1. A sample of size n is drawn without replacement from an urn containing N balls, m of which are red and $N - m$ are black; the balls are otherwise indistinguishable. Let X denote the number of red balls in the sample of size n. In what follows we treat N, n as known and m as unknown.

   (a) Find $P_m(X = x)$.

   (b) Show that
   $$\widehat{m} = \min\{\lfloor X(N+1)/n \rfloor, N\}. \tag{1}$$
   is an MLE of m.

   (c) Define
   $$\underline{x}_{m,\alpha} = \max\{x \in \mathbb{Z} : P_m(X \le x) \le \alpha\}$$
   $$\overline{x}_{m,\alpha} = \min\{x \in \mathbb{Z} : P_m(X \le x) \ge \alpha\}.$$
   Show that
   $$\{m : \underline{x}_{m,\alpha/2} < X \le \overline{x}_{m,1-\alpha/2}\} \tag{2}$$
   is a $100(1-\alpha)\%$ confidence interval for m, possibly conservative. *Hint:* Invert a hypothesis test of $H_0 : m = m_0$ vs. $H_1 : m \ne m_0$, and note that one of the inequalities in (2) is strict.

2. (a) Let $S_1 \sim \text{Bin}(n_1, p)$ and $S_2 \sim \text{Bin}(n_2, p)$ be two independent binomial random variables, and let $S = S_1 + S_2$. Identify the distribution of S_1 conditional on $S = s$, and give its parameter values in terms of an urn model.

   (b) Now let $S_1 \sim \text{Bin}(n_1, p_1)$ and $S_2 \sim \text{Bin}(n_2, p_2)$ be independent, and $S = S_1 + S_2$ as above. Fisher's Exact Test of $H_0 : p_1 = p_2$ versus $H_1 : p_1 > p_2$ rejects H_0 when S_1 is large.

      i. Show that, under H_0, S is a sufficient statistic.

      ii. Write down an expression for the p-value of Fisher's Exact Test, conditional on $S = s$, in terms of the density $f(s_1|s)$ of the distribution in part 2a.

1. Let $X_1, \ldots, X_n$ be a random sample from a distribution with variance $\mathrm{Var}(X_1) = \sigma^2 < \infty$, and let $T_n = T_n(X_1, \ldots, X_n)$ be some statistic.

   (a) Write down an expression for the jackknife estimator V_n of $\mathrm{Var}(T_n)$ in terms of

   $$T_{n-1,i} = T_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n), \quad i = 1, \ldots, n.$$

   (b) Now let $T_n = \overline{X}_n = n^{-1} \sum_{i=1}^n X_i$ be the sample mean. Show that:

      i. $\mathrm{Var}(T_n) = \sigma^2/n$

      ii.
      $$W_n = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \overline{X}_n)^2$$
      is an unbiased estimator of $\mathrm{Var}(T_n)$

      iii. $V_n = W_n$

2. Given

   - an index set S;
   - a distribution $\pi = (\pi_i)$ on S, and $\pi_i > 0$ for all $i \in S$;
   - a Markov chain on S with transition matrix $\mathbf{Q} = (q_{ij})$, where $q_{ij} > 0$ for all $i \neq j$ (the reference chain).

   We construct a new Markov chain whose transition matrix $\mathbf{P} = (p_{ij})$ is given by

   $$p_{ij} = q_{ij} \frac{\pi_j \, q_{ji}}{\pi_i \, q_{ij} + \pi_j \, q_{ji}}$$

   (a) Show that

      i. This new chain is reversible.

      ii. The stationary distribution of this chain is π.

   (b) Sketch an algorithm that generates random samples whose marginal distribution is π.

1. (a) Let $f_\theta(x)$, $\theta \in \Theta \subseteq \mathbb{R}$, be a family of density functions with respect to some common measure. If we say that this family has the monotone likelihood ratio (MLR) property in the real-valued statistic $T = T(x)$, what two properties must hold?

   (b) Taking $\Theta = (0, \infty)$, let $f_\theta(x)$, $x = (x_1, \ldots, x_n)$, be the joint density of a random sample of n i.i.d. uniform $(0, \theta)$ observations:

$$f_\theta(x) = \begin{cases} \theta^{-n}, & \text{if } x_i < \theta \text{ for all } i = 1, \ldots, n \\ 0, & \text{otherwise.} \end{cases}$$

   Show that this family has the MLR property, and give the statistic T.

   (c) Given $\alpha \in (0, 1)$ and $\theta_0 > 0$, find a uniformly most powerful level-α test of

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

   in terms of $T(X)$. Find any critical values and randomization constants explicitly.

2. Recall that a log-normal distribution $\ln \mathcal{N}(x|\mu, \sigma^2)$ is a continuous probability distribution of a random variable whose logarithm is normally distributed $\mathcal{N}(x|\mu, \sigma^2)$. That is, if $X \sim \ln \mathcal{N}(x|\mu, \sigma^2)$, then $\log X \sim \mathcal{N}(x|\mu, \sigma^2)$. Suppose the only random number generator that you have is the one for log-normal distributions $\ln \mathcal{N}(x|\mu, \sigma^2)$. Propose an MCMC algorithm for estimating the following integral

$$I = \int_0^\infty e^{-x^4 - x^6 - x^8} \frac{e^x}{\alpha} dx,$$

   where $\alpha = \int_0^\infty e^{-x^4 - x^6 - x^8} dx$ (is unknown). Describe the algorithm in detail.

Spring 2014 Math 541b Exam

1. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed samples from the Normal distribution $\mathcal{N}(\theta, \sigma^2)$ having mean θ and variance σ^2.

   (a) Does a Uniformly Most Powerful, or UMP, level α test of $H_0 : \sigma^2 \leq 1$ versus $H_1 : \sigma^2 > 1$ exist if the mean θ is known? If so, find the form of the rejection region of the UMP test, and if not, explain why not.

   (b) Does a UMP level α test of $H_0 : \sigma^2 \leq 1$ versus $H_1 : \sigma^2 > 1$ exist, if both θ and σ^2 are unknown, with the restriction $\theta/\sigma^2 = 2$?

