Solutions to HW10

Note: These solutions are based on those generated by R. D. Yates and D. J. Goodman, the authors of our textbook. I have extensively rewritten them.

Problem 7.1.1 •

$X_1, \ldots, X_n$ is an iid sequence of exponential random variables, each with expected value 5.

(a) What is $\text{Var}[M_9(X)]$, the variance of the sample mean based on nine trials?

(b) What is $P[X_1 > 7]$, the probability that one outcome exceeds 7?

(c) Estimate $P[M_9(X) > 7]$, the probability that the sample mean of nine trials exceeds 7? Hint: Use the central limit theorem.

Problem 7.1.1 Solution

We are given that $X_1, X_2 \ldots X_n$ are independent exponential random variables with mean value $\mu_X = 5 = 1/\lambda$ so that for $x \geq 0$, $F_X(x) = 1 - e^{-\lambda x} = 1 - e^{-x/5}$ and $\sigma_X^2 = 1/\lambda^2 = 25$.

(a) By Theorem 7.1, $\sigma^2_{M_n(x)} = \sigma^2_X/n$, so

$$\text{Var}[M_9(X)] = \frac{\sigma_X^2}{9} = \frac{25}{9}. \tag{1}$$

(b) A comment is in order here. The question asks "What is the value of $P[X_1 > 7]$ the probability that one outcome exceeds 7". The probability that X_1 exceeds 7 is not, in general, the same as the probability that some X_i exceeds 7, however in this problem, the X_i are iid, so these two quantities are equal.

$$P[X_1 \geq 7] = 1 - P[X_1 \leq 7] \tag{2}$$
$$= 1 - F_X(7) = 1 - (1 - e^{-7/5}) = e^{-7/5} \approx 0.247 \tag{3}$$

(c) First we express $P[M_9(X) > 7]$ in terms of $X_1, \ldots, X_9$.

$$P[M_9(X) > 7] = 1 - P[M_9(X) \leq 7] = 1 - P[(X_1 + \ldots + X_9) \leq 63] \tag{4}$$

Now the probability that $M_9(X) > 7$ can be approximated using the Central Limit Theorem (CLT).

$$P[M_9(X) > 7] = 1 - P[(X_1 + \ldots + X_9) \leq 63] \tag{5}$$
$$\approx 1 - \Phi\left(\frac{63 - 9\mu_X}{\sqrt{9}\sigma_X}\right) = 1 - \Phi(6/5) \tag{6}$$

Consulting Table 3.1 to obtain a value for $\Phi(6/5)$ and substituting into the expression above yields $P[M_9(X) > 7] \approx 0.1151$.

Problem 7.1.2 •
$X_1, \ldots, X_n$ are independent uniform random variables, all with expected value $\mu_X = 7$ and variance $\text{Var}[X] = 3$.

(a) What is the PDF of X_1?

(b) What is $\text{Var}[M_{16}(X)]$, the variance of the sample mean based on 16 trials?

(c) What is $P[X_1 > 9]$, the probability that one outcome exceeds 9?

(d) Would you expect $P[M_{16}(X) > 9]$ to be bigger or smaller than $P[X_1 > 9]$? To check your intuition, use the central limit theorem to estimate $P[M_{16}(X) > 9]$.

Problem 7.1.2 Solution
$X_1, X_2 \ldots X_n$ are independent uniform random variables with mean value $\mu_X = 7$ and $\sigma_X^2 = 3$

(a) Since X_1 is a uniform random variable, it must have a uniform PDF over an interval $[a, b]$. From Appendix A, we have that for a uniform random variable on the interval $[a, b]$, the mean and variance are $\mu_X = (a + b)/2$ and that $\text{Var}[X] = (b - a)^2/12$. Hence, given the mean and variance, we obtain the following equations for a and b.

$$(b - a)^2/12 = 3 \qquad (a + b)/2 = 7 \tag{1}$$

Solving the first of these equations yields $|b - a| = 6$. For a nonempty interval, b must be greater than a so we have that $b = a + 6$. Then $(2a + 6)/2 = 7$ implies that $a = 4$ and thus $b = 10$ so the distribution of X is

$$f_X(x) = \begin{cases} 1/6 & 4 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

