1	2	3	4	Total

Name:
Number: Answers

BLG560E - Statistics and Estimation in Computer Science

Midterm 1

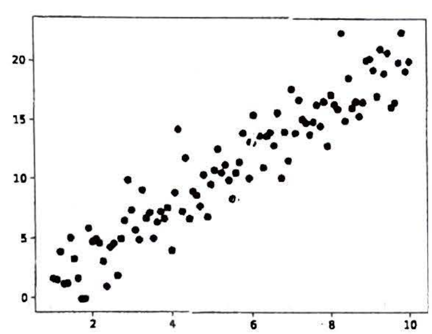
Rules:

- Duration is 90 min.
- Show your work, do not write any result directly.
- Use the allocated space after each question. Do not write answers outside the given frames.

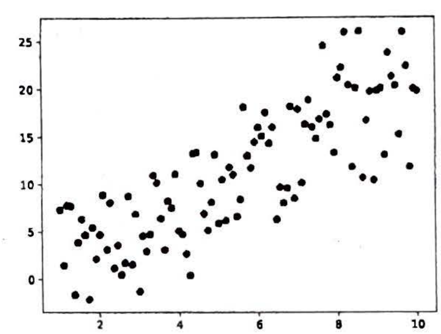
Questions:

1. (20 pts) Consider 4 different sets of bivariate data whose scatter graphs are given below. The correlation coefficientsof these datasets are r_1, r_2, r_3 and r_4, respectively. The coefficients are given below each figure for each dataset.

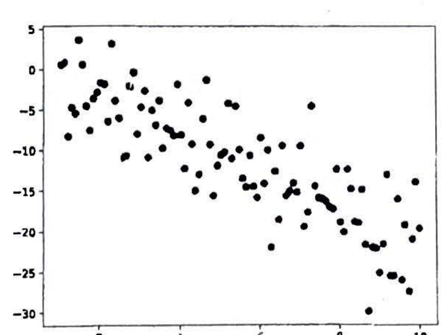
   Remember that the correlation coefficient $r = Cov(X, Y)/(\sigma_X \sigma_Y)$.



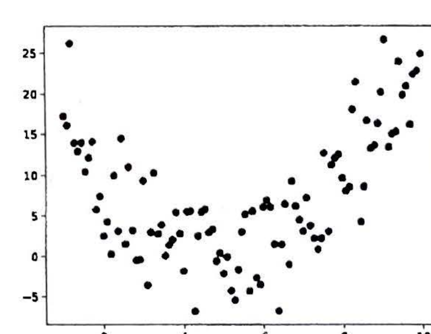
r_1



r_2



r_3



r_4

Sort the correlation coefficients (r_1, r_2, r_3 and r_4) in the ascending order considering their signs.

$r_3 < r_4 < r_2 < r_1$

2. (30 pts) Let X_i be independent and identically distributed (iid) random variables (rv) with exponential distribution, $X_i \sim Exp(\lambda_x)$. Similarly let Y_i are also iid rv, $Y_i \sim Exp(\lambda_y)$. X_i and Y_i are also mutually independent. Consider a new rv such as

$$Z = \frac{1}{N_1}\sum_{i=1}^{N_1} X_i - \frac{1}{N_2}\sum_{i=1}^{N_2} Y_i$$

Note that if a random variable $K \sim Exp(\lambda)$ then

$$f(K = k) = \begin{cases} \lambda \exp\{-\lambda k\} & k \geq 0 \\ 0 & x < 0 \end{cases}$$

with $E(K) = 1/\lambda$ and $\sigma_K^2 = 1/\lambda^2$.

(a) Find the expected value of Z.

$$E(Z) = \frac{1}{N_1}\sum_{i=1}^{N_1} \underbrace{E(X_i)}_{1/\lambda_x} - \frac{1}{N_2}\sum_{i=1}^{N_2} \underbrace{E(Y_i)}_{1/\lambda_y}$$

$$= \frac{N_1/\lambda_x}{N_1} - \frac{N_2/\lambda_y}{N_2} = \frac{1}{\lambda_x} - \frac{1}{\lambda_y}$$

(b) Find the variance of Z.

$$var\left(\frac{1}{N_1}\sum_{i=1}^{N_1} X_i\right) = \frac{1}{N_1} var(X_i) \quad \text{as } X_i \text{ is iid.}$$

$$Z = \mathcal{X} - \mathcal{Y} \quad \text{where } \mathcal{X} = \frac{1}{N_1}\sum_i^{N_1} X_i \ \& \ \mathcal{Y} = \frac{1}{N}\sum_i^{N_2} Y_i$$

$var(Z) = var(\mathcal{X}) + var(\mathcal{Y})$ as $\mathcal{X}$ & $\mathcal{Y}$ are independent

$$var(Z) = \frac{1}{N_1}\frac{1}{\lambda_x^2} + \frac{1}{N_2}\frac{1}{\lambda_y^2}$$

