, calculate standard deviation of {803, 818, 793, 795, 807, 794, 823, 784, 786, 849}], we get $\sigma_{new} = 20.010$

Calculating the F critical value, put the highest variance as the numerator and the lowest variance as the denominator:

F Statistic = $\frac{Variance_{highest}}{Variance_{lowest}} = \frac{62.9^2}{20.01^2} = 9.8811$

The degrees of freedom in the table will be the sample size -1, so:

For $samples_{old}$ with a sample size of 1000, the The degrees of freedom is sample size -1 = 999 and $samples_{new}$ with a sample size of 10, the degree of freedom is 9

And since no alpha was stated in the question, so use 0.05 (the standard "go to" in statistics) and the fact that we are looking for scores "lesser than" a certain point means that this is a one-tailed test.

Finding the critical F Value using the F Table. make sure you look in the alpha = .05 table. Critical F (999,99) at alpha (0.05) = 1.89.

The decision rule for the F-test is as follows:

- If the calculated F-test statistic is greater than the critical value from the F-Table, then reject H0.
- Otherwise, do not reject H0.

F calculated value: 9.8811

F value from table: 1.89

Since the calculated F-test statistic is greater than the critical value from the F-Table, we can reject the null hypothesis.

9. **Calculate the p-value.**

The p-value is the probability of observing a sample standard deviation of 18.98 grams or higher, given that the null hypothesis is true. In this example, the p-value is 0.0007259976652864525.

Python

```
p_value = stats.f.cdf(f_test_statistic, degrees_of_freedom, degrees_of_freedom)
```

https://home.ubalt.edu/ntsbarsh/business-stat/StatistialTables.pdf

Task 5: Regularized Regression

Given the values of an unknown function $f:\mathbb{R}\rightarrow\mathbb{R}$ at some selected points, we try to calculate the parameters of a model function using OLS as a distance and a ridge regularization:

a polynomial model function of eleven α_i parameters:

$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_{10} x^{10}$

Calculate the OLS estimate, and the OLS ridge-regularized estimates for the parameters given the sample points of the graph of f given that the values are y = (-4, 2270146.72), (15, 2597371459585), (-9, 11756031746.56), (4, 4928348.79), (14, 1305326827257.86), (19, 26954912028633.17), (-11, 89948619287.32), (-16, 3914542381288.27), (-1, -13.98), (-3, 79342.22), (16, 4563983667885.92), (-5, 26294495.16), (-8, 3529689518.48), (5, 46525773.07), (-2, -918.13), (6, 264064520.43), (1, 0), (-13, 495396031253.45), (-18, 12916231635918.5), (20, 41703330897817.15), (17, 8258889165333.91), (8, 5044492419.44)

Provide a graphical representation of the graphs of the approximating functions and the data points. Remember to include the steps of your computation which are more important than the actual computations.

SOLUTION

Given the polynomial model function is defined as:

$$f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{10} x^{10}$$

where $\alpha_0, \alpha_1, \alpha_2 \ldots \alpha_{10}$ are the parameters of the model

To calculate the Ordinary Least Squares (OLS) estimate for the parameters of the polynomial model function, we need to minimize the sum of squared residuals between the observed values and the predicted values from the model.

$$\hat{\alpha} = \arg\min_{\alpha} \left[\sum_{i=0}^{n} \left(y_i - f(x_i) \right)^2 \right]$$

where the index i runs over all data points in our data set, y_i is the observed value and $f(x_i)$ is the predicted value.

The sum of squares of n known data points is given by:

$$RSS = \sum_{i=0}^{n} \left(y_i - f(x_i) \right)^2$$

Substituting $f(x)$ in the equation above,

$$RSS = \sum_{i=0}^{n} (y_i - \alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10})^2$$

To find the OLS estimate, we perform a partial derivative of the RSS equation with respect to each parameter α_i and set the derivatives equal to zero. Using power rule

$$\frac{\partial}{\partial \alpha_i} \left(\sum_{i=0}^{n} (y_i - \alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10})^2 \right) = 0$$

$$\frac{\partial RSS}{\partial \alpha_0} = \sum_{i=0}^{n} (2\,(\alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10} - y_i) \frac{\partial}{\partial \alpha_0} (\alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10} - y_i))$$

$$= \sum_{i=0}^{n} (\alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10} - y_i) = 0$$