(b) By Theorem 7.1, we have

$$\text{Var}[M_{16}(X)] = \frac{\text{Var}[X]}{16} = \frac{3}{16} \tag{3}$$

(c) Since $f_X(x) = 0 \ \forall\, x > 10$,

$$P[X_1 \geq 9] = \int_9^\infty f_{X_1}(x)\, dx = \int_9^{10} (1/6)\, dx = 1/6 \tag{4}$$

(d) The variance of $M_{16}(X)$ is much less than $\text{Var}[X_1]$. Hence, the PDF of $M_{16}(X)$ should be much more concentrated about $E[X]$ than the PDF of X_1. Thus we should expect $P[M_{16}(X) > 9]$ to be much less than $P[X_1 > 9]$.

$$P[M_{16}(X) > 9] = 1 - P[M_{16}(X) \leq 9] = 1 - P[(X_1 + \cdots + X_{16}) \leq 16(9)] \tag{5}$$

Applying the Central Limit Theorem to obtain an approximation of this probability yields

$$P\left[M_{16}(X) > 9\right] \approx 1 - \Phi\left(\frac{144 - 16\mu_X}{\sqrt{16}\sigma_X}\right) \approx 1 - \Phi(2.67) \approx 1 - 0.9962 \approx 0.0038 \quad (6)$$

As predicted, $P[M_{16}(X) > 9] \ll P[X_1 > 9]$.

Problem 7.2.1 •

The weight of a randomly chosen Maine black bear has expected value $E[W] = 500$ pounds and standard deviation $\sigma_W = 100$ pounds. Use the Chebyshev inequality to upper bound the probability that the weight of a randomly chosen bear is more than 200 pounds from the expected value of the weight.

Problem 7.2.1 Solution

If the average weight of a Maine black bear is 500 pounds with standard deviation equal to 100 pounds, we can use the Chebyshev inequality to upper bound the probability that a randomly chosen bear will be more then 200 pounds away from the average.

$$P\left[|W - E\left[W\right]| \geq 200\right] \leq \frac{\text{Var}[W]}{200^2} \leq \frac{100^2}{200^2} = 0.25 \quad (1)$$

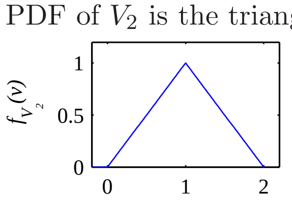
Problem 7.2.3 ■

Let X equal the arrival time of the third elevator in Quiz 7.2. Find the exact value of $P[W \geq 75]$. Compare your answer to the upper bounds derived in Quiz 7.2.

Problem 7.2.3 Solution

First we derive the PDF of the sum $W = X_1 + X_2 + X_3$ of iid uniform $(0, 30)$ random variables, using the techniques of Chapter 6. To simplify our calculations, we find the PDF of $V = Y_1 + Y_2 + Y_3$ where the Y_i are iid uniform $(0, 1)$ random variables, then apply Theorem 3.20 to conclude that $W = 30V$ represents the sum of three iid uniform $(0, 30)$ random variables.

To start, let $V_2 = Y_1 + Y_2$. Since each Y_1 has a PDF shaped like a unit area pulse, the PDF of V_2 is the triangular function



$$f_{V_2}(v) = \begin{cases} v & 0 \leq v \leq 1 \\ 2 - v & 1 < v \leq 2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Then the PDF of $V = V_2 + Y_3$ is the convolution integral

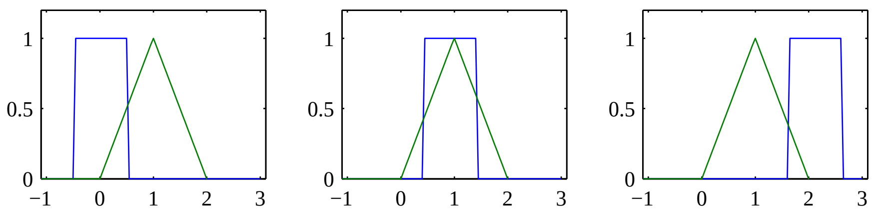
$$f_V(v) = \int_{-\infty}^{\infty} f_{V_2}(y)\, f_{Y_3}(v - y)\, dy \quad (2)$$

$$= \int_0^1 y f_{Y_3}(v - y)\, dy + \int_1^2 (2 - y) f_{Y_3}(v - y)\, dy. \quad (3)$$

Evaluation of these integrals depends on v through the function

$$f_{Y_3}(v-y) = \begin{cases} 1 & v-1 < v < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

To compute the convolution, it is helpful to depict the three distinct cases. In each case, the square "pulse" is $f_{Y_3}(v-y)$ and the triangular pulse is $f_{V_2}(y)$.



