Solutions to Selected Problems In: Detection, Estimation, and Modulation Theory: Part I by Harry L. Van Trees

John L. Weatherwax*

April 16, 2014

Introduction

Here you'll find some notes that I wrote up as I worked through this excellent book. I've worked hard to make these notes as good as I can, but I have no illusions that they are perfect. If you feel that that there is a better way to accomplish or explain an exercise or derivation presented in these notes; or that one or more of the explanations is unclear, incomplete, or misleading, please tell me. If you find an error of any kind – technical, grammatical, typographical, whatever – please tell me that, too. I'll gladly add to the acknowledgments in later printings the name of the first person to bring each problem to my attention.

Special thanks (most recent comments are listed first) to Iman Bagheri, Jeong-Min Choi and Hemant Saggar for their corrections involving chapter 2.

All comments (no matter how small) are much appreciated. In fact, if you find these notes useful I would appreciate a contribution in the form of a solution to a problem that is not yet worked in these notes. Sort of a "take a penny, leave a penny" type of approach. Remember: pay it forward.

*wax@alum.mit.edu

Chapter 2 (Classical Detection and Estimation Theory)

Notes On The Text

Notes on the Bayes' Criterion

Given the books Eq. 8 we have

$$\begin{aligned}\mathcal{R} &= P_0 C_{00} \int_{Z_0} p(R|H_0)dR + P_0 C_{10} \int_{Z-Z_0} p(R|H_0)dR \\ &= P_1 C_{01} \int_{Z_0} p(R|H_1)dR + P_1 C_{11} \int_{Z-Z_0} p(R|H_1)dR \,. \end{aligned} \tag{1}$$

We can use

$$\int_{Z_0} p(R|H_0)dR + \int_{Z-Z_0} p(R|H_0)dR = 1 \,,$$

to replace all integrals over $Z - Z_0$ with (one minus) integrals over Z_0. We get

$$\begin{aligned}\mathcal{R} &= P_0 C_{00} \int_{Z_0} p(R|H_0)dR + P_0 C_{10} \left(1 - \int_{Z_0} p(R|H_0)dR\right) \\ &= P_1 C_{01} \int_{Z_0} p(R|H_1)dR + P_1 C_{11} \left(1 - \int_{Z_0} p(R|H_1)dR\right) \\ &= P_0 C_{10} + P_1 C_{11} \\ &+ \int_{Z_0} \left\{ P_1 (C_{01} - C_{11}) p(R|H_1) - P_0 (C_{10} - C_{00}) p(R|H_0) \right\} dR \end{aligned} \tag{2}$$

If we introduce the probability of false alarm P_F, the probability of detection P_D, and the probability of a miss P_M, as defined in the book, we find that $\mathcal{R}$ given via Equation 1 becomes when we use $\int_{Z_0} p(R|H_0)dR + \int_{Z_1} p(R|H_0)dR = \int_{Z_0} p(R|H_0)dR + P_F = 1$

$$\begin{aligned}\mathcal{R} &= P_0 C_{10} + P_1 C_{11} \\ &\quad + P_1 (C_{01} - C_{11}) P_M - P_0 (C_{10} - C_{00})(1 - P_F) \,. \end{aligned} \tag{3}$$

Since $P_0 = 1 - P_1$ we can consider $\mathcal{R}$ computed in Equation 3 as a function of the prior probability P_1 with the following manipulations

$$\begin{aligned}\mathcal{R}(P_1) &= (1 - P_1) C_{10} + P_1 C_{11} + P_1 (C_{01} - C_{11}) P_M - (1 - P_1)(C_{10} - C_{00})(1 - P_F) \\ &= C_{10} - (C_{10} - C_{00})(1 - P_F) + P_1 \left[-C_{10} + C_{11} + (C_{01} - C_{11}) P_M + (C_{10} - C_{00})(1 - P_F) \right] \\ &= C_{00} + (C_{10} - C_{00}) P_F + P_1 \left[C_{11} - C_{00} + (C_{01} - C_{11}) P_M - (C_{10} - C_{00}) P_F \right] \\ &= C_{00}(1 - P_F) + C_{10} P_F + P_1 \left[(C_{11} - C_{10}) + (C_{01} - C_{11}) P_M - (C_{10} - C_{00}) P_F \right] \,. \end{aligned} \tag{4}$$

Recall that for the Bayes decision test our decision regions Z_0 and Z_1 are determined via

$$\Lambda(R) > \frac{P_0 (C_{10} - C_{00})}{P_1 (C_{01} - C_{11})} \,,$$

for H_1. Thus if P_0 changes the decision regions Z_0 and Z_1 change (via the above expression) and thus both P_F and P_M change since they depend on Z_0 and Z_1. Lets assume that we specify a decision boundary η that then defines classification regions Z_0 and Z_1. These decision regions correspond to a specific value of P_1 denoted via P_1^*. Note that P_1^* is not the true prior probability of the class H_1 but is simply an equivalent prior probability that one could use in the likelihood ratio test to obtain the same decision regions Z_0 and Z_1. The book denotes $\mathcal{R}_B(P_1)$ to be the expression given via Equation 4 where P_F and P_M changes in concert with P_1. The book denotes $\mathcal{R}_F(P_1)$ to be the expression given by Equation 4 but where P_F and P_M are *fixed* and are held constant as we change the value of P_1.

In the case where we do not fix P_F and P_M we can evaluate $\mathcal{R}_B(P_1)$ at its two end points of P_1. If $P_1 = 0$ then from Equation 4 $\mathcal{R}_B(0) = C_{00}(1 - P_F) + C_{10}P_F$. When we have $P_1 = 0$ then we see that

$$\eta \equiv \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \to +\infty\,,$$

for all R the function $\Lambda(R)$ is always less than η, and all classifications are H_0. Thus Z_1 is the empty set and $P_F = 0$. Thus we get $\mathcal{R}_B(0) = C_{00}$.

The other extreme is when $P_1 = 1$. In that case $P_0 = 0$ so $\eta = 0$ and we would have that $\Lambda(R) > 0$ for all R implying that all points are classified as H_1. This implies that $P_M = 0$. The expression for $\mathcal{R}_B(P_1)$ from Equation 4 is given by

$$\mathcal{R}_B(1) = C_{00}(1 - P_F) + C_{10}P_F + (C_{11} - C_{00}) - (C_{10} - C_{00})P_F = C_{11}\,,$$

when we simplify. These to values for $\mathcal{R}(0)$ and $\mathcal{R}(1)$ give the end point conditions on $\mathcal{R}_B(P_1)$ seen in Figure 2.7. If we do not *know* the value of P_1 then one can still design a hypothesis test by specifying values of P_M and P_F such that the coefficient of P_1 in the expression for $\mathcal{R}_F(P_1)$ in Equation 4 *vanishes*. The idea behind this procedure is that this will make $\mathcal{R}_F$ a horizontal line for all values of P_1 which is an upper bound on the Bayes' risk. To make the coefficient of P_1 vanish requires that

$$(C_{11} - C_{00}) + (C_{01} - C_{11})P_M - (C_{10} - C_{00})P_F = 0\,. \tag{5}$$

This is also known as the *minimax* test. The decision threshold η can be introduced into the definitions of P_F and P_M and the above gives an equation that can be used to determine its value. If we take $C_{00} = C_{11} = 0$ and introduce the shorthands $C_{01} = C_M$ (the cost of a miss) and $C_{01} = C_F$ (the cost of a false alarm) so we get our constant minimax risk of

$$\mathcal{R}_F = C_F P_F + P_1[C_M P_M - C_F P_F] = P_0 C_F P_F + P_1 C_M P_M\,. \tag{6}$$

Receiver operating characteristics: Example 1 (Gaussians with different means)

Under H_1 each sample R_i can be written as $R_i = m + n_i$ with n_i a Gaussian random variable (with mean 0 and variance σ^2). Thus $R_i \sim N(m, \sigma^2)$. The statistic l which is the sum of individual random variables is also normal. The mean of l is given by the sum of the N means (multiplied by the scaling factor $\frac{1}{\sqrt{N}\sigma}$) or

$$\left(\frac{1}{\sqrt{N}\sigma}\right)(Nm) = \frac{\sqrt{N}}{\sigma}m\,,$$

and a variance given by the sum of the N variances (multiplied by the square of the scaling factor $\frac{1}{\sqrt{N}\sigma}$) or

$$\left(\frac{1}{N\sigma^2}\right) N\sigma^2 = 1\,.$$

These two arguments have shown that $l \sim N\left(\frac{\sqrt{N}}{\sigma}m, 1\right)$.

We now derive $\Pr(\epsilon)$ for the case where we have measurements from Gaussians with different means (Example 1). To do that we need to note the following symmetry identity about $\text{erfc}_*(X)$ function. We have that

$$\begin{aligned}
\text{erfc}_*(-X) &\equiv \int_{-X}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\
&= 1 - \int_{-\infty}^{-X} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1 - \int_{X}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\
&= 1 - \text{erfc}_*(X)\,.
\end{aligned} \tag{7}$$

Then using this result we can derive the given expression for $\Pr(\epsilon)$ under the case when $P_0 = P_1 = \frac{1}{2}$ and $\eta = 1$. We have

$$\begin{aligned}
\Pr(\epsilon) &= \frac{1}{2}(P_F + P_M) \quad \text{since } \eta = 1 \text{ this becomes} \\
&= \frac{1}{2}\left(\text{erfc}_*\left(\frac{d}{2}\right) + 1 - P_D\right) = \frac{1}{2}\left(\text{erfc}_*\left(\frac{d}{2}\right) + 1 - \text{erfc}_*\left(-\frac{d}{2}\right)\right) \\
&= \frac{1}{2}\left(\text{erfc}_*\left(\frac{d}{2}\right) + 1 - \left(1 - \text{erfc}_*\left(\frac{d}{2}\right)\right)\right) \\
&= \text{erfc}_*\left(\frac{d}{2}\right)\,,
\end{aligned} \tag{8}$$

the expression given in the book.

Receiver operating characteristics: Example 2 (Gaussians with $\sigma_0^2 \neq \sigma_1^2$)

Following the arguments in the book we end up wanting to evaluate the expression $P_F = \Pr(r_1^2 + r_2^2 \geq \gamma | H_0)$. By definition this is just the integral over the region of $r_1 - r_2$ space where $r_1^2 + r_2^2 \geq$ hold true. This is

$$P_F = \int_{\theta=0}^{2\pi} \int_{Z=\sqrt{\gamma}}^{\infty} p(R_1|H_0) P(R_2|H_0) dR_1 dR_2\,.$$

When we put in the expressions for $p(R_1|H_0)$ and $P(R_2|H_0)$ we see why converting to polar coordinates is helpful. We have

$$\begin{aligned}
P_F &= \int_{\theta=0}^{2\pi} \int_{Z=\sqrt{\gamma}}^{\infty} \left(\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}\frac{R_1^2}{\sigma_0^2}}\right) \left(\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}\frac{R_2^2}{\sigma_0^2}}\right) dR_1 dR_2 \\
&= \frac{1}{2\pi\sigma_0^2} \int_{\theta=0}^{2\pi} \int_{Z=\sqrt{\gamma}}^{\infty} \exp\left\{-\frac{1}{2}\frac{R_1^2 + R_2^2}{\sigma_0^2}\right\} dR_1 dR_2\,.
\end{aligned}$$

When we change to polar we have the differential of area change via $dR_1 dR_2 = Z d\theta dZ$ and thus get for P_F the following

$$P_F = \frac{1}{2\pi\sigma_0^2} \int_{\theta=0}^{2\pi} \int_{Z=\sqrt{\gamma}}^{\infty} Z \exp\left\{ -\frac{1}{2} \frac{Z^2}{\sigma_0^2} \right\} d\theta dZ = \frac{1}{\sigma_0^2} \int_{Z=\sqrt{\gamma}}^{\infty} Z \exp\left\{ -\frac{1}{2} \frac{Z^2}{\sigma_0^2} \right\} dZ \,.$$

If we let $v = \frac{Z^2}{2\sigma_0^2}$ then $dv = \frac{Z}{\sigma_0^2} dZ$ so P_F becomes

$$P_F = \frac{1}{\sigma_0^2} \int_{\frac{\gamma}{2\sigma_0^2}}^{\infty} \sigma_0^2 e^{-v} dv = -e^{-v}\Big|_{\frac{\gamma}{2\sigma_0^2}}^{\infty} = e^{-\frac{\gamma}{2\sigma_0^2}} \,, \tag{9}$$

the expression for P_F given in the book. For P_D the only thing that changes in the calculation is that the normal has a variance of σ_1^2 rather than σ_0^2. Making this change gives the expression for P_D in the book.

We can compute the ROC curve for this example by writing $\gamma = -2\sigma_0^2 \ln(P_F)$ and putting this into the expression for P_D. Where we find

$$P_D = \exp\left(-\frac{\gamma}{2\sigma_1^2} \right) = \exp\left(\frac{\sigma_0^2}{\sigma_1^2} \ln(P_F) \right) \,.$$

This gives

$$\ln(P_D) = \frac{\sigma_0^2}{\sigma_1^2} \ln(P_F) \quad \text{or} \quad P_D = P_F^{\frac{\sigma_0^2}{\sigma_1^2}} \,. \tag{10}$$

From Equation 10 if $\frac{\sigma_1^2}{\sigma_0^2}$ increases then $\frac{\sigma_0^2}{\sigma_1^2}$ decreases, thus $P_F^{\sigma_0^2/\sigma_1^2}$ get larger (since P_F is less than 1). This in tern makes P_D gets larger.

Notes on properties of ROC curves

Recall that a randomized rule applied between thresholds at two points on the ROC curve (say A and B) allows a system designer to obtain (P_F, P_D) performance for all points on the straight line between A and B. This comment allows one to argue that a ROC curve must be concave down. For if the ROC curve were concave up, then by using a randomized rule this linear approximation would have better performance than the ROC curve. Since we know that the ROC curve expresses the optimal performance characteristics this is a contradiction.

Notes on the M hypothesis decision problem

In this section we derive (with more detail) some of the results presented in the book in the case where there are at total of M hypothesis to choose from. We start with the definition of the Bayes' risk $\mathcal{R}$ or

$$\begin{aligned}\mathcal{R} &= \sum_{j=0}^{M-1}\left(P_j\sum_{i=0}^{M-1}C_{ij}P(\text{choose } i|j \text{ is true})\right) = \sum_{j=0}^{M-1}\left(P_j\sum_{i=0}^{M-1}C_{ij}\int_{Z_i}p(R|H_j)dR\right)\\ &= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1}P_jC_{ij}\int_{Z_i}p(R|H_j)dR\,. \end{aligned} \tag{11}$$

Lets consider the case where there are three classes $M = 3$ and expand our $\mathcal{R}$ where we replace the integration region over the "correct" regions with their complement in terms of Z i.e.

$$\begin{aligned}\mathcal{R} &= P_0C_{00}\int_{Z_0=Z-Z_1-Z_2}p(R|H_0)dR + P_0C_{10}\int_{Z_1}p(R|H_0)dR + P_0C_{20}\int_{Z_2}p(R|H_0)dR\\ &= P_1C_{01}\int_{Z_0}p(R|H_1)dR + P_1C_{11}\int_{Z_1=Z-Z_0-Z_2}p(R|H_1)dR + P_1C_{21}\int_{Z_2}p(R|H_1)dR\\ &= P_2C_{02}\int_{Z_0}p(R|H_2)dR + P_2C_{12}\int_{Z_1}p(R|H_2)dR + P_2C_{22}\int_{Z_2=Z-Z_0-Z_1}p(R|H_2)dR\,. \end{aligned} \tag{12}$$

If we then simplify by breaking the intervals over segments into their component pieces for example we simplify the first integral above as

$$\int_{Z-Z_1-Z_2}p(R|H_0)dR = 1 - \int_{Z_1}p(R|H_0)dR - \int_{Z_2}p(R|H_0)dR\,.$$

Doing this in three places gives us

$$\begin{aligned}\mathcal{R} &= P_0C_{00} + P_1C_{11} + P_2C_{22}\\ &+ \int_{Z_0}\{P_1(C_{01}-C_{11})p(R|H_1) + P_2(C_{02}-C_{22})p(R|H_2)\}dR\\ &+ \int_{Z_1}\{P_0(-C_{00}+C_{10})p(R|H_0) + P_2(C_{12}-C_{22})p(R|H_2)\}dR\\ &+ \int_{Z_2}\{P_0(-C_{00}+C_{20})p(R|H_0) + P_2(-C_{11}+C_{21})p(R|H_1)\}dR\\ &= P_0C_{00} + P_1C_{11} + P_2C_{22}\\ &+ \int_{Z_0}\{P_2(C_{02}-C_{22})p(R|H_2) + P_1(C_{01}-C_{11})p(R|H_1)\}dR\\ &+ \int_{Z_1}\{P_0(C_{10}-C_{00})p(R|H_0) + P_2(C_{12}-C_{22})p(R|H_2)\}dR\\ &+ \int_{Z_2}\{P_0(C_{20}-C_{00})p(R|H_0) + P_1(C_{21}-C_{11})p(R|H_1)\}dR\,. \end{aligned} \tag{13}$$

If we define the *integrands* of the above integrals at the point R as I_1, I_2, and I_3 such that

$$\begin{aligned}I_0(R) &= P_2(C_{02}-C_{22})p(R|H_2) + P_1(C_{01}-C_{11})p(R|H_1)\\ I_1(R) &= P_0(C_{10}-C_{00})p(R|H_0) + P_2(C_{12}-C_{22})p(R|H_2)\\ I_2(R) &= P_0(C_{20}-C_{00})p(R|H_0) + P_1(C_{21}-C_{11})p(R|H_1)\,. \end{aligned} \tag{14}$$

Then the optimal decision is made based on the relative magnitude of $I_i(R)$. For example, our decision rule should be

$$\begin{array}{lll} \text{if} & I_0(R) \le \min(I_1(R), I_2(R)) & \text{decide } 0 \\ \text{if} & I_1(R) \le \min(I_0(R), I_2(R)) & \text{decide } 1 \\ \text{if} & I_2(R) \le \min(I_0(R), I_1(R)) & \text{decide } 2\,. \end{array} \tag{15}$$

Based on the results above it seems that for a general M decision hypothesis Bayes test we can write the risk as

$$\mathcal{R} = \sum_{i=0}^{M-1} P_i C_{ii} + \sum_{i=0}^{M-1} \int_{Z_i} \left(\sum_{j=0; j\neq i}^{M-1} P_j (C_{ij} - C_{jj}) p(R|H_j) \right) dR\,.$$

Note that in the above expression the first term, $\sum_{i=0}^{M-1} P_i C_{ii}$, is a fixed cost and cannot be changed regardless of the decision region selected. The second term in $\mathcal{R}$ or

$$\sum_{i=0}^{M-1} \int_{Z_i} \left(\sum_{j=0; j\neq i}^{M-1} P_j (C_{ij} - C_{jj}) p(R|H_j) \right) dR\,,$$

is the average cost accumulated when we incorrectly assign a sample to the regions $i = 0, 1, 2, \cdots, M-1$. Thus we should define Z_i to be the points R such that the integrand evaluated at that R is smaller than all other possible integrand. Thus if we define

$$I_i(R) \equiv \sum_{j=0; j\neq i}^{M-1} P_j (C_{ij} - C_{jj}) p(R|H_j)\,,$$

a point R will be classified as from Z_i if it has $I_i(R)$ the smallest from all possible values of $I_\cdot(R)$. That is we classify R as from H_i when

$$I_i(R) \le \min(I_j(R)) \quad \text{for} \quad 0 \le j \le M-1\,. \tag{16}$$

The book presents this decision region as the equations for the three class case $M = 3$ as

$$P_1(C_{01} - C_{11})\Lambda_1(R) > P_0(C_{10} - C_{00}) + P_2(C_{12} - C_{02})\Lambda_2(R) \quad \text{then } H_1 \text{ or } H_2 \tag{17}$$

$$P_2(C_{02} - C_{22})\Lambda_2(R) > P_0(C_{20} - C_{00}) + P_1(C_{21} - C_{01})\Lambda_1(R) \quad \text{then } H_2 \text{ or } H_1 \tag{18}$$

$$P_2(C_{12} - C_{22})\Lambda_2(R) > P_0(C_{20} - C_{10}) + P_1(C_{21} - C_{11})\Lambda_1(R) \quad \text{then } H_2 \text{ or } H_0\,. \tag{19}$$

We can determine the decision regions in Λ_1 and Λ_2 space when we replace all inequalities with equalities. In that case each of the above equalities would then be a line. We can then solve for the values of Λ_1 and Λ_2 that determine the intersection point by solving these three equations for Λ_1 and Λ_2. In addition, the linear decision regions in Λ_1 and Λ_2 space can be plotted by taking each inequality as an equality and plotting the given lines.

Notes on a degenerate test

For a three class classification problem with cost assignments given by

$$C_{12} = C_{21} = 0 \tag{20}$$
$$C_{01} = C_{10} = C_{20} = C_{02} = C \tag{21}$$
$$C_{00} = C_{11} = C_{22} = 0\,, \tag{22}$$

when we use Equation 17 we get

$$\text{if} \quad P_1 C \Lambda_1(R) > P_0 C + P_2(-C)\Lambda_2(R) \quad \text{then } H_1 \text{ or } H_2 \text{ else } H_0 \text{ or } H_2\,,$$

While Equation 18 gives

$$\text{if} \quad P_2 C \Lambda_2(R) > P_0 C + P_1(-C)\Lambda_1(R) \quad \text{then } H_2 \text{ or } H_1 \text{ else } H_0 \text{ or } H_1\,,$$

If we divide both of these by C we see that they are equivalent to

$$\begin{aligned} &\text{if} \quad P_1\Lambda_1(R) + P_2\Lambda_2(R) > P_0 \quad \text{then } H_1 \text{ or } H_2 \text{ else } H_0 \text{ or } H_2 \\ &\text{if} \quad P_1\Lambda_1(R) + P_2\Lambda_2(R) > P_0 \quad \text{then } H_2 \text{ or } H_1 \text{ else } H_0 \text{ or } H_1\,. \end{aligned}$$

These two expressions combine to give the single expression

$$\text{if} \quad P_1\Lambda_1(R) + P_2\Lambda_2(R) > P_0 \quad \text{then } H_1 \text{ or } H_2 \text{ else } H_0\,.$$

Notes on a dummy hypothesis test

We take $P_0 = 0$ and then $P_1 + P_2 = 1$ with $C_{12} = C_{02}$ and $C_{21} = C_{01}$. Then when we put these simplifications into the M dimensional decision problem we get

$$\begin{aligned} &P_1(C_{21} - C_{11})\Lambda_1(R) > 0 \\ &P_2(C_{12} - C_{22})\Lambda_2(R) > 0 \\ &P_2(C_{12} - C_{22})\Lambda_2(R) > P_1(C_{21} - C_{11})\,. \end{aligned} \tag{23}$$

The first two equations state that we should pick H_1 or H_2 depending on the magnitudes of the costs.

Notes on Random Parameters: Bayes Estimation with a Uniform Cost

We start with the definition risk given a cost function $C(\cdot)$ and a method at estimating A (i.e. the function $\hat{a}(R)$) given by

$$\mathcal{R}_{\text{uniform}} = \int_{-\infty}^{\infty} dR\, p(R) \int_{-\infty}^{\infty} dA\, C(A - \hat{a}(R)) p(A|R)\,,$$

where $C(\cdot)$ is the uniform cost which is 1 except in a window of size Δ centered around where $C(\cdot)$ is zero. That is

$$\int_{-\infty}^{\infty} dA\, C(A - \hat{a}(R))p(A|R) = 1 - \int_{|A-\hat{a}(R)|\leq\frac{\Delta}{2}} dAp(A|R)$$
$$= 1 - \int_{\hat{a}(R)-\frac{\Delta}{2}}^{\hat{a}(R)+\frac{\Delta}{2}} p(A|R)dA\,, \tag{24}$$

which is used to derive the equation for $\mathcal{R}_{\text{uniform}}$ in the book.

Notes on Estimation Theory: Example 2

The expression for $p(A|R)$ via the books equation 141 follow from the arguments given in the book. Once that expression is accepted we can manipulate it by first writing it as

$$p(A|R) = k'(R)\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_n^2}\sum_{i=1}^{N}R_i^2 - \frac{2A}{\sigma_n^2}\sum_{i=1}^{N}R_i + \frac{NA^2}{\sigma_n^2} + \frac{A^2}{\sigma_a^2}\right]\right\}.$$

Note that the coefficient of A^2 in the above is $\frac{N}{\sigma_n^2} + \frac{1}{\sigma_a^2}$. If we define

$$\sigma_p^2 \equiv \left(\frac{1}{\sigma_a^2} + \frac{N}{\sigma_n^2}\right)^{-1} = \left(\frac{\sigma_n^2 + N\sigma_a^2}{\sigma_a^2\sigma_n^2}\right)^{-1} = \frac{\sigma_a^2\sigma_n^2}{N\sigma_a^2 + \sigma_n^2}\,,$$

then the expression for $p(A|R)$ becomes

$$\begin{aligned}
p(A|R) &= k'(R)\exp\left\{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{N}R_i^2\right\}\exp\left\{-\frac{1}{2}\left[-\frac{2A}{\sigma_n^2}\sum_{i=1}^{N}R_i + \frac{A^2}{\sigma_p^2}\right]\right\} \\
&= k'(R)\exp\left\{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{N}R_i^2\right\}\exp\left\{-\frac{1}{2\sigma_p^2}\left[A^2 - \frac{2\sigma_p^2}{\sigma_n^2}\left(\sum_{i=1}^{N}R_i\right)A\right]\right\} \\
&= k'(R)\exp\left\{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{N}R_i^2\right\}\exp\left\{-\frac{1}{2\sigma_p^2}\left[A^2 - \frac{2\sigma_p^2}{\sigma_n^2}\left(\sum_{i=1}^{N}R_i\right)A + \left(\frac{\sigma_p^2\sum_{i=1}^{N}R_i}{\sigma_n^2}\right)^2 - \left(\frac{\sigma_p^2\sum_{i=1}^{N}R_i}{\sigma_n^2}\right)^2\right]\right\} \\
&= k'(R)\exp\left\{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{N}R_i^2\right\}\exp\left\{-\frac{1}{2\sigma_p^2}\left(A - \frac{\sigma_p^2\sum_{i=1}^{N}R_i}{\sigma_n^2}\right)^2\right\}\exp\left\{-\frac{1}{2\sigma_p^2}\left[-\left(\frac{\sigma_p^2\sum_{i=1}^{N}R_i}{\sigma_n^2}\right)^2\right]\right\} \\
&= k''(R)\exp\left\{-\frac{1}{2\sigma_p^2}\left(A - \frac{\sigma_p^2\sum_{i=1}^{N}R_i}{\sigma_n^2}\right)^2\right\}.
\end{aligned}$$

Note that the mean value of the above density can be observed by inspection where we have

$$\hat{a}_{ms}(R) = \frac{\sigma_p^2}{\sigma_n^2}\sum_{i=1}^{N}R_i = \left(\frac{\sigma_a^2}{\sigma_a^2 + \frac{\sigma_n^2}{N}}\right)\left(\frac{1}{N}\sum_{i=1}^{N}R_i\right). \tag{25}$$

Notes on the optimality of the mean-square estimator

The next section of the text will answer the question, about what is the risk if we use a different estimator, say $\hat{a}$, rather than the one that we argue is optimal or $\hat{a}_{ms}$. We start with the definition of the Bayes risk in using $\hat{a}$ or

$$\mathcal{R}_{\mathcal{B}}(\hat{a}|R) = E_a[C(\hat{a}-a)|R] = \int_{-\infty}^{\infty} C(\hat{a}-a)p(a|R)da$$

If we write this in terms of $\hat{a}_{ms}$ with $z \equiv a - \hat{a}_{ms} = a - E[a|R]$ we have $a = z + \hat{a}_{ms}$ and since $p(a|R) = p(z|R)$ i.e. that the densities of a and z are the same the above is given by

$$\int_{-\infty}^{\infty} C(\hat{a}-\hat{a}_{ms}-z)p(z|R)dz\,. \tag{26}$$

If $p(z|R) = p(-z|R)$ the above is equal to

$$\int_{-\infty}^{\infty} C(\hat{a}-\hat{a}_{ms}+z)p(z|R)dz\,. \tag{27}$$

If the cost function is symmetric $C(a_\varepsilon) = C(-a_\varepsilon)$ the above is equal to

$$\int_{-\infty}^{\infty} C(\hat{a}_{ms}-\hat{a}-z)p(z|R)dz\,. \tag{28}$$

Again using symmetry of the a posteriori density $p(z|R)$ we get

$$\int_{-\infty}^{\infty} C(\hat{a}_{ms}-\hat{a}+z)p(z|R)dz\,. \tag{29}$$

We now add 1/2 of Equation 27 and 1/2 of Equation 29 and the convexity of C to get

$$\begin{aligned}
\mathcal{R}_{\mathcal{B}}(\hat{a}|R) &= \frac{1}{2}E_z[C(z+\hat{a}-\hat{a}_{ms})|R] + \frac{1}{2}E_z[C(z+\hat{a}_{ms}-\hat{a})|R] \\
&= E_z\left[\frac{1}{2}C(z+(\hat{a}_{ms}-\hat{a})) + \frac{1}{2}C(z-(\hat{a}_{ms}-\hat{a}))\middle| R\right] \\
&\geq E_z\left[C\left(\frac{1}{2}(z+(\hat{a}_{ms}-\hat{a})) + \frac{1}{2}(z-(\hat{a}_{ms}-\hat{a}))\right)\middle| R\right] \\
&= E_z[C(z)|R] = \mathcal{R}_{\mathcal{B}}(\hat{a}_{ms}|R)\,.
\end{aligned} \tag{30}$$

While all of this manipulations may seem complicated my feeling that the take away from this is that the risk of using any estimator $\hat{a} \neq \hat{a}_{ms}$ will be larger (or worse) than using $\hat{a}_{ms}$ when the cost function is convex. This is a strong argument for using the mean-squared cost function above others.

Notes on Estimation Theory: Example 3: a nonlinear dependence on a

Now the book's equation 137 used to compute the MAP estimate is

$$\left.\frac{\partial l(A)}{\partial A}\right|_{A=\hat{a}(R)} = \left.\frac{\partial \ln(p_{r|a}(R|A))}{\partial A}\right|_{A=\hat{a}(R)} + \left.\frac{\partial \ln(p_a(A))}{\partial A}\right|_{A=\hat{a}(R)} = 0\,. \tag{31}$$

This is equivalent to finding $\hat{a}(R)$ such that

$$\left.\frac{\partial \ln(p_{a|r}(A|R))}{\partial A}\right|_{A=\hat{a}(R)} = 0\,. \tag{32}$$

For this example from the functional form for $p_{a|r}(A|R)$ (now containing a nonlinear function in a) in we have our MAP estimate given by solving the following

$$\begin{aligned}
\frac{\partial \ln(p_{a|r}(A|R))}{\partial A} &= \frac{\partial}{\partial A}\left[\ln(k(R)) - \frac{1}{2}\frac{1}{\sigma_n^2}\sum_{i=1}^{N}[R_i - s(A)]^2 - \frac{1}{2}\frac{A^2}{\sigma_a^2}\right] \\
&= -\frac{1}{\sigma_n^2}\sum_{i=1}^{N}[R_i - s(A)]\left(-\frac{ds(A)}{dA}\right) - \frac{A}{\sigma_a^2} = 0\,,
\end{aligned}$$

equation for A. When we do this (and calling the solution $\hat{a}_{\text{map}}(R)$) we have

$$\hat{a}_{\text{map}}(R) = \frac{\sigma_a^2}{\sigma_n^2}\left(\sum_{i=1}^{N}[R_i - s(A)]\right)\left.\frac{\partial s(A)}{\partial A}\right|_{A=\hat{a}_{\text{map}}(R)} \tag{33}$$

which is the books equation 161.

