Problem Set 1, Number 10

Given the joint probability distribution for $\underline{x}_1$ and $\underline{x}_2$:

$$p(\underline{x}_1, \underline{x}_2) = \frac{1}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}\left(\underline{x}_1^T H_{11} \underline{x}_1 + 2\underline{x}_1^T H_{12}\underline{x}_2 + \underline{x}_2^T H_{22}\underline{x}_2\right)\right]$$

where $\underline{x}_1$ is an n-dimensional vector and $\underline{x}_2$ is an $(m-n)$-dimensional vec' and where

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{bmatrix} = P_x^{-1}$$

Prove that the integral of $p(\underline{x}_1, \underline{x}_2)\,d\underline{x}_1$ over the whole $\underline{x}_1$ space is equal to

$$\frac{(2\pi)^{n/2}}{(2\pi)^{m/2}[\det(P_x)]^{1/2}[\det(H_{11})]^{1/2}} \exp\left[-\frac{1}{2}\underline{x}_2^T(H_{22} - H_{12}^T H_{11}^{-1} H_{12})\underline{x}_2\right]$$

Find $\bar{\underline{x}}_1$ and G to express the argument of the exponential as suggested in the hint:

$$\underbrace{-\cancel{\tfrac{1}{2}}\left(\underline{x}_1^T H_{11}\underline{x}_1 + 2\underline{x}_1^T H_{12}\underline{x}_2 + \underline{x}_2^T H_{22}\underline{x}_2\right)}_{\text{Given}} = \underbrace{-\cancel{\tfrac{1}{2}}(\underline{x}_1 - \bar{\underline{x}}_1)^T H_{11}(\underline{x}_1 - \bar{\underline{x}}_1) - \cancel{\tfrac{1}{2}}\underline{x}_2^T G \underline{x}_2}_{\text{New form (suggested)}}$$

$$\cancel{\underline{x}_1^T H_{11}\underline{x}_1} + 2\underline{x}_1^T H_{12}\underline{x}_2 + \underline{x}_2^T H_{22}\underline{x}_2 = \cancel{\underline{x}_1^T H_{11}\underline{x}_1} - \underline{x}_1^T H_{11}\bar{\underline{x}}_1 - \bar{\underline{x}}_1^T H_{11}\underline{x}_1 + \bar{\underline{x}}_1^T H_{11}\bar{\underline{x}}_1 + \underline{x}_2^T G\underline{x}_2$$

Assume H_{11} and H_{22} are symmetric (because P_x^{-1} must be symmetric, because it is the inverse of a symmetric covariance matrix).

Then $(\underline{x}_1^T H_{11}\bar{\underline{x}}_1)^T = \bar{\underline{x}}_1^T H_{11}^T \underline{x}_1 = \bar{\underline{x}}_1^T H_{11}\underline{x}_1$. Since $\underline{x}_1^T H_{11}\bar{\underline{x}}_1$ and $\bar{\underline{x}}_1^T H_{11}\underline{x}_1$ are transposes of each other and both scalars, they are equal: $\underline{x}_1^T H_{11}\bar{\underline{x}}_1 = \bar{\underline{x}}_1^T H_{11}\underline{x}_1$.

So $\quad 2\underline{x}_1^T H_{12}\underline{x}_2 + 2\underline{x}_1^T H_{11}\bar{\underline{x}}_1 - \bar{\underline{x}}_1^T H_{11}\bar{\underline{x}}_1 = \underline{x}_2^T G\underline{x}_2 - \underline{x}_2^T H_{22}\underline{x}_2 = \underline{x}_2^T(G - H_{22})\underline{x}_2$

Let $H_{11}\bar{\underline{x}}_1 = -H_{12}\underline{x}_2 \Longrightarrow \boxed{\bar{\underline{x}}_1 = -H_{11}^{-1}H_{12}\underline{x}_2} \rightarrow \bar{\underline{x}}_1^T = -\underline{x}_2^T H_{12}^T H_{11}^{-1}$

Then $\quad 2\underline{x}_1^T H_{12}\underline{x}_2 + 2\underline{x}_1^T \cancel{H_{11}}(-\cancel{H_{11}^{-1}}H_{12}\underline{x}_2) - (-\underline{x}_2^T H_{12}^T \cancel{H_{11}^{-1}})\cancel{H_{11}}(-H_{11}^{-1}H_{12}\underline{x}_2) = \underline{x}_2^T(G - H_{22})\underline{x}_2$

$$\cancel{2\underline{x}_1^T H_{12}\underline{x}_2} - \cancel{2\underline{x}_1^T H_{12}\underline{x}_2} = \underline{x}_2^T(G - H_{22})\underline{x}_2 + \underline{x}_2^T H_{12}^T H_{11}^{-1}H_{12}\underline{x}_2$$

$$0 = \underline{x}_2^T(G + H_{12}^T H_{11}^{-1}H_{12} - H_{22})\underline{x}_2$$

$$\underline{x}_2^T G\underline{x}_2 = \underline{x}_2^T(H_{22} - H_{12}^T H_{11}^{-1}H_{12})\underline{x}_2 \Rightarrow \boxed{G = H_{22} - H_{12}^T H_{11}^{-1}H_{12}}$$

