

# 1-st homework

**1. (2 points)** Prove that for a sample  $X_1, \dots, X_n$  with continuous d.f.  $F(y)$ ,  $\forall t \in [0, 1]$ , the follow equality is correct:

$$P \left( \sup_y |F_n(y) - F(y)| > t \right) = P \left( \sup_{u \in [0,1]} |G_n(u) - u| > t \right),$$

where  $G_n(\cdot)$  is the empirical distribution function, based on uniform-distribtuted sample.

**2.(2 points)** Find the probability that  $F_n(s) < F_n(t)$ , where  $F_n(\cdot)$  is empirical distribution function of some sample  $X_1, \dots, X_n$ .

**3.(4 points)** a) Let  $X_1, \dots, X_n$  be i.i.d observations from  $\text{Unif}[a,b]$ . Is the statistic  $X_{(n)} - X_{(1)}$  unbiased (asymptotic or exact) and consistent estimator for  $b - a$

b) let  $X_1, \dots, X_n$  be i.i.d observations from  $\text{Unif}[-3a, a]$ . Is the statistic  $4X_{(n)} + X_{(1)}$  unbiased (asymptotic or exact) and consistent estimator for  $b - a$

**4.(4 points)** Let  $X_1, \dots, X_n$  be i.i.d. observations with finite the second moment and  $E(X)=a$ , which known. Are the following statistics unbiased and consistent:

$$(\bar{X})^2 - a^2, \quad (n-1)^{-1} \sum_i (X_i - a)^2$$

**5. (8 points)** Let  $X_1, \dots, X_n$  be a sample from a two-parameter exponential distribution with density

$$f_{\alpha, \beta}(y) = \begin{cases} \alpha e^{-\alpha(y-\beta)}, & y \geq \beta, \\ 0, & y < \beta. \end{cases}$$

a)(4 points) Let  $a=1$ , are the following statistics  $X_{(1)}$  and  $\bar{X} - 1$  are unbiased (asymptotic or exact) and consistent estimator for  $\beta$ .

b) (4 points) Let  $\beta = 0$  is the estimator for  $\alpha$   $\hat{\beta} = (\bar{X})^{-1}$  unbiased, if not find the bias. Is  $\hat{\beta}$  consistent?

**6\*(2 point per item.** a) Find the distribution of  $\frac{n \cdot X_{(k)}}{\theta}, \frac{n \cdot (\theta - X_{(n-k+1)})}{\theta}$ , if  $X_1, \dots, X_n \sim \text{Unif}[0, \theta]$ .

b) Is the empirical central moment  $\bar{\mu}_4 = \frac{1}{n} \sum_i (X_i - \bar{X})^4$  an unbiased estimator for  $\mu_4 = E(X - E(X))^4$ . If not, find the bias and transform it into an unbiased.

# Homework 1

1. (2 points) Prove that for a sample  $X_1, \dots, X_n$  with continuous d.f.  $F(y)$ ,  $\forall t \in [0, 1]$ , the following equality is correct:

$$P\left(\sup_y |F_n(y) - F(y)| > t\right) = P\left(\sup_{u \in [0,1]} |G_n(u) - u| > t\right),$$

where  $G_n(\cdot)$  is the empirical distribution function, based on uniform-distributed sample.

Pf:  $F_n(y) = \frac{1}{n} \sum \#\{X_i \leq y\}$ ,  $G_n(u) = \frac{1}{n} \sum \#\{F(X_i) \leq u\}$ .

thus  $F_n(y) = \frac{1}{n} \sum \#\{X_i \leq y\} = \frac{1}{n} \sum \#\{F(X_i) \leq F(y)\} = G_n(F(y))$ . (since d.f. not decreasing).

let  $u = F(y)$ . (by def of distribution function and  $u \in [0,1]$ ).

the LHS and RHS has the same expression.

2. (2 points) Find the probability that  $F_n(s) < F_n(t)$ , where  $F_n(\cdot)$  is empirical distribution function of some sample  $X_1, \dots, X_n$ .

Pf:  $\forall s=t$ ,  $F_n(s) = F_n(t)$ ,  $P(F_n(s) < F_n(t)) = 0$

$\forall s > t$ ,  $F_n(s) \geq F_n(t)$ ,  $P(F_n(s) < F_n(t)) = 0$ .

