Exercise Sheet 1 - Probabilities
Prof. Dr. Paul Swoboda

NOTE: For this exercise sheet, you only need to submit a PDF.

Exercise 1: Conditional independence (0.5 Points)

Let $H \in \{1, \ldots, K\}$ be a discrete random variable, and let e_1 and e_2 be the observed values of two other random variables E_1 and E_2. Suppose we wish to calculate the conditional distribution $p(H \mid E_1 = e_1, E_2 = e_2)$, i.e., the numbers

$$p(H = 1 \mid e_1, e_2), \ldots, p(H = K \mid e_1, e_2).$$

a) Which of the following sets of numbers are sufficient for the calculation?

   i) $p(e_1, e_2)$, $p(H)$, $p(e_1 \mid H)$, $p(e_2 \mid H)$

   ii) $p(e_1, e_2)$, $p(H)$, $p(e_1, e_2 \mid H)$

   iii) $p(e_1 \mid H)$, $p(e_2 \mid H)$, $p(H)$

   iv) $p(H, e_1, e_2)$

b) Now suppose that we assume $E_1 \perp E_2 \mid H$, i.e., E_1 and E_2 are conditionally independent given H. Which of the above sets are sufficient now?

Give reasons for your acceptance or rejection of each option. Hint: use the sum rule, product rule, and/or Bayes rule.

Exercise 2: Pairwise and mutual independence (0.5 Points)

We say that two random variables are *pairwise independent* if

$$p(X_1, X_2) = p(X_1)p(X_2).$$

We say that n random variables are *mutually independent* if

$$p(X_{1:n}) = \prod_{i=1}^{n} p(X_i).$$

a) Show that pairwise independence between all pairs of variables does not necessarily imply mutual independence. It suffices to give a counterexample.

b) Show that mutual independence implies pairwise independence.

c) Show that independence is equivalent to $p(X_1|X_2) = p(X_1)$.

Exercise 3: Bias-Variance Decomposition (1 Points)

Assume you are given an estimator $\hat{\theta}$ for a model's true parameters θ^*. Define its

- bias as $\text{bias}[\hat{\theta}] = E[\hat{\theta} - \theta^*]$,
- variance as $\text{var}[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2]$,
- mean squared error (MSE) as $\text{mse}[\hat{\theta}] = E[(\hat{\theta} - \theta^*)^2]$.

Prove that the bias-variance decomposition

$$\text{mse}[\hat{\theta}] = \text{bias}[\hat{\theta}]^2 + \text{var}[\hat{\theta}]$$

holds.

Exercise 4: MLE for Gaussian Distribution (σ^2 known) (1.5 Points)

Assume you are given N independent samples $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ drawn from data distribution $p^* = \mathcal{N}(\mu^*, \sigma^2)$, where μ^* is unknown and $\sigma^2 > 0$ is known. The likelihood of this sequence of data points is given by

$$\mathcal{L}\left(\mu|x_1, \ldots, x_N\right) = \prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2).$$

In this exercise, we derive the maximum likelihood estimate

$$\hat{\mu}_{\text{MLE}} = \frac{1}{N} \sum_i x_i$$

for μ^* and analyze it.

a) We often work with the log-likelihood $\ell = \ln \mathcal{L}$ instead of the likelihood (both functions agree on their maxima). Simplify ℓ.

b) Compute the derivative $\frac{\partial}{\partial \mu}\ell$. Discuss why we prefer to work with the log-likelihood instead of the likelihood.

c) Set the derivative to zero and solve for μ. You should obtain $\hat{\mu}_{\text{MLE}}$ as above.

d) Compute the bias of $\hat{\mu}_{\text{MLE}}$.

e) Compute the variance of $\hat{\mu}_{\text{MLE}}$.

Hint. Recall and use basic properties of variance, i.e., do not compute $\text{var}\left[\hat{\mu}_{\text{MLE}}\right]$ via a lengthy computation.

f) Compute the mean squared error of $\hat{\mu}_{\text{MLE}}$.

Exercise 5: MLE for Binomial Distribution (1.5 Points)

Assume we are given a Binomial distribution

$$\text{Bin}(n_H|n, \theta) = \binom{n}{n_H} \theta^{n_H}(1-\theta)^{n-n_H}$$

for a coin toss, where n_H is the number of heads up events, n is the total number of coin flips where we want to estimate its underlying parameter θ. Show that the MLE estimator for θ is n_H/n, where n_H is the number of heads up.