Using these definitions for $\bar{\underline{x}}_1$ and G,

$$p(\underline{x}_1, \underline{x}_2) = \frac{1}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}_1 - \bar{\underline{x}}_1)^T H_{11}(\underline{x}_1 - \bar{\underline{x}}_1) - \frac{1}{2}\underline{x}_2^T G\underline{x}_2\right]$$

Problem Set 1, Number 10, continued

Separating $\underline{x}_1$ and $\underline{x}_2$ parts of equation,

$$p(\underline{x}_1, \underline{x}_2) = \frac{1}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}_1 - \bar{\underline{x}}_1)^T H_{11} (\underline{x}_1 - \bar{\underline{x}}_1)\right] \exp\left[-\frac{1}{2}\underline{x}_2^T G \underline{x}_2\right]$$

So $$\int_{\mathbb{R}^n}\cdots\int p(\underline{x}_1, \underline{x}_2)\, d\underline{x}_1 = \int_{\mathbb{R}^n}\cdots\int \frac{1}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}_1 - \bar{\underline{x}}_1)^T H_{11}(\underline{x}_1 - \bar{\underline{x}}_1)\right]\exp\left[-\frac{1}{2}\underline{x}_2^T G \underline{x}_2\right] d\underline{x}_1$$

$$= \frac{1}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}\underline{x}_2^T G \underline{x}_2\right] \int_{\mathbb{R}^n}\cdots\int \exp\left[-\frac{1}{2}(\underline{x}_1 - \bar{\underline{x}}_1)^T H_{11}(\underline{x}_1 - \bar{\underline{x}}_1)\right] d\underline{x}_1$$

Now make the quantity inside the integral look like the pdf of a joint Gaussian random vector $\underline{x}_1$, with distribution $\mathcal{N} \sim (\bar{\underline{x}}_1, H_{11}^{-1})$:

(see the standard form for this on Bar-Shalom p.51)

Multiply and divide by: $(2\pi)^{n/2}[\det(H_{11}^{-1})]^{1/2}$:

$$\int_{\mathbb{R}^n}\cdots\int p(\underline{x}_1, \underline{x}_2)\, d\underline{x}_1 = \frac{1}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}\underline{x}_2^T G \underline{x}_2\right] (2\pi)^{n/2}[\det(H_{11}^{-1})]^{1/2} \cdot \int_{\mathbb{R}^n}\cdots\int \frac{1}{(2\pi)^{n/2}[\det(H_{11}^{-1})]^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}_1 - \bar{\underline{x}}_1)^T H_{11}(\underline{x}_1 - \bar{\underline{x}}_1)\right] d\underline{x}_1$$

The quantity inside the integral now looks like the pdf of a $\sim\mathcal{N}(\bar{\underline{x}}_1, H_{11}^{-1})$ random vector, integrated over the entire vector. By the results of Problem Set 1, Problem 9, the quantity thus integrates to 1. Therefore, we have:

$$\int_{\mathbb{R}^n}\cdots\int p(\underline{x}_1, \underline{x}_2)\, d\underline{x}_1 = \frac{(2\pi)^{n/2}[\det(H_{11}^{-1})]^{1/2}}{(2\pi)^{m/2}[\det(P_x)]^{1/2}} \exp\left[-\frac{1}{2}\underline{x}_2^T G \underline{x}_2\right]$$

$$[\det(H_{11}^{-1})]^{1/2} = [1/\det(H_{11})]^{1/2} = [\det(H_{11})]^{-1/2} \quad \checkmark$$

Substituting this, and the expression for $G = H_{22} - H_{12}^T H_{11}^{-1} H_{12}$, we have:

$$\boxed{\int_{\mathbb{R}^n}\cdots\int p(\underline{x}_1, \underline{x}_2)\, d\underline{x}_1 = \frac{(2\pi)^{n/2}}{(2\pi)^{m/2}[\det(P_x)]^{1/2}[\det(H_{11})]^{1/2}} \exp\left[-\frac{1}{2}\underline{x}_2^T (H_{22} - H_{12}^T H_{11}^{-1} H_{12}) \underline{x}_2\right]} \quad \checkmark$$

Solution

The original hypothesis test can be stated as follows:

$$H_0 : \theta = 0$$
$$H_1 : \theta \neq 0$$

with an observation vector $\boldsymbol{z} \in \mathbb{R}^n$ given by

$$\boldsymbol{z} = \theta \mathbf{1} + \boldsymbol{w}$$

where $\mathbf{1} = [1, 1, \ldots, 1]^T \in \mathbb{R}^n$, $\mathbb{E}[\boldsymbol{w}] = \mathbf{0}$, and $\mathbb{E}\left[\boldsymbol{w}\boldsymbol{w}^T\right] = P$.

