ECE531 Homework Assignment Number 9 Solution

Due by 8:50pm on Wednesday 13-Apr-2011

Make sure your reasoning and work are clear to receive full credit for each problem.

1. 4 points. Kay I: Problem 6.5

   Solution: I will use $y_n = x[n]$ for my observations. The mean of one sample from the lognormal pdf given in the problem can be computed as

$$\mathrm{E}[Y_k] = \int_0^\infty t \frac{1}{\sqrt{2\pi}t} \exp\left[-\frac{1}{2}(\ln t - \theta)^2\right] dt$$

and if we do a variable substitution $v = \ln t$ then we have

$$\begin{aligned}
\mathrm{E}[Y_k] &= \int_{-\infty}^\infty e^v \frac{1}{\sqrt{2\pi}e^v} \exp\left[-\frac{1}{2}(v-\theta)^2\right] e^v\, dv \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(v-\theta)^2\right] e^v\, dv \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(v^2 - 2v\theta - 2v + \theta^2)\right] dv \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(v^2 - 2v(\theta+1) + (\theta+1)^2 - 2\theta - 1)\right] dv \\
&= \exp\left[-\frac{1}{2}(-2\theta - 1)\right] \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(v - (\theta+1))^2\right] dv
\end{aligned}$$

where we "completed the square" in the second to last step. The term inside the integral is the pdf of a $\mathcal{N}(\theta+1, 1)$ random variable, so the integral will be one. Hence, we get the desired result

$$\begin{aligned}
\mathrm{E}[Y_k] &= \exp\left[-\frac{1}{2}(-2\theta-1)\right]\cdot 1 \\
&= \exp(\theta + 1/2).
\end{aligned}$$

The mean of the observations is clearly not linear in the unknown parameter, i.e. there is no $H \neq 0$ such that $\mathrm{E}[Y] = H\theta$. So we can not directly apply the BLUE approach to this problem.

If we transform the observations so that $Z_k = \ln Y_k$, we can compute the marginal pdf of the transformed observations as

$$\begin{aligned}
p_{Z_k}(z\,;\,\theta) &= \frac{p_{Y_k}(y = e^z\,;\,\theta)}{|dz/dy|_{y=e^z}} \\
&= \frac{\frac{1}{\sqrt{2\pi}e^z}\exp\left[-\frac{1}{2}(z-\theta)^2\right]}{e^{-z}} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z-\theta)^2\right]
\end{aligned}$$

where $dz/dy = \ln(y)/dy = 1/y$. The pdf of Z_k is clearly $\mathcal{N}(\theta, 1)$ and the Z_k are i.i.d. This problem is now suitable for BLUE because $\mathrm{E}[Z] = H\theta$ with $H = [1, \ldots, 1]^\top$. We can also see that $C = I$. Hence, from the results developed in lecture, we have

$$\begin{aligned}\hat{\theta}_{\mathsf{BLUE}}(z) &= (H^\top C^{-1} H)^{-1} H^\top C^{-1} z \\ &= \frac{1}{N} \sum_{k=0}^{N-1}\end{aligned}$$

which is clearly the sample mean of the transformed observations. Also, it should be clear that BLUE=MVU given then transformed observations since the transformed observations are Gaussian.

The point here is that, in some cases, it may be a good idea to perform a nonlinear operation on the observations prior to BLUE to make the problem suitable for BLUE. Although the overall estimator is no longer linear on the original observations, the combination of the nonlinear operator and the BLUE might be easier to derive (and implement) than the full-blown MVU estimator.