$0 \le v < 1$ $\quad$ $1 \le v < 2$ $\quad$ $2 \le v < 3$

From the graphs, we can compute the convolution for each case:

$$0 \le v < 1: \qquad f_{V_3}(v) = \int_0^v y\,dy = \frac{1}{2}v^2 \tag{5}$$

$$1 \le v < 2: \qquad f_{V_3}(v) = \int_{v-1}^1 y\,dy + \int_1^v (2-y)\,dy = -\frac{v^2}{2} + 3v - 2 \tag{6}$$

$$2 \le v < 3: \qquad f_{V_3}(v) = \int_{v-1}^2 (2-y)\,dy = \frac{(3-v)^2}{2} \tag{7}$$

To complete the problem, we use Theorem 3.20 to observe that $W = 30V_3$ is the sum of three iid uniform $(0, 30)$ random variables. From Theorem 3.19,

$$f_W(w) = \frac{1}{30} f_{V_3}(v_3)\, v/30 = \begin{cases} (w/30)^2/60 & 0 \le w < 30, \\ [-(w/30)^2/2 + 3(w/30) - 2]/30 & 30 \le w < 60, \\ [3-(w/30)]^2/60 & 60 \le w < 90, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Finally, we can compute the exact probability

$$P[W \ge 75] = \frac{1}{60}\int_{75}^{90} [3-(w/30)]^2\,dw = -\left.\frac{(3-w/30)^3}{6}\right|_{75}^{90} = \frac{1}{48} \tag{9}$$

For comparison, the Markov inequality indicated that $P[W < 75] \le 3/5$ and the Chebyshev inequality showed that $P[W < 75] \le 1/4$. As we see, both inequalities are quite weak in this case.

Problem 7.3.1 •

When X is Gaussian, verify the claim of Equation (7.16) that the sample mean is within one standard error of the expected value with probability 0.68.

Problem 7.3.1 Solution

For an an arbitrary Gaussian (μ, σ) random variable Y,

$$\begin{aligned} P[\mu-\sigma \leq Y \leq \mu+\sigma] &= P[-\sigma \leq Y-\mu \leq \sigma] && (1)\\ &= P\left[-1 \leq \frac{Y-\mu}{\sigma} \leq 1\right] && (2)\\ &= \Phi(1)-\Phi(-1) = 2\Phi(1)-1 = 0.6827. && (3)\end{aligned}$$

Note that Y can be any Gaussian random variable, including, for example, $M_n(X)$ when X is Gaussian. When X is not Gaussian, the same claim holds to the extent that the central limit theorem promises that $M_n(X)$ is nearly Gaussian for large n.

Problem 7.4.1 •

$X_1, \ldots, X_n$ are n independent identically distributed samples of random variable X with PMF

$$P_X(x) = \begin{cases} 0.1 & x = 0, \\ 0.9 & x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) How is $E[X]$ related to $P_X(1)$?

(b) Use Chebyshev's inequality to find the confidence level α such that $M_{90}(X)$, the estimate based on 90 observations, is within 0.05 of $P_X(1)$. In other words, find α such that

$$P[|M_{90}(X) - P_X(1)| \geq 0.05] \leq \alpha.$$

(c) Use Chebyshev's inequality to find out how many samples n are necessary to have $M_n(X)$ within 0.03 of $P_X(1)$ with confidence level 0.1. In other words, find n such that

$$P[|M_n(X) - P_X(1)| \geq 0.03] \leq 0.1.$$

Problem 7.4.1 Solution

We are given that $X_1, \ldots, X_n$ are n independent identically distributed samples of the random variable X having PMF

$$P_X(x) = \begin{cases} 0.1 & x = 0 \\ 0.9 & x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

(a) $E[X]$ is in fact the same as $P_X(1)$ because X is a Bernoulli random variable.

