# Problem 1a

Thursday, 26 January 2023 4:02 PM

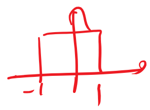
**Problem 1.** (Uncorrelated Random Variables)

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

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

**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 $\text{Cov}[X, Y] = 0$?

$X \sim \text{Uniform}(-1, 1)$



$Y = X^2$

$$\text{Cov}[XY] = \mathbb{E}[X\overset{X^2}{Y}] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$= \underbrace{\mathbb{E}[X^3]}_{0} - \underbrace{\underbrace{\mathbb{E}[X]}_{0}\mathbb{E}[X^2]}_{0}$$

$$= 0$$

$$\mathbb{E}[X] = \int_{-1}^{1} x p(x)\, dx = \int_{-1}^{1} \frac{x}{2}\, dx = 0$$

$$\mathbb{E}[X^3] = \int_{-1}^{1} x^3 p(x)\, dx = \int_{-1}^{1} \frac{x^3}{2}\, dx = 0$$

# Problem 1b

Thursday, 26 January 2023 4:03 PM

**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

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

Is Donald correct? Justify your answer.

$$\lg \prod_i p(x_i, y_i)$$

$$\Downarrow$$

$$\sum_i \lg p(x_i, y_i)$$

$$x_1, x_2, \ldots, x_N \sim p(x)$$

$$y_1, y_2, \ldots, y_N \sim p(y)$$

$$(x_i, y_i) \sim p(x, y)$$

$$p_\theta(x, y) = p(x, y \mid \theta)$$

# Problem 2a

Thursday, 26 January 2023 4:04 PM

**Problem 2.a.** We'll begin with the **exponential distribution**[1]. Consider a process where events occur continuously and independently at some average rate $\lambda$ ($\lambda > 0$). 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 $x$:

$$\boxed{p(x|\lambda) = \begin{cases} \lambda \exp(-\lambda x) & x \geq 0, \\ 0 & x < 0 \end{cases}} \tag{7}$$

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.

$$p_\eta(x) = \frac{h(x) \exp[\eta^T s(x)]}{Z(\eta)}$$

$$p(x|\lambda) = \lambda \exp(-\lambda x)$$

$h(x) = 1$ for $x \geq 0$

$= 0$ otherwise.

$s(x) = -x$

$\eta = \lambda$

$Z(\eta) = \frac{1}{\lambda}$ $A(\eta) = \log \frac{1}{\lambda}$

# Problem 2b

Thursday, 26 January 2023 4:04 PM

**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:

$$\mathbb{E}[s(x)] = \nabla \log Z(\eta) = \nabla A(\eta) \tag{8}$$

and that the maximum likelihood estimator $\eta_{\text{MLE}}$ satisfies the condition that

$$\boxed{\nabla A(\eta_{\text{MLE}}) = \frac{1}{N}\sum_{i=1}^{N} s(x_i)} \tag{9}$$

$\nabla A(\eta)$, $\eta \neq \eta_{MLE}$

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

$$\mathbb{E}_{\lambda_{MLE}}[s(x)] = \frac{1}{N}\sum_{i=1}^{N} s(x_i)$$

$$A(\eta) = \lg \frac{1}{\eta}$$

$$\frac{1}{N}\sum_{i=1}^{N} s(x_i) = \frac{\partial}{\partial \eta} A(\eta_{MLE})$$

$$= \eta\left(-\frac{1}{\eta^2}\right)$$

$$= -\frac{1}{\eta_{MLE}}$$

$$\frac{\partial}{\partial \eta} A(\eta_{MLE}) = -\frac{1}{\eta_{MLE}} = \frac{1}{N}\sum_{i=1}^{N} s(x_i)$$

$$\boxed{\eta_{MLE} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N} x_i} = \frac{1}{\bar{x}}}$$

