Statistics II: Introduction to Inference

Problem set 1

In all the following problems, $X_1, \ldots, X_n$ represent a random sample of size n from a distribution $F_{\boldsymbol{\theta}}$. The specific form of the family $F_{\boldsymbol{\theta}}$ will be provided in each individual problem.

1. Let $F_{\boldsymbol{\theta}}$ be `Binomial`(m, ρ) where $\theta = \rho$.
   (a) Consider the class of linear estimators of ρ, i.e., the estimators of the form $T_{\mathbf{l}}(\mathbf{X}) = \sum_{j=1}^{n} l_j X_j$, $l_j \in \mathbb{R}$, $j = 1, \ldots, n$. Find conditions on $\mathbf{l} = (l_1, \ldots, l_n)'$ under which $T_{\mathbf{l}}(\mathbf{X})$ is an unbiased estimator of ρ.

   [*Let* $\mathbf{l}^\star$ *be a choice of* $\mathbf{l}$ *which satisfies the condition obtained in part (a). Then the estimator obtained by replacing* $\mathbf{l}$ *by* $\mathbf{l}^\star$ *in* $T_{\mathbf{l}}(\mathbf{X})$ *is called a* linear unbiased estimator *of* ρ.]

   (b) Find the variance of $T_{\mathbf{l}}(\mathbf{X})$. Denote it by $\sigma^2_{\mathbf{l}}$.
   (c) Minimize $\sigma^2_{\mathbf{l}}$ with respect to $\mathbf{l}$ subject to the constraint obtained in part (a).

   [*Let the solution obtained in part (c) be* $\mathbf{l}^\star$. *The estimator obtained by replacing* $\mathbf{l}$ *by* $\mathbf{l}^\star$ *in* $T_{\mathbf{l}}(\mathbf{X})$ *is called the* Best Liner Unbiased Estimator (BLUE).]

   (d) Is the BLUE same as the UMVUE of ρ?

2. Let $F_{\boldsymbol{\theta}}$ be some distribution with mean μ and variance σ^2 (i.e., $E(X_1) = \mu$ and $\mathrm{var}(X_1) = \sigma^2$), and $\boldsymbol{\theta} = (\mu, \sigma^2)$. Find the BLUE for μ.

3. Let $F_{\boldsymbol{\theta}}$ be `exponential`(λ) distribution with the pdf

$$f_\lambda(x) = \lambda \exp\{-\lambda x\}, \qquad \lambda > 0,\ x > 0,$$

   and $\theta = \lambda$.
   (a) Find the UMVUE of the population mean $\psi(\theta) = \theta^{-1}$.
   (b) Is it an efficient estimator?

4. Let $F_{\boldsymbol{\theta}}$ be `Poisson`(λ) distribution and $\theta = \lambda$.
   (a) Find the UMVUE of θ.
   (b) Is it an efficient estimator?

5. Let $F_{\boldsymbol{\theta}}$ be `uniform`$(0, \theta)$ distribution. The highest order statistics $X_{(n)} = \max\{X_1, \ldots, X_n\}$ is a complete and sufficient statistic (CSS) for this class of distributions. Can you identify the UMVUE of θ?

6. Let $F_{\boldsymbol{\theta}}$ be `normal`(μ, σ^2) and $\boldsymbol{\theta} = (\mu, \sigma^2)$. Consider the class of estimators of the form

$$S^2_{a_n} = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{a_n}, \quad a_n > 0.$$

   Find the estimator which minimizes the MSE in estimating σ^2 in this class. What is the bias associated with the best estimator?
   [*Let* $\chi = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. *You may use the fact that* $E(\chi) = (n-1)\sigma^2$ *and* $\mathrm{var}(\chi) = 2(n-1)\sigma^4$]

7. Let $F_{\boldsymbol{\theta}}$ be $\texttt{normal}(\mu, \sigma^2)$ and $\boldsymbol{\theta} = (\mu, \sigma^2)$.

   (a) Find the UMVUE of μ.

   (b) Find the UMVUE of σ^2.

8. Let $F_{\boldsymbol{\theta}}$ be $\texttt{normal}(\mu, 1)$ and $\theta = \mu$. Consider the class of linear estimators $T_{\mathbf{l}}(\mathbf{X})$ for μ. Show that minimizing the MSE of $T_{\mathbf{l}}(\mathbf{X})$ with respect to $\mathbf{l}$ does not lead to a valid estimator.

9. The waiting time (in whole minutes) for a bus is distributed as a Poisson(λ) distribution. We are interested in estimating the probability that the waiting time is at least one minute, i.e.,

$$\psi(\lambda) = P(X \geq 1) = 1 - \exp(-\lambda).$$

   To estimate this probability, we propose collecting the waiting times of n individuals, denoted as $X_1, \ldots, X_n$. Assuming $X_1, \ldots, X_n$ are independently distributed, find an unbiased estimator of $\psi(\lambda)$.