Notes on Estimation Theory: Example 4

When the parameter A has a exponential distribution

$$p_a(A) = \begin{cases} \lambda e^{-\lambda A} & A > 0 \\ 0 & \text{otherwise} \end{cases}\,,$$

and the likelihood is given by a Poisson distribution the posteriori distribution looks like

$$\begin{aligned}
p_{a|n}(A|N) &= \frac{\Pr(n = N|a = A)p_a(A)}{\Pr(n = N)} = \frac{1}{\Pr(n = N)}\left(\frac{A^N}{N!}e^{-A}\right)\lambda e^{-\lambda A} \\
&= k(N)A^N \exp(-(1+\lambda)A)\,.
\end{aligned} \tag{34}$$

To find $k(N)$ such that this density integrate to one we need to evaluate

$$k(N)\int_0^\infty A^N \exp(-(1+\lambda)A)dA\,.$$

To do so let $v = (1+\lambda)A$ so $dv = (1+\lambda)dA$ to get

$$k(N)\int_0^\infty \left(\frac{v}{1+\lambda}\right)^N e^{-v}\frac{dv}{1+\lambda} = \frac{k(N)}{(1+\lambda)^{N+1}}\int_0^\infty v^N e^{-v}dv = \frac{k(N)N!}{(1+\lambda)^{N+1}}\,.$$

To make this equal one we need that

$$k(N) = \frac{(1+\lambda)^{N+1}}{N!}\,. \tag{35}$$

Now that we have the expression for $k(N)$ we can evaluate $\hat{a}_{ms}(N)$. We find

$$\begin{aligned}
\hat{a}_{ms}(N) &\equiv \int_0^\infty Ap(A|N)dA \\
&= \frac{(1+\lambda)^{N+1}}{N!} \int_0^\infty A^{N+1} e^{-(1+\lambda)A} dA \\
&= \frac{(1+\lambda)^{N+1}}{N!} \cdot \frac{1}{(1+\lambda)^{N+2}} \int_0^\infty v^{N+1} e^{-v} dv \\
&= \frac{N+1}{\lambda+1} \,.
\end{aligned} \quad (36)$$

To evaluate $\hat{a}_{map}(N)$ we first note from Equation 34 that

$$\ln(p(A|N)) = N\ln(A) - A(1+\lambda) + \ln(k(N)) \,,$$

so that setting the first derivative equal to zero we get

$$\frac{\partial \ln(p(A|N))}{\partial A} = \frac{N}{A} - (1+\lambda) = 0 \,.$$

Solving for A we get $\hat{a}_{map}(N)$

$$\hat{a}_{map}(N) = \frac{N}{1+\lambda} \,. \quad (37)$$

Nonrandom Parameter Estimation: The expression $\text{Var}[\hat{a}(R) - A]$

We can compute an equivalent representation of $\text{Var}[\hat{a}(R) - A]$ using its definition when A is non-random as (but R is due to measurement noise) and $E[\hat{a}(A)] = A + B(A)$ as

$$\begin{aligned}
\text{Var}[\hat{a}(R) - A] &= E[(\hat{a}(R) - A - E[\hat{a}(R) - A])^2] \\
&= E[(\hat{a}(R) - A - E[\hat{a}(R)] + A)^2] \\
&= E[(\hat{a}(R) - A - B(A))^2] \\
&= E[(\hat{a}(R) - A)^2] - 2E[\hat{a}(R) - A]B(A) + B(A)^2 \,.
\end{aligned}$$

Now the expectation of the second term is given by

$$\begin{aligned}
E[\hat{a}(R) - A] &= E[\hat{a}(R)] - A \\
&= A + B(A) - A = B(A) \,,
\end{aligned}$$

so using this the above becomes

$$\text{Var}[\hat{a}(R) - A] = E[(\hat{a}(R) - A)^2] - 2B(A)^2 + B(A)^2 = E[(\hat{a}(R) - A)^2] - B(A)^2 \,. \quad (38)$$

which is equation 173 in the book.

Notes on The Cramer-Rao Inequality Derivation

The Schwarz inequality is

$$\int f(x)g(x)dx \leq \left|\int f(x)g(x)dx\right| \leq \left(\int f(x)^2dx\right)^{1/2}\left(\int g(x)^2dx\right)^{1/2}. \tag{39}$$

We will have equality if $f(x) = kg(x)$. If we take for f and g the functions

$$\begin{aligned} f(R) &\equiv \frac{\partial \ln(p_{r|a}(R|A))}{\partial A}\sqrt{p_{r|a}(R|A)} \\ g(R) &\equiv \sqrt{p_{r|a}(R|A)}(\hat{a}(R) - R)\,, \end{aligned}$$

as the component functions in the Schwarz inequality then we find a right-hand-side (RHS) of this inequality given by

$$\text{RHS} = \left(\int\left[\frac{\partial \ln(p_{r|a}(R|A))}{\partial A}\right]^2 p_{r|a}(R|A)dR\right)^{1/2}\left(\int p_{r|a}(R|A)(\hat{a}(R) - A)^2dR\right)^{1/2},$$

and a left-hand-side (LHS) given by

$$\text{LHS} = \int \frac{\partial \ln(p_{r|a}(R|A))}{\partial A} p_{r|a}(R|A)(\hat{a}(R) - A)dR\,.$$

From the derivation in the book this LHS expression is on the right-hand-side is equal to the value of 1. Squaring both sides of the resulting inequality

$$1 \leq \left(\int\left[\frac{\partial \ln(p_{r|a}(R|A))}{\partial A}\right]^2 p_{r|a}(R|A)dR\right)^{1/2}\left(\int p_{r|a}(R|A)(\hat{a}(R) - A)^2dR\right)^{1/2},$$

gives the books equation 186. Simply dividing by the integral with the derivative and recognizing that these integrals are expectations gives

$$E[(\hat{a}(R) - A)^2] \geq \left\{E\left[\frac{\partial \ln(p_{r|a}(R|A))}{\partial A}\right]^2\right\}^{-1}, \tag{40}$$

which is one formulation of the Crammer-Rao lower bound on the value of the expression $\text{Var}[\hat{a}(R) - A]$ and is the books equation 188. From the above proof we will have an efficient estimator (one that achieve the Cramer-Rao lower bound) and the Schwarz inequality is tight when $f = kg$ or in this case

$$\frac{\partial \ln(p_{r|a}(R|A))}{\partial A}\sqrt{p_{r|a}(R|A)} = k(A)\sqrt{p_{r|a}(R|A)}\,(\hat{a}(R) - R)\,.$$

or

$$\frac{\partial \ln(p_{r|a}(R|A))}{\partial A} = k(A)(\hat{a}(R) - A)\,. \tag{41}$$

If we can write our estimator $\hat{a}(R)$ in this form then we can state that we have an efficient estimator.

Notes on Example 2: Using The Cramer-Rao Inequality

The expression for $p(R|A)$ for this example is given via the books equation 139 or

$$p_{r|a}(R|A) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(R_i - A)^2}{2\sigma_n^2}\right). \quad (42)$$

The logarithm of this is then given by

$$\ln(p_{r|a}(R|A)) = -\frac{N}{2}\ln(2\pi) - N\ln(\sigma_n) - \frac{1}{2\sigma_n^2}\sum_{i=1}^{N}(R_i - A)^2 .$$

To find the maximum likelihood solution we need to find the maximum of the above expression with respect to the variable A. The A derivative of this expression is given by

$$\frac{\ln(p_{r|a}(R|A))}{\partial A} = \frac{2}{2\sigma_n^2}\sum_{i=1}^{N}(R_i - A) = \frac{1}{\sigma_n^2}\left(\sum_{i=1}^{N} R_i - AN\right) = \frac{N}{\sigma_n^2}\left(\frac{1}{N}\sum_{i=1}^{N} R_i - A\right). \quad (43)$$

Setting this equal to zero and solving for A we get

$$\hat{a}_{ml}(R) = \frac{1}{N}\sum_{i=1}^{N} R_i . \quad (44)$$

An efficient estimator (equal to the Crammer-Rao lower bound) will have

$$\frac{\partial \ln(p_{r|a}(R|A))}{\partial A} = k(A)(\hat{a}(R) - A) .$$

we see from Equation 43 that our estimator $\hat{a}_{ml}(R)$ is of this form. As we have an efficient estimator we can evaluate the variance of it by using the Crammer-Rao inequality as an equality. The needed expression in the Crammer-Rao inequality is

$$\frac{\partial^2 \ln(p_{r|a}(R|A))}{\partial A^2} = -\frac{N}{\sigma_n^2} . \quad (45)$$

Thus we find

$$\text{Var}[\hat{a}_{\text{ml}}(R) - A] = \left(-E\left[\frac{\partial^2 \ln(p_{r|a}(R|A))}{\partial A^2}\right]\right)^{-1} = \left(\frac{N}{\sigma_n^2}\right)^{-1} = \frac{\sigma_n^2}{N} \quad (46)$$

which is the books equation 201.

Notes on Example 4: Using The Cramer-Rao Inequality

The likelihood of a for Example 4 is a Poisson random variable, given by the books equation 162 or

$$\Pr(n\,\text{events}|a = A) = \frac{A^n}{n!}\exp(-A) \quad \text{for} \quad n = 0, 1, 2, \dots . \quad (47)$$

The maximum likelihood estimate of A, after we observe the number n events, is given by finding the maximum of the density above $\Pr(n\,\text{events}|a = A)$. We can do this by setting $\frac{\partial \ln(p(n=N|A))}{\partial A}$ equal to zero and solving for A. This derivative is

$$\begin{aligned}\frac{\partial}{\partial A}\ln(\Pr(n = N|A)) &= \frac{\partial}{\partial A}(N\ln(A) - A - \ln(N!)) \\ &= \frac{N}{A} - 1 = \frac{1}{A}(N - A)\,. \end{aligned} \tag{48}$$

Setting this equal to zero and solving for A gives

$$\hat{a}_{\text{ml}}(N) = N\,. \tag{49}$$

Note that in Equation 48 we have written $\frac{\partial \ln(p(n=N|A))}{\partial A}$ in the form $k(A)(\hat{a} - A)$ and thus $\hat{a}_{ml}$ is an efficient estimator (one that achieves the Crammer-Rao bounds). Computing the variance of this estimator using this method we then need to compute

$$\frac{\partial^2 \ln(p(n = N|A))}{\partial A^2} = -\frac{N}{A^2}\,.$$

Thus using this we have

$$\text{Var}[\hat{a}_{\text{ml}}(N) - A] = \frac{1}{-E\left\{\frac{\partial^2 \ln(p(R|A))}{\partial A^2}\right\}} = \frac{1}{E\left(\frac{N}{A^2}\right)} = \frac{A^2}{E[N]} = \frac{A^2}{A} = A\,. \tag{50}$$

A bit of explanation might be needed for these manipulations. In the above $E[N]$ is the expectation of the observation N with A a fixed parameter. The distribution of N with A a fixed parameter is a Poisson distribution with mean A given by Equation 47. From facts about the Poisson distribution this expectation is A.

Note that these results can be obtained from MAP estimates in the case where our prior information is infinitely weak. For example, in example 2 weak prior information means that we should take $\sigma_a \to \infty$ in the MAP estimate of a. Using Equation 25 since $\hat{a}_{map}(R) = \hat{a}_{ms}(R)$ for this example this limit gives

$$\hat{a}_{\text{map}}(R) = \frac{\sigma_a^2}{\sigma_a^2 + (\sigma_n^2/N)}\left(\frac{1}{N}\sum_{i=1}^{N} R_i\right) \to \frac{1}{N}\sum_{i=1}^{N} R_i\,.$$

which matches the maximum likelihood estimate of A as shown in Equation 44.

In example 4, since A is distributed as an exponential with parameter λ it has a variance given by $\text{Var}[A] = \frac{1}{\lambda^2}$ see [1], so to remove any prior dependence in the MAP estimate we take $\lambda \to 0$. In that case the MAP estimate of A given by Equation 37 limits to

$$\hat{a}_{\text{map}} = \frac{N}{1 + \lambda} \to N\,,$$

which is the same as the maximum likelihood estimate Equation 49.

Notes on Example 3: Using The Cramer-Rao Inequality

Consider the expression for $p(A|R)$ for this example given in the books equation 160

$$p(A|R) = k(R) \exp\left\{ -\frac{1}{2} \left(\frac{1}{\sigma_n^2} \sum_{i=1}^{N} [R_i - s(A)]^2 + \frac{1}{\sigma_a^2} A^2 \right) \right\} . \tag{51}$$

From this expression for $p(A|R)$ to get $p(R|A)$ we would need to drop $p_a(A) \propto \exp\left\{-\frac{A^2}{2\sigma_a^2}\right\}$. When we do this and then take the logarithm of $p(R|A)$ we get

$$\ln(p(R|A)) = \ln(k'(R)) - \frac{1}{2\sigma_n^2} \sum_{i=1}^{N} [R_i - s(A)]^2 .$$

To compute $\hat{a}_{ml}(R)$ we compute $\frac{\partial \ln(p(R|A))}{\partial A}$, set this expression equal to zero and then solve for $\hat{a}_{ml}(R)$. We find the needed equation to solve given by

$$\frac{1}{\sigma_n^2} \left(\frac{\partial s(A)}{\partial A} \right) \left[\frac{1}{N} \sum_{i=1}^{N} R_i - s(A) \right] \Bigg|_{A=\hat{a}_{ml}(R)} = 0 . \tag{52}$$

To satisfy this equation either $\frac{ds}{dA} = 0$ is zero or $s(A) = \frac{1}{N} \sum_{i=1}^{N} R_i$. The second equation has a solution for $\hat{a}_{ml}(R)$ given by

$$\hat{a}_{ml}(R) = s^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} R_i \right) , \tag{53}$$

which is the books equation 209. If this estimates is unbiased we can evaluate the Kramer-Rao lower bound on $\text{Var}[\hat{a}_{ml}(R) - A]$ by computing the second derivative

$$\begin{aligned} \frac{\partial^2 \ln(p(R|A))}{\partial A^2} &= \frac{1}{\sigma_n^2} \frac{\partial^2 s}{\partial A^2} \left(\sum_{i=1}^{N} [R_i - s(A)] \right) + \frac{1}{\sigma_n^2} \frac{\partial s}{\partial A} \left(-N \frac{\partial s}{\partial A} \right) \\ &= \frac{1}{\sigma_n^2} \frac{\partial^2 s}{\partial A^2} \left(\sum_{i=1}^{N} [R_i - s(A)] \right) - \frac{N}{\sigma_n^2} \left(\frac{\partial s}{\partial A} \right)^2 . \end{aligned} \tag{54}$$

Taking the expectation of the above expression and using the fact that $E(R_i - s(A)) = E(n_i) = 0$ the first term in the above expression vanishes and we are left with the expectation of the second derivative of the log-likelihood given by

$$-\frac{N}{\sigma_n^2} \left(\frac{\partial s}{\partial A} \right)^2 .$$

Using this expectation the Cramer-Rao lower bound gives

$$\text{Var}[\hat{a}_{ml}(R) - A] \geq \frac{-1}{E\left\{ \frac{\partial^2 p(R|A)}{\partial A^2} \right\}} = \frac{\sigma_n^2}{N \left(\frac{ds(A)}{dA} \right)^2} . \tag{55}$$

We can see why we need to divide by the derivative squared when computing the variance of a nonlinear transformation form the following simple example. If we take $Y = s(A)$ and Taylor expand Y about the point $A = A_A$ where $Y_A = s(A_A)$ we find

$$Y = Y_A + \left.\frac{ds}{dA}\right|_{A=A_A} (A - A_A) + O((A - A_A)^2)\,.$$

Computing $Y - Y_A$ we then have

$$Y - Y_A = (A - A_A) \left.\frac{ds}{dA}\right|_{A=A_A}\,.$$

From this we can easily compute the variance of our nonlinear function Y in terms of the variance of the input and find

$$\text{Var}[Y - Y_A] = \left(\left.\frac{ds(A)}{dA}\right|_{A=A_A} \right)^2 \text{Var}[A - A_A]\,.$$

Which shows that a nonlinear transformation "expands" the variance of the mapped variable Y by a multiple of the derivative of the mapping.

Notes on The Cramer-Rao Bound in Estimating a Random Parameter

Starting with the conditional expectation of the error given A given by

$$B(A) = \int_{\infty}^{\infty} [\hat{a}(R) - A] p(R|A) dR\,, \quad (56)$$

when we multiply by the a priori density of A or $p(A)$ we get

$$p(A)B(A) = \int_{\infty}^{\infty} [\hat{a}(R) - A] p(R|A) p(A) dR\,.$$

Taking the A derivative of the above gives

$$\frac{d}{dA} p(A)B(A) = -\int_{-\infty}^{\infty} p(R, A) dR + \int_{-\infty}^{\infty} \frac{\partial p(R, A)}{\partial A} [\hat{a}(R) - A] dR\,. \quad (57)$$

Next we integrate the above over all space to get

$$0 = -1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial p(R, A)}{\partial A} [\hat{a}(R) - A] dR\,. \quad (58)$$

Then using the Schwarz inequality as was done on Page 13 we can get the stated lower bound on the variance of our estimator $\hat{a}(R)$. The Schwarz inequality will hold with equality if and only if

$$\frac{\partial^2 \ln(p(R, A))}{\partial A^2} = -k\,. \quad (59)$$

Since $p(R, A) = p(A|R)p(R)$ we have $\ln(p(R, A)) = \ln(p(A|R)) + \ln(p(R))$ so Equation 59 becomes

$$\frac{\partial^2 \ln(p(R, A))}{\partial A^2} = \frac{\partial^2 \ln(p(A|R))}{\partial A^2} = -k\,.$$

Then integrating this expression twice gives that $p(A|R)$ must satisfy

$$p(A|R) = \exp(-kA^2 + c_1(R)A + c_2(R))\,.$$

Notes on the proof that $\sigma_{\varepsilon_i}^2 = \text{Var}[\hat{a}_i(R) - A_i] \geq J^{ii}$

Lets verify some of the elements of $E[xx^T]$. We find

$$
\begin{aligned}
E[x_1 x_2] &= \int_{-\infty}^{\infty} (\hat{a}_1(R) - A_1) \frac{\partial \ln(p(R|A))}{\partial A_1} p(R|A) dR \\
&= \int_{-\infty}^{\infty} \hat{a}_1(R) \frac{\partial p(R|A)}{\partial A_1} dR - A_1 \int_{-\infty}^{\infty} \frac{\partial p(R|A)}{\partial A_1} dR \\
&= 1 - A_1 \frac{\partial}{\partial A_1} \int_{-\infty}^{\infty} p(R|A) dR \quad \text{using the book's equation 264 for the first term} \\
&= 1 - A_1 \frac{\partial}{\partial A_1} 1 = 1 - 0 = 1 \,. \qquad (60)
\end{aligned}
$$

Notes on the general Gaussian problem: Case 3

The book has shown that $l(R) \equiv \Delta m^T Q R$ and the transformation from primed to "unprimed" variables looks like

$$\Delta m = W^{-1} \Delta m' \quad \text{and} \quad R = W^{-1} R' \,,$$

thus in the primed coordinate system we have

$$l(R') = \Delta m'^T W^{-T} Q W^{-1} R' \,.$$

Recall that W^T is the matrix containing the eigenvectors of K as its column values, since Q is defined as $Q = K^{-1}$ we can conclude that

$$K W^T = W^T \Lambda \quad \text{is the same as} \quad Q^{-1} W^T = W^T \Lambda \,,$$

so inverting both sides gives $W^{-T} Q = \Lambda^{-1} W^{-T}$. Multiply this last expression by W^{-1} on the right gives $W^{-T} Q W^{-1} = \Lambda^{-1} W^{-T} W^{-1}$. Since the eigenvectors are orthogonal $W W^T = I$ so $W^{-T} W^{-1} = I$ and we obtain

$$W^{-T} Q W^{-1} = \Lambda^{-1} \,.$$

Using this expression we see that $l(R')$ becomes

$$l(R') = \Delta m'^T \Lambda^{-1} R' = \sum_{i=1}^{N} \frac{\Delta m_i' R_i'}{\lambda_i} \,. \qquad (61)$$

In the same way we find that d^2 becomes

$$
\begin{aligned}
d^2 &= \Delta m^T Q \Delta m = \Delta m'^T W^{-T} Q W^{-1} \Delta w' \\
&= \Delta m'^T \Lambda^{-1} \Delta m' = \sum_{i=1}^{N} \frac{(\Delta m_i')^2}{\lambda_i} \,. \qquad (62)
\end{aligned}
$$

If $\rho > 0$ then $m'_{11} = 0$ and $m'_{12} = 1$ will maximize d^2. In terms of m_{11} and m_{12} this means that

$$\frac{m_{11} + m_{12}}{\sqrt{2}} = 0 \quad \text{and} \quad \frac{m_{11} - m_{12}}{\sqrt{2}} = 1\,.$$

Solving for m_{11} and m_{12} we get $\left[\begin{array}{c} m_{11} \\ m_{12} \end{array}\right] = \frac{1}{\sqrt{2}} \left[\begin{array}{c} 1 \\ -1 \end{array}\right] = \phi_2$ Note that ϕ_2 is the eigenvector corresponding to the smaller eigenvalue (when $\rho > 0$).

If $\rho < 0$ then $m'_{11} = 1$ and $m'_{12} = 0$ will maximize d^2. In terms of m_{11} and m_{12} this means that

$$\frac{m_{11} + m_{12}}{\sqrt{2}} = 1 \quad \text{and} \quad \frac{m_{11} - m_{12}}{\sqrt{2}} = 0\,.$$

Solving for m_{11} and m_{12} we get $\left[\begin{array}{c} m_{11} \\ m_{12} \end{array}\right] = \frac{1}{\sqrt{2}} \left[\begin{array}{c} 1 \\ 1 \end{array}\right] = \phi_1$ Note that ϕ_1 is again the eigenvector corresponding to the smaller eigenvalue (when $\rho < 0$).

When $m_1 = m_2 = m$ we get for the H_1 decision boundary

$$\frac{1}{2}(R-m)^T Q_0 (R-m) - \frac{1}{2}(R-m)^T Q_1 (R-m) > \ln(\eta) + \frac{1}{2}\ln|K_1| - \frac{1}{2}\ln|K_0| \equiv \gamma^*\,.$$

We can write the left-hand-side of the above as

$$\frac{1}{2}\left[(R-m)^T Q_0 - (R-m)^T Q_1\right](R-m) = \frac{1}{2}(R-m)^T (Q_0 - Q_1)(R-m)\,. \tag{63}$$

$$Q_n = \frac{1}{\sigma_n^2}\left(I + \frac{1}{\sigma_n^2}K_s\right)^{-1} = \frac{1}{\sigma_n^2}[I - H]\,,$$

so

$$\left(I + \frac{1}{\sigma_n^2}K_s\right)^{-1} = I - H\,,$$

solving for H we get

$$\begin{aligned} H &= I - \left(I + \frac{1}{\sigma_n^2}K_s\right)^{-1} \qquad (64) \\ &= \left(I + \frac{1}{\sigma_n^2}K_s\right)^{-1}\left[\left(I + \frac{1}{\sigma_n^2}K_s\right) - I\right] \\ &= \left(I + \frac{1}{\sigma_n^2}K_s\right)^{-1}\left(\frac{1}{\sigma_n^2}K_s\right) \\ &= \left(\sigma_n^2 I + K_s\right)^{-1} K_s\,. \end{aligned}$$

Also factor the inverse out on the right of Equation 64 to get

$$\left(\frac{1}{\sigma_n^2}\right)\left(I + \frac{1}{\sigma_n^2}K_s\right)^{-1} = K_s(\sigma_n^2 I + K_s)^{-1}\,.$$

Notice that from the functional forms of Q_0 and Q_1 we can write this as

$$I - \left(I + \frac{1}{\sigma_n^2} K_s \right)^{-1} = \sigma_n^2 Q_0 - \sigma_n^2 Q_1 = \sigma_n^2 \Delta Q \,, \tag{65}$$

We have

$$P_F = \int_{\gamma''}^{\infty} (2^{N/2} \sigma_n^N \Gamma(N/2))^{-1} L^{N/2-1} e^{-L/2\sigma_n^2} dL = 1 - \int_0^{\gamma''} (2^{N/2} \sigma_n^N \Gamma(N/2))^{-1} L^{N/2-1} e^{-L/2\sigma_n^2} dL \,,$$

Since the integrand is a density and must integrate to one. If we assume N is even and let $M = \frac{N}{2} - 1$ (an integer) then

$$\Gamma\left(\frac{N}{2}\right) = \Gamma(M+1) = M! \,.$$

Lets change variables in the integrand by letting $x = \frac{L}{2\sigma_n^2}$ (so that $dx = \frac{dL}{2\sigma_n^2}$) and then the expression for P_F becomes

$$\begin{aligned} P_F &= 1 - \int_0^{\frac{\gamma''}{2\sigma_n^2}} (2^{N/2} \sigma_n^N)^{-1} \left(\frac{1}{M!} \right) (2\sigma_n^2)^{\frac{N}{2}-1} x^M e^{-x} (2\sigma_n^2) dx \\ &= 1 - \int_0^{\frac{\gamma''}{2\sigma_n^2}} \frac{x^M}{M!} e^{-x} dx \,. \end{aligned} \tag{66}$$

Using the same transformation of the integrand used above (i.e. letting $x = \frac{L}{2\sigma_n^2}$) we can write P_F in terms of x as

$$P_F = \int_{\gamma'''}^{\infty} \frac{x^M}{M!} e^{-x} dx \,.$$

We next integrate this by parts M times as

$$\begin{aligned} P_F &= \frac{1}{M!} \left[-x^M e^{-x} \Big|_{\gamma'''}^{\infty} + M \int_{\gamma'''}^{\infty} x^{M-1} e^{-x} dx \right] \\ &= \frac{1}{M!} \left[\gamma'''^M e^{-\gamma'''} + M \int_{\gamma'''}^{\infty} x^{M-1} e^{-x} dx \right] \\ &= \frac{\gamma'''^M}{M!} e^{-\gamma'''} + \frac{1}{(M-1)!} \int_{\gamma'''}^{\infty} x^{M-1} e^{-x} dx \quad \text{first integration by parts} \\ &= \frac{\gamma'''^M}{M!} e^{-\gamma'''} + \frac{1}{(M-1)!} \left[-x^{M-1} e^{-x} \Big|_{\gamma'''}^{\infty} + (M-1) \int_{\gamma'''}^{\infty} x^{M-2} e^{-x} dx \right] \\ &= \frac{\gamma'''^M}{M!} e^{-\gamma'''} + \frac{\gamma'''^{M-1}}{(M-1)!} e^{-\gamma'''} + \frac{1}{(M-2)!} \int_{\gamma'''}^{\infty} x^{M-2} e^{-x} dx \quad \text{second integration by parts} \\ &= e^{-\gamma'''} \left(\sum_{k=M,M-1,M-2,\cdots}^{2} \frac{\gamma'''^k}{k!} \right) + \frac{1}{(M-(M-1))} \int_{\gamma'''}^{\infty} x e^{-x} dx \\ &= e^{-\gamma'''} \left(\sum_{k=2}^{M} \frac{\gamma'''^k}{k!} \right) - x e^{-x} \Big|_{\gamma'''}^{\infty} + \int_{\gamma'''}^{\infty} e^{-x} dx = e^{-\gamma'''} \left(\sum_{k=2}^{M} \frac{\gamma'''^k}{k!} \right) + \gamma''' e^{-\gamma'''} + e^{-\gamma'''} \,. \end{aligned}$$

Thus

$$P_F = e^{-\gamma'''} \sum_{k=0}^{M} \frac{\gamma'''^k}{k!} \,. \quad (67)$$

If $\gamma''' \gg 1$ and M is not too large then the largest term is $\frac{\gamma'''^M}{M!}$ is the largest and we can factor it out to get

$$\begin{aligned} P_F &= \frac{(\gamma''')^M e^{-\gamma'''}}{M!} \sum_{k=0}^{M} (\gamma''')^{k-M} \left(\frac{M!}{k!} \right) \\ &= \frac{(\gamma''')^M e^{-\gamma'''}}{M!} \left(1 + \frac{M}{\gamma'''} + \frac{M(M-1)}{\gamma'''^2} + \frac{M(M-1)(M-2)}{\gamma'''^3} + \cdots \right) . \quad (68) \end{aligned}$$

If we drop the terms after the second in the above expansion and recall that $(1+x)^{-1} \approx 1-x$ when $x \ll 1$ we can write

$$P_F \approx \frac{(\gamma''')^M e^{-\gamma'''}}{M!} \left(1 - \frac{M}{\gamma'''} \right)^{-1} . \quad (69)$$

Problem Solutions

The conventions of this book dictate that *lower* case letters (like y) indicate a random variable while *capital* case letters (like Y) indicate a particular realization of the random variable y. To maintain consistency with the book I'll try to stick to that notation. This is mentioned because other books use the *opposite* convention like [1] which could introduce confusion.

Problem 2.2.1 (a Likelihood Ratio Test (LRT))

Part (1): We assume a hypothesis of

$$H_1 \;:\; r = s + n \quad (70)$$
$$H_0 \;:\; r = n \,, \quad (71)$$

where both s and n are exponentially distributed. For example for s we have

$$p_s(S) = \begin{cases} ae^{-aS} & S \geq 0 \\ 0 & S < 0 \end{cases} . \quad (72)$$

A similar expression holds for $p_n(N)$. Now from properties of the exponential distribution the mean of s is $\frac{1}{a}$ and the mean of n is $\frac{1}{b}$. We hope in a physical application that the mean of s is larger than the mean of n. Thus we should expect that $b > a$. I'll assume this condition in what follows.

We now need to compute $p(R|H_0)$ and $p(R|H_1)$. The density is $p(R|H_0)$ is already given. To compute $p(R|H_1)$ we can use the fact that the probability density function (PDF) of the sum of two random variables is the convolution of the individual PDFs or

$$p(R|H_1) = \int_{-\infty}^{\infty} p_s(R-n)p_n(n)dn\,.$$

Since the domain of n is $n \geq 0$ the function $p_n(N)$ vanishes for $N < 0$ and the lower limit on the above integral becomes 0. The upper limit is restricted by recognizing that the argument to $p_s(R-n)$ will be negative for n sufficiently larger. For example, when $R-n<0$ or $n>R$ the density $p_S(R-n)$ will cause the integrand to vanish. Thus we need to evaluate

$$\begin{aligned} p(R|H_1) &= \int_0^R p_s(R-n)f_n(n)dn = \int_0^R ae^{-a(R-n)}be^{-bn}dn \\ &= \frac{ab}{b-a}(e^{-aR}-e^{-bR})\,, \end{aligned}$$

when we integrate and simplify some. This function has a domain given by $0 < R < \infty$ and is zero otherwise. As an aside, we can check that the above density integrates to one (as it must). We have

$$\begin{aligned} \int_0^{\infty} \frac{ab}{b-a}(e^{-aR}-e^{-bR})dR &= \frac{ab}{b-a}\left[\frac{e^{-aR}}{-a}+\frac{e^{-bR}}{b}\right|_0^{\infty} \\ &= 0 - \frac{ab}{b-a}\left(-\frac{1}{a}+\frac{1}{b}\right) = 1\,, \end{aligned}$$

when we simplify. The likelihood ratio then decides H_1 when

$$\Lambda(R) = \frac{\frac{ab}{b-a}(e^{-aR}-e^{-bR})}{be^{-bR}} > \eta \equiv \frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})}\,,$$

and H_0 otherwise. We can write the above LRT as

$$\frac{a}{b-a}(e^{(b-a)R}-1) > \eta\,.$$

If we solve the above for R we get

$$R > \frac{1}{b-a}\ln\left[\left(\frac{b-a}{a}\right)\eta+1\right] \equiv \gamma\,.$$

If R is not larger than γ we declare H_0.

