CS 215 : Data Analysis and Interpretation

(Instructor : Suyash P. Awate)

End-Semester Examination (Closed Book)

Roll Number: ____________________

Name: ____________________

For all questions, if you feel that some information is missing, make justifiable assumptions, state them clearly, and answer the question.

Relevant Formulae

• Poisson: $P(k|\lambda) := \lambda^k \exp(-\lambda)/(k!)$

• Exponential: $P(x;\lambda) = \lambda \exp(-\lambda x); \forall x > 0$

• Gamma:

$$P(x;\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}$$

where $x > 0, \alpha > 0, \beta > 0$

• Gamma function: $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$ for real-valued z.
When z is integer valued, then $\Gamma(z) = (z-1)!$, where ! denotes factorial.
For all z, $\Gamma(z+1) = z\Gamma(z)$.

• Univariate Gaussian:

$$P(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-0.5\frac{(x-\mu)^2}{\sigma^2}\right)$$

• Product of two univariate Gaussians: $G(z;\mu_1,\sigma_1^2)G(z;\mu_2,\sigma_2^2) \propto G(z;\mu_3,\sigma_3^2)$
where

$$\mu_3 = \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}; \sigma_3^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

• Multivariate Gaussian:

$$G(x;\mu,C) = \frac{1}{(2\pi)^{D/2}|C|^{0.5}} \exp(-0.5(x-\mu)^\top C^{-1}(x-\mu))$$

• $d(Ax) = Adx$
• $d(x^\top Ax) = x^\top(A+A^\top)dx$
• Laplace: $P(x;\mu,b) := (0.5/b)\exp(-|x-\mu|/b)$, where $b > 0$ and $x \in \mathbb{R}$
• "i.i.d." stands for "independent and identically distributed"
• "PDF" stands for "probability density function"
• "PMF" stands for "probability mass function"

---

1. [10 points]

   • (5 points) Given the analytical form for the Gamma PDF underlying random variable X, mathematically derive the analytical form for the random variable $Y := 1/X$.

   ---

   inverse Gamma. see lecture slides.

   • (5 points) Mathematically derive (from scratch, without using any known result) a closed-form mathematical expression for the mode of the PDF of Y.

   `https://en.wikipedia.org/wiki/Inverse-gamma_distribution#Characterization`

   `https://www.youtube.com/watch?v=TA97OE5RDSM`

   ---

2. [10 points]

- (1 point) Mathematically define the principal mode of variation for a 2D dataset $\{(x_n \in \mathbb{R}, y_n \in \mathbb{R})\}_{n=1}^{N}$.

---

see lecture slides for the formulation as an optimization problem

---

For each of the following PDFs, indicate the directions corresponding to the first/principal and the second mode of variation. Also indicate if they are unique or not. Give a clear mathematical justification (including a proof where applicable) in each case.

- (3 points) Uniform PDF over the interior of the quadrilateral with vertices at $(0, 0)$, $(2, 0)$, $(2, 1)$, and $(0, 1)$.

---

independence of X and Y implies covariance matrix is diagonal. (1 point)

diagonal entries on covariance matrix, i.e., variances of X and Y, are unequal. (1 point)

so, unique directions/eigenvectors. principal direction along X axis. (1 point)

- (3 points) Uniform PDF over the interior of the quadrilateral with vertices at $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$.

---

see lecture slides.

covariance matrix is proportional to identity matrix. use property of change in covariance upon rotation. see lecture slides. (2 points)

non-unique directions/eigenvectors. any direction is a principal mode of variation. (1 point)

- (3 points) Uniform PDF over the interior of the ellipse with end points of its major/minor axes at $(2, 0)$, $(0, 1)$, $(-2, 0)$, and $(0, -1)$.

---

covariance matrix is diagonal, but diagonal entries on covariance matrix, i.e., variances of X and Y, are unequal. (2 points)

so, unique directions/eigenvectors. principal direction along X axis. (1 point)

[think about a special case of the ellipse where one of the major axes has length tending to zero]

---

3. [20 points]

• (5 points) For a Gaussian PDF $\mathcal{N}(\mu, \theta)$ with known mean μ and unknown variance θ, what prior will you choose on the variance so that it is a conjugate prior ? Using this conjugate prior, mathematically derive the mode of the posterior PDF on θ.

---

inverse Gamma. show that it is conjugate; can rely on previous problem to find its functional form. (2 points)

deriving posterior mode (3 points)

see lecture slides.

• (8 points) For a Gaussian PDF $\mathcal{N}(\mu, \theta)$ with known mean μ and unknown variance θ, what prior will you choose on the variance so that it is "non-informative" ? How is a non-informative prior defined/constructed ? Derive this prior mathematically.