2. 4 points. Kay I: Problem 6.9

Solution: I will use $y_n = x[n]$ for my observations and $\theta = A$ for the unknown parameter. The observations are given as

$$Y = \begin{bmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_{N-1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ \cos(2\pi f_1) \\ \vdots \\ \cos(2\pi f_1(N-1)) \end{bmatrix}}_{H} \theta + \underbrace{\begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{N-1} \end{bmatrix}}_{W}$$

You should recognize that this problem fits the form of the linear Gaussian model with $\mathrm{E}[Y] = H\theta$ and $C = \sigma^2 I$. Hence,

$$\begin{aligned} \hat{\theta}_{\mathsf{BLUE}}(y) &= (H^\top C^{-1} H)^{-1} H^\top C^{-1} y \\ &= (H^\top H)^{-1} H^\top y \\ &= \frac{\sum_{k=0}^{N-1} \cos(2\pi f_1 k) y_k}{\sum_{k=0}^{N-1} \cos^2(2\pi f_1 k)}. \end{aligned}$$

Interpretation: The amplitude BLUE is an inner product of the received signal with $s_1(t)$, i.e. a matched filter with the expected signal when a one is sent, scaled by the total energy in $s_1(t)$. In the absence of noise, when a zero is sent, the numerator will be zero. When a one is sent, the numerator will be $\sum_{k=0}^{N-1} \theta \cos^2(2\pi f_1 k)$, which will yield $\hat{\theta}_{\mathsf{BLUE}}(y) = \theta$ after the denominator is canceled. Hence, this estimator seems reasonable.

The variance of this estimator can be calculated as

$$\begin{aligned} \mathrm{var}[\hat{\theta}_{\mathsf{BLUE}}(Y)] &= (H^\top C^{-1} H)^{-1} \\ &= \frac{\sigma^2}{\sum_{k=0}^{N-1} \cos^2(2\pi f_1 k)}. \end{aligned}$$

To determine value of f_1 that maximizes the denominator (and, thus, minimizes the variance), we can write

$$\begin{aligned} \frac{d}{df_1} \sum_{k=0}^{N-1} \cos^2(2\pi f_1 k) &= \frac{d}{df_1} \frac{1}{2} \sum_{k=0}^{N-1} (1 + \cos(4\pi f_1 k)) \\ &= -\frac{1}{2} \sum_{k=0}^{N-1} 4\pi k \sin(4\pi f_1 k)). \end{aligned}$$

If we set this to zero, there are lots of solutions, but an obvious one is when $f_1 = 0$. It should be clear that this maximizes the denominator since $\sum_{k=0}^{N-1} \cos^2(2\pi f_1 k) \leq N$ and we can make this inequality into an equality when $f_1 = 0$. Hence, the best transmit frequency over the allowed range $0 \leq f_1 < 1/2$ is $f_1 = 0$, in which case we have

$$\mathrm{var}[\hat{\theta}_{\mathsf{BLUE}}(Y)] = \frac{\sigma^2}{N}.$$

The reason this is the best transmit frequency is that the energy of the signal is maximized, which makes it easier to distinguish the presence of the signal from the absence of the signal. We probably wouldn't do this in a wireless communication system, however, because antennas tend to be inefficient at propagating DC.

3. 4 points. Kay I: Problem 6.16

Solution: This problem is about *model mismatch* in that it considers the impact of using the wrong covariance matrix to compute your BLUE. This can (and often does) happen in practice when your knowledge of the statistics of the observations is not perfect. We denote the actual covariance matrix as C and the covariance matrix that you used to compute your BLUE as $\hat{C}$. We assume the linear model in theorem 6.1 with $Y = H\theta + W$ where W is zero mean. We also a scalar parameter $\theta = A$ with $H = [1, 1]^\top$ as in example 6.2.

First, let's check if the erroneous BLUE is biased:

$$\begin{aligned}
\mathrm{E}\left\{\tilde{\theta}_{\mathsf{BLUE}}(Y)\right\} &= \mathrm{E}\left\{(H^\top \hat{C}^{-1} H)^{-1} H^\top \hat{C}^{-1} Y\right\} \\
&= \underbrace{(H^\top \hat{C}^{-1} H)^{-1} H^\top \hat{C}^{-1}}_{\tilde{A}} \mathrm{E}\{Y\} \\
&= (H^\top \hat{C}^{-1} H)^{-1} H^\top \hat{C}^{-1} H\theta \\
&= \theta.
\end{aligned}$$

So even though we used the wrong covariance matrix, the erronious BLUE is unbiased irrespective of H and $\hat{C}$.