This is a composite hypothesis test because, under H_1, θ can take on a range of values.

To eliminate the pesky cross-correlation terms in P, one can transform the measurement model as

$$\boldsymbol{z}_a = R_a^{-T}\boldsymbol{z}, \quad \boldsymbol{h}_a = R_a^{-T}\mathbf{1}, \quad \boldsymbol{w}_a = R_a^{-T}\boldsymbol{w}$$

where $R_a^T R_a = P$ is the Cholesky factorization of the symmetric, positive definite matrix P. The transformed measurement model is

$$\boldsymbol{z}_a = \theta \boldsymbol{h}_a + \boldsymbol{w}_a$$

This transformation does not change the nature of the test, but the fact that $\mathbb{E}\left[\boldsymbol{w}_a\boldsymbol{w}_a^T\right] = I$ simplifies subsequent analysis.

Under H_1 we do not know whether θ is positive or negative; we only know it is different from zero. Let us denote the unknown value of θ under H_1 as θ_1. Now the problem can be re-stated as

$$H_0 : \boldsymbol{z}_a \sim \mathcal{N}(\mathbf{0}, I)$$
$$H_1 : \boldsymbol{z}_a \sim \mathcal{N}(\theta_1 \boldsymbol{h}_a, I)$$

This is a special case of the general Gaussian problem that was discussed in lecture. The optimal test has the form

$$\boldsymbol{z}_a^T\boldsymbol{z}_a - (\boldsymbol{z}_a - \theta_1\boldsymbol{h}_a)^T(\boldsymbol{z}_a - \theta_1\boldsymbol{h}_a) \underset{H_0}{\overset{H_1}{\gtrless}} \nu$$

which can be simplified to

$$\theta_1 \boldsymbol{h}_a^T \boldsymbol{z}_a \underset{H_0}{\overset{H_1}{\gtrless}} \nu'$$

which may be written

$$\theta_1 q \underset{H_0}{\overset{H_1}{\gtrless}} \nu'$$

where

$$q \triangleq \boldsymbol{h}_a^T \boldsymbol{z}_a$$

is the sufficient statistic. If $\theta_1 > 0$, the test reduces to

$$q \underset{H_0}{\overset{H_1}{\gtrless}} \nu^+$$

The superscript $^+$ on ν indicates our assumption that $\theta_1 > 0$. If $\theta_1 < 0$, the test reduces to

$$q \underset{H_1}{\overset{H_0}{\gtrless}} \nu^-$$

We don't know which of these two cases holds because we don't know the value of θ_1, but we do know that the test is governed by the sufficient statistic q, which, as a linear combination of jointly Gaussian elements, is itself a Gaussian random variable whose density under H_0 is easily shown to be

$$p(q|H_0) = \mathcal{N}(q; 0, \sigma_0^2)$$

where $\sigma_0^2 = \boldsymbol{h}_a^T \boldsymbol{h}_a$. Without knowledge of θ_1, we cannot find $p(q|H_1)$, the density of q under H_1, but we can summarize what we do know by writing

$$\begin{aligned} H_0 &: q \sim \mathcal{N}(0, \sigma_0^2) \\ H_1 &: q \nsim \mathcal{N}(0, \sigma_0^2) \end{aligned}$$

Thus, our test boils down to determining whether the sufficient statistic q appears to be drawn from $\mathcal{N}(0, \sigma_0^2)$ or not. If q appears too extreme—toward either tail of the Gaussian distribution—then we should reject H_0. We can realize this test by taking the absolute value of q:

$$|q| \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

To satisfy a false alarm probability $P_F = \alpha$, we set

$$\lambda = \texttt{norminv}(1\text{ - }\tfrac{\alpha}{2}, 0, \sigma_0)$$

Note that, without knowledge of $p(q|H_1)$, we cannot determine the probability of detection P_D associated with $P_F = \alpha$.

For $n = 2$, $\alpha = 0.01$, and

$$P = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

we have

$$\boldsymbol{h}_a = R_a^{-T}\mathbf{1} = \begin{bmatrix} 1 & 0 \\ -0.5774 & 1.1547 \end{bmatrix}\mathbf{1} = \begin{bmatrix} 1 \\ 0.5774 \end{bmatrix}$$

from which $\sigma_0^2 = \boldsymbol{h}_a^T\boldsymbol{h}_a = 1.333$. The test of the statistic $q = \boldsymbol{h}_a^T\boldsymbol{z}_a$ boils down to

$$|q| \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

with $\lambda = \texttt{norminv}(1\ \text{-}\ \frac{0.01}{2}, 0, \sigma_0) = 2.9743$.

