PRINT your name: __________________________.  __________________________
(last)  (first)

All answers should be justified unless otherwise stated.

You may consult a single, double-sided sheet of paper with notes. Apart from that, you may not look at books, notes, electronic devices etc.

You have 180 minutes. There are 7 questions, of varying credit (100 points total). The questions are of varying difficulty, so avoid spending too long on any one question.

Good luck!
A List of Useful Formulas:

- **Discrete-Time Fourier Transform**:

<table>
<thead>
<tr>
<th>Time domain $x[n]$</th>
<th>Fourier transform $X(f)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta[n-M]$</td>
<td>$e^{-i2\pi f M}$</td>
</tr>
<tr>
<td>$a^n u[n]$ (0 &lt; $</td>
<td>a</td>
</tr>
<tr>
<td>$W \text{sinc} (W n)$</td>
<td>$\text{rect} \left(\frac{f}{W}\right)$</td>
</tr>
<tr>
<td>$W \text{sinc}^2 (W n)$</td>
<td>$\text{tri} \left(\frac{f}{W}\right)$</td>
</tr>
<tr>
<td>$e^{-i\alpha n}$</td>
<td>$\delta (f + \alpha)$</td>
</tr>
</tbody>
</table>

where $\text{tri}(t) = \max \{1 - |t|, 0\}$ and $\text{rect}(t) = \begin{cases} 1, & |t| < 0.5, \\ 0, & \text{else}. \end{cases}$

- **Kalman Filter Recursion**: Suppose that the state evolves as

$$X_{n+1} = \alpha X_n + W_n, \quad W_n \sim \mathcal{N} (0, \sigma_W^2),$$

and the observation satisfies

$$Y_n = hW_n + Z_n, \quad Z_n \sim \mathcal{N} (0, \sigma_Z^2).$$

where $W_1, W_2, \cdots, Z_1, \cdots$ are independent. Then the Kalman filter recursion:

$$\hat{x}_n (y_{n-1}^i) = \alpha \hat{x}_{n-1} (y_{n-1}^i),$$

$$\sigma_{\hat{x}_n}^2 = \alpha^2 \sigma_{\hat{x}_{n-1}}^2 + \sigma_W^2,$$

$$\hat{x}_n (y_{1}^i) = \hat{x}_n (y_{1}^{i-1}) + \frac{h \sigma_{\hat{x}_{\hat{x}_n}}^2 [y_n - h \hat{x}_n (y_{n-1}^i) \}}{h^2 \sigma_{\hat{x}_n}^2 + \sigma_Z^2},$$

$$\sigma_{\hat{x}_n}^2 = \frac{\sigma_{\hat{x}_{\hat{x}_n}}^2 \sigma_Z^2}{h^2 \sigma_{\hat{x}_n}^2 + \sigma_Z^2},$$

where $\xi_n = X_n - \hat{X}_n (Y_1^n)$, and $\zeta_n = X_n - \hat{X}_n (Y_1^{n-1})$.

- **Cramer’s rule**:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}^{-1} = \frac{1}{ac - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$$
Problem 1 (4 points)

Let $U_1$ and $U_2$ be zero-mean Gaussian variables with variances $\sigma_1^2$, $\sigma_2^2$ and covariance $\text{Cov}(U_1, U_2)$. Compute the MMSE estimate and the MAP estimate of $U_1$ given $U_2$. Are they the same or different?

Problem 2 (6 points)

Consider an estimation problem: Estimate $X \sim \mathcal{N}(0, 1)$ given an observed random vector in two dimensions $Y = Xa + Z$. $Z = [Z_1, Z_2]^T$ is zero-mean jointly Gaussian with $\sigma_1^2 = \sigma_2^2 = 1$, $\text{Cov}(Z_1, Z_2) = 0.5$, and is independent of $X$. Choose a vector $a$, subject to the constraint $\|a\| = 1$, to minimize the MSE of estimating $X$ given $Y$, and compute the resulting MSE.

Problem 3 (9 points)

You have access to a random number generator which, by repeatedly calling, can be used to generate a sequence of independent random variables each uniformly distributed in $[0, 1]$.

(a) (3 points) Describe how you would use the random number generator to simulate a symmetric two-state Markov chain $\{X_n\}$ with crossover transition probability $\alpha$ and initial state distribution $\Pr\{X_0 = 1\} = p$.

(b) (3 points) Describe how you would use the random number generator to simulate a discrete-time i.i.d. $\mathcal{N}(0, 1)$ process.

