Name: ______________________________

Question:	1	2	3	4	Total
Points:	9	10	18	8	45
Score:					

E1 244 - Detection & Estimation Theory (2019) Final Exam

Instructions

- This exam has a total of 4 questions with a maximum score of 45 points. The total time is 3 hours.
- There are partial marks for subquestions, so please attempt as many parts as possible.
- Write your name at the top of this cover sheet.
- Attach your solution sheets to this cover sheet and return everything including rough work.
- No class notes, calculators or electronic aids are permitted.
- Academic dishonesty will not be tolerated.

Useful formulas, definitions, etc.

- **Generalized Likelihood Ratio Test (GLRT).** The GLRT for testing the composite hypotheses $H_0 : Y \sim p_\theta, \theta \in \Lambda_0$ vs $H_1 : Y \sim p_\theta, \theta \in \Lambda_1$ has the form

$$\frac{\max_{\theta \in \Lambda_1} p_\theta(y)}{\max_{\theta \in \Lambda_0} p_\theta(y)} \gtreqless \eta.$$

- **Constrained quadratic minimization.** For $f(x) = ax^2 + bx + c$ and $u \in \mathbb{R}$,

$$\arg\min_{x \geq u} f(x) = \max\left\{u, \arg\min_{x \in \mathbb{R}} f(x)\right\} \quad \text{and}$$
$$\arg\min_{x \leq u} f(x) = \min\left\{u, \arg\min_{x \in \mathbb{R}} f(x)\right\}.$$

- **Geometric probability distribution.** For $0 \leq \theta \leq 1$, the Geom(θ) distribution has pmf $p[i] = \theta(1-\theta)^{i-1}$, $i = 1, 2, \ldots$
- **Negative binomial probability distribution.** For $r = 1, 2, \ldots$ and $0 \leq \theta \leq 1$, the NegBin(r, θ) distribution has pmf $p[i] = \binom{i+r-1}{i} \cdot \theta^r \cdot (1-\theta)^i$, $i = 0, 1, 2, \ldots$

1. **Mixture testing**
   Consider testing the hypotheses

$$H_0 : (Y_1, \dots, Y_n) \overset{\text{iid}}{\sim} \text{Unif}([k]), \quad \text{vs.}$$
$$H_1 : (Y_1, \dots, Y_n) \sim P_\epsilon,$$

where k is a known positive integer, Unif($[k]$) is the uniform probability distribution over the alphabet $[k] = \{1, \dots, k\}$, $\epsilon \in (0, 1)$ is known, and P_ϵ is the joint probability distribution of $Y_1, \dots, Y_n$ defined as follows: First, $Z \sim$ Unif($[k]$) is sampled uniformly. Then (given Z), $Y_1, \dots, Y_n$ are sampled iid from the probability distribution P_ϵ^Z over $[k]$, under which each outcome $j \in [k]$ has probability

$$P_\epsilon^Z[j] = \begin{cases} \frac{1}{k} + \epsilon, & \text{if } j = Z \\ \frac{1}{k} - \frac{\epsilon}{k-1}, & \text{if } j \in [k] \setminus \{Z\}. \end{cases}$$

(a) **(2 points)** For a general sequence $(y_1, \dots, y_n)$ with elements from $[k]$, write down the probability of the sequence under each hypothesis.

Solution: The probability of $(y_1, \dots, y_n)$ occurring under H_0 is k^{-n}. The probability of $(y_1, \dots, y_n)$ occurring under H_1 is

$$\frac{1}{k} \sum_{\ell=1}^{k} \left(\frac{1}{k} + \epsilon \right)^{N_\ell} \left(\frac{1}{k} - \frac{\epsilon}{k-1} \right)^{n - N_\ell}$$

where N_ℓ denotes the number of symbols that are equal to ℓ, i.e., $N_\ell = \sum_{i=1}^{n} \mathbb{1}\{Y_i = \ell\}$.

(b) **(3 points)** Write down, in the shortest possible way, a Bayes-optimal test statistic for deciding between H_0 and H_1. How does it depend on the observed sequence $y_1, \dots, y_n$?