Part (2): We would replace η in the above expression with $\frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})}$.

Part (3): For Neyman-Pearson test (in the standard form) we fix a value of

$$P_F \equiv \Pr(\text{say } H_1|H_0 \text{ is true})$$

say α and seek to maximize P_D. Since we have shown that the LRT in this problem is equivalent to $R > \gamma$ we have

$$\Pr(\text{say } H_1|H_0 \text{ is true}) = \Pr(R > \gamma|H_0 \text{ is true})\,.$$

We can calculate the right-hand-side as

$$\begin{aligned}
\Pr(R > \gamma | H_0 \text{ is true}) &= \int_\gamma^\infty p_n(N) dN \\
&= \int_\gamma^\infty b e^{-bN} dN = b \left(- \frac{e^{-bN}}{b} \right|_\gamma^\infty = e^{-b\gamma} \,.
\end{aligned}$$

At this equals P_F we can write γ as a function of P_F as $\gamma = -\frac{1}{b} \ln(P_F)$.

Problem 2.2.2 (exponential and Gaussian hypothesis test)

Part (1): We find

$$\begin{aligned}
\Lambda(R) &= \frac{p(R|H_1)}{p(R|H_0)} = \frac{\sqrt{2\pi}}{2} \exp\left\{ -|R| + \frac{1}{2} R^2 \right\} \\
&= \sqrt{\frac{\pi}{2}} \exp\left\{ \frac{1}{2}(R^2 - 2|R| + 1) - \frac{1}{2} \right\} = \sqrt{\frac{\pi}{2}} \exp\left\{ \frac{1}{2}(|R| - 1)^2 - \frac{1}{2} \right\} \\
&= \sqrt{\frac{\pi}{2}} e^{-\frac{1}{2}} \exp\left\{ \frac{1}{2}(|R| - 1)^2 \right\} \,.
\end{aligned}$$

Part (2): The LRT says to decide H_1 when $\Lambda(R) > \eta$ and decide H_0 otherwise. From the above expression for $\Lambda(R)$ this can be written as

$$\frac{1}{2}(|R| - 1)^2 > \ln\left(\sqrt{\frac{2}{\pi}} e^{\frac{1}{2}} \eta \right) \,,$$

or simplifying some

$$|R| > \pm \sqrt{2 \ln\left(\sqrt{\frac{2}{\pi}} e^{\frac{1}{2}} \eta \right)} + 1 \,.$$

If we plot the two densities we get the result shown in Figure 1. See the caption on that plot for a description.

Problem 2.2.3 (nonlinear hypothesis test)

Our two hypothesis are

$$\begin{aligned}
H_1 &: y = x^2 \\
H_0 &: y = x^3 \,,
\end{aligned}$$

where $x \sim N(0, \sigma)$. The LRT requires calculating the ratio $\frac{p(Y|H_1)}{p(Y|H_0)}$ which we will do by calculating each of the conditional densities $p(Y|H_0)$ and $p(Y|H_1)$. For this problem, the

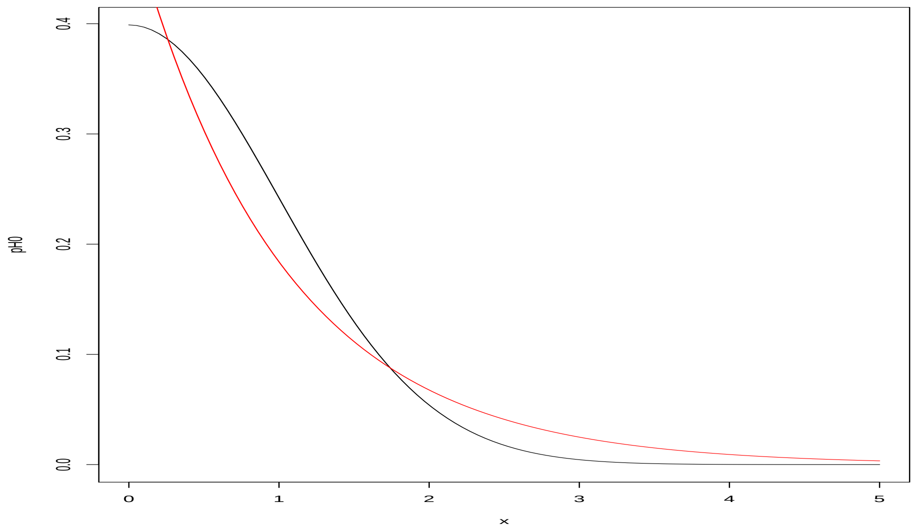

Figure 1: Plots of the density $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}R^2\right)$ (in black) and $\frac{1}{2}\exp(-|R|)$ (in red). Notice that the exponential density has fatter tails than the Gaussian density. A LRT where if $|R|$ is greater than a threshold we declare H_1 makes sense since the exponential density (from H_1) is much more likely to have large valued samples.

distribution function for y under the hypothesis H_0 is given by

$$\begin{aligned} P(Y|H_0) &= \Pr\{y \le Y|H_0\} = \Pr\{x^2 \le Y|H_0\} \\ &= \Pr\{-\sqrt{Y} \le x \le \sqrt{Y}|H_0\} = 2\int_0^{+\sqrt{Y}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\xi^2}{2\sigma^2}\right\} d\xi \,. \end{aligned}$$

The density function for Y (under H_0) is the derivative of this expression with respect to Y. We find

$$p(Y|H_0) = \frac{2}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{Y}{2\sigma^2}\right\}\left(\frac{1}{2\sqrt{Y}}\right) = \frac{1}{\sqrt{2\pi\sigma^2 Y}} \exp\left\{-\frac{Y}{2\sigma^2}\right\} .$$

Next the distribution function for y under the hypothesis H_1 is given by

$$\begin{aligned} P(Y|H_1) &= \Pr\{y \le Y|H_1\} = \Pr\{x^3 \le Y|H_1\} \\ &= \Pr\{x \le Y^{1/3}|H_1\} = \int_{-\infty}^{Y^{1/3}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\xi^2}{2\sigma^2}\right\} d\xi \,. \end{aligned}$$

Again the density function for y (under H_1) is the derivative of this expression with respect to Y. We find

$$p(Y|H_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{Y^{2/3}}{2\sigma^2}\right\}\left(\frac{1}{3Y^{2/3}}\right) = \frac{1}{3\sqrt{2\pi\sigma^2 Y^{2/3}}} \exp\left\{-\frac{Y^{2/3}}{2\sigma^2}\right\} .$$

Using these densities the LRT then gives

$$\Lambda(Y) = \frac{\frac{1}{3Y^{2/3}} \exp\left\{-\frac{Y^{2/3}}{2\sigma^2}\right\}}{\frac{1}{Y^{1/2}} \exp\left\{-\frac{Y}{2\sigma^2}\right\}} = Y^{-1/6} \exp\left\{-\frac{1}{2\sigma^2}(Y^{2/3} + Y)\right\} .$$

After receiving the measurement Y, the decision as to whether H_0 or H_1 occurred is based on the value of $\Lambda(Y)$ defined above. If $\Lambda(Y) > \eta \equiv \frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})}$ then we say H_1 occurred otherwise we say H_0 occurred.

Problem 2.2.4 (another nonlinear hypothesis test)

The distribution function for $y|H_0$ can be computed as

$$
\begin{aligned}
P(Y|H_0) &= \Pr\{x^2 \leq Y|H_0\} \\
&= \Pr\{-\sqrt{Y} \leq x \leq \sqrt{Y}|H_0\} = \int_{-\sqrt{Y}}^{\sqrt{Y}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\xi - m)^2}{2\sigma^2}\right\} d\xi \,.
\end{aligned}
$$

The density function for $y|H_0$ is the derivative of this expression. To evaluate that derivative we will use the identity

$$
\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} f(x,t)dx = \frac{d\beta}{dt} f(\beta,t) - \frac{d\alpha}{dt} f(x,t) + \int_{\alpha(t)}^{\beta(t)} \frac{\partial f}{\partial t}(x,t)dx \,. \tag{73}
$$

With this we find

$$
\begin{aligned}
p(Y|H_0) &= \frac{1}{\sqrt{2\pi\sigma^2}}\left(\exp\left\{-\frac{(\sqrt{Y}-m)^2}{2\sigma^2}\right\}\left(\frac{1}{2\sqrt{Y}}\right) - \exp\left\{-\frac{(-\sqrt{Y}-m)^2}{2\sigma^2}\right\}\left(-\frac{1}{2\sqrt{Y}}\right)\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\left(\frac{1}{2\sqrt{Y}}\right)\left(\exp\left\{-\frac{(\sqrt{Y}-m)^2}{2\sigma^2}\right\} + \exp\left\{-\frac{(\sqrt{Y}+m)^2}{2\sigma^2}\right\}\right) .
\end{aligned}
$$

The distribution function for $y|H_1$ can be computed as

$$
\begin{aligned}
P(Y|H_1) &= \Pr\{e^x \leq Y|H_1\} \\
&= \Pr\{x \leq \ln(Y)|H_1\} = \int_{-\infty}^{\ln(Y)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\xi - m)^2}{2\sigma^2}\right\} d\xi \,.
\end{aligned}
$$

Taking the derivative to get $p(y|H_1)$ we have

$$
p(Y|H_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln(Y) - m)^2}{2\sigma^2}\right\}\left(\frac{1}{Y}\right) .
$$

Using these two expressions we find the likelihood ratio test would be if $\frac{p(Y|H_1)}{p(Y|H_0)} > \eta$ then decide H_1 otherwise decide H_0.

Problem 2.2.5 (testing samples with different variances)

Part (1-2): Note that this problem is exactly the *same* as considered in Example 2 in the book but where we have taken K independent observations rather than N. Thus all formulas derived in the book are valid here after we replace N with K. The LRT expressions desired for this problem are then given by the book's Eq. 29 or Eq. 31.

Part (3): Given the decision region $l(\mathbf{R}) > \eta$ for H_1 and the opposite inequality for deciding H_0, we can write P_F and $P_M = 1 - P_D$ as

$$P_F = \Pr(\text{choose } H_1 | H_0 \text{ is true}) = \Pr\left(\sum_{i=1}^{K} R_i^2 > \gamma | H_0 \text{ is true}\right)$$
$$P_D = \Pr(\text{choose } H_1 | H_1 \text{ is true}) = \Pr\left(\sum_{i=1}^{K} R_i^2 > \gamma | H_1 \text{ is true}\right) .$$

The book discusses how to evaluate these expressions in section 6.

Part (5): According to the book, when $C_{00} = C_{11} = 0$ the minimax criterion is $C_M P_M = C_F P_F$. If $C_M = C_F$ then the minimax criterion reduces to $P_M = P_F$ or $1 - P_D = P_F$. Since both P_D and P_F are functions of γ we would solve the above expression for γ to determine the threshold γ to use in the minimax LRT.

Problem 2.2.6 (multiples of the mean)

Part (1): Given the two hypothesis H_0 and H_1 to compute the LRT we need to have the conditional probabilities $p(R|H_0)$ and $p(R|H_1)$. The density for $p(R|H_0)$ is the same as that of $p_n(N)$. To determine $p(R|H_1)$ we note that as m_1 is fixed and $b \sim N(0, \sigma_b)$ that the product $bm_1 \sim N(0, m_1\sigma_b)$. Adding an independent zero mean random variable n gives another Gaussian random variable back with a larger variance. Thus the distribution of $R|H_1$ given by

$$R|H_1 = bm_1 + n \sim N\left(0, \sqrt{\sigma_b^2 m_1^2 + \sigma_n^2}\right) .$$

Using the above density we have that the LRT is given by

$$\begin{aligned}\Lambda(R) \equiv \frac{p(R|H_1)}{p(R|H_0)} &= \frac{\frac{1}{\sqrt{2\pi}\sqrt{m_1^2\sigma_b^2+\sigma_n^2}} \exp\left(-\frac{R^2}{2(m_1^2\sigma_b^2+\sigma_n^2)}\right)}{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_n^2}} \exp\left(-\frac{R^2}{2\sigma_n^2}\right)} \\ &= \sqrt{\frac{\sigma_n^2}{m_1^2\sigma_b^2 + \sigma_n^2}} \exp\left\{\frac{1}{2}\frac{m_1^2\sigma_b^2}{(m_1^2\sigma_b^2 + \sigma_n^2)\sigma_n^2} R^2\right\} ,\end{aligned}$$

when we simplify. We pick H_1 when $\Lambda(R) > \eta$ or

$$R^2 > \left(\frac{2(m_1^2\sigma_b^2 + \sigma_n^2)\sigma_n^2}{m_1^2\sigma_b^2}\right) \ln\left(\eta\sqrt{\frac{m_1^2\sigma_b^2 + \sigma_n^2}{\sigma_n^2}}\right) .$$

Taking the square root we decide H_1 when

$$|R| > \sqrt{\left(\frac{2(m_1^2\sigma_b^2 + \sigma_n^2)\sigma_n^2}{m_1^2\sigma_b^2}\right) \ln\left(\eta\sqrt{\frac{m_1^2\sigma_b^2 + \sigma_n^2}{\sigma_n^2}}\right)} \equiv \gamma . \tag{74}$$

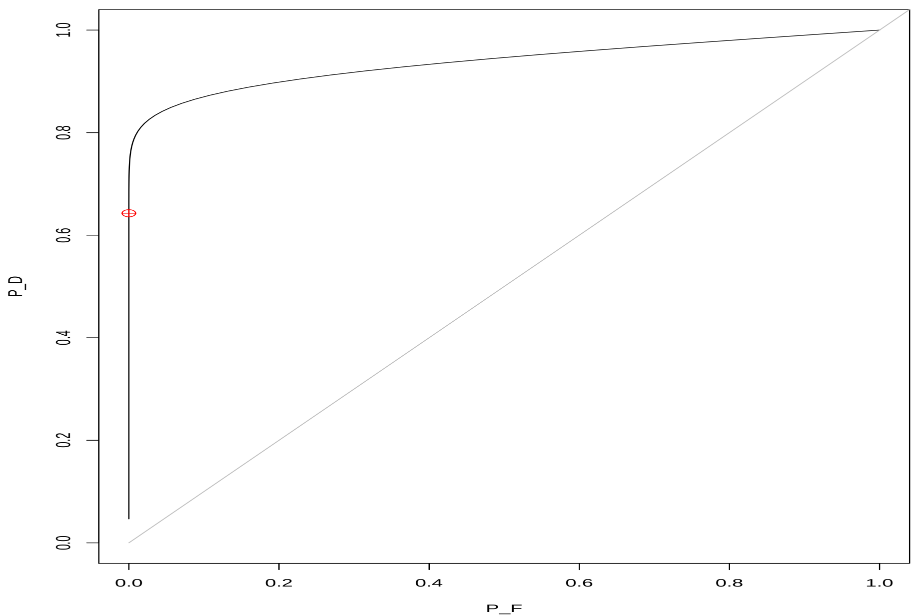

Figure 2: The ROC curve for Problem 2.2.6. The minimum probability of error is denoted with a red marker.

The optimal processor takes the absolute value of the number R and compares its value to a threshold γ.

Part (2): To draw the ROC curve we need to compute P_F and P_D. We find

$$P_F = \int_{|R|>\gamma} p(R|H_0)dR = 2\int_{R=\gamma}^{\infty} p(R|H_0)dR = 2\int_{R=\gamma}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{1}{2}\frac{R^2}{\sigma_n^2}\right\} dR \,.$$

Let $v = \frac{R}{\sigma_n}$ so that $dR = \sigma_n dv$ and we find

$$P_F = 2\int_{v=\frac{\gamma}{\sigma_n}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}v^2\right\} dv = 2\text{erfc}_*\left(\frac{\gamma}{\sigma_n}\right)$$

For P_D we find

$$\begin{aligned} P_D &= 2\int_{R=\gamma}^{\infty} p(R|H_1)dR = 2\int_{R=\gamma}^{\infty} \frac{1}{\sqrt{m_1^2\sigma_b^2 + \sigma_n^2}} \exp\left\{-\frac{1}{2}\frac{R^2}{m_1^2\sigma_b^2 + \sigma_n^2}\right\} dR \\ &= 2\int_{v=\frac{\gamma}{\sqrt{m_1^2\sigma_b^2+\sigma_n^2}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}v^2\right\} dv = 2\text{erfc}_*\left(\frac{\gamma}{\sqrt{m_1^2\sigma_b^2 + \sigma_n^2}}\right) \,. \end{aligned}$$

To plot the ROC curve we plot the points (P_F, P_D) as a function of γ. To show an example of this type of calculation we need to specify some parameters values. Let $\sigma_n = 1$, $\sigma_b = 2$, and $m_1 = 5$. With these parameters in the `R` code `chap_2_prob_2.2.6.R` we obtain the plot shown in Figure 2.

Part (3): When we have $P_0 = P_1 = \frac{1}{2}$ we find

$$\Pr(\epsilon) = P_0 P_F + P_1 P_M = \frac{1}{2} P_F(\gamma) + \frac{1}{2}(1 - P_D(\gamma)) = \frac{1}{2} + \frac{1}{2}(P_F(\gamma) - P_D(\gamma)) .$$

The minimum probability of error (MPE) rule has $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$ and thus $\eta = 1$. With this value of η we would evaluate Equation 74 to get the value of γ of the MPE decision threshold. This threshold gives $\Pr(\epsilon) = 0.1786$.

Problem 2.2.7 (a communication channel)

Part (1): Let S_1 and S_2 be the events that our data was generated via the source 1 or 2 respectively. Then we want to compute $P_F \equiv \Pr(\text{Say } S_2|S_1)$ and $P_D \equiv \Pr(\text{Say } S_1|S_2)$. Now all binary decision problems are likelihood ratio problems where we must compute $\frac{p(R|S_2)}{p(R|S_1)}$. We will assume that R in this case is a sequence of N outputs from the communication system. That is an example R (for $N = 9$) might look like

$$\mathbf{R} = \left[\begin{array}{ccccccccc} a & a & b & a & b & b & a & a & a \end{array} \right] .$$

We assume that the source is held constant for the entire length of the sequence of R. Let r_i be one of the N samples of the vector $\mathbf{R}$. Since each of the r_i outputs are independent of the others given the source we can evaluate $p(R|S_1)$ as

$$p(R|S_1) = \prod_{i=1}^{N} p(r_i|S_1) = p(r = a|S_1)^{N_a} p(r = b|S_1)^{N-N_a} .$$

Here N_a is the number of a output and $N - N_a = N_b$ is the number of b output in our sample of N total outputs. A similar type of an expression will hold for $p(R|S_2)$. Based on these expressions we need to compute $p(r|S_i)$ which we can do from the numbers given in the problem. Let I_0 and I_1 be the events that the given source emits a 0 or a 1. Then we have

$$\begin{aligned} p(r = a|S_1) &= p(r = a|I_0, S_1)p(I_0|S_1) + p(r = a|I_1, S_1)p(I_1|S_1) = (0.4)(0.5) + (0.7)(0.5) = 0.55 \\ p(r = b|S_1) &= p(r = b|I_0, S_1)p(I_0|S_1) + p(r = b|I_1, S_1)p(I_1|S_1) = (0.6)(0.5) + (0.3)(0.5) = 0.45 \\ p(r = a|S_2) &= p(r = a|I_0, S_2)p(I_0|S_2) + p(r = a|I_1, S_2)p(I_1|S_2) = (0.4)(0.4) + (0.7)(0.6) = 0.58 \\ p(r = b|S_2) &= p(r = b|I_0, S_2)p(I_0|S_2) + p(r = b|I_1, S_2)p(I_1|S_2) = (0.6)(0.4) + (0.3)(0.6) = 0.42 . \end{aligned}$$

With what we have thus far the LRT (will say to decide S_2) if

$$\begin{aligned} \Lambda(R) \equiv \frac{p(r = a|S_2)^{N_a} p(r = b|S_2)^{N-N_a}}{p(r = a|S_1)^{N_a} p(r = b|S_1)^{N-N_a}} &= \left(\frac{p(r = a|S_2)}{p(r = a|S_1)} \right)^{N_a} \left(\frac{p(r = b|S_1)}{p(r = b|S_2)} \right)^{N_a} \left(\frac{p(r = b|S_2)}{p(r = b|S_1)} \right)^{N} \\ &= \left(\frac{p(r = a|S_2)}{p(r = b|S_2)} \frac{p(r = b|S_1)}{p(r = a|S_1)} \right)^{N_a} \left(\frac{p(r = b|S_2)}{p(r = b|S_1)} \right)^{N} > \eta . \end{aligned}$$

For the numbers in this problem we find

$$\frac{p(r = a|S_2)}{p(r = b|S_2)} \frac{p(r = b|S_1)}{p(r = a|S_1)} = \frac{(0.58)(0.45)}{(0.42)(0.55)} = 1.129870 > 1 .$$

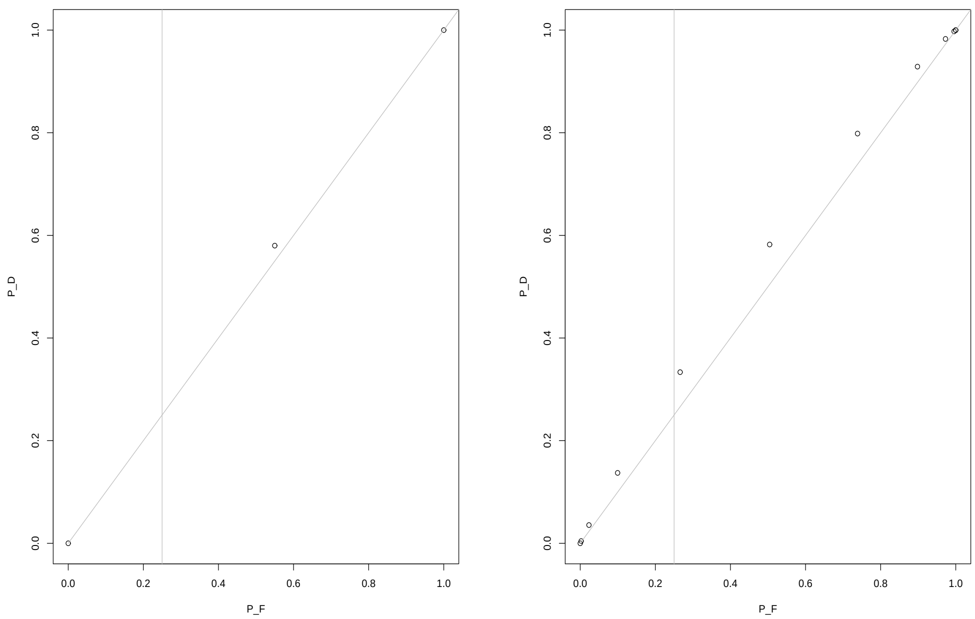

Figure 3: The ROC curve for Problem 2.2.7. **Left:** When $N = 1$. **Right:** When $N = 10$. A vertical line is drawn at the desired value of $P_F = \alpha = 0.25$.

We can solve for N_a (and don't need to flip the inequality since the log is positive) to find

$$N_a > \frac{\eta \left(\frac{p(r=b|S_1)}{p(r=b|S_2)} \right)^N}{\ln\left(\frac{p(r=a|S_2)}{p(r=b|S_2)} \frac{p(r=b|S_1)}{p(r=a|S_1)} \right)} \equiv \gamma \,.$$

Since N_a can only be integer valued for $N_a = 0, 1, 2, \cdots, N$ we only need to consider integer values of γ say for $\gamma_I = 0, 1, 2, \cdots, N+1$. Note that at the limit of $\gamma_I = 0$ the LRT $N_a \geq 0$ is always true, so we always declare S_2 and obtain the point $(P_F, P_D) = (1, 1)$. The limit of $\gamma_I = N+1$ the LRT of $N_a \geq N+1$ always fails so we always declare S_1 and obtain the point $(P_F, P_D) = (0, 0)$. Since N_a is the count of the number of as from N it is a binomial random variable (under both S_1 and S_2) and once γ_I is specified, we have P_F and P_D given by

$$P_F = \Pr\{N_a \geq \gamma_I | S_1\} = \sum_{k=\gamma_I}^{N} \binom{N}{k} p(r=a|S_1)^k p(r=b|S_1)^{N-k}$$

$$P_D = \Pr\{N_a \geq \gamma_I | S_2\} = \sum_{k=\gamma_I}^{N} \binom{N}{k} p(r=a|S_2)^k p(r=b|S_2)^{N-k} \,.$$

To plot the ROC curve we evaluate P_F and P_D for various values of γ_I. This is done in the R code `chap_2_prob_2.2.7.R`. Since the value of $P_F = \alpha = 0.25$ does not exactly fall on a integral value for γ_I we must use a randomized rule to achieve the desired performance. Since we are not told what N is (the number of samples observed we will consider two cases $N = 1$ and $N = 10$.

In the case $N = 1$ the desired value $P_F = \alpha = 0.25$ falls between the two points $P_F(\gamma_I = 2) = 0$ and $P_F(\gamma_I = 1) = 0.55$. To get the target value of 0.25 we need to introduce the probability that we will use the threshold $\gamma_I = 2$ as $p_{\gamma=2}$. The complement of this probability or $1 - p_{\gamma=2}$ is the probability that we use the threshold $\gamma_I = 1$. Then to get the desired false alarm rate we need to take $p_{\gamma=2}$ to satisfy

$$p_{\gamma=2} P_F(\gamma_I = 2) + (1 - p_{\gamma=2}) P_F(\gamma_I = 1) = 0.25 .$$

Putting in what we know for $P_F(\gamma_I = 2) = 0$ and $P_F(\gamma_I = 1) = 0.55$ this gives $p_{\gamma=2} = 0.54$. The randomized procedure that gets $P_F = \alpha$ while maximizing P_D is to observe N_a and then with a probability $p_{\gamma=2} = 0.54$ return the result from the test $N_a \geq 2$ (which will always be false causing us to returning S_1). With probability $1 - p_{\gamma=2} = 0.45$ return the result from the test $N_a \geq 1$.

In the case where $N = 10$ we find that $\alpha = 0.25$ falls between the two points $P_F(\gamma_I = 8) = 0.09955965$ and $P_F(\gamma_I = 7) = 0.2660379$. We again need a randomized rule where we have to pick $p_{\gamma=8}$ such that

$$p_{\gamma=8} P_F(\gamma_I = 8) + (1 - p_{\gamma=8}) P_F(\gamma_I = 7) = 0.25 .$$

Solving this gives $p_{\gamma_I=8} = 0.096336$. The randomized decision to get $P_F = 0.25$ while maximizing P_D is of the same form as in the $N = 1$ case. That is to observe N_a and then with a probability $p_{\gamma=8} = 0.096336$ return the result from the test $N_a \geq 8$. With probability $1 - p_{\gamma=8} = 0.9036635$ return the result from the test $N_a \geq 7$.

Problem 2.2.8 (a Cauchy hypothesis test)

Part (1): For the given densities the LRT would state that if

$$\Lambda(R) = \frac{p(X|H_1)}{p(X|H_0)} = \frac{1 + (X - a_0)^2}{1 + (X - a_1)^2} = \frac{1 + X^2}{1 + (X - 1)^2} > \eta ,$$

we decide H_1 (and H_0 otherwise). We can write the above as a quadratic expression in X on the left-hand-side as

$$(1 - \eta)X^2 + 2\eta X + 1 - 2\eta > 0 .$$

Using the quadratic formula we can find the values of X where the left-hand-side of this expression *equals* zero. We find

$$X_\pm = \frac{-2\eta \pm \sqrt{4\eta^2 - 4(1 - \eta)(1 - 2\eta)}}{2(1 - \eta)} = \frac{-\eta \pm \sqrt{-(1 - 3\eta + \eta^2)}}{1 - \eta} . \tag{75}$$

In order for a real value of X above to exist we must have that the expression $1 - 3\eta + \eta^2 < 0$. If that expression is not true then $(1 - \eta)X^2 + 2\eta X + 1 - 2\eta$ is either always positive or always negative. This expression $1 - 3\eta + \eta^2$ is zero at the values

$$\frac{3 \pm \sqrt{9 - 4}}{2} = \frac{3 \pm \sqrt{5}}{2} = \{0.381966, 2.618034\} .$$

For various values of η we have

- If $\eta < 0.381966$ (say $\eta = 0$) then $1 - 3\eta + \eta^2 > 0$ and $(1-\eta)X^2 + 2\eta X + 1 - 2\eta$ is always positive indicating we always choose H_1. This gives the point $(P_F, P_D) = (1, 1)$ on the ROC curve.

- If $\eta > 2.618034$ (say $\eta = 3$) then $1 - 3\eta + \eta^2 > 0$ and $(1-\eta)X^2 + 2\eta X + 1 - 2\eta$ is always negative indicating we always choose H_0. This gives the point $(P_F, P_D) = (0, 0)$ on the ROC curve.

- If $0.381966 < \eta < 2.618034$ (say $\eta = 0.75$) then $1 - 3\eta + \eta^2 < 0$ and there are two points X, given by Equation 75, where $(1-\eta)X^2 + 2\eta X + 1 - 2\eta$ changes sign. Note that from Equation 75 $X_- < X_+$ if $\eta < 1$. For example, if $\eta = 0.75$ then the two points X_- and X_+ are
$$X_- = -6.316625 \quad \text{and} \quad X_+ = 0.3166248\,.$$
When $X < X_-$ one finds that the expression $(1-\eta)X^2 + 2\eta X + 1 - 2\eta$ is always positive so we choose H_1, when $X_- < X < X_+$ the expression is negative so we choose H_0, and when $X > X_+$ the expression is positive again so we choose H_0.

  If $\eta > 1$ say $\eta = 1.25$ then the two points X_- and X_+ are
$$X_- = 9.358899 \quad \text{and} \quad X_+ = 0.641101\,.$$
When $X < X_+$ one finds that the expression $(1-\eta)X^2 + 2\eta X + 1 - 2\eta$ is always negative so we choose H_0, when $X_+ < X < X_-$ the expression is positive so we choose H_1, and when $X > X_-$ the expression is positive again so we choose H_1.

Part (2): Using the above information we can express P_F and P_D as a function of η. We will increase η from 0 to $+\infty$ and plot the point (P_F, P_D) as a function of η.

For all points $0 < \eta < 0.381966$ we get the ROC point $(P_F, P_D) = (1, 1)$. After we increase η past $\eta > 2.618034$ we get the ROC point $(P_F, P_D) = (0, 0)$. For values of η between 0.381966 and 1.0 these two values we have

$$\begin{aligned}
P_F &= \Pr(\text{choose } H_1 | H_0 \text{ is true}) = \int_{-\infty}^{X_-} p(X|H_0)dX + \int_{X_+}^{\infty} p(X|H_0)dX \\
&= 1 - \int_{X_-}^{X_+} \frac{1}{\pi(1+X^2)} dX = 1 - \frac{1}{\pi}\tan^{-1}(X_+) + \frac{1}{\pi}\tan^{-1}(X_-) \\
P_D &= \Pr(\text{choose } H_1 | H_1 \text{ is true}) = 1 - \int_{X_-}^{X_+} \frac{1}{\pi(1+(X-1)^2)} dX \\
&= 1 - \frac{1}{\pi}\tan^{-1}(X_+ - 1) + \frac{1}{\pi}\tan^{-1}(X_- - 1)\,.
\end{aligned}$$

In the case where η between 1.0 and 2.618034 we would have

$$\begin{aligned}
P_F &= \int_{X_+}^{X_-} p(X|H_0)dX = \frac{1}{\pi}\tan^{-1}(X_-) - \frac{1}{\pi}\tan^{-1}(X_+) \\
P_D &= \int_{X_+}^{X_-} p(X|H_1)dX = \frac{1}{\pi}\tan^{-1}(X_- - 1) - \frac{1}{\pi}\tan^{-1}(X_+ - 1)\,.
\end{aligned}$$

All of these calculations are done in the R script `chap_2_prob_2.2.8.R`. When that script is run we get the result shown in Figure 4.