$\forall s < t$ ,  $F_n(s) \leq F_n(t)$ .

$$F_n(t) - F_n(s) = \frac{1}{n} \sum \#\{X_i \leq s\} - \frac{1}{n} \sum \#\{X_i \leq t\} = \frac{1}{n} \sum \#\{X_i \in (s, t]\}.$$

$$\text{denote } p = P\{X_i \in (s, t]\}. \quad P(F_n(s) < F_n(t)) = 1 - (1-p)^n$$

3. (4 points) a) Let  $X_1, \dots, X_n$  be i.i.d observations from  $\text{Unif}[a, b]$ . Is the statistic  $X_{(n)} - X_{(1)}$  unbiased (asymptotic or exact) and consistent estimator for  $b - a$

- b) Let  $X_1, \dots, X_n$  be i.i.d observations from  $\text{Unif}[-3a, a]$ . Is the statistic  $4X_{(n)} + X_{(1)}$  unbiased (asymptotic or exact) and consistent estimator for  $b - a$

a) firstly consider  $X_i \sim \text{Unif}[0, 1]$ .

$$F_{X_{(1)}} = 1 - P(X_{(1)} > x) = 1 - (1-x)^n$$

$$f_{X_{(1)}} = n(1-x)^{n-1} \quad E[X_{(1)}] = \int x f_{X_{(1)}} = n \cdot \frac{1}{n(n+1)} = \frac{1}{n+1}.$$

$$\text{similarly, } E[X_{(n)}] = \frac{n}{n+1}.$$

$$\text{then we apply the linear transformation. } E[X_{(n)}] = a + \frac{n}{n+1}(b-a). \quad E[X_{(1)}] = a + \frac{1}{n+1}(b-a).$$

$$E[X_{(n)} - X_{(1)}] = \frac{n-1}{n+1}(b-a) = b-a - \frac{2}{n+1}(b-a). \quad \text{thus } E[X_{(n)} - X_{(1)}] \neq b-a \quad \text{but} \quad \xrightarrow{n \rightarrow \infty} b-a \text{ when } n \rightarrow \infty.$$

$$\lim_{n \rightarrow \infty} P(|X_{(n)} - X_{(1)} - (b-a)|) \leq \lim_{n \rightarrow \infty} P(|X_{(n)} - b| + |X_{(1)} - a|) = 0.$$

thus  $X_{(n)} - X_{(1)}$  is asymptotic unbiased and consistent for  $b - a$ .

b) similarly as a),  $E[X_{(n)}] = -3a + \frac{n}{n+1}(4a) = a \cdot \frac{n-3}{n+1}$

$$E[X_{(1)}] = -3a + \frac{1}{n+1}(4a) = a \cdot \frac{1-3n}{n+1}$$

$$E[4X_{(n)} + X_{(1)}] = a \cdot \left[ \frac{4n-12+1-3n}{n+1} \right] = a \cdot \frac{n-11}{n+1} \neq 4a \quad \text{and} \quad \xrightarrow{n \rightarrow \infty} a \neq 4a.$$

$$X_{(1)} \rightarrow -3a, \quad X_{(n)} \rightarrow a.$$

$$T = 4X_{(1)} + X_{(n)} \rightarrow a \neq 4a \quad \text{thus not unbiased or consistent for } 4a.$$

4. (4 points) Let  $X_1, \dots, X_n$  be i.i.d. observations with finite the second moment and  $E(X)=a$ , which known. Are the following statistics unbiased and consistent:

$$(\bar{X})^2 - a^2, \quad (n-1)^{-1} \sum_i (X_i - a)^2$$

Sol: denote that  $Var(X) = \sigma^2 < \infty$ .

$$E[(\bar{X})^2 - a^2] = Var(\bar{X}) + [E(\bar{X})]^2 - a^2 = \frac{\sigma^2}{n}.$$

for estimator  $a$ . asymptotic unbiased and consistent. (since  $\bar{X} \rightarrow a$ ). ; for others neither.

$$E\left[\frac{1}{n-1} \sum (X_i - a)^2\right] = \frac{1}{n-1} \cdot \sum [E(X_i^2) - 2a^2 + a^2] = \frac{1}{n-1} \sum [Var(X_i) + [E(X_i)]^2 - a^2] = \frac{n\sigma^2}{n-1}$$

$$\frac{1}{n-1} \sum (X_i - a)^2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum (X_i - a)^2 \rightarrow \frac{1}{n} \sum (X_i - a)^2 \rightarrow Var(X_i) = \sigma^2$$

for estimator  $\sigma^2$ . asymptotic unbiased and consistent ; for others neither.

