Problem sheet 2

1. Let X and Y be two random variables (RVs). Their covariance $\mathrm{COV}(X,Y)$ is defined as:

$$\mathrm{COV}(X,Y) \quad = \quad \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

   (a) Show that $\mathrm{COV}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

   (b) Show that if X and Y are independent $\mathrm{COV}(X,Y) = 0$.

   (c) Consider a RV $X \in \{-1, 0, 1\}$, which takes on each of its three possible values with equal probability. Let a second RV Y be defined by $Y = X^2$. Calculate the covariance of X and Y. Interpret your result. (4 marks)

2. (*Markov inequality*) Let X be a continuous, non-negative random variable (RV) and a a positive constant. Show that:

$$P(X \geq a) \quad \leq \quad \frac{\mathbb{E}[X]}{a}$$

   (Hint: write down the expectation of X as an integral, and split the integral at a.)

3. (*Chebyshev inequality*) Let X be any RV with mean μ_X and whose variance σ_X^2 exists. Show that for any positive constant a:

$$P(|X - \mu_X| \geq a) \quad \leq \quad \frac{\sigma_X^2}{a^2}$$

4. An estimator $\hat{\theta}_n$ of a parameter is said to be consistent if it converges in probability to the true parameter value θ, that is if:

$$\forall \epsilon > 0, \quad \lim_{n \to \infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) \quad = \quad 0$$

   Using the Chebyshev inequality, show (informally) that the conditions

$$\begin{aligned} \mathbb{E}[\hat{\theta}_n] &= \theta \\ \lim_{n \to \infty} \mathrm{VAR}(\hat{\theta}_n) &= 0 \end{aligned}$$

   are sufficient to establish consistency.

5. (*Law of Large Numbers*) Let $X_1, X_2 \ldots X_n$ be a set of independently and identically distributed RVs (a *random sample*) with $\mathbb{E}[X_i] = \mu_X$ and $\mathrm{VAR}(X_i) = \sigma_X^2 < \infty$. Show that the *sample mean*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges in probability to the mean μ_X.

6. $\mathbf{X}_1 \ldots \mathbf{X}_n$, $\mathbf{X}_i \in \mathbb{R}^d$ are independently and identically distributed multivariate Normal random vectors, each having pdf:

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

   (a) Write down the log-likelihood function for this model.

   (b) Derive maximum likelihood estimators for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. (The easiest way to solve this problem is using matrix derivatives. See Appendix C in Bishop for an introduction. You may also find the note "Matrix Identities" by Roweis useful; this is available on the course website. The derivatives

$$\frac{\partial}{\partial \mathbf{A}} \log(|\mathbf{A}|) = (\mathbf{A}^{-1})^T$$
$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^{-1}\mathbf{A}) = -\mathbf{X}^{-1}\mathbf{A}^T\mathbf{X}^{-1}$$

   may prove particularly useful. Here, $\mathrm{Tr}(\cdot)$ denotes the *trace* of its matrix argument.)