Weekly Exercise 4

Dr. Csaba Domokos
Technische Universität München, Computer Vision Group
May 22nd, 2017 (submission deadline: May 22nd, 2017)

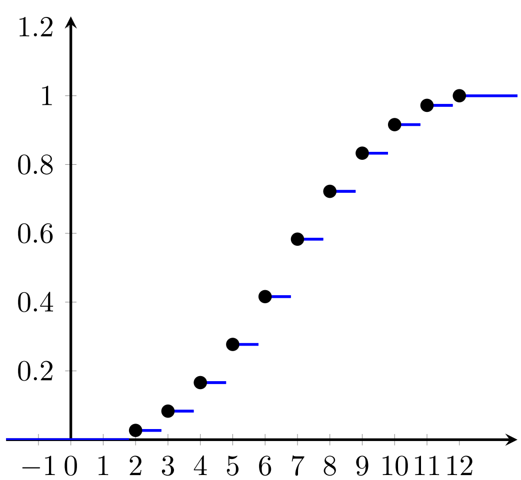
Probability distributions (8 Points)

Exercise 1 (**Probability distribution**, 2 points)**.** We throw two "fair" dice. Let us define a random variable X as the sum of the numbers showing on the dice. Define and draw the cumulative distribution function F_X.

Solution. The cumulative distribution function F_X is defined as

$$F_X(x) = \begin{cases} 0 & \text{if } x < 2 \\ \frac{1}{36} & \text{if } 2 \le x < 3 \\ \frac{1}{12} & \text{if } 3 \le x < 4 \\ \frac{1}{6} & \text{if } 4 \le x < 5 \\ \frac{5}{18} & \text{if } 5 \le x < 6 \\ \frac{5}{12} & \text{if } 6 \le x < 7 \\ \frac{7}{12} & \text{if } 7 \le x < 8 \\ \frac{13}{18} & \text{if } 8 \le x < 9 \\ \frac{5}{6} & \text{if } 9 \le x < 10 \\ \frac{11}{12} & \text{if } 10 \le x < 11 \\ \frac{35}{36} & \text{if } 11 \le x < 12 \\ 1 & \text{if } 12 \le x \ . \end{cases}$$

The graph of F_X looks as



Exercise 2 (Density function, 1 point). Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined as follows

$$f(x) = \begin{cases} x, & \text{if } 0 < x < 1 \ , \\ 2 - x, & \text{if } 1 < x < 2 \ , \\ 0, & \text{otherwise} \ . \end{cases}$$

Is it possible that f is a density function?

Solution. $f(x)$ is obviously non-negative. We need to check whether $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ holds.

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\mathrm{d}x &= \int_{-\infty}^{0} 0 \ \mathrm{d}x + \int_{0}^{1} x \ \mathrm{d}x + \int_{1}^{2} 2 - x \ \mathrm{d}x + \int_{2}^{\infty} 0 \ \mathrm{d}x \\ &= \left[\frac{x^2}{2}\right]_0^1 + 2(2-1) - \left[\frac{x^2}{2}\right]_1^2 \\ &= \frac{1}{2} + 2 - \frac{3}{2} = 1 \ . \end{aligned}$$

Therefore the answer is positive that is $f(x)$ can be a density function.

Exercise 3 (Random variable and expectation, 2 points). Let X be a discrete random variable with the possible values of 1, 2 and 3, where the corresponding probabilities are given as

$$P(X = 1) = \frac{1}{3} \ , \quad P(X = 2) = \frac{1}{2} \ , \quad P(X = 3) = \frac{1}{6} \ .$$

a) Define and draw the cumulative distribution function F_X.

b) What is the expected value of X?

Solution. a) The cumulative distribution function $F_X : \mathbb{R} \to \mathbb{R}$ is defined as

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \ , \\ \frac{1}{3} & \text{if } 1 \leq x < 2 \ , \\ \frac{5}{6} & \text{if } 2 \leq x < 3 \ , \\ 1 & \text{if } 3 \leq x \ . \end{cases}$$

b) The expected value is calculated as

$$\mathbb{E}[X] = 1\frac{1}{3} + 2\frac{1}{2} + 3\frac{1}{6} = \frac{11}{6} \ .$$

Exercise 4 (Random variable and expectation, 3 points). In order to express his gratitude, Siegfried invites Eduard to a pub for a couple of beers. There, they start playing a friendly game of darts. The dart board is a perfect disk of radius 10cm. If a dart falls within 1cm of the center, 100 points are scored. If the dart hits the board between 1 and 3cm from the center, 50 points are scored, if it is at a distance of 3 to 5cm 25 points are scored and if it is further away than 5cm 10 points are scored. As Siegfried and Eduard are both quite experienced dart players, they hit the dart board every time.

a) Define a random variable X corresponding to the score of throws.

b) What is the expected value of the scores?

