Group Assignments

12.1 Group Assignment 1

1. Suppose that A and B are independent events. Show that A^c and B^c are independent.

2. The probability that a child has brown hair is 1/4. Assume independence between children and assume there are three children.

   (a) If it is known that at least one child has brown hair, what is the probability that at least two children have brown hair?

   (b) If it is known that the oldest child has brown hair, what is the probability that at least two children have brown hair?

3. Let (X, Y) be uniformly distributed on the unit disc, $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$. Set $R = \sqrt{X^2 + Y^2}$. What is the CDF and PDF of R?

4. A fair coin is tossed until a head appears. Let X be the number of tosses required. What is the expected value of X?

5. Let $X_1, \ldots, X_n$ be i.i.d. from Bernoulli(p).

   (a) Let $\alpha > 0$ be fixed and define

   $$\epsilon_n = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}.$$

   Let $\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ and define the confidence interval $I_n = [\hat{p}_n - \epsilon_n, \hat{p}_n + \epsilon_n]$. Use Hoeffding's inequality to show that $P(p \in I_n) \geq 1 - \alpha$.

   (b) Let $\alpha = 0.05$ and $p = 0.4$. Conduct a simulation study to see how often the confidence interval I_n contains p (called coverage). Do this for $n = 10, 100, 1000, 10000$. Plot the coverage as a function of n.

   (c) Plot the length of the confidence interval as a function of n.

   (d) Say that $X_1, \ldots, X_n$ represents if a person has a disease or not. Let us assume that unbeknownst to us the true proportion of people with the disease has changed from $p = 0.4$ to $p = 0.5$. We use the confidence interval to make a decision, that is, when presented with evidence (samples) we calculate I_n and our decision is that the true proportion of people with the disease is in I_n. Conduct a simulation study to answer the following question: Given that the true proportion has changed, what is the probability that our decision is correct? Again using $n = 10, 100, 1000, 10000$.

12.2 Group Assignment 2

1. Consider a supervised learning problem where we assume that $Y \mid X$ is Poisson distributed. That is, the conditional density of $Y \mid X$ is given by
$$f_{Y|X}(y, x) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad \lambda(x) = \exp(\alpha \cdot x + \beta).$$
Here α is a vector (slope) and β is a number (intercept). Follow the calculations from Section 4.2.1 to derive a loss that needs to be minimized with respect to α and β. Note: do we really need the factorial term?

2. Let $X_1, \ldots, X_n$ be i.i.d. from Uniform$(0, \theta)$. Let $\hat{\theta} = \max(X_1, \ldots, X_n)$. First, find the distribution function of $\hat{\theta}$. Then compute the bias$(\hat{\theta})$, se$(\hat{\theta})$ and MSE$_n(\hat{\theta})$.

3. Consider the continuous distribution with density
$$p(x) = \frac{1}{2}\cos(x), \quad -\frac{\pi}{2} < x < \frac{\pi}{2}.$$
   (a) Find the distribution function F.
   (b) Find the inverse distribution function F^{-1}.
   (c) To sample using an Accept-Reject sampler, Algorithm 1, we need to find a density g such that $p(x) \leq Mg(x)$ for some $M > 0$. Find such a density g and find the value of M.

4. Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of i.i.d. discrete random variables, where $P(Y_i = 0) = 0.1$, $P(Y_i = 1) = 0.3$, $P(Y_i = 2) = 0.2$, and $P(Y_i = 3) = 0.4$. Let $X_n = \max\{Y_1, \ldots, Y_n\}$. Let $X_0 = 0$ and verify that $X_0, X_1, \ldots, X_n$ is a Markov chain. Find the transition matrix P.

5. Let $X_1, \ldots, X_n$ be i.i.d. from some distribution F that is unknown. Let $\hat{F}_n$ be the empirical distribution function, use this to find an estimate of the p quantile of F (call it q). Use Theorem 5.28 to find a confidence interval for q.

12.3 Group Assignment 3

1. Consider a three-state $(1, 2, 3)$ Markov chain with transition matrix

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}.$$

   (a) Draw the transition diagram.

   (b) Find the stationary distribution π.

   (c) Given that the chain is in state 1 at time 1, what is the probability that the chain is in state 2 at time 4?

   (d) Given that the chain is in state 1 at time 1, what is the expected time until the chain is in state 3 for the first time?

   (e) What is the period of each state?

2. Assume that we are trying to classify a binary outcome Y, i.e., our data is of the form $(X, Y) \sim F_{X,Y}$, where $Y \in \{0, 1\}$ and $X \in \mathbb{R}^d$. We have used data to train a classifier $g(X)$. We can evaluate the performance of the classifier using i.i.d. testing data, $(X_1, Y_1), \ldots, (X_n, Y_n)$. We are interested in estimating the following quantities:

$$\text{Precision} : P(Y = 1 \mid g(X) = 1),$$
$$\text{Recall} : P(g(X) = 1 \mid Y = 1).$$

   (a) Write down the empirical version of the precision and recall.

   (b) Let us now think that the variable Y denotes if a battery's health has deteriorated or not, and let X denote a bunch of constructed health indicators about the battery. If the model $g(X)$ predicts that the battery has deteriorated, you need to run a test to confirm this. The cost of running the test is c when the battery is not deteriorated. On the other hand, if the battery is in fact deteriorated and the test is not run, the battery will die during use and the cost of this is d. Define a random variable representing the cost of the decision $g(X)$ and write down the formula for the expected cost in terms of the precision and recall.

(c) Advanced question: Can you produce a confidence interval for the expected cost? What about the precision and recall?

3. Let X and Y be two d-dimensional zero mean, unit variance Gaussian random vectors. Show that X and Y are nearly orthogonal by calculating their dot product. Can you, for instance, also bound the probability that the dot product is larger than ϵ?

4. Let $u_1, \ldots, u_r$ be $n \times 1$ unit length vectors that are linearly independent, i.e.,

$$\sum_{i=1}^{r} \alpha_i u_i = 0 \implies \alpha_i = 0 \text{ for all } i.$$

   (a) Verify that the matrix $u_i u_i^T$ is a rank one matrix for all i. What is the null-space and range of $u_i u_i^T$?

   (b) Verify that the matrix $U = \sum_{i=1}^{r} u_i u_i^T$ is a rank r matrix.

   (c) i. If we perform SVD on U, are the vectors $u_1, \ldots, u_r$ the same as the right singular vectors? If not, can you give an example?

      ii. What if the vectors $u_1, \ldots, u_r$ are all orthogonal? In this case, what are the singular values of U?

5. Let $X \sim \text{Uniform}(B_1)$ and define $Y = \|X\|_2$ (the Euclidean norm).

   (a) Find the distribution function of Y.

   (b) What is the distribution of $\ln(1/Y)$?

   (c) Calculate $E[\ln(1/Y)]$, first by using the distribution function of Y and then by using the distribution function of $\ln(1/Y)$.