(b) By Chebyshev's inequality,

$$P[|M_{90}(X) - P_X(1)| \geq .05] = P[|M_{90}(X) - E[X]| \geq .05] \leq \frac{\text{Var}[Y]}{(0.5)^2} = \alpha \qquad (2)$$

so

$$\alpha = \frac{\sigma_X^2}{90(.05)^2} = \frac{.09}{90(.05)^2} = 0.4 \qquad (3)$$

(c) Now we wish to find the value of n such that $P[|M_n(X) - P_X(1)| \geq .03] \leq .1$. From the Chebyshev inequality, we write

$$0.1 = \frac{\sigma_X^2}{n(.03)^2}. \tag{4}$$

Since $\sigma_X^2 = 0.09$, solving for n yields $n = 100$.

Problem 7.4.2 •

Let $X_1, X_2, \ldots$ denote an iid sequence of random variables, each with expected value 75 and standard deviation 15.

(a) How many samples n do we need to guarantee that the sample mean $M_n(X)$ is between 74 and 76 with probability 0.99?

(b) If each X_i has a Gaussian distribution, how many samples n' would we need to guarantee $M_{n'}(X)$ is between 74 and 76 with probability 0.99?

Problem 7.4.2 Solution

$X_1, X_2, \ldots$ are iid random variables each with mean 75 and standard deviation 15.

(a) We would like to find the value of n such that

$$P[74 \leq M_n(X) \leq 76] = 0.99 \tag{1}$$

When we know only the mean and variance of X_i, our only real tool is the Chebyshev inequality which says that

$$\begin{aligned} P[74 \leq M_n(X) \leq 76] &= 1 - P[|M_n(X) - E[X]| \geq 1] & (2)\\ &\geq 1 - \frac{\operatorname{Var}[X]}{n} = 1 - \frac{225}{n} \geq 0.99 & (3) \end{aligned}$$

This yields $n \geq 22{,}500$.

(b) If each X_i is a Gaussian, the sample mean, $M_n(X)$ will also be Gaussian with mean and variance

$$\begin{aligned} E[M_{n'}(X)] &= E[X] = 75 & (4)\\ \operatorname{Var}[M_{n'}(X)] &= \operatorname{Var}[X]/n' = 225/n' & (5) \end{aligned}$$

In this case,

$$\begin{aligned} P[74 \leq M_{n'}(X) \leq 76] &= \Phi\left(\frac{76-\mu}{\sigma}\right) - \Phi\left(\frac{74-\mu}{\sigma}\right) & (6)\\ &= \Phi(\sqrt{n'}/15) - \Phi(-\sqrt{n'}/15) & (7)\\ &= 2\Phi(\sqrt{n'}/15) - 1 = 0.99 & (8) \end{aligned}$$

so $\Phi(\sqrt{n'}/15) = 1.99/2 = .995$. Then from the table, $\sqrt{n'}/15 \approx 2.58$ so $n' \approx 1{,}498$.

Since even under the Gaussian assumption, the number of samples n' is so large that even if the X_i are not Gaussian, the sample mean may be approximated by a Gaussian. Hence, about 1500 samples probably is about right. However, in the absence of any information about the PDF of X_i beyond the mean and variance, we cannot make any guarantees stronger than that given by the Chebyshev inequality.

Problem 7.4.3 •

Let X_A be the indicator random variable for event A with probability $P[A] = 0.8$. Let $\hat{P}_n(A)$ denote the relative frequency of event A in n independent trials.

(a) Find $E[X_A]$ and $\mathrm{Var}[X_A]$.

(b) What is $\mathrm{Var}[\hat{P}_n(A)]$?