# Problem 2c

Thursday, 26 January 2023 4:27 PM

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

**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.

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

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

$$\Rightarrow p_\theta(x) = h(x)\exp[\eta^T s(x) - A(\eta)]$$

$$\frac{1}{Z(\eta)}$$

$$\exp\left(-\log\sigma - \frac{x^2+\mu^2-2x\mu}{2\sigma^2}\right)$$

$$A(\eta) = \log Z(\eta)$$

$$\exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2} - \log\sigma\right)$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}}}_{h(x)} \exp\left(\underbrace{\begin{bmatrix}\mu/\sigma^2 \\ -\frac{1}{2\sigma^2}\end{bmatrix}}_{\eta}^T \underbrace{\begin{bmatrix} x \\ x^2\end{bmatrix}}_{s(x)} - \underbrace{\left(\frac{\mu^2}{2\sigma^2} + \log\sigma\right)}_{A(\eta)}\right) \qquad \eta = \begin{bmatrix}\eta_1 \\ \eta_2\end{bmatrix} = \begin{bmatrix}\mu/\sigma^2 \\ -\frac{1}{2\sigma^2}\end{bmatrix}$$

$$A(\eta) = \frac{\mu^2}{2\sigma^2} + \log\sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

---

$$\nabla A(\eta_{MLE}) = \frac{1}{N}\sum_{i=1}^{N} s(x_i) \qquad \frac{\partial A}{\partial \eta_1} = \frac{\eta_1}{2\eta_2} = \mu$$

$$\begin{bmatrix} \frac{\partial A}{\partial \eta_1}(\eta_{MLE}) \\ \frac{\partial A}{\partial \eta_2}(\eta_{MLE}) \end{bmatrix} = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} x \\ x^2\end{bmatrix} \qquad \boxed{\mu_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i}$$

$$= \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x \\ \frac{1}{N}\sum_{i=1}^{N} x^2 \end{bmatrix} \qquad \frac{\partial A}{\partial \eta_2} = \underbrace{\frac{\eta_1^2}{4\eta_2^2}}_{\mu^2} - \frac{1}{2\eta_2} \quad \mu^2 + \sigma^2$$

$$\mu_{MLE}^2 + \sigma_{MLE}^2 = \frac{1}{N}\sum_{i=1}^{N} x^2$$

$$\sigma^2_{MLE} = \frac{1}{N}\sum_{i}^{N} x_i^2 - \mu_{MLE}^2$$

$$= \frac{1}{N}\sum_{i}^{N} x_i^2 - 2\mu^2_{MLE} + \mu^2_{MLE} \qquad \mu_{MLE} = \frac{1}{N}\sum_{i}^{N} x_i$$

$$\eta_{\ldots} = \begin{bmatrix}\eta_{1,MLE} \\ \ldots\end{bmatrix}$$

$$\eta_{MLE} = \begin{bmatrix} \eta_{1\,MLE} \\ \eta_{2\,MLE} \end{bmatrix}$$

$$= \frac{1}{N}\sum_{i}^{N} x_i^2 - 2\underbrace{\mu_{MLE}^2}_{\downarrow} + \mu_{MLE}^2 \qquad \mu_{MLE} = \frac{1}{N}\sum_{i} x_i$$

$$= \frac{1}{N}\sum_{i}^{N} x_i^2 - 2\overbrace{\mu_{MLE}\, x_i} + \mu_{MLE}^2$$

$$\boxed{\sigma_{MLE}^2 = \frac{1}{N}\sum_{i}^{N}(x_i - \mu_{MLE})^2}$$

# Problem 3a

Thursday, 26 January 2023 5:11 PM

**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 $\mathbf{x} = [1 \quad 0 \quad 0 \quad 0]^\top$.

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

$$\text{Cat}[\lambda] \qquad \lambda = [\underbrace{\lambda_1, \lambda_2, \lambda_3, \lambda_4}]$$

$$\lambda_{MLE} \qquad \sum_{j=1}^{4} \lambda_j = 1$$

$$\lambda_{MLE} = \underset{\lambda}{\text{argmax}} \log p(D \mid \lambda)$$

$$p(D \mid \lambda_1, \dots, \lambda_k) = \prod_{i=1}^{n} \prod_{j=1}^{k} \lambda_j^{[x_{ij}=1]} \qquad \text{s.t} \quad \sum_{j=1}^{k} \lambda_j = 1$$

$$= \prod_{j=1}^{k} \lambda_j^{N_j}$$

$$\mathcal{L} = \sum_{j=1}^{k} N_j \log \lambda_j + \nu \left( \sum_{j=1}^{k} \lambda_j - 1 \right)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = \frac{N_j}{\lambda_j} + \nu = 0$$

$$\Rightarrow \lambda_j = -\frac{N_j}{\nu}$$

$$\sum_j \lambda_j = 1$$

$$\sum_j -\frac{N_j}{\nu} = 1$$

$$\nu = -\sum_j N_j$$

$$\Rightarrow \boxed{\lambda_j = \frac{N_j}{\sum_{\ell=1}^{k} N_\ell}}$$

# Problem 3b

Thursday, 26 January 2023 5:24 PM

**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 3.

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

**Table 3**: 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 $p(\mathbf{x}^* = \begin{bmatrix}0 & 0 & 1 & 0\end{bmatrix}^\top |\mathcal{X})$.