5. (8 points) Let  $X_1, \dots, X_n$  be a sample from a two-parameter exponential distribution with density

$$f_{\alpha, \beta}(y) = \begin{cases} \alpha e^{-\alpha(y-\beta)}, & y \geq \beta, \\ 0, & y < \beta. \end{cases}$$

a) (4 points) Let  $\alpha=1$ , are the following statistics  $X_{(1)}$  and  $\bar{X} - 1$  are unbiased (asymptotic or exact) and consistent estimator for  $\beta$ .

b) (4 points) Let  $\beta = 0$  is the estimator for  $\alpha$ .  $\hat{\beta} = (\bar{X})^{-1}$  unbiased, if not find the bias. Is  $\hat{\beta}$  consistent?

$$Pf: a). f_{\beta}(x) = e^{-(x-\beta)} \cdot \# \{x \geq \beta\}.$$

$$\text{let } Y_{(1)} = X_{(1)} - \beta. \quad Y_{(1)} \sim Exp(1). \quad P(Y_{(1)} > t) = (e^{-t})^n = e^{-nt}$$

$$F_{Y_{(1)}}(t) = 1 - e^{-nt} \quad f_{Y_{(1)}}(t) = n e^{-nt}. \quad E(Y_{(1)}) = \int_0^\infty nt \cdot e^{-nt} = -t \cdot e^{-nt} + \int e^{-nt} = \frac{1}{n}.$$

$$E[Y_{(1)}^2] = \int_0^\infty nt^2 e^{-nt} dt = \frac{2}{n^2} \quad \text{Var}(Y_{(1)}) = \frac{1}{n^2}$$

$$X_{(1)} = Y_{(1)} + \beta \Rightarrow E[X_{(1)}] = \frac{1}{n} + \beta \quad P(|X_{(1)} - \beta| > \varepsilon) = P(X_{(1)} > \beta + \varepsilon) = P(Y_{(1)} > \varepsilon) = e^{-n\varepsilon} \rightarrow 0.$$

$$E[\bar{X}] = E[X_i] = 1 + \beta \Rightarrow E[\bar{X} - 1] = \beta \text{ unbiased.}$$

by the LLN.  $\bar{X} \xrightarrow{P} E[X_i]$ .  $\Rightarrow$  consistent.

Thus.  $X_{(1)}$  asymptotic unbiased and consistent for  $\beta$ ;  $\bar{X} - 1$ . exact unbiased and consistent for  $\beta$ .

$$b). \beta=0. \quad f_{\alpha}(y) = \alpha e^{-\alpha y} \cdot \# \{y \geq 0\}. \quad \text{denote } S_n = \sum_i X_i. \quad f_{S_n}(x) = \frac{\alpha^n}{(n-1)!} x^{n-1} e^{-\alpha x} \quad x > 0.$$

$$E[\bar{X}^{-1}] = E\left[\frac{n}{S_n}\right] = n \cdot \int_0^\infty \frac{1}{t} \cdot \frac{\alpha^n}{(n-1)!} t^{n-1} e^{-\alpha t} dt = \frac{n\alpha^n}{(n-1)!} \int_0^\infty t^{n-2} e^{-\alpha t} dt.$$

$$\stackrel{s=at}{=} \frac{n\alpha}{(n-1)!} \cdot \int_0^\infty s^{n-2} e^{-s} ds = \frac{n\alpha}{(n-1)!} \cdot \Gamma(n-1) = \frac{n\alpha}{n-1} \cdot (n-2)! = \frac{n\alpha}{n-1}$$

the bias:  $\frac{1}{n} \cdot \alpha \rightarrow 0$  ( $n \rightarrow \infty$ ).

by the LLN.  $\bar{X} \xrightarrow{P} \frac{1}{\alpha}$ . by the continuity of  $g(x) = x^{-1}$ .  $\frac{1}{\bar{X}} \rightarrow \alpha$ .

thus.  $\bar{X}^{-1}$  is. asymptotic unbiased and consistent estimator for  $\alpha$ .

6\*(2 point per item. a) Find the distribution of  $\frac{n \cdot X_{(k)}}{\theta}, \frac{n \cdot (\theta - X_{(n-k+1)})}{\theta}$ , if  $X_1, \dots, X_n \sim \text{Unif}[0, \theta]$ .

b) Is the empirical central moment  $\bar{\mu}_4 = \frac{1}{n} \sum_i (X_i - \bar{X})^4$  an unbiased estimator for  $\mu_4 = E(X - E(X))^4$ . If not, find the bias and transform it into an unbiased.

