[afterskip=.3]subsection

BASIC PROBABILITY: THEORY
Master of Logic, University of Amsterdam, 2016
TEACHERS Christian Schaffner and Philip Schulz TA Bas Cornelissen

Homework problem set 5

Your homework must be handed in **electronically via Moodle before Wednesday October 5th, 21:00h**. This deadline is strict and late submissions are graded with a 0. At the end of the course, the lowest of your 7 weekly homework grades will be dropped. You are strongly encouraged to work together on the exercises, including the homework. However, after this discussion phase, you have to write down and submit your own individual solution. Numbers alone are never sufficient, always motivate your answers.

Problem 1: Markov's inequality (2pt)

Let X be a random variable and $a > 0$ some real constant. Prove *Markov's Inequality*:

$$P(|X| \geq a) \leq \frac{E[|X|]}{a},$$

where $|X|$ is the absolute value of X.

Problem 2: Poisson distribution (3pt)

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

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

where λ is the distribution's only parameter. Let $X_1, \ldots, X_N$ be n independent, identically distributed (i.i.d.) random variables following a Poisson(λ) distribution.

(a) **1pt** Show that $T(X_1, \ldots, X_N) = \sum_{i=1}^{N} X_i$ is a sufficient statistic for the Poisson distribution. *Hint: use the Factorization Theorem (5.13).*

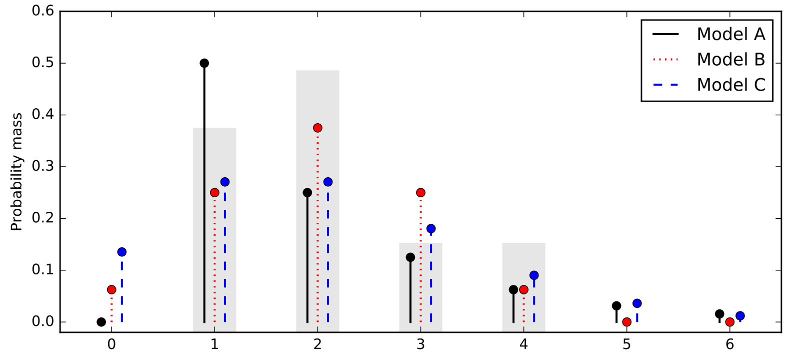

Figure 1: Distributions of the three models using the ML-parameters. The grey bars in the background show the empirical distribution of the observations, i.e. the relative frequency of all observations in $\mathcal{D}$.

(b) **0.5pt** Find the log-likelihood $\mathcal{L}_x(\lambda) = \ln P(X_1, \ldots, X_n \mid \lambda)$ of the data.

(c) **0.5pt** Find the derivative $\frac{\partial}{\partial \lambda}\mathcal{L}(\lambda)$ of the log-likelihood.

(d) **1pt** Show that the maximum likelihood estimate for λ is

$$\lambda_{\text{ML}} = \text{argmax}_\lambda P(x_1, \ldots, x_N \mid \lambda) = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Problem 3: Multiple binomials (2pt)

In accordance with our notation, I lowercased all RVs that are arguments to a pmf. Let $X_1, \ldots, X_N$ be i.i.d. Binom(n, θ) random variables. (Be careful not to confuse n and N).

(a) **0.5pt** Find the log-likelihood $\ln P(x_1, \ldots, x_N \mid n, \theta)$ of the data.

(b) **0.5pt** Find the derivative $\frac{\partial}{\partial \theta} \ln P(x_1, \ldots, x_N \mid n, \theta)$.