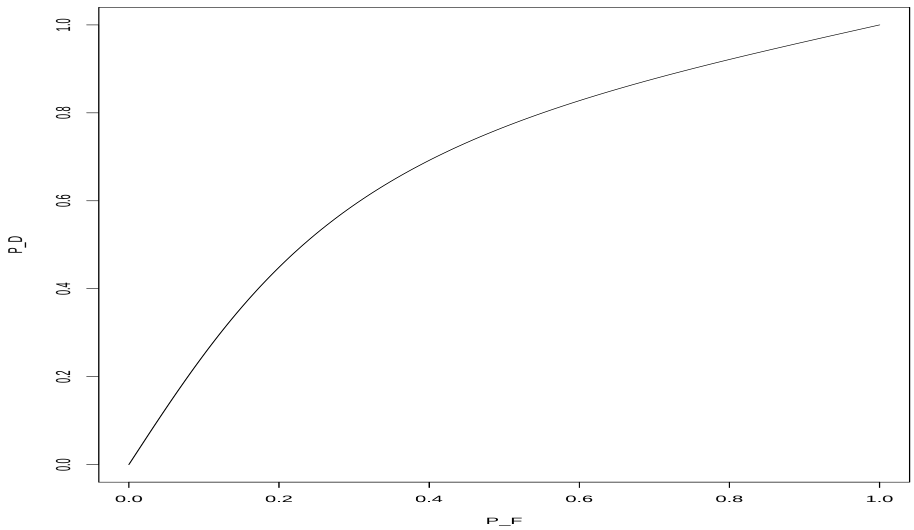

Figure 4: Plots of the ROC curve for Problem 2.2.8.

Problem 2.2.9 (coin flipping)

For this problem we assume that a coin is flipped N times and we are told the number of heads N_H. We wish to decide if the last flip of the coin landed heads (denoted as H_1) or tails (denoted H_0). Since this is a binary decision problem we need to evaluate $p(R|H_i)$ for some received measurement vector $\mathbf{R}$. The most complete set of information one could have for the N coin flips would be the sequence of complete outcomes. For example in the case where $N = 9$ we might observe the sequence

$$\mathbf{R} = \left[\begin{array}{ccccccccc} h & h & t & h & t & t & h & h & h \end{array} \right] .$$

Lets derive the LRT under the case of complete observability and then show that all that is needed to make a decision is N_H. In the same way as Problem 2.2.7 on page 28 we have

$$\begin{aligned} p(R|H_1) &= P_1^{N_H}(1-P_1)^{N-N_H} \\ p(R|H_0) &= P_0^{N_H}(1-P_0)^{N-N_H} . \end{aligned}$$

Using these the LRT is given by

$$\frac{p(R|H_1)}{p(R|H_0)} = \frac{P_1^{N_H}(1-P_1)^{N-N_H}}{P_0^{N_H}(1-P_0)^{N-N_H}} > \eta ,$$

decide H_1 (otherwise decide H_0).

Problem 2.2.10 (a Poisson counting process)

Part (1): For a Poisson counting process, under the hypothesis H_0 and H_1, the probabilities we have n events by the time T are given by

$$\Pr\{N(T) = n|H_0\} = \frac{e^{-k_0 T}(k_0 T)^n}{n!}$$
$$\Pr\{N(T) = n|H_1\} = \frac{e^{-k_1 T}(k_1 T)^n}{n!}.$$

Part (2): In the case where $P_0 = P_1 = \frac{1}{2}$, $C_{00} = C_{11} = 0$, and equal error costs $C_{10} = C_{01}$ we have $\eta = \frac{P_0(C_{10}-C_{00})}{P_1(C_{01}-C_{11})} = 1$ and our LRT says to decide H_1 when

$$\frac{\Pr\{N(T) = n|H_1\}}{\Pr\{N(T) = n|H_0\}} = e^{-k_1 T} e^{k_0 T} \left(\frac{k_1}{k_0}\right)^n > 1.$$

If we assume that $k_1 > k_0$ indicating that the second source has events that happen more frequently. Then the LRT can be written as

$$n > \frac{T(k_1 - k_0)}{\ln\left(\frac{k_1}{k_0}\right)} \equiv \gamma.$$

Since n can only take on integer values the only possible values for γ are $0, 1, 2, \cdots$. Thus our LRT reduces to $n \geq \gamma_I$ for $\gamma_I = 0, 1, 2, \cdots$. Note we consider the case where $n = \gamma_I$ to cause us to state that the event H_1 occurred.

Part (3): To determine the probability of error we use

$$\Pr(\epsilon) = P_0 P_F + P_1 P_M = P_0 P_F + P_1(1 - P_D) = \frac{1}{2} + \frac{1}{2}(P_F - P_D).$$

Here we would have

$$P_F = \Pr\{\text{Say } H_1|H_0\} = \sum_{i=\gamma_I}^{\infty} \frac{e^{-k_0 T}(k_0 T)^i}{i!}$$
$$P_D = \Pr\{\text{Say } H_1|H_1\} = \sum_{i=\gamma_I}^{\infty} \frac{e^{-k_1 T}(k_1 T)^i}{i!}.$$

The above could be plotted as a function of γ_I in the (P_F, P_D) plane to obtain the ROC curve for this problem.

Problem 2.2.11 (adding up Gaussian random variables)

For a LRT we need to compute $p(Y|H_0)$ and $p(Y|H_1)$ since Y is our observed variable. For the two hypothesis given we have

$$\begin{aligned} p(Y|H_1) &= p(Y|N \leq 1) \\ &= p(Y|N=0)P(N=0) + p(Y|N=1)P(N=1) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} e^{-\lambda} + \frac{1}{\sqrt{2\pi}\sqrt{2\sigma^2}} e^{-\frac{1}{2}\frac{y^2}{2\sigma^2}} \lambda e^{-\lambda} . \end{aligned}$$

and for $p(Y|H_0)$ we have

$$\begin{aligned} p(Y|H_0) = p(Y|N > 1) &= \sum_{k=2}^{\infty} p(Y|N=k)P(N=k) \\ &= \sum_{k=2}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{(k+1)\sigma^2}} e^{-\frac{1}{2}\frac{y^2}{(k+1)\sigma^2}} \left(\frac{\lambda^k e^{-\lambda}}{k!}\right) . \end{aligned}$$

With these densities the LRT is simply to decide H_1 if $\frac{p(Y|H_1)}{p(Y|H_0)} > \eta$ and H_0 otherwise.

Problem 2.2.13 (the expectation of $\Lambda(R)$)

To begin, recall that $\Lambda(R) \equiv \frac{p(R|H_1)}{p(R|H_0)}$.

Part (1): We have

$$\begin{aligned} E(\Lambda^{n+1}|H_0) &= \int \Lambda^{n+1}(R) p(R|H_0) dR = \int \left(\frac{p(R|H_1)}{p(R|H_0)}\right)^{n+1} p(R|H_0) dR \\ &= \int \left(\frac{p(R|H_1)}{p(R|H_0)}\right)^{n} p(R|H_1) dR = E(\Lambda^n|H_1) . \end{aligned}$$

Part (2): We have

$$E(\Lambda|H_0) = \int \frac{p(R|H_1)}{p(R|H_0)} p(R|H_0) dR = \int p(R|H_1) dR = 1 .$$

Part (3): Recall that $\mathrm{Var}(\Lambda^2|H_0) = E(\Lambda^2|H_0) - E(\Lambda|H_0)^2$ from #1 in this problem we have shown that $E(\Lambda^2|H_0) = E(\Lambda|H_1)$. From #2 of this problem $E(\Lambda|H_0) = 1$ so

$$E(\Lambda|H_0)^2 = 1^2 = 1 = E(\Lambda|H_0) ,$$

thus

$$\mathrm{Var}(\Lambda|H_0) = E(\Lambda|H_1) - E(\Lambda|H_0) .$$

We can work this problem in a different way. Consider the difference

$$
\begin{aligned}
E(\Lambda|H_1) - E(\Lambda|H_0) &= \int \frac{p(R|H_1)}{p(R|H_0)} p(R|H_1) dR - \int \frac{p(R|H_1)}{p(R|H_0)} p(R|H_0) dR \\
&= \int \frac{p(R|H_1)}{p(R|H_0)} \left(\frac{p(R|H_1)}{p(R|H_0)} - 1 \right) p(R|H_0) dR \\
&= \int \Lambda(R)(\Lambda(R) - 1) p(R|H_0) dR \\
&= E(\Lambda^2|H_0) - E(\Lambda|H_0) \quad \text{using \#1 from this problem} \\
&= E(\Lambda|H_1) - E(\Lambda|H_0)\,.
\end{aligned}
$$

Problem 2.2.14 (some mathematical results)

Part (2): From Part (3) of the previous problem we have that

$$
\begin{aligned}
\text{Var}(\Lambda|H_0) &= E(\Lambda|H_1) - E(\Lambda|H_0) = E(\Lambda|H_1) - 1 \\
&= \int \frac{p(R|H_1)}{p(R|H_0)} p(R|H_1) dR - 1\,.
\end{aligned}
$$

Thus to evaluate $\text{Var}(\Lambda|H_0) + 1$ we need to evaluate the integral

$$
I = \int \frac{p(R|H_1)}{p(R|H_0)} p(R|H_1) dR\,.
$$

Using the books Eq. 24 we find

$$
\begin{aligned}
I &= \int \exp\left\{ \frac{m}{\sigma^2} \sum_{i=1}^{N} R_i - \frac{Nm^2}{2\sigma^2} \right\} p(R|H_1) dR \\
&= e^{-\frac{Nm^2}{2\sigma^2}} \int_{R_1, R_2, \cdots, R_{N-1}, R_N} e^{\frac{m}{\sigma^2} \sum_{i=1}^{N} R_i} \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(R_i - m)^2}{2\sigma^2}} \right) dR_1 dR_2 \cdots dR_{N-1} dR_N \\
&= e^{-\frac{Nm^2}{2\sigma^2}} \prod_{i=1}^{N} \int_{R_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{m}{\sigma^2} R_i - \frac{(R_i - m)^2}{2\sigma^2} \right\} dR_i\,.
\end{aligned}
$$

The exponent can be simplified as

$$
\begin{aligned}
\frac{m}{\sigma^2} R_i - \frac{R_i^2}{2\sigma^2} + \frac{2R_i m}{2\sigma^2} - \frac{m^2}{2\sigma^2} &= -\frac{R_i^2}{2\sigma^2} + \frac{2R_i m}{\sigma^2} - \frac{m^2}{2\sigma^2} \\
&= -\frac{1}{2\sigma^2}(R_i - 2m)^2 + \frac{3m^2}{2\sigma^2}\,.
\end{aligned}
$$

With this the integral above becomes

$$
\begin{aligned}
I &= e^{-\frac{Nm^2}{2\sigma^2}} \prod_{i=1}^{N} e^{\frac{3m^2}{2\sigma^2}} \int_{R_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(R_i - 2m)^2 \right\} dR_i \\
&= e^{-\frac{Nm^2}{2\sigma^2}} e^{\frac{3Nm^2}{2\sigma^2}} = e^{\frac{Nm^2}{\sigma^2}}\,.
\end{aligned}
$$

Taking the logarithm of this last expression gives $\frac{Nm^2}{\sigma^2}$ which is the definition of d^2 showing the requested result.

Problem 2.2.15 (bounds on $\text{erfc}_*(X)$**)**

Part (1): Recall that from the definition of the function $\text{erfc}_*(X)$ we have that

$$\text{erfc}_*(X) = \int_X^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \,.$$

We will integrate this by parts by writing it as

$$\frac{1}{\sqrt{2\pi}} \int_X^\infty \frac{1}{x} (x e^{-x^2/2}) dx \,.$$

Then using the integration by parts lemma $\int v du = vu - \int u dv$ with

$$v = \frac{1}{x} \quad \text{and} \quad du = x e^{-x^2/2} dx \,,$$

where

$$dv = -\frac{1}{x^2} dx \quad \text{and} \quad u = -e^{-x^2/2} \,,$$

we have $\text{erfc}_*(X)$ given by

$$\begin{aligned} \text{erfc}_*(X) &= \frac{1}{\sqrt{2\pi}} \left(-\frac{1}{x} e^{-\frac{x^2}{2}} \bigg|_X^\infty - \int_X^\infty \frac{1}{x^2} e^{-\frac{x^2}{2}} dx \right) \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{1}{X} e^{-\frac{X^2}{2}} - \int_X^\infty \frac{1}{x^2} e^{-\frac{x^2}{2}} dx \right) \,. \end{aligned} \tag{76}$$

Since the second integral term $\frac{1}{\sqrt{2\pi}} \int_X^\infty \frac{1}{x^2} e^{-\frac{x^2}{2}} dx$ is positive (and nonzero) if we drop this term from the above sum we will have an expression that is larger in value that $\text{erfc}_*(X)$ or an upper bound. This means that

$$\text{erfc}_*(X) < \frac{1}{\sqrt{2\pi} X} e^{-\frac{X^2}{2}} \,. \tag{77}$$

We can continue by integrating this second term by parts as

$$\begin{aligned} \int_X^\infty \frac{1}{x^3} (x e^{-\frac{x^2}{2}}) dx &= -\frac{1}{x^3} e^{-\frac{x^2}{2}} \bigg|_X^\infty - \int_X^\infty \left(-\frac{3}{x^4} \right) \left(-e^{-\frac{x^2}{2}} \right) dx \\ &= \frac{1}{X^3} e^{-\frac{X^2}{2}} - \int_X^\infty \frac{3}{x^4} e^{-\frac{x^2}{2}} dx \,. \end{aligned}$$

Remembering the factor of $\frac{1}{\sqrt{2\pi}}$ and combining these results we have shown that

$$\text{erfc}_*(X) = \frac{1}{\sqrt{2\pi} X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi} X^3} e^{-\frac{X^2}{2}} + \frac{1}{\sqrt{2\pi}} \int_X^\infty \frac{3}{x^4} e^{-\frac{x^2}{2}} dx \,. \tag{78}$$

Since the last expression is positive (and nonzero) if we drop it the remaining terms will then sum to something smaller than $\text{erfc}_*(X)$. Thus we have just shown

$$\frac{1}{\sqrt{2\pi} X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi} X^3} e^{-\frac{X^2}{2}} < \text{erfc}_*(X) \,. \tag{79}$$

Combining expressions 77 and 79 we get the desired result.

Part (2): We will sketch the solution to this part of the problem but not work it out in full. Starting from Equation 76 we will derive a recursive expression for the second integral which will expand to give the series presented in the book. To begin, in that integral we let $t = \frac{x^2}{2}$, so $x = \sqrt{2t}$ and $dx = \frac{1}{\sqrt{2}t^{1/2}}dt$ and the integral becomes

$$\int_{\frac{X^2}{2}}^{\infty} (\sqrt{2}t^{1/2})^{-2} e^{-t} \frac{1}{\sqrt{2}t^{1/2}} dt = 2^{-3/2} \int_{\frac{X^2}{2}}^{\infty} t^{-3/2} e^{-t} dt \,.$$

Thus with this we have written $\text{erfc}_*(X)$ as

$$\text{erfc}_*(X) = \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} \int_{\frac{X^2}{2}}^{\infty} t^{-\left(\frac{1}{2}+1\right)} e^{-t} dt \,.$$

We will derive an asymptotic expansion to this second integral. Define $I_n(x)$ as

$$I_n(x) \equiv \int_x^{\infty} t^{-\left(\frac{1}{2}+n\right)} e^{-t} dt \,.$$

The integral we initially have is $I_1(\frac{X^2}{2})$. We can write $I_n(x)$ recursively using integration by part as

$$\begin{aligned} I_n(x) &= -t^{-\left(\frac{1}{2}+n\right)} e^{-t}\Big|_x^{\infty} + \int_x^{\infty} -\left(\frac{1}{2}+n\right) t^{-\left(\frac{1}{2}+n+1\right)} e^{-t} dt \\ &= x^{-\left(n+\frac{1}{2}\right)} e^{-x} - \left(\frac{1}{2}+n\right) \int_x^{\infty} t^{-\left(\frac{1}{2}+n+1\right)} e^{-t} dt \\ &= x^{-\left(n+\frac{1}{2}\right)} e^{-x} - \left(n+\frac{1}{2}\right) I_{n+1}(x) \,. \end{aligned} \quad (80)$$

Let $\tilde{x} \equiv \frac{X^2}{2}$. Then using this we have shown that we can write $\text{erfc}_*(X)$ as

$$\begin{aligned} \text{erfc}_*(X) &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} I_1(\tilde{x}) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} \left[\tilde{x}^{-\frac{3}{2}} e^{-\tilde{x}} - \frac{3}{2} I_2(\tilde{x})\right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} \left[\tilde{x}^{-\frac{3}{2}} e^{-\tilde{x}} - \frac{3}{2}\left(\tilde{x}^{-\frac{5}{2}} e^{-\tilde{x}} - \frac{5}{2} I_3(\tilde{x})\right)\right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} \left[\tilde{x}^{-\frac{3}{2}} e^{-\tilde{x}} - \frac{3}{2}\tilde{x}^{-\frac{5}{2}} e^{-\tilde{x}} + \frac{3\cdot 5}{2^2} I_3(\tilde{x})\right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} \left[\tilde{x}^{-\frac{3}{2}} e^{-\tilde{x}} - \frac{3}{2}\tilde{x}^{-\frac{5}{2}} e^{-\tilde{x}} + \frac{3\cdot 5}{2^2}\tilde{x}^{-\frac{7}{2}} e^{-\tilde{x}} - \frac{3\cdot 5\cdot 7}{2^3} I_4(\tilde{x})\right] \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{1}{\sqrt{2\pi}2^{3/2}} \left[\sum_{k=1}^{N} (-1)^{k-1} \frac{3\cdot 5\cdot 7\cdots(2k-1)}{2^{k-1}} \tilde{x}^{-\left(k+\frac{1}{2}\right)} e^{-\tilde{x}}\right] + (-1)^{N+1} \frac{3\cdot 5\cdot 7\cdots(2N+1)}{\sqrt{2\pi}2^{3/2}2^{N-1}} I_{N+1}(\tilde{x}) \,. \end{aligned}$$

When we put in what we know for $\tilde{x}$ we get

$$\begin{aligned} \text{erfc}_*(X) &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} - \frac{e^{-\frac{X^2}{2}}}{\sqrt{2\pi}X} \left[\sum_{k=1}^{N} (-1)^{k-1} \frac{3\cdot 5\cdot 7\cdots(2k-1)}{X^{2k}}\right] + (-1)^N \frac{3\cdot 5\cdot 7\cdots(2N+1)}{2^{N-1}} I_{N+1}(\tilde{x}) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{X} e^{-\frac{X^2}{2}} \left[1 + \sum_{k=1}^{N} (-1)^k \frac{1\cdot 3\cdot 5\cdot 7\cdots(2k-1)}{X^{2k}}\right] + (-1)^N \frac{3\cdot 5\cdot 7\cdots(2N+1)}{2^{N-1}} I_{N+1}\left(\frac{X^2}{2}\right) \,. \end{aligned} \quad (81)$$

Problem 2.2.16 (an upper bound on the complementary error function)

Part (1): From the definition of the function $\mathrm{erfc}_*(X)$ we have that

$$\mathrm{erfc}_*(X) = \int_X^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \,,$$

which we can simplify by the following change of variable. Let $v = x - X$ (then $dv = dx$) and the above becomes

$$\begin{aligned}
\mathrm{erfc}_*(X) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(v+X)^2/2} dv \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(v^2+2vX+X^2)/2} dv \\
&= \frac{e^{-\frac{X^2}{2}}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{v^2}{2}} e^{-vX} dv \\
&\leq \frac{e^{-\frac{X^2}{2}}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{v^2}{2}} dv \,,
\end{aligned}$$

since $e^{-vX} \leq 1$ for all $v \in [0, \infty)$ and $X > 0$. Now because of the identity

$$\int_0^\infty e^{-\frac{v^2}{2}} dv = \sqrt{\frac{\pi}{2}} \,,$$

we see that the above becomes

$$\mathrm{erfc}_*(X) \leq \frac{1}{2} e^{-\frac{X^2}{2}} \,,$$

as we were to show.

Part (2): We want to compare the expression derived above with the bound

$$\mathrm{erfc}_*(X) < \frac{1}{\sqrt{2\pi} X} e^{-\frac{X^2}{2}} \,. \tag{82}$$

Note that these two bounds only differ in the coefficient of the exponential of $-\frac{X^2}{2}$. These two coefficients will be equal at the point X^* when

$$\frac{1}{\sqrt{2\pi} X^*} = \frac{1}{2} \quad \Rightarrow \quad X^* = \sqrt{\frac{2}{\pi}} = 0.7978 \,.$$

Once we know this value where the coefficients are equal we know that when $X > 0.7978$ then Equation 82 would give a tighter bound since in that case $\frac{1}{\sqrt{2\pi} X} < \frac{1}{2}$, while if $X < 0.7978$ the bound

$$\mathrm{erfc}_*(X) < \frac{1}{2} e^{-\frac{X^2}{2}} \,,$$

is better. One can observe the truth in these statements by looking at the books Figure 2.10 which plots three approximations to $\mathrm{erfc}_*(X)$. In that figure one can visually observe that $\frac{1}{2} e^{-\frac{X^2}{2}}$ is a better approximation to $\mathrm{erfc}_* X$ than $\frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}$ is in the range $0 < X < 0.7978$. The opposite statement holds for $X > 0.7978$

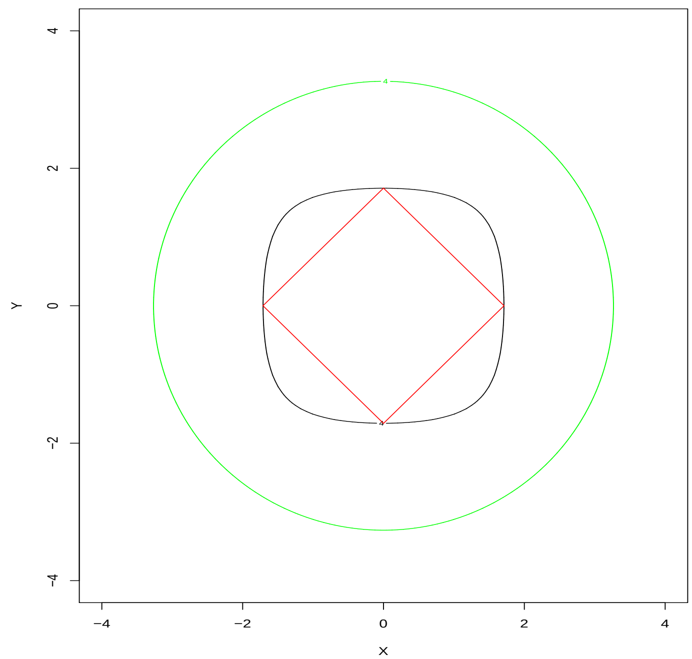

Figure 5: For $\gamma = 4$, the points that satisfy $\Omega(\gamma)$ outside of the black curve, points that satisfy $\Omega_{\text{subset}}(\gamma)$ outside of the green curve, and points that satisfy $\Omega_{\text{super}}(\gamma)$ outside of the red curve. When integrated over these regions we obtain exact, lower bound, and upper bound values for P_F and P_D.

Problem 2.2.17 (multidimensional LRTs)

Part (1): For the given densities we have that when

$$\begin{aligned}\Lambda(X_1, X_2) &= \frac{2\pi\sigma_0^2}{4\pi\sigma_1\sigma_0}\left[\exp\left\{-\frac{1}{2}X_1^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right\} + \exp\left\{-\frac{1}{2}X_2^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right\}\right] \\ &= \frac{\sigma_0}{2\sigma_1}\left[\exp\left\{\frac{1}{2}X_1^2\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2\sigma_1^2}\right)\right\} + \exp\left\{\frac{1}{2}X_2^2\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2\sigma_1^2}\right)\right\}\right] > \eta\,,\end{aligned}$$

we decide H_1. Solving for the function of X_1 and X_2 in terms of the parameter of this problem we have

$$\exp\left\{\frac{1}{2}X_1^2\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2\sigma_1^2}\right)\right\} + \exp\left\{\frac{1}{2}X_2^2\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2\sigma_1^2}\right)\right\} > \frac{2\eta\sigma_1}{\sigma_0} \equiv \gamma \qquad (83)$$

For what follows lets assume that $\sigma_1 > \sigma_0$. This integration region is like a nonlinear ellipse in that the points that satisfy this inequality are outside of a squashed ellipse. See Figure 5 where we draw the contour (in black) of Equation 83 when $\gamma = 4$. The points outside of this squashed ellipse are the ones that satisfy $\Omega(\gamma)$.

Part (2): The probabilities we are looking for will satisfy

$$P_F = \Pr\{\text{Say } H_1|H_0\} = \int_{\Omega(\gamma)} p(X_1, X_2|H_0) dX_1 dX_2$$
$$P_D = \Pr\{\text{Say } H_1|H_1\} = \int_{\Omega(\gamma)} p(X_1, X_2|H_1) dX_1 dX_2 \,.$$

Here $\Omega(\gamma)$ is the region of the (X_1, X_2) plane that satisfies Equation 83. Recalling the identity that $X < e^X$ when $X > 0$ the region of the (X_1, X_2) plane where

$$\frac{1}{2} X_1^2 \left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2 \sigma_1^2} \right) + \frac{1}{2} X_2^2 \left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2 \sigma_1^2} \right) > \gamma \,, \tag{84}$$

will be a *smaller* set than $\Omega(\gamma)$. The last statement is the fact that X_1 and X_2 need to be larger to make their "squared sum" bigger than γ. Thus there are point (X_1, X_2) closer to the origin where Equation 83 is true while Equation 84 is not. Thus if we integrate over this new region rather than the original $\Omega(\gamma)$ region we will have an lower bound on P_F and P_D. Thus

$$P_F \geq \int_{\Omega_{\text{subset}}(\gamma)} p(X_1, X_2|H_0) dX_1 dX_2$$
$$P_D \geq \int_{\Omega_{\text{subset}}(\gamma)} p(X_1, X_2|H_1) dX_1 dX_2 \,.$$

Here $\Omega_{\text{subset}}(\gamma)$ is the region of the (X_1, X_2) plane that satisfies Equation 84. This region can be written as the circle

$$X_1^2 + X_2^2 > \frac{2\gamma\sigma_0\sigma_1}{\sigma_1^2 - \sigma_0^2} \,.$$

See Figure 5 where we draw the contour (in green) of Equation 84 when $\gamma = 4$. The points outside of this circle are the ones that belong to $\Omega_{\text{subset}}(\gamma)$ and would be integrated over to obtaining the approximations to P_F and P_D for the value of $\gamma = 4$.

One might think that one could then integrate in polar coordinates to evaluate these integrals. This appears to be true for the lower bound approximation for P_F (where we integrate against $p(X_1, X_2|H_0)$ which has an analytic form with the polar expression $X_1^2 + X_2^2$) but the integral over $p(X_1, X_2|H_1)$ due to its functional form (even in polar) appears more difficult. If anyone knows how to integrate this analytically please contact me.

To get an upper bound on P_F and P_D we want to construct a region of the (X_1, X_2) plane that is a superset of the points in $\Omega(\gamma)$. We can do this by considering the "internal polytope" or the "box" one gets by taking $X_1 = 0$ and solving for the two points X_2 on Equation 83 (and the same thing for X_2) and connecting these points by straight lines. For example, when we take $\gamma = 4$ and solve for these four points we get

$$(1.711617, 0) \,, \quad (0, 1.711617) \,, \quad (-1.711617, 0) \,, \quad (0, -1.711617) \,.$$

One can see lines connecting these points drawn Figure 5. Let the points in the (X_1, X_2)

space outside of these lines be denoted $\Omega_{\text{super}}(\gamma)$. This then gives the bounds

$$P_F < \int_{\Omega_{\text{super}}(\gamma)} p(X_1, X_2|H_0) dX_1 dX_2$$
$$P_D < \int_{\Omega_{\text{super}}(\gamma)} p(X_1, X_2|H_1) dX_1 dX_2 .$$

The R script `chap_2_prob_2.2.17.R` performs plots needed to produce these figures.

Problem 2.2.18 (more multidimensional LRTs)

Part (1): From the given densities in this problem we see that H_1 represents "the idea" that one of the coordinates of the vector $\mathbf{X}$ will have a non zero mean of m. The probability that the coordinate that has a non zero mean is given by p_k. The LRT for this problem is

$$\begin{aligned}\frac{p(X|H_1)}{p(X|H_0)} &= \sum_{k=1}^{M} p_k \exp\left\{-\frac{(X_k - m)^2}{2\sigma^2}\right\} \exp\left\{\frac{X_k^2}{2\sigma^2}\right\} \\ &= \sum_{i=1}^{M} p_k \exp\left\{\frac{2X_k m - m^2}{2\sigma^2}\right\} = e^{-\frac{m^2}{\sigma^2}} \sum_{i=1}^{M} p_k e^{\frac{m}{\sigma^2} X_k} .\end{aligned}$$

If the above ratio is greater than some threshold η we decide H_1 otherwise we decide H_0.

Part (2): If $M = 2$ and $p_1 = p_2 = \frac{1}{2}$ the above LRT becomes

$$\frac{1}{2} e^{-\frac{m^2}{\sigma^2}} (e^{\frac{m}{\sigma^2} X_1} + e^{\frac{m}{\sigma^2} X_2}) > \eta .$$

Thus we decide H_1 when

$$e^{\frac{m}{\sigma^2} X_1} + e^{\frac{m}{\sigma^2} X_2} > 2\eta e^{\frac{m^2}{\sigma^2}} \equiv \gamma . \tag{85}$$

Note that $\gamma > 0$. For the rest of the problem we assume that $m > 0$. Based on this expression we will decide H_1 when the magnitude of (X_1, X_2) "is large" and the inequality in Equation 85 is satisfied. For example, when $\gamma = 4$ the region of the (X_1, X_2) plane classified as H_1 are the points to the North and East of the boundary line (in black) in Figure 6. The points classified as H_0 are the points to the South-West in Figure 6. **Part (3):** The exact expressions for P_F and P_D involve integrating over the region of (X_1, X_2) space defined by Equation 85 but with different integrands. For P_F the integrand is $p(X|H_0)$ and for P_D the integrand is $p(X|H_1)$.