Sol: a) denote  $U_{(k)} = \frac{X_{(k)}}{\theta}$ ,  $U_{(k)} \sim \text{Unif}[0, 1]$ ,  $Z_1 = \frac{n X_{(k)}}{\theta}$ ,  $Z_2 = \frac{n(\theta - X_{(n-k+1)})}{\theta}$

$$F_{U_{(k)}}(u) = \mathbb{P}(U_{(k)} \leq u) = \sum_{j=k}^n \binom{n}{j} u^j (1-u)^{n-j}$$

$$f_{U_{(k)}}(u) = \frac{n!}{(k-1)!(n-k)!} u^{k-1} (1-u)^{n-k}, \quad U_{(k)} = \frac{Z_1}{n}$$

$$f_{Z_1}(z) = f_{U_{(k)}}\left(\frac{z}{n}\right) \cdot \frac{1}{n} = \frac{n!}{(k-1)!(n-k)!} \frac{z^{k-1}}{n^k} \left(1 - \frac{z}{n}\right)^{n-k}$$

$$\Rightarrow F_{Z_1}(z) = \sum_{j=k}^n \binom{n}{j} \left(\frac{z}{n}\right)^j \left(1 - \frac{z}{n}\right)^{n-j}$$

let  $Y_i = \theta - X_i$ ,  $Y_i \sim \text{Unif}[0, \theta]$  and  $Y_{(k)} = \theta - X_{(n-k+1)}$ , thus  $Z_2 = \frac{n Y_{(k)}}{\theta}$

that is, we get the completely same form:  $\begin{cases} Z_2 = \frac{\theta Y_{(k)}}{n} & Y_i \sim \text{Unif}[0, \theta] \\ Z_1 = \frac{\theta X_{(k)}}{n} & X_i \sim \text{Unif}[0, \theta] \end{cases}$

the distribution should be the same.

$$f_{Z_2}(z) = \frac{n!}{(k-1)!(n-k)!} \frac{z^{k-1}}{n^k} \left(1 - \frac{z}{n}\right)^{n-k}, z \in (0, n), \text{ and } F_{Z_2}(z) = \sum_{j=k}^n \binom{n}{j} \left(\frac{z}{n}\right)^j \left(1 - \frac{z}{n}\right)^{n-j}$$

b). denote that  $Y_i = X_i - \mathbb{E}X$ .

$$\bar{Y} = \bar{X} - \mathbb{E}X = \frac{1}{n} \sum Y_i, \quad \mathbb{E}Y_i = 0.$$

$$\begin{aligned} \bar{M}_4 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^4 = \frac{1}{n} \sum_{i=1}^n (Y_i^4 - 4\bar{Y}Y_i^3 + 6\bar{Y}^2Y_i^2 - 4\bar{Y}^3Y_i + \bar{Y}^4) \\ &= \frac{1}{n} \left( \sum Y_i^4 - 4\bar{Y} \sum Y_i^3 + 6\bar{Y}^2 \sum Y_i^2 - 3n\bar{Y}^4 \right) \end{aligned}$$

take the expectation.

$$\mathbb{E}[\sum Y_i^4] = n \cdot M_4.$$

$$\begin{aligned} \mathbb{E}[\bar{Y} \sum Y_i^3] &= \frac{1}{n} \mathbb{E}[(\sum Y_i)(\sum Y_i^3)] = \frac{1}{n} \mathbb{E}[\sum Y_i^4] + \frac{1}{n} \mathbb{E}[\sum_{i=1}^n Y_i \cdot \sum_{i \neq j} Y_j^3] \\ &= \frac{1}{n} \cdot n \cdot M_4 = M_4. \end{aligned}$$

independent and  $\mathbb{E}Y_i = 0$ .

