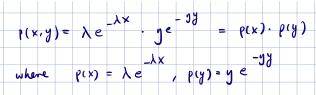
## Exercise 1: Maximum-Likelihood Estimation (5+5+5+5 P)

We consider the problem of estimating using the maximum-likelihood approach the parameters  $\lambda, \eta > 0$  of the probability distribution:

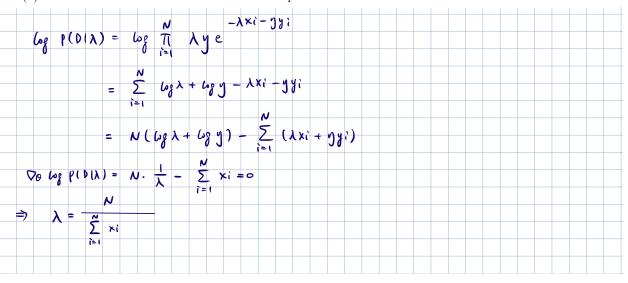
$$p(x,y) = \lambda \eta e^{-\lambda x - \eta y}$$

supported on  $\mathbb{R}^2_+$ . We consider a dataset  $\mathcal{D} = ((x_1, y_1), \dots, (x_N, y_N))$  composed of N independent draws from this distribution.

(a) Show that x and y are independent.



(b) Derive a maximum likelihood estimator of the parameter  $\lambda$  based on  $\mathcal{D}$ .



(c) Derive a maximum likelihood estimator of the parameter  $\lambda$  based on  $\mathcal{D}$  under the constraint  $\eta = 1/\lambda$ .

(d) Derive a maximum likelihood estimator of the parameter  $\lambda$  based on  $\mathcal{D}$  under the constraint  $\eta = 1 - \lambda$ .

$$Cog P(O|\lambda) = N(Ug\lambda + Ug(U-\lambda)) - \sum_{i=1}^{N} (\lambda x_i + y_i)$$

$$= N(Ug\lambda + Ug(U-\lambda)) - \sum_{i=1}^{N} (\lambda x_i + y_i - \lambda y_i)$$

$$\Rightarrow (x_i - y_i) = 0$$

$$\Rightarrow \lambda - \lambda^{2} = \lambda \sum_{i=1}^{N} (x_i - y_i) = 1$$

$$\Rightarrow \lambda = \frac{1-2\lambda}{\lambda - \lambda^{2}} = \lambda (x_i - y_i) = 1$$

$$\Rightarrow \lambda = \frac{2\pi\lambda + (x_i - y_i)}{2\pi}$$

## Exercise 2: Maximum Likelihood vs. Bayes (5+10+15 P)

An unfair coin is tossed seven times and the event (head or tail) is recorded at each iteration. The observed sequence of events is

$$\mathcal{D} = (x_1, x_2, \dots, x_7) = (\text{head}, \text{head}, \text{tail}, \text{tail}, \text{head}, \text{head}, \text{head}).$$

We assume that all tosses  $x_1, x_2, \ldots$  have been generated independently following the Bernoulli probability distribution

$$P(x \mid \theta) = \begin{cases} \theta & \text{if } x = \text{head} \\ 1 - \theta & \text{if } x = \text{tail,} \end{cases}$$

where  $\theta \in [0, 1]$  is an unknown parameter.

(a) State the likelihood function  $P(\mathcal{D}|\theta)$ , that depends on the parameter  $\theta$ .

$$p(D|O) = \pi p(x; |O) = O^{S}((-O)^{2}$$

(b) Compute the maximum likelihood solution  $\hat{\theta}$ , and evaluate for this parameter the probability that the next two tosses are "head", that is, evaluate  $P(x_8 = \text{head} \mid \hat{\theta})$ .

