Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers
Roy D. Yates and David J. Goodman

Problem Solutions : Yates and Goodman,9.1.2 9.2.2 9.3.2 9.3.6 9.3.7 9.4.3 9.4.4 9.5.8 9.5.9 9.6.4 9.7.4 and 9.7.6

Problem 9.1.2

(a) We wish to develop a hypothesis test of the form

$$P[|K - E[K]| > c] = 0.05$$

to determine if the coin we've been flipping is indeed a fair one. We would like to find the value of c, which will determine the upper and lower limits on how many heads we can get away from the expected number out of 100 flips and still accept our hypothesis. Under our fair coin hypothesis, the expected number of heads, and the standard deviation of the process are

$$E[K] = 50$$
$$\sigma_K = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5$$

Now in order to find c we make use of the central limit theorem and divide the above inequality through by σ_K to arrive at

$$P\left[\frac{|K - E[K]|}{\sigma_k} > \frac{c}{\sigma_K}\right] = 0.05$$

Taking the complement, we get

$$P\left[-\frac{c}{\sigma_K} \le \frac{K - E[K]}{\sigma_k} \le \frac{c}{\sigma_K}\right] = 0.95$$

Using the Central Limit Theorem we can write

$$\Phi\left(\frac{c}{\sigma_K}\right) - \Phi\left(\frac{-c}{\sigma_K}\right) = 2\Phi\left(\frac{c}{\sigma_K}\right) - 1 = 0.95$$

This implies $\Phi(c/\sigma_K) = 0.975$ or $c/5 = 1.96$. That is, $c = 9.8$ flips. So we see that if we observe more then $50 + 10 = 60$ or less then $50 - 10 = 40$ heads, then with significance level $\alpha \approx 0.05$ we should reject the hypothesis that the coin is fair.

(b) Now we wish to develop a test of the form

$$P[K > c] = 0.01$$

Thus we need to find the value of c that makes the above probability true. This value will tell us that if we observe more than c heads, then with significance level $\alpha = 0.01$, we should reject the hypothesis that the coin is fair. To find this value of c we look to evaluate the CDF

$$F_K(k) = \begin{cases} 0 & k < 0 \\ \sum_{i=0}^{k} \binom{100}{i} (1/2)^{100} & k = 0, 1, \ldots, 100 \\ 1 & k > 100 \end{cases}$$

Computation reveals that $c \approx 62$ flips. So if we observe 62 or greater heads, then with a significance level of 0.01 we should reject the fair coin hypothesis. Another way to obtain this result is to use a Central Limit Theorem approximation. First, we express our rejection region in terms of a zero mean, unit variance random variable.

$$P[K > c] = 1 - P[K \le c] = 1 - P\left[\frac{K - E[K]}{\sigma_K} \le \frac{c - E[K]}{\sigma_K}\right] = 0.01$$

Since $E[K] = 50$ and $\sigma_K = 5$, the CLT approximation is

$$P[K > c] \approx 1 - \Phi\left(\frac{c - 50}{5}\right) = 0.01$$

From Table 4.1, we have $(c - 50)/5 = 2.35$ or $c = 61.75$. Once again, we see that we reject the hypothesis if we observe 62 or more heads.

Problem 9.2.2

Given hypothesis H_i, K has the binomial PMF

$$P_{K|H_i}(k) = \begin{cases} \binom{n}{k} q_i^k (1 - q_i)^{n-k} & k = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

(a) The ML rule is

- $k \in A_1$ if $P_{K|H_1}(k) \ge P_{K|H_0}(k)$.
- $k \in A_0$ if $P_{K|H_1}(k) < P_{K|H_0}(k)$.

When we observe $K = k \in \{0, 1, \ldots, n\}$, plugging in the conditional PMF's yields the rule

- $k \in A_1$ if $\binom{n}{k} q_1^k (1 - q_1)^{n-k} \ge \binom{n}{k} q_0^k (1 - q_0)^{n-k}$
- $k \in A_0$ if $\binom{n}{k} q_1^k (1 - q_1)^{n-k} < \binom{n}{k} q_0^k (1 - q_0)^{n-k}$

Cancelling common factors, taking the logarithm of both sides, and rearranging yields

- $k \in A_1$ if $k \ln q_1 + (n - k) \ln(1 - q_1) \ge k \ln q_0 + (n - k) \ln(1 - q_0)$
- $k \in A_0$ if $k \ln q_1 + (n - k) \ln(1 - q_1) < k \ln q_0 + (n - k) \ln(1 - q_0)$

By combining all terms with k, the rule can be simplified to

- $k \in A_1$ if $k \ln\left(\frac{q_1/(1-q_1)}{q_0/(1-q_0)}\right) \ge n \ln\left(\frac{1-q_0}{1-q_1}\right)$
- $k \in A_0$ if $k \ln\left(\frac{q_1/(1-q_1)}{q_0/(1-q_0)}\right) < n \ln\left(\frac{1-q_0}{1-q_1}\right)$

Note that $q_1 > q_0$ implies $q_1/(1 - q_1) > q_0/(1 - q_0)$. Thus, we can rewrite our ML rule as

- $k \in A_1$ if $k \ge k^* = n \frac{\ln[(1-q_0)/(1-q_1)]}{\ln[q_1/q_0] + \ln[(1-q_0)/(1-q_1)]}$
- $k \in A_0$ if $k < k^*$

(b) Let k^* denote the threshold given in part (a). Using $n = 500$, $q_0 = 10^{-4}$, and $q_1 = 10^{-2}$, we have

$$k^* = 500 \frac{\ln[(1-10^{-4})/(1-10^{-2})]}{\ln[10^{-2}/10^{-4}] + \ln[(1-10^{-4})/(1-10^{-2})]} \approx 1.078$$