Now let's compute the variance of the erroneous BLUE with $\tilde{\theta}_{\mathsf{BLUE}}(Y) = \tilde{A}Y$:

$$\begin{aligned}
\mathrm{var}[\tilde{\theta}_{\mathsf{BLUE}}(Y)] &= \mathrm{E}\left\{(\tilde{\theta}_{\mathsf{BLUE}}(Y) - \theta)^2\right\} \\
&= \mathrm{E}\left\{(\tilde{A}Y - \theta)^2\right\} \\
&= \mathrm{E}\left\{(\tilde{A}Y - \tilde{A}H\theta)^2\right\} \\
&= \tilde{A}\mathrm{E}\left\{(Y - H\theta)^2\right\}\tilde{A}^\top \\
&= \tilde{A}C\tilde{A}^\top \\
&= (H^\top \hat{C}^{-1} H)^{-1} H^\top \hat{C}^{-1} C \hat{C}^{-1} H (H^\top \hat{C}^{-1} H)^{-1}.
\end{aligned}$$

Now we plug in for $\hat{C} = [1, 0; 0, \alpha]$, $C = I$, and $H = [1, 1]^\top$ to write

$$\begin{aligned}
\mathrm{cov}[\tilde{\theta}_{\mathsf{BLUE}}(Y)] &= (1 + 1/\alpha)^{-1}(1 + 1/\alpha^2)(1 + 1/\alpha)^{-1} \\
&= \frac{1 + 1/\alpha^2}{(1 + 1/\alpha)^2} \\
&= \frac{\alpha^2 + 1}{(\alpha + 1)^2}.
\end{aligned}$$

When $\alpha = 1$, there is no model mismatch and $\mathrm{var}[\tilde{\theta}_{\mathsf{BLUE}}(Y)] = 1/2$, which agrees with the result in example 6.2 when $N = 2$.

When $\alpha \to 0$, our assumed model differs from the real model in that we are now assuming there is no noise in the second sample. So our estimator, if it is doing the "right thing" will throw out the first sample and only use the second one. We can expect the actual variance of the estimator to get worse in this case because we shouldn't be throwing away that second sample. In fact, when $\alpha \to 0$, we have $\mathrm{var}[\tilde{\theta}_{\mathsf{BLUE}}(Y)] = 1$, which would be the variance of the correct BLUE if we only had one sample. So this makes sense.

When $\alpha \to \infty$, our assumed model differs from the real model in that we are now assuming there is massive noise in the second sample. So our estimator, if it is doing the "right thing" will throw out the second sample and only use the first one. We can compute $\lim_{\alpha\to\infty} \mathrm{var}[\tilde{\theta}_{\mathsf{BLUE}}(Y)] = 1$, which is the same as when $\alpha \to 0$. So this makes sense too.

4. 4 points. Kay I: Problem 7.3

Solution to part (a): The joint pdf of the observations is

$$p_Y(y\,;\,\theta) = \frac{1}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}(y-1\theta)^\top(y-1\theta)\right\},$$

To find the MLE of θ, we compute

$$\begin{aligned}
\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) &= -\frac{1}{2}\frac{\partial}{\partial\theta}(y-1\theta)^\top(y-1\theta) \\
&= -\frac{1}{2}\frac{\partial}{\partial\theta}y^\top y - 2\theta 1^\top y + \theta^2 1^\top 1 \\
&= -\frac{1}{2}\left(-21^\top y + 2\theta 1^\top 1\right) \\
&= \sum_k y_k - N\theta.
\end{aligned}$$

The last step is to set this equal to zero, substitute $\theta = \hat{\theta}_{\mathsf{ml}}(y)$, and solve for $\hat{\theta}_{\mathsf{ml}}(y)$. If we do this, we get

$$\hat{\theta}_{\mathsf{ml}}(y) = \frac{1}{N}\sum_k y_k = \bar{y}.$$

In other words, the MLE is the sample mean, which is the same as the MVU and BLUE estimators.