10. Suppose $n = 10$ individuals throw a biased die independently, continuing until they roll a *six*. The number of tosses for each person are as follows:

$$15 \quad 12 \quad 9 \quad 34 \quad 5 \quad 1 \quad 15 \quad 44 \quad 16 \quad 5.$$

   We are interested in estimating the probability of rolling a *six*, denoted as p. Find the UMVUE of p.

Statistics II: Introduction to Inference

Problem set 2

1. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the $N(\mu, \sigma^2)$ distribution. Find the Fisher information matrix, $\mathbf{I}_n(\boldsymbol{\theta})$, for the parameter $\boldsymbol{\theta} = (\mu, \sigma^2)$.

   [Note: The Fisher information matrix for a vector valued parameter $\boldsymbol{\theta}$ is defined as

$$\mathbf{I}_n(\boldsymbol{\theta}) = E\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\theta})\right)^{\top}\right] = -E\left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}} \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\theta})\right)\right]]$$

   Further, from the above definition of $\mathbf{I}_n(\boldsymbol{\theta})$ show that, when $X_1, \ldots, X_n$ are i.i.d., then $\mathbf{I}_n(\boldsymbol{\theta}) = n\mathbf{I}_1(\boldsymbol{\theta})$, where $\mathbf{I}_1(\boldsymbol{\theta})$ is the Fisher information matrix for one sample.

2. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the following distributions. In each case, find the method of moments estimator (MOME) for $g(\theta)$:

   (a) `Gamma`(α, β), and $g(\boldsymbol{\theta}) = (\alpha, \beta)^{\top}$.

   (b) `Beta`(α, β) and $g(\theta) = \alpha/\beta$.

   (c) `Poisson`(λ) and $g(\theta) = \exp\{-\lambda\}$.

   (d) `Location-scale Exponential`(μ, σ) with p.d.f.

$$f_{\mathbf{X}}(x; \mu, \sigma) = \begin{cases} \sigma^{-1} \exp\{-\sigma^{-1}(x - \mu)\} & x > \mu, \\ 0 & \text{otherwise,} \end{cases}$$

   and $g(\theta) = (\mu, \sigma)$.

3. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the following distributions. In each case, find the MLE for $g(\theta)$:

   (a) `Binomial`(m, θ), and $g(\theta) = \theta$.

   (b) `Binomial`(θ, p), and $g(\theta) = \theta$ when $n = 1$.

   (c) `Binomial`(m, θ), and $g(\theta) = P(X_1 + X_2 = 0)$.

   (d) `Hypergeometric`(m, r, θ) with p.m.f.

$$f_X(x; m, r, \theta) = \frac{\binom{m}{x}\binom{\theta - m}{r - x}}{\binom{\theta}{r}}, \quad \theta = m+1, m+2, \ldots; \quad \max\{0, r + m - \theta\} \le x \le \min\{m, r\},$$

   $g(\theta) = \theta$ and $n = 1$.

   (e) Double exponential: pdf $f_X(x; \theta) = 2^{-1} \exp\{-|x - \theta|\}$; with $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$.

   (f) `Uniform`(α, β), and $g(\theta) = \alpha + \beta$.

   (g) `Normal`(θ, θ^2), and $g(\theta) = \theta$.

   (h) `Inverse Gaussian`(θ_1, θ_2) and $g(\boldsymbol{\theta}) = (\theta_1, \theta_2)$.

4. Suppose that the random variables $Y_1, \dots, Y_n$ satisfy $Y_i = \beta x_i + \epsilon_i,\ i = 1, \dots, n$ where $x_1, \dots, x_n$ are fixed constants, and $\epsilon_1, \dots, \epsilon_n$ are iid $\mathrm{N}\left(0, \sigma^2\right), \sigma^2$ unknown.

   (a) Find a two-dimensional sufficient statistic for $\left(\beta, \sigma^2\right)$.

   (b) Find the MLE of β, and show that it is an unbiased estimator of β.

   (c) Find the distribution of the MLE of β.

   (d) Show that $\sum Y_i / \sum x_i$ is an unbiased estimator of β.

   (e) Calculate the exact variance of $\sum Y_i / \sum x_i$ and compare it to the variance of the MLE.

   (f) Show that $\left[\sum (Y_i/x_i)\right]/n$ is also an unbiased estimator of β.

   (g) Calculate the exact variance of $\left[\sum (Y_i/x_i)\right]/n$ and compare it to the variances of the estimators in the previous two estimates.

5. Let $W_1, \dots, W_k$ be unbiased estimators of a parameter θ with known variances $\mathrm{var}(W_i) = \sigma_i^2,\ i = 1, \dots, k$. Find the best unbiased estimator of θ of the form $\sum_{i=1}^k a_i W_i$.

6. Suppose that when the radius of a circle is measured, a random error is made, which is modeled as $N(0, \sigma^2)$. If n repeated independent measurements are made, then find an unbiased estimator of area of the circle. Is it the UMVUE?