Solution: A Bayes-optimal test must threshold the likelihood ratio statistic

$$\begin{aligned} L_n &= \frac{p_1(y_1, \dots, y_n)}{p_0(y_1, \dots, y_n)} \\ &= \frac{\frac{1}{k} \sum_{\ell=1}^{k} \left(\frac{1}{k} + \epsilon \right)^{N_\ell} \left(\frac{1}{k} - \frac{\epsilon}{k-1} \right)^{n - N_\ell}}{k^{-n}} \\ &= \frac{1}{k} \left(1 - \frac{\epsilon k}{k-1} \right)^n \sum_{\ell=1}^{k} \left(\frac{1 + \epsilon k}{1 - \frac{\epsilon k}{k-1}} \right)^{N_\ell}, \end{aligned}$$

which is equivalent to thresholding the statistic

$$T(y_1, \dots, y_n) = \sum_{\ell=1}^{k} \left(\frac{1 + \epsilon k}{1 - \frac{\epsilon k}{k-1}} \right)^{N_\ell}.$$

Hence, a Bayes-optimal test statistic only depends on the number of symbols for each alphabet (N_ℓ, $\ell = 1, \dots, k$) in the sequence $y_1, \dots, y_n$.

(c) **(4 points)** Suppose k (the alphabet size) is quite large ($k \gg 1$), and $\epsilon \gg \frac{1}{k}$. Explain why the test statistic you devised above should reliably detect the true hypothesis as n (the sample size) becomes large.
(Hint: Approximate the test statistic in this regime of k and ϵ, and think about what happens to it under H_0 and H_1 as $n \to \infty$).

Solution: When $k \gg 1$ and $\epsilon k \gg 1$, we can approximate $1 + \epsilon k \approx \epsilon k$, $\frac{k}{k-1} \approx 1$ and $1 - \frac{\epsilon k}{k-1} \approx 1$, using which

$$\begin{aligned} T(y_1, \ldots, y_n) &= \sum_{\ell=1}^{k} \left(\frac{1+\epsilon k}{1 - \frac{\epsilon k}{k-1}} \right)^{N_\ell} \\ &\approx \sum_{\ell=1}^{k} \left(\frac{\epsilon k}{1-\epsilon} \right)^{N_\ell} \\ &\approx \sum_{\ell=1}^{k} (\epsilon k)^{N_\ell} . \end{aligned}$$

Under H_0, since every alphabet is equiprobable, by the law of large numbers,

$$T(y_1, \ldots, y_n) \approx \sum_{\ell=1}^{k} (\epsilon k)^{n/k} = k\,(\epsilon k)^{n/k}, \tag{1}$$

while under H_1, again by the law of large numbers,

$$T(y_1, \ldots, y_n) \approx \sum_{\ell=1}^{k} (\epsilon k)^{N_\ell} \geq (\epsilon k)^{N_Z} = (\epsilon k)^{n\left(\frac{1}{k}+\epsilon\right)} = (\epsilon k)^{n\epsilon} (\epsilon k)^{n/k} . \tag{2}$$

Comparing (1) and (2), as $n \to \infty$, T grows exponentially faster under H_1 due to the presence of the term $(\epsilon k)^{n\epsilon}$ in (2) which grows with n, instead of k in (1) which does not grow with n.

2. **Composite hypothesis testing**
Consider testing the hypotheses

$$H_0 : Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1), \quad \mu \in \left(-\infty, -\frac{\epsilon}{2}\right)$$

vs.

$$H_1 : Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1), \quad \mu \in \left(\frac{\epsilon}{2}, \infty\right).$$

(a) **(5 points)** Write down the form of the Generalized Likelihood Ratio Test (GLRT) for this problem, in the shortest possible way. What real-valued function of the observations $y_1, \ldots, y_n$ does the test statistic depend on?

Solution: The GLRT test statistic is

$$
\begin{aligned}
T(y_1,\ldots,y_n) &= \frac{\max_{\mu\geq\frac{\epsilon}{2}} p_\mu(y_1,\ldots y_n)}{\max_{\mu\leq-\frac{\epsilon}{2}} p_\mu(y_1,\ldots y_n)} \\
&= \frac{\max_{\mu\geq\frac{\epsilon}{2}} \prod_{i=1}^n \exp\left(-\frac{(y_i-\mu)^2}{2}\right)}{\max_{\mu\leq-\frac{\epsilon}{2}} \prod_{i=1}^n \exp\left(-\frac{(y_i-\mu)^2}{2}\right)} \\
&= \frac{\max_{\mu\geq\frac{\epsilon}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2\right)}{\max_{\mu\leq-\frac{\epsilon}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2\right)} \\
&= \frac{\exp\left(-\frac{1}{2}\min_{\mu\geq\frac{\epsilon}{2}} \sum_{i=1}^n (y_i-\mu)^2\right)}{\exp\left(-\frac{1}{2}\min_{\mu\leq-\frac{\epsilon}{2}} \sum_{i=1}^n (y_i-\mu)^2\right)}.
\end{aligned}
$$