To confirm this maximizes the likelihood function, note that $p_Y(y\,;\,\theta)$ is maximized when

$$(y-1\theta)^\top(y-1\theta) = \sum_k (y_k-\theta)^2 = \sum_k y_k^2 - 2y_k\theta + \theta^2$$

is minimized. This is quadratic in θ, taking the derivative with respect to θ, setting the result equal to zero, and solving for θ will give the global minimum (which maximizes $p_Y(y\,;\,\theta)$).

Solution to part (b): The joint pdf of the observations is

$$p_Y(y\,;\,\theta) = \begin{cases} \theta^N \exp\left\{-\theta\sum_k y_k\right\} & \min y_k > 0 \\ 0 & \text{otherwise.} \end{cases}$$

To find the MLE of θ, we compute (assuming $\min y_k > 0$)

$$\begin{aligned}
\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) &= \frac{\partial}{\partial\theta}(N\ln\theta) - \frac{\partial}{\partial\theta}\left(\theta\sum_k y_k\right) \\
&= \frac{N}{\theta} - \sum_k y_k
\end{aligned}$$

The last step is to set this equal to zero, substitute $\theta = \hat{\theta}_{\mathsf{ml}}(y)$, and solve for $\hat{\theta}_{\mathsf{ml}}(y)$. If we do this, we get

$$\hat{\theta}_{\mathsf{ml}}(y) = \frac{N}{\sum_k y_k} = \frac{1}{\bar{y}}.$$

To confirm this maximizes the likelihood function, note that $p_Y(y\,;\,\theta)$ is maximized when $\ln p_Y(y\,;\,\theta)$ is maximized. We can compute

$$\frac{\partial^2}{\partial\theta^2}\ln p_Y(y\,;\,\theta) = -\frac{N}{\theta^2}$$

which is strictly less than zero for all $\theta > 0$. Hence $\ln p_Y(y\,;\,\theta)$ is a concave function and $\hat{\theta}_{\mathsf{ml}}(y)$ must be the maximum.

5. 4 points. Kay I: Problem 7.8

Solution: Letting the unknown parameter $\theta = p$, we can write the joint pdf (really pmf) of the i.i.d. observations as

$$\begin{aligned} p_Y(y\,;\,\theta) &= \prod_{k=0}^{N-1} \theta^{y_k}(1-\theta)^{(1-y_k)} \\ &= \theta^{\sum_k y_k}(1-\theta)^{\sum_k (1-y_k)} \\ &= \theta^{N\bar{y}}(1-\theta)^{N-N\bar{y}} \end{aligned}$$

To find the MLE of θ, we compute

$$\begin{aligned} \frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) &= \frac{\partial}{\partial\theta}\left\{N\bar{y}\ln\theta + (N-N\bar{y})\ln(1-\theta)\right\} \\ &= \frac{N\bar{y}}{\theta} + \frac{N-N\bar{y}}{1-\theta}\cdot(-1) \\ &= \frac{N\bar{y}}{\theta} - \frac{N-N\bar{y}}{1-\theta} \end{aligned}$$

The last step is to set this equal to zero, substitute $\theta = \hat{\theta}_{\mathsf{ml}}(y)$, and solve for $\hat{\theta}_{\mathsf{ml}}(y)$. If we do this, we get

$$\begin{aligned} & N\bar{y}(1-\hat{\theta}_{\mathsf{ml}}(y)) - (N-N\bar{y})\hat{\theta}_{\mathsf{ml}}(y) = 0 \\ \Leftrightarrow\quad & N\bar{y} - N\bar{y}\hat{\theta}_{\mathsf{ml}}(y) - N\hat{\theta}_{\mathsf{ml}}(y) + N\bar{y}\hat{\theta}_{\mathsf{ml}}(y) = 0 \\ \Leftrightarrow\quad & \hat{\theta}_{\mathsf{ml}}(y) = \bar{y}. \end{aligned}$$

In other words, the MLE of θ is the sample mean.

