CS5340: Uncertainty Modeling in AI

Tutorial 2

Released: Jan. 19, 2023

Problem 1. (Uncorrelated Random Variables)

Consider two random variables X and Y, where $\mathrm{Cov}[X, Y] = 0$. Recall that

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Problem 1.a. Your friend Donald says, "Since the covariance of X and Y is zero, then it must be case that that X and Y are independent!" Is Donald correct? If yes, provide a proof. If not, give a counterexample. *Hint:* Can you find some function f such that $Y = f(X)$ and $\mathrm{Cov}[X, Y] = 0$?

Problem 1.b. Consider we obtain samples $x_1, x_2, \ldots, x_N$ drawn iid from $p(X)$ and another batch of samples $y_1, y_2, \ldots, y_N$ drawn iid from $p(Y)$. Suppose we want to model the joint distribution $p_\theta(X, Y)$. Donald suggests we perform maximum likelihood estimation by finding

$$\arg\max_\theta \sum_i^N \log p_\theta(x_i, y_i).$$

Is Donald correct? Justify your answer.

Problem 2. (Exponential Family)

In Lecture 2, we learned about the Exponential Family (ExpFam). In this tutorial, we'll make the general concept more concrete with some examples. Recall that ExpFam distributions have the following (natural) form:

$$p_\eta(x) = \frac{h(x)\exp[\eta^\top s(x)]}{Z(\eta)} \quad \text{or} \quad p_\theta(x) = h(x)\exp[\eta^\top s(x) - A(\eta)]$$

where $\eta = \eta(\theta)$ are the natural parameters, $s(x)$ are the sufficient statistics, $h(x)$ is the base measure, and $Z(\eta)$ is the partition function. In an alternative form (on the right above), $A(\eta)$ is the log partition function or cumulant function.

Problem 2.a. We'll begin with the **exponential distribution**[1]. Consider a process where events occur continuously and independently at some average rate λ ($\lambda > 0$). Real-world examples include radioactive decay, customer arrival times, and machine failure times. Let x be the time between events. The exponential distribution models the probability distribution of x:

$$p(x|\lambda) = \begin{cases} \lambda\exp(-\lambda x) & x \geq 0, \\ 0 & x < 0, \end{cases} \tag{1}$$

where $\lambda > 0$. Show that the Exponential distribution is ExpFam.
Hint: By rearranging elements, try to rewrite the exponential distribution in terms of the natural parameters, sufficient statistics, base measure, and (log) partition function. To get you started, the base measure is $h(x) = 1$ for $x \geq 0$, which we will assume for this example.

[1] Note that the exponential distribution is *not* the same as the exponential family. The exponential distribution is a specific distribution, but the exponential family is a more general form.

Problem 2.b. Assume that you have access to data $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$. Use facts about ExpFam to help you derive the MLE of the natural parameters of the exponential distribution. Recall that:

$$\mathbb{E}[s(x)] = \nabla \log Z(\eta) = \nabla A(\eta) \tag{2}$$

and that the maximum likelihood estimator η_{MLE} satisfies the condition that

$$\nabla A(\eta_{\text{MLE}}) = \frac{1}{N} \sum_{i=1}^{N} s(x_i) \tag{3}$$

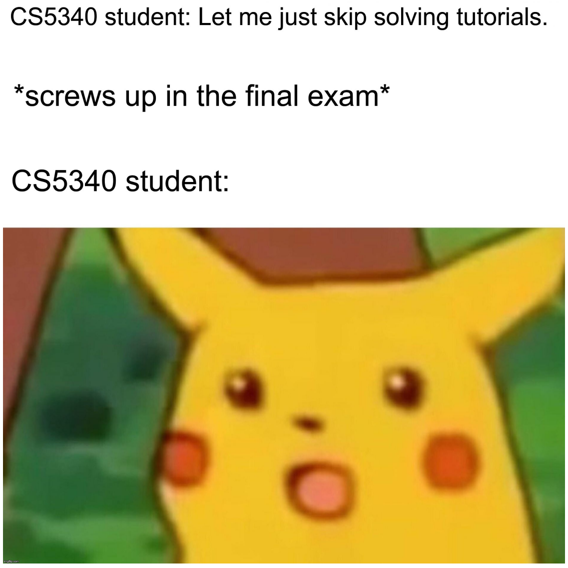
Hint: The MLE estimator for the exponential satisfies $\mathbb{E}_{\lambda_{\text{MLE}}}[s(x)] = \frac{1}{N} \sum_{i=1}^{N} s(x_i)$.

Problem 2.c. Repeat the two problems above for the (univariate) **Gaussian** distribution, i.e.,

1. show that the Gaussian is exponential family, and
2. derive the MLE of its natural parameters by leveraging the properties of ExpFam distributions.

(*Challenge*) If you are feeling like a challenge, show that the *multivariate Gaussian* is also exponential family, and derive the MLE of its natural parameters.

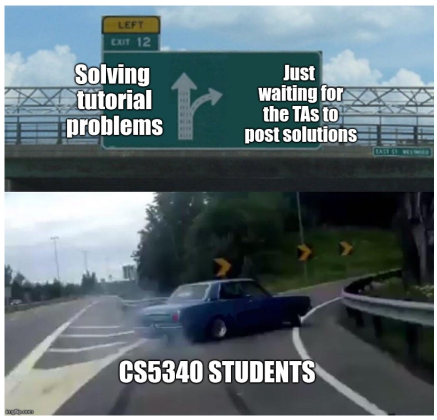
Problem 3. (Meme of the Year)
A poll was conducted amongst CS5340 students to pick the best meme template in 2018. The four meme templates that were in the run for *meme of the year* 2018 are shown in Fig. 1. The votes received by each meme template are tabulated in Table 1. Denote the vote of i-th student by a one-hot vector $\mathbf{x}_i$ and the entire dataset by $\mathcal{X} = \{\mathbf{x}\}_{i=1}^{N}$. For example, someone who voted for "Surprised Pikachu" will have $\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\top$.



(a) Surprised Pikachu

(b) Two Buttons Dilemma

(c) Distracted Boyfriend



(d) Left Exit 12

Figure 1: The four meme templates in the run for meme of the year 2018

ID	Template Name	# Votes
1	Surprised Pikachu	25
2	Two Buttons Dilemma	12
3	Distracted Boyfriend	30
4	Left Exit 12	10

Table 1: Votes received by each template by CS5340 students

Problem 3.a. Fit an appropriate distribution to the data given above and compute the parameters of the distribution using maximum likelihood estimation.

Problem 3.b. Imagine that before looking at the poll results from the class you had access to poll results from another such poll that was conducted on Reddit, the results of which are shown in Table 2.

ID	Template Name	# Votes
1	Surprised Pikachu	250
2	Two Buttons Dilemma	110
3	Distracted Boyfriend	280
4	Left Exit 12	140

Table 2: Votes received by each template on Reddit

- Choose an appropriate distribution to incorporate this prior knowledge into your model. Derive an expression for the posterior distribution.
- Derive an expression for posterior predictive distribution $p(\mathbf{x}^*|\mathcal{X})$. Specifically, compute the probability $p(\mathbf{x}^* = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^\top |\mathcal{X})$.

Problem 3.c. Choose an appropriate prior distribution for the likelihood distribution chosen in 2.a. and estimate the parameters using maximum a posteriori (MAP) estimation.

Problem 3.d. (*Challenge*) Show that the categorical distribution is exponential family.