(c) (3 points) Describe how you would use the random number generator to simulate a discrete-time zero-mean Gaussian process with power spectral density $S(f) = 1$ for $|f| < 0.25$ and $H(f) = 0$ for $0.25 \leq |f| < 0.5$.

Your descriptions should be explicit enough to be directly implementable in a language like MATLAB or Python.
Problem 4 (12 points)

We observe a nonnegative real valued random variable. Under the null hypothesis it is distributed like an exponential random variable of mean 1. Under the alternate hypothesis it is distributed as the sum of two independent exponentially distributed random variables of mean 1.

(a) (6 points) Compute explicitly and plot the Neyman-Pearson error curve, i.e. the optimal tradeoff between the probability of missed detection and the probability of false alarm.

(b) (3 points) Consider now the MAP rule under the prior that the two hypotheses are equally likely. What is the probability of missed detection and what is the probability of false alarm of this rule? Is the operating point of this rule on, below or above the error curve you plotted in part (a)? Explain.

(c) (3 points) Consider now a rule that maximizes the minimum of the probability of missed detection and the probability of false alarm. Show on the plot in part (a) how you can get the operating point of this rule. (No need to compute explicitly.) Is it on, below or above the error curve? Explain.

[ Useful facts:

- Pdf of an exponential random variable of mean 1: \( f_X(x) = e^{-x}, x \geq 0. \)
- Pdf of the sum of two independent exponential random variables with mean 1: \( f_Y(y) = ye^{-y}, y \geq 0. \)
- \( \int xe^{-x}dx = -(x+1)e^{-x}+C. \)
]
Problem 5 (25 points)

(a) (i) (3 points) You have a biased coin with probability of getting a Head equals $p$. The parameter $p$ is unknown. How should you estimate $p$ based on $n$ independent tosses of the coin, such that your estimate converges to $p$ as $n \to \infty$? Make precise your notion of convergence and justify why your estimate converges under this notion.

(ii) (4 points) Now suppose your eye-sight is not so good so with probability 0.1 a Head will be seen as a Tail and with probability 0.1 a Tail will be seen as a Head. Let us further assume the errors your eyes make are independent from flip to flip. Find a procedure to estimate the coin bias $p$ from $n$ independent flips of the coin, such that your estimate converges to $p$ as $n \to \infty$. Justify your answer.

(b) The RNA found in a cell can be described by $N$ length-$L$ sequences $s_1, s_2, \ldots, s_N$, called transcripts, with each transcript $s_i$ appearing with abundance $\rho_i$, where $\sum_i \rho_i = 1$. In the RNA-Seq protocol, the sequencing machine generates a set of reads, with each read being $s_i$ with probability $\rho_i$, independent from every other read. The goal is to estimate from the reads the abundance parameters $\rho_i$’s, which yield gene expression information to the biologist.

(i) (2 points) Find a procedure to estimate the abundance parameters such that your estimates converge to the true parameter values as the number of reads increases to infinity.

(ii) Now suppose there is noise in the sequencing machine and there are read errors. For simplicity, let us assume there are only two transcript $s_1 = a, s_2 = b$ with abundance parameters $\rho$ and $1 - \rho$ respectively. Furthermore, we will assume each transcript is a binary sequence. The errors on a read are independent flips of the symbols on the chosen underlying transcript, with flip probability 0.01.

(1) (2 points) Consider an example where $a = [0, 1, 0, 0, 0]^{T}$ and $b = [1, 0, 0, 0, 1]^{T}$, $\rho = 1/4$. Compute the probability of observing the $i$th read to be $[0, 1, 0, 0, 1]^{T}$.

(2) (3 points) In general, write down the distribution of the $i$th read $R_i$ in terms of $a = [a_1, \ldots, a_L]^{T}$, $b = [b_1, \ldots, b_L]^{T}$ and $\rho$.

(3) (2 points) We want to estimate $\rho$ by first detecting for each read which is the underlying transcript ($a$ or $b$) and then using the decisions to estimate $\rho$. Explain why one cannot perform MAP detection for each read.

(4) (4 points) Instead we will perform ML detection. State the ML rule, and give expressions for the probability of error given the underlying transcript is $a$ and for the probability of error given the underlying transcript is $b$.

(5) (5 points) Suppose the ML decision of the $i$th read is $\hat{S}_i$. Explain how you would use $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_n$ to estimate $\rho$ such that the estimate converges to $\rho$ as $n \to \infty$. (Hint: part (a)(ii) may be useful here.)
Problem 6 (22 points)

Consider a continuous-time LTI system with frequency response $H(f) = 1$ if $|f| \leq 1$ and 0 otherwise. Suppose that the input to the system is zero-mean Gaussian white noise process $\{W(t)\}$ with constant power spectral density $N_0/2$ and let $\{Z(t)\}$ be the output of the system.

