INDIAN INSTITUTE OF TECHNOLOGY MADRAS
Department of Chemical Engineering

CH5115 Parameter and State Estimation
Quiz 2 Solutions

1

Given the individual probability mass function of $y[k]$,

$$f(y,\theta) = \begin{cases} \theta^y(1-\theta)^{1-y} & \text{if } y \in \{0,1\}, 0 < \theta < 1 \\ 0 & \text{Otherwise} \end{cases}$$

The likelihood for the obtained dataset is expressed as

$$f(\mathbf{y}|\theta) = \theta^{\left(\sum_{k=1}^{N} y[k]\right)} (1-\theta)^{\left(N - \sum_{k=1}^{N} y[k]\right)} \tag{1}$$

Claim: $T(y[k]) = \sum_{k=1}^{N} y[k]$ is a sufficient statistic for θ.

Proof: Firstly, it is to be checked whether $f(\mathbf{y}|\theta)$ can be written as a product of $g(T(y[k]),\theta)$ and $h(\mathbf{y})$.

$$f(\mathbf{y}|\theta) = \theta^{\left(\sum_{k=0}^{N} y[k]\right)} (1-\theta)^{\left(N - \sum_{k=0}^{N} y[k]\right)} \tag{2}$$

$$\implies (1) \times \underbrace{\theta^{\left(\sum_{k=0}^{N} y[k]\right)} (1-\theta)^{\left(N - \sum_{k=0}^{N} y[k]\right)}}_{g(T(y[k]),\theta)} \tag{3}$$

Since the likelihood in this scenario is composable in the afforementioned manner according to Fisher-Neaman factorisation theorem, it can be concluded that $T(y[k]) = \sum_{k=1}^{N} y[k]$ is a sufficient statistics.

From the mass function, $E(y[k])$ can be calculated as θ. Further, $E(T(y[k]))$ can also be obtained as $N\theta$. From the corollary of R-B theorem, the realizable, MVUE of θ is $\boxed{\hat{\theta} = \frac{1}{N}\sum_{k=1}^{N} y[k]}$.

2

The log-likelihood of the associated distribution is

$$L = lnf(\mathbf{y}|\theta) = \sum_{k=1}^{N} y[k]\ln\theta + \left(N - \sum_{k=1}^{N} y[k]\right)(\ln(1-\theta))$$

$$\frac{\partial L}{\partial\theta} = S = \frac{\sum\limits_{k=1}^{N} y[k]}{\theta} - \frac{N - \sum\limits_{k=1}^{N} y[k]}{1-\theta}$$

$$E\left(-\frac{d^2L}{d\theta^2}\right) = \frac{\sum\limits_{k=1}^{N} E(y[k])}{\theta^2} + \frac{E(N - \sum\limits_{k=1}^{N} y[k])}{(1-\theta)^2}$$

$$E\left(-\frac{d^2L}{d\theta^2}\right) = \frac{N}{\theta} + \frac{N}{1-\theta}$$

$$I(\theta) = \frac{N}{\theta(1-\theta)}$$

$$\sigma_{\theta_{CR}} = \frac{\theta(1-\theta)}{N}$$

The variance for the MVUE estimate for θ is given by

$$\hat{\theta} = \frac{1}{N}\sum_{k=1}^{N} y[k]$$

$$\sigma_{\hat{\theta}} = \frac{\theta(1-\theta)}{N}$$

The MVUE ($\hat{\theta}_{CR}$) according to the Cramer-Rao theorem is obtained as

$$\hat{\theta}_{CR} = \frac{S}{I} + \theta \tag{4}$$

$$\hat{\theta}_{CR} = \frac{(1-\theta)\sum\limits_{k=1}^{N} y[k] - \theta(N - \sum\limits_{k=1}^{N} y[k])}{N} + \theta \tag{5}$$

$$\hat{\theta}_{CR} = \theta_{MVUE} = \frac{1}{N}\sum_{k=1}^{N} y[k] \tag{6}$$

Since i) $T(y)$ is complete and ii) the loss function *i.e* the likelihood in this case is concave (this can be verified in the steps for calculation of $I(\theta)$) the solution of θ_{MVUE} is unique which is why both the solutions are one and same.

