

CS 215 : Data Analysis and Interpretation : Assignment 3

Instructor : Suyash P. Awate

Due Date : 21 Oct 2018, Sun, 11:55 pm; Maximum Points 40

Submission Instructions:

- Submit your solution, i.e., code, resulting graphs, and the report (in Adobe PDF format), for each question on moodle.
- Submit a single zip file that contains the solution to each problem below in a separate folder.
- To get partial credit for the code, ensure that the code is very well documented.
- To get partial credit for the derivations, include all derivation steps in their full details.
- 5 points for submission in the proper format.

1. (10 points) Use the Matlab function `randn()` to generate a data sample of N points drawn from a Gaussian distribution with mean $\mu_{\text{true}} = 10$ and standard deviation $\sigma_{\text{true}} = 4$. Consider the problem of using the data to get an estimate $\hat{\mu}$ of this Gaussian mean, assuming it is unknown, when the standard deviation σ_{true} is known.

Consider using one of the two prior distributions on the mean: (i) a Gaussian prior with mean $\mu_{\text{prior}} = 10.5$ and standard deviation $\sigma_{\text{prior}} = 1$ and (ii) a uniform prior over $(9.5, 11.5)$.

Consider various sample sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$. For each sample size N , repeat the following experiment $M \geq 100$ times: generate the data, get the maximum likelihood estimate $\hat{\mu}^{\text{ML}}$, get the maximum-a-posteriori estimates $\hat{\mu}^{\text{MAP}1}$ and $\hat{\mu}^{\text{MAP}2}$, and measure the relative errors $|\hat{\mu} - \mu_{\text{true}}|/\mu_{\text{true}}$ for all three estimates.

- Plot a single graph that shows the relative errors for each value of N as a box plot (use the Matlab `boxplot()` function), for each of the three estimates.
 - Interpret what you see in the graph. (i) What happens to the error as N increases ? (ii) Which of the three estimates will you prefer and why ?
2. (10 points) Use the Matlab function `rand()` to generate a data sample of N points from the uniform distribution on $(0, 1)$. Transform the resulting data x to generate a transformed data sample where each datum $y := (-1/\lambda) \log(x)$ with $\lambda = 5$. The transformed data y will have some distribution with parameter λ ; what is its analytical form ? Use a Gamma prior on the parameter λ , where the Gamma distribution has parameters $\alpha = 5.5$ and $\beta = 1$.
- Consider various sample sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$. For each sample size N , repeat the following experiment $M \geq 100$ times: generate the data, get the maximum likelihood estimate $\hat{\lambda}^{\text{ML}}$, get the Bayesian estimate as the posterior mean $\hat{\lambda}^{\text{PosteriorMean}}$, and measure the relative errors $|\hat{\lambda} - \lambda_{\text{true}}|/\lambda_{\text{true}}$ for both the estimates.
- Derive a formula for the posterior mean.

- Plot a single graph that shows the relative errors for each value of N as a box plot (use the Matlab `boxplot()` function), for both the estimates.
 - Interpret what you see in the graph. (i) What happens to the error as N increases ? (ii) Which of the two estimates will you prefer and why ?
3. (5 points) Consider a 2-dimensional data sample (assuming an extremely large sample size N) such that the data are drawn from a uniform distribution on a circle (i.e., a ring; the boundary of a disc, but *not* its interior) with center as the origin and radius r . Suppose you decide to fit a multivariate (2-dimensional) Gaussian distribution to this data by maximizing the likelihood function. The multivariate Gaussian is $P(x; \mu, C) := 1/\sqrt{2\pi|C|} \exp(-(x - \mu)^\top C^{-1}(x - \mu))$, where x and μ are vectors of size 2×1 , C is a matrix of size 2×2 , and $|C|$ is the determinant of C .
- Derive the mathematical formula, in terms of r , for the estimated mean and the estimated covariance matrix. You may use <http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html> to compute derivatives of the likelihood function with respect to the vector parameter μ and matrix parameter C .
 - Where is the mode of this Gaussian situated within \mathbb{R}^2 ? Do you think this Gaussian fits the data well ? Is it a good model ? Why or why not ?
 - Generate a large sample that is uniformly distributed on a circle with center origin and radius r . Compute the maximum likelihood estimates for the mean and covariance and report them, along with the sample size. Do they match the theoretically predicted values ?
4. (10 points) Suppose random variable X has a uniform distribution over $(0, \theta)$, where the parameter θ is unknown. Consider a Pareto distribution prior on θ , with a scale parameter $\theta_m > 0$ and a shape parameter $\alpha > 1$, as $P(\theta) \propto (\theta_m/\theta)^\alpha$ for $\theta \geq \theta_m$ and $P(\theta) = 0$ otherwise.
- Find the maximum-likelihood estimate $\hat{\theta}^{\text{ML}}$ and the maximum-a-posteriori estimate $\hat{\theta}^{\text{MAP}}$.
 - Does $\hat{\theta}^{\text{MAP}}$ tend to $\hat{\theta}^{\text{ML}}$ as the sample size tends to infinity ? Is this desirable or not ?
 - Find an estimator of the mean of the posterior distribution $\hat{\theta}^{\text{PosteriorMean}}$.
 - Does $\hat{\theta}^{\text{PosteriorMean}}$ tend to $\hat{\theta}^{\text{ML}}$ as the sample size tends to infinity ? Is this desirable or not ?