2. Consider a vector $\mathbf{X} = (X_1, X_2, X_3)$ of counts with distribution given by the multinomial distribution with probabilities

$$P(\mathbf{X} = \mathbf{x}) = \binom{n}{x_1, x_2, x_3} \prod_{i=1}^{3} p_i^{x_i}$$

   for $\mathbf{x} = (x_1, x_2, x_3)$, a vector of non-negative integers summing to n, and

$$(p_1, p_2, p_3) = \left(\frac{1}{3} + \frac{\theta}{3}, \frac{2\theta}{3}, \frac{2}{3} - \theta \right) \quad \text{for some } \theta \in (0, 1).$$

   (a) Write out the equation that would need to be solved in order to obtain the maximum likelihood estimate of θ.

   (b) Show that if additional 'missing data' is now introduced to form a 'full model' that a simpler equation then that in part (a) results, and solve it explicitly. Hint: Consider the first cell.

   (c) Specify the steps of an EM algorithm that takes advantage of the simplification obtained by treating the situation as a missing data problem as in part (b).

Fall 2014 Math 541b Exam

1. (a) Let $q_{x,y}$ be a Markov transition function, and π_x a probability distribution on a finite state space S. Show that the Markov chain that accepts moves made according to $q_{x,y}$ with probability

$$p_{x,y} = \min\left\{\frac{\pi_y q_{y,x}}{\pi_x q_{x,y}}, 1\right\},$$

and otherwise remains at x, has stationary distribution π_x. Show that if $q_{x,y}$ and π_x are positive for all $x, y \in S$ then the chain so described has unique stationary distribution π_x.

(b) Let $f(y)$ and $g(y)$ be two probability mass functions, both positive on $\mathbb{R}$. With X_1 generated according to g, consider the Markov chain $X_1, X_2, \ldots$ that for at stage $n \geq 1$ generates an independent observation Y_n from density g, and accepts this value as the new state X_{n+1} with probability

$$\min\left\{\frac{f(Y_n)g(X_n)}{f(X_n)g(Y_n)}, 1\right\}$$

and otherwise sets X_{n+1} to be X_n. Prove that the chain converges in distribution to a random variable with distribution f.

(c) The accept/reject method. Let f and g be density functions on $\mathbb{R}$ such that the support of f is a subset of the support of g, and suppose that there exists a constant M such that $f(x) \leq Mg(x)$. Consider the procedure that generates a random variable with distribution g, an independent random variable with the uniform distribution U on $[0,1]$ and sets $Y = X$ when $U \leq f(X)/Mg(X)$. Show that Y has density f.

2. Let f be a real valued function on $\mathbb{R}^n$, and $Z = f(X_1, \ldots, X_n)$ for $X_1, \ldots, X_n$ independent random variables.

(a) With $E^{(i)}(\cdot) = E(\cdot|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ show the following version of the Efron-Stein inequality

$$\text{Var}(Z) \leq E\left(\sum_{i=1}^{n}(Z - E^{(i)}Z)^2\right). \tag{1}$$

Hint: With $E_i(\cdot) = E(\cdot|X_1, \ldots, X_i)$, show that

$$Z - EZ = \sum_{i=1}^{n} \Delta_i \quad \text{where} \quad \Delta_i = E_i Z - E_{i-1} Z,$$

compute the variance of Z in this form, use properties of conditional expectation such as $E_i(E^{(i)}(\cdot)) = E_{i-1}(\cdot)$, and (conditional) Jensens' inequality.

(b) Letting $(X_1', \ldots, X_n')$ be an independent copy of $(X_1, \ldots, X_n)$, with

$$Z_i' = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n),$$

show that

$$\text{Var}(Z) \leq \frac{1}{2} E\left(\sum_{i=1}^{n} (Z - Z_i')^2\right).$$

Hint: Express the right hand side of (1) in terms of conditional variances, and justify and use the conditional version of the fact that if X and Y are independent and have the same distribution then the variance of X can be expresses in terms of $E(X - Y)^2$.

Spring 2015 Math 541b Exam

1. Let $X_1, X_2, \ldots, X_n$ be independent Cauchy random variable with density
$$f(x|\theta) = \frac{1}{\pi(1 + (x-\theta)^2)},$$
and let $\widetilde{X_n}$ = median of $\{X_1, X_2, \ldots, X_n\}$.

   (a) Prove that $\sqrt{n}(\widetilde{X_n} - \theta)$ is asymptotically normal with mean 0 and variance $\pi^2/4$ by showing that as n tends to infinity,
$$P(\sqrt{n}(\widetilde{X_n} - \theta) \leq a) \longrightarrow P(Z \geq -2a/\pi)$$
where Z is a standard normal random variable. *Hint:* If we define Bernoulli random variables $Y_i = 1_{\{X_i \leq \theta + a/\sqrt{n}\}}$, the event $\{\widetilde{X_n} \leq \theta + a/\sqrt{n}\}$ is equivalent to $\{\sum_i Y_i \geq (n+1)/2\}$ when n is odd. Applying the CLT might also be needed.

   (b) Using the result from part (a), find an *approximate* α-level large sample test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

2. We observe independent Bernoulli variables $X_1, X_2, \ldots, X_n$, which depend on unobservable variables $Z_1, \ldots, Z_n$ which, given $\theta_1, \ldots, \theta_n$, are distributed independently as $N(\theta_i, 1)$, where
$$X_i = \begin{cases} 0 & \text{if } Z_i \leq u \\ 1 & \text{if } Z_i > u. \end{cases}$$
The values $\theta_1, \theta_2, \ldots, \theta_n$ are distributed independently as $N(\xi, \sigma^2)$. Assuming that u and σ^2 are known, we are interested in the maximum likelihood estimate of ξ.

   (a) Show that for any for given values of ξ and σ^2, and all $i = 1, \ldots, n$, the random variable Z_i is normally distributed with mean ξ and variance $\sigma^2 + 1$.

   (b) Write down the likelihood function for the complete data $Z_1, \ldots, Z_n$ when these values are observed.

(c) Now assume that only $X_1, \ldots, X_n$ are observed, and show that the EM sequence for the estimation of the unknown ξ is given by

$$\xi^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} E(Z_i | X_i, \xi^{(t)}, \sigma^2).$$

Start by computing the expected log likelihood of the complete data.

(d) Show that

$$E(Z_i | X_i, \xi^{(t)}, \sigma^2) = \xi^{(t)} + \sqrt{\sigma^2 + 1} \cdot H_i \left(\frac{u - \xi^{(t)}}{\sqrt{\sigma^2 + 1}} \right)$$

where

$$H_i(t) = \begin{cases} \frac{\phi(t)}{1 - \Phi(t)} & \text{if } X_i = 1 \\ -\frac{\phi(t)}{\Phi(t)} & \text{if } X_i = 0, \end{cases}$$

and $\Phi(t)$ and $\phi(t)$ are cumulative distribution and density function of a standard normal variable, respectively.