Thus the ML rule is that if we observe $K \leq 1$, then we choose hypothesis H_0; otherwise, we choose H_1. The conditional error probabilities are

$$\begin{aligned} P[E|H_0] = P[A_1|H_0] &= P[K > 1|H_0] \\ &= 1 - P_{K|H_0}(0) - P_{K|H_1}(1) \\ &= 1 - (1-q_0)^{500} - 500q_0(1-q_0)^{499} = 0.0012 \end{aligned}$$

and

$$\begin{aligned} P[E|H_1] = P[A_0|H_1] &= P[K \leq 1|H_1] \\ &= P_{K|H_1}(0) + P_{K|H_1}(1) \\ &= (1-q_1)^{500} + 500q_1(1-q_1)^{499} = 0.0398 \end{aligned}$$

(c) In the test of Example 9.7, the geometric random variable N, the number of tests needed to find the first failure, was used. In this problem, the binomial random variable K, the number of failures in 500 tests, was used. We will call these two procedures the geometric and the binomial tests. Also, we will use $P[E_N|H_i]$ and $P[E_K|H_i]$ to denote the respective conditional error probabilities. From Example 9.7, we have the following comparison:

geometric test	binomial test
$P[E_N|H_0] = 0.045$	$P[E_K|H_0] = 0.0012$
$P[E_N|H_1] = 0.0095$	$P[E_K|H_1] = 0.0398$

When making comparisons between tests, we want to judge both the reliability of the test as well as the cost of the testing procedure. With respect to the cost of the test, the geometric test is guaranteed to never require more than 464 tests because if we observe 464 consecutive working devices, then we choose H_0. Moreover, the geometric test may require far fewer than 464 tests, particularly if hypothesis H_1 happens to be true. On the other hand, the binomial test always uses 500 tests. Consequently, the geometric test is likely to cost a little less, although the difference may not be very significant.

When we examine the test reliability, we see that the conditional error probabilities appear to be comparable in that $P[E_K|H_0] \ll P[E_N|H_0]$ while $P[E_N|H_1] \ll P[E_K|H_1]$. If we knew the a priori probabilities $P[H_i]$ and also the relative costs of the two type of errors, then we could determine which test procedure was better. However, in the absence of that kind of information, we make the following observation. Given that the product is a pacemaker whose installation requires heart surgery, we would like the device to be very very reliable. Hence, if H_1 is true and the failure rate is $q_1 = 0.01$, we would like very much to know this fact before the pacemaker is available for sale. On the other hand, if H_0 is true and we make an error and guess that the failure rate is $q_1 = 0.01$, then we may go back and redesign either the pacemaker or the assembly process to improve reliability. It seems likely that this cost may be significantly lower than the cost of installing many faulty pacemakers. These arguments suggest that the conditional error probabilities given H_1 are far more important. On this basis, it would appear that the geometric test is a better test.

Problem 9.3.2

Let the components of $\mathbf{s}_{ijk}$ be denoted by $s_{ijk}^{(1)}$ and $s_{ijk}^{(2)}$ so that given hypothesis H_{ijk},

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} s_{ijk}^{(1)} \\ s_{ijk}^{(2)} \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \end{bmatrix}$$

As in Example 9.9, we will assume N_1 and N_2 are iid zero mean Gaussian random variables with variance σ^2. Thus, given hypothesis H_{ijk}, X_1 and X_2 are independent and the conditional joint PDF of X_1 and X_2 is

$$\begin{aligned} f_{X_1,X_2|H_{ijk}}(x_1,x_2) &= f_{X_1|H_{ijk}}(x_1)\, f_{X_2|H_{ijk}}(x_2) \\ &= \frac{1}{2\pi\sigma^2} e^{-(x_1-s_{ijk}^{(1)})^2/2\sigma^2} e^{-(x_2-s_{ijk}^{(2)})^2/2\sigma^2} \\ &= \frac{1}{2\pi\sigma^2} e^{-[(x_1-s_{ijk}^{(1)})^2+(x_2-s_{ijk}^{(2)})^2]/2\sigma^2} \end{aligned}$$

In terms of the distance $\left\|\mathbf{x}-\mathbf{s}_{ijk}\right\|$ between vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mathbf{s}_{ijk} = \begin{bmatrix} s_{ijk}^{(1)} \\ s_{ijk}^{(2)} \end{bmatrix}$$

we can write

$$f_{X_1,X_2|H_i}(x_1,x_2) = \frac{1}{2\pi\sigma^2} e^{-\left\|\mathbf{x}-\mathbf{s}_{ijk}\right\|^2/2\sigma^2}$$

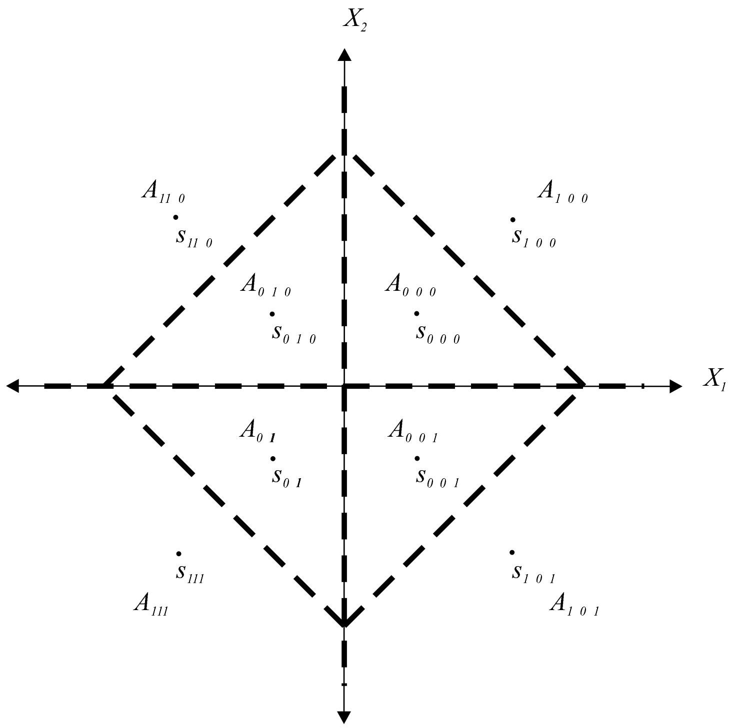
Since all eight symbols $s_{000},\ldots,s_{111}$ are equally likely, the MAP and ML rules are

- $\mathbf{x} \in A_{ijk}$ if $f_{X_1,X_2|H_{ijk}}(x_1,x_2) \geq f_{X_1,X_2|H_{i'j'k'}}(x_1,x_2)$ for all other hypotheses $H_{i'j'k'}$.