---

derive fisher information, in its simplest analytical form (without involving any integrals) (5 points)

state form of Jeffreys prior using fisher information (3 points)

• (7 points) For a Gaussian PDF $\mathcal{N}(\mu, \sigma^2)$ with known mean μ and unknown standard deviation σ, what prior will you choose on the standard deviation so that it is "non-informative" ? Derive this prior mathematically, from its definition as a non-informative prior.

---

derive fisher information, in its simplest analytical form (5 points)

state form of Jeffreys prior using fisher information (2 points)

---

4. [15 points]

   • (4 points) For an exponential distribution, mathematically derive the maximum-likelihood estimator (MLE) for its parameter λ (as per formula on first page). Mathematically prove if this MLE is biased or unbiased.

   MLE derivation: (2 points)

   biased, using Jensen's inequality `https://karwailim.github.io/pubs/Lim_MLE_2017.pdf` (2 points)

   • (3 points) For a Poisson distribution, derive the maximum-likelihood estimator (MLE) for its variance.

   ---

   sample mean. see the lecture slides.

   • (8 points) Is this MLE an efficient estimator ? Mathematically prove from scratch, i.e., without assuming any properties of the MLE for the Poisson.

   ---

   show that MLE is unbiased (2 points)

   find variance of MLE (2 points)

   find Fisher information w.r.t. parameter (2 points)

   apply CRLB to show that the MLE is efficient (2 points)

   `https://karwailim.github.io/pubs/Lim_MLE_2017.pdf`

   ---

5. [15 points]

   Consider the i.i.d. data sample $\{x_n \in \mathbb{R}^D\}_{n=1}^N$ where dimension D is large.

   • (10 points) Your goal is to be able to reduce the dimensionality of representation of each datum x_n from D dimensions to E (fixed; given; $E \ll D$) dimensions (yielding the dataset $\{y_n \in \mathbb{R}^E\}_{n=1}^N$) using principal component analysis. Give the pseudocode for such an algorithm that can be implemented using standard operations we covered in class.

   ---

   please see lecture slides/notes and the assignment.

   form covariance matrix (2 points)

   perform eigen decomposition and get eigenvalues and eigenvectors. sort. (3 points)

   perform projection onto E dimensions of eigenbasis (5 points)

   • (5 points) Give an algorithm for regenerating/reconstructing the image using those E coordinates (and the knowledge of the designed E-dimensional basis in the D-dimensional space).

   ---

   please see lecture slides/notes and the assignment.

   append zeros (2 points)

   re-rotate and re-translate (2+1 points)

   ---

6. [10 points]

   Consider a numerical library with functions available for singlar-value decomposition (SVD), i.e., $svd()$, and for eigen-value decomposition (EVD), i.e., $eig()$.

   • (5 points) Can you compute a SVD of a real-valued matrix A by only using calls to $eig()$ (and without using any other decomposition), along with simple deterministic operations ? If so, give a clear algorithm along with a clear mathematical justification. If not, mathematically justify why not ?

   `https://web.mit.edu/be.400/www/SVD/Singular_Value_Decomposition.htm`

   stating that $A = USV^\top$ with properties of all matrices (1 point)

   analyzing $A \top A$ (2 points)

   analyzing $AA^\top$ (2 points)

   • (5 points) Can you compute an EVD of a real-valued symmetric positive-definite matrix B (assuming it has all distinct eigenvalues) by only using calls to $svd()$ (and without using any other decomposition), along with simple deterministic operations ? If so, give a clear algorithm with some mathematical justification. If not, mathematically justify why not ?

   B has an eigen-decomposition of the form $Q\Lambda Q^\top$ where all diagonal values on Λ are non-negative (2 points)

   this form also satisfies the criterion for SVD of B (2 points)

   use uniqueness properties of the SVD, given all distinct eigenvalues (1 point)

   `http://hua-zhou.github.io/teaching/biostatm280-2017spring/slides/16-eigsvd/eigsvd.html`

   ---

7. [10 points]

   • (6 points) For a paired i.i.d. data sample $\{(x_n \in \mathbb{R}, y_n \in \mathbb{R})\}_{n=1}^{N}$, derive the maximum-likelihood estimates for the parameters (i.e., for slope and intercept) of a linear-regression model where Y is the dependent variable. Assume that the errors, in measuring Y, are zero-mean i.i.d. Gaussian random variables.

   ---

   see lecture slides

   state the model (1 point): slope β, intercept α

   state the likelihood function (2 points)

   solve for α (1 point)

   solve for β (2 points)

   • (2 points) What is the Euclidean distance between the fitted line and the centroid of the dataset ?

   ---

   see lecture slides

   • (2 points) Evaluate the sum of the residuals between the fitted line and each y_n ?

   ---

   see lecture slides