Similarly,

$$\frac{\partial RSS}{\partial \alpha_1} = \sum_{i=0}^{n} (2\,(\alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10} - y_i) \frac{\partial}{\partial \alpha_1} (\alpha_0 + \alpha_1 x_i + \alpha_2 {x_i}^2 + \cdots + \alpha_{10} {x_i}^{10} - y_i))$$

$$= \sum_{i=0}^{n} (\alpha_0 x_i + \alpha_1 {x_i}^2 + \alpha_2 {x_i}^3 + \cdots + \alpha_{10} {x_i}^{11} - x_i y_i) = 0$$

Continuing in this manner,

$$\frac{\partial RSS}{\partial \alpha_2} = \sum_{i=0}^{n} (\alpha_0 {x_i}^2 + \alpha_1 {x_i}^3 + \alpha_2 {x_i}^4 + \cdots + \alpha_{10} {x_i}^{12} - {x_i}^2 y_i) = 0$$

$$\frac{\partial RSS}{\partial \alpha_{10}} = \sum_{i=0}^{n} (\alpha_0 {x_i}^{10} + \alpha_1 {x_i}^{11} + \alpha_2 {x_i}^{12} + \cdots + \alpha_{10} {x_i}^{20} - {x_i}^{10} y_i) = 0$$

We can represent the above equations in matrix form use a matrix method, however we could use scipy.optimize.curve_fit

The OLS ridge-regularized estimates is

$$\hat{\alpha} = \arg\min_{\alpha} \left[\sum_{i=0}^{n} \left(y_i - f(x_i)\right)^2 + penalty(\alpha) \right]$$

$$= \arg\min_{\alpha} \left[\sum_{i=0}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \sum_{k=0}^{K} {\alpha_k}^2 \right]$$

where λ is a positive constant.

Use the task5 code to get the estimate

https://data36.com/polynomial-regression-python-scikit-learn/

https://www.statology.org/polynomial-regression-python/

https://muthu.co/maths-behind-polynomial-regression/

https://muthu.co/simple-example-of-polynomial-regression-using-python/

https://www.public.asu.edu/~gwaissi/ASM-e-book/module403.html

https://www2.stat.duke.edu/~hc95/Teaching/STA103/lec14_notes.pdf

https://www.geo.fu-berlin.de/en/v/soga-py/Basics-of-statistics/Linear-Regression/Polynomial-Regression/Polynomial-Regression---An-example/index.html

https://towardsdatascience.com/maximum-likelihood-vs-bayesian-estimation-dd2eb4dfda8a

https://home.iitk.ac.in/~kundu/gamma-bayes-rev-1.pdf

https://home.iitk.ac.in/~kundu/gamma-bayes-rev-1.pdf

https://math.stackexchange.com/questions/4303041/derive-bayes-estimator-with-a-gamma-prior

Question 6:

Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size n = 10 from a gamma distribution with α = 3 and β = 1/θ. Suppose we believe that θ has a gamma distribution with α = 77 and β = 91.

(a) Find the posterior distribution of θ.

(b) If the observed x = 18.2, what is the Bayes point estimate associated with square-error loss function?

(c) What is the Bayes point estimate using the mode of the posterior distribution?

Hint: Can the posterior distribution be related to a chi-square distribution?

Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a gamma distribution with α=3 and β=1/θ. Suppose we believe that θ follows a gamma-distribution with α = 77 and β = 91 and suppose we have a trial $(x_1,\ldots,x_n)$ with an observed $\bar{x}$=62.88.

a) Find the posterior distribution of θ.

b) What is the Bayes point estimate of θ associated with the square-error loss function?

c) What is the Bayes point estimate of θ using the mode of the posterior distribution?

https://www.scielo.br/j/aabc/a/FptKv7xMmRC3BjsVbSsMDkv/?format=pdf&lang=en

https://michael-franke.github.io/intro-data-analysis/ch-03-03-estimation-bayes.html#

https://math.stackexchange.com/questions/3992755/posterior-for-gamma-prior-and-gamma-likelihood-with-known-shape.

Solution:

The observed data X_i has a gamma distribution with the shape parameter $\alpha > 0$ and an (inverse) scale parameter $\beta > 0, and\ \theta = \frac{1}{\beta}$. The Gamma density can be written as follows:

$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$; $x > 0$. Here $\Gamma(\alpha)$ is the gamma function.