This rule simplifies to

- $\mathbf{x} \in A_{ijk}$ if $\left\|\mathbf{x}-\mathbf{s}_{ijk}\right\| \leq \left\|\mathbf{x}-\mathbf{s}_{ijk}\right\|$ for all other $i'j'k'$.

This means that A_{ijk} is the set of all vectors $\mathbf{x}$ that are closer to $\mathbf{s}_{ijk}$ than any other signal. Graphically, to find the boundary between points closer to $\mathbf{s}_{ijk}$ than $\mathbf{s}_{i'j'k'}$, we draw the line segment connecting $\mathbf{s}_{ijk}$ and $\mathbf{s}_{i'j'k'}$. The boundary is then the perpendicular bisector. The resulting boundaries are shown in the following figure:

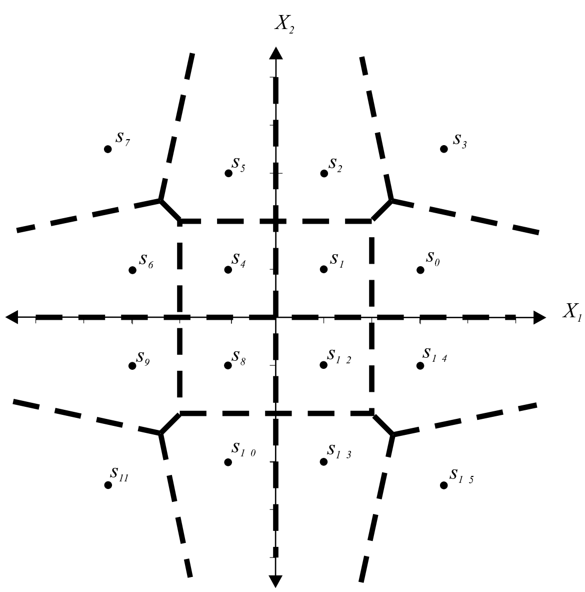


Problem 9.3.6

(a) In Problem 9.3.4, we found that in terms of the vectors $\mathbf{x}$ and $\mathbf{s}_i$, the acceptance regions are defined by the rule

- $\mathbf{x} \in A_i$ if $\|\mathbf{x}-\mathbf{s}_i\|^2 \leq \left\|\mathbf{x}-\mathbf{s}_j\right\|^2$ for all j

Just as in the case of QPSK, the acceptance region A_i is the set of vectors $\mathbf{x}$ that are closest to $\mathbf{s}_i$. Graphically, these regions are easily found from the sketch of the signal constellation:



(b) For hypothesis A_1, we see that the acceptance region is

$$A_1 = \{(X_1, X_2) | 0 < X_1 \le 2, 0 < X_2 \le 2\}$$

Given H_1, a correct decision is made if $(X_1, X_2) \in A_1$. Given H_1, $X_1 = 1 + N_1$ and $X_2 = 1 + N_2$. Thus,

$$\begin{aligned} P[C|H_1] &= P[(X_1, X_2) \in A_1 | H_1] \\ &= P[0 < 1 + N_1 \le 2, 0 < 1 + N_2 \le 2] \\ &= (P[-1 < N_1 \le 1])^2 \\ &= (\Phi(1/\sigma_N) - \Phi(-1/\sigma_N))^2 \\ &= (2\Phi(1/\sigma_N) - 1)^2 \end{aligned}$$

(c) Surrounding each signal s_i is an aceptance region A_i that is no smaller than the acceptance region A_1. That is,

$$\begin{aligned} P[C|H_i] &= P[(X_1, X_2) \in A_i | H_1] \\ &\ge P[-1 < N_1 \le 1, -1 < N_2 \le 1] \\ &= (P[-1 < N_1 \le 1])^2 \\ &= P[C|H_1] \end{aligned}$$

This implies

$$P[C] = \sum_{i=0}^{15} P[C|H_i]P[H_1] \ge \sum_{i=0}^{15} P[C|H_1]P[H_i] = P[C|H_1] \sum_{i=0}^{15} P[H_i] = P[C|H_1]$$

Problem 9.3.7

Theorem 9.7, the MAP multiple hypothesis test is

- $(x_1, x_2) \in A_i$ if $p_i f_{X_1,X_2|H_i}(x_1, x_2) \ge p_j f_{X_1,X_2|H_j}(x_1, x_2)$ for all j

From Example 9.9, the conditional PDF of X_1, X_2 given H_i is

$$f_{X_1,X_2|H_i}(x_1, x_2) = \frac{1}{2\pi\sigma^2} e^{-[(x_1 - \sqrt{E}\cos\theta_i)^2 + (x_2 - \sqrt{E}\sin\theta_i)^2]/2\sigma^2}$$

Using this conditional joint PDF, the MAP rule becomes

- $(x_1, x_2) \in A_i$ if for all j

$$\frac{-(x_1 - \sqrt{E}\cos\theta_i)^2 - (x_2 - \sqrt{E}\sin\theta_i)^2 + (x_1 - \sqrt{E}\cos\theta_j)^2 + (x_2 - \sqrt{E}\sin\theta_j)^2}{2\sigma^2} \ge \ln \frac{p_j}{p_i}$$

