Homework #2
Due February 25

Notice: Please bring your completed homework with you to your appointment.

1. *Jointly Gaussian.* Given a Gaussian random vector $X \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = [\ 1\ \ 5\ \ 2\ ]^T$ and

$$\Sigma \ = \ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

   Find the distributions of the following random variables.

   a. X_1
   b. $X_2 + X_3$
   c. $2X_1 + X_2 - X_3$
   d. X_3 given (X_1, X_2)
   e. (X_2, X_3) given X_1
   f. X_1 given (X_2, X_3)
   g. AX where $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$.

2. *Noise cancelation.* A classic problem in statistical signal processing involves estimating a weak signal (e.g., the heart beat of a fetus) in the presence of a strong interference (the heart beat of its mother) by making two observations—one with the weak signal present and one without, by placing one microphone on the mothers belly and another close to her heart. The observations can then be combined to estimate the weak signal by "canceling out" the interference. The following is a simple version of this application.

   Let the weak signal X be a random variable with mean zero and variance P. Let the observations be $Y_1 = X + Z_1$ and $Y_2 = Z_1 + Z_2$, where Z_1 is the strong interference and Z_2 is measurement noise. Assume that Z_1 and Z_2 are zero-mean with variances N_1 and N_2, respectively. Further assume that X, Z_1, and Z_2 are uncorrelated.

   a. Find the best linear MSE estimate of X given Y_1 and Y_2 and the corresponding MSE. Interpret the results.
   b. How about estimating the signal in noise, without having a second signal to cancel some of the noise? Find the best linear MSE estimate of X given Y_1 and the corresponding MSE.

3. *Multiple looks with Gaussian noise.* Let $Y_i = X + Z_i$ for $i = 1, 2, ..., n$ be n observations of a signal $X \sim \mathcal{N}(0, P)$. The additive noise components $Z_1, Z_2, ..., Z_n$ are zero-mean jointly Gaussian random variables that are independent of X. For each of the following two noise correlations, find the best MSE estimate of X given $Y_1, Y_2, ..., Y_n$ and its MSE. It might be convenient to assume a form of the estimator and use the orthogonality principle to claim optimality.

   a. The noise components $Z_1, ..., Z_n$ are uncorrelated, each with variance N.
   b. The noise components $Z_1, ..., Z_n$ have correlation $\mathbf{E}(Z_i Z_j) = N 2^{-|i-j|}$ for $1 \le i, j \le n$. (Hint: try a linear estimator with coefficients that are of the form $h^T = [\ a\ \ b\ \ b\ ...\ b\ \ b\ \ a\ ]$.)

4. *Noisy measurements.* Consider noisy linear measurements Y of a Gaussian source $X \sim \mathcal{N}(0, \Sigma_X)$ corrupted by independent noise $W \sim \mathcal{N}(0, \Sigma_W)$, given by

$$Y = AX + W,$$

where A is a matrix.

a. What is the MMSE estimate of X given Y, and what is the MMSE?

b. State the estimator in part a. for the case $\Sigma_X = \sigma_X^2 I$ and $\Sigma_W = \sigma_W^2 I$.

c. Recall that the least-squares fit for an overdetermined linear system is given by $\hat{X} = (A^T A)^{-1} A^T Y$ (this is also the ML estimator when $\Sigma_W = I$) and the least-norm solution for an underdetermined linear system is given by $\hat{X} = A^T (AA^T)^{-1} Y$. Interpret these in comparison to the solutions to parts a. and b. (When the system is overdetermined, it may help to multiply the estimator by the term $(A^T A)^{-1} A^T A$.)

5. *Logistic regression from Poisson.* Suppose $X \sim Bern(p)$ and Y is conditionally a Poisson random variable with mean λ_0 if $X = 0$ and mean λ_1 if $X = 1$. That is, $Y \in \mathbf{N}$ has the following conditional probability mass function:

$$p(y|x=0) = \frac{\lambda_0^y}{y!} e^{-\lambda_0},$$
$$p(y|x=1) = \frac{\lambda_1^y}{y!} e^{-\lambda_1}.$$

a. What is the MMSE estimate of X given Y? Express this using the logistic function

$$h(x) = \frac{1}{1 + e^{-x}}.$$

Don't worry about calculating the MSE.

b. What is the estimator that minimizes the probability of error?

c. Without knowledge of the distribution, logistic regression attempts to form a similar estimator from training data consisting of pairs (X_i, Y_i) for $i = 1, ..., n$. The method does not place a distribution on the Y_i's but uses a model with parameters a and b (for the one dimensional case) where the distribution on $X_i \in \{0, 1\}$ is assumed to be $\mathbf{P}(X = 1) = h(aY_i + b)$, and the X_i's are independent. The parameters a and b are selected by maximum likelihood.

The result of logistic regression is an estimator that will be similar to the estimators in parts a. and b. Consider qualitatively the following question. What will happen to the logistic regression ML fit if all of the data point pairs in the training data where $X_i = 1$ are duplicated? For example, if half of the data had $X = 0$ and half had $X = 1$, then after duplication there will be 50% more data points, two-thirds of which will have $X_i = 1$. How does this relate to the Bernoulli-Poisson model above? (Don't worry about being very precise with this.)

6. *Logistic regression from Gaussian.* Suppose $X \sim Bern(p)$ and Y is conditionally a unit variance Gaussian random variable with mean -1 if $X = 0$ and mean 1 if $X = 1$. That is, $Y \in \Re$ has the following conditional probability distribution:

$$\begin{aligned} p(y|x=0) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y+1)^2}, \\ p(y|x=1) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-1)^2}. \end{aligned}$$

a. What is the MMSE estimate of X given Y? Express this using the logistic function

$$h(x) = \frac{1}{1+e^{-x}}.$$

Don't worry about calculating the MSE.

b. What is the MAP estimate of X given Y?

c. What is the minimum probability of error when estimating X as a function of Y? Plot this as a function of p. If you're curious, check what happens when you adjust the variance of Y conditioned on X, as well as p.