$$\log P(0|0) = \log 0^{5} (1-0)^{\frac{1}{2}} = 5 \cdot \log 0 + 2 \log (1-0)^{\frac{1}{2}}$$

$$\nabla_{\theta} \log P(0|0) = \frac{5}{0} - \frac{2}{1-0} = 0$$

$$\Rightarrow \frac{5-50-20}{0(1-0)} = 0$$

$$\Rightarrow 0 = \frac{5}{7}$$

$$\Rightarrow P(x_8 = head, x_9 = head(0) = P(x_8 = head(0) \cdot P(x_9 = head(0)))$$

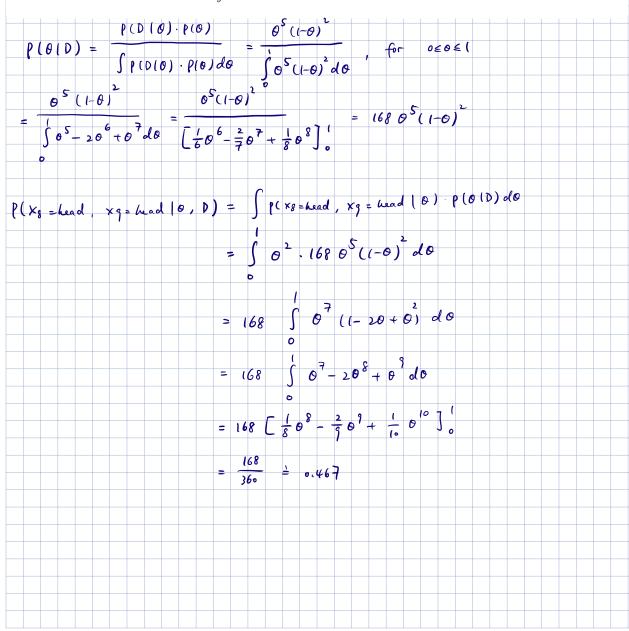
$$= 0.51$$

(c) We now adopt a Bayesian view on this problem, where we assume a prior distribution for the parameter  $\theta$  defined as:

$$p(\theta) = \begin{cases} 1 & \text{if } 0 \le \theta \le 1\\ 0 & \text{else.} \end{cases}$$

Compute the posterior distribution  $p(\theta|\mathcal{D})$ , and evaluate the probability that the next two tosses are head, that is,

$$\int P(x_8 = \text{head}, x_9 = \text{head} \mid \theta) p(\theta \mid \mathcal{D}) d\theta.$$

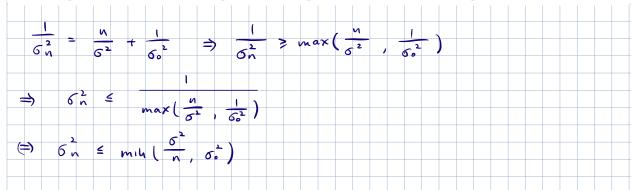


## Exercise 3: Convergence of Bayes Parameter Estimation (5+5 P)

We consider Section 3.4.1 of Duda et al., where the data is generated according to the univariate probability density  $p(x|\mu) \sim \mathcal{N}(\mu, \sigma^2)$ , where  $\sigma^2$  is known and where  $\mu$  is unknown with prior distribution  $p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2)$ . Having sampled a dataset  $\mathcal{D}$  from the data-generating distribution, the posterior probability distribution over the unknown parameter  $\mu$  becomes  $p(\mu|\mathcal{D}) \sim \mathcal{N}(\mu_n, \sigma_n^2)$ , where

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \qquad \frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2} \hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \qquad \hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n x_k.$$

(a) Show that the variance of the posterior can be upper-bounded as  $\sigma_n^2 \leq \min(\sigma^2/n, \sigma_0^2)$ , that is, the variance of the posterior is contained both by the uncertainty of the data mean and of the prior.



(b) Show that the mean of the posterior can be lower- and upper-bounded as  $\min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq \max(\hat{\mu}_n, \mu_0)$ , that is, the mean of the posterior distribution lies somewhere on the segment between the mean of the prior distribution and the sample mean.