Consider a hypothesis test about the music coming through a loudspeaker. Is the music a Brahms cello sonata (soft and sonorous) or Stravinsky's Rite of Spring second movement (loud and percussive)? For some reason, you can't hear the loudspeaker but you can measure the deflection of its diaphragm. Let the random variable z represent a measurement of the diaphragm deflection. The Brahms-vs-Stravinsky hypothesis can then be stated as

$$H_0: \; z \sim \mathcal{N}(0, \sigma_0^2)$$
$$H_1: \; z \sim \mathcal{N}(0, \sigma_1^2)$$

with $\sigma_0 < \sigma_1$.

(a) Find a simple detection statistic for this hypothesis test.
(b) Find the value of the detection threshold for $P_F = 0.03$.
(c) Find the probability of detection P_D associated with this threshold for the following ratios of σ_1/σ_0: 4, 6, 8, 10. Do you think the P_D is satisfactory when $\sigma_1/\sigma_0 = 4$?
(d) Suppose we decide to take more measurements to improve P_D. We space these measurements far enough apart in time so that they are independent of one another. Find a simple detection statistic that makes use of N measurements $z_1, z_2, \ldots, z_N$ of the diaphragm deflection.
(e) Assuming $\sigma_1/\sigma_0 = 4$, how large must N be to ensure $P_D > 0.99$ for $P_F = 0.001$?

4. **[10 points]** Problem Set 2, Number 5 except make two slight changes: (1) in part (b) use $P_F = 0.01$ instead of $P_F = 0.03$, and (2) in part (e) assume $\sigma_1/\sigma_0 = 6$.

We can recognize this problem as a special case of the general Gaussian problem discussed in lecture:

$$H_0: \; z \sim N(\mu_0, P_0) = p(z|H_0)$$
$$H_1: \; z \sim N(\mu_1, P_1) = p(z|H_1)$$

$$\Lambda(z) = \left[\frac{p(z|H_1)}{p(z|H_0)}\right] = \frac{|P_0|^{1/2} \exp\left[-\frac{1}{2}(z-\mu_1)^T P_1^{-1}(z-\mu_1)\right]}{|P_1|^{1/2} \exp\left[-\frac{1}{2}(z-\mu_0)^T P_0^{-1}(z-\mu_0)\right]} \underset{H_0}{\overset{H_1}{\gtrless}} \nu$$

$$\Lambda'(z) = \log[\Lambda(z)] \underset{H_0}{\overset{H_1}{\gtrless}} \log(\nu) = \nu'$$

$$\Lambda''(z) = (z-\mu_0)^T P_0^{-1}(z-\mu_0) - (z-\mu_1)^T P_1^{-1}(z-\mu_1)$$

$$\underset{H_0}{\overset{H_1}{\gtrless}} 2\nu' + \log|P_1| - \log|P_0| = \nu''$$

This will be our starting point.

(a) This is the scalar case with $\mu_0 = \mu_1 = 0$ and $P_0 = \sigma_0^2$, $P_1 = \sigma_1^2$.

Thus, $$\Lambda''(z) = z^2\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_1^2\sigma_0^2}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \nu''$$

and since we know that $\sigma_1^2 > \sigma_0^2$, we can be sure the quantity in parentheses is positive. Thus, without changing the inequalities, we can write:

$$\boxed{\gamma = \Lambda'''(z) = z^2 \underset{H_0}{\overset{H_1}{\gtrless}} \nu''' = \lambda}$$

(b) For $P_F = 0.01$, we need to solve for λ:

$$P_F = 0.01 = \int_{\lambda'''}^{\infty} p(\gamma \mid H_0)\, d\gamma$$

Note that $\gamma = z^2$, $z \sim N(0, \sigma_i^2)$ $i = 0,1$

$$= \sigma_i^2 x^2, \quad x \sim N(0,1)$$

Thus, $x^2 \sim \chi_1^2$ and $P[\lambda < \gamma] = P[\lambda/\sigma_i^2 < x^2]$

$$\Rightarrow \quad P_F = 0.01 = \int_{\lambda/\sigma_0^2}^{\infty} p(x^2)\, d(x^2)$$

(Use σ_0^2 because P_F is calculated assuming H_0)

Solve in Matlab as: $\lambda = \sigma_0^2 \cdot \text{chi2inv}(1-0.01, 1)$

$$\boxed{\Rightarrow \quad \lambda = \sigma_0^2 \cdot 6.6349}$$

(c) $P_D = \int_{\lambda}^{\infty} p(\gamma|H_1)\, d\gamma = \int_{l}^{\infty} p(x^2)\, d(x^2)$

where $l = \lambda/\sigma_1^2 = \frac{\sigma_0^2}{\sigma_1^2} \cdot 6.6349$

Solve in Matlab by:

$P_D = 1 - \text{chi2cdf}(\, l,\ 1\ )$

σ_1/σ_0	P_D
4	0.5196
6	0.6677
8	0.74747
10	0.79673

$P_D = 0.5196$ does not seem satisfactory

(this is the value of P_D when $\sigma_1/\sigma_0 = 4$).

