Homework 2

Due date: Friday, October 21

1. Let (X, Y) be a pair of random variables with some known joint distribution. In class, we have seen that for predicting Y from X, the estimator $T(X) = \hat{Y} = E(Y|X)$ is that which minimizes the Mean-Squared Error $E(\hat{Y} - Y)^2$. What is the estimator which now minimizes $E|\hat{Y} - Y|$?

2. Consider the simple stochastic process defined by

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \quad t = 1, \dots L;$$

here $\rho = 0.7$ and the ε_t's are i.i.d. standard normal.

   (a) It can be shown that $\mathcal{L}(Y_t|Y_{-t}) = \mathcal{L}(Y_t|Y_{t-1}, Y_{t+1})$, where $\mathcal{L}$ means *law* or "*the distribution of*", and the notation Y_{-t} means all the Y_s except Y_t (you will get some *extra credit* if you manage to prove this fact). Show that $\mathcal{L}(Y_t|Y_{t-1}, Y_{t+1}) = N(\mu_t, \gamma^2)$, and calculate μ_t and γ (*Hint:* First compute $\mathcal{L}(Y_t, Y_{t+1}|Y_{t-1})$ [easy], and then use the standard conditioning formula for multivariate Gaussians to get $\mathcal{L}(Y_t|Y_{t-1}, Y_{t+1})$.)

   (b) Show that $\mathcal{L}(Y_1, \dots, Y_L|Y_0 = y_0, Y_{L+1} = y_{L+1}) \sim N(\mu, \Sigma)$ where $\mu = \mu(y_0, y_{L+1})$ is a L vector, and $\Sigma = \Sigma(L, \rho)$ an $L \times L$ covariance matrix. Derive expressions for μ and Σ.

   (c) Using this distribution, sample a sequence with $L = 10$ subject to

$$y_0 = 1 > Y_1 > \dots > Y_L > y_{L+1} = 0; \tag{1}$$

   that is, we want to sample from the multivariate density

$$g(y) \propto f_Y(y) 1_{\{1>y_1>\dots>y_L>0\}}, \tag{2}$$

   where f_Y is the density of the multivariate normal $N(\mu, \Sigma)$ and where the constant of proportionality makes sure that g integrates up to one.

   You may use the rejection method for this question: you sample $Y_1, \dots, Y_L$ from the distribution $\mathcal{L}(Y_1, \dots, Y_L|Y_0 = y_0, Y_{L+1} = y_{L+1})$; if your vector satisfies the constraint (1) you accept the sample, otherwise repeat. Record the number of samples drawn. Repeat this procedure 100 times and record the total number of samples drawn.

   (d) Consider the alternative sampling strategy to simulate $Y_1, \dots, Y_L$ subject to (1). Start with the initial value

$$(y_1, y_2, \dots, y_L) = (L/(L+1), (L-1)/L + 1, \dots, 1/(L+1))$$

   Then

   - Step 1:

- update y_1 by drawing a sample from $\mathcal{L}(Y_1|Y_0 = 1, Y_2 = y_2)$ subject to the constraint (1).
- then update y_2 by drawing a sample from $\mathcal{L}(Y_2|Y_1 = y_1, Y_3 = y_3)$ subject to the constraint (1). Here y_1 is the new value (the one we just calculated) of Y_1.
- ...
- then update y_L by drawing a sample from $\mathcal{L}(Y_L|Y_{L-1} = y_{L-1}, Y_{L+1} = 0)$ subject to the constraint (1). Here y_{L-1} is the new value of Y_{L-1}.

- Step 2: Repeat Step 1 with the new sample as your initial value.
- Step 3: Stop this loop after B = 100 iterations. The sampled vector is the current value of $(y_1, y_2, \ldots, y_L)$.

It is possible to prove that the y has a limiting distribution, namely (2), as the number of iterations tends to infinity.

Write some code for implementing this new strategy. Sample a sequence with $L = 10$ subject to the constraint. Record the number of samples drawn. Repeat this procedure 100 times and record the total number of samples drawn.

(e) Which procedure is more efficient and why?

3. Consider the Gaussian process introduced in the last question:

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t = 1, \ldots T.$$

where $-1 < \rho < 1$ and the ε_t's are iid standard normal (We will take $X_0 = 0$ so that $X_1 = \varepsilon_1$.

(a) Find the covariance matrix Σ of X.

(b) Explain how you would generate a sample from X.

(c) In the remainder of this exercise, we will take $\rho = .8$. Generate a sample from the observed process Y

$$Y_t = X_t + 1/4W_t, \quad t = 1, \ldots T = 1024,$$

where X is as before and W is a standard white noise.

(d) Construct the Wiener Filter and plot X and your estimate $\hat{X}$ on a same figure. Calculate the estimation error

$$1/T \sum (X_t - \hat{X}_t)^2.$$

(e) Repeat (d) a thousand times (without plotting the results each time) and summarize your results with a histogram. Compute the average estimation error.

(f) Calculate the exact Mean Squared Error for this estimation problem and compare this value with your findings in the previous question. Comment.

(g) *Bonus Question:* What is the standard deviation of the average you calculated in (e)?