3

Given, $y[k] = \frac{1}{\beta}u[k] + e[k]$ where, $e[k] \sim \mathcal{N}(0, \sigma_e^2)$.

Calculation of α^*: The likelihood of $y[k]$ can be written as

$$L(\mathbf{y}|\alpha) = (2\pi\sigma_e)^{-\frac{N}{2}} \exp\left(- \frac{\sum_{k=1}^{N} (y[k] - \alpha u[k])^2}{2\sigma_e^2} \right) \tag{7}$$

$$l(\mathbf{y}|\alpha) = log(L(\mathbf{y}|\alpha)) = -\frac{N}{2} log(2\pi\sigma_e) - \frac{\sum_{k=1}^{N} (y[k] - \alpha u[k])^2}{2\sigma_e^2} \tag{8}$$

$$\left.\frac{dl(\mathbf{y}|\alpha)}{d\alpha}\right|_{\alpha=\alpha^*} = 0 = \sum_{k=1}^{N} (y[k] - \alpha u[k]) u[k] \big|_{\alpha=\alpha^*} \tag{9}$$

$$\alpha^* = \frac{\sum_{k=1}^{N} y[k]u[k]}{\sum_{k=1}^{N} u^2[k]} \tag{10}$$

$$I(\alpha) = \frac{\sum_{k=1}^{N} u^2[k]}{\sigma_e^2} \tag{11}$$

The expression for $\hat{\beta}_{MV}$ can be obtained as $\hat{\beta}_{MV} = \frac{1}{\alpha^*} = \frac{\sum_{k=1}^{N} u^2[k]}{\sum_{k=1}^{N} y[k]u[k]}$.

The following observations are made by re-parameterizing the model:

1. The point estimates of β obtained from MLE and by inverting $\hat{\alpha}$ are the same.
2. From the efficiency view-point, re-parameterization of α as $1/\beta$ do not result in the most efficient estimate of β. Even though $\hat{\alpha}$ is an efficient estimate, $\hat{\beta} = \frac{1}{\hat{\alpha}}$ is not an efficient estimate.
3. $\hat{\beta}$ is a biased estimate where as $\hat{\alpha}$ is unbiased estimate.
4. Both $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically efficient.

Therefore, it can be said that re-parameterizing of the model for finite sample size results in biased and inefficient estimate even though the point estimates of both the, original and re-parameterized, models are same.

MATLAB codes to check the consistency of $\hat{\beta}_{MV}$

```
N = 100000; % Sample size
ek = randn(N,1); % Generation of e[k]
k = 1:1:N;
uk = k>=0; % Generation of known deterministic input u[k]
beta = 1.5;
alpha = 1/beta;
yk = (1/beta).*uk' + ek; % Generation of y[k]
% Consistency
for i = 1:length(yk)
    beta_hat(i) = sum(uk(1:i).*uk(1:i))./sum(yk(1:i).*uk(1:i)');
    alpha_hat(i) = 1/beta_hat(i);
end
% Ploting the estimated beta and alpha with true values
```

```matlab
figure;subplot(2,1,1);plot(beta_hat)
hold on;plot(beta*ones(length(beta_hat),1))
title('Consistency of esitmated parameters')
ylabel('Estimated beta')
subplot(2,1,2);plot(alpha_hat)
hold on;plot(alpha*ones(length(alpha_hat),1))
ylabel('Estimated alpha')
xlabel('Sample size (N)');
```