Expanding the squares and using the identity $\cos^2\theta + \sin^2\theta = 1$ yields the simplified rule

- $(x_1, x_2) \in A_i$ if for all j,

$$x_1[\cos\theta_i - \cos\theta_j] + x_2[\sin\theta_i - \sin\theta_j] \ge \frac{\sigma^2}{\sqrt{E}} \ln \frac{p_j}{p_i}$$

Note that the MAP rules define linear constraints in x_1 and x_2. Since $\theta_i = \pi/4 + i\pi/2$, we use the following table to enumerate the constraints:

	$\cos\theta_i$	$\sin\theta_i$
$i=0$	$1/\sqrt{2}$	$1/\sqrt{2}$
$i=1$	$-1/\sqrt{2}$	$1/\sqrt{2}$
$i=2$	$-1/\sqrt{2}$	$-1/\sqrt{2}$
$i=3$	$1/\sqrt{2}$	$-1/\sqrt{2}$

To be explicit, to determine whether $(x_1, x_2) \in A_i$, we need to check the MAP rule for each $j \neq i$. Thus, each A_i is defined by three constraints. Using the above table, the acceptance regions are

- $(x_1, x_2) \in A_0$ if

$$x_1 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_1}{p_0} \qquad x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_3}{p_0} \qquad x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_0}$$

- $(x_1, x_2) \in A_1$ if

$$x_1 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_1}{p_0} \qquad x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_1} \qquad -x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_3}{p_1}$$

- $(x_1, x_2) \in A_2$ if

$$x_1 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_3} \qquad x_2 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_1} \qquad x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_0}$$

- $(x_1, x_2) \in A_3$ if

$$x_1 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_3} \qquad x_2 \leq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_3}{p_0} \qquad -x_1 + x_2 \geq \frac{\sigma^2}{\sqrt{2E}} \ln \frac{p_2}{p_3}$$

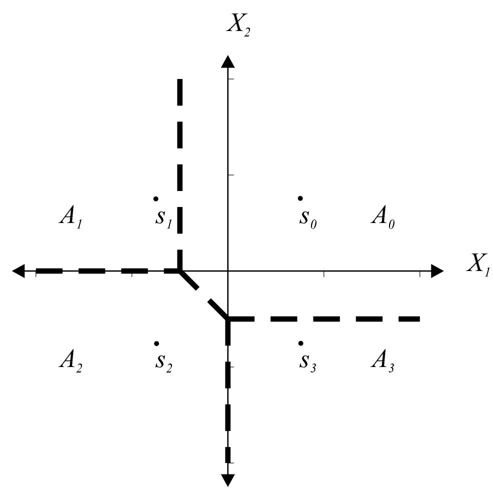
Using the parameters

$$\sigma = 0.8 \qquad E = 1 \qquad p_0 = 1/2 \qquad p_1 = p_2 = p_3 = 1/6$$

the acceptance regions for the MAP rule are

$$\begin{aligned}
A_0 &= \{(x_1, x_2) | x_1 \geq -0.497, x_2 \geq -0.497, x_1 + x_2 \geq -0.497\} \\
A_1 &= \{(x_1, x_2) | x_1 \leq -0.497, x_2 \geq 0, -x_1 + x_2 \geq 0\} \\
A_2 &= \{(x_1, x_2) | x_1 \leq 0, x_2 \leq 0, x_1 + x_2 \geq -0.497\} \\
A_3 &= \{(x_1, x_2) | x_1 \geq 0, x_2 \leq -0.497, -x_1 + x_2 \geq 0\}
\end{aligned}$$

Here is a sketch of these acceptance regions:



Note that the boundary between A_1 and A_3 defined by $-x_1+x_2 \geq 0$ plays no role because of the high value of p_0.

Problem 9.4.3

(a) The marginal PMFs of X and Y are listed below

$$P_X(x) = \begin{cases} 1/3 & x = -1,0,1 \\ 0 & \text{otherwise} \end{cases} \qquad P_Y(y) = \begin{cases} 1/4 & y = -3,-1,0,1,3 \\ 0 & \text{otherwise} \end{cases}$$

(b) No, the random variables X and Y are not independent since

$$P_{X,Y}(1,-3) = 0 \neq P_X(1)\,P_Y(-3)$$

(c) Direct evaluation leads to

$$E[X] = 0 \qquad \text{Var}[X] = 2/3$$
$$E[Y] = 0 \qquad \text{Var}[Y] = 5$$

This implies

$$\sigma_{X,Y} = \text{Cov}[X,Y] = E[XY] - E[X]E[Y] = E[XY] = 7/6$$

(d) From Theorem 9.11, the optimal linear estimate of X given Y is

$$\hat{X}_L(Y) = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(Y-\mu_Y)+\mu_X = \frac{7}{30}Y+0$$

Therefore, $a^* = 7/30$ and $b^* = 0$.