(d) This is the vector case of the original generalized Gaussian problem with $\underline{\mu}_0 = \underline{\mu}_1 = 0$ and $P_0 = \sigma_0^2 I$, $P_1 = \sigma_1^2 I$.

Thus, we have:

$$\Lambda''(\underline{z}) = \underline{z}^T\underline{z} \left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_1^2 \sigma_0^2} \right) \underset{H_0}{\overset{H_1}{\gtrless}} \nu''$$

As before, we can divide out the quantity in parentheses since $\sigma_1^2 > \sigma_0^2$ makes it positive. Thus, our detection statistic is

$$\boxed{\gamma = \Lambda'''(\underline{z}) = \underline{z}^T\underline{z} \underset{H_0}{\overset{H_1}{\gtrless}} \nu''' = \lambda}$$

(e) Our new statistic can be expressed as

$$\gamma = \sigma_i^2 \sum_{j=1}^{N} x_j^2 \quad , \quad x_j \sim N(0,1) \quad , \quad i = 0,1$$

Thus, $\gamma = \sigma_i^2 \xi$ where $\xi \sim \chi_N^2$

For $P_F = 0.001$ our threshold λ must be:

$$\lambda = \sigma_0^2 \cdot \text{chi2inv}(1 - P_F, N)$$

And the corresponding P_D for various values of N and for $\sigma_1 / \sigma_0 = 6$ is:

N	P_D
5	0.989
6	0.996

Thus, we would require $N \geq 6$ to satisfy $P_D > 0.99$ for $P_F = 0.001$.

You've been given a coin that when tossed shows heads with probability θ and tails with probability $1-\theta$. You wish to estimate θ by experimentation. You toss the coin N times, yielding N independent observations $y_1, y_2, ..., y_N$, where

$$y_k = \begin{cases} 1 & \text{if } k\text{th toss shows heads} \\ 0 & \text{if } k\text{th toss shows tails} \end{cases}$$

The observations are packaged together in a vector $\boldsymbol{y} = [y_1, y_2, ..., y_N]^T$.

1. Suppose $N = 6$ and $\boldsymbol{y}^* = [0, 0, 1, 1, 0, 1]^T$ is the outcome of your experiment. Fill out the table below with the probability of the outcome $\boldsymbol{y}^*$ given θ.

θ	$P(\boldsymbol{y} = \boldsymbol{y}^*\|\theta)$
0.2	0.004096
0.4	0.013824
0.8	0.004096
1.0	0

2. Let $n_1 = \sum_{k=1}^{N} y_k$ be the number of heads and $n_0 = N - n_1$ be the number of tails. Write an expression for $P(\boldsymbol{y}|\theta)$ in terms of θ, n_1, and n_0.

   One could either write out the formula directly, based on intuition:

$$P(\boldsymbol{y}|\theta) = \theta^{n_1}(1-\theta)^{n_0}$$

   Or one could derive it from first principles:

$$\begin{aligned} P(y_k = 1|\theta) =& \theta \\ P(y_k = 0|\theta) =& 1 - \theta \end{aligned}$$

   This can be written concisely as

$$P(y_k|\theta) = \theta^{y_k}(1-\theta)^{1-y_k}$$

   Given the independence of the coin tosses,

$$\begin{aligned} P(\boldsymbol{y}|\theta) =& \prod_{k=1}^{N} P(y_k|\theta) \\ =& \prod_{k=1}^{N} \theta^{y_k}(1-\theta)^{1-y_k} \\ =& \theta^{\sum_{k=1}^{N} y_k}(1-\theta)^{\sum_{k=1}^{N} 1-y_k} \\ =& \theta^{n_1}(1-\theta)^{n_0} \end{aligned}$$

3. Derive $\hat{\theta}_{\text{ML}}$, the maximum likelihood estimate of θ, from the expression in part 2.

$$\begin{aligned} \hat{\theta}_{\text{ML}} &= \arg\max_{\theta} P(\boldsymbol{y}|\theta) \\ &= \arg\max_{\theta} \log P(\boldsymbol{y}|\theta) \\ &= \arg\max_{\theta} l(\theta) \end{aligned}$$

where

$$l(\theta) \triangleq \log P(\boldsymbol{y}|\theta) = n_1 \log \theta + n_0 \log(1-\theta)$$