Thus, the likelihood of the observed data is

$$P(X|\theta) = \mathcal{L}(x_1, x_1, \ldots x_1|\theta) = P(x_1|\theta)P(x_2|\theta) \ldots P(x_n|\theta)$$

$$\frac{\theta^{n\alpha}}{\left(\Gamma(\alpha)\right)^n} e^{-\theta \sum_{i=1}^{n} x_i} \left(\prod_i^n x_i\right)^{\alpha-1}$$

The prior pdf θ is a gamma distribution given by

$$P(\theta) = \frac{\theta^{(\alpha-1)}\, e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$$

The equation used for Bayesian estimation takes on the same form as bayes theorem.

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_0^\infty P(X|\theta)P(\theta)d\theta}$$

Where $P(\theta|X)$ the posterior distribution, $P(X|\theta)$ the likelihood function, and $P(\theta)$ the prior distribution.

Hence,

$$P(\theta|X) = \frac{\frac{\theta^{n\alpha}}{\left(\Gamma(\alpha)\right)^n} e^{-\theta \sum_{i=1}^{n} x_i} (\prod_i^n x_i)^{\alpha-1} * \frac{\theta^{(\alpha-1)}\, e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}}}{\int_0^\infty (\frac{\theta^{n\alpha}}{\left(\Gamma(\alpha)\right)^n} e^{-\theta \sum_{i=1}^{n} x_i} \left(\prod_i^n x_i\right)^{\alpha-1} * \frac{\theta^{(\alpha-1)}\, e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}}) d\theta}$$

By definition, $Posterior \propto Prior * Likelihood$ and wasting away any quantity not depending on θ thus

$$P(\theta|X) \propto \theta^{n\alpha} e^{-\theta \sum_{i=1}^{n} x_i} * \theta^{(\alpha-1)}\, e^{-\theta/\beta}$$

$$= \theta^{n\alpha} e^{-\theta \sum_{i=1}^{n} x_i} * \theta^{(\alpha-1)}\, e^{-\theta/\beta}$$

$$= \theta^{30} e^{-\theta \sum_{i=1}^{n} x_i} * \theta^{76}\, e^{-\theta/91}$$

Therefore $P(\theta|X) \propto \theta^{106} * e^{-\theta(\frac{1}{91} + \sum_{i=1}^{n} x_i)}$

Now, comparing the above expression with Gamma (α, β) distribution's PDF, we immediately recognize that the posterior is a *Gamma*[107; $\frac{1}{\frac{1}{91} + \sum_{i=1}^{n} x_i}$]

The Bayes point estimate associated with the square-error loss function is the posterior mean:

$$\hat{\theta} = \mathrm{E}[\theta|X] = \alpha\beta = \frac{107}{\frac{1}{91} + \sum_{i=1}^{n} x_i}$$

Since $\bar{x} = 18.2$, then $\sum_{i=1}^{n} x_i \cong n\bar{x} = 10 * 18.2 = 182$

Finally the desired point estimate is

$$\mathrm{E}[\theta|X] = \frac{107}{\frac{1}{91} + 182}$$

the Bayes point estimate of θ using the mode of the posterior distribution is given by: The mode of the posterior distribution is a value which maximizes the posterior density function.

θ_{max} = $(\alpha - 1)\beta$ = $\frac{106}{\frac{1}{91}+\sum_{i=1}^{n} x_i}$

c) The mode of the posterior distribution is a value which maximizes the posterior density function.

From the a) part we see that the posterior density function is

$$k(\theta \mid \mathbf{x}) = \frac{1}{\Gamma(40) \cdot \widetilde{\beta}^{40}} \cdot \theta^{39} \cdot e^{-\theta \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)}.$$

We see that maximizing this posterior density (in terms of θ) is the same as maximizing

$$g(\theta) = \theta^{39} \cdot e^{-\theta \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)},$$

since the constants do not affect maximization.

The derivative of g is

$$g'(\theta) = 39 \cdot \theta^{38} \cdot e^{-\theta \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)} - \theta^{39} \cdot e^{-\theta \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)} \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)$$

$$= \theta^{38} \cdot e^{-\theta \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)} \cdot \left(39 - \theta \cdot \left(\frac{1}{2} + \sum_{i=1}^{n} x_i\right)\right).$$

Stationary points are the null-points of the above derivative, so

$$g'(\theta) = 0 \iff \theta = 0 \quad \text{or} \quad \theta = \frac{39}{\frac{1}{2} + \sum_{i=1}^{n} x_i}.$$

Obviously, $g(0) = 0$, so g cannot reach its global maximum at $\theta = 0$. Thus, the point where it reaches its maximum is

$$\theta_{\max} = \frac{39}{\frac{1}{2} + \sum_{i=1}^{n} x_i} = \frac{39}{\frac{1}{2} + 182} = \boxed{0.214}.$$