1. Let $X_1, \ldots, X_n$ be a sample from distribution F, let $X_{(1)} \leq \ldots \leq X_{(n)}$ be the corresponding order statistics, and let θ and $\tilde{\theta}$ be the population and sample median, respectively. Assume that the sample size is 3 ($n = 3$),

   (a) Find the distribution of the ordered bootstrap sample $(X^*_{(1)}, X^*_{(2)}, X^*_{(3)})$, where X^*_i's are randomly selected from the sample with replacement.

   (b) Determine the bootstrap estimator $\widehat{\lambda_1}$ of the bias of sample median, $\lambda_1 = E(\tilde{\theta}) - \theta$.

   (c) Determine the bootstrap estimator $\widehat{\lambda_2}$ of the variance of sample median, $\lambda_2 = Var(\tilde{\theta})$.

2. Denote $\mathbf{z} \in \mathbb{R}^2$ by $\mathbf{z} = (x, y)$, and let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent with distribution $\mathcal{N}(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \text{for } \rho \in (-1, 1), \text{ unknown.}$$

   a. Write down the $\mathcal{N}(0, \Sigma)$ density function, and the likelihood

$$L(\rho) = f(\mathbf{x}_1, \ldots, \mathbf{x}_n; \rho)$$

   of the sample.

   b. Determine the Neyman Pearson procedure for testing $H_0 : \rho = 0$ versus $H_1 : \rho = \rho_0$ at level $\alpha \in (0, 1)$ for some $\rho_0 \neq 0$ in $(0, 1)$. (You do not need to explicitly write down any null distributions arising.)

   c. Determine if the test in b) is uniformly most powerful for testing $H_0 : \rho = 0$ versus $H_1 : \rho > 0$, and justify your conclusion.

1. Let $X_1, \dots, X_n$ be i.i.d. from a normal distribution with unknown mean μ and variance 1. Suppose that negative values of X_i are truncated at 0, so that instead of X_i, we actually observe

$$Y_i = \max(0, X_i), \qquad i = 1, 2, \dots, n,$$

from which we would like to estimate μ. By reordering, assume that $Y_1, \dots, Y_m > 0$ and $Y_{m+1} = \dots = Y_n = 0$.

   (a) Explain how to use the EM algorithm to estimate μ from $Y_1, \dots, Y_n$. Specifically, give the details about E-step and M-step. Show that a recursive formula for the successive EM estimates $\mu^{(k+1)}$ is

$$\mu^{(k+1)} = \frac{1}{n}\sum_{i=1}^{m} Y_i + \frac{n-m}{m}\mu^{(k)} - \frac{n-m}{m}\frac{\phi(\mu^{(k)})}{\Phi(-\mu^{(k)})},$$

   where $\phi(x)$ is probability density function and $\Phi(x)$ is cumulative density function of the standard normal distribution.

   (b) Find the log-likelihood function $\log L(\mu)$ based only on observed data, and use it to write down a (nonlinear) equation which the MLE $\widehat{\mu}$ satisfies.

   (c) Use the equation in part (b) to verify that $\widehat{\mu}$ is indeed a fixed point of the recursion found in (a).

   (d) Prove that $\mu^{(k)} \longrightarrow \widehat{\mu}$ for any starting point $\mu^{(0)}$, providing at least one of the observations is not truncated. To do this, prove that the difference between $\mu^{(k)}$ and $\widehat{\mu}$ gets smaller as k gets larger. *Hint:* The Mean Value Theorem and the following inequalities, which you can use without proof, might be useful.

$$0 < \frac{\phi(x)[\phi(x) - x\Phi(-x)]}{\Phi^2(-x)} < 1, \qquad \text{for all } x.$$

   Note: The Mean Value Theorem says that if f is continuous and differentiable on the interval (a, b), then there is a number c in (a, b) such that $f(b) - f(a) = f'(c)(b - a)$.

2. Let $X_1, \dots, X_n$ be iid Unif$(0, \theta)$, where $\theta > 0$ is unknown.

   (a) Find the MLE $\widehat{\theta}$, its c.d.f. $F_\theta(u) = P_\theta(\widehat{\theta} \le u)$, and its expected value $E_\theta(\widehat{\theta})$.

   (b) Consider a confidence interval for θ of the form

$$[a\widehat{\theta}, b\widehat{\theta}], \quad \text{where } 1 \le a \le b \text{ are constants.} \tag{1}$$

   For given $0 < \alpha < 1$, characterize all $1 \le a \le b$ making $[a\widehat{\theta}, b\widehat{\theta}]$ a $(1-\alpha)$ confidence interval.

   (c) Find values $1 \le a \le b$ minimizing the expected length $E_\theta(b\widehat{\theta} - a\widehat{\theta})$ among all $(1-\alpha)$ confidence intervals of the form (1), uniformly in θ.

1. Suppose that out of n i.i.d. Bernoulli trials, each with probability p of success, there are zero successes.

   (a) Given $\alpha \in (0,1)$, derive an exact upper $(1-\alpha)$-confidence bound for p by either pivoting the c.d.f. of the Binomial distribution or inverting the appropriate hypothesis test.

   (b) There is a famous rule of thumb called the "Rule of Threes" which says that, when n is large, $3/n$ is an approximate upper 95%-confidence bound for p in the above situation. Justify the Rule of Threes by applying a large-n first order Taylor approximation to your answer from Part 1a, and use the fact that $|\log(.05)| \approx 3$.

2. Let $w_1, \ldots, w_n$ be i.i.d. from the mixture distribution

$$f(w;\psi) = \sum_{i=1}^{g} \pi_i f_i(w),$$

   where $\psi = (\pi_1, \ldots, \pi_g)$ is a vector of unknown probabilities summing to one, and $f_1, \ldots, f_g$ are *known* density functions.

   (a) Write an equation one would solve to find the maximum likelihood estimate of ψ.

   (b) To implement the EM algorithm, write down the full likelihood when in addition to the sample $w_1, \ldots, w_n$, the 'missing data'

$$Z_{ij} = \mathbf{1}(\text{the } j\text{th observation } w_j \text{ comes from } i\text{th group } f_i),$$

   is also observed.