The maximizing θ must satisfy

$$\frac{dl(\theta)}{d\theta} = \frac{n_1}{\theta} - \frac{n_0}{1-\theta} = 0$$

which implies

$$\hat{\theta}_{\mathrm{ML}} = \frac{n_1}{n_0 + n_1} = \frac{n_1}{N}$$

Bar Shalom 2-1

1) $p(x|z) = \dfrac{p(z|x)\,p(x)}{p(z)}$ by Bayes's rule

$$p(z|x) = p_w(z-x) = N(z; x, \sigma^2)$$

$$p(z) = \int_{-\infty}^{\infty} p(z|x)\,p(x)\,dx = \int_{-\infty}^{\infty} N(z; x, \sigma^2)\left[p_1\,\delta(x-1) + p_2\,\delta(x-2)\right] dx$$

$$= p_1 N(z; 1, \sigma^2) + p_2 N(z; 2, \sigma^2)$$

$$p(x) = \sum_{i=1}^{2} p_i\,\delta(x-i)$$

$$\Longrightarrow p(x|z) = \frac{N(z; x, \sigma^2)\left[p_1\,\delta(x-1) + p_2\,\delta(x-2)\right]}{p_1 N(z; 1, \sigma^2) + p_2 N(z; 2, \sigma^2)}$$

2) Map estimate: $\hat{x}_{MAP}(z) = \arg\max_x \left[p(z|x)\,p(x)\right]$

$p(z|x)p(x)$ is only nonzero at two values of x, namely, $x=1$ and $x=2$. But at these values $p(z|x)p(x)$ is infinite. To determine which value of x to choose, let us recognize that the value of $p(z|x)p(x)$ can be taken for decision-making purposes to be

$$p(z|x)\,p(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{x-\epsilon/2}^{x+\epsilon/2} p(z|\xi)\,p(\xi)\,dx$$

This leads to $\hat{x}_{MAP}(z) = \arg\max_x \left[p(z|x)\,p(x)\right]$

$$\Longrightarrow \hat{x}_{MAP}(z) = \begin{cases} 1 & \text{if } N(z; 1, \sigma^2)\,p_1 > N(z; 2, \sigma^2)\,p_2 \\ 2 & \text{if } N(z; 1, \sigma^2)\,p_1 < N(z; 2, \sigma^2)\,p_2 \\ 1 \text{ or } 2 & \text{if } N(z; 1, \sigma^2)\,p_1 = N(z; 2, \sigma^2)\,p_2 \end{cases}$$

$$= \arg\max_i\, p_i N(z; i, \sigma^2)$$

The MSE of $\hat{x}_{MAP}(z)$ conditioned on z is:

$$E[(x-\hat{x}_{MAP}(z))^2|z] = E[x^2|z] - 2E[x|z]\hat{x}_{MAP}(z) + \hat{x}^2_{MAP}(z)$$

where $E[x^2|z] = \int_{-\infty}^{\infty} x^2 p(x|z) = \frac{1}{p(z)}\int_{-\infty}^{\infty} x^2 N(z;x,\sigma^2)\left[p_1\delta(x-1)+p_2\delta(x-2)\right]dx$

$$= \frac{1}{p(z)}\left[N(z;1,\sigma^2)p_1 + 4N(z;2,\sigma^2)p_2\right] \quad \left(\begin{array}{c}p(z) \text{ defined} \\ \text{on prev. page}\end{array}\right)$$

and $E[x|z] = \int_{-\infty}^{\infty} x\,p(x|z)\,dx = \frac{1}{p(z)}\left[N(z;1,\sigma^2)p_1 + 2N(z;2,\sigma^2)p_2\right]$

3) $\hat{x}_{MMSE}(z) = E[x|z] =$

MSE: $E[(x-\hat{x}_{MMSE}(z))^2|z] = E\{(x-E[x|z])^2|z\} = Var(x|z)$

$$= E[x^2|z] - \hat{x}^2_{MMSE}(z)$$

but $E[x^2|z]$ was found in part (2).

4) Cases and results:

Case	p_1, σ, z	$\hat{x}_{MAP}$	σ^2_{MAP}	$\hat{x}_{MMSE}$	σ^2_{MMSE}
A	0.5, 1, 1.5	2 or 1	0.5	1.5	0.25
B	0.5, 1, 3	2	0.182	1.82	0.15
C	0.3, 1, 1.5	2	0.3	1.7	0.21
D	0.5, 0.1, 1.8	2	0	2	0

5) As one would expect, the MSE for $\hat{x}_{MMSE}$ is always less than or equal to the MSE for $\hat{x}_{MAP}$. However, the MAP estimator is best suited for this problem because maximizing $p(x|z)$ is more proper when x is discrete.