posterior $Dir[\tilde{\alpha}]$

$$\tilde{\alpha} = [\alpha_1 + N_1, \alpha_2 + N_2, \ldots, \alpha_k + N_k]^T$$

$$p(x^* | D) = \int p(x^* | \lambda_1, \ldots, \lambda_k)\, p(\lambda_1, \ldots \lambda_k | D)\, d\lambda$$

$$\boxed{\Gamma(z+1) = z\Gamma(z)}$$

$$= \int \prod_j \lambda_j^{[x_j^* = 1]} \cdot \frac{1}{B(\tilde{\alpha})} \prod_{j=1}^{k} \lambda_j^{\tilde{\alpha}_j - 1}\, d\lambda$$

multiply and divide by $\boxed{\dfrac{\prod_j \Gamma(\tilde{\alpha}_j + [x_j^* = 1])}{\Gamma(\sum_j \tilde{\alpha}_j + [x_j^* = 1])}} = C$

$$= \underbrace{\frac{1}{B(\tilde{\alpha})} \cdot \frac{\prod_j \Gamma(\tilde{\alpha}_j + [x_j^* = 1])}{\Gamma(\sum_j \tilde{\alpha}_j + [x_j^* = 1])}} \underbrace{\int \frac{1}{C} \prod_j^{k} \lambda_j^{\tilde{\alpha}_j + [x_j^* = 1] - 1}\, d\lambda}_{=1 \text{ Drichlet PDF}}$$

$$= \frac{\Gamma(\sum_j \tilde{\alpha})}{\prod_j \Gamma(\tilde{\alpha}_j)} \cdot \underbrace{\frac{\prod_j \Gamma(\tilde{\alpha}_j + [x_j = 1])}{\Gamma(\sum_j \tilde{\alpha}_j + [x_j^* = 1])}}$$

$$= \frac{\cancel{\Gamma(\sum_j \tilde{\alpha})}}{\prod_j \Gamma(\tilde{\alpha}_j)} \cdot \frac{\prod_j \Gamma(\tilde{\alpha}_j + [x_j^* = 1])}{\sum_j \tilde{\alpha}_j \cdot \cancel{\Gamma(\sum_j \tilde{\alpha}_j)}}$$

assume that $x^*$ has 1 at the m-th element.

$$= \frac{1}{\prod_j \Gamma(\tilde{\alpha}_j)} \cdot \frac{\Gamma(\tilde{\alpha}_m + 1) \cdot \prod_{j \neq m} \Gamma(\tilde{\alpha}_j)}{\sum_j \tilde{\alpha}_j}$$

$\Gamma(z+1) = z\Gamma(z)$

$$= \frac{1}{\quad} \quad \frac{\tilde{\alpha}_m \boxed{\cancel{\Gamma(\tilde{\alpha}_m) \prod_{j \neq m} \Gamma(\tilde{\alpha}_j)}}}{\quad}$$

$$= \frac{1}{\cancel{\boxed{\prod_j \Gamma(\tilde{\alpha}_j)}}} \frac{\tilde{\alpha}_m \cancel{\boxed{\Gamma(\tilde{\alpha}_m) \prod_{j \neq m} \Gamma(\tilde{\alpha}_j)}}}{\sum_j \tilde{\alpha}_j}$$

$$= \frac{\tilde{\alpha}_m}{\sum_j \tilde{\alpha}_j} \qquad [\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_k]$$

# Problem 3c

Thursday, 26 January 2023 5:26 PM

**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.

$$\lambda_{MAP} = \underset{\lambda}{\mathrm{argmax}}\ p(\lambda | D)$$

$$\sum_j^L \lambda_j = 1$$

$$\lambda_j = \frac{\alpha_j + N_j - 1}{\sum_{l=1}^{L} \alpha_j + N_l - 1}$$

# Problem 2c

Thursday, 26 January 2023 4:27 PM

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

**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.