   (c) Write down the estimate of ψ using the full data likelihood in part (2b).

   (d) Write down the E and M steps of the EM algorithm.

1. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a vector of i.i.d. $N(\mu, \sigma^2)$ random variables, where both μ and σ are unknown.

   (a) Given $\alpha_1 \in (0,1)$, write down an exact $(1-\alpha_1)$ confidence interval for μ.

   (b) Given $\alpha_2 \in (0,1)$, write down an exact $(1-\alpha_2)$ confidence interval for σ^2.

   (c) Letting $\mathcal{I}_{\alpha_1}(\boldsymbol{X})$ and $\mathcal{J}_{\alpha_2}(\boldsymbol{X})$ denote the confidence intervals in parts 1a and 1b, respectively, for given $\alpha \in (0,1)$ show how to choose α_1, α_2 so that the overall coverage probability satisfies

   $$P_{\mu,\sigma^2}\left(\mu \in \mathcal{I}_{\alpha_1}(\boldsymbol{X}) \text{ and } \sigma^2 \in \mathcal{J}_{\alpha_2}(\boldsymbol{X})\right) \geq 1-\alpha \quad \text{for all} \quad \mu, \sigma^2.$$

   The inequality does not have to be sharp.

2. Let P_0 and P_1 be probability distributions on $\mathbb{R}$ with densities p_0 and p_1 with respect to Lebesgue measure, and let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables.

   (a) Let β denote the power of the most powerful test of size α, $0 < \alpha < 1$, for testing the null hypothesis $H_0 : X_1, \ldots, X_n \sim P_0$ against the alternative $H_a : X_1, \ldots, X_n \sim P_1$. Show that $\alpha < \beta$ unless $P_0 = P_1$.

   (b) Let P_0 be the uniform distribution on the interval $[0,1]$ and P_1 be the uniform distribution on $[1/3, 2/3]$. Find the Neyman-Pearson test of size α for testing H_0 against H_a (consider all possible values of $0 < \alpha < 1$).

1. Suppose lifetimes $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ of $n+m$ lightbulbs are independent and have the exponential distribution

$$p(x;\theta) = (1/\theta)\exp(-x/\theta)\mathbf{1}(x > 0)$$

with unknown parameter θ. The first n lifetimes have been observed precisely, but the only information recorded on the final m observations is whether or not the bulb lasted longer than some time t. We consider using the EM algorithm to compute the maximum likelihood estimate of θ.

   a. Write down the full likelihood function, that is, had all failure times been observed, and the full log likelihood.

   b. Write down the maximum likelihood estimate that is obtained by using the full likelihood.

   c. For X an exponential variable with the same distribution as the data, compute

$$E[X|X > t] \quad \text{and} \quad E[X|X < t].$$

   d. Describe the E and M step for computing the maximum likelihood estimate of θ under the data as observed.

2. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, 1)$ random variables.

   (a) Find the likelihood ratio test of size $0 < \alpha < 1$ for testing $H_0 : \mu = 0$ against $H_a : \mu \neq 0$.

   (b) Is this test uniformly most powerful? Justify your answer.

   (c) Does a uniformly most powerful test exist for this problem? Either exhibit such a test or prove that none exists.

1. Let $M(n; p_1, \ldots, p_k)$ denote the multinomial distribution with n trials and cell probabilities $p_1, \ldots, p_k$. Now let $X = (x_1, x_2, x_3, x_4)$ have multinomial distribution

$$M(200; 1/2 + \theta/2, 1/4 - \theta/4, 1/4 - \theta/3, \theta/12)$$

for some $\theta \in (0, 3/4)$. Write down the E-step and M-step of the Expectation Maximization algorithm for estimating θ by assuming that there is actually data $(x_{11}, x_{12}, x_{21}, x_{22}, x_3, x_4)$ from 6 cells rather than 4, but that the first 4 cells are unobserved but $x_1 = x_{11} + x_{12}$ and $x_2 = x_{21} + x_{22}$ are observed. *Hint:* Choose convenient cell probabilities for the unobserved cells in the complete model.

2. Recall that the bivariate normal distribution with mean $(\mu_1, \mu_2)^T$, variances σ_1^2, σ_2^2, and correlation coefficient ρ has density

$$f(y_1, y_2) \propto \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - \frac{2\rho(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right\}\right] \quad (1)$$

for $y_1, y_2 \in (-\infty, \infty)$.

Now let θ be a random variable with (univariate) normal distribution $N(\mu, \tau^2)$ and, given θ, let $X_1, \ldots, X_n$ be i.i.d. with (univariate) normal distribution $N(\theta, \sigma^2)$. Let $\overline{X} = (1/n)\sum_{i=1}^n X_i$ denote the sample mean. For this problem assume that θ is unobserved, the X_i are observed, τ and σ are known values, and μ is an unknown parameter.

   (a) Find the joint distribution of θ and $\overline{X}$. Give the name of the distribution and the values of any parameters in terms of quantities defined above. *Hint:* It may help to use the change of variables $\widetilde{\theta} = \theta - \mu$ and $\widetilde{X} = \overline{X} - \mu$.

   (b) Using your answer from part (2a), find the marginal distribution of $\overline{X}$. Give the name of the distribution and the values of any parameters in terms of quantities defined above.

   (c) Using your answer from part (2b), for arbitrary $\alpha \in (0, 1)$ write down an exact $(1 - \alpha)$ confidence interval for μ in terms of observed random variables and known quantities. Compare its width with the standard interval for θ and comment.

1. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p$ be known vectors (all vectors are column vectors in this problem) with $p \le n$, let $\boldsymbol{\beta} \in \mathbb{R}^p$ be an unknown parameter vector, and let $Y_1, \ldots, Y_n$ be independent Bernoulli random variables such that

$$P_{\boldsymbol{\beta}}(Y_i = 1) = 1 - P_{\boldsymbol{\beta}}(Y_i = 0) = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta}), \quad i = 1, \ldots, n, \tag{1}$$

where Φ denotes the c.d.f. of the standard normal distribution and "prime" denotes transpose. This problem concerns using the EM algorithm to find the MLE of $\boldsymbol{\beta}$ from the data $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$.