Statistics II: Introduction to Inference

Problem set 3

1. (**Additive properties**) *Prove the following statements using moment generating functions.*

   (a) Let $X_i \overset{ind}{\sim} \texttt{binomial}(n_i, p)$, for $i = 1, \ldots, k$, then $T = \sum_{i=1}^{k} X_i$ follows $\texttt{binomial}(\sum_i n_i, p)$.

   (b) Let $X_i \overset{ind}{\sim} \texttt{Poisson}(\lambda_i)$, for $i = 1, \ldots, n$, then $T = \sum_{i=1}^{n} X_i$ follows $\texttt{Poisson}(\sum_i \lambda_i)$.

   (c) Let $X_i \overset{ind}{\sim} \texttt{normal}(\mu_i, \sigma_i^2)$, for $i = 1, \ldots, n$, then $T = \sum_{i=1}^{n} X_i$ follows $\texttt{normal}\left(\sum_i \mu_i, \sum_i \sigma_i^2\right)$.

   (d) Let $X_i \overset{ind}{\sim} \texttt{Gamma}(\alpha_i, \beta)$, for $i = 1, \ldots, n$, then $T = \sum_{i=1}^{n} X_i$ follows $\texttt{Gamma}\left(\sum_i \alpha_i, \beta\right)$.

   (e) Let $X_i \overset{ind}{\sim} \chi^2_{n_i}$, for $i = 1, \ldots, k$, then $T = \sum_{i=1}^{k} X_i$ follows χ^2_N where $N = \sum_i n_i$.

2. Let $X \sim \texttt{normal}(\mu, \sigma^2)$ distribution, then $T = aX + b \sim \texttt{normal}(a\mu + b, a^2\sigma^2)$.

3. Let $X \sim \texttt{Gamma}(\alpha, \beta)$ distribution, then $T = aX \sim \texttt{Gamma}(\alpha, \beta/a)$.

4. Let $X \sim \texttt{beta}(n/2, m/2)$ distribution, then $T = mX/\{n(1-X)\} \sim F_{n,m}$.

5. Let $X \sim \texttt{uniform}(0, 1)$ distribution, and $\alpha > 0$ then $T = X^{1/\alpha} \sim \texttt{beta}(\alpha, 1)$.

6. Let $X \sim \texttt{Cauchy}(0, 1)$ distribution, then $T = 1/(1 + X^2) \sim \texttt{beta}(0.5, 0.5)$.

7. Let $X \sim \texttt{uniform}(0, 1)$ distribution, then $T = -2 \log X \sim \chi^2_2$.

8. Let X be distributed as some absolutely continuous distribution with cdf G_X, then $T = G_X(X) \sim \texttt{uniform}(0, 1)$.

9. Let the random variable X have pdf

$$f(x) = \sqrt{\frac{2}{\pi}} \exp\{-x^2/2\}, \qquad x > 0.$$

   (a) Find $E(X)$ and $\text{var}(X)$.

   (b) Find an appropriate transformation $Y = g(X)$ and α, β, so that $Y \sim \texttt{Gamma}(\alpha, \beta)$.

10. Let X is distributed as $\texttt{Gamma}(\alpha, \beta)$ distribution, $\alpha, \beta > 0$. Then show that the r-th order population moment

$$E\left(X^r\right) = \beta^{-r} \frac{\Gamma(r + \alpha)}{\Gamma(\alpha)}, \qquad r > -\alpha.$$

11. Let the bivariate random variable (X, Y) has a joint pdf

$$f_{X,Y}(x, y) = \begin{cases} C(x + 2y) & \text{if} \quad 0 < y < 1,\ 0 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the marginal distribution of Y.

   (b) Find the conditional distribution of Y given $X = 1$.

(c) Compare the expectations of the above two distributions of Y.

(d) Find the covariance between X and Y.

(e) Find the distribution of $Z = 9/(2Y+1)^2$.

(f) What is $P(X > Y)$?

12. Let $X \sim \texttt{normal}(0,1)$. Define $Y = -X\mathbb{I}(|X| \leq 1) + X\mathbb{I}(|X| > 1)$. Find the distribution of Y.

    (Hint: Apply the CDF approach)

13. Let $X \sim \texttt{normal}(0,1)$. Define $Y = \text{sign}(X)$ and $Z = |X|$. Here $\text{sign}(\cdot)$ is a $\mathbb{R} \to \{0,1\}$ function such that $\text{sign}(a) = 1$ if $a \geq 0$, and $\text{sign}(a) = -1$ otherwise.

    (a) Find the marginal distributions of Y and Z.

    (b) Find the joint CDF of (Y, Z). Hence or otherwise prove that Y and Z are independently distributed.