To confirm this achieves the maximum of the likelihood function, we can compute the second derivative

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}\ln p_Y(y\,;\,\theta) &= \frac{\partial}{\partial\theta}\left\{\frac{N\bar{y}}{\theta} - \frac{N-N\bar{y}}{1-\theta}\right\} \\ &= \frac{-N\bar{y}}{\theta^2} - \frac{N-N\bar{y}}{(1-\theta)^2}. \end{aligned}$$

Substitute in $\theta = \bar{y}$ to get

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}\ln p_Y(y\,;\,\theta) &= \frac{-N\bar{y}}{\bar{y}^2} - \frac{N-N\bar{y}}{(1-\bar{y})^2} \\ &= -\left(\frac{N}{\bar{y}} + \frac{N}{1-\bar{y}}\right) \end{aligned}$$

which is strictly less than zero for all $0 < \theta < 1$. Hence $\ln p_Y(y\,;\,\theta)$ is a concave function and $\hat{\theta}_{\mathsf{ml}}(y) = \bar{y}$ must be the maximum.

6. 5 points. Kay I: Problem 7.13. Please use Matlab (or Octave) for your computer simulation and include your code in your homework submission.

Solution: My Matlab code is below.

```
% ECE531 Spring 2011
% HW9 Problem 6
% DRB 07-Apr-2011
% ---------------------------------------------------------------
% USER PARAMETERS
% ---------------------------------------------------------------
sig2 = 0.1;                    % sigma^2 (variance of AWGN)
A = 1;                         % value of unknown parameter
N = 50;                        % number of observations
M = 5000;                      % number of realizations for averaging
% ---------------------------------------------------------------

theta_hat_mle = zeros(1,M); % allocate space for estimates
Abar = ones(1,N)/N;            % sample mean estimator

for m=1:M,
    y = A + sqrt(sig2)*randn(N,1);   % generate vector observation
    theta_hat_mle(m) = Abar*y;
end

[n,x] = hist(theta_hat_mle,[0.8:0.01:1.2]);
plot(x,n/(M*0.01),x,normpdf(x,A,sqrt(sig2/N)));
legend('experimental','theoretical');
axis([0.8 1.2 0 10]);
grid on
xlabel('y');
ylabel('pdf of MLE for parameter A');
title(['M = ',num2str(M)]);
text(0.81,9,['E[MLE] = ',num2str(mean(theta_hat_mle)), ' (experimental)']);
text(0.81,8,['var[MLE] = ',num2str(var(theta_hat_mle)), ' (experimental)']);
```

The results for $M = 1000$ and $M = 5000$ are shown in Figures 1 and 2, respectively. What we see here is that, even at $N = 50$, the asymptotic properties of the MLE are valid. We know the sample mean is unbiased and achieves the CRLB, so it isn't surprising to see the mean and variance both matching the theoretical predictions of $\mathrm{E}_\theta[\hat{\theta}_{\mathsf{ml}}(Y)] = 1$ and $\mathrm{var}_\theta[\hat{\theta}_{\mathsf{ml}}(Y)] = \frac{\sigma^2}{N} = \frac{0.1}{50} = 0.002$. These pictures also show that the distribution of the estimation errors is also Guassian, with zero mean and variance given by the CRLB, which also isn't really very surprising because the sample mean in this problem is Gaussian distributed.

So, while none of these results are particularly surprising, this problem gives a template for how one would go about numerically evaluating the performance of an estimator.