(a) Show that if $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. standard normals and $W_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$, $i = 1, \ldots, n$, then

$$Y_i = \mathbf{1}\{W_i > 0\}, \quad i = 1, \ldots, n \tag{2}$$

have the distribution (1). For the rest of this problem assume the Y_i are defined by (2) and observed but that the W_i and ε_i are unobserved.

(b) Show that the complete log-likelihood function $\ell_c(\boldsymbol{\beta})$ of $W_1, \ldots, W_n$, if they were observed, is given by

$$\ell_c(\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}(W_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2$$

up to additive constants.

(c) The E step of the EM algorithm involves computing

$$Q(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta}) = E_{\boldsymbol{\beta}}\left[\ell_c(\widetilde{\boldsymbol{\beta}}) \middle| \boldsymbol{Y}\right].$$

Show that

$$E_{\boldsymbol{\beta}}\left[(W_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2 \middle| \boldsymbol{Y}\right] = E_{\boldsymbol{\beta}}\left[(W_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2 \middle| Y_i\right], \tag{3}$$

and, up to additive terms C, C' that don't depend on $\widetilde{\boldsymbol{\beta}}$, that

$$E_{\boldsymbol{\beta}}\left[(W_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2 \middle| Y_i = 1\right] = (\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2 - 2\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \frac{\phi(\boldsymbol{x}_i'\boldsymbol{\beta})}{\Phi(\boldsymbol{x}_i'\boldsymbol{\beta})}\right) + C, \quad \text{and} \tag{4}$$

$$E_{\boldsymbol{\beta}}\left[(W_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2 \middle| Y_i = 0\right] = (\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}})^2 - 2\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}\left(\boldsymbol{x}_i'\boldsymbol{\beta} - \frac{\phi(\boldsymbol{x}_i'\boldsymbol{\beta})}{\Phi(-\boldsymbol{x}_i'\boldsymbol{\beta})}\right) + C', \tag{5}$$

where ϕ is the standard normal density function. You don't have to explicitly find C, C'.

(d) Let $\boldsymbol{X}$ be the $(n \times p)$ matrix with rows $\boldsymbol{x}_1', \ldots, \boldsymbol{x}_n'$, assumed to be of full rank p, and let $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{\beta})$ be the n-long vector with entries

$$v_i = \begin{cases} \phi(\boldsymbol{x}_i'\boldsymbol{\beta})/\Phi(\boldsymbol{x}_i'\boldsymbol{\beta}), & \text{if } Y_i = 1 \\ -\phi(\boldsymbol{x}_i'\boldsymbol{\beta})/\Phi(-\boldsymbol{x}_i'\boldsymbol{\beta}), & \text{if } Y_i = 0. \end{cases}$$

Show that the recursion for the EM iterates $\{\boldsymbol{\beta}^{(k)}\}$ given by maximizing $Q(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta})$ over $\widetilde{\boldsymbol{\beta}}$ is

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{v}(\boldsymbol{\beta}^{(k)}). \tag{6}$$

Hint: Write $Q(\widetilde{\boldsymbol{\beta}}, \boldsymbol{\beta})$ in matrix notation using $\boldsymbol{X}$ and $\boldsymbol{v}$ and then maximize over $\widetilde{\boldsymbol{\beta}}$.

2. Let $\theta \sim N(0, 1)$ and, conditional on θ, let $Y_i \sim N(\theta, 1), i = 1, 2, \cdots, n$ be i.i.d. Let $\overline{Y} = \sum_{i=1}^{n} Y_i/n$.

(a) Compute the density function of $\overline{Y}$.

(b) Compute the posterior distribution of θ given $\overline{Y}$.

(c) Compute the conditional expectation $E[\theta|\overline{Y}]$, and determine the behavior of the posterior distribution of θ as $n \to \infty$.

1. Let $X_1, \ldots, X_n$ be an i.i.d. sample from $N(\theta, \theta)$ distribution, where $\theta > 0$.

   (a) Find a non-constant pivotal quantity (that is, a function of $X_1, \ldots, X_n$ and θ whose distribution does not depend on θ). *Hint:* See if you can make $\sum_{i=1}^n X_i$ into a pivot via simple transformations involving θ.

   (b) Use the pivotal quantity from part (a) to construct a confidence interval for θ.

2. For nonnegative integers θ, n, N with $n \leq N$ and $\theta \leq N$, the hypergeometric distribution $\text{Hyper}(\theta, n, N)$ has p.d.f.

$$f_\theta(x) = P_\theta(X = x) = \binom{\theta}{x}\binom{N-\theta}{n-x} \Big/ \binom{N}{n} \tag{1}$$

   for nonnegative integers x and values of the parameters such that the above quotient in (1) is defined, with $f_\theta(x) = 0$ otherwise. Recall that this is the distribution of the number of white balls in a simple random sample without replacement of size n from an urn containing θ white balls and $N - \theta$ black balls. Throughout this problem we shall treat n (i.e., the sample size) and N (i.e., the population size) as fixed and known, and θ as an unknown parameter.

   (a) If $X \sim \text{Hyper}(\theta, n, N)$, show that this family of distributions has the monotone likelihood ratio property in X.

   (b) Given $\alpha \in (0, 1)$, give the form of an exactly level-α UMP test of $H_0 : \theta = N$ vs. $H_1 : \theta < N$ in as simple a form as possible, involving X. Your test may need to involve randomization to achieve the exact level α. Justify that the test is UMP. *Hint:* For this you only need to consider $f_N(x)$, which takes a particularly simple form.

   (c) Suppose that among the urn's N balls, only 1 is black. Given $\alpha, \beta \in (0, 1)$, find an expression for the smallest sample size n guaranteeing that your level-α UMP test will reject H_0 with probability at least $1 - \beta$.

1. Let $X_1, \dots, X_n$ be i.i.d. with exponential distribution with parameter $\theta > 0$, meaning that that pdf of X_1 is $f_\theta(x) = \begin{cases} \theta e^{-x\theta}, & x \geq 0, \\ 0, & x < 0. \end{cases}$

   (a) Find the Maximum Likelihood estimator of θ.

   (b) Find the Likelihood ratio test for testing $H_0 : \theta = \theta_0$ against the alternative $H_a : \theta \neq \theta_0$ to make the test approximately level $\alpha \in (0,1)$. Specify the test statistic and the rejection region (based on its asymptotic distribution).

   (c) Show that the Likelihood ratio test for testing $H_0 : \theta \leq \theta_0$ against the alternative $H_a : \theta > \theta_0$ (where $\theta_0 > 0$) rejects when $\bar{X}_n \leq C$ for C depending on the desired size of the test, where $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$.

