TEACHERS Alexandre Cremers and Jakub Dotlačil TA Bas Cornelissen

Practice problem set 5

This week's exercises deal with joint distributions, covariance and correlation. You only have to hand in the homework problems; these exercises are optional and for practicing only. If you have questions about them, please post them to the discussion forum and try to help each other. We will also keep an eye on that.

Problem 1: Properties of the Poisson distribution

The *Poisson distribution* models the number of times some event happens in a given period of time. It's probability mass function is given by

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad \lambda \in \mathbb{R}_{>0}, \quad k = 0, 1, 2, \dots$$

where λ is the distribution's only parameter. Let $X \sim \text{Poisson}(\lambda)$.

(a) Show that $E[X] = \lambda$.

(b) Show that $\text{Var}(X) = \lambda$.

Problem 2: Maximum a posteriori estimates

The *Beta distribution* is a continuous probability distribution over $[0, 1]$. You don't need to know anything about continuous distributions for this exercise. The important thing is that if you sample from a Beta distribution, you will get a real number θ in the interval $[0, 1]$, with probability

$$P(\Theta = \theta|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1},$$

Where $\alpha > 0$ and $\beta > 0$ are the two parameters of the distribution and $B(\alpha, \beta)$ is a normalizing constant that does not depend on θ.

Let $\Theta \sim \text{Beta}(\alpha, \beta)$. Since Θ takes values in $[0, 1]$ we use Θ as the parameter for a Bernoulli distribution. Let $X_1, \dots, X_N \sim \text{Bern}(\Theta)$ be i.i.d. random variables — we use the beta distribution as a *prior* for the Bernoulli.

(a) Check that the joint probability of i.i.d. Bernoulli variables $X_1, \dots X_N$ for a given θ is given by

$$P(X_1 = x_1, \dots, X_N = x_n \mid \theta) = \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i}$$

(b) Find the unnormalized posterior log-probability

$$\mathcal{P}(\theta) := \ln\big[P(X_1 = x_1, \dots, X_N = x_n \mid \Theta = \theta) \cdot P(\Theta = \theta|\alpha, \beta)\big]$$

(c) Find the derivative $\frac{\partial}{\partial\theta}\mathcal{P}(\theta)$.

(d) Show that the maximum a posteriori estimate for θ is

$$\theta_{\text{MAP}} = \text{argmax}_\theta P(\Theta = \theta|X_1, \dots, X_N) = \frac{K + \alpha - 1}{N + \alpha + \beta - 2},$$

where $K = \sum_{i=1}^{N} X_i$.

(e) For which values of the parameters α and β do you get the MLE estimate $\theta_{\text{ML}} = \text{argmax}_\theta P(X_1, \dots X_N \mid \theta)$ back?

(f) The Beta prior effectively adds some more observations of $X = 0$ and $X = 1$ to the data. How many of each?

Problem 3: Multiple binomials

Let $X_1, \dots, X_N$ be i.i.d. $\text{Binom}(n, \theta)$ random variables. Show that the maximum likelihood estimator for λ is:

$$\theta_{\text{MLE}} = \text{argmax}_\lambda P(X_1, \dots, X_N \mid \lambda) = \frac{\sum_{i=1}^{N} x_i}{n \cdot N}.$$

Note that this differs from the script since we now have N rather than 1 random variable. (Be careful not to confuse n and N).

(a) Find the log-likelihood $\ln P(X_1^N = x_1^N \mid n, \theta)$ of the data.

(b) Find the derivative $\frac{\partial}{\partial\theta} \ln P(X_1^N = x_1^N \mid n, \theta)$.

(c) Show that the maximum likelihood estimator for θ is (keeping n fixed):

$$\theta_{\text{MLE}} = \text{argmax}_\theta P(X_1^N = x_1^N \mid n, \theta) = \frac{\sum_{i=1}^N x_i}{n \cdot N}.$$

CREDITS Some questions are from MIT course 18-05 by Jeremy Orloff and Jonathan Bloom, see ocw.mit.edu.