National University of Computer and Emerging Sciences

FAST School of Computing **Spring-2023** **Islamabad Campus**

MT-2002: Statistical Modeling

Serial No:

Sessional Exam-II
Total Time: 1 Hour
Total Marks: 55

Monday, 10th April, 2023

Course Instructors

Mukhtar Ullah, Shahzad Saleem, Muhammad Almas Khan

Signature of Invigilator

Asadullah Nawaz	20I-0761	C	Asad
Student Name	Roll No.	Course Section	Student Signature

DO NOT OPEN THE QUESTION BOOK OR START UNTIL INSTRUCTED.

Instructions:

1. Attempt on question paper. Attempt all of them. Read the question carefully, understand the question, and then attempt it.
2. No additional sheet will be provided for rough work. Use the back of the last page for rough work.
3. If you need more space, write on the back side of the paper and clearly mark question and part number etc.
4. After asked to commence the exam, please verify that you have **eight (8)** different printed pages including this title page. There are a total of **3** questions.
5. Calculator sharing is strictly prohibited.
6. Use permanent ink pens only. Any part done using soft pencil will not be marked and cannot be claimed for rechecking.

	Q-1	Q-2	Q-3	Total
Marks Obtained	19	18	15	52
Total Marks	20	20	15	55

Scanned with CamScanner

National University of Computer and Emerging Sciences

FAST School of Computing Spring-2023 Islamabad Campus

19

Question 1 [20 Marks]

The following code snippets implement six models labelled **Model_1** through **Model_6**.

```
with pm.Model() as Model_1:
    α = pm.Normal('α', mu=0, sd=10)
    β = pm.Normal('β', mu=0, sd=1)
    ϵ = pm.HalfNormal('ϵ', 5)

    μ = pm.Deterministic('μ', α + β * x)

    y_pred = pm.Normal('y_pred', mu=μ, sd=ϵ, observed=y)
    idata_g = pm.sample(2000, tune=2000, return_inferencedata=True)
```

```
with pm.Model() as Model_2:
    α = pm.Normal('α', mu=0, sd=10)
    β = pm.Normal('β', mu=0, sd=10)

    μ = α + pm.math.dot(x_c, β)
    θ = pm.Deterministic('θ', pm.math.sigmoid(μ))
    bd = pm.Deterministic('bd', -α/β)

    yl = pm.Bernoulli('yl', p=θ, observed=y_0)

    idata_0 = pm.sample(1000, return_inferencedata=True)
```

```
with pm.Model() as model_3:
    α_tmp = pm.Normal('α_tmp', mu=0, sd=10)
    β = pm.Normal('β', mu=0, sd=1, shape=2)
    ϵ = pm.HalfCauchy('ϵ', 5)

    μ = α_tmp + pm.math.dot(X_centered, β)

    α = pm.Deterministic('α', α_tmp - pm.math.dot(X_mean, β))

    y_pred = pm.Normal('y_pred', mu=μ, sd=ϵ, observed=y)

    idata = pm.sample(2000, return_inferencedata=True)
```

```
with pm.Model() as model_4:
    α = pm.Normal('α', mu=0, sd=10)
    β = pm.Normal('β', mu=0, sd=2, shape=len(x_n))

    μ = α + pm.math.dot(x_1, β)
    θ = pm.Deterministic('θ', 1 / (1 + pm.math.exp(-μ)))
    bd = pm.Deterministic('bd', -α/β[1] - β[0]/β[1] * x_1[:,0])

    yl = pm.Bernoulli('yl', p=θ, observed=y_1)

    idata_1 = pm.sample(2000, target_accept=0.9, return_inferencedata=True
```

```
with pm.Model() as model_5:
    α = pm.Normal('α', mu=y_2.mean(), sd=1)
    β1 = pm.Normal('β1', mu=0, sd=1)
    β2 = pm.Normal('β2', mu=0, sd=1)
    ϵ = pm.HalfCauchy('ϵ', 5)

    mu = α + β1 * x_2 + β2 * x_2**2

    y_pred = pm.Normal('y_pred', mu=mu, sd=ϵ, observed=y_2)

    idata= pm.sample(2000, return_inferencedata=True)
```

```
with pm.Model() as Model_6:
    ψ = pm.Beta('ψ', 1, 1)
    α = pm.Normal('α', 0, 10)
    β = pm.Normal('β', 0, 10, shape=2)
    θ = pm.math.exp(α + β[0] * data['x1'] + β[1] * data['x2'])
    yl = pm.ZeroInflatedPoisson('yl', ψ, θ, observed=data['y'])
    idata= pm.sample(1000, return_inferencedata=True)
az.plot_trace(idata);
```

Answer the following. [1 + 1 + 3 × 6]

a) Which (of the above) models qualify as linear models and why?