(c) State the approximate distribution of Z given that N_1 and N_2 are sufficiently large. State your reason.

$$Z \sim \mathcal{N}\left(\frac{1}{\lambda_x} - \frac{1}{\lambda_y}, \frac{1}{N_1}\frac{1}{\lambda_x^2} + \frac{1}{N_2}\frac{1}{\lambda_y^2}\right)$$ due to central limit theorem

1	2	3	4	Total

Name:
Number: Answers

3. (30 pts) Consider a software company where issues (such as bugs etc.) are resolved every day. The administrators are curious about the expected value of number-of-issues resolved per day. Hence, they analyze the frequency of days versus number-of-issues resolved in 2018.

number of days in 2018 (out of 365 days)	number of issues resolved in that day
100	0
150	1
60	2
40	3
5	4

Lets assume that the number-of-issues resolved per day can be modelled using Poisson distribution. Hence

$$P(\#\text{ of resolved issues per day}=k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where λ is the parameter of interest ie. the expected value of number-of-issues resolved per day.

a) Find the formula for the maximum likelihood estimator of λ, ($\hat{\lambda}_{MLE}$).

Let $k_i = \#$ of issues solved at day i

$N = 365$

then likelihood is

$$L(\lambda) = \prod_{i=1}^{N} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!}$$

log likelihood

$$\ell(\lambda) = \sum_{i=1}^{N} k_i \log\lambda - \lambda - \log(k_i!)$$

$$\frac{\partial}{\partial\lambda}\ell(\lambda) = \sum_{i=1}^{N} \frac{k_i}{\lambda} - 1 = 0$$

$$= \frac{1}{\lambda}\left(\sum_{i=1}^{N} k_i\right) - N = 0 \Rightarrow \hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{N} k_i}{N}$$

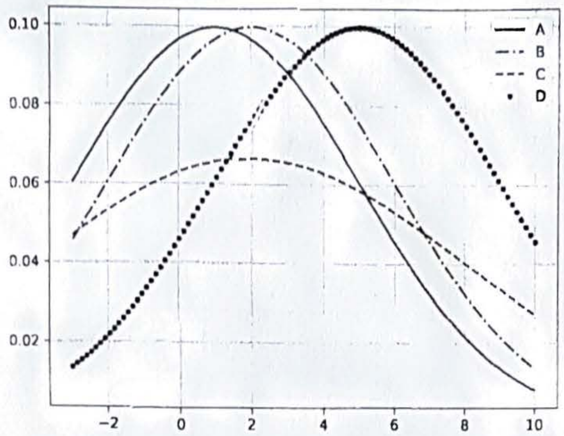
b) Using the data in the table given above, compute the maximum likelihood estimate of number-of-issues resolved per day ($\hat{\lambda}_{MLE}$).

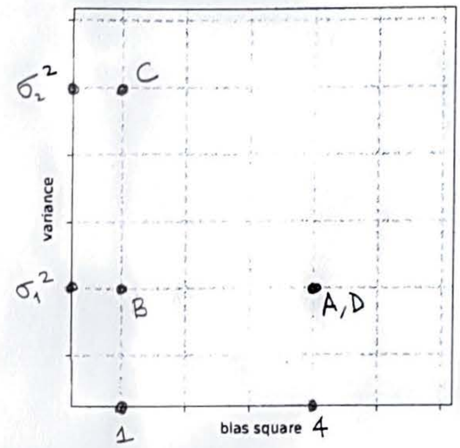
From part (a) and table

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{N} k_i}{365} = \frac{1}{365}\left(100\times0 + 150\times1 + 60\times2 + 40\times3 + 5\times4\right)$$

$$= \frac{410}{365}$$

4. (20 pts) Consider 4 different estimators A, B, C, D whose distributions are given below. Assume that the correct population parameter is 3.





(a) Mark the estimators on the bias^2 vs variance graph given above. The expected values of A, B, C, D estimators are $\mu_A = 1, \mu_B = 2, \mu_C = 2, \mu_D = 5$. Furthermore, assume the variances of the estimators are $\sigma_A^2 = \sigma_1^2, \sigma_B^2 = \sigma_1^2, \sigma_C^2 = \sigma_2^2, \sigma_D^2 = \sigma_1^2$. Mark important values on the horizontal and vertical axis.

(b) Show that mean square error (MSE) is equal to bias sqaured plus variance ($MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta})$) where $\hat{\theta}$ is the estimator of θ.

$$MSE = E\left((\theta - \hat{\theta})^2\right) = \underbrace{E(\theta^2)}_{\theta^2} - 2\theta E(\hat{\theta}) + \overbrace{E(\hat{\theta}^2)}$$

$$= \underbrace{\theta^2 - 2\theta E(\hat{\theta}) + E^2(\hat{\theta})} + var(\hat{\theta})$$

$$= bias^2(\hat{\theta}) + var(\hat{\theta})$$