3.1 Maximum likelihood estimation

Exercise 3.1 (Order statistics). Let $X_1, \ldots, X_n$ be an i.i.d. sample of the random variable $X$. Suppose that $X$ has cumulative distribution function $F$ and density $f$. We call order statistics the ordered values of $X_1, \ldots, X_n$, denoted by

$$X^{(1)} \leq X^{(2)} \leq \cdots \leq X^{(n)}.$$ 

In particular, $X^{(1)} = \min_{1 \leq i \leq n} X_i$ and $X^{(n)} = \max_{1 \leq i \leq n} X_i$.

1. Show that the $k$th order statistic has cumulative distribution function given by

$$F_k(t) = \sum_{j=k}^{n} \binom{n}{j} F(t)^j (1 - F(t))^{n-j}.$$ 

**Hint:** for a given $i$, start by writing the probability that exactly $k-1$ elements are smaller than $X_i$.

2. Deduce from the previous question that the density of $X^{(k)}$ is given by

$$f_k(t) = \frac{n!}{(k-1)!(n-k)!} F(t)^{k-1} (1 - F(t))^{n-k} f(t).$$

Exercise 3.2 (Maximum likelihood estimation, exponential random variables). Let $X_1, \ldots, X_n$ be an i.i.d. sample from $X \sim \mathcal{E}(\lambda)$. Recall that, in that case, $X$ has density

$$f_\lambda(t) = \lambda \exp(-\lambda t) \mathbf{1}_{t \geq 0}.$$ 

What is the maximum likelihood estimator for $\lambda$? Is it consistent?

3.2 Consistency of $M$-estimators

Exercise 3.3 (Pathological behaviors for convergence of functions). Find a sequence of fixed (non-random) functions $M_n : \mathbb{R} \to \mathbb{R}$ such that:

1. $M_n$ converges pointwise and each $M_n$ has a unique maximum at a point $\theta_n$, but the sequence $\theta_n$ does not converge. Can you also find a sequence $M_n$ that converges uniformly?

2. $M_n$ converges pointwise but not uniformly to a limit $M$, each $M_n$ has a unique maximum at a point $\theta_n$, and the sequence $\theta_n$ converges to $\theta_0$.

Exercise 3.4 (Maximum likelihood, uniform random variable). Consider $\hat{\theta}_n$ the maximum likelihood estimator for the uniform distribution on $[0, \theta]$. We proved together that

$$\hat{\theta}_n = \max_{1 \leq i \leq n} X_i.$$ 

Show that it is consistent for $\theta$, but this time by checking carefully the hypotheses of the consistency theorem given in the lecture.
**Exercise 3.5 (Kullback-Leibler divergence).** Let $P$ and $Q$ be two distributions with probability density functions $p$ and $q$. The Kullback-Leibler divergence from the distribution $Q$ to the distribution $P$ is defined as

$$D_{KL}(P \parallel Q) = \int p(x) \log \frac{p(x)}{q(x)} dx.$$ 

1. Prove that $D_{KL}(P \parallel Q) \geq 0$, this time using Jensen’s inequality. *Hint:* what can you say about the function $x \mapsto x \log x$?

2. Suppose that $P_1$ and $P_2$ are independent distributions with joint distribution $P(x,y) = P_1(x)P_2(y)$, and $Q, Q_1, Q_2$ likewise. In that case, prove that

$$D_{KL}(P \parallel Q) = D_{KL}(P_1 \parallel Q_1) + D_{KL}(P_2 \parallel Q_2).$$

3. Suppose that $P$ (resp. $Q$) is exponential of parameter $\lambda > 0$ (resp. $\nu > 0$). Compute $D_{KL}(P \parallel Q)$ in closed-form.