# Homework 2

**Question 1:** Let $X_1, \ldots, X_n$ be iid sampled from Poisson($\lambda$). Let $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Find the bias and variance of $\hat{\lambda}$.

**Question 2:** Let $X_1, \ldots, X_n$ be iid sampled from Uniform$(0, \theta)$. Let $\hat{\theta} = \max\{X_1, \ldots, X_n\}$. Find the bias and variance of $\hat{\theta}$.

**Question 3:** A number $M(X)$ is a median of a random variable $X$ if $P(X > M(X)) \leq \frac{1}{2}$ and $P(X < M(X)) \leq \frac{1}{2}$.

(1) Show that $M(X)$ always exists, and that if $X$ is absolutely continuous with strictly positive density function, then the median is unique

(2) If $X$ has a finite second moment, show that

$$M(X) = \mathbb{E}(X) + O(\sqrt{\text{Var } X})$$

for any median $M(X)$.

Complete the following exercises from Wainwright's book.

**Exercise 4.3** (Maximum likelihood and uniform laws) Recall from Example 4.8 our discussion of empirical and population risks for maximum likelihood over a family of densities $\{p_\theta, \theta \in \Omega\}$.

(a) Compute the population risk $R(\theta, \theta^*) = \mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(X)}{p_\theta(X)}\right]$ in the following cases:

(i) Bernoulli: $p_\theta(x) = \frac{e^{\theta x}}{1+e^{\theta x}}$ for $x \in \{0, 1\}$;

(ii) Poisson: $p_\theta(x) = \frac{e^{\theta x} e^{-\exp(\theta)}}{x!}$ for $x \in \{0, 1, 2, \ldots\}$;

(iii) multivariate Gaussian: $p_\theta$ is the density of an $\mathcal{N}(\theta, \Sigma)$ vector, where the covariance matrix $\Sigma$ is known and fixed.

(b) For each of the above cases:

(i) Letting $\widehat{\theta}$ denote the maximum likelihood estimate, give an explicit expression for the excess risk $E(\widehat{\theta}, \theta^*) = R(\widehat{\theta}, \theta^*) - \inf_{\theta \in \Omega} R(\theta, \theta^*)$.

(ii) Give an upper bound on the excess risk in terms of an appropriate Rademacher complexity.

**Exercise 4.6** (Too many linear classifiers) Consider the function class

$$\mathscr{F} = \{x \mapsto \text{sign}(\langle \theta, x \rangle) \mid \theta \in \mathbb{R}^d, \ \|\theta\|_2 = 1\},$$

corresponding to the $\{-1, +1\}$-valued classification rules defined by linear functions in $\mathbb{R}^d$. Supposing that $d \geq n$, let $x_1^n = \{x_1, \ldots, x_n\}$ be a collection of vectors in $\mathbb{R}^d$ that are linearly independent. Show that the empirical Rademacher complexity satisfies

$$\mathcal{R}(\mathscr{F}(x_1^n)/n) = \mathbb{E}_\varepsilon\left[\sup_{f \in \mathscr{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(x_i)\right|\right] = 1.$$

Discuss the consequences for empirical risk minimization over the class $\mathscr{F}$.

**Exercise 4.12** (VC dimension of left-sided intervals) Consider the class of left-sided half-intervals in $\mathbb{R}^d$:

$$\mathcal{S}^d_{\text{left}} := \{(-\infty, t_1] \times (-\infty, t_2] \times \cdots \times (-\infty, t_d] \mid (t_1, \ldots, t_d) \in \mathbb{R}^d\}.$$

Show that for any collection of $n$ points, we have $\text{card}(\mathcal{S}^d_{\text{left}}(x_1^n)) \leq (n+1)^d$ and $\nu(\mathcal{S}^d_{\text{left}}) = d$.

**Exercise 4.13** (VC dimension of spheres) Consider the class of all spheres in $\mathbb{R}^2$—that is

$$\mathcal{S}^2_{\text{sphere}} := \{S_{a,b},\ (a,b) \in \mathbb{R}^2 \times \mathbb{R}_+\}, \tag{4.34}$$

where $S_{a,b} := \{x \in \mathbb{R}^2 \mid \|x - a\|_2 \leq b\}$ is the sphere of radius $b \geq 0$ centered at $a = (a_1, a_2)$.

(a) Show that $\mathcal{S}^2_{\text{sphere}}$ can shatter any subset of three points that are not collinear.

(b) Show that no subset of four points can be shattered, and conclude that the VC dimension is $\nu(\mathcal{S}^2_{\text{sphere}}) = 3$.