We can find a *lower bound* for P_F and P_D by noting that for points *on* the decision boundary threshold we have

$$e^{\frac{m}{\sigma^2} X_1} = \gamma - e^{\frac{m}{\sigma^2} X_2}$$

Thus

$$e^{\frac{m}{\sigma^2} X_1} < \gamma \quad \text{or} \quad X_1 < \frac{\sigma^2}{m} \ln(\gamma) .$$

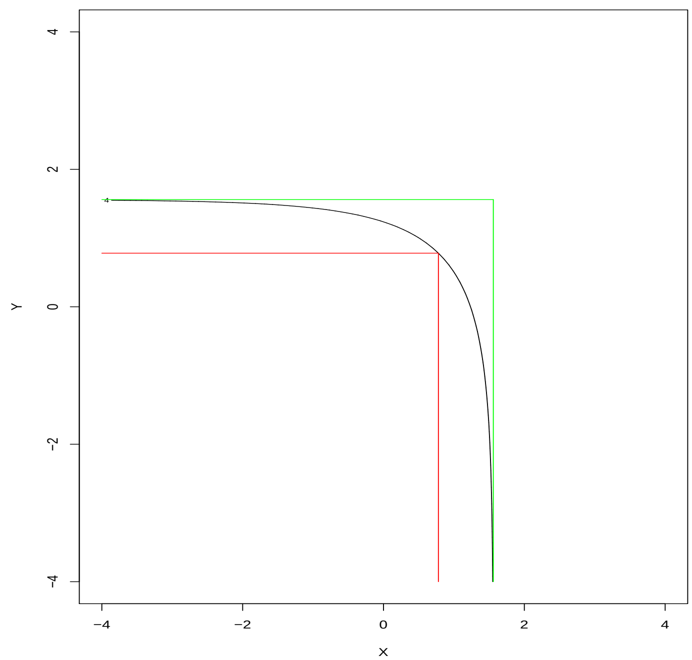

Figure 6: For $\gamma = 4$, the points that are classified as H_1 are to the North and East past the black boundary curve. Points classified as H_0 are in the South-West direction and are "below" the black boundary curve. A lower bound for P_F and P_D can be obtained by integrating to the North and East of the green curve. A upper bound for P_F and P_D can be obtained by integrating to the North and East of the red curve.

The same expression holds for X_2. Thus if we integrate instead over the region of $X_1 > \frac{\sigma^2}{m}\ln(\gamma)$ and $X_2 > \frac{\sigma^2}{m}\ln(\gamma)$ we will get a lower bound for P_F and P_D. To get an upper bound we find the point where the line $X_1 = X_2$ intersects the decision boundary. This is given at the location of

$$2e^{\frac{m}{\sigma^2}X_1} = \gamma \quad \text{or} \quad X_1 = \frac{\sigma^2}{m}\ln\left(\frac{\gamma}{2}\right).$$

Using this we can draw an integration region to compute an *upper bound* for P_F and P_D. We would integrate over the (X_1, X_2) points to the North-East of the red curve in Figure 6. The R script `chap_2_prob_2.2.18.R` performs plots needed to produce these figures.

Problem 2.2.19 (different means and covariances)

Part (1): For this problem, we have N samples from two different hypothesis each of which has a different mean m_k and variance σ_k^2. Given these densities the LRT test for this problem is given by

$$\begin{aligned}
\Lambda(R) = \frac{p(R|H_1)}{p(R|H_0)} &= \prod_{i=1}^{N} \frac{\sigma_0}{\sigma_1} \exp\left\{-\frac{(R_i - m_1)^2}{2\sigma_1^2} + \frac{(R_i - m_0)^2}{2\sigma_0^2}\right\} \\
&= \left(\frac{\sigma_0}{\sigma_1}\right)^N \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{N}(R_i - m_1)^2 + \frac{1}{2\sigma_0^2}\sum_{i=1}^{N}(R_i - m_0)^2\right\} \\
&= \left(\frac{\sigma_0}{\sigma_1}\right)^N \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{N}(R_i^2 - 2m_1R_i + m_1^2) + \frac{1}{2\sigma_0^2}\sum_{i=1}^{N}(R_i^2 - 2R_im_0 - m_0^2)\right\} \\
&= \left(\frac{\sigma_0}{\sigma_1}\right)^N \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{i=1}^{N}R_i^2 + \left(\frac{m_1}{\sigma_1^2} - \frac{m_0}{\sigma_0^2}\right)\sum_{i=1}^{N}R_i - \frac{Nm_1^2}{2\sigma_1^2} + \frac{Nm_0^2}{2\sigma_0^2}\right\} \\
&= \left(\frac{\sigma_0}{\sigma_1}\right)^N e^{\left\{-\frac{N}{2}\left(\frac{m_1^2}{\sigma_1^2} - \frac{m_0^2}{\sigma_0^2}\right)\right\}} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)l_\beta + \left(\frac{m_1}{\sigma_1^2} - \frac{m_0}{\sigma_0^2}\right)l_\alpha\right\}.
\end{aligned}$$

We decide H_1 if the above ratio is greater than our threshold η. The above can be written

$$-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)l_\beta + \left(\frac{m_1}{\sigma_1^2} - \frac{m_0}{\sigma_0^2}\right)l_\alpha > \ln\left\{\eta\left(\frac{\sigma_1}{\sigma_0}\right)^N e^{\left\{\frac{N}{2}\left(\frac{m_1^2}{\sigma_1^2} - \frac{m_0^2}{\sigma_0^2}\right)\right\}}\right\}. \tag{86}$$

The above is the expression for a *line* in the (l_α, l_β) space. We take the right-hand-side of the above expression be equal to γ (a parameter we can change to study different possible detection trade offs).

Part (2): If $m_0 = \frac{1}{2}m_1 > 0$ and $\sigma_0 = 2\sigma_1$ the above LRT, when written in terms of m_1 and σ_1, becomes if

$$\frac{7m_1}{8\sigma_1^2}l_\alpha - \frac{3}{8\sigma_1^2}l_\beta > \gamma,$$

then we decide H_1. Note that this decision region is a *line* in (l_α, l_β) space. An example of this decision boundary is drawn in Figure 7 for $m_1 = 1$, $\sigma_1 = \frac{1}{2}$, and $\gamma = 4$.

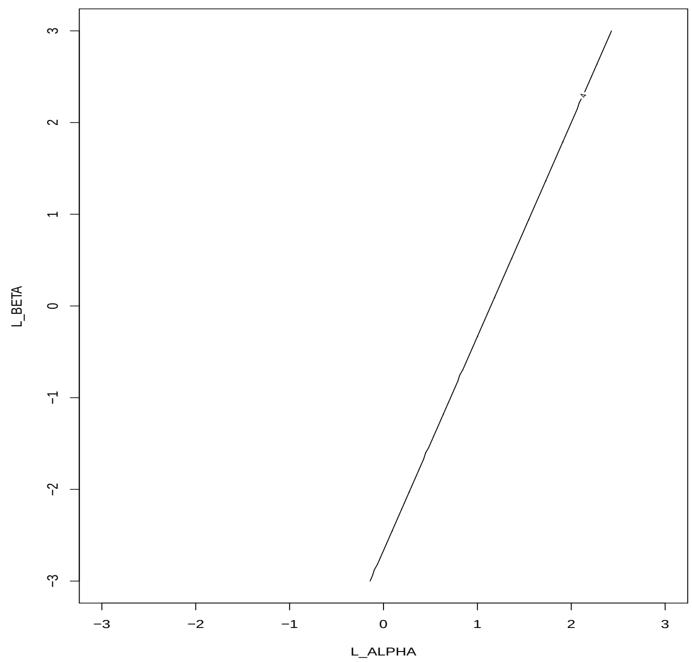

Figure 7: For $m_1 = 1$, $\sigma_1 = \frac{1}{2}$, and $\gamma = 4$, the points that are classified as H_1 are the ones in the South-East direction across the black decision boundary.

Problem 2.2.20 (specifications of different means and variances)

Part (1): When $m_0 = 0$ and $\sigma_0 = \sigma_1$ Equation 86 becomes

$$\frac{m_1}{\sigma_1^2} l_\alpha > \gamma \, .$$

If we assume that $m_1 > 0$ this is equivalent to $l_\alpha > \frac{\gamma\sigma_1^2}{m_1}$, which is a vertical line in the l_α, l_β plane. Points to the right of the constant $\frac{\gamma\sigma_1^2}{m_1}$ are classified as H_1 and points to the left of that point are classified as H_0. To compute the ROC we have

$$P_F = \int_{\frac{\gamma\sigma_1^2}{m_1}}^{\infty} p(L|H_0)dL$$
$$P_D = \int_{\frac{\gamma\sigma_1^2}{m_1}}^{\infty} p(L|H_1)dL \, .$$

Recall that L in this case is $l_\alpha \equiv \sum_{i=1}^{N} R_i$ we can derive the densities $p(L|H_0)$ and $p(L|H_1)$ from the densities for R_i in each case. Under H_0 each R_i is a Gaussian with mean 0 and variance σ_1^2. Thus $\sum_{i=1}^{N} R_i$ is a Gaussian with mean 0 and variance $N\sigma_1^2$. Under H_1 each R_i is a Gaussian with a mean of m_1 and a variance σ_1^2. Thus $\sum_{i=1}^{N} R_i$ is a Gaussian with mean

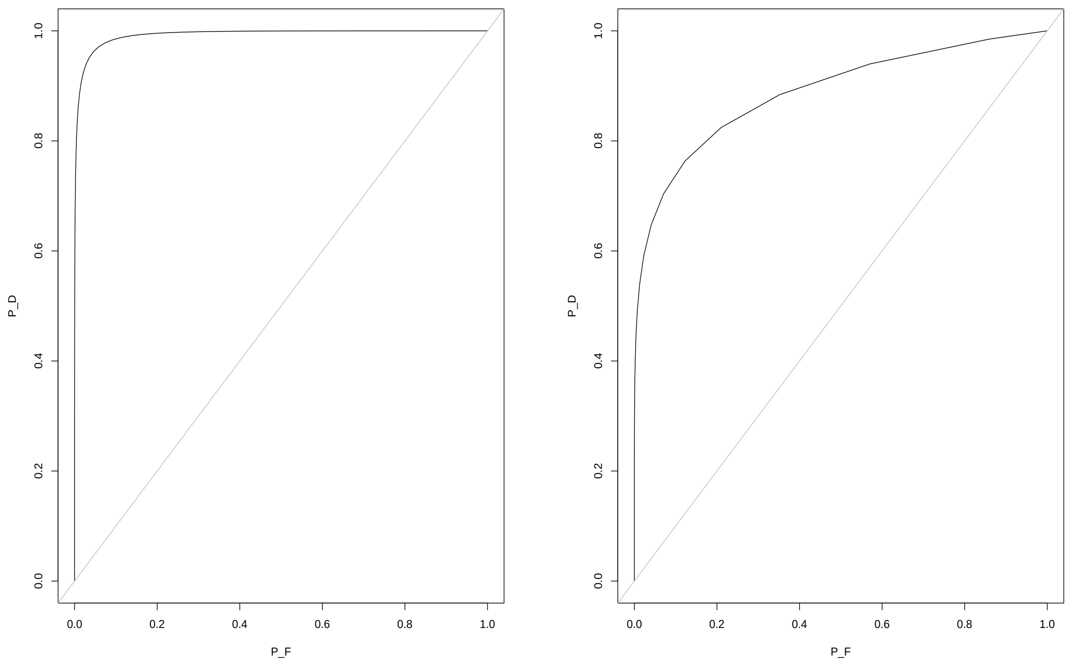

Figure 8: **Left:** The ROC curve for Problem 2.2.20. Part 1. **Right:** The ROC curve for Problem 2.2.20. Part 2.

Nm_1 and variance $N\sigma_1^2$. Thus we have

$$
\begin{aligned}
P_F &= \int_{\frac{\gamma\sigma_1^2}{m_1}}^{\infty} \frac{1}{\sqrt{2\pi N\sigma_1^2}} e^{-\frac{1}{2}\frac{L^2}{N\sigma_1^2}} dL \\
&= \int_{\frac{\gamma\sigma_1}{\sqrt{N}m_1}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}V^2} dV = \text{erfc}_*\left(\frac{\gamma\sigma_1}{\sqrt{N}m_1}\right) \\
P_D &= \int_{\frac{\gamma\sigma_1^2}{m_1}}^{\infty} \frac{1}{\sqrt{2\pi N\sigma_1^2}} e^{-\frac{1}{2}\frac{(L-Nm_1)^2}{N\sigma_1^2}} dL \\
&= \int_{\frac{\gamma\sigma_1}{\sqrt{N}m_1} - \frac{\sqrt{N}m_1}{\sigma_1}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}V^2} dV = \text{erfc}_*\left(\frac{\gamma\sigma_1}{\sqrt{N}m_1} - \frac{\sqrt{N}m_1}{\sigma_1}\right) .
\end{aligned}
$$

In the R script `chap_2_prob_2.2.20.R` we plot the given ROC curve for $m_1 = 1$, $\sigma_1 = \frac{1}{2}$, and $N = 3$. When this script is run it produces the plot given in Figure 8 (left).

Part (2): When $m_0 = m_1 = 0$, $\sigma_1^2 = \sigma_s^2 + \sigma_n^2$, and $\sigma_n^2 = \sigma_0^2$ Equation 86 becomes

$$
-\frac{1}{2}\left(\frac{1}{\sigma_s^2 + \sigma_n^2} - \frac{1}{\sigma_n^2}\right) l_\beta > \gamma ,
$$

or simplifying some we get

$$
l_\beta > \frac{2\sigma_n^2(\sigma_s^2 + \sigma_n^2)}{\sigma_s^2} \gamma . \tag{87}
$$

This is a horizontal line in the l_α, l_β plane. Lets define the constant on the right-hand-side as γ^*. Points in the l_α, l_β plane above this constant are classified as H_1 and points below this constant are classified as H_0. To compute the ROC we have

$$P_F = \int_{\gamma^*}^{\infty} p(L|H_0)dL$$

$$P_D = \int_{\gamma^*}^{\infty} p(L|H_1)dL\,.$$

Recall that L in this case is $l_\beta \equiv \sum_{i=1}^{N} R_i^2$. We can derive the densities $p(L|H_0)$ and $p(L|H_1)$ from the densities for R_i under H_0 and H_1.

Under H_0 each R_i is a Gaussian random variable with a mean of 0 and a variance of σ_n^2. Thus $\frac{1}{\sigma_n^2}\sum_{i=1}^{N} R_i^2$ is a chi-squared random variable with N degrees of freedom and we should write Equation 87 as

$$\frac{l_\beta}{\sigma_n^2} > \frac{2(\sigma_s^2+\sigma_n^2)}{\sigma_s^2}\gamma\,,$$

so

$$P_F = \int_{\frac{2(\sigma_s^2+\sigma_n^2)}{\sigma_s^2}\gamma}^{\infty} p(L|H_0)dL\,.$$

Here $p(L|H_0)$ is the chi-squared probability density with N degrees of freedom.

Under H_1 each R_i is a Gaussian with a mean of 0 and a variance $\sigma_s^2+\sigma_n^2$. Thus $\frac{1}{\sigma_s^2+\sigma_n^2}\sum_{i=1}^{N} R_i^2$ is another chi-squared random variable with N degrees of freedom and we should write Equation 87 as

$$\frac{l_\beta}{\sigma_s^2+\sigma_n^2} > \frac{2\sigma_n^2}{\sigma_s^2}\gamma\,,$$

so

$$P_D = \int_{\frac{2\sigma_n^2}{\sigma_s^2}\gamma}^{\infty} p(L|H_1)dL\,.$$

Here $p(L|H_1)$ is again the chi-squared probability density with N degrees of freedom.

In the `R` script `chap_2_prob_2.2.20.R` we plot the given ROC curve for $\sigma_n = 1$, $\sigma_s = 2$, and $N = 3$. When this script is run it produces the plot given in Figure 8 (right).

Problem 2.2.21 (error between to points)

Part (1): Since the book does not state exactly how we should compare the true impact point denoted via (x, y, z) and either of the two target located at the points (x_0, y_0, z_0) and (x_1, y_1, z_1). If we consider as our measurement the normalized *squared* distance between the impact point and a target point say (x_0, y_0, z_0) under hypothesis H_0 then the distribution of this sum is given by a chi-squared distribution with three degrees of freedom. Thus if our impact point is at (x, y, z) and we compute

$$D^2 = \frac{(x-x_i)^2}{\sigma^2} + \frac{(y-y_i)^2}{\sigma^2} + \frac{(z-z_i)^2}{\sigma^2}\,, \tag{88}$$

for $i = 0, 1$ to perform our hypothesis test we can use a chi-squared distribution for the distributions $p(D^2|H_0)$ and $p(D^2|H_1)$.

If we want to use the *distance* (rather than the square distance) as our measure of deviation between the impact point and one of the hypothesis points, it turns out that the Euclidean distance between two points is given by a *chi* distribution (not chi-squared). That is if X_i are normal random variables with means μ_i and variance σ_i^2 then

$$Y = \sqrt{\sum_{i=1}^{N} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2},$$

is given by a chi distribution. The probability density function for a chi distribution looks like

$$f(x; N) = \frac{2^{1-\frac{N}{2}} x^{N-1} e^{-\frac{x^2}{2}}}{\Gamma\left(\frac{N}{2}\right)}. \tag{89}$$

If we remove the mean μ_i from the expression for Y we get the noncentral chi distribution. That is if Z looks like

$$Z = \sqrt{\sum_{i=1}^{N} \left(\frac{X_i}{\sigma_i} \right)^2},$$

and we define $\lambda = \sqrt{\sum_{i=1}^{N} \left(\frac{\mu_i}{\sigma_i} \right)^2}$ then the probability density function for Z is given by

$$f(x; N, \lambda) = \frac{e^{-\frac{x^2+\lambda^2}{2}} x^N \lambda}{(\lambda x)^{N/2}} I_{\frac{N}{2}-1}(\lambda x), \tag{90}$$

where $I_\nu(z)$ is the modified Bessel function of the first kind. Since the chi distribution is a bit more complicated than the chi-squared we will consider the case where we use and use expression Equation 88 to measure distances. For reference, the chi-squared probability density looks like

$$f(x; N) = \frac{x^{\frac{N}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{N}{2}} \Gamma\left(\frac{N}{2}\right)}. \tag{91}$$

Using this the LRT for H_1 against H_0 look like

$$\begin{aligned}
\frac{p(R|H_1)}{p(R|H_0)} &= \frac{\left(\frac{(x-x_1)^2}{\sigma^2} + \frac{(y-y_1)^2}{\sigma^2} + \frac{(z-z_1)^2}{\sigma^2} \right)^{\frac{N}{2}-1}}{\left(\frac{(x-x_0)^2}{\sigma^2} + \frac{(y-y_0)^2}{\sigma^2} + \frac{(z-z_0)^2}{\sigma^2} \right)^{\frac{N}{2}-1}} \\
&\quad \times \exp\left\{ -\frac{1}{2} \left(\frac{(x-x_1)^2}{\sigma^2} + \frac{(y-y_1)^2}{\sigma^2} + \frac{(z-z_1)^2}{\sigma^2} \right) + \frac{1}{2} \left(\frac{(x-x_0)^2}{\sigma^2} + \frac{(y-y_0)^2}{\sigma^2} + \frac{(z-z_0)^2}{\sigma^2} \right) \right\} \\
&= \left(\frac{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2}{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} \right)^{\frac{N}{2}-1} \\
&\quad \times \exp\left\{ -\frac{1}{2\sigma^2} \left((x-x_1)^2 - (x-x_0)^2 + (y-y_1)^2 - (y-y_0)^2 + (z-z_1)^2 - (z-z_0)^2 \right) \right\}.
\end{aligned}$$

We would decide H_1 when this ratio is greater than a threshold η.

Part (2): The time variable is another independent measurement and the density function for the combined time and space measurement would simply be the product of the spatial density functions for H_1 and H_0 discussed above and the Gaussian for time.

Problem 2.3.1 (the dimension of the M hypothesis Bayes test)

Part (1): The general M hypothesis test is solved by computing the minimum of M expressions as demonstrated in Equation 16. As there are M expressions to find the minimum of these M expressions requires $M-1$ comparisons i.e. we have a decision space that is $M-1$.

Part (2): In the next problem we show that the decision can make based on β_i, which can be computed in terms $\Lambda_k(R)$ when we divide by $p(R|H_0)$.

Problem 2.3.2 (equivalent form for the Bayes test)

Part (1): When we use $P_j p(R|H_j) = p(H_j|R)p(R)$ we can write Equation 11 as

$$\mathcal{R} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} \int_{Z_i} p(H_j|R)p(R)dR = \sum_{i=0}^{M-1} \int_{Z_i} p(R) \sum_{j=0}^{M-1} C_{ij} p(H_j|R)dR\,.$$

Given a sample R the risk is given by evaluating the above over the various Z_i's. We can make this risk as small as possible by picking Z_i such that it only integrates over points R where the integrand above is as small as possible. That means pick R to be from class H_i if

$$p(R) \sum_{j=0}^{M-1} C_{ij} p(H_j|R) < p(R) \sum_{j=0}^{M-1} C_{i'j} p(H_j|R) \quad \text{for all} \quad i' \neq i\,.$$

We can drop $p(R)$ from both sides to get the optimal decision to pick the smallest value of (over i)

$$\sum_{j=0}^{M-1} C_{ij} p(H_j|R)\,.$$

This is the definition of β_i and is what we minimize to find the optimal decision rule.

Part (2): When the costs are as given we see that

$$\beta_i = \sum_{j=0;j\neq i}^{M-1} C p(H_j|R) = C(1 - p(H_i|R))\,.$$

Thus minimizing β_i is the same as maximizing $p(H_i|R)$ as a function of i.

Problem 2.3.3 (Gaussian decision regions)

Part (1): The minimum probability of error is the same as picking the class/hypothesis corresponding to the largest a-posterior probability. Thus we need to compute

$$p(H_k|R) \quad \text{for} \quad k = 1, 2, 3, 4, 5 \,.$$

Since each hypothesis is equally likely the above is equivalent to picking the class with the maximum likelihood or the expression $p(R|H_k)$ which is the largest. The decision boundaries are then the points where two likelihood functions $p(R|H_k)$ and $p(R|H_{k'})$ meet. For example, the decision boundary between H_1 and H_2 is given when $p(R|H_1) = p(R|H_2)$ or when we simplify that we get

$$|R + 2m| = |R + m| \,.$$

By geometry we must have $R+m < 0$ and $R+2m > 0$ so we need to solve $R+2m = -(R+m)$ or $R = -\frac{3}{2}m$. In general, the decision boundary will be the midpoint between each Gaussians mean and are located at $-\frac{3m}{2}$, $-\frac{m}{2}$, $\frac{m}{2}$, and $\frac{3m}{2}$.

Part (2): We can compute the probability of error in the following way

$$\begin{aligned}
\Pr(\varepsilon) &= 1 - P(\text{correct decision}) \\
&= 1 - \int_{-\infty}^{-\frac{3m}{2}} p(R|H_1)dR - \int_{-\frac{3m}{2}}^{-\frac{m}{2}} p(R|H_2)dR \\
&- \int_{-\frac{m}{2}}^{+\frac{m}{2}} p(R|H_3)dR - \int_{\frac{m}{2}}^{+\frac{3m}{2}} p(R|H_4)dR - \int_{\frac{3m}{2}}^{\infty} p(R|H_5)dR \,.
\end{aligned}$$

These integrals could be "evaluated" by converting to expressions involving the error function.

Problem 2.3.4 (Gaussians with different variances)

Part (1): As in Problem 2.3.4 the requested criterion is equivalent to the maximum likelihood classification. That is given R we pick the hypothesis H_i such that $p(R|H_i)$ is largest.

Part (2): If $\sigma_\beta^2 = 2\sigma_\alpha^2$ and $\sigma_\alpha = m$ our conditional densities look like

$$\begin{aligned}
p(R|H_1) &= \frac{1}{\sqrt{2\pi}m} \exp\left\{-\frac{R^2}{2m^2}\right\} \\
p(R|H_2) &= \frac{1}{\sqrt{2\pi}m} \exp\left\{-\frac{(R-m)^2}{2m^2}\right\} \\
p(R|H_3) &= \frac{1}{\sqrt{2\pi}(\sqrt{2}m)} \exp\left\{-\frac{R^2}{2(2m^2)}\right\} \,.
\end{aligned}$$

If we plot these three densities on the R-axis for $m = 2.5$ we get the plot in Figure 9. We now need to calculate the decision boundaries or the points where we would change the

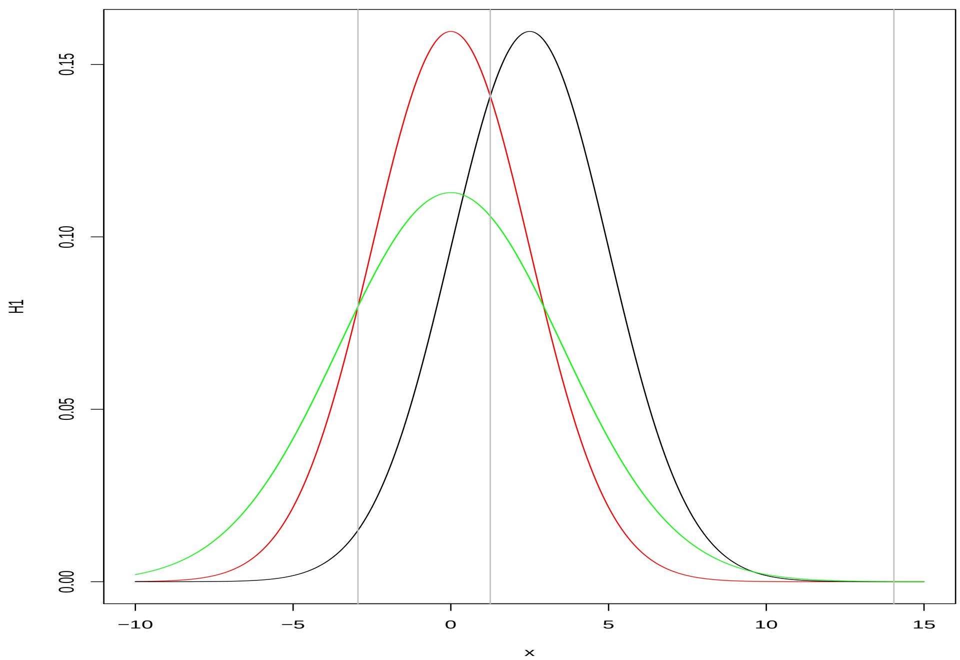

Figure 9: The three densities for Problem 2.3.4 for $m = 2.5$. Here $p(R|H_1)$ is plotted in red, $p(R|H_2)$ is plotted in black, and $p(R|H_3)$ is plotted in green. The hypothesis, H_i, that is chosen for any value of R corresponds to the index of the largest likelihood $p(R|H_i)$ at the given point R. The decision regions are drawn in the figure as light gray lines.

decision made. From the plot it looks like the classification is to pick "green" or H_3 then "red" or H_1, then "black" or H_2, and finally "green" or H_3 again. Each of these points where the decision changes is given by solving equations of the form $p(R|H_i) = p(R|H_j)$. We will compute these decision boundaries here. For the first decision boundary we need to find R such that $p(R|H_3) = p(R|H_1)$ or

$$\frac{1}{\sqrt{2}} \exp\left\{-\frac{R^2}{2(2m^2)}\right\} = \exp\left\{-\frac{R^2}{2m^2}\right\} .$$

When we take the logarithm and solve for R we find $R = \pm m\sqrt{2\ln(2)} = \pm 2.943525$ when we consider with $m = 2.5$. Lets denote this numeric value as m_{13}. For the second decision boundary we need to find R such that $p(R|H_1) = p(R|H_2)$ or

$$-\frac{R^2}{2m^2} = -\frac{(R-m)^2}{2m^2} \quad \text{or} \quad R = \frac{m}{2} .$$

We will denote this boundary as m_{12}. Finally, for the third decision boundary we need to find R such that $p(R|H_2) = p(R|H_3)$ or

$$\exp\left\{-\frac{(R-m)^2}{2m^2}\right\} = \frac{1}{\sqrt{2}} \exp\left\{-\frac{R^2}{2(2m^2)}\right\} .$$

This can be written as

$$R^2 - 4mR + 2m^2(1 - \ln(2)) = 0 ,$$

or solving for R we get

$$R = 2m \pm \sqrt{2}m\sqrt{1 + 8\ln(2)} = \{-4.045, 14.045\} ,$$

when $m = 2.5$. We would take the larger value (and denote it m_{23}) since we are looking for a positive decision boundary.

Part (3): To evaluate this probability we will compute

$$\Pr(\varepsilon) = 1 - \Pr(\text{correct}) ,$$

with Pr(correct) given by

$$3\Pr(\text{correct}) = \int_{-\infty}^{m_{13}} p(R|H_3)dR + \int_{m_{13}}^{m_{12}} p(R|H_1)dR + \int_{m_{12}}^{m_{23}} p(R|H_2)dR + \int_{m_{23}}^{\infty} p(R|H_3)dR .$$

The factor of three in the above expression is to account for the fact that all three hypothesis are equally likely i.e. $P(H_i) = \frac{1}{3}$ for $i = 1, 2, 3$.

Some of this algebra is done in the `R` code `chap_2_prob_2.3.4.R`.

Problem 2.3.5 (hypothesis testing with 2 dimensional densities)

Part (1): Since this is a three class problem we can directly use the results where the decision region is written in terms of likelihood ratios $\Lambda_1(R)$ and $\Lambda_2(R)$ as found in Equations 17, 18, and 19. To do that we need to compute these likelihood ratios. We find

$$
\begin{aligned}
\Lambda_1(R) &\equiv \frac{p(R|H_2)}{p(R|H_1)} \\
&= \frac{\sigma_{11}\sigma_{21}}{\sigma_{12}\sigma_{22}} \exp\left\{-\frac{1}{2}\left(\frac{R_1^2}{\sigma_{12}^2} + \frac{R_2^2}{\sigma_{22}^2} - \frac{R_1^2}{\sigma_{11}^2} - \frac{R_2^2}{\sigma_{21}^2}\right)\right\} \\
&= \frac{\sigma_n^2}{\sigma_n\sqrt{\sigma_s^2+\sigma_n^2}} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_s^2+\sigma_n^2} - \frac{1}{\sigma_n^2}\right) l_1 - \frac{1}{2}\left(\frac{1}{\sigma_n^2} - \frac{1}{\sigma_n^2}\right) l_2\right\} \\
&= \frac{\sigma_n}{\sqrt{\sigma_s^2+\sigma_n^2}} \exp\left\{\frac{\sigma_s^2}{2\sigma_n^2(\sigma_s^2+\sigma_n^2)} l_1\right\}.
\end{aligned}
$$

For $\Lambda_2(R)$ we have

$$
\begin{aligned}
\Lambda_2(R) &\equiv \frac{p(R|H_3)}{p(R|H_1)} \\
&= \frac{\sigma_{11}\sigma_{21}}{\sigma_{13}\sigma_{23}} \exp\left\{-\frac{1}{2}\left(\frac{R_1^2}{\sigma_{13}^2} + \frac{R_2^2}{\sigma_{23}^2} - \frac{R_1^2}{\sigma_{11}^2} - \frac{R_2^2}{\sigma_{21}^2}\right)\right\} \\
&= \frac{\sigma_n^2}{\sigma_n\sqrt{\sigma_s^2+\sigma_n^2}} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_n^2} - \frac{1}{\sigma_n^2}\right) l_1 - \frac{1}{2}\left(\frac{1}{\sigma_s^2+\sigma_n^2} - \frac{1}{\sigma_n^2}\right) l_2\right\} \\
&= \frac{\sigma_n}{\sqrt{\sigma_s^2+\sigma_n^2}} \exp\left\{\frac{\sigma_s^2}{2\sigma_n^2(\sigma_s^2+\sigma_n^2)} l_2\right\}.
\end{aligned}
$$

The class priors are specified as $P_0 = 1 - 2p$ and $P_1 = P_2 = p$ so using Equations 17, 18, and 19 we have that the decision region is based on

$$
\begin{aligned}
&\text{if} \quad p(1-0)\Lambda_1(R) > (1-2p)(1-0) + p(\alpha-1)\Lambda_2(R) \quad \text{then } H_1 \text{ or } H_2 \text{ else } H_0 \text{ or } H_2 \\
&\text{if} \quad p(1-0)\Lambda_2(R) > (1-2p)(1-0) + p(\alpha-1)\Lambda_1(R) \quad \text{then } H_2 \text{ or } H_1 \text{ else } H_0 \text{ or } H_1 \\
&\text{if} \quad p(\alpha-0)\Lambda_2(R) > (1-2p)(1-1) + p(\alpha-1)\Lambda_1(R) \quad \text{then } H_2 \text{ or } H_0 \text{ else } H_1 \text{ or } H_0 .
\end{aligned}
$$

Simplifying these some we get

$$
\begin{aligned}
&\text{if} \quad \Lambda_1(R) > \frac{1}{p} - 2 + (\alpha-1)\Lambda_2(R) \quad \text{then } H_1 \text{ or } H_2 \text{ else } H_0 \text{ or } H_2 \\
&\text{if} \quad \Lambda_2(R) > \frac{1}{p} - 2 + (\alpha-1)\Lambda_1(R) \quad \text{then } H_2 \text{ or } H_1 \text{ else } H_0 \text{ or } H_1 \\
&\text{if} \quad \Lambda_2(R) > \Lambda_1(R) \quad \text{then } H_2 \text{ or } H_0 \text{ else } H_1 \text{ or } H_0 .
\end{aligned}
$$

To find the decision regions in the $l_1 - l_2$ plane we can first make the inequalities above equal to equalities to find the decision boundaries. One decision boundary is easy, when we put in what we know for for $\Lambda_1(R)$ and $\Lambda_2(R)$ the last equation becomes $l_2 = l_1$, which is a diagonal line in the $l_1 - l_2$ space. To determine the full decision boundaries lets take values for p and α say $p = 0.25$ and $\alpha = 0.75$ and plot the first two decision boundaries above. This is done in Figure 10. Note that the first and second equation are symmetric (give the other equation) if we switch l_1 and l_2. Thus the decision boundary in the $l_1 - l_2$ space expressed by these two expressions is the *same*. Based on the inequalities in the above expressions we would get the classification regions given in Figure 10. Some of this algebra and our plots is performed in the `R` code `chap_2_prob_2.3.5.R`.