2. Let X_1 and X_2 be two independent observations from some distribution F with mean θ (nothing else is known about F). Assume that we estimate the mean via the sample mean $\hat{\theta} = \frac{X_1+X_2}{2}$. We are interested in estimating the probability $\lambda(t) = P_F\left(\hat{\theta} - \theta \leq t\right)$ for all t.

   (a) We are going to estimate $\lambda(t)$ using bootstrap. Assume that observed values are $X_1 = 1$ and $X_2 = 3$. Let $\hat{X}_1$ and $\hat{X}_2$ be the bootstrap sample, and find its distribution (for better readability, you may draw a small table with possible values of $(\hat{X}_1, \hat{X}_2)$ and their probabilities).

   (b) Let $\hat{\theta}^* = \frac{\hat{X}_1+\hat{X}_2}{2}$. Find the distribution of $\hat{\theta}^* - \hat{\theta}$ (conditionally on the values of X_1 and X_2), and estimate $\lambda(-0.5)$.

   (c) Now assume that F has density $f_\theta(x) = \frac{1}{2}e^{-|x-\theta|}$. Again, we would like to estimate the probability $P_F\left(\hat{\theta} - \theta \leq t\right)$ via bootstrap. How would you proceed in this case? Your solution has to use the additional information about the density.

1. For given density functions p and q we consider the Neyman-Pearson tests of level $\alpha \in (0,1)$ for the hypotheses $H_0 : r = p$ versus $H_1 : r = q$, when observing n independent observations from density r.

   (a) Derive the Neyman-Pearson test for a given level $\alpha \in (0,1)$ when p and q are the density functions of the $\mathcal{N}(\mu, \sigma^2)$ distribution with means $\mu = \mu_0$ and $\mu = \mu_1$, respectively, with known variance σ^2. Make the form of the test as simple as you can.

   (b) Find the power $\beta(\mu)$ of the test in a) as function of μ.

   (c) For any given density functions p and q such that the Kullback Liebler Divergence

$$D(p||q) = E_p\left[\log\frac{p(X)}{q(X)}\right] \quad \text{and the variance} \quad \tau^2_{p||q} = \mathrm{Var}_p\left[\log\frac{p(X)}{q(X)}\right]$$

   are finite, use the Central Limit Theorem to derive an approximation to the Neyman Pearson test for a given level $\alpha \in (0,1)$.

   (d) Assuming in addition that $D(q||p)$ and $\tau^2_{q||p}$ are finite, likewise approximate the power function of the test in (c), and show that it recovers the normal case in (b).

2. Throughout this problem let $a > 0$ be known constant. Let f be the density function

$$f(x) \propto a - x, \quad \text{for} \quad 0 < x < a.$$

   Assuming the only random variables you have access to are i.i.d. $U(0,1)$, give the details of a way to simulate a random variable with distribution f.

1. Let F be a given strictly increasing cumulative distribution function with corresponding density f on the real line. Assume that n i.i.d. random variables are generated from F, but we are told only $X := \max(X_1, \dots, X_n)$. In particular, the positive integer n itself is unknown.

   (a) Find the probability density function $f_n(\cdot)$ of X.

   (b) Show that the family $\{f_n(\cdot),\ n \geq 1\}$ has monotone likelihood ratio with respect to some statistic, and make sure to identify the statistic involved.

   (c) Find the uniformly most powerful test for testing $H_0 : n \leq 5$ against $H_a : n > 5$.

2. Suppose that $X_1, \dots, X_n$ are i.i.d. random variables with mean μ, variance σ^2 and finite moments of all orders. We are interested in estimating $g(\mu)$ where $g : \mathbb{R} \mapsto \mathbb{R}$ is some smooth function with bounded third derivative.

   (a) The "plug-in" estimator of $g(\mu)$ is $g(\overline{X}_n)$ where $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ is the sample mean. Using Taylor's expansion and assuming that n is large, find the leading term in the expression for the bias $b := Eg(\overline{X}_n) - g(\mu)$ of $g(\overline{X}_n)$ for estimating $g(\mu)$.

   (b) Since μ is unknown, the bias is also unknown. Explain how one can use the non-parametric bootstrap to estimate the bias b. Specifically, write down the expression for the bootstrap estimator of the bias.

   (c) Let $g(x) = x^2$ and $n = 2$. Suppose that the observed values are $X_1 = 1$ and $X_2 = 3$. Find the exact value of the bootstrap estimator of the bias in this case. Which value would you use to estimate $g(\mu)$?

1. (Wilk's Theorem) Let $X_1, \ldots, X_n$ be independent random variables distributed as $\mathcal{P}(\lambda), \lambda \in \Lambda = (0, \infty)$, that is, with probability mass function

$$P_\lambda(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

For $\lambda_0 \in \Lambda$, we wish to test $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$. It may be useful to recall that the mean and variance of the $\mathcal{P}(\lambda)$ distribution are both equal to λ.

1.1. Write out the Generalized likelihood ratio test statistic G_n for this instance.

1.2. Verify directly that the conclusion of Wilk's theorem holds, that is, asymptotically as $n \to \infty$ that $-2 \log G_n$ has a χ^2 distribution, and specify its degrees of freedom. You may use the fact that $\overline{X}/\lambda_0 = 1 + (\overline{X} - \lambda_0)/\lambda_0$, and also apply $x - x^2/2$ as an approximation to $\log(1 + x)$ for x small, without further justification.

1.3. Using direct methods as in the previous part, develop an asymptotic for the power of the test in terms of the non-central χ^2 statistic for the sequence of alternatives of the form $\lambda_n = \lambda_0 + \delta/\sqrt{n}$; make sure you specify both the degrees of freedom and the non-centrality parameter. (Recall that the sum $(Z_1 + \mu_1)^2 + \cdots + (Z_k + \mu_k)^2$, where $Z_1, \ldots, Z_k$ are iid $\mathcal{N}(0, 1)$, has a non central χ^2 distribution on k degrees of freedom with non-centrality parameter $\mu_1^2 + \cdots + \mu_k^2$.)

2. (Jackknife) Let $X_1, \ldots, X_n$ be independent with common distribution function depending on $\theta \in \Theta \subset \mathbb{R}$, unknown.

