

BASIC PROBABILITY: THEORY

Master of Logic, University of Amsterdam, 2016

TEACHERS Christian Schaffner and Philip Schulz 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 time some event happens in a given period of time. Its probability mass function is given by

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

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

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

Where $\alpha > -1$ and $\beta > -1$ are the two parameters of the distribution and $B(\alpha, \beta)$ is a normalizing constant that does not depend on θ .

Is this correctly parametrized? Normally you use $\alpha, \beta > 0$ and $\alpha - 1, \beta - 1$ in the exponent, right?

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

$$\mathcal{P}(\theta) = \ln[P(X_1, \dots, X_N \mid \Theta = \theta) \cdot P(\Theta = \theta)]$$

- (c) Find the derivative $\frac{\partial}{\partial \theta} \mathcal{P}(\theta)$.
- (d) Show that the maximum a posteriori estimate for θ is

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

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?