(e) The conditional probability mass function is

$$P_{X|Y}(x|-3) = \frac{P_{X,Y}(x,-3)}{P_Y(-3)} = \begin{cases} \frac{1/6}{1/4} = 2/3 & x = -1 \\ \frac{1/12}{1/4} = 1/3 & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

(f) The minimum mean square estimator of X given that $Y = 3$ is

$$\hat{x}_M(-3) = E[X|Y = -3] = \sum_x x P_{X|Y}(x|-3) = -2/3$$

(g) The mean squared error of this estimator is

$$\begin{aligned}\hat{e}_M(-3) = E\left[(X - \hat{x}_M(-3))^2 | Y = -3\right] &= \sum_x (x + 2/3)^2 P_{X|Y}(x|-3) \\ &= (-1/3)^2(2/3) + (2/3)^2(1/3) = 2/9\end{aligned}$$

Problem 9.4.4

to each other. In particular, completing the row sums and column sums shows that each random variable has the same marginal PMF. That is,

$$\begin{aligned}P_X(x) &= P_Y(x) = P_U(x) = P_V(x) = P_S(x) = P_T(x) = P_Q(x) = P_R(x) \\ &= \begin{cases} 1/3 & x = -1, 0, 1 \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

This implies

$$E[X] = E[Y] = E[U] = E[V] = E[S] = E[T] = E[Q] = E[R] = 0$$

and that

$$E\left[X^2\right] = E\left[Y^2\right] = E\left[U^2\right] = E\left[V^2\right] = E\left[S^2\right] = E\left[T^2\right] = E\left[Q^2\right] = E\left[R^2\right] = 2/3$$

Since each random variable has zero mean, the second moment equals the variance. Also, the standard deviation of each random variable is $\sqrt{2/3}$. These common properties will make it much easier to answer the questions.

(a) Random variables X and Y are independent since for all x and y,

$$P_{X,Y}(x, y) = P_X(x) P_Y(y)$$

Since each other pair of random variables has the same marginal PMFs as X and Y but a different joint PMF, all of the other pairs of random variables must be dependent. Since X and Y are independent, $\rho_{X,Y} = 0$. For the other pairs, we must compute the covariances.

$$\begin{aligned}\text{Cov}[U, V] &= E[UV] = (1/3)(-1) + (1/3)(-1) = -2/3 \\ \text{Cov}[S, T] &= E[ST] = 1/6 - 1/6 + 0 + -1/6 + 1/6 = 0 \\ \text{Cov}[Q, R] &= E[QR] = 1/12 - 1/6 - 1/6 + 1/12 = -1/6\end{aligned}$$

The correlation coefficient of U and V is

$$\rho_{U,V} = \frac{\text{Cov}[U, V]}{\sqrt{\text{Var}[U]}\sqrt{\text{Var}[V]}} = \frac{-2/3}{\sqrt{2/3}\sqrt{2/3}} = -1$$

In fact, since the marginal PMF's are the same, the denominator of the correlation coefficient will be $2/3$ in each case. The other correlation coefficients are

$$\rho_{S,T} = \frac{\text{Cov}[S, T]}{2/3} = 0 \qquad \rho_{Q,R} = \frac{\text{Cov}[Q, R]}{2/3} = -1/4$$

(b) From Theorem 9.11, the least mean square linear estimator of U given V is

$$\hat{U}_L(V) = \rho_{U,V}\frac{\sigma_U}{\sigma_V}(V - E[V]) + E[U] = \rho_{U,V}V = -V$$

Similarly for the other pairs, all expected values are zero and the ratio of the standard deviations is always 1. Hence,

$$\hat{X}_L(Y) = \rho_{X,Y}Y = 0$$
$$\hat{S}_L(T) = \rho_{S,T}T = 0$$
$$\hat{Q}_L(R) = \rho_{Q,R}R = -R/4$$

From Theorem 9.11, the mean square errors are

$$e_L^*(X,Y) = \text{Var}[X](1-\rho_{X,Y}^2) = 2/3$$
$$e_L^*(U,V) = \text{Var}[U](1-\rho_{U,V}^2) = 0$$
$$e_L^*(S,T) = \text{Var}[S](1-\rho_{S,T}^2) = 2/3$$
$$e_L^*(Q,R) = \text{Var}[Q](1-\rho_{Q,R}^2) = 5/8$$

Problem 9.5.8

The minimum mean square error linear estimator is given by Theorem 9.11 in which X_n and Y_{n-1} play the roles of X and Y in the theorem. That is, our estimate $\hat{X}_n$ of X_n is

$$\hat{X}_n = \hat{X}_L(Y_{n-1}) = \rho_{X_n,Y_{n-1}}\left(\frac{\text{Var}[X_n]}{\text{Var}[Y_{n-1}]}\right)^{1/2}(Y_{n-1} - E[Y_{n-1}]) + E[X_n]$$

By recursive application of $X_n = cX_{n-1} + Z_{n-1}$, we obtain

$$X_n = a^n X_0 + \sum_{j=1}^{n} a^{j-1} Z_{n-j}$$

The expected value of X_n is $E[X_n] = a^n E[X_0] + \sum_{j=1}^{n} a^{j-1} E\left[Z_{n-j}\right] = 0$. The variance of X_n is

$$\text{Var}[X_n] = a^{2n}\,\text{Var}[X_0] + \sum_{j=1}^{n}[a^{j-1}]^2\,\text{Var}\left[Z_{n-j}\right] = a^{2n}\,\text{Var}[X_0] + \sigma^2\sum_{j=1}^{n}[a^2]^{j-1}$$

Since $\text{Var}[X_0] = \sigma^2/(1-c^2)$, we obtain

$$\text{Var}[X_n] = \frac{c^{2n}\sigma^2}{1-c^2} + \frac{\sigma^2(1-c^{2n})}{1-c^2} = \frac{\sigma^2}{1-c^2}$$

Note that $E[Y_{n-1}] = dE[X_{n-1}] + E[W_n] = 0$. The variance of Y_{n-1} is

$$\text{Var}[Y_{n-1}] = d^2\,\text{Var}[X_{n-1}] + \text{Var}[W_n] = \frac{d^2\sigma^2}{1-c^2} + \eta^2$$

Since X_n and Y_{n-1} have zero mean, the covariance of X_n and Y_{n-1} is

$$\text{Cov}[X_n, Y_{n-1}] = E[X_n Y_{n-1}] = E[(cX_{n-1} + Z_{n-1})(dX_{n-1} + W_{n-1})]$$