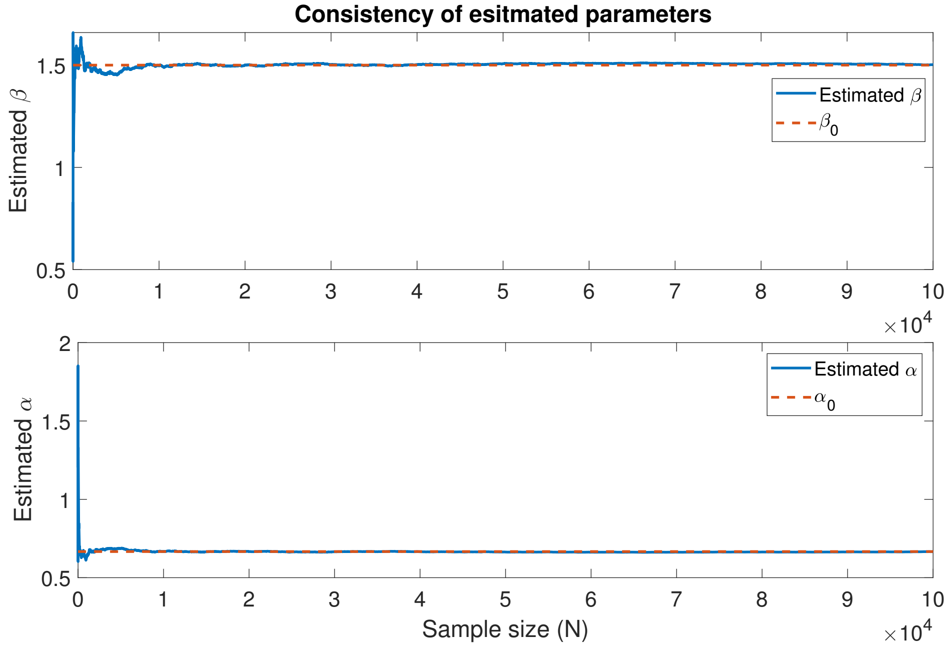

Figure 1: Consistency of $\hat{\beta}_{MV}$ and $\hat{\alpha}^*$

As observed from the Figure 1, $\hat{\beta}_{MV}$ converges with the increasing sample size, however, $\hat{\alpha}^*$ is not a consistent estimate of α.

4

Given the model $y[k] = f(x[k], \theta) + e[k]$ with known variance and $f(.,.)$. Since the particular form of the nonlinearity is known along with the knowledge of the optimal value of θ-$\hat{\theta}_N$ one approach to implement BLUE can be through Taylor series expansion of $f(x[k], \theta)$

Linearised model: From Taylor series expansion

$$f(x[k], \theta) = f(x[k], \hat{\theta}_N) + \left.\frac{\partial f(x[k], \theta)}{\partial \theta}\right|_{\hat{\theta}_N} \delta\theta + h.o.t \tag{12}$$

Since we are using Taylor series expansion, we assume that $f(x, \theta)$ is analytic with respect to θ.

Ignoring higher order terms, we get,

$$y[k] - \underbrace{f(x[k], \hat{\theta}_N)}_{\hat{y}[k]} = \underbrace{\frac{\partial f(x[k], \theta)}{\partial \theta} x[k]\bigg|_{\hat{\theta}_N}}_{g(x[k])} \delta\theta + e[k] \tag{13}$$

$$y^{'}[k] = y[k] - \hat{y}[k] = g(x[k]) + \varepsilon[k] \tag{14}$$

$$\mathbf{y}^{'} = \mathbf{g}(\mathbf{x})\delta\theta + \boldsymbol{\varepsilon} \tag{15}$$

$$\delta\theta_{\mathsf{BLUE}} = \frac{\mathbf{y}^{'\mathbf{T}}\mathbf{g}(\mathbf{x})}{||\mathbf{g}(\mathbf{x})||_2^2} \tag{16}$$

$$\hat{\theta}_{N+1} = \hat{\theta}_N + \delta\theta_{\mathsf{BLUE}}$$

$\delta\theta_{\mathsf{BLUE}}$ has to estimated recursively.