14. Suppose $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \texttt{normal}\left(\mu_x, \sigma^2\right)$, $Y_1, \cdots, Y_m \overset{\text{IID}}{\sim} \texttt{normal}\left(\mu_y, \sigma^2\right)$, and all the random variables $\{X_1, \cdots, X_n, Y_1, \cdots, Y_m\}$ are mutually independent. Then find the distribution of $T := S_X^{\star 2}/S_Y^{\star 2}$, where $S_X^{\star 2}$ and $S_Y^{\star 2}$ are the unbiased sample variances of X and Y, respectively.

15. Let $X_1, \cdots, X_n$ be iid random variables with continuous CDF F_X, and suppose $E(X_1) = \mu$. Define the random variables $Y_1, \cdots, Y_n$ as follows:

$$Y_i = \begin{cases} 1 & \text{if} X_i > \mu \\ 0 & \text{otherwise.} \end{cases}$$

    (a) Find $E(Y_1)$.

    (b) Find the distribution of $\sum_{i=1}^n Y_i$.

16. Let $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \texttt{normal}\left(\mu, \sigma^2\right)$, and S_n^2 be the sample variance. Find a function of S_n^2, say $g\left(S_n^2\right)$, which satisfies $E\left[g\left(S_n^2\right)\right] = \sigma$. (Hint: You may use problem 2.)

17. [1] Let $X_1, \cdots, X_n$ be iid with pdf f_X and CDF F_X. Find the CDF of r-th order statistics $X_{(r)}$. Hence derive the pdf of $X_{(r)}$.

18. Let Y have a $\texttt{Cauchy}(0,1)$ distribution.

    (a) Find the CDF of Y.

    (b) Hence provide a method of simulating random samples from $\texttt{Cauchy}(0,1)$ distribution, starting from $\texttt{uniform}(0,1)$ random variables.

[1] You may skip this problem.

Statistics II: Introduction to Inference

Problem set 4

1. For a random variable X, the following are known

$$P(X \geq 0) = 1 \qquad \text{and} \qquad P(X \geq 10) = 1/5.$$

   Prove that $E(X) \geq 2$.

2. Suppose that X is a random variable for which $E(X) = 10$, $P(X \leq 7) = 0.2$ and $P(X \geq 13) = 0.3$. Then show that $\text{var}(X) \geq 9/2$.

3. Consider a probability distribution with CDF F, expectation μ and variance σ^2. Find the minimum size, say n, of random samples which ensures at least 0.99 probability of the event that the sample mean $\bar{X}_n$ will lie within 2σ limit of the expectation μ, i.e., $\mu - 2\sigma \leq \bar{X}_n \leq \mu + 2\sigma$.

4. Let $Z_1, Z_2, \ldots$ be a sequence of random variables, and suppose that $n = 1, 2, \ldots$, the distribution Z_n is as follows

$$P(Z_n = 0) = 1 - \frac{1}{n}, \qquad \text{and} \qquad P(Z_n = n^2) = \frac{1}{n}, \qquad \text{for} \quad n = 1, 2, \ldots.$$

   Show that

$$\lim_{n\to\infty} E(Z_n) = \infty, \quad \text{and} \quad Z_n \xrightarrow{p} 0, \quad \text{i.e., for any } \epsilon > 0, \quad P(|Z_n| > \epsilon) \to 0, \quad \text{as} \quad n \to \infty.$$

   [Note: This example shows that the sufficient condition for consistency of a sequence of estimators is not *necessary* (i.e., the converse is not true).]

5. Suppose that 75% of the people in a certain metropolitan area live in the city and 25% of the people live in the suburbs. If 1200 people attending a certain concert represent a random sample from the metropolitan area, what is the probability that the number of people from the suburbs attending the concert will be fewer than 270?

6. Suppose that a random sample of size n is to be taken from a distribution with mean is μ and the standard deviation is $\sigma = 3$. Use the central limit theorem to determine approximately the smallest value of n for which the following relation will be satisfied:

$$P\left(|\bar{X}_n - \mu| < 0.3\right) \geq 0.95.$$

7. Suppose that the proportion of defective items in a large manufactured lot is 0.1. What is the smallest random sample of items that must be taken from the lot in order for the probability to be at least 0.99 that the proportion of defective items in the sample will be less than 0.13?

8. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{uniform}(0, \theta)$ distribution and $T_n = X_{(n)}$ be the maximum order statistic. Show that $Z_n = n(T_n - \theta) \xrightarrow{d} Z$ where Z has the CDF F_Z

$$F_Z = \begin{cases} \exp\{z/\theta\} & \text{if} \quad z < 0, \\ 1 & \text{if} \quad z \geq 0. \end{cases}$$

9. Consider the random variable with X with the following specifications $E(X) = \mu$ and the r-th order central moments μ_r are $\mu_2 = 5/4, \;\; \mu_4 = 125/2$. What is the best possible upperbound of the probability of the event that $\bar{X}_n \in [\mu - 1, \mu + 1]$ which can be obtained using a random sample of size $n = 20$.