From the problem statement, we learn that

$$E[X_{n-1}W_{n-1}] = 0 \qquad E[X_{n-1}]E[W_{n-1}] = 0$$
$$E[Z_{n-1}X_{n-1}] = 0 \qquad E[Z_{n-1}W_{n-1}] = 0$$

Hence, the covariance of X_n and Y_{n-1} is

$$\text{Cov}[X_n, Y_{n-1}] = cd\,\text{Var}[X_{n-1}]$$

The correlation coefficient of X_n and Y_{n-1} is

$$\rho_{X_n,Y_{n-1}} = \frac{\text{Cov}[X_n, Y_{n-1}]}{\sqrt{\text{Var}[X_n]\,\text{Var}[Y_{n-1}]}}$$

Since $E[Y_{n-1}]$ and $E[X_n]$ are zero, the linear predictor for X_n becomes

$$\hat{X}_n = \rho_{X_n,Y_{n-1}} \left(\frac{\text{Var}[X_n]}{\text{Var}[Y_{n-1}]} \right)^{1/2} Y_{n-1} = \frac{\text{Cov}[X_n, Y_{n-1}]}{\text{Var}[Y_{n-1}]} Y_{n-1} = \frac{cd\,\text{Var}[X_{n-1}]}{\text{Var}[Y_{n-1}]} Y_{n-1}$$

Substituting the above result for $\text{Var}[X_n]$, we obtain the optimal linear predictor of X_n given Y_{n-1}.

$$\hat{X}_n = \frac{c}{d} \frac{1}{1+\beta^2(1-c^2)} Y_{n-1}$$

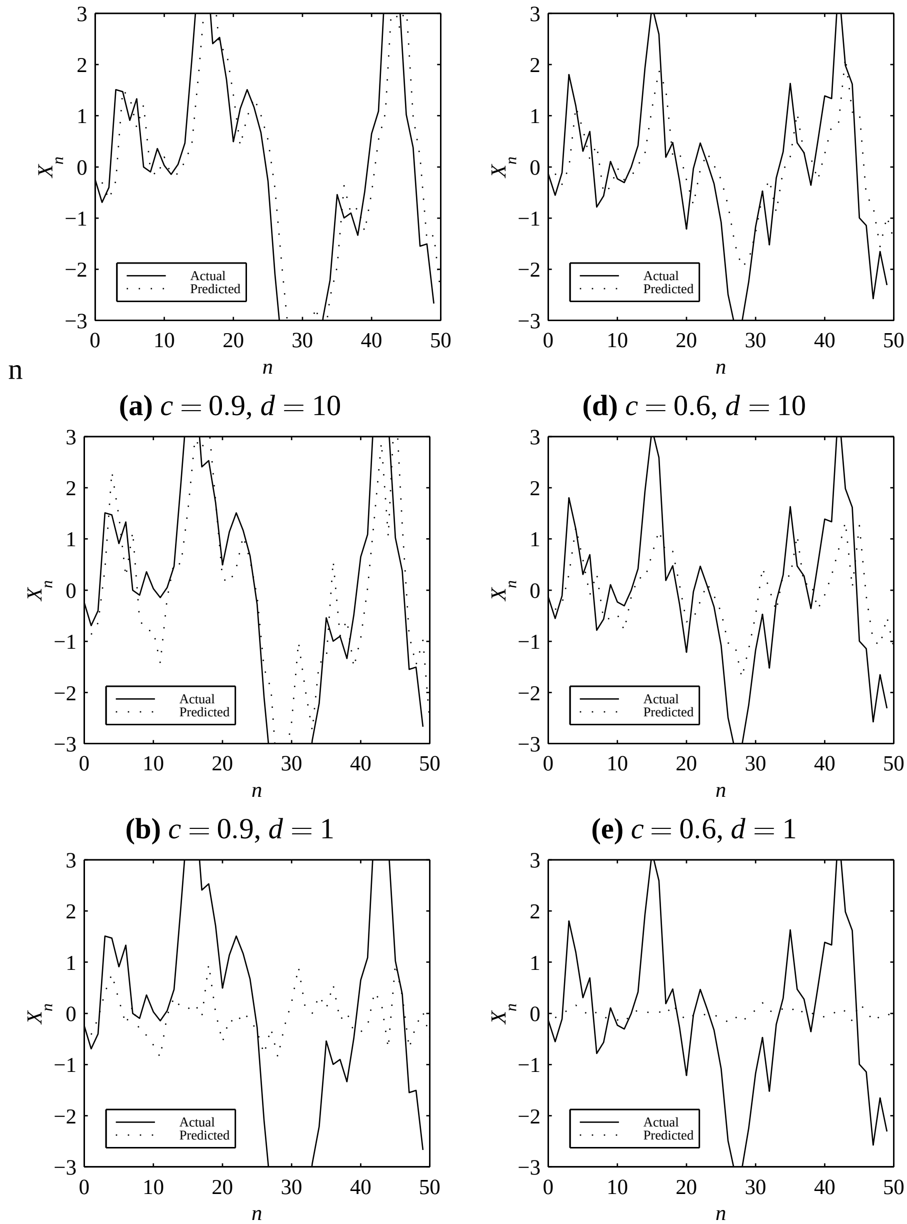
where $\beta^2 = \eta^2/(d^2\sigma^2)$. From Theorem 9.11, the mean square estimation error at step n

$$e_L^*(n) = E\left[(X_n - \hat{X}_n)^2\right] = \text{Var}[X_n] \left(1 - \rho^2_{X_n,Y_{n-1}}\right) = \sigma^2 \frac{1+\beta^2}{1+\beta^2(1-c^2)} \tag{1}$$

We see that mean square estimation error $e_L^*(n) = e_L^*$, a constant for all n. In addition, e_L^* is an increasing function β.