The unconstrained minimum of $\sum_{i=1}^n (y_i-\mu)^2$ over $\mu\in\mathbb{R}$ is achieved at $\mu^* = \frac{1}{n}\sum_{i=1}^n y_i = \bar{y}$, so the GLRT statistic becomes

$$
T(y_1,\ldots,y_n) = \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left[(y_i-\bar{y}_u)^2 - (y_i-\bar{y}_l)^2\right]\right\}
$$

where $\bar{y}_u = \max\left\{\bar{y}, \frac{\epsilon}{2}\right\}$ and $\bar{y}_l = \min\left\{\bar{y}, -\frac{\epsilon}{2}\right\}$. This is further equivalent to thresholding the test statistic

$$
\begin{aligned}
T_2(y_1,\ldots,y_n) &= -\frac{1}{2n}\sum_{i=1}^n \left[(y_i-\bar{y}_u)^2 - (y_i-\bar{y}_l)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^n \left(y_i - \frac{\bar{y}_u+\bar{y}_l}{2}\right)(\bar{y}_u-\bar{y}_l) \\
&= \left(\bar{y} - \frac{\bar{y}_u+\bar{y}_l}{2}\right)(\bar{y}_u-\bar{y}_l).
\end{aligned}
$$

The GLRT test statistic thus depends only on the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$.

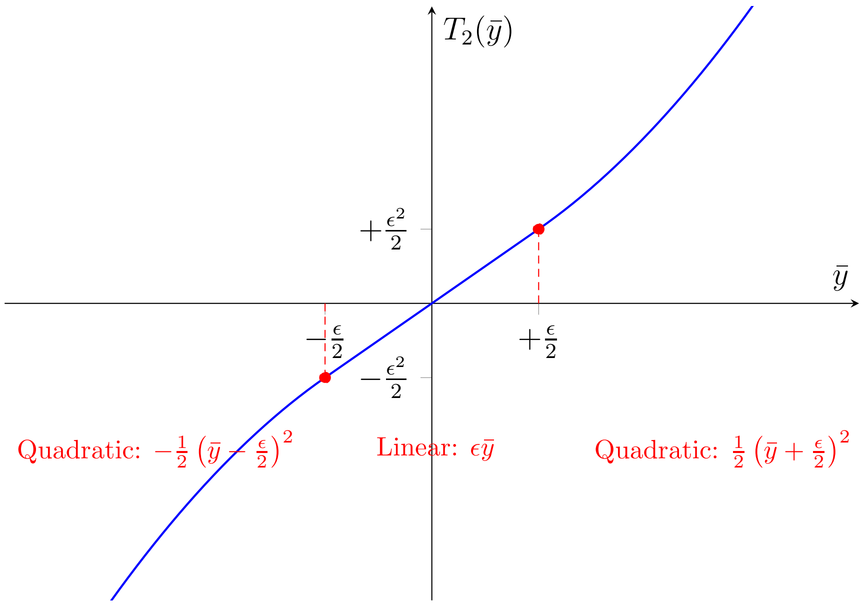
(b) **(5 points)** Sketch graphically the GLRT test statistic vs. the real-valued function of $y_1,\ldots,y_n$, that you found above, as clearly as you can.

Solution: We consider three cases for T_2 depending on the location of $\bar{y}$:

1. $\bar{y}\leq -\frac{\epsilon}{2}$: Here $T_2 = \left(\bar{y} - \frac{\frac{\epsilon}{2}+\bar{y}}{2}\right)\left(\frac{\epsilon}{2}-\bar{y}\right) = -\frac{1}{2}\left(\bar{y}-\frac{\epsilon}{2}\right)^2$.

2. $\bar{y}\geq \frac{\epsilon}{2}$: Here $T_2 = \left(\bar{y} - \frac{\bar{y}-\frac{\epsilon}{2}}{2}\right)\left(\bar{y}+\frac{\epsilon}{2}\right) = \frac{1}{2}\left(\bar{y}+\frac{\epsilon}{2}\right)^2$.

3. $-\frac{\epsilon}{2} < \bar{y} < \frac{\epsilon}{2}$: Here $T_2 = (\bar{y}-0)\,\epsilon = \epsilon\bar{y}$.



3. **Geometric distribution estimation**
Consider the family of geometric probability distributions $\{\text{Geom}(\theta) : \theta \in [0, 1]\}$.

(a) **(3 points)** Is this family complete? Why/why not?