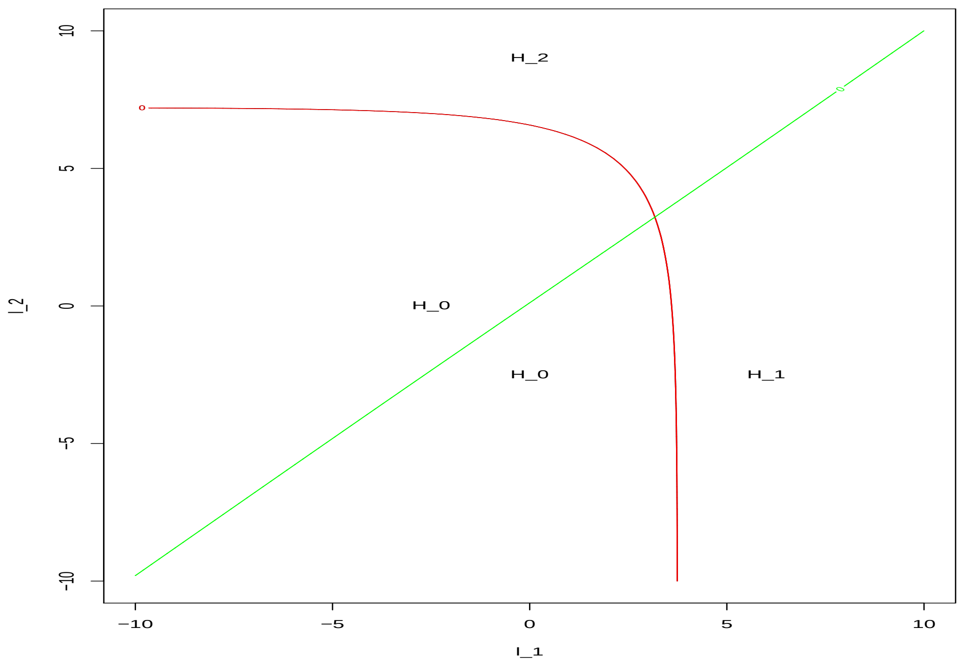

Figure 10: The decision boundaries for Problem 2.3.5 with the final classification label denoted in each region.

Problem 2.3.6 (a discrete M class decision problem)

Part (1): We are given

$$\Pr(r = n|H_m) = \frac{k_m^n}{n!} e^{-k_m} = \frac{(km)^n e^{-km}}{n!} \,,$$

for $m = 1, 2, 3, \ldots, M$ and k a fixed positive constant. Thus each hypothesis involves the expectation of progressively more samples being observed. For example, the *mean* number of events for the hypothesis $H_1, H_2, H_3, H_4, \ldots$ are $k, 2k, 3k, 4k, \ldots$. Since each hypothesis is equally likely and the coefficients of the errors are the same the optimal decision criterion corresponds to picking the hypothesis with the maximum likelihood i.e. we pick H_m if

$$\Pr(r = n|H_m) = \frac{(km)^n e^{-km}}{n!} \geq \Pr(r = n|H_{m'}) = \frac{(km')^n e^{-km'}}{n!} \,, \tag{92}$$

for all $m' \neq m$. For the value of $k = 3$ we plot $\Pr(r = n|H_m)$ as a function of n for several values of m in Figure 11. In that plot you can see that for progressively larger values of r we would select larger hypothesis values H_m. In the plot presented, as r, increased we would decide on the hypothesis $H_1, H_2, H_3, \cdots$. Our decision as to when the selected hypothesis changes is when two sequential likelihood functions have equal values. If we take the inequality in Equation 92 as an equality, the boundaries of the decision region between two hypothesis m and m' are where (canceling the factor $n!$)

$$(km)^n e^{-km} = (km')^n e^{-km'} \,.$$

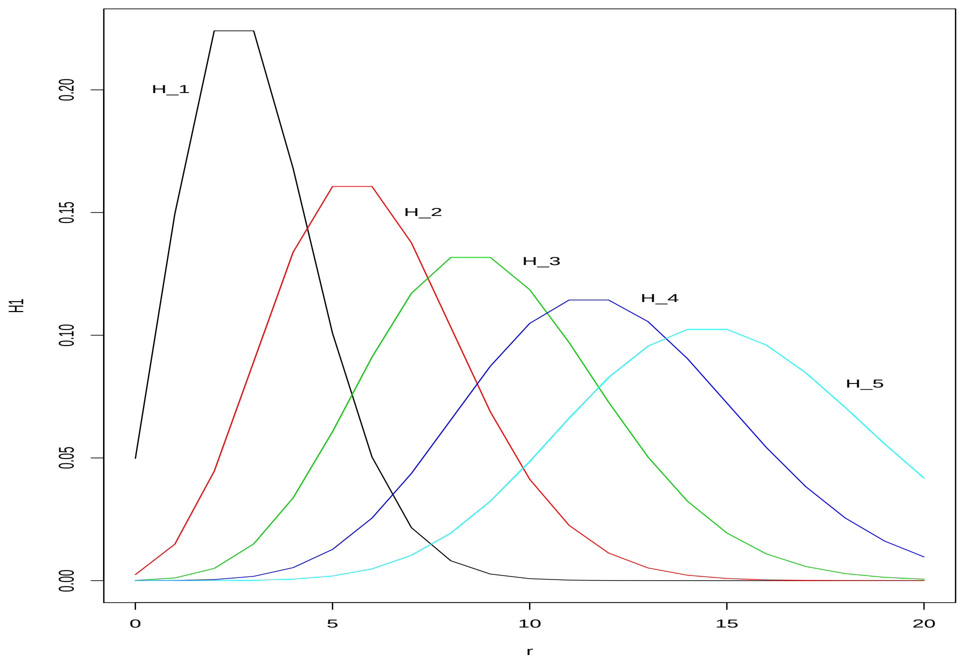

Figure 11: Plots of the likelihoods $\Pr(r = n|H_m)$ as a function of r in Problem 2.3.6 for various hypothesis or values of m. The optimal decision (given a measurement r) is to take the hypothesis that has the largest likelihood.

Solving for n (the decision boundary between $H_{m'}$ and H_m) in the above we get

$$n = \frac{k(m - m')}{\ln\left(\frac{m}{m'}\right)} .$$

Since the change in hypothesis happens between sequential $H_{m'}$ and H_m we have that $m = m' + 1$ and the above simplifies to

$$n_{m' \to m} = \frac{k}{\ln\left(1 + \frac{1}{m'}\right)} .$$

We can evaluate this expression for $m' = 1$ to find the locations where we switch between H_1 and H_2, for $m' = 2$ to find the location where we switch between H_2 and H_3, etc. For $m' = 1, 2, 3, 4, 5$ (here $M = 5$) we find

```
[1]  4.328085  7.398910 10.428178 13.444260 16.454445
```

These numbers match very well the cross over points in Figure 11. Since the number of counts received must be a natural number we would round the numbers above to the following

$$
\begin{aligned}
H_1 &: 1 \le r \le 4 \\
H_2 &: 5 \le r \le 7 \\
H_3 &: 8 \le r \le 10 \\
&\vdots \\
H_m &: \left\lfloor \frac{k}{\ln\left(1+\frac{1}{m-1}\right)} \right\rfloor + 1 \le r \le \left\lfloor \frac{k}{\ln\left(1+\frac{1}{m}\right)} \right\rfloor \\
&\vdots \\
H_M &: \left\lceil \frac{k}{\ln\left(1+\frac{1}{M-1}\right)} \right\rceil \le r \le \infty\,.
\end{aligned}
\tag{93}
$$

Part (2): We would calculate $\Pr(\epsilon) = 1 - \Pr(\text{correct})$ where $\Pr(\text{correct})$ is calculated by summing over regions in r where we would make the correct classification. This is given by

$$
\begin{aligned}
\Pr(\text{correct}) &= \sum_{m=1}^{M} \sum_{n \in Z_m} \Pr(r = n | H_m) \\
&= \sum_{n=1}^{4} \Pr(r = n|H_1) + \sum_{n=5}^{7} \Pr(r = n|H_2) + \sum_{n=8}^{10} \Pr(r = n|H_3) + \cdots + \sum_{n=n_l}^{n_u} \Pr(r = n|H_m)\,,
\end{aligned}
$$

where the lower and upper summation endpoints n_l and n_u are based on the decision boundary computed as given in Equation 93.

Problem 2.3.7 (M-hypothesis classification different mean vectors)

Part (1): For a three class problem we need to consider Equations 17, 18, and 19 to decide the decision boundaries. Thus we should compute

$$
\begin{aligned}
\Lambda_1(R) = \frac{p(R|H_1)}{p(R|H_0)} &= \frac{\exp\left\{-\frac{1}{2}\frac{(r-m_1)^T(r-m_1)}{\sigma^2}\right\}}{\exp\left\{-\frac{1}{2}\frac{(r-m_0)^T(r-m_0)}{\sigma^2}\right\}} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left[(r-m_1)^T(r-m_1) - (r-m_0)^T(r-m_0)\right]\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left[2r^T(m_0 - m_1) + m_1^T m_1 - m_0^T m_0\right]\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(m_1^T m_1 - m_0^T m_0\right)\right\} \exp\left\{\frac{1}{\sigma^2} r^T (m_1 - m_0)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(m_1^T m_1 - m_0^T m_0\right)\right\} e^{l_1}\,,
\end{aligned}
$$

with l_1 defined as

$$l_1 = \sum_{i=1}^{3} c_i r_i = \frac{1}{\sigma^2} \sum_{i=1}^{3} (m_{1i} - m_{0i}) r_i \quad \text{so} \quad c_i = \frac{1}{\sigma^2}(m_{1i} - m_{0i}) \,.$$

In the same way we have

$$\begin{aligned}
\Lambda_2(R) &= \frac{p(R|H_2)}{p(R|H_0)} \\
&= \exp\left\{ -\frac{1}{2\sigma^2}\left(m_2^T m_2 - m_0^T m_0\right) \right\} \exp\left\{ \frac{1}{\sigma^2} r^T (m_2 - m_0) \right\} \\
&= \exp\left\{ -\frac{1}{2\sigma^2}\left(m_2^T m_2 - m_0^T m_0\right) \right\} e^{l_2} \,,
\end{aligned}$$

with l_2 defined as

$$l_2 = \sum_{i=1}^{3} d_i r_i = \frac{1}{\sigma^2} \sum_{i=1}^{3} (m_{2i} - m_{0i}) r_i \quad \text{so} \quad d_i = \frac{1}{\sigma^2}(m_{2i} - m_{0i}) \,.$$

Part (2): For the given cost assignment given here when we use Equations 17, 18, and 19 (and cancel out the common cost) we have

$$\begin{aligned}
&P_1 \Lambda_1(R) > P_0 - P_2 \Lambda_2(R) \quad \text{then } H_1 \text{ or } H_2 \text{ else } H_0 \text{ or } H_2 \\
&P_2 \Lambda_2(R) > P_0 \quad \text{then } H_2 \text{ or } H_1 \text{ else } H_0 \text{ or } H_1 \\
&P_2 \Lambda_2(R) > P_0 + P_1 \Lambda_1(R) \quad \text{then } H_2 \text{ or } H_0 \text{ else } H_1 \text{ or } H_0 \,.
\end{aligned}$$

To make the remaining problem simpler, we will specify values for $P_0 = P_1 = P_2 = \frac{1}{3}$ and values for m_0, m_1, m_2, and σ so that every expression in the decision boundary can be evaluated numerically. This is done in the `R` script `chap_2_prob_2.3.6.R` and the result plotted in Figure 12.

Problem 2.4.1 (estimation in $r = ab + n$)

Part (1): The maximum a posterior estimate of a is derived from Bayes' rule

$$p(A|R) = \frac{p(R|A)p(A)}{p(R)} \,,$$

from which we see that we need to compute $p(R|A)$ and $p(A)$. Now when the variable a is *given* the expression we are observing or $r = ab + n$ is the sum of two independent random variables ab and n. These two terms are independent normal random variables with zero means and variances $a^2\sigma_b^2$ and σ_n^2 respectively. The variance of r is then the sum of the variance of ab and n. With these arguments we have that the densities we need in the above given by

$$\begin{aligned}
p(R|A) &= \frac{1}{\sqrt{2\pi}(A^2\sigma_b^2 + \sigma_n^2)^{1/2}} \exp\left\{ -\frac{R^2}{2(A^2\sigma_b^2 + \sigma_n^2)} \right\} \\
p(A) &= \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left\{ -\frac{A^2}{2\sigma_a^2} \right\} \,.
\end{aligned}$$

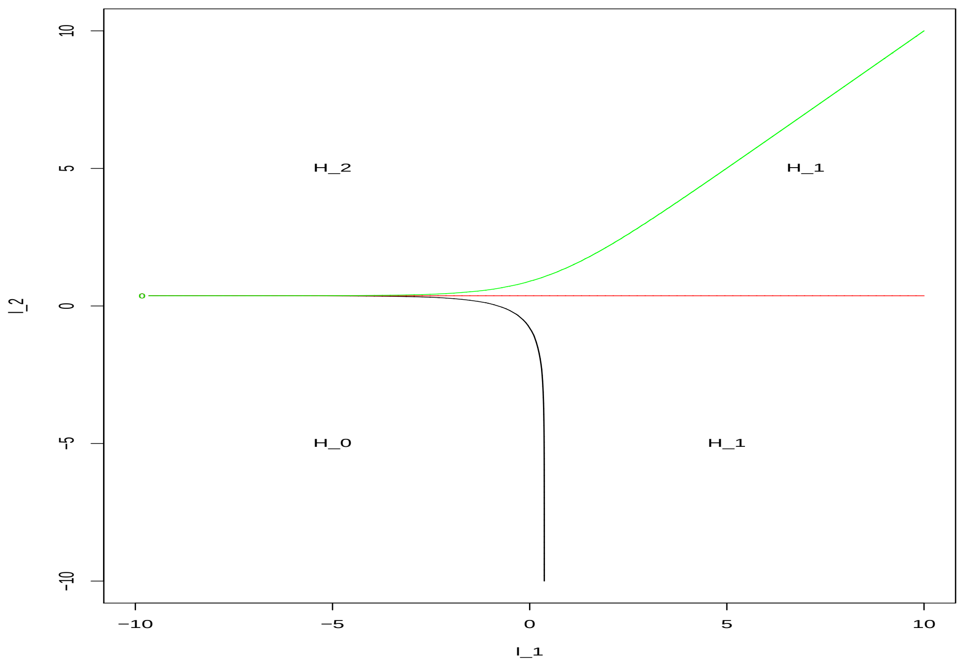

Figure 12: Plots of decision regions for Problem 2.3.6. The regions into which we would classify each set are labeled.

Note to calculate our estimate $\hat{a}_{\text{map}}(R)$ we don't need to explicitly calculate $p(R)$, since it is not a function of A. From the above densities

$$\begin{aligned}\ln(p(R|A)) &= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(A^2\sigma_b^2 + \sigma_n^2) - \frac{R^2}{2(A^2\sigma_b^2 + \sigma_n^2)} \quad \text{and} \\ \ln(p(A)) &= -\frac{1}{2}\ln(2\pi) - \ln(\sigma_a) - \frac{A^2}{2\sigma_a^2}\,,\end{aligned}$$

We now need to find the value A such that

$$\frac{\partial \ln(p(A|R))}{\partial A} = 0\,.$$

This equation is

$$-\frac{A\sigma_b^2}{A^2\sigma_b^2 + \sigma_n^2} + \frac{A\sigma_b^2 R^2}{(A^2\sigma_b^2 + \sigma_n^2)^2} - \frac{A}{\sigma_a^2} = 0\,.$$

One solution to the above is $A = 0$. If $A \neq 0$ and let $v \equiv A^2\sigma_b^2 + \sigma_n^2$ then the equation above is equivalent to

$$-\frac{\sigma_b^2}{v} + \frac{\sigma_b^2 R^2}{v^2} - \frac{1}{\sigma_a^2} = 0\,,$$

or

$$v^2 + \sigma_a^2\sigma_b^2 v - \sigma_a^2\sigma_b^2 R^2 = 0\,.$$

Using the quadratic equation the solution for v is given by

$$v = \frac{-\sigma_a^2\sigma_b^2 \pm \sqrt{\sigma_a^4\sigma_b^4 + 4\sigma_a^2\sigma_b^2 R^2}}{2} = \frac{-\sigma_a^2\sigma_b^2 \pm \sigma_a^2\sigma_b^2\sqrt{1 + \frac{4R^2}{\sigma_a^2\sigma_b^2}}}{2}\,.$$

From the definition of v we expect $v > 0$ thus we must take the expression with the positive sign. Now that we know v we can solve for $A^2\sigma_b^2 + \sigma_n^2 = v$ for A from which we would find two additional solutions for $\hat{a}_{map}(R)$.

Part (2): If we desire to simultaneously find $\hat{a}_{\text{map}}$ and $\hat{b}_{\text{map}}$ then we must treat our two unknowns a and b in a vector (a, b) and then find them both at the same time. Thus our unknown "**A**" is now the vector (a, b) and we form the log-likelihood l given by

$$
\begin{aligned}
l &= \ln(p(R|A,B)) + \ln(p(A,B)) - \ln(p(R)) \\
&= \ln(p(R|A,B)) + \ln(p(A)) + \ln(p(B)) - \ln(p(R)) \\
&= \ln\left(\frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{1}{2}\frac{(R-AB)^2}{\sigma_n^2}}\right) + \ln\left(\frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{1}{2}\frac{A^2}{\sigma_a^2}}\right) + \ln\left(\frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{1}{2}\frac{B^2}{\sigma_b^2}}\right) - \ln(p(R)) \,.
\end{aligned}
$$

We want to maximize l with respect to A and B so we take derivatives of l with respect to A and B, set the result equal to zero, and then solve for A and B. The A and B derivative of l are given by

$$
\begin{aligned}
\frac{(R-AB)B}{\sigma_n^2} - \frac{B}{\sigma_b^2} &= 0 \\
\frac{(R-AB)A}{\sigma_n^2} - \frac{A}{\sigma_a^2} &= 0 \,.
\end{aligned}
$$

One solution to these equations is $A = B = 0$. This is the only solution for if A and B are both non-zero then the above two equations become

$$
R - AB = \frac{\sigma_n^2}{\sigma_b^2} \quad \text{and} \quad R - AB = \frac{\sigma_n^2}{\sigma_a^2} \,,
$$

thus $R - AB$ is equal to two different things implying contradicting equations. Even if $\sigma_b^2 = \sigma_a^2$ so that the two equations above are identical, we would then have only one equation for the two unknowns A and B. Thus it looks like in this case that the only solutions are $\hat{a}_{map} = \hat{b}_{map} = 0$. This is *different* than the estimate of $\hat{a}_{map}$ we obtained in Part (1) above.

Part (3-a): For the expression for $r = a + \sum_{i=1}^{k} b_i + n$ we have that

$$
p(R|A) = \frac{1}{\sqrt{2\pi}(k\sigma_b^2 + \sigma_n^2)^{1/2}} \exp\left\{-\frac{(R-A)^2}{2(k\sigma_b^2 + \sigma_n^2)}\right\} \,.
$$

Thus

$$
\ln(p(R|A)) = \ln\left(\frac{1}{\sqrt{2\pi}(k\sigma_b^2 + \sigma_n^2)^{1/2}}\right) - \frac{(R-A)^2}{2(k\sigma_b^2 + \sigma_n^2)} \,.
$$

With $l \equiv \ln(p(R|A)) + \ln(p(A)) - \ln(p(R))$ taking the A derivative of l, and setting it equal to zero, gives

$$
\frac{R-A}{k\sigma_b^2 + \sigma_n^2} - \frac{A}{\sigma_a^2} = 0 \,.
$$

If we solve the above for A we find

$$
A = \frac{\sigma_a^2 R}{\sigma_a^2 + k\sigma_b^2 + \sigma_n^2} \,, \tag{94}
$$

for the expression for $\hat{a}_{map}(R)$.

Part (3-b): In this case we have $\mathbf{A} = [a, b_1, b_2, \cdots, b_{k-1}, b_k]^T$ and the densities we need are given by

$$p(R|\mathbf{A}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{ -\frac{(R - A - \sum_{i=1}^{l} B_i)^2}{2\sigma_n^2} \right\}$$

$$p(\mathbf{A}) = \left(\frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{A^2}{2\sigma_a^2}} \right) \left(\prod_{i=1}^{l} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{B_i^2}{2\sigma_b^2}} \right) .$$

Then computing the log-likelihood $l \equiv \ln(p(R|\mathbf{A})) + \ln(p(\mathbf{A})) - \ln(p(R))$ we find

$$l = -\frac{(R - A - \sum_{i=1}^{l} B_i)^2}{2\sigma_n^2} - \frac{A^2}{2\sigma_a^2} - \frac{1}{2\sigma_b^2} \sum_{i=1}^{l} B_i^2 + \text{constants} .$$

If we take the derivative of l with respect to A and set it equal to zero we get

$$\frac{1}{\sigma_n^2} \left(R - A - \sum_{i=1}^{l} B_i \right) - \frac{A}{\sigma_a^2} = 0 .$$

If we take the derivative of l with respect to B_i and set it equal to zero we get

$$\frac{1}{\sigma_n^2} \left(R - A - \sum_{i'=1}^{l} B_{i'} \right) - \frac{B_i}{\sigma_b^2} = 0 ,$$

for $i = 1, 2, \cdots, k-1, k$. As a system of equations this is

$$\begin{bmatrix} \left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma_a^2}\right) & \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} & \cdots & \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} \\ \frac{1}{\sigma_n^2} & \left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma_b^2}\right) & \frac{1}{\sigma_n^2} & \cdots & \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} \\ \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} & \left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma_b^2}\right) & \cdots & \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} \\ \vdots & & & \ddots & & \vdots \\ \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} & \cdots & \left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma_b^2}\right) & \frac{1}{\sigma_n^2} \\ \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} & \cdots & \frac{1}{\sigma_n^2} & \left(\frac{1}{\sigma_n^2} + \frac{1}{\sigma_b^2}\right) \end{bmatrix} \begin{bmatrix} A \\ B_1 \\ B_2 \\ \vdots \\ B_{k-1} \\ B_k \end{bmatrix} = \frac{R}{\sigma_n^2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}$$

This system must have a solution that depends on k (the number of b_i terms). While it might be hard to solve the full system above analytically we can solve smaller systems, look for a pattern and hope it generalizes. If anyone knows how to solve this system analytically "as is" please contact me. When $k = 1$ we have the system

$$\begin{bmatrix} \frac{1}{\sigma_n^2} + \frac{1}{\sigma_a^2} & \frac{1}{\sigma_n^2} \\ \frac{1}{\sigma_n^2} & \frac{1}{\sigma_n^2} + \frac{1}{\sigma_b^2} \end{bmatrix} \begin{bmatrix} A \\ B_1 \end{bmatrix} = \frac{R}{\sigma_n^2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

Solving for A and B we get

$$A = \frac{\tau_n \tau_b R}{\tau_a \tau_b + \tau_b \tau_n + \tau_a \tau_n} \quad \text{and} \quad B_1 = \frac{\tau_n \tau_a R}{\tau_a \tau_b + \tau_b \tau_n + \tau_a \tau_n} .$$

Here we have introduced the *precision* τ which is defined as one over the variance. For instance we have $\tau_a \equiv \frac{1}{\sigma_a^2}$, $\tau_b \equiv \frac{1}{\sigma_b^2}$, and $\tau_n \equiv \frac{1}{\sigma_n^2}$. If we do the same thing when $k = 2$ we find solutions for A, B_1, and B_2 given by

$$A = \frac{\tau_n \tau_b R}{\tau_a \tau_b + 2\tau_a \tau_n + \tau_b \tau_n} \quad \text{and} \quad B_1 = B_2 = \frac{\tau_n \tau_a R}{\tau_a \tau_b + 2\tau_a \tau_n + \tau_b \tau_n} .$$

If we do the same thing for $k = 3$ we get

$$A = \frac{\tau_n \tau_b R}{\tau_a \tau_b + 3\tau_a \tau_n + \tau_b \tau_n} \quad \text{and} \quad B_1 = B_2 = B_3 = \frac{\tau_n \tau_a R}{\tau_a \tau_b + 3\tau_a \tau_n + \tau_b \tau_n} .$$

For general k it looks like the solution is

$$A = \frac{\tau_n \tau_b R}{\tau_a \tau_b + k\tau_a \tau_n + \tau_b \tau_n} \quad \text{and} \quad B_1 = B_2 = \cdots = B_{k-1} = B_k = \frac{\tau_n \tau_a R}{\tau_a \tau_b + k\tau_a \tau_n + \tau_b \tau_n} .$$

See the Mathematical file `chap_2_prob_2_4_1.nb` where we solve the above linear system for several values of k. If we take the above expression for A (assuming k terms B_i in the summation) and multiply by $\frac{1}{\tau_n \tau_a \tau_b}$ on top and bottom of the fraction we get

$$A = \frac{\sigma_a^2 R}{\frac{1}{\tau_n} + \frac{k}{\tau_b} + \frac{1}{\tau_a}} = \frac{\sigma_a^2 R}{\sigma_n^2 + k\sigma_b^2 + \sigma_a^2} ,$$

which is the *same* as Equation 94.

Problem 2.4.2 (a first reproducing density example)

Part (1): Before any measurements are made we are told that the density of λ is a gamma distribution defined by

$$p(\lambda | n_*, l_*) = \frac{{l_*}^{n_*}}{\Gamma(n_*)} e^{-\lambda l_*} \lambda^{n_* - 1} \quad \text{for} \quad \lambda \geq 0 , \tag{95}$$

and is zero for $\lambda < 0$. This density has an expectation and variance [1] given by

$$E(\lambda) = \frac{n_*}{l_*} \tag{96}$$

$$\text{Var}(\lambda) = \frac{n_*}{{l_*}^2} . \tag{97}$$

Part (2): After the first measurement is made we need to compute the a posterior distribution of λ using Bayes' rule

$$p(\lambda | X) = \frac{p(X|\lambda) p(\lambda)}{p(X)} .$$

For the densities given we have

$$p(\lambda | X) = \frac{1}{p(X)} \left(\lambda e^{-\lambda X} \frac{{l_*}^{n_*}}{\Gamma(n_*)} e^{-\lambda l_*} \lambda^{n_* - 1} \right) = \left(\frac{{l_*}^{n_*}}{p(X) \Gamma(n_*)} \right) e^{-\lambda (l_* + X)} \lambda^{n_*} .$$

In the above expression the factor $p(X)$ just contributes to the normalization constant in that $p(\lambda|X)$ when integrated over valid values of λ should evaluate to one. Now notice that the functional form for $p(\lambda|X)$ derived above is of the same form as a gamma distribution just like the priori distribution $p(\lambda)$ was. That is (with the proper normalization) we see that $p(\lambda|X)$ must be

$$p(\lambda|X) = \frac{(l_* + X)^{n_*+1}}{\Gamma(n_* + 1)} e^{-\lambda(l_*+X)} \lambda^{n_*} .$$

Then $\hat{\lambda}_{ms}$ is just the mean of λ with the above density. From Equations 96 and 97 we see that for the a posteriori density that we have here

$$\hat{\lambda}_{ms} = \frac{n_* + 1}{l_* + X} \quad \text{and} \quad E[(\hat{\lambda}_{ms} - \lambda)^2] = \frac{n_* + 1}{(l_* + X)^2} .$$

Part (3): When we now have several measurements X_i the calculation of the a posteriori density is as follows

$$\begin{aligned} p(\lambda|X) &= \frac{p(X|\lambda)p(\lambda)}{p(X)} = \frac{1}{p(X)} \left(\prod_{i=1}^{n} p(X_i|\lambda) \right) p(\lambda) = \frac{1}{p(X)} \left(\prod_{i=1}^{n} \lambda e^{-\lambda X_i} \right) \left(\frac{{l_*}^{n_*}}{\Gamma(n_*)} e^{-\lambda l_*} \lambda^{n_*-1} \right) \\ &= \frac{{l_*}^{n_*}}{p(X)\Gamma(n_*)} \exp\left\{ -\lambda \left(l_* + \sum_{i=1}^{n} X_i \right) \right\} \lambda^{n+n_*-1} . \end{aligned}$$

Again we recognize this is a gamma distribution with different parameters so the normalization constant can be determined and the full a posteriori density is

$$p(\lambda|X) = \frac{(l_* + \sum_{i=1}^{n} X_i)^{n+n_*}}{\Gamma(n + n_*)} \exp\left\{ -\lambda \left(l_* + \sum_{i=1}^{n} X_i \right) \right\} \lambda^{n+n_*-1} .$$

Again using Equations 96 and 97 our estimates of λ and its variance are now given by

$$\hat{\lambda}_{ms} = \frac{n_* + n}{l_* + \sum_{i=1}^{n} X_i} \quad \text{and} \quad E[(\hat{\lambda}_{ms} - \lambda)^2] = \frac{n_* + n}{(l_* + \sum_{i=1}^{n} X_i)^2} .$$

Part (4): We now compute $\hat{\lambda}_{map}$ for the density $p(\lambda|X)$ computed in Part 3 above. Taking the λ derivative of $\ln(p(\lambda|X))$ and using the l' and n' notation in the book we have

$$\frac{d}{d\lambda} \ln(p(\lambda|X)) = \frac{d}{d\lambda} \ln\left(\frac{l'^{n'}}{\Gamma(n')} e^{-\lambda l'} \lambda^{n'-1} \right) = \frac{d}{d\lambda} \left[-\lambda l' + (n' - 1) \ln(\lambda) \right] = -l' + \frac{(n' - 1)}{\lambda} .$$

Setting this expression equal to zero and solving for λ then gives

$$\hat{\lambda}_{map} = \frac{n' - 1}{l'} = \frac{n_* + n - 1}{l_* + \sum_{i=1}^{n} X_i} .$$

Note that this is different from $\hat{\lambda}_{ms}$ computed in Part (3) by the term $\frac{-1}{l_* + \sum_{i=1}^{n} X_i}$.