(a) (2 points) Find the variance of $Z(t_0)$ at any given time $t$.

(b) (4 points) Which of the following statements are correct?

(i) $Z(t_1)$ and $Z(t_2)$ are independent for all $t_1, t_2$ with $t_1 \neq t_2$.

(ii) $Z(t_1)$ and $Z(t_2)$ are dependent for all $t_1, t_2$.

(iii) $Z(t_1)$ and $Z(t_2)$ are independent for some $t_1, t_2$ and dependent for some $t_1, t_2$.

Fully justify your answer.

(c) We want to estimate a scalar signal $X \sim \mathcal{N}(0, 1)$ which is independent of $\{W(t)\}$. This scalar signal is modulated on a rectangular pulse:

$$s(t) = \begin{cases} 
1, & 0 < t < 1 \\
0, & \text{otherwise}
\end{cases}$$

The received signal $Y_1(t)$ is a noisy version of the signal corrupted by white noise $\{W(t)\}$:

$$Y_1(t) = Xs(t) + W(t).$$

(i) (2 points) Suppose we observe $Y_1(0.5)$. What is the MMSE in estimating $X$?

(ii) (6 points) Now suppose we observe $Y_1(t)$ from $t = -\infty$ to $t = \infty$. Propose an estimator of $X$ that would have lower MSE than in part (c)(i), and compute the MSE of your estimator.

(d) Now suppose the received signal $Y_2(t)$ is a noisy version of the signal in part (c) but is corrupted by the noise $\{Z(t)\}$:

$$Y_2(t) = Xs(t) + Z(t).$$

(i) (2 points) Suppose we observe $Y_2(0.5)$. What is the MMSE in estimating $X$? Compare and contrast with part (c)(i).

(ii) (6 points) Now suppose we observe in addition to $Y_2(0.5)$ also $Y_2(\tau)$, where $\tau > 1$. Can there be a time $\tau$ such that this addition observation decreases the estimation error in $X$? If so, choose $\tau$ to minimize the estimation error. If not, explain why not.
Problem 7 (22 points)

Let \( \{ U_n \}, \{ V_n \} \) be two independent i.i.d. \( N(0, 1) \) processes. The processes \( \{ X_n \} \) and \( \{ Y_n \} \) are described by the following equations:

\[
X_n = \alpha X_{n-1} + U_n, \\
Y_n = \beta X_n + V_n.
\]

for \( n \geq 1 \).

(a) (3 points) Under what conditions on the distribution of \( X_0 \) and on the values of \( \alpha \) and \( \beta \) are the processes \( \{ X_n \} \) and \( \{ Y_n \} \) jointly stationary? Fully justify your answer. We will assume these conditions are satisfied in the rest of the question.

(b) (3 points) Compute the cross-covariance function \( K_{XY}(m) \).

(c) (6 points) For any process \( \{ A_n \} \), let \( e_{A_n|A_n^{-1}} \) be the MMSE prediction error of \( A_n \) given the past \( A_n^{-1} = [A_0, A_1, \ldots, A_{n-1}]^T \). Compute \( e_{X_n|X_n^{-1}} \) and \( e_{Y_n|Y_n^{-1}} \). Do they depend on \( n \)? Explain.

(d) (4 points) For any two processes \( \{ A_n \} \) and \( \{ B_n \} \), let \( e_{A_n|A_n^{-1},B_n^{-1}} \) be the MMSE prediction error of \( A_n \) given the past \( A_n^{-1} \) and the past \( B_n^{-1} \). Compute \( e_{X_n|X_n^{-1},Y_n^{-1}} \) and \( e_{Y_n|Y_n^{-1},X_n^{-1}} \).

(e) (4 points) The Granger causality of a process \( \{ A_n \} \) on another process \( \{ B_n \} \) is defined to be:

\[
G(A \rightarrow B) := \lim_{n \to \infty} e_{B_n|B_n^{-1}} - e_{B_n|B_n^{-1},A_n^{-1}},
\]

assuming the limit exists. Compute \( G(X \rightarrow Y) \) and \( G(Y \rightarrow X) \). Are they equal?

(f) (2 points) Based on part (e), discuss why \( G(A \rightarrow B) \) is a reasonable measure of causality of the process \( \{ A_n \} \) on the process \( \{ B_n \} \), rather than say \( K_{AB}(m) \).