2.1. Let $W_1, W_2, \ldots$ be a sequence of estimators for a parameter $\theta \in \Theta$ so that for any $n \geq 1$, $W_n = t_n(X_1, \ldots, X_n)$ for some $t_n : \mathbb{R}^n \to \Theta$. For any $n \geq 1$, define the **jackknife estimator** of θ to be

$$Z_n := nW_n - \frac{n-1}{n} \sum_{i=1}^{n} t_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

Assume that $W_1, W_2, \ldots$ are asymptotically unbiased in the sense that there exists $a, b \in \mathbb{R}$ such that

$$EW_n = \theta + a/n + b/n^2 + O(1/n^3), \qquad \forall\, n \geq 1. \qquad (*)$$

Show that if $b = 0$ and the $O(1/n^3)$ term is zero in $(*)$, then Z_n is unbiased for θ.

2.2. Show, generally that when $(*)$ holds,

$$EZ_n = \theta + O(1/n^2), \quad \forall\, n \geq 1.$$

2.3. Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables with parameter $0 < \theta < 1$. The MLE for θ is the sample mean, so by the Functional Equivariance Property of the MLE, the MLE for θ^2 is

$$W_n := \Big(\frac{1}{n} \sum_{i=1}^{n} X_i\Big)^2.$$

(You can take this fact as given.) Show that W_n is a biased estimate of θ^2, but that the jackknife estimator of θ^2 is unbiased.

1. Let $\boldsymbol{X} \sim \mathcal{N}_n(\mu, \Sigma)$, that is, $\boldsymbol{X}$ has the multivariate normal distribution in $\mathbb{R}^n$ with mean μ and covariance matrix Σ, and for all $\alpha \in (0,1)$ defined z_α by $P(Z \le z_\alpha) = \alpha$ where $Z \sim \mathcal{N}(0,1)$.

   (a) Let Σ be positive definite, and for fixed μ_1 and μ_2, distinct vectors in $\mathbb{R}^n$, express the Neyman-Pearson test with type I error $\alpha \in (0,1)$ of $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$ in terms of a test statistic $T(\boldsymbol{X})$ that has the $\mathcal{N}(0,1)$ distribution under H_0.

   (b) Find the power function β of the test in (a) and determine what it specializes to in the case where the components of $\boldsymbol{X}$ are independent and identically distributed univariate normal distributions.

   (c) Let Σ be a non-zero, non-negative definite covariance matrix that is not positive definite. Show that for some values of μ_0, μ_1 that there exists tests for the hypotheses in part (a) that have type I error $\alpha = 0$ and power $\beta = 1$. Determine a set of necessary and sufficient conditions on μ_0, μ_1 for that to be the case, prove that your conditions are as claimed, and give the form of these $\alpha = 0, \beta = 1$ tests.

2. Let $X_1, \ldots, X_n$ be i.i.d. with Bernoulli(p) distribution for some $p \in (0,1)$, meaning that $P(X_1 = 1) = p,\ P(X_1 = 0) = 1 - p$.

   (a) Show that the maximum likelihood estimator of p is $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$.

   (b) For some $p_0 \in (0,1)$, we would like to test the null hypothesis $H_0 : p = p_0$ against the alternative $H_a : p \neq p_0$ using the likelihood ratio test. The log-likelihood ratio test statistic is defined as $\Lambda_n = \log\left(\frac{\sup_{p\in[0,1]} L_n(p)}{L_n(p_0)}\right)$ where $L_n(p)$ is the likelihood function, that is, the joint density of the observations, considered as a function of p. Show that the explicit expression for this test statistic is

   $$\Lambda_n = n\left(\bar{X}_n \log\left(\frac{\bar{X}_n}{p_0}\right) + (1-\bar{X}_n)\log\left(\frac{1-\bar{X}_n}{1-p_0}\right)\right).$$

   (c) Prove that

   $$2\Lambda_n = \frac{n\left(\bar{X}_n - p_0\right)^2}{p_0(1-p_0)} + o_P(1)$$

   where $o_P(1)$ is a term that converges to 0 in probability as $n \to \infty$. Deduce from this expression the approximate distribution of $2\Lambda_n$ for large n.

   For this task, you may use the following facts without proving them: $\bar{X}_n \log\left(\frac{\bar{X}_n}{p_0}\right) = (\bar{X}_n - p_0 + p_0)\log\left(1 + \frac{\bar{X}_n - p_0}{p_0}\right)$, $\log(1+x) = x - x^2/2 + o(x^2)$ when $x \to 0$ and $n \cdot o\left((\bar{X}_n - p_0)^2\right)$ converges to 0 in probability as $n \to \infty$ (here, $o(\cdot)$ stands for "small-O").

1. Let $X_1, \ldots, X_n$ be i.i.d. with Poisson distribution $P(\lambda)$ where $\lambda \in (0, \infty)$, that is, for all $i = 1, \ldots, n$ that $P(X_i = k) = e^{-\lambda}\lambda^k/k!, k = 0, 1, \ldots$ and are independent.

   (a) Show that this family of distributions has the monotone likelihood ratio property with respect to an appropriately chosen statistic $T(X_1, \ldots, X_n)$.

   (b) For the statistical model described, give an example of a hypothesis testing problem $H_0 : \theta \in \Theta_0,\ H_a : \theta \in \Theta_a$ where Θ_0, Θ_a are two subsets of $(0, \infty)$ that satisfy $\Theta_0 \cap \Theta_a = \emptyset,\ \Theta_0 \cup \Theta_a = (0, \infty)$, that admits a uniformly most powerful test of any size $\alpha \in [0, 1]$. Justify your answer.

   (c) For the statistical model described above, give an example of a hypothesis testing problem $H_0 : \theta \in \Theta_0,\ H_a : \theta \in \Theta_a$, with $\Theta_0,\ \Theta_a$ satisfying the same properties as in part (b), that **does not** admit a uniformly most powerful test of given size $\alpha \in (0, 1)$. Justify your answer.

