Homework 2

Due Octorber 1 at 11 pm

Unless stated otherwise, justify any answers you give. You can work in groups, but each student must write their own solution based on their own understanding of the problem.

When uploading your homework to Gradescope you will have to select the relevant pages for each question. Please submit each problem on a separate page (i.e., 1a and 1b can be on the same page but 1 and 2 must be on different pages). We understand that this may be cumbersome but this is the best way for the grading team to grade your homework assignments and provide feedback in a timely manner. Failure to adhere to these guidelines may result in a loss of points. Note that it may take some time to select the pages for your submission. Please plan accordingly. We suggest uploading your assignment at least 30 minutes before the deadline so you will have ample time to select the correct pages for your submission. If you are using LaTeX, consider using the minted or listings packages for typesetting code.

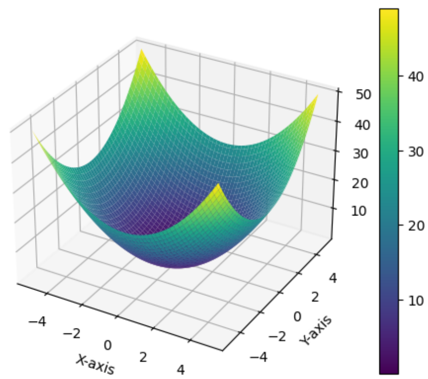
1. (Geometric random variable) Let $\tilde{a}$ be a geometric random variable with parameter α. What is the probability that $\tilde{a}$ equals a for $a = 1, 2, 3, \dots$ if we condition on the event $\tilde{a} > 5$? Justify your answer mathematically, and also explain why it makes sense intuitively (for example by referring to the coin flip example that we used to derive the geometric pmf).

   For $a = 1, 2, 3, 4, 5$, we have $P(\tilde{a} = a | \tilde{a} > 5) = \frac{P(\tilde{a}=a \cap \tilde{a}>5)}{P(\tilde{a}>5)} = 0$
   For $a > 5$, we will have $P(\tilde{a} = a | \tilde{a} > 5) = \frac{P(\tilde{a}=a \cap \tilde{a}>5)}{P(\tilde{a}>5)} = \frac{P(\tilde{a}=a)}{1-P(\tilde{a}\leq 5)}$ where we compute the numerator and denominator.
   $P(\tilde{a} = a) = (1-\alpha)^{a-1} \times \alpha$ and $P(\tilde{a} \leq 5) = \sum_{i=1}^{5} P(\tilde{a} = i) = (1-\alpha)^{i-1} \times \alpha$. Thus we have the result as $P(\tilde{a} = a | \tilde{a} > 5) = \frac{(1-\alpha)^{a-1}\times\alpha}{1-\{\alpha+(1-\alpha)\times\alpha+(1-\alpha)^2\times\alpha+(1-\alpha)^3\times\alpha+(1-\alpha)^4\times\alpha+(1-\alpha)^5\times\alpha\}}$.

2. (Chess games) Garry and Anish decide to play 10 chess games. Garry wins 4, they draw 4, and Anish wins 2. We decide to model the games probabilistically, assuming that they are independent and in each game Garry has a probability θ of winning and Anish has a probability α of winning.

   (a) Plot the log-likelihood function of the parametric model.

   We first compute the log-likelihood function:
   $\mathcal{L}(\theta) = log\{\theta^4(1-\theta-\alpha)^4\alpha^2\} = 4log\theta + 4log(1-\theta-\alpha) + 2log\alpha$



   We see that it is a convex function and that the global minimum point exists.

   (b) What is the maximum likelihood estimate of θ and α?

   We compute the first order condition for the function:
   $\frac{\partial}{\partial\theta}\mathcal{L}(\theta) = \frac{4}{\theta} - \frac{4}{1-\theta-\alpha} = 0$
   $\frac{\partial}{\partial\alpha}\mathcal{L}(\alpha) = -\frac{4}{1-\theta-\alpha} + \frac{2}{\alpha} = 0$
   Finally, we will have $\begin{cases} \hat{\theta} = \frac{2}{5} = 0.4 \\ \hat{\alpha} = \frac{1}{5} = 0.2 \end{cases}$

   (c) Model the data as realizations from a discrete random variable and compute its empirical pmf. Compare this nonparametric model to the parametric model from the previous questions.

