MATH 357: Assignment 1

Due date: Friday, January 31th, at 11:55 pm.

To be submitted online via myCourses

Winter 2025

(i) Provide detailed answers to the following problems.

(ii) You can no doubt find the answers to many of these questions on the internet or in text books. However, I advise you to attempt these questions without such consultation. The only way to learn the concepts taught in Math 357 is by banging your head on the wall.

Problem 1. Suppose $X_1, X_2, \ldots$ are continuous and identically distributed random variables with pdf and cdf f and F, respectively. Each X_i represents annual rainfall at a given location or region in year i.

(a) Find the distribution of the number of years until the first year's rainfall X_1 is exceeded for the first time.

(b) Show that the expected value of the number of years until the first year's rainfall X_1 is exceeded for the first time is ∞.

Problem 2. (Finite mixture models) Let $Z \in \{1, \ldots, K\}$, for a pre-specified integer $K \in \mathbb{N}$, be a discrete random variable with pmf

$$\pi_j = P(Z = j), j = 1, \ldots, K,$$

where $0 < \pi_j < 1$ and $\sum_{j=1}^{K} \pi_j = 1$. Also assume a random variable X has the following conditional pmf/pdf,

$$(X|Z = j) \sim f_j \quad , \quad j = 1, \ldots, K.$$

(a) Find the marginal pmf/pdf of X in terms of π_j's and f_j's.

(b) (Gaussian finite mixtures) Assume that f_j in the above formulation is the pdf of $N(\mu_j, \sigma_j^2)$, for each j. Find the (marginal) expected value, $\mathrm{E}(X)$, and variance of X, $\mathrm{Var}(X)$, in terms of $(\pi_j, \mu_j, \sigma_j^2), j = 1, \ldots, K$.

Some insight: Finite mixtures are very flexible and powerful statistical models for capturing (hidden) heterogeneity in a population. Focusing on the battery life of a given brand of electric car discussed in class: there could be cars that die quite early, those that will have normal life time as claimed or expected, and those that go beyond expectation and survive more than expected! This is an example of heterogeneous behavior in a population or distribution of a random variable X. A mixture of $K = 3$ Gaussians or Exponentials could be used to model the distribution of X in this example! In practice, the main challenge is that when we have data $X_1, \ldots X_n$ from such populations, we won't know the labels $Z_1, \ldots, Z_n$ for each data point!

Problem 3. Let X be a real-valued random variable with an unknown distribution F, where $F(x) = P(X \leq x), \forall x \in \mathbb{R}$. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from F. Recall the *empirical cdf* (ecdf) $F_n(x)$ discussed in class which is used as an "educated guess" of $F(x)$, for any $x \in \mathbb{R}$. Assume that we are also interested in the general probabilities $P(X \in B)$, for any Borel set B, which is called the probability distribution of X. Use the above random sample to answer the following questions.

(a) For any $t_1, t_2 \in \mathbb{R}$, find $\mathrm{Cov}\Big(F_n(t_1), F_n(t_2)\Big)$.

(b) By extending the definition of the ecdf $F_n(x)$, provide an "educated guess" for the probability distribution of X and call it $P_n(B)$, for any Borel set B.

(c) Find the expected value and variance of $P_n(B)$.

(d) Choose a_n and b_n such that the limiting distribution of $a_n\{P_n(B) - b_n\}$ is $N(0, 1)$, as $n \to \infty$.

Problem 4. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution F with a pdf f. Consider the two order statistics $X_{(1)} = \min_{1\le i\le n} X_i$ and $X_{(n)} = \max_{1\le i\le n} X_i$, for any $n \ge 1$. The sample range is defined as $R_n = X_{(n)} - X_{(1)}$.

(a) Derive the joint cdf and pdf of $(X_{(1)}, X_{(n)})$ in terms of f and F.

(b) Find the cdf and pdf of the sample range R_n.

(c) For $\alpha > 0$, assume that $\lim_{x\to\infty} x^\alpha P(X_1 > x) = b > 0$. Find the limiting distribution (both cdf and pdf) of $U_n = (bn)^{-1/\alpha} X_{(n)}$, as $n \to \infty$.

(d) Assume that $\lim_{x\to\infty} e^x P(X_1 > x) = b > 0$. Find the limiting distribution (both cdf and pdf) of $V_n = X_{(n)} - \log(bn)$, as $n \to \infty$.

Problem 5. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$.

(a) Show that $Y_i = \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(1)}$, for $i = 1, \ldots, n$.

(b) Using the result in part (a), show that

$$\sum_{i=1}^n Y_i \sim \chi^2_{(n)}.$$

(c) In class, using a transformation technique we proved that the sample mean $\bar{X}_n$ and the sample variance S_n^2 are independent random variables (statistics). Verify the same claim using the moment generating technique.

Problem 6. Assume that $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ are two independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively.

(a) Find the distribution of $\bar{X}_m - \bar{Y}_n$.

(b) Find the distribution of $U(n,m) = \frac{(m-1)S_m^2}{\sigma_1^2} + \frac{(n-1)S_n^2}{\sigma_2^2}$.

(c) Find the constant $c(n,m) > 0$ such that

$$V(n,m) = \frac{c(n,m)\ \{\bar{X}_m - \bar{Y}_n - (\mu_1 - \mu_2)\}}{\sqrt{U(n,m)}} \sim t(m+n-2).$$

The constant $c(n,m) > 0$ is also a function of (σ_1^2, σ_2^2).

(d) Assume that $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ are two independent random samples from, respectively, the distributions F_1 and F_2, such that $\mu_1 = E(X_1), \sigma_1^2 = Var(X_1) < \infty$ and $\mu_2 = E(Y_1), \sigma_1^2 = Var(Y_1) < \infty$. Without using the t distribution in part (a), directly find the limiting distribution of the random sequence $V(n,m)$, as both $(m,n) \to \infty$. Note that you still need your choice of $c(n,m)$ from part (c) to answer this part.

Problem 7.

(a) Let $X \sim t(\nu)$. Show that $X^2 \sim F(1, \nu)$.

(b) Let $Y \sim F(p, q)$. Show that,

(i) $Y^{-1} \sim F(q, p)$.

(ii) $\frac{rX}{1+rX} \sim \text{Beta}(p/2, q/2)$, where Beta is the beta distribution and $r = p/q$.

Note: the pdf of a $Y \sim Beta(a, b)$ distribution:

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{(a-1)}(1-y)^{(b-1)} \quad , \quad 0 \leq y \leq 1$$

where $\Gamma(\cdot)$ is the gamma function, and $a, b > 0$ two parameters of a Beta distribution.