2. Let $X = (X_1, \ldots, X_n)$ be a random sample of size n from a family of probability densities $\{f_\theta : \theta \in \mathbb{R}\}$, so that $f_\theta : \mathbb{R}^n \to (0, \infty)$ for any $\theta \in \mathbb{R}$, and $X_1, \ldots, X_n$ are i.i.d. Fix $\theta_0 \in \mathbb{R}$. Suppose we test the hypothesis H_0 that $\{\theta = \theta_0\}$ versus the alternative $\{\theta \neq \theta_0\}$. Let $Y = Y_n$ denote the MLE of θ, and assume that under P_{θ_0} that

$$Y_n \to_p \theta_0 \quad \text{and} \quad \sqrt{n}(Y - \theta_0) \to_d \mathcal{N}(0, I_{X_1}^{-1})$$

   where $\to_p$ and $\to_d$ denote convergence in probability and in distribution respectively, $\mathcal{N}$ denote the normal distribution and I_X the Fisher information, which we assume exists.

   Let

$$\lambda(X) := \frac{\sup_{\theta \in \mathbb{R}} f_\theta(X)}{f_{\theta_0}(X)}$$

   denote the generalized likelihood ratio statistic. If H_0 is true, we will ask you to show that $-2 \log \lambda(X)$ converges in distribution as $n \to \infty$ to a chi-squared random variable with one degree of freedom, under some additional assumptions.

   Fix $x \in \mathbb{R}^n$ and denote $\ell_n(\theta) := \log f_\theta(x)$. Below you may assume whatever smoothness conditions you require for your argument, but please note them when they are applied, in the form of, say: here we assume that a second order Taylor expansion for $\ell_n(\theta)$ holds with appropriate form of remainder.

   (a) Using Taylor series, show that

$$-2 \log \lambda(X) = -\ell_n''(\widehat{Y})(\theta_0 - Y)^2$$

   where $\widehat{Y}$ is some point in an interval with endpoints Y and θ_0.

   (b) Using the weak law of large numbers, show that, for any $\theta \in \mathbb{R}$, $\frac{1}{n}\ell_n''(\theta)$ converges in probability to the constant $I_{X_1}(\theta)$ as $n \to \infty$, and that the same conclusion holds for $\frac{1}{n}\ell_n''(\widehat{Y})$.

   (c) Combining the above observations, conclude that $-2 \log \lambda(X)$ converges in distribution to a chi-squared random variable with one degree of freedom.

1. Let $X_1, ..., X_n$ be an i.i.d. sample from the uniform distribution $U[0, \theta]$, $\theta > 0$.

   (a) Find a sufficient statistic for θ and construct a pivotal quantity for θ based on this sufficient statistic. Finally, construct a lower $(1-\alpha)$-confidence bound $\widehat{\theta}_L$ for θ based on this pivotal quantity, that is, $P(\theta \geq \widehat{\theta}_L) \geq 1 - \alpha$.

   (b) Show that the family of uniform distributions defined above has the monotone likelihood ratio property (with respect to which statistic?)

   (c) Find the most powerful test of size $\alpha = 0.1$ for testing $H_0 : \theta = \theta_0$ against $H_a : \theta > \theta_0$ for some $\theta_0 > 0$. Invert the rejection region of the test to construct a confidence interval for θ. Compare your result to part (a).

2. Let $X_1, \ldots, X_n$ be i.i.d. normal $N(0, \sigma^2)$ for some $\sigma^2 > 0$.

   (a) Find a maximum likelihood estimator (MLE) of σ^2.

   (b) Find a MLE of σ (without the square!), denoted $\widehat{\sigma}_n$, and find the asymptotic distribution (as $n \to \infty$), including the asymptotic variance, of $\sqrt{n}(\widehat{\sigma}_n - \sigma)$ using the general asymptotic properties of the MLE.
   (hint: you can use the fact that $\mathbb{E}X_1^4 = 3\sigma^4$)

   (c) Assume that the sample is of size $n = 2$ and $X_1 = -2$, $X_2 = 4$. Find the upper 75% confidence bound for σ using the non-parametric bootstrap. When would the bootstrap approach be advantageous, compared to using the asymptotic pivotal quantity constructed in part (b)?

Spring 2024 Math 541b Exam

1. Let X, Y be exponential random variables with densities $f_X(t) = \lambda_1 e^{-\lambda_1 t}$, $t \geq 0$ and $f_Y(t) = \lambda_2 e^{-\lambda_2 t}$, $t \geq 0$, where $\lambda_1, \lambda_2 > 0$ are parameters. We would like to test $H_0 : \lambda_1 \leq \lambda_2$ against the alternative $H_a : \lambda_1 > \lambda_2$. Note that the null and the alternative do not change under transformation of the parameters given by $(\lambda_1, \lambda_2) \mapsto (c\lambda_1, c\lambda_2)$, $c > 0$.

   (a) What is the distribution of $\frac{X}{c}$ for $c > 0$?

   (b) Suggest a function $T(x, y)$ such that

   $$T(x, y) = T(x', y') \iff x' = cx,\ y' = cy \text{ for some } c > 0,$$

   and argue that $T(X, Y)$ is a natural choice of the test statistic for the problem of testing H_0 against H_a.

   (c) Show that the density corresponding to the distribution of $T(X, Y)$ is given by $p_T(t) = \frac{\lambda_2}{\lambda_1} \frac{1}{\left(t + \frac{\lambda_2}{\lambda_1}\right)^2}$, $t \geq 0$.

   (d) Note that the family of densities in part (c) depends only on one parameter $\tau = \frac{\lambda_2}{\lambda_1}$. Show that this family has the monotone likelihood ratio, and find the uniformly most powerful test for the problem $H_0' : \tau \geq 1$ against $H_a' : \tau < 1$. Which test for the original problems does this give?

2. Let $X_1, \ldots, X_n$ be an i.i.d. sample from normal distribution with mean μ and variance 1. It is known that $\mu \geq 0$.

   (a) Does the uniformly most powerful test for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ exist for any values of μ_0?

   (b) Find the simplest possible form of the Likelihood Ratio test for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ (please remember to take the fact that $\mu \geq 0$ into account!)

   (c) Let $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ be the sample mean. Assume that $\mu_0 > 0$ and prove that $\mathbb{P}\big(\bar{X}_n < 0\big) \to 0$ as $n \to \infty$ (for example, you can use Chebyshev's inequality). Use this fact to show directly that the asymptotic distribution of $2 \log \Lambda_n$, where $\Lambda_n = \frac{\sup_{\mu \geq 0} L_n(\mu)}{L_n(\mu_0)}$ and $L_n(\mu)$ is the likelihood function, is chi-squared with 1 degree of freedom (again, assuming that $\mu_0 > 0$).

   (d) Find the asymptotic distribution of $2 \log \Lambda_n$ when $\mu_0 = 0$.