1 All models qualify as linear models except
3 model_5 which is for polynomial regression.
5 The models are linear because they use $\alpha + \beta x$
i.e the equation of line as mean of likelihood.

1

b) Which models qualify as generalized linear models and why?

Model_2, Model_4 and Model_6 are generalized linear models because they will work for discrete output variable as well and use inverse link function such as sigmoid and exponential function

Scanned with CamScanner

National University of Computer and Emerging Sciences

FAST School of Computing Spring-2023 Islamabad Campus

c) Which models are the ones for simple linear regression? Reproduce here only the relevant pieces of code for each model to justify your answer.

$\boxed{\text{Model - 1}}$ is for simple linear regression as there is only one independent variable and one continuous dependent variable. we are also just using identity and no other inverse link function so it is simple regression model

code => μ = pm.Deterministic ('μ', $\alpha + \beta * x$)
when β is unidimensional and x is vector

d) Which models are the ones for multiple linear regression? Reproduce here only the relevant pieces of code for each model to justify your answer.

$\boxed{\text{Model - 3}}$ is for multiple linear regression as there are more than one independent variable. We are using vector of coefficients (shape = 2) and matrix for independent variables and computing dot product for μ.

$$\mu = \alpha + \beta X \quad \text{where} \quad \beta X = \sum_{i=1}^{n} b_i x_i = b_1x_1 + b_2x_2 + \dots b_m x_m$$

e) Which models are the ones for simple logistic regression? Reproduce here only the relevant pieces of code for each model to justify your answer.

$\boxed{\text{Model - 2}}$ is for simple logistic regression as it is using single independent variable with logistic / sigmoid function and we are feeding it to bernoulli likelihood

code:

```
β = pm.Normal('β', mu=0, sd=10);
θ = pm.Deterministic('θ', pm.math.sigmoid(μ))
y1 = pm.Bernoulli('y1', p=θ, observed=y-0);
```


Code:

```
β = pm.Normal('β', mu=0, sd=1, shape=2);
μ = α+tmp + pm.math.dot(x-centred, β)
```

Scanned with CamScanner

f) Which models are the ones for multiple logistic regression? Reproduce here only the relevant pieces of code for each model to justify your answer.

Model-4 is used for multiple logistic regression as it is for multiple independent variables, i.e. vector of β coefficients, which are then used in sigmoid function and θ is then feded to Bernoulli likelihood.

```
        β = pm.Normal('β', mu=0, sd=2, shape(len-xn));
code => θ = pm.Deterministic('θ', 1/(1+pm.math.exp(-μ)));
        y1 = pm.Bernoulli('y', p=θ, observed=y1);
```

g) Which models are the ones for polynomial regression? Reproduce here only the relevant pieces of code for each model to justify your answer.

Model-5 is used for polynomial regression as we have multiple β coefficients used for increasing power of same independent variable

```
        β1 = pm.Normal('β1', mu=0, sd=1);
code => β2 = pm.Normal('β2', mu=0, sd=1);
        mu = α + β1 * x_2 + β2 * x_2 ** 2;
```

h) Which models are extensions for Poisson regression? Reproduce here only the relevant pieces of code for each model to justify your answer.

Model-6 is used for poisson regression. Specifically zero inflated poisson regression (ZIP). It is being used for count of data, with multiple independent variables. The inverse link function is simply exponential function

```
        β = pm.Normal('β', 0, 10, shape=2)
code => θ = pm.math.exp(α + β[0]*data['x1'] + β[1]*data['x2']);
        y1 = pm.ZeroInflatedPoisson('y1', ψ, θ, observed=data['y']);
```


Scanned with CamScanner

National University of Computer and Emerging Sciences

FAST School of Computing **Spring-2023** **Islamabad Campus**

Question 2 [20 Marks]

18

a) A sample of students is measured for height each year for 3 years. After the third year, you want to fit a linear regression predicting height using year as a predictor. Write down the model definition (using statistical notations) for this regression, using any variable names and priors you choose. **[10]**

Note: You also need to justify your choice of priors.

Since, height is continuous variable, and we only have one independent variable / predictor, we can use both simple linear regression model or robust linear regression model. Since, we are unaware of data, it is better choice to go for robust linear regression model as it deals with outliers as well.