$$\mathbb{E}[\bar{Y}^2 \sum Y_i^2] = \frac{1}{n^2} \mathbb{E}[(\sum Y_i)^2 (\sum Y_i^2)] = \frac{1}{n^2} \sum_{i=1}^n Y_i^2 Y_j Y_k$$

$$\left\{ \begin{array}{ll} \text{if } i=j=k, \mathbb{E}[Y_i^4] = M_4 \\ i=j+k, \mathbb{E}[Y_i^3] \cdot \mathbb{E}[Y_k] = 0 \\ i+j=k, \mathbb{E}[Y_i^2] \cdot \mathbb{E}[Y_j^2] = M_2^2 \\ i+j+k, \mathbb{E}[Y_i^2] \mathbb{E}[Y_j] \mathbb{E}[Y_k] = 0. \end{array} \right. \Rightarrow \mathbb{E}[\bar{Y}^2 \sum Y_i^2] = \frac{1}{n^2} (n M_4 + n(n-1) M_2^2).$$

$$= \frac{M_4 + (n-1)M_2^2}{n}$$

$$\mathbb{E}[\bar{Y}^4] = \frac{1}{n^4} \mathbb{E}[\sum Y_i]^4 = \frac{1}{n} \mathbb{E}(\sum Y_a Y_b Y_c Y_d).$$

$$= \frac{1}{n^4} (n M_4 + C_4^2 C_n^2 M_2^2) = \frac{M_4 + 3(n-1)M_2^2}{n^3}$$

$$\begin{aligned}
\mathbb{E}[\bar{M}_4] &= \frac{1}{n} \left[ nM_4 - 4M_4 + b \frac{M_4 + (n-1)M_2^2}{n} - \beta n \cdot \frac{M_4 + 3(n-1)M_2^2}{n^3} \right] \\
&= M_4 \left[ 1 - \frac{4}{n} + \frac{6}{n^2} - \frac{3}{n^3} \right] + M_2^2 \left[ \frac{6(n-1)}{n^2} - \frac{9(n-1)}{n^3} \right] \\
&= \frac{(n-1)(n^2-3n+3)}{n^3} M_4 + \frac{3(n-1)(2n-3)}{n^3} M_2^2
\end{aligned}$$

thus  $\bar{M}_4$  has bias, but asymptotic unbiased.

To construct unbiased estimator

$$\begin{aligned}
\bar{M}_2^2 &= \frac{1}{n^2} \left( \sum Y_i^2 \right)^2 - \frac{2}{n} \left( \sum Y_i^2 \right) \bar{Y}^2 + \bar{Y}^4 \\
\mathbb{E}[\bar{M}_2^2] &= \frac{1}{n^2} \left( nM_4 + n(n-1)M_2^2 \right) - \frac{2}{n} \left( \frac{M_4}{n} + \frac{n-1}{n} M_2^2 \right) + \frac{M_4}{n^3} + \frac{3(n-1)}{n^3} M_2^2 \\
&= M_4 \left( \frac{1}{n} - \frac{2}{n^2} + \frac{1}{n^3} \right) + M_2^2 \left( \frac{n-1}{n} - \frac{2(n-1)}{n^2} + \frac{3(n-1)}{n^3} \right) \\
&= \frac{(n-1)^2}{n^3} M_4 + \frac{(n-1)(n^2-2n+3)}{n^3} M_2^2
\end{aligned}$$

denote the new estimator for  $M_4$  by  $\hat{M}_4$ .

if we let  $\hat{M}_4 = \alpha \bar{M}_4 + \beta \bar{M}_2^2$  then. solve the LS to find the parameter:

$$\begin{cases} \alpha \cdot \frac{(n-1)(n^2-3n+3)}{n^3} + \beta \cdot \frac{(n-1)^2}{n^3} = 1 \\ \alpha \cdot \frac{3(n-1)(2n-3)}{n^3} + \beta \cdot \frac{(n-1)(n^2-2n+3)}{n^3} = 0 \end{cases}$$

$$\begin{aligned}
\Delta &= \frac{(n-1)^2(n^2-3n+3)(n^2-2n+3)}{n^6} - \frac{3(n-1)^3(2n-3)}{n^6} \\
&= \frac{(n-1)^2}{n^6} \cdot [n^4 - 5n^3 + 12n^2 - 15n + 9] - 6n^2 + 15n - 9 \\
&= \frac{(n-1)^2}{n^6} \cdot n^2(n-2)(n-3) = \frac{(n-1)^2(n-2)(n-3)}{n^4}
\end{aligned}$$

$$\Delta \alpha = \frac{(n-1)(n^2-2n+3)}{n^3} \quad \Delta \beta = - \frac{3(n-1)(2n-3)}{n^3}$$

$$\alpha = \frac{\Delta \alpha}{\Delta} = \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \quad \beta = \frac{\Delta \beta}{\Delta} = - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)}$$

$$\hat{M}_4 = \alpha \bar{M}_4 + \beta \bar{M}_2^2 = \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} \bar{M}_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \bar{M}_2^2$$