10. Let $X_1, \ldots, X_n$ be a random sample from some distribution with expectation μ, variance σ^2 and finite forth order raw moment μ_4'. Determine the sequence of real numbers $\{a_n\}$ and the random variable Z such that the sequence of random variables $Z_n = \sum_{i=1}^n X_i^2/\sqrt{n} + \bar{X}_n^2$ satisfies $Z_n - a_n \xrightarrow{d} Z$. What is the distribution of Z?

Statistics II: Introduction to Inference

Problem set 5

1. Consider the testing problem $H_0 : X \sim f_0$ against $H_1 : X \sim f_1$, where

x :	-4	-3	0	1	2	5
f_0 :	0.05	0.20	0.30	0.15	0.25	0.05
f_1 :	0.15	0.30	0.05	0.05	0.25	0.20

   Using Neyman Pearson (NP) lemma find the most powerful (MP) test at level 0.05 and level 0.075 for testing H_0 and H_1.

2. Let $\phi^\star$ be an MP size α test for $H_0 : \mathbf{X} \sim f_0(\mathbf{x})$ against $H_1 : \mathbf{X} \sim f_1(\mathbf{x})$, and let $\beta^\star$ be the power of $\phi^\star$, $0 < \beta^\star < 1$.

   (a) Show that $\phi^{\star\star} = (1 - \phi^\star)$ is an MP test for testing $H_0 : \mathbf{X} \sim f_1(\mathbf{x})$ against $H_1 : \mathbf{X} \sim f_0(\mathbf{x})$ at level $(1 - \beta^\star)$.

   (b) Further, show that if $\phi^\star$ is an unbiased test, then $\phi^{\star\star}$ is also unbiased.

3. Let ϕ_1 and ϕ_2 be 2 size-α tests for testing $H_0 : \theta \in \Theta_0$ against $H_A : \theta \in \Theta_A$ and $\phi^\star$ is a convex combination of ϕ_1 and ϕ_2. Show that $\phi^\star$ is a level-α test. What can say about the power function of $\phi^\star$?

4. The lifetime of equipment is normally distributed with mean θ and standard distribution 5. For testing the null hypothesis $H_0 : \theta \leq 30$ against the alternative hypothesis $H_A : \theta > 30$, a random sample of size n is chosen. Determine n and the cutoff c such that the test

$$\phi(\mathbf{x}) = 1, \text{ if } \bar{x} \geq c, \qquad \phi(\mathbf{x}) = 0, \text{ if } \bar{x} < c$$

   has power function values 0.1 and 0.9 at the points $\theta = 30$ and $\theta = 35$ respectively. Draw the power function of the resultant test.

5. Let X be distributed as $U(0, \theta)$ and $X_{(n)}$ denote the largest order statistic based on a random sample of size n from this distribution. We reject $H_0 : \theta = 1$ and accept $H_1 : \theta \neq 1$ if either $x_{(n)} \leq 1/2$ or $x_{(n)} \geq 1$. Find the power function of the test.

6. Based on a random sample $X_1, \ldots, X_n$, derive the MP size-α test for testing $H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1 (> \theta_0)$ for the population with the following pdf

$$f_X(x;\theta) = (\sqrt{2\pi}\theta)^{-1} e^{-x^2/2\theta^2}; -\infty < x < \infty;\ \theta > 0, \quad \text{and} \quad f_X(x;\theta) = 0, \quad \text{otherwise.}$$

   Will the MP test obtained above be UMP for testing $H_0 : \theta \leq \theta_0$, against $H_1 : \theta > \theta_0$.

   [Hint: To generalize to $H_0 : \theta \leq \theta_0$ you need to show that the power function of the MP test obtained above is of the form

$$\beta_\phi(\theta) = P\left(T_n \geq K_0^2/\theta^2 \mid T_n \sim \chi^2_n\right),$$

   for some fixed value of K_0. Then from the monotonicity of the CDF you can show that $\beta_\phi(\theta)$ is an increasing function of θ. This in turn will lead to the fact that the size of the generalized test is $\beta_\phi(\theta_0)$.]

7. Let X be distributed as $f_X(x;\theta) = \theta x^{\theta-1},\ 0 < x < 1, \theta > 0,$ and $f_X(x;\theta) = 0,$ otherwise, $\theta > 0.$ To test $H_0 : \theta = \theta_1$ against $H_A : \theta > \theta_1$ based on a sample of size n , the following critical region is proposed $\mathbb{C} = \{\mathbf{x} : \prod_{i=1}^{n} x_i \geq 0.5\}$. Find the power function of the above test.

   [Hint: It might be helpful to consider the transformation $Y = -\log X$.]

8. Let X be an observation in $(0, 1)$. Find an

   MP level-α test of $H_0 : X \sim f_0$ against $H_A : X \sim \texttt{Uniform}(0, 1)$, where, $f_0(x) = 4x$ if $0 < x < \frac{1}{2}$, or $f_0(x) = 4(1 - x)$ if $\frac{1}{2} \leq x < 1$.

