Md Abul Hayat
University of Arkansas

spring 2018
mahayat@uark.edu

Homework 2 : STAT 5113 (Statistical Inference)

Question A1: Consider a sample $X_1, \cdots, X_n$ from the Unif$(0, \theta)$ distribution. The MLE of θ is given

$$\hat{\theta} = \max_{1 \le i \le n} X_i$$

a. Find the cdf of $\hat{\theta}$, and use it to find the pdf of $\hat{\theta}$.
b. Derive an expression for the bias of $\hat{\theta}$.
c. Suppose the sample consisted of the following numbers:

6.83	8.85	1.46	7.81	5.89	7.20	6.60	11.98	10.55	8.12	7.59	4.50
10.51	0.18	8.62	9.58	6.89	2.30	7.55	4.12	10.67	1.08	0.53	9.47

Provide an estimate of θ and of the bias of the estimator.
d. Using the data provided above, give an estimate of the MSE of $\hat{\theta}$.

Answer:

(a)

$$F_{\hat{\theta}}(t) = P(\hat{\theta} \le t) = P(\max_{1 \le i \le n} X_i \le t) = P(X_1 \le t, X_2 \le t, \cdots, X_n \le t) = \prod_{i=1}^{n} P(X_i \le t) = \frac{t^n}{\theta^n}$$

$$f_{\hat{\theta}}(t) = \frac{F_{\hat{\theta}}(t)}{dt} = \frac{nt^{n-1}}{\theta^n}$$

(b)

$$E(\hat{\theta}) = \int_0^{\theta} t \cdot f_{\hat{\theta}}(t)dt = \int_0^{\theta} \frac{nt^n}{\theta^n} dt = \frac{n}{\theta^n} \int_0^{\theta} t^n dt = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta$$

So, $\hat{\theta}$ is a biased estimate of θ. We can define the unbiased estimate of θ as $\tilde{\theta} = \frac{n+1}{n}\hat{\theta}$ as in that case $E(\tilde{\theta}) = \theta$

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = \frac{n}{n+1}\theta - \theta = -\frac{\theta}{n+1}$$

(c) R code :

```
d = c(6.83,8.85,1.46,7.81,5.89,7.20,6.60,11.98,10.55,8.12,7.59,4.50,
      10.51,0.18,8.62,9.58,6.89,2.30,7.55,4.12,10.67,1.08,0.53,9.47)
n = length(d) # n = 24
theta_hat_biased = max(d)
theta_hat_unbiased = theta_hat_biased*(n+1)/n
bias = theta_hat_biased - theta_hat_unbiased
```

Output at console :

```
> theta_hat_biased
[1] 11.98
> theta_hat_unbiased
[1] 12.47917
> bias
[1] -0.4991667
```

(d) At first we want to find out the variance of $\hat{\theta}$ i.e. $\text{Var}(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2$

$$E(\hat{\theta}^2) = \int_0^{\theta} t^2 \frac{nt^{n-1}}{\theta^n} dt = \frac{n}{\theta^n} \int_0^{\theta} t^{n+1} dt = \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2$$

$$E(\hat{\theta}) = \frac{n}{n+1}\theta \quad \text{so,} \quad E(\hat{\theta})^2 = \frac{n^2}{(n+1)^2}\theta^2$$

$$\text{Var}(\hat{\theta}) = E(\hat{\theta}^2) - E(\hat{\theta})^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$$

from Bias-Variance trade-off we know that MSE is related as

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \Big[\text{Bias}(\hat{\theta})\Big]^2$$

$$\text{MSE}(\hat{\theta}) = \frac{n\theta^2}{(n+1)^2(n+2)} + \Big[-\frac{\theta}{n+1}\Big]^2 = \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+1)(n+2)}$$

Assuming $\theta = 12.47917$ and $n = 24$, we get $\text{MSE}(\hat{\theta}) = 0.47916$

Answer A2: (a) Here, $X_i \sim N(\mu_1, \sigma^2), \quad i = 1, \cdots, n_1$ and $Y_j \sim N(\mu_2, \sigma^2), \quad j = 1, \cdots, n_2$

$$\widehat{\mu_1} = \bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$$

$$\widehat{\mu_2} = \bar{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j$$

So, $\widehat{\mu_1}$ and $\widehat{\mu_2}$ are also distributed normally as they are linear combinations of Normal random variables.

$$E(\widehat{\mu_1}) = \frac{1}{n_1}\sum_{i=1}^{n_1} E(X_i) = \frac{\mu_1 n_1}{n_1} = \mu_1, \quad \text{Var}(\widehat{\mu_1}) = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\text{Var}(X_i) = \frac{n_1\sigma^2}{n_1^2} = \frac{\sigma^2}{n_1}$$

$$E(\widehat{\mu_2}) = \frac{1}{n_2}\sum_{j=1}^{n_2} E(Y_j) = \frac{\mu_2 n_2}{n_2} = \mu_2, \quad \text{Var}(\widehat{\mu_2}) = \frac{1}{n_2^2}\sum_{j=1}^{n_2}\text{Var}(Y_j) = \frac{n_2\sigma^2}{n_2^2} = \frac{\sigma^2}{n_2}$$

$$\widehat{\mu_1} \sim N(\mu_1, \frac{\sigma^2}{n_1}), \quad \widehat{\mu_2} \sim N(\mu_2, \frac{\sigma^2}{n_2})$$

$$\text{Also,} \sum_{i=1}^{n_1}\frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n_1-1} = \text{Gamma}\Big(\frac{n_1-1}{2}, \frac{1}{2}\Big)\text{(from class note)}$$