Problem 9.5.9

requested parameters are:



(a) $c = 0.9$, $d = 10$ **(d)** $c = 0.6$, $d = 10$

(b) $c = 0.9$, $d = 1$ **(e)** $c = 0.6$, $d = 1$

(c) $c = 0.9$, $d = 0.1$ **(f)** $c = 0.6$, $d = 0.1$

For $\sigma = \eta = 1$, the solution to Problem 9.5.8 showed that the optimal linear predictor of X_n given Y_{n-1} is

$$\hat{X}_n = \frac{cd}{d^2 + (1 - c^2)} Y_{n-1}$$

The mean square estimation error at step n was found to be

$$e_L^*(n) = e_L^* = \sigma^2 \frac{d^2 + 1}{d^2 + (1 - c^2)}$$

We see that the mean square estimation error is $e_L^*(n) = e_L^*$, a constant for all n. In addition, e_L^* is a decreasing function of d. In graphs (a) through (c), we see that the predictor tracks X_n less well as β increases. Decreasing d corresponds to decreasing the contribution of X_{n-1} to the measurement Y_{n-1}. Effectively, the impact of measurement noise variance η^2 is increased. As d decreases, the predictor places less emphasis on the measurement Y_n and instead makes predictions closer to $E[X] = 0$. That is, when d is small in graphs (c) and (f), the predictor stays close to zero. With respect to c, the performance of the predictor is less easy to understand. In Equation 3, the mean square error e_L^* is the product of

$$\operatorname{Var}[X_n] = \frac{\sigma^2}{1-c^2} \qquad 1-\rho^2_{X_n,Y_{n-1}} = \frac{(d^2+1)(1-c^2)}{d^2+(1-c^2)}$$

As a function of increasing c^2, $\operatorname{Var}[X_n]$ increases while $1-\rho^2_{X_n,Y_{n-1}}$ decreases. Overall, the mean square error e_L^* is an increasing function of c^2. However, $\operatorname{Var}[X]$ is the mean square error obtained using a blind estimator that always predicts $E[X]$ while $1-\rho^2_{X_n,Y_{n-1}}$ characterizes the extent to which the optimal linear predictor is better than the blind predictor. When we compare graphs (a)-(c) with $a = 0.9$ to graphs (d)-(f) with $a = 0.6$, we see greater variation in X_n for larger a but in both cases, the predictor worked well when d was large.

Note that the performance of our predictor is limited by the fact that it is based on a single observation Y_{n-1}. Generally, we can improve our predictor when we use all of the past observations $Y_0, \ldots, Y_{n-1}$.

Problem 9.6.4

Example 9.18.

(a) Given $Q = q$, the conditional PMF of K is

$$P_{K|Q}(k|q) = \begin{cases} \binom{n}{k} q^k (1-q)^{n-k} & k = 0,1,\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

The ML estimate of Q given $K = k$ is

$$\hat{q}_{\text{ML}}(k) = \arg\max_{0 \le q \le 1} P_{Q|K}(q|k)$$

Differentiating $P_{Q|K}(q|k)$ with respect to q and setting equal to zero yields

$$\frac{dP_{Q|K}(q|k)}{dq} = \binom{n}{k} \left(kq^{k-1}(1-q)^{n-k} - (n-k)q^k(1-q)^{n-k-1} \right) = 0$$

The maximizing value is $q = k/n$ so that

$$\hat{Q}_{\text{ML}}(K) = \frac{K}{n}$$

(b) To find the PMF of K, we average over all q.

$$P_K(k) = \int_{-\infty}^{\infty} P_{K|Q}(k|q) f_Q(q)\, dq = \int_0^1 \binom{n}{k} q^k (1-q)^{n-k}\, dq$$

We can evaluate this itegral by expressing it in terms of the integral of a beta PDF. Since $\beta(k+1,n-k+1) = \frac{(n+1)!}{k!(n-k)!}$, we can write

$$P_K(k) = \frac{1}{n+1}\int_0^1 \beta(k+1,n-k+1)q^k(1-q)^{n-k}\,dq = \frac{1}{n+1}$$

That is, K has the uniform PMF

$$P_K(k) = \begin{cases} 1/(n+1) & k=0,1,\ldots,n \\ 0 & \text{otherwise} \end{cases}$$

Hence, $E[K] = n/2$.

(c) The conditional PDF of Q given K is

$$f_{Q|K}(q|k) = \frac{P_{K|Q}(k|q)\, f_Q(q)}{P_K(k)} = \begin{cases} \frac{(n+1)!}{k!(n-k)!}q^k(1-q)^{n-k} & 0 \le q \le 1 \\ 0 & \text{otherwise} \end{cases}$$

That is, given $K=k$, Q has a beta $(k+1,n-k+1)$ PDF.

(d) The MMSE estimate of Q given $K=k$ is the conditional expectation $E[Q|K=k]$. From the beta PDF described in Appendix A, $E[Q|K=k] = (k+1)/(n+2)$. The MMSE estimator is

$$\hat{Q}_M(K) = E[Q|K] = \frac{K+1}{n+2}$$

Problem 9.7.4

(a) This part is just algebra and doesn't require any probabilities. By expanding the square, the random variable V_n can be written as

$$\begin{aligned}
V_n &= \frac{1}{n}\sum_{i=1}^{n}\left[X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n}X_j + \frac{1}{n^2}\left(\sum_{j=1}^{n}X_j\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \frac{2}{n^2}\left(\sum_{i=1}^{n}X_i\right)\sum_{j=1}^{n}X_j + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n^2}\left(\sum_{j=1}^{n}X_j\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \frac{2}{n^2}\left(\sum_{i=1}^{n}X_j\right)^2 + \frac{1}{n^2}\left(\sum_{j=1}^{n}X_j\right)^2
\end{aligned}$$

(b) From the previous part,

$$V_n = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - [M_n(X)]^2$$

Taking expectations, we have

$$E[V_n] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2\right] - E\left[(M_n(X))^2\right] = E\left[X_i^2\right] - E\left[(M_n(X))^2\right]$$