9. Suppose that $X_1, \ldots, X_n$ are iid with a common pdf $f(x)$, which takes one of the following forms:

$$f_0(x) = \begin{cases} 3x^2/64 & 0 < x < 4 \\ 0 & \text{otherwise,} \end{cases} \qquad f_1(x) = \begin{cases} 3\sqrt{x}/16 & 0 < x < 4 \\ 0 & \text{otherwise.} \end{cases}$$

   Find the MP level-α test for testing $H_0 : X \sim f_0(x)$ against $H_A : X \sim f_1(x)$.

10. Based on a random sample $X_1, \ldots, X_n$, derive the UMP size- α test for testing $H_0 : \theta \leq \theta_0$ against $H_A : \theta > \theta_0$ for the location exponential population with pdf $f_X(x;\theta) = \theta^{-1} \exp\{-(x - \mu)/\theta\}$; with $x > \mu$, $\mu \in \mathbb{R}$ and $\theta > 0$, when μ is known.

    [Hint: First consider the MP test for testing $H_0 : \theta = \theta_0$ against $H_0 : \theta = \theta_1 (> \theta_0)$, and then generalize. Towards that, show that the power function of the MP test is

$$\beta_\phi(\theta) = P\left(T_n > k_0/\theta \mid T_n \sim \texttt{Gamma}(n, 1)\right).$$

    Hence show that $\beta_\phi(\theta)$ is an increasing function of θ.]

11. Let X be an observation from $\texttt{Poisson}(\theta)$. Find an UMP level-α test for testing $H_0 : \theta \leq \theta_0$ against $H_A : \theta > \theta_0$.

12. Suppose $X_1, \cdots, X_m$ be a random sample of size m from $B(n, p)$, find the UMP level- α test for testing $H_0 : p \leq p_0$ against $H_A : p > p_0$.

Statistics II: Introduction to Inference

Problem set 6

1. Show that the following distributions belong to either location, or scale, or location scale families. Hence find the appropriate transformation of a random variable belonging to these distributions, such that the transformed random variable has a distribution free of the parameters θ.

   (a) $\texttt{normal}(\mu, \sigma^2)$,

   (b) $\texttt{exponential}(\lambda)$,

   (c) $\texttt{uniform}(-\theta, \theta)$,

   (d) $\texttt{exponential}(\lambda)$,

   (e) $\texttt{Gamma}(n, \theta)$, with n known,

   (f) $\texttt{normal}(\theta, \theta^2)$,

   (g) $\texttt{uniform}(\theta - 1/2, \theta + 1/2)$.

2. (a) Let $X_1, \ldots, X_n$ be a random sample of size n from a location family of distribution with location parameter θ. Show that the distribution of any function of $Y_{i,j} = \{X_i - X_j\}$, $i, j = 1, \ldots, n$ is free of θ.

   (b) Let $X_1, \ldots, X_n$ be a random sample of size n from a scale family of distribution with scale parameter θ. Show that the distribution of any function of $Y = X_1^2 / \sum_{j=1} X_j^2$ and $Z = X_{(1)}/X_{(n)}$ are free of θ.

3. The independent random variables $X_1, \ldots, X_n$ have common distribution

$$P(X_i \le x) = \begin{cases} 0 & \text{if} \quad x \le 0 \\ (x/\beta)^\gamma & \text{if} \quad 0 < x < \beta \\ 1 & \text{if} \quad x \ge \beta \end{cases}.$$

   Find a $(1 - \alpha)100\%$ confidence interval for β based on the MLE of β, when γ is known.

4. Find a $(1 - \alpha)$ confidence interval for θ, given $X_1, \ldots, X_n$ iid with pdf

   (a) $f_X(x; \theta) = 1$, $\theta - 0.5 \le x \le \theta + 0.5$.
   [Instead of the exact confidence interval, you may find a approximate confidence interval based on the asymptotic distribution of an appropriate pivot.]

   (b) $f_X(x; \theta) = 2x\theta^{-2}$, $0 < x < \theta$, $\theta > 0$.

5. Let $X_1, \cdots, X_n$ be iid from $\texttt{Uniform}(\theta, 1)$ distribution, $\theta < 1$.

   (a) Obtain a suitable pivot for finding a confidence interval for θ.

   (b) Based on this test find a $(1 - \alpha)$ confidence set.

6. Let T be a statistic with continuous strictly decreasing CDF $F_T(\cdot; \theta)$, and α_1, α_2 be such that $\alpha_1 + \alpha_2 = \alpha$, for some fixed $\alpha \in (0, 1)$. Suppose that for each t the functions $L(t)$ and $U(t)$ are defined as

$$F_T(T; U(T)) = \alpha_1, \quad \text{and} \quad F_T(T; L(T)) = 1 - \alpha_2.$$

   Then show that the random interval $[L(t), U(t)]$ is a $(1 - \alpha)$ confidence interval for θ.