(c) Use the Chebyshev inequality to find the confidence coefficient $1-\alpha$ such that $\hat{P}_{100}(A)$ is within 0.1 of $P[A]$. In other words, find α such that

$$P\left[\left|\hat{P}_{100}(A) - P[A]\right| \leq 0.1\right] \geq 1 - \alpha.$$

(d) Use the Chebyshev inequality to find out how many samples n are necessary to have $\hat{P}_n(A)$ within 0.1 of $P[A]$ with confidence coefficient 0.95. In other words, find n such that

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \leq 0.1\right] \geq 0.95.$$

Problem 7.4.3 Solution

(a) Since X_A is a Bernoulli ($p = P[A]$) random variable,

$$E[X_A] = P[A] = 0.8, \qquad \mathrm{Var}[X_A] = P[A](1 - P[A]) = 0.16. \tag{1}$$

(b) Let $X_{A,i}$ denote X_A for the ith trial. Since $\hat{P}_n(A) = M_n(X_A) = \frac{1}{n}\sum_{i=1}^n X_{A,i}$,

$$\mathrm{Var}[\hat{P}_n(A)] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[X_{A,i}] = \frac{P[A](1 - P[A])}{n}. \tag{2}$$

(c) Since $\hat{P}_{100}(A) = M_{100}(X_A)$, we can use Theorem 7.12(b) to write

$$P\left[\left|\hat{P}_{100}(A) - P[A]\right| < c\right] \geq 1 - \frac{\mathrm{Var}[X_A]}{100c^2} = 1 - \frac{0.16}{100c^2} = 1 - \alpha. \tag{3}$$

For $c = 0.1$, $\alpha = 0.16/[100(0.1)^2] = 0.16$. Thus, with 100 samples, our confidence coefficient is $1 - \alpha = 0.84$.

(d) In this case, the number of samples n is unknown. Once again, we use Theorem 7.12(b) to write

$$P\left[\left|\hat{P}_n(A) - P[A]\right| < c\right] \geq 1 - \frac{\mathrm{Var}[X_A]}{nc^2} = 1 - \frac{0.16}{nc^2} = 1 - \alpha. \tag{4}$$

For $c = 0.1$, we have confidence coefficient $1 - \alpha = 0.95$ if $\alpha = 0.16/[n(0.1)^2] = 0.05$, or $n = 320$.

Problem 7.4.5 •
In n independent experimental trials, the relative frequency of event A is $\hat{P}_n(A)$. How large should n be to ensure that the confidence interval estimate

$$\hat{P}_n(A) - 0.05 \leq P[A] \leq \hat{P}_n(A) + 0.05$$

has confidence coefficient 0.9?

Problem 7.4.5 Solution
First we observe that the interval estimate can be expressed as

$$\left|\hat{P}_n(A) - P[A]\right| < 0.05. \tag{1}$$

Since $\hat{P}_n(A) = M_n(X_A)$ and $E[M_n(X_A)] = P[A]$, we can use Theorem 7.12(b) to write

$$P\left[\left|\hat{P}_n(A) - P[A]\right| < 0.05\right] \geq 1 - \frac{\mathrm{Var}[X_A]}{n(0.05)^2}. \tag{2}$$

Note that $\mathrm{Var}[X_A] = P[A](1 - P[A]) \leq \max_{x \in (0,1)} x(1-x) = 0.25$. Thus for confidence coefficient 0.9, we require that

$$1 - \frac{\mathrm{Var}[X_A]}{n(0.05)^2} \geq 1 - \frac{0.25}{n(0.05)^2} \geq 0.9. \tag{3}$$

This implies $n \geq 1{,}000$ samples are needed.

Problem 8.1.1 •
Let L equal the number of flips of a coin up to and including the first flip of heads. Devise a significance test for L at level $\alpha = 0.05$ to test the hypothesis H that the coin is fair. What are the limitations of the test?

Problem 8.1.1 Solution
To test the hypothesis H that the coin is fair. Then, we must choose a rejection region R such that, given that H is true, the probability that the outcome s is in the rejection region R is0.05, *i.e* $\alpha = P[s \in R|H] = 0.05$. Our outcome in this experiment is the value of the random variable L. We will define the rejection region by picking a threshold l^* such that rejection region $R = \{l > l^*\}$. What remains is to choose l^* so that $P[L > l^*|H] = 0.05$. Note that $L > l$ if we have observed l tails in a row before observing the first heads. Under the hypothesis that the coin is fair, l tails in a row occurs with probability

$$P[L > l] = (1/2)^l \tag{1}$$

Thus, we need

$$P[R] = P[L > l^*] = 2^{-l^*} = 0.05 \tag{2}$$

Thus, $l^* = -\log_2(0.05) = \log_2(20) = 4.32$. In this case, we reject the hypothesis that the coin is fair if $L \geq 5$. The significance level of the test is $\alpha = P[L > 4] = 2^{-4} = 0.0625$ which close to but not exactly 0.05.