if $X \sim \text{Gamma}(\alpha, \beta)$ then $cX \sim \text{Gamma}(\alpha, \frac{\beta}{c})$, using this

$$S_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 = \frac{\sigma^2}{n_1}\sum_{i=1}^{n_1}\frac{(X_i - \bar{X})^2}{\sigma^2} \sim \text{Gamma}\Big(\frac{n_1-1}{2}, \frac{n_1}{2\sigma^2}\Big)$$

$$S_2^2 = \frac{1}{n_2}\sum_{j=1}^{n_2}(Y_j - \bar{Y})^2 = \frac{\sigma^2}{n_2}\sum_{j=1}^{n_2}\frac{(Y_j - \bar{Y})^2}{\sigma^2} \sim \text{Gamma}\Big(\frac{n_2-1}{2}, \frac{n_2}{2\sigma^2}\Big)$$

$$\widehat{\sigma^2} = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2} = \frac{n_1}{n_1+n_2}S_1^2 + \frac{n_2}{n_1+n_2}S_2^2$$

$$\frac{n_1}{n_1+n_2}S_1^2 \sim \text{Gamma}\Big(\frac{n_1-1}{2}, \frac{n_1}{2\sigma^2}\cdot\frac{n_1+n_2}{n_1}\Big) = \text{Gamma}\Big(\frac{n_1-1}{2}, \frac{n_1+n_2}{2\sigma^2}\Big)$$

$$\text{Similarly,} \quad \frac{n_2}{n_1+n_2}S_2^2 \sim \text{Gamma}\Big(\frac{n_2-1}{2}, \frac{n_1+n_2}{2\sigma^2}\Big)$$

if $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$ then $X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$ using this

$$\widehat{\sigma^2} \sim \text{Gamma}\Big(\frac{n_1+n_2-2}{2}, \frac{n_1+n_2}{2\sigma^2}\Big)$$

Now, we want to find out if the three random variables $\widehat{\mu_1}, \widehat{\mu_2}, \widehat{\sigma^2}$ are pairwise independent or not. Here, $\widehat{\mu_1}$ depends only on $\bar{X}$, $\widehat{\mu_2}$ depends only on $\bar{Y}$ and $\widehat{\sigma^2}$ depends on both $X_j - \bar{X}$ and $Y_j - \bar{Y}$ terms.

$$\text{Cov}(\bar{X}, X_j - \bar{X}) = \text{Cov}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1}X_i, X_j - \frac{1}{n_1}\sum_{i=1}^{n_1}X_i\Big) = \text{Cov}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1}X_i, X_j\Big) - \text{Cov}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1}X_i, \frac{1}{n_1}\sum_{i=1}^{n_1}X_i\Big)$$

$$\text{Cov}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1}X_i, X_j\Big) = \frac{1}{n_1}\sum_{i=1}^{n_1}\text{Cov}(X_i, X_j) = \frac{1}{n_1}\text{Cov}(X_j, X_j) = \frac{1}{n_1}\text{Var}(X_j) = \frac{\sigma^2}{n_1}$$

$$\text{Cov}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \frac{1}{n_1}\sum_{i=1}^{n_1} X_i\Big) = \text{Cov}(\bar{X}, \bar{X}) = \text{Var}(\bar{X}) = \frac{\sigma^2}{n_1}$$

Hence, $\text{Cov}(\bar{X}, X_j - \bar{X}) = 0$ similarly, we can prove that $\text{Cov}(\bar{Y}, Y_j - \bar{Y}) = 0$, So, in summary, $\widehat{\mu_1} \perp\!\!\!\perp \widehat{\sigma^2}, \widehat{\mu_2} \perp\!\!\!\perp \widehat{\sigma^2}$. Also, $\widehat{\mu_1} \perp\!\!\!\perp \widehat{\mu_2}$ because they are from two different Normal distributions. If the pdf of $\widehat{\mu_1}, \widehat{\mu_2}, \widehat{\sigma^2}$ is respectively $f_1(x), f_2(x)$ and $f_3(x)$ then their joint distribution is $f_1(x)f_2(x)f_3(x)$. Where, $f_1(x), f_2(x)$ and $f_3(x)$ can be found from the distributions of $\widehat{\mu_1}, \widehat{\mu_2}, \widehat{\sigma^2}$ written above.

(b)

$$\text{Bias}(\mu_1) = E(\widehat{\mu_1}) - \mu_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} E(X_i) - \mu_1 = \frac{\mu_1 n_1}{n_1} - \mu_1 = \mu_1 - \mu_1 = 0$$

$\widehat{\mu_1}$ is an **unbiased** estimate of μ_1

$$\text{Bias}(\mu_2) = E(\widehat{\mu_2}) - \mu_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} E(Y_j) - \mu_2 = \frac{\mu_2 n_2}{n_2} - \mu_2 = \mu_2 - \mu_2 = 0$$

$\widehat{\mu_2}$ is an **unbiased** estimate of μ_2

$$\text{if,} \quad X \sim \text{Gamma}(\alpha, \beta) \quad \text{then,} \quad E(X) = \frac{\alpha}{\beta}$$

$$\text{Bias}(\sigma^2) = E(\widehat{\sigma^2}) - \sigma^2 = \frac{n_1 + n_2 - 2}{2} \cdot \frac{2\sigma^2}{n_1 + n_2} - \sigma^2 = -\frac{2\sigma^2}{n_1 + n_2}$$

$\widehat{\sigma^2}$ is an **biased** estimate of σ^2