7. Let X be a single observation from a $\texttt{beta}(\theta, 1)$ distribution. Let $Y = -(\log X)^{-1}$. Evaluate the confidence coefficient, β, of the interval $[y/2, y]$.

8. A confidence interval $[L(\mathbf{X}), U(\mathbf{X})]$ for the parameter θ with confidence coefficient at least $1 - \alpha$ is called *unbiased* if $P_\theta\left(L(\mathbf{X}) < \theta < U(\mathbf{X})\right) \geq 1 - \alpha$, and $P_\theta\left(L(\mathbf{X}) < \theta' < U(\mathbf{X})\right) \leq 1 - \alpha$ for all $\theta' \neq \theta$.

   Based on a random sample of size n from $\texttt{uniform}(0, \theta)$, show that the symmetric confidence interval obtained from the pivot $X_{(n)}/\theta$ is unbiased for sufficiently large n.

9. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_0, \sigma^2)$ where μ_0 is known.

   (a) Find the UMVUE of σ^2.

   (b) Using the UMVUE, find an appropriate pivot for σ^2, and its distribution.

   (c) Using this pivot find a $(1 - \alpha)$-confidence interval for σ^2.

10. Let $X_1, \ldots, X_n \overset{iid}{\sim} \texttt{exponential}(\theta)$.

   (a) Find an appropriate transformation $W_i = T(X_i, \theta)$ such that the distribution of W_i does not depend on $1/\theta$.

   (b) Consider two pivots, one based on W_1 only, and another involving the UMVUE of $1/\theta$.

   (c) Find the symmetric $(1 - \alpha)$-confidence intervals using the two pivots obtained in part (b).

   (d) Generate $n = 10$ IID samples from $\texttt{exponential}(5)$ distribution. Based on the samples obtain realizations of the two confidence intervals for $\alpha = 0.1, 0.05, 0.025$.

Statistics II: Introduction to Inference

Problem set 7

1. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$.

   (a) Consider testing the hypotheses

   $$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0.$$

   Suppose we want the power of the MP test of size α to be at least $(1-\beta)$, for some fixed $0 < \beta < 1$ when $\mu = \mu_1$ $(\mu_1 > \mu_0)$. Determine the minimum sample size n required to achieve this power level.

   (b) Find the confidence interval for μ based on the sample. Then, determine the minimum sample size n such that the length of the confidence interval for μ is at most l.

2. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$, where σ_1^2, σ_2^2 known.

   (a) Consider testing the hypotheses

   $$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 > \mu_2.$$

   Suppose we want the power of the MP test of size α to be at least $(1-\beta)$, for some fixed $0 < \beta < 1$ when $\mu_1 - \mu_2 = \delta$ $(\delta > 0)$. Determine the minimum sample size n required to achieve this power level.

   (b) Construct a confidence interval for the difference $\mu_1 - \mu_2$. Determine the minimum sample size n such that length of the confidence interval does not exceed a specified value l.

3. Let $X_1, \ldots, X_n \overset{iid}{\sim} \texttt{Uniform}(0, \theta)$.

   (a) Consider testing the hypotheses

   $$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$

   Find the minimum sample size n such that the most powerful (MP) test of size α has power at least $1 - \beta$ for some fixed $0 < \beta < 1$ when $\theta = \theta_1$ $(\theta_1 > \theta_0)$.

   (b) Construct a confidence interval (L, U) for θ. Determine the minimum sample size n such that the ratio of the upper bound (U) to the lower bound (L) of the confidence interval does not exceed a specified value l.

4. Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Construct a confidence interval for σ^2. Determine the minimum sample size n (numerically) such that the ratio of the upper bound and lower bound of the confidence interval does not exceed a specified value $l = 1.5$. The following table provides critical values of the chi-squared distribution for various sample sizes:

Sample Size n	187	188	189	190	191	192	193
$\chi^2_{n-1;\,1-\alpha/2}$	150.126	151.024	151.923	152.822	153.721	154.621	155.521
$\chi^2_{n-1;\,\alpha/2}$	225.660	226.761	227.863	228.964	230.064	231.165	232.265

Here, $\chi^2_{n-1;\,\alpha/2}$ and $\chi^2_{n-1;\,1-\alpha/2}$ denote the upper and lower critical values, respectively, of the chi-squared distribution with $n-1$ degrees of freedom, respectively.

6 Exercises

1. Suppose that the proportion p of defective items in a large population of items is unknown, and that it is desired to test the following hypotheses:

$$H_0 : p = 0.2 \quad \text{vs} \quad H_1 : p \neq 0.2.$$

Suppose also that a random sample of 20 items is drawn from the population. Let Y denote the number of defective items in the sample, and consider a test procedure ϕ such that the critical region contains all the outcomes for which either $Y \geq 7$ or $Y \leq 1$.