9

Priors :-

$\alpha \sim$ Normal $(\mu_\alpha, \sigma_\alpha)$

$\beta \sim$ Normal $(\mu_\beta, \sigma_\beta)$

$\epsilon \sim$ Half Cauchy (σ_ϵ)

$v \sim$ Exponential (λ)

$\mu = \alpha + \beta x$ → Deterministic

Likelihood :-

~~y ~ Norm~~

$y \sim$ StudentT (mu $= \mu$, sd $= \epsilon$, v)

The choice of priors is simple as we are unaware of data so we can use weakly informative priors α and β. ϵ can be calculated using both half normal and half-cauchy but half-cauchy suits better. v is normality parameter calculated better using exponential. we can also use shifted exp...

Scanned with CamScanner

b) Rank in order (from lowest to highest on a scale of 0 to 6) based on correlation in the following figure. You should write your answer in the boxes below the figure. **[7]**

$\boxed{r=1}$ $\boxed{r=-1}$ $r \to -1$ $\boxed{r=0}$ $r \to 1$ $r \to -1$ $r \to 1$

Well Done.

6 0 2 3 5 1 4

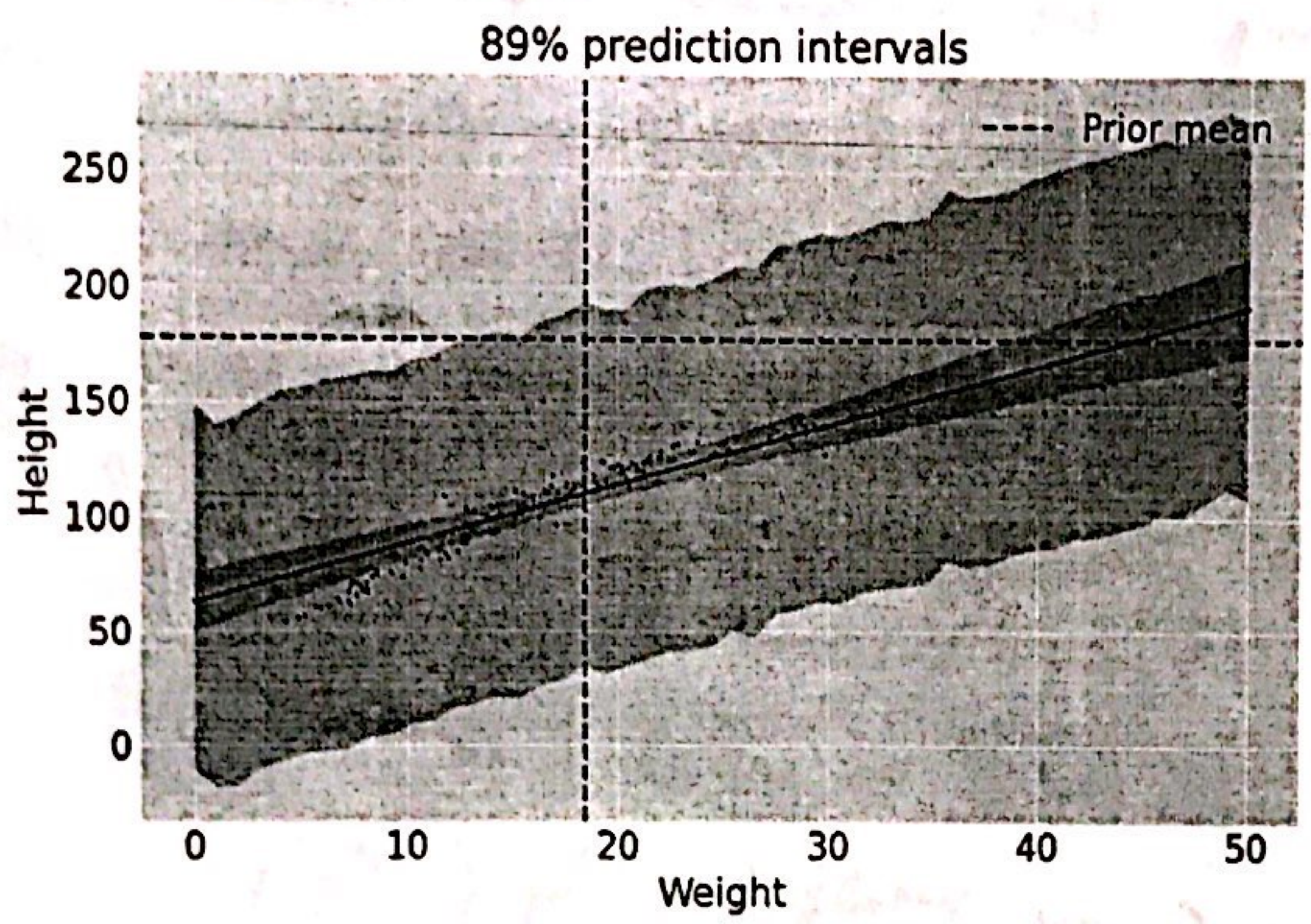
6	0	2	3	5	1	4

I have ordered using pearlson's coefficie

$r \approx -1 < 0 < 1$

c) Following model resulted in a prediction interval plot shown below

$$\alpha \sim Normal(178, 20)$$
$$\beta \sim Normal(0, 10)$$
$$\sigma \sim Uniform(0, 50)$$
$$\mu_i = \alpha + \beta(x_i - \bar{x}), \text{ where } \bar{x} \text{ is the mean of } x_i$$
$$y_i \sim Normal(\mu_i, \sigma)$$



As can be seen, this gives absurdly large 89% prediction interval. Suggest a solution to reduce the prediction interval. [3]

We need to make the model more robust using student's t distribution. Also, sd being calculated using uniform distribution puts strong limit on data spread. Hence, we can ~~define~~ redefine prior for sd as:

See Solution

$\sigma \sim HalfCauchy(\delta_6)$

we will need a normality parameter v:

$v \sim Exponential(\lambda)$

Finally, we can use student's T as likelihood:

$y_i \sim StudentT(\mu_i, \delta, v)$

Scanned with CamScanner

National University of Computer and Emerging Sciences

FAST School of Computing **Spring-2023** **Islamabad Campus**

Question 3 [15 Marks]

a) Instead of a logistic regression we use a simple linear regression model to solve a classification problem. Under the simple linear regression model, will we get the same or different results? **[5]**

15