Problem 2-14 in Bar Shalom (with the clarification offered in the problem set).

Do this problem in two parts: (1) by assuming a Gaussian distribution for the error in the estimate of the variance, as described in the book's hint, and then (2) by assuming that the original error measurements from which $\hat{\sigma}^2$ is derived are Gaussian distributed and using this fact to derive the *actual* distribution for the error in the estimate of the variance (which is not exactly Gaussian). Compare the results.

Let e_i be the measurement error of the i^{th} trial.
From problem statement: $E[e_i] = 0$, $E[e_i^2] = \sigma^2$

I. Gaussian approximation for $p(\hat{\sigma}^2)$

In this case, the actual distribution of the e_i is unimportant. We only need to know that $e_i \sim (0, \sigma^2)$

Maximum Likelihood estimate of σ^2 would be

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \qquad \text{(sample variance when mean is } \emptyset \text{ and known)}$$

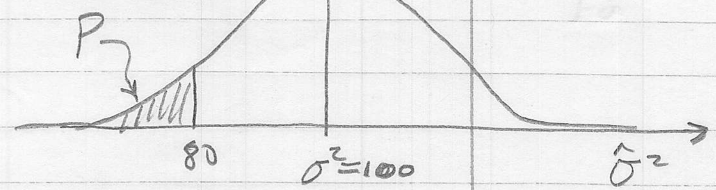
Then $E[\hat{\sigma}^2] = \frac{1}{N}\sum_{i=1}^{N} E[e_i^2] = \sigma^2$ (unbiased)

and $E[(\hat{\sigma}^2 - \sigma^2)^2] = \dfrac{2\sigma^4}{N}$ (Bar Shalom 2.6.3-3)

Now we assume that $\hat{\sigma}^2$ is Gaussian distributed:

$$p(\hat{\sigma}^2) = \mathcal{N}\left(\sigma^2, \frac{2\sigma^4}{N}\right)$$

For borderline $\sigma^2 = 100$ case:



$$P = \text{normcdf}\left(80, 100, \sqrt{\frac{2 \cdot 100^2}{N}}\right) \leq 0.05$$

is satisfied for $N \geq 136$.

II. Actual distribution of $\hat{\sigma}^2$

In this case, we assume $e_i \sim N(0,\sigma^2)$ and compute the actual distribution of $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \quad , \quad e_i \sim N(0,\sigma^2) \quad (e_i \text{ independent})$$

$$= \frac{\sigma^2}{N}\sum_{i=1}^{N} z_i^2 \quad , \quad z_i \sim N(0,1) \quad (z_i \text{ independent})$$

Let $q \triangleq \frac{N\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{N} z_i^2$. Then $q \sim \chi_N^2$.

Consider equivalent events:

$$P[\hat{\sigma}^2 \leq 80 \mid \sigma^2 \geq 100] = P[q \leq \tfrac{80 \cdot N}{\sigma^2} \mid \sigma^2 \geq 100]$$

$$= \int_{-\infty}^{\frac{80\cdot N}{\sigma^2}} p(q)\, dq$$

We can solve for N such that this quantity is less than 0.05 for borderline case of $\sigma^2 = 100$:

$$\text{chi2cdf}(80 \cdot N/100, N) \leq 0.05$$

Satisfied for $N \geq 124$.

<u>Comparison</u>

We need fewer trials if we use the actual distribution of $\hat{\sigma}^2$ based on knowledge that $e_i \sim N(0,\sigma^2)$. Thus, in this case, approximating $p(\hat{\sigma}^2)$ as Gaussian when it is not exactly so leads to a conservative (unnecessarily safe) value for the required N.

Solution

Consider the problem of estimating an unknown frequency parameter ω from complex measurements

$$z_k = \exp(j\omega kT) + n_k, \quad k = 0, 1, ..., N-1$$

where $j = \sqrt{-1}$ and where $n_k = a_k + jb_k$ is a sequence of independent, identically-distributed complex zero-mean Gaussian noise samples with variance σ^2:

$$a_k, b_k \sim \mathcal{N}(0, \sigma^2), \quad \mathbb{E}[a_k a_j] = \sigma^2 \delta_{kj}, \quad \mathbb{E}[b_k b_j] = \sigma^2 \delta_{kj}, \quad \mathbb{E}[a_k b_j] = 0 \ \ \forall k, j$$

Here, δ_{kj} is the Kronecker delta (equal to unity for $k = j$ and otherwise zero).

Suppose the N measurements are stacked as a complex vector $\boldsymbol{z} = \boldsymbol{x} + j\boldsymbol{y}$, where $\boldsymbol{z} = [z_0, z_1, ..., z_{N-1}]^T$, and where $\boldsymbol{x} = [x_0, x_1, ..., x_{N-1}]^T$ and $\boldsymbol{y} = [y_0, y_1, ..., y_{N-1}]^T$ are the real and imaginary components of $\boldsymbol{z}$.