Solution: The geometric family is complete. The pmf of the geometric distribution is of the exponential family form:

$$p_\theta(i) = \frac{\theta}{1-\theta} \cdot e^{i \log(1-\theta)}$$

with the parameter space $\theta \in [0, 1] \subset \mathbb{R}$ containing a 1-dimensional rectangle, so the family is complete.

(b) **(3 points)** Based on n iid samples $X_1, \ldots, X_n$ from $\text{Geom}(\theta)$, $\theta \in [0, 1]$, can you find a sufficient statistic for θ?

Solution: The probability of an iid sequence under $\text{Geom}(\theta)$ is

$$p_\theta(X_1, \ldots, X_n) = \left(\frac{\theta}{1-\theta}\right)^n \cdot \exp\left(\log(1-\theta)\sum_{i=1}^n X_i\right),$$

so by the factorization theorem (alternatively, by the exponential family theorem), a sufficient statistic for θ is $\sum_{i=1}^n X_i$.

(c) **(4 points)** What is the Fisher information for a single sample? Write down the Cramér-Rao lower bound for the variance of an unbiased estimator of θ based on n iid samples from the $\text{Geom}(\theta)$ distribution.

Solution: The Fisher information of $X \sim \text{Geom}(\theta)$ can be calculated as

$$\begin{aligned} -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log p_\theta(X)\right] &= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log\theta(1-\theta)^{X-1}\right] \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\{\log\theta+(X-1)\log(1-\theta)\}\right] \\ &= -\mathbb{E}\left[-\frac{1}{\theta^2}-\frac{X-1}{(1-\theta)^2}\right] \\ &= \frac{1}{\theta^2}+\frac{\frac{1}{\theta}-1}{(1-\theta)^2} = \frac{1}{\theta^2(1-\theta)}. \end{aligned}$$

Consequently, the Cramér-Rao lower bound for the variance of any unbiased estimator of θ based on n iid samples from $\text{Geom}(\theta)$ is the reciprocal of n times the Fisher information of a single sample: $\theta^2(1-\theta)/n$.

(d) **(3 points)** Suppose $X_1, \ldots, X_n$ are iid samples from $\text{Geom}(\theta)$, for some $\theta \in [0,1]$. Find the maximum likelihood estimator (MLE) of θ as a function of $X_1, \ldots, X_n$.

Solution: The maximum likelihood estimator is

$$\begin{aligned} \hat{\theta} &= \arg\max_{0\le\theta\le1} p_\theta(X_1,\ldots,X_n) \\ &= \arg\max_{0\le\theta\le1} \theta^n\,(1-\theta)^{-n+\sum_{i=1}^n X_i} \\ &= \arg\max_{0\le\theta\le1} n\log\theta+\left(-n+\sum_{i=1}^n X_i\right)\log(1-\theta) \\ &= \frac{n}{\sum_{i=1}^n X_i} = (\bar{X})^{-1}. \end{aligned}$$

(e) **(5 points)** Suppose $X_1, \ldots, X_n$ are iid samples from $\text{Geom}(\theta)$, for some $\theta \in [0,1]$. Find the best unbiased estimator of θ as a function of $X_1, \ldots, X_n$. Provide supporting arguments as clearly as possible.
(Hint: Start with a crude unbiased estimator of θ, then improve it. You can use the fact that the sum $\sum_{i=1}^m (Z_i - 1)$, where Z_i are iid $\text{Geom}(\theta)$ random variables, has the $\text{NegBin}(m,\theta)$ distribution.)

Solution: An unbiased estimator of θ is $W_1(X) = \mathbb{1}\{X_1 = 1\}$ which is a Bernoulli random variable with mean θ under p_θ.

We also have a sufficient statistic $T(X) = \sum_{i=1}^n X_i$ for θ from part (b). Moreover, the distribution of $T(X) - n$ is a negative binomial distribution with parameters (n, θ), which can be verified to be an exponential family distribution whose parameter set is the 1-dimensional rectangle $[0,1]$. So $T(X)$ is a complete statistic.

Let us improve the estimator W_1 by conditioning on the complete, sufficient statistic T; by the Lehmann-Scheffe theorem we know that this must be the best unbiased estimator of θ.