The shortcoming of this test is that we always accept the hypothesis that the coin is fair whenever heads occurs on the first, second, third or fourth flip. If the coin was biased such that the probability of heads was much higher than 1/2, say 0.8 or 0.9, we would hardly ever reject the hypothesis that the coin is fair. In that sense, our test cannot identify that kind of biased coin. This means that this particular test is only suited to the case that we know that the coin is fair or has a higher probability of tails than heads. (If we wanted to design a test that considered deviations in either direction from fairness, we'd want to have both lower and upper threshold values for our rejection region. However, we'd also want to change the structure of our experiment so that we observed at least a minimum number of flips. You can calculate the probabilities of error for different approaches to see why.)

Problem 8.1.4 •

The duration of a voice telephone call is an exponential random variable T with expected value $E[T] = 3$ minutes. Data calls tend to be longer than voice calls on average. Observe a call and reject the null hypothesis that the call is a voice call if the duration of the call is greater than t_0 minutes.

(a) Write a formula for α, the significance of the test as a function of t_0.

(b) What is the value of t_0 that produces a significance level $\alpha = 0.05$?

Problem 8.1.4 Solution

(a) The rejection region is $R = \{T > t_0\}$. The duration of a voice call has exponential PDF

$$f_T(t) = \begin{cases} (1/3)e^{-t/3} & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The significance level of the test is

$$\alpha = P[T > t_0] = \int_{t_0}^{\infty} f_T(t)\, dt = e^{-t_0/3}. \tag{2}$$

(b) The significance level is $\alpha = 0.05$ if $t_0 = -3\ln\alpha = 8.99$ minutes.

Problem 8.2.1 •

In a random hour, the number of call attempts N at a telephone switch has a Poisson distribution with a mean of either α_0 (hypothesis H_0) or α_1 (hypothesis H_1). For a priori probabilities $P[H_i]$, find the MAP and ML hypothesis testing rules given the observation of N.

Problem 8.2.1 Solution

For the MAP test, we must choose acceptance regions A_0 and A_1 for the two hypotheses H_0 and H_1. From Theorem 8.2, the MAP rule is

$$n \in A_0 \text{ if } \frac{P_{N|H_0}(n)}{P_{N|H_1}(n)} \geq \frac{P[H_1]}{P[H_0]}; \qquad n \in A_1 \text{ otherwise.} \tag{1}$$

Since $P_{N|H_i}(n) = \lambda_i^n e^{-\lambda_i}/n!$, where $\lambda_i = 1/\alpha_i$, the MAP rule becomes

$$n \in A_0 \text{ if } \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{-(\lambda_0-\lambda_1)} \geq \frac{P[H_1]}{P[H_0]}; \qquad n \in A_1 \text{ otherwise.} \tag{2}$$

We obtain the threshold n^* by substituting n^* for n in (2) and isolating n^*. Taking logarithms we obtain

$$n^* (\ln \lambda_0 - \ln \lambda_1) - (\lambda_0 - \lambda_1) \geq \ln (P[H_1]/P[H_2]). \tag{3}$$

Rearranging yields

$$n \geq \frac{\ln (P[H_1]/P[H_2]) + \lambda_0 - \lambda_1}{\ln \lambda_0 - \ln \lambda_1}. \tag{4}$$

Now, in order to determine whether n^* should be a lower bound or an upper bound for our rejection region, we need to know which is larger, α_0 or α_1. Suppose that $\alpha_0 > \alpha_1$. Then $\lambda_0 < \lambda_1$ and we state the MAP rule as

$$n \in A_0 \text{ if } n \leq n^* = \frac{\lambda_0 - \lambda_1 + \ln(P[H_0]/P[H_1])}{\ln(\lambda_0/\lambda_1)}; \qquad n \in A_1 \text{ otherwise.} \tag{5}$$

From the MAP rule, we can get the ML rule by setting the a priori probabilities to be equal. This yields the ML rule

$$n \in A_0 \text{ if } n \leq n^* = \frac{\lambda_0 - \lambda_1}{\ln(\lambda_0/\lambda_1)}; \qquad n \in A_1 \text{ otherwise.} \tag{6}$$