We wish to find the CRLB for an ML estimate of ω. Since $\boldsymbol{z}$ is Gaussian, the likelihood function is simply

$$\Lambda(\omega) \triangleq p(\boldsymbol{z}|\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^N \exp\left(-\frac{1}{2\sigma^2}[\boldsymbol{z} - \boldsymbol{\mu}_z]^H [\boldsymbol{z} - \boldsymbol{\mu}_z]\right)$$

where $\boldsymbol{\mu}_z = \mathbb{E}[\boldsymbol{z}]$ and where $\boldsymbol{x}^H$ represents the conjugate transpose of the vector $\boldsymbol{x}$. Expressing the inner product as a summation, and recognizing that $\exp(j\omega kT) = \cos(\omega kT) + j\sin(\omega kT)$, $\Lambda(\omega)$ may be written

$$\Lambda(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^N \exp\left(-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}[x_k - \cos(\omega kT)]^2 + [y_k - \sin(\omega kT)]^2\right)$$

Taking the natural log, we have

$$\log\Lambda(\omega) = C - \frac{1}{2\sigma^2}\sum_{k=0}^{N-1}[x_k - \cos(\omega kT)]^2 + [y_k - \sin(\omega kT)]^2$$

where C is a constant that does not depend on ω. Now taking the derivative with respect to ω, we have

$$H_\omega \triangleq \frac{d\log\Lambda(\omega)}{d\omega} = -\frac{T}{\sigma^2}\sum_{k=0}^{N-1} k a_k \sin(\omega kT) - k b_k \cos(\omega kT)$$

Recognizing that the cross-terms involving a_k and b_k vanish when calculating $\mathbb{E}[H_\omega^2]$, we have

$$J \triangleq \mathbb{E}\left[H_\omega^2\right] = \frac{T^2}{\sigma^4}\sum_{k=0}^{N-1} k^2\sigma^2\left[\sin^2(\omega kT) + \cos^2(\omega kT)\right]$$

which reduces to $J = QT^2/\sigma^2$, where

$$Q = \sum_{k=0}^{N-1} k^2 = \frac{N(N-1)(2N-1)}{6}$$

Thus, the CRLB is

$$\text{CRLB} \triangleq J^{-1} = \frac{\sigma^2}{QT^2}$$

Solution

The `sribls.m` function can be implemented as

```
% sribls : Square-root-information batch least squares routine.  Given the
%          linear measurement model
%
%          zprime = Hprime*x + w,  w ~ N(0,R),
%
%          sribls returns xhat, the least squares (Maximum Likelihood)
%          estimate of x, and Rotilde, the associated square root information
%          matrix.
%
% INPUTS
%
% Hprime ----- nz-by-nx measurement sensitivity matrix.
%
% zprime ----- nz-by-1 measurement vector.
%
% R ---------- nz-by-nz measurement noise covariance matrix.
%
%
% OUTPUTS
%
% xhat ------- nx-by-1 Maximum Likelihood estimate of x.
%
% Rotilde ---- nx-by-nx square root information matrix, where P =
%              inv(Rotilde'*Rotilde) is the estimation error covariance matrix.
%
%+------------------------------------------------------------------------------+
% References:
%
%
% Author:
%+==============================================================================+

 [nz,nx] = size(Hprime);
 Rc = inv(chol(R)');
 H = Rc*Hprime;
 z = Rc*zprime;
 [Qt,Rt] = qr(H);
 zt = Qt'*z;
 Rotilde = Rt(1:nx,:);
 zotilde = zt(1:nx);
 xhat = inv(Rotilde)*zotilde;
```

The solution produced by the normal equations is

```
xhatNE =
         -21.0323602904203
          16.1530201120834
          23.1137222239126
```

Note that, because the columns of H are nearly linearly dependent, the solution is highly sensitive to the values in `z`; thus your results may differ.

The solution produced by the square-root information least-squares solver is

```
xhatSRI =
         -21.1179643159471
          16.4092391227941
          23.9465687541611
```

with associated SRI matrix

```
Rotilde =
         -1441001.65958447           1427641.9845944          -6546030.89155809
                         0         -1821432.06559233           2575894.09838084
                         0                         0         -0.328600410898828
```

Note the small value of the (3,3) element of `Rotilde` relative to the other diagonal elements. This is indicative of a poorly conditioned matrix. But the condition number of `Rotilde` is only on the order of the square root of the condition number of the $H^T R^{-1} H$ matrix used in the normal equations.

Note also that both approaches lead to poor estimates for the given measurements $\boldsymbol{z}$. The true value of $\boldsymbol{x}$ for this problem was

```
x =
        -18.8598151735416
         15.3927145351384
         23.2277775745102
```