   If we use empirical approach to model our probability, we use uniform sample space where $\hat{\theta} = \frac{4}{10} = 0.4$ and $\hat{\alpha} = \frac{2}{10} = 0.2$, since Garry wins 4 out of 10 and Anish wins 2 out of 10. We find that the results from parametric model and non-parametric model coincide.

3. (Darts) In a game of darts, a player needs to hit a certain number k times. Assume that all attempts are mutually independent, and that the probability of success in each attempt is θ. Derive the pmf of a random variable representing the number of required attempts.

This can be modeled as a bernoulli sequence. Here we let the random variable X denote the number of attempts until the k-th hit shows up, which can be translated into: there are k hits that happened during the first $X-1$ attempts. Thus we would have $P(X = x) = \binom{x-1}{k-1}\theta^{k-1}(1-\theta)^{n-k} \times \theta = \binom{x-1}{k-1}\theta^{k}(1-\theta)^{x-k}$ where $k = 0, 1, 2, \cdots$.

4. (Air Traffic) The tables in *train.csv* and *test.csv* record the numbers of arriving flights at London Heathrow airport every 10 minutes between 18:00 and 19:30 every day. The training and test data are collected from 2010-06-01 to 2010-12-31 and 2011-01-01 to 2011-06-28, respectively.

   (a) Use the training set to estimate the pmf using a parametric model and a nonparametric model. Explain any assumptions you make to choose the parametric model. Plot the nonparametric and parametric pmfs.

   We assume that the data points are independent and follow possion distribution where the pmf for it is $P_\lambda(x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$, and the log-likelihood function is $\mathcal{L}(\lambda) = -\sum_{i=1}^{n} log(x_i!) + n\overline{x_n} log\lambda - n\lambda$ we then take the derivative for λ and could get $\frac{n\overline{x_n}}{\lambda} - n = 0$ and that $\hat{\lambda} = \overline{x_n}$.

   The non-parametric model will say that the probability that a datapoint appear is the fraction of the number of times it shows up in the dataset and the total number of rows.

   The plots is shown below:

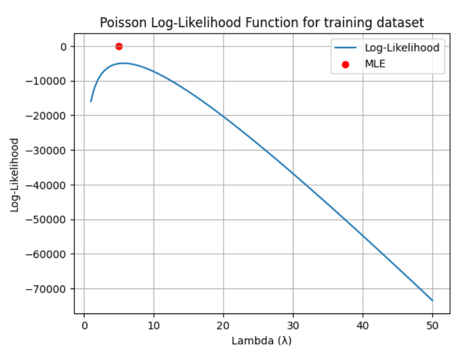


Figure 1: Parametric Pmf

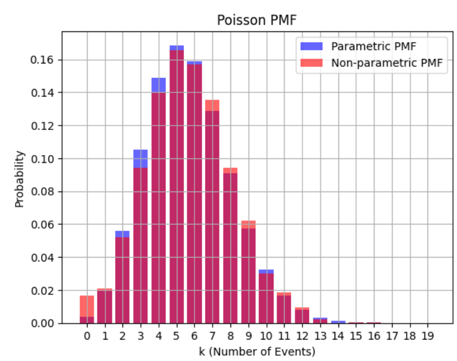


Figure 2: Parametric/Non-parametric PMF

   (b) Evaluate the performance of your two models on the test set. Compute the root mean square error between the estimated pmfs and the empirical pmf of the test set. Which model performs better?

   We calculate by code that the RMSE for paramtric model is 0.000444228 and the RMSE for non-parametric model is 0.000271174. So we see that the empirical pmf model performs better. This is reasonable since we use the empirical model on the test dataset, the results from the two non-parametric empirical model should be similar.