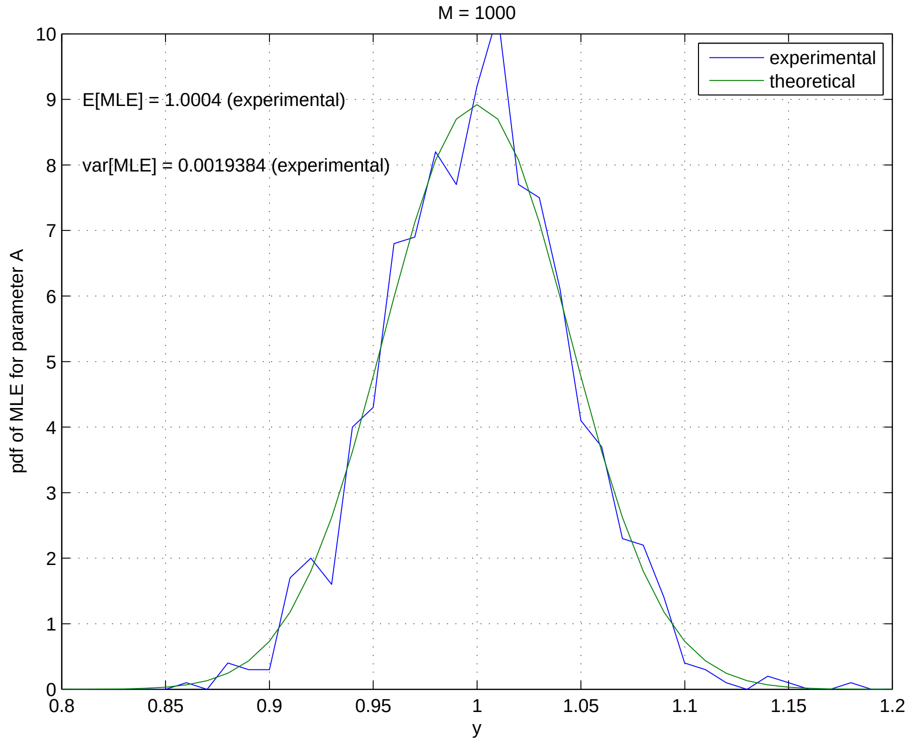

Figure 1: An example at $M = 1000$ realizations.

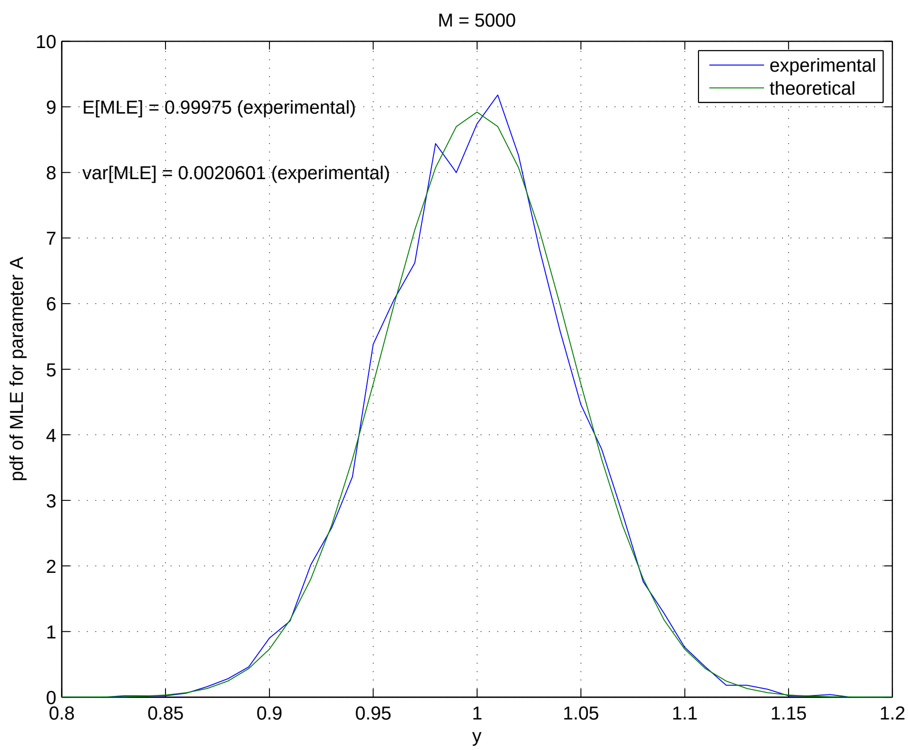

Figure 2: An example at $M = 5000$ realizations.