Problem 2.4.3 (normal conjugate densities)

Part (1): If we assume that the priori distribution of a is $N\left(m_0, \frac{\sigma_n}{k_0}\right)$, then the a posteriori density is given using Bayes' rule and the density for $p(R|A)$. We find

$$
\begin{aligned}
p(A|R) = \frac{p(R|A)p(A)}{p(R)} &= \frac{1}{p(R)}\left(\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{(R-A)^2}{2\sigma_n^2}\right]\right)\left(\frac{k_0}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{(A-m_0)^2}{2\left(\frac{\sigma_n^2}{k_0^2}\right)}\right]\right) \\
&= \frac{k_0}{2\pi\sigma_n^2 p(R)}\exp\left\{-\frac{1}{2\sigma_n^2}\left((R-A)^2 + k_0^2(A-m_0)^2\right)\right\} \\
&= \alpha\exp\left\{-\frac{1}{2\sigma_n^2}\left(A^2 - 2RA + R^2 + k_0^2A^2 - 2k_0^2 m_0 A + k_0^2 m_0^2\right)\right\} \\
&= \alpha'\exp\left\{-\frac{1}{2\sigma_n^2}\left((1+k_0^2)A^2 - 2(R+k_0^2 m_0)A\right)\right\} = \alpha'\exp\left\{-\frac{1}{2\left(\frac{\sigma_n^2}{1+k_0^2}\right)}\left(A^2 - 2\left(\frac{R+k_0^2 m_0}{1+k_0^2}\right)A\right)\right\} \\
&= \alpha'\exp\left\{-\frac{1}{2\left(\frac{\sigma_n^2}{1+k_0^2}\right)}\left(A^2 - 2\left(\frac{R+k_0^2 m_0}{1+k_0^2}\right)A + \left(\frac{R+k_0^2 m_0}{1+k_0^2}\right)^2 - \left(\frac{R+k_0^2 m_0}{1+k_0^2}\right)^2\right)\right\} \\
&= \alpha''\exp\left\{-\frac{1}{2\left(\frac{\sigma_n^2}{1+k_0^2}\right)}\left(A - \left(\frac{R+k_0^2 m_0}{1+k_0^2}\right)\right)^2\right\}.
\end{aligned}
$$

Here α, α', and α'' are constants that don't depend on A. Notice that this expression is a normal density with a mean and variance given by

$$
m_1 = \frac{R + k_0^2 m_0}{1+k_0^2} \quad \text{and} \quad \sigma_1^2 = \frac{\sigma_n^2}{1+k_0^2}.
$$

Part (2): If we now have N independent observations of R denoted $R_1, R_2, R_3, \cdots, R_{N-1}, R_N$ then the likelihood $p(R|A)$ is now given by the product of the individual densities or

$$
p(R|A) = \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left[-\frac{(R_i - A)^2}{2\sigma_n^2}\right].
$$

Using this the a posteriori density is given by

$$\begin{aligned}
p(A|R) &= \alpha \exp\left[-\frac{1}{2\sigma_n}\sum_{i=1}^{N}(R_i - A)^2\right]\exp\left[-\frac{(A-m_0)^2}{2\left(\frac{\sigma_n^2}{k_0^2}\right)}\right] \\
&= \alpha \exp\left[-\frac{1}{2\sigma_n}\left(\sum_{i=1}^{N}(R_i^2 - 2AR_i + A^2)\right)\right]\exp\left[-\frac{(A^2-2Am_0+m_0^2)}{2\left(\frac{\sigma_n^2}{k_0^2}\right)}\right] \\
&= \alpha' \exp\left[-\frac{1}{2\sigma_n}\left(NA^2 - 2A\sum_{i=1}^{N}R_i\right)\right]\exp\left[-\frac{(A^2-2Am_0)}{2\left(\frac{\sigma_n^2}{k_0^2}\right)}\right] \\
&= \alpha' \exp\left[-\frac{1}{2}\left\{\left(\frac{N}{\sigma_n^2} + \frac{k_0^2}{\sigma_n^2}\right)A^2 - 2\left(\frac{\sum_{i=1}^{N}R_i}{\sigma_n^2} + \frac{m_0 k_0^2}{\sigma_n^2}\right)A\right\}\right] \\
&= \alpha' \exp\left[-\frac{1}{2\left(\frac{\sigma_n^2}{N+k_0^2}\right)}\left\{A^2 - \frac{2}{\sigma_n^2}\left(\frac{\sigma_n^2}{N+k_0^2}\right)\left(\sum_{i=1}^{N}R_i + m_0k_0^2\right)A\right\}\right] \\
&= \alpha'' \exp\left[-\frac{1}{2\left(\frac{\sigma_n^2}{N+k_0^2}\right)}\left\{A - \left(\frac{1}{N+k_0^2}\right)\left(\sum_{i=1}^{N}R_i + m_0k_0^2\right)\right\}^2\right].
\end{aligned}$$

When we put the correct normalization we see that the above is a normal density with mean and variance given by

$$m_N = \frac{m_0k_0^2 + N\left(\frac{1}{N}\sum_{i=1}^{N}R_i\right)}{k_0^2+N} \quad \text{and} \quad \sigma_N^2 = \frac{\sigma_n^2}{N+k_0^2}.$$

Problem 2.4.4 (more conjugate priors)

Part (1): For the densities given we find that

$$\begin{aligned}
p(A|R) = \frac{p(R|A)p(A)}{p(R)} &= \frac{1}{p(R)}\left(\prod_{i=1}^{N}\frac{A^{1/2}}{(2\pi)^{1/2}}\exp\left(-\frac{A}{2}(R_i-m)^2\right)\right)\left(cA^{\frac{k_1}{2}-1}\exp\left(-\frac{1}{2}Ak_1k_2\right)\right) \\
&= \left(\frac{c}{p(R)(2\pi)^{N/2}}\right)A^{\frac{N}{2}+\frac{k_1}{2}-1}\exp\left(-\frac{1}{2}Ak_1k_2 - \frac{A}{2}\sum_{i=1}^{N}(R_i-m)^2\right) \\
&= \left(\frac{c}{p(R)(2\pi)^{N/2}}\right)A^{\frac{N}{2}+\frac{k_1}{2}-1}\exp\left(-\frac{1}{2}A\left(k_1k_2 + \sum_{i=1}^{N}(R_i-m)^2\right)\right).
\end{aligned} \tag{98}$$

This will have the same form as the a priori distribution for A with parameters k_1' and k_2' if

$$\frac{k_1'}{2} - 1 = \frac{k_1+N}{2} - 1 \quad \text{or} \quad k_1' = k_1 + N,$$

and

$$k_1'k_2' = k_1k_2 + \sum_{i=1}^{N}(R_i-m)^2.$$

Solving for k_2' (since we know the value of k_1') we get

$$k_2' = \frac{1}{k_1'}\left(k_1k_2 + N\left(\frac{1}{N}\sum_{i=1}^{N}(R_i-m)^2\right)\right).$$

These are the same expressions in the book.

Part (2): To find $\hat{a}_{ms}$ we must find the conditional mean of the density given by Equation 98 which is of the same form of the a priori distribution (but with different parameters). From the above, the a posteriori distribution has the form

$$p(A|k_1', k_2') = cA^{\frac{k_1'}{2}-1} \exp\left(-\frac{1}{2}Ak_1'k_2'\right) .$$

As the gamma distribution looks like

$$p(A|\lambda, r) = \frac{1}{\Gamma(r)}(\lambda A)^{r-1}\lambda e^{-\lambda A} , \tag{99}$$

we see that the a posteriori distribution, computed above, is also a gamma distribution with $\lambda = \frac{1}{2}k_1'k_2'$ and $r = \frac{k_1'}{2}$. Using the known expression for the mean of a gamma distribution in the form of Equation 99 we have

$$E[A] = \frac{r}{\lambda} = \frac{\frac{k_1'}{2}}{\frac{1}{2}k_1'k_2'} = \frac{1}{k_2'} .$$

Thus the conditional mean of our a posteriori distribution is given by

$$\hat{a}_{ms} = \frac{k_1'}{k_1k_2 + Nw} = \frac{k_1 + N}{k_1k_2 + Nw} .$$

Problem 2.4.5 (recursive estimation)

Part (1): Recall that from Problem 2.4.3 we have that with K observations of R_i that $p(A|\mathbf{R}) = p(A|\{R_i\}_{i=1}^K)$ is a normal density with a mean m_K and a variance σ_K^2 given by

$$m_K = \frac{m_0k_0^2 + Kl}{K + k_0^2} \quad \text{and} \quad \sigma_K^2 = \frac{\sigma_n^2}{K + k_0^2} \quad \text{with} \quad l = \frac{1}{K}\sum_{i=1}^K R_i . \tag{100}$$

To make the a priori model for this problem match the one from Problem 2.4.3 we need to take $m_0 = 0$ and $\frac{\sigma_n}{k_0} = \sigma_a$ or $k_0 = \frac{\sigma_n}{\sigma_a}$. Thus the mean m_K of the a posteriori distribution for a becomes (after K measurements)

$$m_K = \hat{a}_{ms} = \frac{K\left(\frac{1}{K}\sum_{i=1}^K R_i\right)}{K + \frac{\sigma_n^2}{\sigma_a^2}} = \left(\frac{1}{1 + \frac{\sigma_n^2}{K\sigma_a^2}}\right)\left(\frac{1}{K}\sum_{i=1}^K R_i\right) . \tag{101}$$

Part (2): The MAP estimator takes the priori information on a to be infinitely weak i.e. $\sigma_a^2 \to +\infty$. If we take that limit in Equation 101 we get $\hat{a}_{map} = \frac{1}{K}\sum_{i=1}^K R_i$.

Part (3): The mean square error is the integral of the conditional variance over all possible values for R i.e. the book's equation 126 we get

$$R_{ms} = \int_{-\infty}^{\infty} dRp(R)\int_{-\infty}^{\infty}[A - \hat{a}_{ms}(R)]p(A|R)dA = \int_{-\infty}^{\infty} dRp(R)\left(\frac{\sigma_n^2}{K + k_0^2}\right) = \frac{\sigma_n^2}{K + k_0^2} ,$$

or the conditional variance again (this true since the conditional variance is independent of the measurement R).

Part (4-a): Arguments above show that each density over only the first j measurements or $p(A|\{R_i\}_{i=1}^{j})$ should be normal with a mean and a variance which we will denote by

$$\hat{a}_j(R_1, R_2, \cdots, R_j) \quad \text{and} \quad \sigma_j^2 \,.$$

We can now start a recursive estimation procedure with our initial mean and variance estimates given by $\hat{a}_0 = 0$ and $\sigma_0^2 = \sigma_a^2$ and using Bayes' rule to link mean and variance estimates as new measurements arrive. For example, we will use

$$p(A|R_1, R_2, \cdots, R_{j-1}, R_j) = \frac{p(A|R_1, R_2, \cdots, R_{j-1}) p(R_j|A, R_1, R_2, \cdots, R_{j-1})}{p(R_1, R_2, \cdots, R_{j-1}, R_j)} = \frac{p(A|R_1, R_2, \cdots, R_{j-1}) p(R_j|A)}{p(R_1, R_2, \cdots, R_{j-1}, R_j)} \,.$$

This is (using the densities defined above)

$$\begin{aligned}
p(A|R_1, R_2, \cdots, R_{j-1}, R_j) &= \frac{1}{p(R_1, \cdots, R_j)} \left(\frac{1}{\sqrt{2\pi}\sigma_{j-1}} \exp\left\{ -\frac{1}{2\sigma_{j-1}^2}(A - \hat{a}_{j-1})^2 \right\} \right) \left(\frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{ -\frac{1}{2\sigma_n^2}(A - R_j)^2 \right\} \right) \\
&= \frac{1}{2\pi\sigma_{j-1}\sigma_n p(R_1, \cdots, R_j)} \exp\left\{ -\frac{1}{2\sigma_{j-1}^2}(A^2 - 2\hat{a}_{j-1}A + \hat{a}_{j-1}^2) - \frac{1}{2\sigma_n^2}(A^2 - 2R_jA + R_j^2) \right) \\
&= \alpha \exp\left\{ -\frac{1}{2} \left[\left(\frac{1}{\sigma_{j-1}^2} + \frac{1}{\sigma_n^2} \right) A^2 - 2 \left(\frac{\hat{a}_{j-1}}{\sigma_{j-1}^2} + \frac{R_j}{\sigma_n^2} \right) A \right] \right\} \\
&= \alpha \exp\left\{ -\frac{1}{2} \left(\frac{1}{\sigma_{j-1}^2} + \frac{1}{\sigma_n^2} \right) \left[A^2 - 2 \frac{\left(\frac{\hat{a}_{j-1}}{\sigma_{j-1}^2} + \frac{R_j}{\sigma_n^2} \right)}{\left(\frac{1}{\sigma_{j-1}^2} + \frac{1}{\sigma_j^2} \right)} A \right] \right\} .
\end{aligned}$$

From the above expression we will define the variance of $p(A|\{R_i\}_{i=1}^{j})$ or σ_j^2 as

$$\frac{1}{\sigma_j^2} \equiv \frac{1}{\sigma_{j-1}^2} + \frac{1}{\sigma_n^2} \quad \text{so} \quad \sigma_j^2 = \frac{\sigma_{j-1}^2 \sigma_n^2}{\sigma_n^2 + \sigma_{j-1}^2} \,. \tag{102}$$

Using this the above becomes

$$p(A|R_1, R_2, \cdots, R_{j-1}, R_j) = \alpha' \exp\left\{ -\frac{1}{2\sigma_j^2} \left[A - \left(\frac{\hat{a}_{j-1}}{\sigma_{j-1}^2} + \frac{R_j}{\sigma_n^2} \right) \sigma_j^2 \right]^2 \right\} .$$

With the correct normalization this is a normal density with a mean given by

$$\begin{aligned}
\hat{a}_j(R_1, R_2, \cdots, R_{j-1}, R_j) &= \left(\frac{\hat{a}_{j-1}}{\sigma_{j-1}^2} + \frac{R_j}{\sigma_n^2} \right) \sigma_j^2 \\
&= \frac{\sigma_n^2}{\sigma_n^2 + \sigma_{j-1}^2} \hat{a}_{j-1}(R_1, R_2, \cdots, R_{j-1}) + \frac{\sigma_{j-1}^2}{\sigma_n^2 + \sigma_{j-1}^2} R_j \,.
\end{aligned}$$

This expresses the new mean $\hat{a}_j$ as a function of $\hat{a}_{j-1}$, σ_{j-1}^2 and the new measurement R_j.

From Equation 102 with $j = 1$ and $\sigma_0^2 = \sigma_a^2$ when we iterate a few of these terms we find

$$\begin{aligned}
\frac{1}{\sigma_1^2} &= \frac{1}{\sigma_a^2} + \frac{1}{\sigma_n^2} \\
\frac{1}{\sigma_2^2} &= \frac{1}{\sigma_1^2} + \frac{1}{\sigma_n^2} = \frac{1}{\sigma_a^2} + \frac{2}{\sigma_n^2} \\
\frac{1}{\sigma_3^2} &= \frac{1}{\sigma_a^2} + \frac{2}{\sigma_n^2} + \frac{1}{\sigma_n^2} = \frac{1}{\sigma_a^2} + \frac{3}{\sigma_n^2} \\
&\vdots \\
\frac{1}{\sigma_j^2} &= \frac{1}{\sigma_a^2} + \frac{j}{\sigma_n^2} \,.
\end{aligned}$$

Problem 2.4.6

Part (1): Using Equation 29 we can write our risk as

$$\mathcal{R}(\hat{a}|R) = \int_{-\infty}^{\infty} C(\hat{a}_{ms} - \hat{a} + Z) p(Z|R) dZ \,.$$

Introduce the variable $Y = \hat{a}_{ms} - \hat{a} + Z$ and the above becomes

$$\int_{-\infty}^{\infty} C(Y) p(Y + \hat{a} - \hat{a}_{ms}|R) dY \,, \tag{103}$$

which is the books P.4.

Part (2): We first notice that due to symmetry the MS risk is

$$\mathcal{R}(\hat{a}_{ms}|R) = \int_{-\infty}^{\infty} C(Z) p(Z|R) dZ = 2\int_{0}^{\infty} C(Z) p(Z|R) dZ \,,$$

and that using Equation 103 we can write this as

$$\begin{aligned}
R(\hat{a}|R) &= \int_{-\infty}^{0} C(Z) p(Z + \hat{a} - \hat{a}_{ms}|R) dZ + \int_{0}^{\infty} C(Z) p(Z + \hat{a} - \hat{a}_{ms}|R) dZ \\
&= \int_{0}^{\infty} C(Z) p(-Z + \hat{a} - \hat{a}_{ms}|R) dZ + \int_{0}^{\infty} C(Z) p(Z + \hat{a} - \hat{a}_{ms}|R) dZ \\
&= \int_{0}^{\infty} C(Z) p(Z - \hat{a} + \hat{a}_{ms}|R) dZ + \int_{0}^{\infty} C(Z) p(Z + \hat{a} - \hat{a}_{ms}|R) dZ
\end{aligned}$$

Then $\Delta\mathcal{R} \equiv \mathcal{R}(\hat{a}|R) - \mathcal{R}(\hat{a}_{ms}|R)$ can be computed by subtracting the two expressions above. Note that this would give the expression stated in the book except that the product of the densities would need to be subtraction.

Problem 2.4.7 (bias in the variance calculation?)

We are told to assume that $R_i \sim N(m, \sigma^2)$ for all i and that R_i and R_j are uncorrelated. We first compute some expectation of powers of R_i. Recalling the fact from probability of

$$E[X^2] = \text{Var}(X) + E[X]^2 \,,$$

and the assumed distribution of the R_i we have that

$$\begin{aligned} E[R_i] &= m \\ E[R_i R_j] &= \left\{ \begin{array}{cc} E[R_i]E[R_j] = m^2 & i \neq j \\ \text{Var}(R_i) + E[R_i]^2 = \sigma^2 + m^2 & i = j \end{array} \right. \,, \end{aligned}$$

note that we have used the fact that the R_i are uncorrelated in the evaluation of $E[R_i R_j]$ when $i \neq j$. Next define

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} R_i \,,$$

using this we can write our expression for V as

$$\begin{aligned} V &= \frac{1}{n} \sum_{i=1}^{n} (R_i - \bar{R})^2 = \frac{1}{n} \sum_{i=1}^{n} (R_i^2 - 2\bar{R}R_i + \bar{R}^2) \qquad (104) \\ &= \frac{1}{n} \sum_{i=1}^{n} R_i^2 - 2\frac{\bar{R}}{n} \sum_{i=1}^{n} R_i + \frac{n}{n}\bar{R}^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} R_i^2 - \bar{R}^2 \,. \qquad (105) \end{aligned}$$

We now ask if V is an unbiased estimator of σ^2 or whether $E[V] = \sigma^2$. Thus we take the expectation of V given by Equation 105. We find

$$E[V] = \frac{1}{n} \sum_{i=1}^{n} E[R_i^2] - E[\bar{R}^2] = \frac{1}{n} n(\sigma^2 + m^2) - E[\bar{R}^2] = \sigma^2 + m^2 - E[\bar{R}^2] \,.$$

To evaluate this we now need to compute the expectation of $\bar{R}^2$. From the definition of $\bar{R}$ we have

$$\bar{R}^2 = \left(\frac{1}{n} \sum_{i=1}^{n} R_i \right)^2 = \frac{1}{n^2} \sum_{i,j} R_i R_j \,.$$

Thus we have

$$E[\bar{R}^2] = \frac{1}{n^2} \sum_{i,j} E[R_i R_j] = \frac{1}{n^2} \left(n(\sigma^2 + m^2) + n(n-1)m^2 \right) = \frac{\sigma^2}{n} + m^2 \,. \qquad (106)$$

With this expression we can now compute $E[V]$ to find

$$E[V] = \sigma^2 + m^2 - \left(\frac{\sigma^2}{n} + m^2 \right) = \sigma^2 \frac{(n-1)}{n} \neq \sigma^2 \,. \qquad (107)$$

Thus the given expression for V is *not* unbiased.

Lets check that if we normalize "correctly" (divide the sum in Equation 104 by $n-1$ rather than n) we get an unbiased estimate of σ^2. Let this estimate of the variance be denoted by $\tilde{V}$. Using similar steps as in the above we get

$$\begin{aligned} \tilde{V} &\equiv \frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n} R_i^2 - 2\frac{\bar{R}}{n-1}\sum_{i=1}^{n} R_i + \frac{n}{n-1}\bar{R}^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n} R_i^2 - \frac{n}{n-1}\bar{R}^2 . \end{aligned}$$

In this case then we find

$$\begin{aligned} E[\tilde{V}] &= \frac{1}{n-1}\sum_{i=1}^{n} E[R_i^2] - \frac{n}{n-1}E[\bar{R}^2] \\ &= \frac{1}{n-1}n(\sigma^2 + m^2) - \frac{n}{n-1}\left(\frac{\sigma^2}{n} + m^2\right) = \sigma^2 , \end{aligned}$$

when we simplify. Thus $\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2$ is an *unbiased* estimator of σ^2.

Problem 2.4.8 (ML estimation of a binomial distribution)

We are told that R is distributed as a binomial random variable with parameters (A, n). This means that the probability we observe the value of R after n trials is given by

$$p(R|A) = \left(\begin{array}{c} n \\ R \end{array}\right) A^R (1-A)^{n-R} \quad \text{for} \quad 0 \le R \le n .$$

We desire to estimate the probability of success, A, from the measurement R.

Part (1): To compute the maximum likelihood (ML) estimate of A we compute

$$\hat{a}_{ml}(R) = \text{argmax}_A p(R|A) = \text{argmax}_A \left(\begin{array}{c} n \\ R \end{array}\right) A^R (1-A)^{n-R} .$$

To compute this maximum we can take the derivative of $p(R|A)$ with respect to A, set the resulting expression equal to zero and solve for A. We find the derivative equal to

$$\left(\begin{array}{c} n \\ R \end{array}\right) \left(RA^{R-1}(1-A)^{n-R} + A^R(n-R)(1-A)^{n-R-1}(-1)\right) = 0 .$$

Dividing by $A^{R-1}(1-A)^{n-R-1}$ to get

$$R(1-A) + A(n-R)(-1) = 0 ,$$

and solving for A gives or ML estimate of

$$\hat{a}_{ml}(R) = \frac{R}{n} \,. \tag{108}$$

Lets compute the bias and variance of this estimate of A. The bias, $b(A)$, is defined as

$$\begin{aligned} b(A) &= E[\hat{a} - A|A] = E[\hat{a}|A] - A \\ &= E\left[\left. \frac{R}{n} \right| A \right] - A = \frac{1}{n} E[R|A] - A \,. \end{aligned}$$

Now since R is drawn from a binomial random variable with parameters (n, A), the expectation of R is An, from which we see that the above equals zero and our estimator is unbiased. To study the conditional variance of our error (defined as $e = \hat{a} - A$) consider

$$\begin{aligned} \sigma_e^2(A) &= E[(e - E[e])^2|A] = E[e^2|A] = E[(\hat{a} - A)^2|A] \\ &= E[\left(\frac{1}{n}R - A\right)^2 |A] = \frac{1}{n^2} E[(R - nA)^2|A] \\ &= \frac{1}{n^2}(nA(1-A)) = \frac{A(1-A)}{n} \,. \end{aligned} \tag{109}$$

In the above we have used the result that the variance of a binomial random variable with parameters (n, A) is $nA(1 - A)$. In developing a ML estimator A is not considered random and as such the above expression is the desired variance of our estimator.

Part (2): Efficiency means that our estimator is unbiased and must *satisfy* the Cramer-Rao lower bound. One form of which is

$$\text{Var}[\hat{a}(R) - A] = \frac{1}{E\left\{ \left[\frac{\partial \ln(p(R|A))}{\partial A} \right]^2 \right\}} \,. \tag{110}$$

For the density considered here the derivative of the above is given by

$$\frac{\partial \ln(p(R|A))}{\partial A} = \frac{R}{A} - \frac{n-R}{1-A} = \frac{R - nA}{A(1-A)} \,. \tag{111}$$

Squaring this expression gives

$$\frac{R^2 - 2nAR + n^2A^2}{A^2(1-A)^2} \,.$$

Taking the expectation with respect to R and recalling that $E[R] = nA$ and

$$E[R^2] = \text{Var}[R] + E[R]^2 = nA(1-A) + n^2A^2$$

we have

$$\frac{(nA(1-A) + n^2A^2) - 2nA(nA) + n^2A^2}{A^2(1-A)^2} = \frac{n}{A(1-A)} \,,$$

when we simplify. As this is equal to $1/\sigma_e^2(A)$ computed in Equation 109 this estimator is efficient. Another way to see this argument is to note that in addition Equation 111 can be written in the efficiency required form of

$$\frac{\partial \ln(p(R|A))}{\partial A} = [\hat{a}(R) - A]k(A) \,.$$

by writing it as

$$\frac{\partial \ln(p(R|A))}{\partial A} = \left[\frac{R}{n} - A \right] \left(\frac{n}{A(1-A)} \right) \,.$$

Problem 2.4.11 (ML estimation of the mean and variance of a Gaussian)

Part (1): For the maximum likelihood estimate we need to find values of (A_1, A_2) that maximizes

$$p(\mathbf{R}|A_1, A_2) = \prod_{i=1}^{n} p(R_i|A_1, A_2) = \prod_{i=1}^{n} \frac{1}{(2\pi A_2)^{1/2}} \exp\left\{-\frac{(R_i - A_1)^2}{2A_2}\right\}$$
$$= \frac{1}{(2\pi)^{n/2}} \frac{1}{A_2^{n/2}} \exp\left\{-\frac{1}{2A_2}\sum_{i=1}^{n}(R_i - A_1)^2\right\}.$$

Taking logarithm of the above expression we desire to maximize we get

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(A_2) - \frac{1}{2A_2}\sum_{i=1}^{n}(R_i - A_1)^2.$$

To maximize this we must evaluate

$$\nabla_A[\ln(p(R|A))]|_{A=\hat{a}_{ml}(R)} = 0.$$

As a system of equations this is

$$\nabla_A[\ln(p(R|A))] = \left[\begin{array}{c} \frac{1}{A_2}\sum_{i=1}^{n}(R_i - A_1) \\ -\frac{n}{2}\frac{1}{A_2} + \frac{1}{2A_2^2}\sum_{i=1}^{n}(R_i - A_1)^2 \end{array}\right] = 0.$$

The first equation gives

$$\hat{a}_1 = \frac{1}{n}\sum_{i=1}^{n} R_i.$$

When we put this expression into the second equation and solve for A_2 we get

$$\hat{a}_2 = \frac{1}{n}\sum_{i=1}^{n}(R_i - \hat{a}_1)^2.$$

Part (2): The estimator for A_1 is not biased while the estimator for A_2 is biased as was shown in Problem 2.4.7 above.

Part (3): The estimators are seen to be coupled since the estimate of A_2 depends on the value of A_1.

Problem 2.4.12 (sending a signal)

Part (1): The likelihood for $\mathbf{A} = (A_1, A_2)$ for this problem is given by

$$p(\mathbf{R}|\mathbf{A}) = p(R_1|\mathbf{A})p(R_2|\mathbf{A})$$
$$= \left(\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left\{-\frac{1}{2\sigma_n^2}(R_1 - S_1)^2\right\}\right)\left(\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left\{-\frac{1}{2\sigma_n^2}(R_2 - S_2)^2\right\}\right).$$

The logarithm of this expression is given by

$$\ln(p(R|A)) = -\frac{1}{2\sigma_n^2}(R_1 - S_1)^2 - \frac{1}{2\sigma_n^2}(R_2 - S_2)^2 + \text{constants}\,.$$

To find the maximum likelihood estimate of $\mathbf{A}$ we need to take the derivatives of $\ln(p(R|A))$ with respect to A_1 and A_2. We find

$$\begin{aligned}
\frac{\partial \ln(p(R|A))}{\partial A_1} &= \frac{1}{\sigma_n^2}(R_1 - S_1)\frac{\partial S_1}{\partial A_1} + \frac{1}{\sigma_n^2}(R_2 - S_2)\frac{\partial S_2}{\partial A_1} \\
&= \frac{1}{\sigma_n^2}(R_1 - S_1)x_{11} + \frac{1}{\sigma_n^2}(R_2 - S_2)x_{21} \quad \text{and} \\
\frac{\partial \ln(p(R|A))}{\partial A_2} &= \frac{1}{\sigma_n^2}(R_1 - S_1)\frac{\partial S_1}{\partial A_2} + \frac{1}{\sigma_n^2}(R_2 - S_2)\frac{\partial S_2}{\partial A_2} \\
&= \frac{1}{\sigma_n^2}(R_1 - S_1)x_{12} + \frac{1}{\sigma_n^2}(R_2 - S_2)x_{22}\,.
\end{aligned}$$

Each derivative would need to be equated to zero and the resulting system solved for A_1 and A_2. To do this we first solve for $\mathbf{S}$ and then with these known values we solve for $\mathbf{A}$. As a first step write the above as

$$\begin{aligned}
x_{11}S_1 + x_{21}S_2 &= x_{11}R_1 + x_{21}R_2 \\
x_{12}S_1 + x_{22}S_2 &= x_{12}R_1 + x_{22}R_2\,.
\end{aligned}$$

In matrix notation this means that $\mathbf{S}$ is given by

$$\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.$$

Here we have assumed the needed matrix inverses exits. Now that we know the estimate for $\mathbf{S}$ the estimate for $\mathbf{A}$ is given by

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.$$

We now seek to answer the question as to whether $\hat{a}_1(R)$ and $\hat{a}_2(R)$ are unbiased estimators of A_1 and A_2. Consider the expectation of the vector $\begin{bmatrix} \hat{a}_1(R) \\ \hat{a}_2(R) \end{bmatrix}$. We find

$$\begin{aligned}
E\begin{bmatrix} \hat{a}_1(R) \\ \hat{a}_2(R) \end{bmatrix} &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} E\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} E\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \\
&= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},
\end{aligned}$$

showing that our estimator is unbiased.