(c) **1pt** Show that the maximum likelihood estimator for θ is (keeping n

fixed):

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

Problem 4: Three models (3pt)

Once upon a time, three data enthusiasts, Alice, Bob and Charlie, were asked to look into a dataset that was as small as it was confidential. The client could not provide any information about the dataset, so just gave the data:

$$\mathcal{D} = \{1, 1, 2, 4, 2, 1, 3, 2, 2\}.$$

Puzzled, all three came up with different explanations of the data. Alice thought the data was drawn from a geometric distribution, Bob proposed a binomial distribution and Charlie a Poisson distribution (see problem 2). More precisely, they assume that the data is generated by $N = |\mathcal{D}|$ i.i.d. random variables, but everyone proposes different ones:

$$\text{Model Alice:} \quad A_i \sim \text{Geom}(\pi), \tag{1}$$

$$\text{Model Bob:} \quad B_i \sim \text{Binom}(n, \theta), \quad n = 4 \tag{2}$$

$$\text{Model Charlie:} \quad C_i \sim \text{Poisson}(\lambda), \tag{3}$$

for $i = 1, \ldots, N$. So Alice, Bob and Charlie suppose that the observation $x_1 \in \mathcal{D}$ is the value taken by the RV A_1, B_1 and C_1 respectively.

(a) **0.5pt** Bob has already decided on one of the parameters: $n = 4$. But all remaining parameters have to be estimated from the data. Calculate the ML-estimates for π, θ and λ from the data. The distributions of the resulting models are shown in Figure 1. *Hint: use homework problem 5.2, 5.3 and this week's board questions.*

(b) **0.5pt** Calculate the log-likelihood of the data $\mathcal{D}$ for each of the three models, using the ML-estimates from (a) as parameters. Which model gives the best explanation of the data, i.e. assigns the data the highest likelihood? *Hint: you might find Table 1 useful.*

(c) **1pt** Alice, Bob and Charlie wonder how their models differ. Show that all models have the same expectation when parametrised with their MLE parameters. Observe that this expectation is equal to the sample mean $\overline{x} := \frac{1}{N}\sum_{j=1}^N x_j$, where x_j is an observations. However, the models have different variances: show that $\text{Var}(B_i) < \text{Var}(C_i) \leq \text{Var}(A_i)$ if $\overline{x} \geq 2$. I think it is $Var(C_i) \leq Var(B_i)$ because the MLE parameters are 0.5 for both the binomial and geometric and 2 for the Poisson. Thus $\frac{1}{\sqrt{2}} < 4 \times 0.25$.

	$X \sim \text{Geom}(\pi_{\text{ML}})$	$X \sim \text{Binom}(4, \theta_{\text{ML}})$	$X \sim \text{Poisson}(\lambda_{\text{ML}})$
$\ln P(X=0)$	$-\infty$	-2.7726	-2.0000
$\ln P(X=1)$	-0.6931	-1.3863	-1.3069
$\ln P(X=2)$	-1.3863	-0.9808	-1.3069
$\ln P(X=3)$	-2.0794	-1.3863	-1.7123
$\ln P(X=4)$	-2.7726	-2.7726	-2.4055
$\ln P(X=5)$	-3.4657	$-\infty$	-3.3218

Table 1: Some log-probabilities for the three models

(d) **0.5pt** The differences between their models are of course much larger than their expectations suggest I do not understand this sentence. What do the expectations suggest concretely?. This becomes clear when the client sends some more observations:

$$\mathcal{D}' = \mathcal{D} \cup \{0, 4, 5\}$$

Which model(s) cannot account for this data? In other words, under which model(s) does this data get likelihood 0?

(e) **1pt** Let's call the parameter of the one remaining model γ_{ML}. It's inventor actually had a competing hypothesis for the parameter: $\gamma'_{\text{ML}} = 1.25 \times \gamma_{\text{ML}}$, which s/he deemed slightly less likely: $P(\gamma'_{\text{ML}}) = 0.4$ while $P(\gamma_{\text{ML}}) = 0.6$. Which of these two parameter should s/he prefer after observing the new data $\{0, 4, 5\}$? Answer that question by calculating the posterior $P(\gamma \mid \{0, 4, 5\})$. This exercise as formulated does not make sense. If you want to compute the posterior, you should do it over the entire data set, i.e. $\mathcal{D} \cup \{0, 4, 5\}$.

Working with actual data means many computations. That's why you normally want to do this using a computer — not the human, but the electronic one. You can learn more about that in the follow-up course Basic Probability: programming.