Since we know the mean and variance of the sample mean $M_n(X)$, we can calculate the second moment of the sample mean.

$$E\left[(M_n(X))^2\right] = \text{Var}[M_n(X)] + (E[M_n(X)])^2 = \frac{\text{Var}[X]}{n} + (E[X])^2$$

This implies

$$E[V_n] = E[X]^2 - \frac{\text{Var}[X]}{n} - (E[X])^2 = \text{Var}[X] - \frac{\text{Var}[X]}{n} = \frac{n-1}{n}\text{Var}[X]$$

Problem 9.7.6

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$$

The joint PDF of $X_1, \ldots, X_n$ is

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-(x_1^2+\cdots+x_n^2)/2\sigma^2}$$

(a) The ML estimate of σ, maximizes the joint PDF $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$. We find this by taking the derivative of the joint PDF with respect to σ. Using $w^2 = x_1^2 + \cdots + x_n^2$ to simplify our expressions, we have

$$\frac{d\, f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{d\sigma} = \frac{\sigma^n e^{-w^2/2\sigma^2}(w^2/\sigma^3) - e^{-w^2/2\sigma^2} n\sigma^{n-1}}{(2\pi)^{n/2}\sigma^{2n}} = 0$$

Solving for σ yields

$$\sigma = \frac{\sqrt{w^2}}{\sqrt{n}} = \frac{\sqrt{x_1^2 + \cdots + x_n^2}}{\sqrt{n}}$$

The ML estimator is

$$\hat{\sigma}_{\text{ML}}(n) = \frac{\sqrt{X_1^2 + \cdots + X_n^2}}{\sqrt{n}}$$

(b) First we observe that

$$V_n = \hat{\sigma}_{\text{ML}}^2(n) = \frac{X_1^2 + \cdots + X_n^2}{n}$$

Since $E[X_i] = 0$, we know that $E\left[X_i^2\right] = \sigma^2$. In this case, we see that V_n is a sample mean estimate of $E\left[X_i^2\right] = \sigma^2$. By Theorem 9.14, V_n is an unbiased consistent sequence of estimates of σ^2.

(c) To determine whether $\hat{\sigma}_{\text{ML}}(n)$ is an unbiased estimator, we check to see if

$$E[\hat{\sigma}_{\text{ML}}(n)] = \sigma$$

For arbitrary n, this is diffcult. For example for $n = 1$, $\hat{\sigma}_{\text{ML}}(1) = \sqrt{X_1^2} = |X_1|$. The mean value is

$$E[\hat{\sigma}_{\text{ML}}(1)] = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\, dx = \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^{\infty} x e^{-x^2/2\sigma^2}\, dx$$

Making the variable substitution $u = x^2/2\sigma^2$, we obtain

$$E[\hat{\sigma}_{\text{ML}}(1)] = \sigma\sqrt{2/\pi} \int_0^{\infty} e^{-u}\, du = \sigma\sqrt{2/\pi}$$

Since $\hat{\sigma}_{\text{ML}}(1) \neq \sigma$, we see that the estimator is biased. For $n > 1$, this estimator of the standard deviation is biased since

$$E\left[\sqrt{\frac{X_1^2 + \cdots + X_n^2}{n}}\right] \neq \sqrt{E\left[\frac{X_1^2 + \cdots + X_n^2}{n}\right]}$$

To determine consistency of the estimator, we use Definition 9.5 and check whether for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P[|\hat{\sigma}_{\text{ML}}(n) - \sigma| \geq \varepsilon] = 0$$

Equivalently, we can check if

$$\lim_{n\to\infty} P[|\hat{\sigma}_{\text{ML}}(n) - \sigma| \leq \varepsilon] = 1$$

For our estimate of the standard deviation, we observe that

$$\begin{aligned} P[|\hat{\sigma}_{\text{ML}}(n) - \sigma| \leq \varepsilon] &= P[\sigma - \varepsilon \leq \hat{\sigma}_{\text{ML}}(n) \leq \sigma + \varepsilon] \\ &= P\left[\sigma^2 - 2\varepsilon\sigma + \varepsilon^2 \leq V_n \leq \sigma^2 + 2\varepsilon\sigma + \varepsilon^2\right] \\ &= P\left[-(2\varepsilon\sigma - \varepsilon^2) \leq V_n - \sigma^2 \leq 2\varepsilon\sigma + \varepsilon^2\right] \end{aligned}$$

For a nontrivial problem, $\sigma > 0$. Hence, we can assume that ε is sufficiently small to ensure that $\varepsilon < \sigma$. This implies that

$$2\sigma\varepsilon + \varepsilon^2 > 2\sigma\varepsilon - \varepsilon^2 > 0$$

This implies

$$\begin{aligned} P[|\hat{\sigma}_{\text{ML}}(n) - \sigma| \leq \varepsilon] &\geq P\left[-(2\varepsilon\sigma - \varepsilon^2) \leq V_n - \sigma^2 \leq 2\varepsilon\sigma - \varepsilon^2\right] \\ &= P\left[\left|V_n - \sigma^2\right| \leq 2\varepsilon\sigma - \varepsilon^2\right] \end{aligned}$$

Let $\varepsilon' = 2\varepsilon\sigma - \varepsilon^2 > 0$. Since V_n is a consistent estimator of σ^2, for any $\varepsilon' > 0$,

$$\lim_{n\to\infty} P[|\hat{\sigma}_{\text{ML}}(n) - \sigma| \leq \varepsilon] \geq \lim_{n\to\infty} = P\left[\left|V_n - \sigma^2\right| \leq \varepsilon'\right] = 1$$

Hence, $\hat{\sigma}_{\text{ML}}(n)$ is a consistent sequence of estimates.