Part (2): Lets compute the variance of our estimator $\begin{bmatrix} \hat{a}_1(R) \\ \hat{a}_2(R) \end{bmatrix}$. To do this we first let matrix $\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$ be denoted as X and note that

$$\begin{bmatrix} \hat{a}_1(R) - A_1 \\ \hat{a}_2(R) - A_2 \end{bmatrix} = X^{-1} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} - X^{-1} X \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = X^{-1} \left(\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} - X \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \right)$$
$$= X^{-1} \left(X \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} - X \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \right) = X^{-1} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}.$$

Using this we know that

$$\mathrm{Var} \begin{bmatrix} \hat{a}_1(R) \\ \hat{a}_2(R) \end{bmatrix} = X^{-1} \left(\mathrm{Var} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \right) X^{-T}.$$

Since

$$\mathrm{Var} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & \sigma_n^2 \end{bmatrix} = \sigma_n^2 I,$$

the above becomes

$$\begin{aligned}
\mathrm{Var} \begin{bmatrix} \hat{a}_1(R) \\ \hat{a}_2(R) \end{bmatrix} &= \sigma_n^2 \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^{-1} \\
&= \sigma_n^2 \left(\frac{1}{x_{11}x_{22} - x_{12}x_{21}} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix} \right) \left(\frac{1}{x_{11}x_{22} - x_{12}x_{21}} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{bmatrix} \right) \\
&= \frac{\sigma_n^2}{(x_{11}x_{22} - x_{12}x_{21})^2} \begin{bmatrix} x_{22}^2 + x_{12}^2 & -x_{22}x_{21} - x_{12}x_{11} \\ -x_{22}x_{21} - x_{12}x_{11} & x_{21}^2 + x_{11}^2 \end{bmatrix}.
\end{aligned} \tag{112}$$

Part (3): Lets compute the Cramer-Rao bound for any unbiased estimator. To do this we need to evaluate the matrix J which has components given by

$$J_{ij} = E\left[\frac{\partial \ln(p(R|A))}{\partial A_i} \frac{\partial \ln(p(R|A))}{\partial A_j} \right] = -E\left[\frac{\partial^2 \ln(p(R|A))}{\partial A_i \partial A_j} \right].$$

Using the above expressions we compute

$$\begin{aligned}
\frac{\partial^2 \ln(p(R|A))}{\partial {A_1}^2} &= -\frac{1}{\sigma_n^2} x_{11} \frac{\partial S_1}{\partial A_1} - \frac{1}{\sigma_n^2} x_{21} \frac{\partial S_2}{\partial A_1} = -\frac{1}{\sigma_n^2}(x_{11}^2 + x_{21}^2) \\
\frac{\partial^2 \ln(p(R|A))}{\partial A_1 \partial A_2} &= -\frac{1}{\sigma_n^2} x_{11} \frac{\partial S_1}{\partial A_2} - \frac{1}{\sigma_n^2} x_{21} \frac{\partial S_2}{\partial A_2} = -\frac{1}{\sigma_n^2}(x_{11}x_{12} + x_{21}x_{22}) \\
\frac{\partial^2 \ln(p(R|A))}{\partial {A_2}^2} &= -\frac{1}{\sigma_n^2} x_{12} \frac{\partial S_1}{\partial A_2} - \frac{1}{\sigma_n^2} x_{22} \frac{\partial S_2}{\partial A_2} = -\frac{1}{\sigma_n^2}(x_{12}^2 + x_{22}^2).
\end{aligned}$$

Thus as a matrix we find that J is given by

$$J = \frac{1}{\sigma_n^2} \begin{bmatrix} x_{11}^2 + x_{21}^2 & x_{11}x_{12} + x_{21}x_{22} \\ x_{11}x_{12} + x_{21}x_{22} & x_{12}^2 + x_{22}^2 \end{bmatrix}.$$

The Cramer-Rao bound involves J^{-1} which we compute to be

$$J^{-1} = \frac{\sigma_n^2}{(x_{11}^2 + x_{21}^2)(x_{12}^2 + x_{22}^2) - (x_{11}x_{12} + x_{21}x_{22})^2} \begin{bmatrix} x_{11}^2 + x_{22}^2 & -(x_{11}x_{12} + x_{21}x_{22}) \\ -(x_{11}x_{12} + x_{21}x_{22}) & x_{11}^2 + x_{21}^2 \end{bmatrix}.$$

We can see how this expression compares with the one derived in Equation 112 if we expand the denominator D of the above fraction to get

$$\begin{aligned} D &= x_{11}^2 x_{12}^2 + x_{11}^2 x_{22}^2 + x_{21}^2 x_{12}^2 + x_{21}^2 x_{22}^2 - (x_{11}^2 x_{12}^2 + 2x_{11}x_{12}x_{21}x_{22} + x_{21}^2 x_{22}^2) \\ &= x_{11}^2 x_{22}^2 - 2x_{11}x_{12}x_{21}x_{22} + x_{21}^2 x_{12}^2 = (x_{11}x_{22} - x_{12}x_{21})^2 . \end{aligned}$$

Thus the above expression matches that from Equation 112 and we have that our estimate *is* efficient.

Problem 2.4.14 (a Poisson random variable)

Part (1): Here we are told that $P(R|A) = \frac{A^R e^{-A}}{R!}$ and A is nonrandom. The logarithm is given by

$$\ln(P(R|A)) = R\ln(A) - A - \ln(R!) .$$

The first two derivatives of the log-likelihood above are given by

$$\begin{aligned} \frac{\partial \ln(P(R|A))}{\partial A} &= \frac{R}{A} - 1 \\ \frac{\partial^2 \ln(P(R|A))}{\partial A^2} &= -\frac{R}{A^2} . \end{aligned}$$

By the Cramer-Rao bound we have that

$$\text{Var}[\hat{a}(R)] \geq \frac{-1}{E\left[\frac{\partial^2 \ln(P(R|A))}{\partial A^2}\right]} .$$

Thus we need to evaluate the expectation above. We find

$$E\left[\frac{\partial^2 \ln(P(R|A))}{\partial A^2}\right] = -E\left[\frac{R}{A^2}\right] = -\frac{1}{A^2}E[R] = -\frac{A}{A^2} = -\frac{1}{A} ,$$

using the fact that $E[A] = A$ see [1]. Thus we see that $\text{Var}[\hat{a}(R)] \geq A$.

Part (2): In this case we have

$$p(\mathbf{R}|A) = \prod_{i=1}^{n} p(R_i|A) = \prod_{i=1}^{n}\left(\frac{A^{R_i}e^{-A}}{R_i!}\right) = \frac{1}{\prod_{i=1}^{n} R_i!} e^{-nA} A^{\sum_{i=1}^{n} R_i} .$$

Now the logarithm of the above gives

$$\ln(p(\mathbf{R}|A)) = -nA + \left(\sum_{i=1}^{n} R_i\right)\ln(A) - \ln\left(\prod_{i=1}^{n} R_i!\right) .$$

To compute the maximum likelihood estimate of A we take the first derivative of the log-likelihood, set the result equal to zero and solve for A. This means that we need to solve

$$\frac{\partial \ln(p(\mathbf{R}|A))}{\partial A} = -n + \frac{\sum_{i=1}^{n} R_i}{A} = 0 .$$

This means A is given by

$$A = \frac{1}{n} \sum_{i=1}^{n} R_i \,.$$

To show that this is efficient we will write this as $k(A)(\hat{a}(R) - A)$ as

$$\frac{n}{A} \left(\frac{1}{n} \sum_{i=1}^{n} R_i - A \right) .$$

Thus $\hat{a}(R) = \frac{1}{n} \sum_{i=1}^{n} R_i$ is an efficient estimator.

Problem 2.4.15 (Cauchy random variables)

Part (1): Note that we have

$$p(\mathbf{R}|A) = \frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1 + (R_i - A)^2} \,,$$

so that

$$\ln(p(\mathbf{R}|A)) = - \sum_{i=1}^{n} \ln(1 + (R_i - A)^2) - n \ln(\pi) \,.$$

Taking the first derivative of the above with respect to A we find

$$\frac{\partial \ln(p(\mathbf{R}|A))}{\partial A} = \sum_{i=1}^{n} \frac{2(R_i - A)}{1 + (R_i - A)^2} \,.$$

The second derivative is then given by

$$\frac{\partial^2 \ln(p(\mathbf{R}|A))}{\partial A^2} = 2 \sum_{i=1}^{n} \left(\frac{-1}{1 + (R_i - A)^2} + \frac{(R_i - A)^2}{(1 + (R_i - A)^2)^2} \right) = 2 \sum_{i=1}^{n} \frac{-1 + (R_i - A)^2}{(1 + (R_i - A)^2)^2} \,.$$

By the Cramer-Rao inequality we have that

$$\text{Var}[\hat{a}(R)] \geq \frac{-1}{E\left\{ \frac{\partial^2 \ln((\mathbf{R}|A))}{\partial A^2} \right\}} \,.$$

Thus we need to evaluate the expectation in the denominator of the above fraction. We find

$$\begin{aligned}
E\left\{ \frac{\partial^2 \ln(p(R|A))}{\partial A^2} \right\} &= 2 \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{(-1 + (R_i - A)^2)}{(1 + (R_i - A)^2)^2} p(R_i|A) dR_i \\
&= \frac{2}{\pi} \sum_{i=1}^{n} \left[- \int_{-\infty}^{\infty} \frac{dR_i}{(1 + (R_i - A)^2)^3} + \int_{-\infty}^{\infty} \frac{(R_i - A)^2 dR_i}{(1 + (R_i - A)^2)^3} \right] \\
&= \frac{2}{\pi} \sum_{i=1}^{n} \left[- \int_{-\infty}^{\infty} \frac{dv}{(1 + v^2)^3} + \int_{-\infty}^{\infty} \frac{v^2 dv}{(1 + v^2)^3} \right] .
\end{aligned}$$

When we evaluate the above two integrals using Mathematica we find

$$E\left\{ \frac{\partial^2 \ln(p(R|A))}{\partial A^2} \right\} = \frac{2}{\pi} \sum_{i=1}^{n} \left[-\frac{3\pi}{8} + \left(\frac{\pi}{8} \right) \right] = \frac{2n}{\pi} \left(-\frac{2\pi}{8} \right) = -\frac{n}{2} \,.$$

Thus the Cramer-Rao inequality then gives $\text{Var}[\hat{a}(R)] \geq \frac{2}{n}$ as we were to show.

Problem 2.4.16 (correlated Gaussian random variables)

Part (1): We have

$$\begin{aligned}
p(\mathbf{R}|\rho) &= \prod_{i=1}^{n} p(R_{i1}, R_{i2}|\rho) \\
&= \prod_{i=1}^{n} \left(\frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{ -\frac{(R_{i1}^2 - 2\rho R_{i1}R_{i2} + R_{i2}^2)}{2(1-\rho^2)} \right\} \right) \\
&= \frac{1}{(2\pi)^n (1-\rho^2)^{n/2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} (R_{i1}^2 - 2\rho R_{i1}R_{i2} + R_{i2}^2) \right\} .
\end{aligned}$$

Taking the logarithm of the above we get

$$\ln(p(\mathbf{R}|\rho)) = -n\ln(2\pi) - \frac{n}{2}\ln(1-\rho^2) - \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} (R_{i1}^2 - 2\rho R_{i1}R_{i2} + R_{i2}^2) .$$

The ML estimate is given by solving $\frac{\partial \ln(p(\mathbf{R}|\rho))}{\partial \rho} = 0$ for ρ. The needed first derivative is

$$\begin{aligned}
\frac{\partial \ln(p(\mathbf{R}|\rho))}{\partial \rho} &= -\frac{n}{2}\frac{(-2\rho)}{(1-\rho^2)} + \frac{(-2\rho)}{2(1-\rho^2)^2} \sum_{i=1}^{n} (R_{i1}^2 - 2\rho R_{i1}R_{i2} + R_{i2}^2) - \frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} (-2R_{i1}R_{i2}) \\
&= \frac{n\rho}{1-\rho^2} + \frac{1}{1-\rho^2} \sum_{i=1}^{n} R_{i1}R_{i2} - \frac{\rho}{(1-\rho^2)^2} \sum_{i=1}^{n} (R_{i1}^2 - 2\rho R_{i1}R_{i2} + R_{i2}^2) .
\end{aligned}$$

To simplify expression this lets define some sums. We introduce

$$S_{11} \equiv \frac{1}{n} \sum_{i=1}^{n} R_{i1}^2 , \quad S_{12} \equiv \frac{1}{n} \sum_{i=1}^{n} R_{i1}R_{i2} , \quad S_{22} \equiv \frac{1}{n} \sum_{i=1}^{n} R_{i2}^2 .$$

With these the expression for $\frac{\partial \ln(p(\mathbf{R}|\rho))}{\partial \rho}$ now becomes

$$\begin{aligned}
\frac{\partial \ln(p(\mathbf{R}|\rho))}{\partial \rho} &= \frac{n\rho}{1-\rho^2} + \frac{n}{1-\rho^2} S_{12} - \frac{n\rho}{(1-\rho^2)^2}(S_{11} - 2\rho S_{12} + S_{22}) \\
&= \frac{n}{(1-\rho^2)^2} \left[-\rho^3 + \rho - \rho S_{11} + (1+\rho^2) S_{12} - \rho S_{22} \right] . \qquad (113)
\end{aligned}$$

For later work we will need to compute the second derivative of $\ln(p(\mathbf{R}|\rho))$ with respect to ρ. Using Mathematica we find

$$\frac{\partial^2 \ln(p(\mathbf{R}|\rho))}{\partial \rho^2} = -\frac{n}{(1-\rho^2)^3} \left[-1 + \rho^4 + (1+3\rho^2) S_{11} + (-6\rho - 2\rho^3) S_{12} + (1+3\rho^2) S_{22} \right] . \qquad (114)$$

The equation for the ML estimate of ρ is given by setting Equation 113 equal to zero and solving for ρ.

Part (2): One form of the Cramer-Rao inequality bounds the variance of $\hat{a}(R)$ by a function of the expectation of the expression $\frac{\partial^2 \ln(p(\mathbf{R}|\rho))}{\partial \rho^2}$. In this problem, the samples R_i are mean zero, variance one, and correlated with correlation ρ which means that

$$E[R_{i1}] = E[R_{i2}] = 0 , \quad E[R_{i1}^2] = E[R_{i2}^2] = 1 , \quad \text{and} \quad E[R_{i1}R_{i2}] = \rho .$$

Using these with the definitions of S_{11}, S_{22}, and S_{12} we get that

$$E[S_{11}] = E[S_{22}] = 1 \quad \text{and} \quad E[S_{12}] = \rho \,.$$

Using these expressions, the expectation of Equation 114 becomes

$$E\left\{ \frac{\partial^2 \ln(p(\mathbf{R}|\rho))}{\partial \rho^2} \right\} = -n \frac{1+\rho^2}{(1-\rho^2)^2} \,.$$

Thus using the "second derivative" form of the Cramer-Rao inequality or

$$E[(\hat{a}(R) - A)^2] \geq -\left\{ E\left[\frac{\partial^2 \ln(p_{r|a}(R|A))}{\partial A^2} \right] \right\}^{-1} \,, \tag{115}$$

we see that in this case we have

$$E[(\hat{a}(R) - A)^2] \geq \frac{(1-\rho^2)^2}{n(1+\rho^2)} \,.$$

Problem 2.4.17 (the Cramer-Rao inequality for biased estimates)

If A is nonrandom then

$$E[\hat{a}(R) - A] = \int p(R|A)[\hat{a}(R) - A]dR = \int p(R|A)\hat{a}(R)dR - A = (A + B(A)) - A = B(A) \,.$$

Taking the A derivative of both sides gives

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial A} \{p(R|A)[\hat{a}(R) - A]\} dR = B'(A) \,.$$

Following the steps in the book's proof of the Cramer-Rao bound we get an equation similar to the books 185 of

$$\int_{-\infty}^{\infty} \left[\frac{\partial \ln(p(R|A))}{\partial A} \sqrt{p(R|A)} \right] \left[\sqrt{p(R|A)}(\hat{a}(R) - A) \right] dR = 1 + B'(A) \,.$$

Using the Schwarz inequality 39 on the integral on the left-hand-side we get that

$$1 + B'(A) \leq \left\{ \int_{-\infty}^{\infty} \left[\frac{\partial \ln(p(R|A))}{\partial A} \sqrt{p(R|A)} \right]^2 dR \right\}^{1/2} \left\{ \int_{-\infty}^{\infty} \left[\sqrt{p(R|A)}(\hat{a}(R) - A) \right]^2 dR \right\}^{1/2} \,.$$

Squaring both sides we get

$$(1 + B'(A))^2 \leq \left\{ \int_{-\infty}^{\infty} \left[\frac{\partial \ln(p(R|A))}{\partial A} \sqrt{p(R|A)} \right]^2 dR \right\} \text{Var}[\hat{a}(R)] \,.$$

If we solve for $\text{Var}[\hat{a}(R)]$ we get

$$\text{Var}[\hat{a}(R)] \geq \frac{(1 + B'(A))^2}{E\left\{ \left[\frac{\partial \ln(p(R|A))}{\partial A} \right]^2 \right\}} \,,$$

which is the expression we desired to obtain.

Problem 2.4.21 (an alternative derivation of the Cramer-Rao inequality)

Part (1): The first component of $E[z] = 0$ is true because $\hat{a}(R)$ is an unbiased estimator. That the second component of $E[z]$ is zero can be shown to be true by considering

$$\int \frac{\partial \ln(p(R|A))}{\partial A} p(R|A) dR = \int 1 \frac{\partial \ln(p(R|A))}{\partial A} dR = \frac{\partial}{\partial A} \int p(R|A) dR = \frac{\partial}{\partial A}(1) = 0\,.$$

Part (2): The covariance matrix is $\Lambda_z = E[zz^T]$. The needed outer product is given by

$$zz^T = \left[\begin{array}{c} \hat{a}(R) - A \\ \frac{\partial \ln(p(R|A))}{\partial A} \end{array} \right] \left[\begin{array}{cc} \hat{a}(R) - A & \frac{\partial \ln(p(R|A))}{\partial A} \end{array} \right] = \left[\begin{array}{cc} (\hat{a}(R) - A)^2 & (\hat{a}(R) - A)\frac{\partial \ln(p(R|A))}{\partial A} \\ (\hat{a}(R) - A)\frac{\partial \ln(p(R|A))}{\partial A} & \left(\frac{\partial \ln(p(R|A))}{\partial A}\right)^2 \end{array} \right].$$

Following the same steps leading to Equation 60 we can show that the expectation of the $(1, 2)$ and $(2, 1)$ terms is one and that $E[zz^T]$ takes the form

$$E[zz^T] = \left[\begin{array}{cc} E[(\hat{a}(R) - A)^2] & 1 \\ 1 & E\left[\left(\frac{\partial \ln(p(R|A))}{\partial A}\right)^2\right] \end{array} \right].$$

As Λ_z is positive semidefinite we must have $|\Lambda_z| \geq 0$ which means that

$$E[(\hat{a}(R) - A)^2] E\left[\left(\frac{\partial \ln(p(R|A))}{\partial A}\right)^2\right] \geq 1\,,$$

or

$$E[(\hat{a}(R) - A)^2] \geq \frac{1}{E\left[\left(\frac{\partial \ln(p(R|A))}{\partial A}\right)^2\right]}\,,$$

which is the Cramer-Rao inequality. If equality holds in the Cramer-Rao inequality then this means that $|\Lambda_z| = 0$.

Problem 2.4.22 (a derivation of the random Cramer-Rao inequality)

Note that we can show that $E[z] = 0$ in the same was as was done in Problem 2.4.21. We now compute zz^T and find

$$zz^T = \left[\begin{array}{c} \hat{a}(R) - A \\ \frac{\partial \ln(p(R,A))}{\partial A} \end{array} \right] \left[\begin{array}{cc} \hat{a}(R) - A & \frac{\partial \ln(p(R,A))}{\partial A} \end{array} \right] = \left[\begin{array}{cc} (\hat{a}(R) - A)^2 & (\hat{a}(R) - A)\frac{\partial \ln(p(R,A))}{\partial A} \\ (\hat{a}(R) - A)\frac{\partial \ln(p(R,A))}{\partial A} & \left(\frac{\partial \ln(p(R,A))}{\partial A}\right)^2 \end{array} \right].$$

Consider the expectation of the $(1, 2)$ component of zz^T where using integration by parts we find

$$\begin{aligned}
\int (\hat{a}(R) - A)\frac{\partial \ln(p(R, A))}{\partial A} p(R, A) dR dA &= \int (\hat{a}(R) - A)\frac{\partial p(R, A)}{\partial A} dR dA \\
&= (\hat{a}(R) - A) p(R, A)|_{-\infty}^{\infty} - \int (-1) p(R, A) dR dA \\
&= 0 + \int p(R, A) dR dA = 1\,.
\end{aligned}$$

Using this we find that $E[zz^T]$ is given by

$$E[zz^T] = \begin{bmatrix} E[(\hat{a}(R) - A)^2] & 1 \\ 1 & E\left[\left(\frac{\partial \ln(p(R,A))}{\partial A}\right)^2\right] \end{bmatrix}.$$

As Λ_z is positive semidefinite we must have $|\Lambda_z| \geq 0$ which as in the previous problem means that

$$E[(\hat{a}(R) - A)^2] \geq \frac{1}{E\left[\left(\frac{\partial \ln(p(R,A))}{\partial A}\right)^2\right]},$$

which is the Cramer-Rao inequality when a is a random variable.

Problem 2.4.23 (the Bhattacharyya bound)

Part (1): Note that $E[z] = 0$ for the first and second component as in previous problems. Note that for $i \geq 1$ we have

$$\begin{aligned} E\left[\frac{1}{p(R|A)}\frac{\partial^i p(R|A)}{\partial A^i}\right] &= \int \frac{1}{p(R|A)}\frac{\partial^i p(R|A)}{\partial A^i} p(R|A) dR \\ &= \frac{\partial^i}{\partial A^i}\int p(R|A)dR = \frac{\partial^i}{\partial A^i}(1) = 0\,. \end{aligned}$$

Thus $E[z] = 0$ for all components and $\Lambda_z = E[zz^T]$.

Part (2): The first row of the outer product zz^T has components that look like

$$(\hat{a}(R) - A)^2\,, \quad \left(\frac{\hat{a}(R) - A}{p(R|A)}\right)\frac{\partial p(R|A)}{\partial A}\,, \quad \left(\frac{\hat{a}(R) - A}{p(R|A)}\right)\frac{\partial^2 p(R|A)}{\partial A^2}\,, \quad \cdots \quad \left(\frac{\hat{a}(R) - A}{p(R|A)}\right)\frac{\partial^i p(R|A)}{\partial A^i}\,,$$

for $i \leq n$. When we evaluate the expectation for terms in this row we find

$$\begin{aligned} E\left[\left(\frac{\hat{a}(R) - A}{p(R|A)}\right)\frac{\partial^i p(R|A)}{\partial A^i}\right] &= \int (\hat{a}(R) - A)\frac{\partial^i p(R|A)}{\partial A^i} dR \\ &= (\hat{a}(R) - A)\frac{\partial^{i-1} p(R|A)}{\partial A^{i-1}}\bigg|_{-\infty}^{\infty} - \int (-1)\frac{\partial^{i-1} p(R|A)}{\partial A^{i-1}} dR \\ &= \frac{\partial^{i-1}}{\partial A^{i-1}}\int p(R|A)dR = \frac{\partial^{i-1}}{\partial A^{i-1}}(1) = 0\,. \end{aligned}$$

Thus Λ_z has its $(1,1)$ element given by $E[\hat{a}(R) - A)^2]$, its $(1,2)$ element given by 1, and all other elements in the first row are zero. The first column of Λ_z is similar. Denote the lower-right corner of λ_z as $\tilde{J}$ and note that $\tilde{J}$ has elements for $i \geq 1$ and $j \geq 1$ given by

$$E\left[\frac{1}{p(R|A)}\frac{\partial^i p(R|A)}{\partial A^i} \times \frac{1}{p(R|A)}\frac{\partial^j p(R|A)}{\partial A^j}\right] = \int \frac{1}{p(R|A)}\frac{\partial^i p(R|A)}{\partial A^i}\frac{\partial^j p(R|A)}{\partial A^j} dR\,.$$

We know that Λ_z is nonnegative definite. Expressing the fact that $|\Lambda_z| \geq 0$ by expanding the determinant about the first row gives

$$\sigma_\varepsilon^2 |\tilde{J}| - \text{cofactor}(J_{11}) \geq 0\,.$$

Solving for σ_ε^2 we get

$$\sigma_\varepsilon^2 \geq \frac{\text{cofactor}(J_{11})}{|\tilde{J}|} = J^{11} \,,$$

the same procedure as in the proof of the Cramer-Rao bound proof for non-random variables.

Part (3): When $N = 1$ then zz^T looks like

$$zz^T = \begin{bmatrix} \sigma_\varepsilon^2 & 1 \\ 1 & \left(\frac{1}{p(R|A)} \frac{\partial p(R|A)}{\partial A} \right)^2 \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 & 1 \\ 1 & \left(\frac{\partial \ln(p(R|A))}{\partial A} \right)^2 \end{bmatrix} ,$$

so $\tilde{J} = E\left[\left(\frac{\partial \ln(p(R|A))}{\partial A} \right)^2 \right]$ and thus

$$\tilde{J}^{11} = \frac{1}{E\left[\left(\frac{\partial \ln(p(R|A))}{\partial A} \right)^2 \right]} ,$$

which is the Cramer-Rao inequality.

Part (4): Informally, as N increases then $\tilde{J}$ increases (or its norm "increases") and thus $\tilde{J}^{11}$ decreases providing a tighter bound.

Problem 2.4.27 (some vector derivatives)

Part (1): We have defined

$$\nabla_x = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} , \tag{116}$$

and we want to show

$$\nabla_x(A^T B) = (\nabla_x A^T)B + (\nabla_x B^T)A \,.$$

Lets compute the left-hand-side of this expression. We find that $A^T B$ is given by

$$A^T B = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} = A_1 B_1 + A_2 B_2 + \cdots + A_n B_n \,,$$

a scalar. With this we see that

$$\nabla_x(A^T B) = \begin{bmatrix} \frac{\partial}{\partial x_1}(A_1B_1 + A_2B_2 + \cdots + A_nB_n) \\ \frac{\partial}{\partial x_2}(A_1B_1 + A_2B_2 + \cdots + A_nB_n) \\ \vdots \\ \frac{\partial}{\partial x_n}(A_1B_1 + A_2B_2 + \cdots + A_nB_n) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial A_1}{\partial x_1}B_1 + \frac{\partial A_2}{\partial x_1}B_2 + \cdots + \frac{\partial A_n}{\partial x_1}B_n \\ \frac{\partial A_1}{\partial x_2}B_1 + \frac{\partial A_2}{\partial x_2}B_2 + \cdots + \frac{\partial A_n}{\partial x_2}B_n \\ \vdots \\ \frac{\partial A_1}{\partial x_n}B_1 + \frac{\partial A_2}{\partial x_n}B_2 + \cdots + \frac{\partial A_n}{\partial x_n}B_n \end{bmatrix} + \begin{bmatrix} A_1\frac{\partial B_1}{\partial x_1} + A_2\frac{\partial B_2}{\partial x_1} + \cdots + A_n\frac{\partial B_n}{\partial x_1} \\ A_1\frac{\partial B_1}{\partial x_2} + A_2\frac{\partial B_2}{\partial x_2} + \cdots + A_n\frac{\partial B_n}{\partial x_2} \\ \vdots \\ A_1\frac{\partial B_1}{\partial x_n} + A_2\frac{\partial B_2}{\partial x_n} + \cdots + A_n\frac{\partial B_n}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial A_1}{\partial x_1} & \frac{\partial A_2}{\partial x_1} & \cdots & \frac{\partial A_n}{\partial x_1} \\ \frac{\partial A_1}{\partial x_2} & \frac{\partial A_2}{\partial x_2} & \cdots & \frac{\partial A_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial A_1}{\partial x_n} & \frac{\partial A_2}{\partial x_n} & \cdots & \frac{\partial A_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} + \begin{bmatrix} \frac{\partial B_1}{\partial x_1} & \frac{\partial B_2}{\partial x_1} & \cdots & \frac{\partial B_n}{\partial x_1} \\ \frac{\partial B_1}{\partial x_2} & \frac{\partial B_2}{\partial x_2} & \cdots & \frac{\partial B_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial B_1}{\partial x_n} & \frac{\partial B_2}{\partial x_n} & \cdots & \frac{\partial B_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}.$$

If we define $\nabla_x A^T$ as the matrix

$$\begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} = \begin{bmatrix} \frac{\partial A_1}{\partial x_1} & \frac{\partial A_2}{\partial x_1} & \cdots & \frac{\partial A_n}{\partial x_1} \\ \frac{\partial A_1}{\partial x_2} & \frac{\partial A_2}{\partial x_2} & \cdots & \frac{\partial A_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial A_1}{\partial x_n} & \frac{\partial A_2}{\partial x_n} & \cdots & \frac{\partial A_n}{\partial x_n} \end{bmatrix},$$

the we can write the above as $(\nabla_x A^T)B + (\nabla_x B^T)A$ thus our derivative relationship is then

$$\nabla_x(A^T B) = (\nabla_x A^T)B + (\nabla_x B^T)A\,. \tag{117}$$

Part (2): Use Part (1) from this problem with $B = x$ where we would get

$$\nabla_x(A^T x) = (\nabla_x A^T)x + (\nabla_x x^T)A\,.$$

If A does not depend on x then $\nabla_x A^T = 0$ and we have $\nabla_x(A^T x) = (\nabla_x x^T)A$. We now need to evaluate $\nabla_x x^T$ where we find

$$\nabla_x x^T = \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \cdots & \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_n}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_1}{\partial x_n} & \frac{\partial x_2}{\partial x_n} & \cdots & \frac{\partial x_n}{\partial x_n} \end{bmatrix} = I\,,$$

the identity matrix. Thus we have $\nabla_x(A^T x) = A$. **Part (3):** Note from the dimensions given $x^T C$ is a $(1 \times n) \cdot (n \times m) = 1 \times m$ sized matrix. We can write the product of $x^T C$ as

$$x^T C = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^n x_j c_{j1} & \sum_{j=1}^n x_j c_{j2} & \cdots & \sum_{j=1}^n x_j c_{jm} \end{bmatrix}$$

We now compute $\nabla_x(x^T C)$ to get

$$\nabla_x(x^T C) = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{j=1}^n x_j c_{j2}\right) & \frac{\partial}{\partial x_1}\left(\sum_{j=1}^n x_j c_{j2}\right) & \cdots & \frac{\partial}{\partial x_1}\left(\sum_{j=1}^n x_j c_{jm}\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{j=1}^n x_j c_{j2}\right) & \frac{\partial}{\partial x_2}\left(\sum_{j=1}^n x_j c_{j2}\right) & \cdots & \frac{\partial}{\partial x_2}\left(\sum_{j=1}^n x_j c_{jm}\right) \\ \vdots & & \ddots & \vdots \\ \frac{\partial}{\partial x_n}\left(\sum_{j=1}^n x_j c_{j2}\right) & \frac{\partial}{\partial x_n}\left(\sum_{j=1}^n x_j c_{j2}\right) & \cdots & \frac{\partial}{\partial x_n}\left(\sum_{j=1}^n x_j c_{jm}\right) \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nm} \end{bmatrix} = C\,.$$

Part (4): This result is a specialization of Part (3) where $C = I$.

Problem 2.4.28 (some vector derivatives of quadratic forms)

Part (1): Write our derivative as

$$\nabla_x Q = \nabla_x(A^T(x)\Lambda A(x)) = \nabla_x(A^T(x)\Lambda^{1/2}\Lambda^{1/2}A(x)) = \nabla_x((\Lambda^{1/2}A(x))^T(\Lambda^{1/2}A(x)))\,.$$

Then using Problem 2.4.27 with the vectors A and B defined as $A \equiv \Lambda^{1/2}A(x)$ and $B \equiv \Lambda^{1/2}A(x)$ to get

$$\begin{aligned} \nabla_x Q &= (\nabla_x[\Lambda^{1/2}A(x)]^T)\Lambda^{1/2}A(x) + \nabla_x[\Lambda^{1/2}A(x)]\Lambda^{1/2}A(x) \\ &= 2(\nabla_x[\Lambda^{1/2}A(x)]^T)\Lambda^{1/2}A(x) = 2(\nabla_x A(x)^T)\Lambda^{1/2}\Lambda^{1/2}A(x) \\ &= 2(\nabla_x A(x)^T)\Lambda A(x)\,, \end{aligned} \tag{118}$$

as we were to show.

Part (2): If $A(x) = Bx$ then using Problem 2.4.27 where we computed $\nabla_x x^T$ we find

$$\nabla_x A^T = \nabla_x(x^T B^T) = \nabla_x(x^T)B^T = B^T\,.$$

Using this and the result from Part (1) we get

$$\nabla_x Q = 2B^T\Lambda Bx\,.$$

Part (3): If $Q = x^T\Lambda x$ then $A(x) = x$ so from Part (4) of Problem 2.4.27 we have $\nabla_x x^T = I$ which when we use with Equation 118 derived above we get

$$\nabla_x Q = 2\Lambda x\,,$$

as we were to show.

References

[1] S. Ross. *A First Course in Probability.* Macmillan, 3rd edition, 1988.