STOR 565 Homework

1. Let $\mathcal{P} = \{p_\lambda : \lambda > 0\}$ be the family of Poisson pmfs $p_\lambda(k) = e^{-\lambda}\lambda^k/k!$ for integers $k \geq 0$. Suppose that we draw n samples independently from a fixed distribution $p_{\lambda_0} \in \mathcal{P}$ and observe integers $x_1, \ldots, x_n \geq 0$.

   a. Write down the likelihood $L(\lambda)$ and the log-likelihood $\ell(\lambda)$ for the family $\mathcal{P}$.

   b. Find the maximum likelihood estimate $\hat{\lambda}_n^{\text{MLE}}$ of λ_0.

2. Recall that the moment generating function of a random variable X is defined by $M_X(s) = \mathbb{E}e^{sX}$ for all s such that the expectation is finite. Find the moment generating function (MGF) of the following distributions.

   a. Poisson(λ)

   b. $\mathcal{N}(0, 1)$

3. Let $A \in \mathbb{R}^{k \times k}$ be invertible. Show that if v is an eigenvector of A with eigenvalues λ then v is an eigenvector of A with eigenvalue λ^{-1}.

4. Use Jensen's inequality to show that for $a, b > 0$ and $p \geq 1$,

$$(a + b)^p \leq 2^{p-1}\,[a^p + b^p].$$

Verify this inequality in case $p = 2$ by a direct calculation.

5. (Bivariate normal distribution). Let $X = (X_1, X_2)^t \sim \mathcal{N}_2$ with

$$\mathbb{E}X_1 = \mu_1,\ \mathbb{E}X_2 = \mu_2,\ \text{Var}(X_1) = \sigma_1^2,\ \text{Var}(X_2) = \sigma_2^2,\ \text{Corr}(X_1, X_2) = \rho \in [-1, 1]$$

   (a) Find $\mu = \mathbb{E}X$ and $\Sigma = \text{Var}(X)$ in terms of the quantities above.

   (b) Find the determinant of Σ and conclude that Σ is invertible if and only if $\rho \in (-1, 1)$.

   (c) Find Σ^{-1} when $\rho \in (-1, 1)$.

   (d) Write down the density $f(x)$ of X in the case $\rho \in (-1, 1)$. Feel free to look up the general form of the density in a text-book, or online, and then plug in the values of μ and Σ^{-1} that you found above.

6. Let $g : I \to \mathbb{R}$ be a twice differentiable function with $g'' \geq 0$. Let $x, y \in I$ and $\alpha \in (0, 1)$.

   a. Expand the function g in a second order Taylor series around the point $z = \alpha x + (1 - \alpha)y$. In other words, express $g(v)$ in terms of $g(z)$, $g'(z)$, and $g''(u)$ for u between v and z.

   b. Apply the expression in part (a) to $g(x)$ and $g(y)$ and use it to obtain lower bounds for both values, using the fact that the last term in the Taylor series is non-negative.

   c. Use the bounds in (b) to show that g is convex by considering the sum $\alpha g(x) + (1 - \alpha)g(y)$.

7. Let $D_n = (X_1, Y_1), \ldots, (X_n, Y_n) \in \mathcal{X} \times \{0, 1\}$ be a classification data set. Recall that the empirical risk of a fixed classification rule $\phi : \mathcal{X} \to \{0, 1\}$ is defined by

$$\hat{R}_n(\phi) \;=\; \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\phi(X_i) \neq Y_i)$$

and that the risk of ϕ is $R(\phi) = \mathbb{P}(\phi(X) \neq Y)$.

   a. Find the mean and the variance of $\hat{R}_n(\phi)$.

   b. Find an upper bound on $\text{Var}(\hat{R}_n(\phi))$ that holds for any classification rule ϕ.

   c. Argue carefully that $\sum_{i=1}^{n} \mathbb{I}(\phi(X_i) \neq Y_i)$ is binomially distributed. Identify the parameters of the distribution.

   d. Use the variance bound above and Chebyshev's inequality to bound $\mathbb{P}(|\hat{R}_n(\phi) - R(\phi)| \geq t)$ for $t \geq 0$.

Now suppose that $\mathcal{F} = \{\phi_1, \ldots, \phi_N\}$ is a finite family of classification rules. Use the union bound and the results of (d) to find a bound on

$$\mathbb{P}\left(\max_{1 \leq j \leq N} |\hat{R}_n(\phi_j) - R(\phi_j)| \geq t \right).$$

8. Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a classification data set with $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$. Describe the 1-nearest neighbor classification rule $\hat{\phi}_n$, and carefully argue that its empirical risk is zero. In this case, what does the empirical risk tell you about the true risk $R(\hat{\phi}_n)$?

9. Let U_1, U_2 be uncorrelated random variables with mean zero and variance one. Define $U = (U_1, U_2)^t$. Let $X = (X_1, X_2)^t$ be a random vector with components

$$X_1 \;=\; a\,U_1 + b\,U_2 \;\;\text{and}\;\; X_2 \;=\; c\,U_1 + d\,U_2$$

a. Find $\mathbb{E}[U]$.

b. What is $\mathrm{Var}(U)$?

c. Find $\mathbb{E}X$.

d. Find the matrix $\mathrm{Var}(X)$ by directly calculating each entry using the definitions of X_1 and X_2.

e. Find $\mathbf{A}$ such that $X = \mathbf{A}U$.

f. Find $\mathrm{Var}(X)$ using the formula for $\mathrm{Var}(\mathbf{A}U)$.

g. In terms of a, b, c and d, when is $\mathbf{A}$ invertible?

h. If the random vector U is bivariate normal, what is the distribution of X when $\mathbf{A}$ is invertible?