   (a) Determine the value of the power function $\beta_\phi(p)$ at the points $p = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1; and sketch the power function.

   (b) Determine the size of the test.

2. An auto manufacturer gives a warranty for 3 years for its new vehicles. In a random sample of 60 of its vehicles, 20 of them needed five or more major repairs within the warranty period. Estimate the 95% (large sample) confidence interval of the true proportion of vehicles from this manufacturer that need five or more major repairs during the warranty period, with confidence coefficient 0.95. Interpret the result.

3. Suppose that a random sample of 10,000 observations is taken from the normal distribution with unknown mean μ and known variance is 1, and it is desired to test the following hypotheses at the level of significance 0.05:

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0.$$

Suppose also that the test procedure specifies rejecting H_0 when $|\bar{X}_n| \geq c$, where the constant c is chosen so that $P(|\bar{X}_n| \geq c \mid \mu = 0) = 0.05$. Find the probability that the test will reject H_0 if (a) the actual value of μ is 0.01, and (b) the actual value of μ is 0.02.

4. A random sample of size 50 from a particular brand of tea packets produced a mean weight of 15.65 ounces and standard deviation of 0.59 ounce. Assume that the weights of these brands of tea packets are normally distributed. Find a 95% confidence interval for the true mean μ.

5. Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 73.3 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.2$. Construct a 90% confidence interval for the true mean speed μ, of the vehicles on this highway. Interpret the result.

6. Studies have shown that the risk of developing coronary disease increases with the level of obesity, or accumulation of body fat. A study was conducted on the effect of exercise on losing weight. Fifty men who exercised lost an average of 11.4 lb, with a standard deviation of 4.5 lb. Construct a 95% confidence interval for the mean weight loss through exercise. Interpret the result and state any assumptions you have made.

7. Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students. A random sample of 50 scores from each section produced the following results:

   (a) Section I: $\bar{x}_1 = 77.01, s_1 = 10.32$

   (b) Section II: $\bar{x}_2 = 72.22, s_2 = 11.02$

Calculate 95% confidence intervals for each of these three samples.

8. Suppose that $X_1, \ldots, X_m$ form a random sample from the normal distribution with unknown mean μ_1 and unknown variance σ_1^2, and that $Y_1, \ldots, Y_n$ form an independent random sample from the normal distribution with unknown mean μ_2 and unknown variance σ_2^2. Suppose also that it is desired to test the following hypotheses with the usual F-test at the level of significance $\alpha = 0.05$:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 > \sigma_2^2.$$

Assuming that $m = 16$ and $n = 21$, show that the power of the test when $\sigma_1^2 = 2\sigma_2^2$ is given by $P(V \geq 1.1)$, where V is a random variable having the F-distribution with 15 and 20 degrees of freedom.

9. The scores of a random sample of 16 people who took the TOEFL (Test of English as a Foreign Language) had a mean of 540 and a standard deviation of 50. Construct a 95% confidence interval for the population mean μ of the TOEFL score, assuming that the scores are normally distributed.

10. The following data represent the rates (micrometers per hour) at which a razor cut made in the skin of anesthetized newts is closed by new cells.

$$28, 20, 21, 39, 32, 23, 18, 31, 14, 23, 18, 22, 28, 24, 33, 12, 23, 21, 25, \text{ and } 25$$

   (a) Find the 95% confidence interval for population mean rate μ for the new cells to close a razor cut made in the skin of anesthetized newts.

   (b) Find a 99% confidence interval for μ. Is the 95% CI wider or narrower than the 99% CI? Briefly explain why.

   (c) Find the 95% confidence interval for population variance σ^2.

11. A study of two kinds of machine failures shows that 58 failures of the first kind took on the average 79.7 minutes to repair with a standard deviation of 18.4 minutes, whereas 71 failures of the second kind took on average 87.3 minutes to repair with a standard deviation of 19.5 minutes. Find a 99% confidence interval for the difference between the true average amounts of time it takes to repair failures of the two kinds of machines.

12. The management of a supermarket wanted to study the spending habits of its male and female customers. A random sample of 16 male customers who shopped at this supermarket showed that they spent an average of \$55 with a standard deviation of \$12. Another random sample of 25 female customers showed that they spent \$85 with a standard deviation of \$20.50. Assuming that the amounts spent at this supermarket by all its male and female customers were approximately normally distributed, construct a 90% confidence interval for the ratio of variance in spending for males and females, σ_1^2/σ_2^2.

13. The following information was obtained from two independent samples selected from two normally distributed populations with unknown but equal variances.

   - Sample I: 14, 15, 12, 13, 6, 14, 11, 12, 17, 19, 23.
   - Sample II: 16, 18, 12, 20, 15, 19, 15, 22, 20, 18, 23, 12, 20.

   Test whether the difference of the population means is equal to zero or not. Construct a 95% confidence interval for the difference between the population means and interpret.