Solution. a) The probability space $(\Omega, \mathcal{A}, P)$ is given by

$$\Omega = \{(x,y) \in \mathbb{R}^2 \mid \sqrt{x^2+y^2} \leq 10\} \ ,$$
$$\mathcal{A} = \left\{ A \subset \Omega \mid \int_\Omega \chi_A(x) \, \mathrm{d}x \text{ exists.} \right\} \ ,$$

and $P : \mathcal{A} \to [0,1]$, where

$$P(A) = \frac{\int_\Omega \chi_A(x)\mathrm{d}x}{100\pi} \ .$$

The random variable corresponding to the score of throws is defined as $X : \Omega \to \{10, 25, 50, 100\}$, where

$$X(x) = \begin{cases} 100, & \text{if } 0 \leq \|x\|_2 \leq 1, \\ 50, & \text{if } 1 \leq \|x\|_2 \leq 3, \\ 25, & \text{if } 3 \leq \|x\|_2 \leq 5, \\ 10, & \text{if } 5 \leq \|x\|_2 \leq 10. \end{cases}$$

b) The expected value of the scores is calculated as follows:

$$\begin{aligned} \mathbb{E}[X] &= 10 \cdot P(X=10) + 25 \cdot P(X=25) + 50 \cdot P(X=50) + 100 \cdot P(X=100) \\ &= 10\frac{75}{100} + 25\frac{16}{100} + 50\frac{8}{100} + 100\frac{1}{100} = 16.5 \ . \end{aligned}$$

Expectation-maximization algorithm (4 Points)

Exercise 5 (**Expectation-maximization algorithm**, 4 Points)**.** An alternative route in the derivation of the *Expectation-maximization algorithm* is to maximize the expected the *log-posterior* $\ln p(\boldsymbol{\theta} \mid \mathbf{X})$ instead of the expected *log-likelihood*. Show that for this case that the M step yields

$$\boldsymbol{\theta}^{(t+1)} \in \operatorname{argmax}_{\boldsymbol{\theta}} \left(Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) + \ln p(\boldsymbol{\theta}) \right) \ .$$

Therefore, one can assume the prior distribution of the parameters $\boldsymbol{\theta}$ (for example, to avoid singularities).

Hint: consider the maximization problem

$$\boldsymbol{\theta}^{(t+1)} \in \operatorname{argmax}_{\boldsymbol{\theta}} \mathbb{E}[\ln p(\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{Z}) \mid \mathbf{X}, \boldsymbol{\theta}^{(t)}] \ .$$

Solution. In our derivation of the *EM algorithm* (i.e. maximizing the *log-likelihood*) we have obtained the M step as follows:

$$\boldsymbol{\theta}^{(t)} \in \operatorname{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}) = \operatorname{argmax}_{\boldsymbol{\theta}} \mathbb{E}[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) \mid \mathbf{X}, \boldsymbol{\theta}^{(t)}] \ .$$

Consider the following maximization (of the *log-posterior*):

$$\begin{aligned}
\boldsymbol{\theta}^{(t+1)} &\in \operatorname{argmax}_{\boldsymbol{\theta}} \mathbb{E}[\ln p(\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{Z}) \mid \mathbf{X}, \boldsymbol{\theta}^{(t)}] \\
&= \operatorname{argmax}_{\boldsymbol{\theta}} \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{(t)}) \ln p(\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{Z}) \\
&= \operatorname{argmax}_{\boldsymbol{\theta}} \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \boldsymbol{\theta}^{(t)}) \big(\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}) - \ln p(\mathbf{X}, \mathbf{Z}) \big) \\
&= \operatorname{argmax}_{\boldsymbol{\theta}} \big(\ln p(\boldsymbol{\theta}) + \mathbb{E}[\ln p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\theta}) \mid \mathbf{X}, \boldsymbol{\theta}^{(t)}] \big) \\
&= \operatorname{argmax}_{\boldsymbol{\theta}} \Big(Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}) + \ln p(\boldsymbol{\theta}) \Big) \ .
\end{aligned}$$