To find the exact form of the best unbiased estimator, we calculate, for any integer $t \geq n$,

$$
\begin{aligned}
\mathbb{E}\left[W_1(X) \mid T(X)=t\right] &= \mathbb{E}\left[\mathbb{1}\{X_1=1\} \;\Big|\; \sum_{i=1}^{n} X_i = t\right] = \mathbb{P}\left[X_1 = 1 \;\Big|\; \sum_{i=1}^{n} X_i = t\right] \\
&= \frac{\mathbb{P}\left[X_1=1, \sum_{i=1}^{n} X_i = t\right]}{\mathbb{P}\left[\sum_{i=1}^{n} X_i = t\right]} = \frac{\mathbb{P}\left[X_1=1, \sum_{i=2}^{n} X_i = t-1\right]}{\mathbb{P}\left[\sum_{i=1}^{n} X_i = t\right]} \\
&= \frac{\mathbb{P}\left[X_1=1, \sum_{i=2}^{n} X_i - (n-1) = t-n\right]}{\mathbb{P}\left[\sum_{i=1}^{n} X_i - n = t-n\right]} \\
&\overset{(a)}{=} \frac{\mathbb{P}\left[X_1=1\right]\mathbb{P}\left[\sum_{i=2}^{n} X_i - (n-1) = t-n\right]}{\mathbb{P}\left[\sum_{i=1}^{n} X_i - n = t-n\right]} \\
&\overset{(b)}{=} \frac{\theta \cdot \binom{t-2}{t-n}\theta^{n-1}(1-\theta)^{t-n}}{\binom{t-1}{t-n}\theta^{n}(1-\theta)^{t-n}} \\
&= \frac{n-1}{t-1},
\end{aligned}
$$

where (a) is by the independence of the X_i and (b) is by using the connection to the negative binomial distribution. So the best unbiased estimator of θ is $\frac{n-1}{\sum_{i=1}^{n} X_i - 1}$.

4. **Linear estimation of an autoregressive process**
A zero-mean discrete-time process $\{X_t\}_{t=-\infty}^{+\infty}$ evolves as

$$X_{t+1} = \alpha X_t + \beta X_{t-1} + W_t, \tag{3}$$

where W_t is iid $\mathcal{N}(0, \sigma_1^2)$ state noise across time t. The observations from this process are given by

$$Y_t = X_t + V_t, \tag{4}$$

where V_t is iid $\mathcal{N}(0, \sigma_2^2)$ observation noise across time t. Assume that the system is steady state (i.e., the process $\{X_t\}_t$ is wide-sense stationary).

(a) **(5 points)** Can you find the steady state 0-step and 1-step autocorrelations $r_0 = \mathrm{Cov}(X_t, X_t)$ and $r_1 = \mathrm{Cov}(X_{t+1}, X_t)$, in terms of α, β and σ_1^2?
(Hint: Use equation (3) creatively.)

Solution: Multiplying (3) by X_t on both sides and taking expectations gives

$$r_1 = \alpha r_0 + \beta r_1, \tag{5}$$

while squaring (3) and taking expectations gives

$$r_0 = \alpha^2 r_0 + \beta^2 r_0 + \sigma_1^2 + 2\alpha\beta r_1. \tag{6}$$

From (5) and (6) we get

$$r_0 = \frac{\sigma_1^2}{1-\alpha^2-\beta^2-\frac{2\alpha^2\beta}{1-\beta}}, \quad r_1 = \frac{\alpha\sigma_1^2}{(1-\beta)\left(1-\alpha^2-\beta^2-\frac{2\alpha^2\beta}{1-\beta}\right)}.$$

(b) **(3 points)** Suppose you want to find the best linear MMSE estimate of X_{t+1} in terms of the immediately preceding observations Y_t and Y_{t-1}. Describe how you would find the optimal linear coefficients in terms of r_0, r_1, α, β and σ_2^2.

Solution: Let the best linear MMSE estimate of X_{t+1} in terms of Y_t and Y_{t-1} be $h_1 Y_t + h_2 Y_{t-1}$. By the Yule-Walker equations,

$$\begin{aligned}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} &= \begin{bmatrix} \mathrm{Cov}(Y_t, Y_t) & \mathrm{Cov}(Y_t, Y_{t-1}) \\ \mathrm{Cov}(Y_t, Y_{t-1}) & \mathrm{Cov}(Y_t, Y_t) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{Cov}(X_{t+1}, Y_t) \\ \mathrm{Cov}(X_{t+1}, Y_{t-1}) \end{bmatrix} \\ &= \begin{bmatrix} r_0 + \sigma_2^2 & r_1 \\ r_1 & r_0 + \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \alpha r_0 + \beta r_1 \\ \alpha r_1 + \beta r_0 \end{bmatrix}.\end{aligned}$$