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:

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

• 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:

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

• $d(Ax) = Adx$
• $d(x^\top Ax) = x^\top(A+A^\top)dx$
• Fisher information:

$$I(\theta_{\text{true}}) := E_{P(X|\theta_{\text{true}})}\left[\left(\frac{\partial}{\partial\theta}\log P(X|\theta)\Big|_{\theta_{\text{true}}}\right)^2\right]$$

• 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)

   Consider two real-valued continuous random variables X and Y, with a joint PDF $P(X, Y)$.

   Suppose you repeat the following two-step experiment infinitely many times:

   (i) first, generate an independent sample point a as a random draw from $P(Y)$, and then

   (ii) given a, generate a sample point b as a random draw from $P(X|Y = a)$.

   Mathematically derive the PDF underlying the set of sample points b obtained across infinite runs of the experiment.

   Answer: $P(X)$

   Let B the random variable underlying the set of observations b.

   Then, $P(B \leq c) = \int_{x=-\infty}^{c} P_{X|Y}(x|y) \int_{y=infty}^{\infty} P_Y(y)dydx$

   $= \int_{x=-\infty}^{c} \int_{y=infty}^{\infty} P_{XY}(x, y)dydx$

   $= \int_{x=-\infty}^{c} \int_{y=infty}^{\infty} P_{XY}(x, y)dydx$

   $= \int_{x=-\infty}^{c} P_X(x)dx$

   ---

2. (15 points)

Consider a dataset $\{x_n\}_{n=1}^N$ where each x_n is drawn independently from some (unknown) PDF.

- [5 points] Derive the bias of the sample-mean estimator $\overline{x}$ for the mean of the PDF.

---

sample mean is an unbiased estimator of the mean.

- [7 points] Derive the bias of the sample-variance estimator for the variance of the PDF, where the estimator is defined as $\sum_{n=1}^N (1/N)(x_n - \overline{x})^2$.

`https://en.wikipedia.org/wiki/Variance#Biased_sample_variance`

- [3 points] Derive the bias of the sample-variance estimator for the variance of the PDF, where the estimator is defined as $\sum_{n=1}^N (1/(N-1))(x_n - \overline{x})^2$.

Unbiased. Follows from proof in previous question. `https://en.wikipedia.org/wiki/Variance#Unbiased_sample_variance`

---

3. (15 points)

Consider a dataset $\{x_n \in \mathbb{R}\}_{n=1}^{N}$ where integer $N > 99$, N is even, each datum x_n is unique (i.e., all data values are distinct), and each x_n is drawn independently from a Laplace PDF $P(x; \mu, b)$.

- [5 points] Formulate an optimization problem for estimating μ and b from the dataset.

ML estimation. Maximize log likelihood.

- [10 points: 7 + 3] Mathematically derive the optimal estimates for both the parameters.

---

1) [3 points]

log-likelihood function is differentiable in b. estimate is mean absolute deviation around the estimate for μ.

2) [7 points]

log-likelihood function has a derivative w.r.t. μ, which is $\sum_n \text{sgn}(x_n - \mu)$ that is well defined at all x except at the data points x_n.

Assume data are sorted.

In the sorted sequence, assume x_i and x_j are the points numbered $(N/2)$ and $(N/2 + 1)$.

Then, any μ within (x_i, x_j) makes the derivative zero. Also, log-likelihood function value remains constant within $[x_i, x_j]$. Thus, any $\mu \in [x_i, x_j]$ can be an estimate.

---

4. (30 points)

Consider a Gaussian PDF $P(x;\mu,C)$ on a 3D Euclidean domain where

(i) the mean vector μ is of unit norm and

(ii) the covariance matrix $C = \mathbf{I}/\alpha$ where $\mathbf{I}$ is the 3×3 identity matrix and $\alpha > 0$.