Simple regression models are not suited for classification problems since they predict continuous variables while classification problem classifies into discrete classes, hence, we won't get same results. we won't be able to classify rather we will get intermediate real values. ~~=~~ These values cannot predict the class. Therefore, we need some inverse link function for classification problem.

5

b) Suppose we have following model that classifies the patients as *Normal* or *Affected* by a viral infection, based on the features selected from their blood samples data.

$$\alpha \sim Normal(\mu, \sigma)$$
$$\beta \sim Normal(\mu, \sigma)$$
$$\mu = \alpha + \beta(x - \bar{x}), \text{where } \bar{x} \text{ is the mean of } x$$
$$\theta \sim sigmoid(\mu)$$
$$bd \sim -\frac{\alpha}{\beta}$$
$$y \sim Bern(\theta)$$

Now if we want to extend our predictions to three outcomes, i.e., *Normal, Mild,* and *Severe,* then redefine the model. **[10]**

Since, the number of classes has increased, we can use softmax regression model with categorical distribution which deals with multiple classes. Softmax regression uses softmax function as inverse link which is given as:-

10

$$softmax(u_i) = \frac{\exp(u_i)}{\sum_k \exp(u_k)}$$

Now, we will define our model:-

Scanned with CamScanner

$\alpha + \beta_1 x_1 + \beta_2 x_2 = 0$

$\beta_2 x_2 = -\alpha - \beta_1 x_1$

Priors:-

The priors are

$$\alpha \sim \text{Normal}(\mu_\alpha, s_\alpha)$$
$$\beta \sim \text{Normal}(\mu_\beta, s_\beta)$$
$$\mu = \alpha + \beta(x - \bar{x})$$
$$\theta \sim \text{softmax}(\mu)$$
$$bd \sim -\frac{\alpha}{\beta}$$

Well Done

Likelihood:-

$$y \sim \text{categorical}(\theta)$$

So, with only ~~one~~ two changes, we can also deal with ~~problem~~ increasing classes i.e change sigmoid to softmax and change Bernoulli to categorical distribution.

We can also extend our softmax regression model to include more than one features or independent variables. For 2 features/independent variables:-

Priors:-

$$\alpha \sim \text{Normal}(\mu_\alpha, s_\alpha)$$
$$\beta_1 \sim \text{Normal}(\mu_{\beta_1}, s_{\beta_1})$$
$$\beta_2 \sim \text{Normal}(\mu_{\beta_2}, s_{\beta_2})$$
$$\mu = \alpha + \beta_1(x_1 - \bar{x}_1) + \beta_2(x_2 - \bar{x}_2)$$
$$\theta \sim \text{softmax}(\mu)$$
$$bd \sim -\frac{\alpha}{\beta_2} - \frac{\beta_1 x_1}{\beta_2}$$

Likelihood:-

$$y \sim \text{categorical}(\theta)$$

Scanned with CamScanner