• [5 points] Consider a conditional PDF $Q(x;\mu,\alpha) := P(X|\|x\|_2 = 1)$ that restricts the Gaussian PDF $P(x;\mu,C)$ to the sphere $\{x : \|x\|_2 = 1\}$. Thus, the PDF $Q(x)$ is a function on a domain that is the sphere where $\|x\|_2 = 1$. Mathematically derive the functional form of the PDF $Q(x)$ (you don't need to find the exact functional form of the normalizing constant).

$Q(x;\mu,\alpha) \propto \exp(\alpha\mu^\top x)$

• [3 points] This family of PDFs will, of course, be represented in terms of μ and α. What will the locations x of (all possible) modes of the PDF $Q(x;\mu,\alpha)$ be ? Derive mathematically.

Maximum at $x = \mu$

• [3 points] What will the location x, on the unit sphere, where the PDF $Q(x;\mu,\alpha)$ takes its minimum value ? Derive mathematically.

Minimum at $x = -\mu$

• [2 points] For the family of such PDFs $Q(x;\cdot)$, what parameter models/captures the spread of the probability mass around the mode ? Give a convincing mathematical argument ("proof" not required).

μ decides the mode. changing μ to $\mu' = R\mu$ (where R is orthogonal) implies rotating the coordinate system in terms of R.

α parameter models spread.

$\alpha = 0$ leads to a uniform distribution on the sphere.

Increasing α makes the distribution more concentrated around the mode at μ.

• [2 points] Give a mathematical argument ("proof" not required) for the normalizing constant of the PDF to be independent of μ.

---

the PDF is rotationally symmetric around the mode.

by a simple rotation, the PDF can be aligned in some canonical form, e.g., where the mode is at $[1,0,0]$.

the normalizing constant is simply the integral of the un-normalized function over the sphere. this integral remains the same irrespective of any rotation-based reparametrization

---

• [5 points] Formulate an optimization problem for estimating μ (of unit norm) given a data sample of size N.

---

let sample be $\{x_n\}_{n=1}^N$.

ML estimation. Seek for the maximum of the likehood function defined over points μ on the sphere (i.e., with the constraint that $\|\mu\|_2 = 1$.)

• [4 points] Propose a conjugate prior $R(\mu;\theta)$, parameterized by a parameter (or parameter set) θ, for the parameter μ of the PDF $Q(x;\mu,\alpha)$ on the sphere. Mathematically justify that the proposed prior is a conjugate prior.

$R(\mu;a,b) \propto \exp(ba^\top \mu)$, where b is scalar non-negative.

• [6 points: 4 + 2] For the proposed conjugate prior $R(\mu;\theta)$, using Bayesian inference and given a sample of size N, mathematically derive the functional form of the posterior PDF (you don't need to find the exact functional form of the normalizing constant). Derive the mode of the posterior PDF.

---

posterior PDF $\propto \exp((ba + \alpha \sum_n x_n)^\top \mu)$
posterior mode at $(ba + \alpha \sum_n x_n)/\|ba + \alpha \sum_n x_n\|_2$

---

5. (25 points)

Consider a Poisson PMF with parameter λ.

• [5 points] Derive the Jeffreys non-informative prior for λ.

`https://en.wikipedia.org/wiki/Jeffreys_prior#Poisson_distribution_with_rate_parameter`

• [3 points] Given a sample (set) S of size $N > 1$ from the Poisson PMF, derive the formula for the maximum-likelihood estimator, say, $T(S)$, of λ.

---

sample mean

• [7 points: 6 + 1] Mathematically derive the distribution of the estimator $T(S)$. Is this distribution Poisson or not ?

---

sample sum is distributed as Poisson($N\lambda$), as derived in class.

derive sample-mean distribution using transformation of random variables.

NOT Poisson (S isn't integer valued).

• [10 points] Is the estimator $T(S)$ an efficient (i.e., minimum-variance unbiased) estimator ? Give a mathematical proof.

[2 points] show estimator is unbiased

[3 points] find variance of estimator: λ / N

[3 points] find Fisher information: N / λ `https://en.wikipedia.org/wiki/Poisson_distribution`

[2 points] apply Cramer-Rao lower bound to check for the inequality/equality. this estimator is efficient.
