Universitat Autònoma de Barcelona
Facultat de Ciències

Modelització i Inferencia
Pràctica 1

Autor:
Gerard Lahuerta Martín
NIU: 1601350

18 de Novembre del 2021

Índex

1 Presentació del problema

Suposa una mostra aleatòria $X_1, \cdots, X_n \sim Binomial(1, p)$. Considera l'estimador

$$T = \frac{\sqrt{n}}{2(n+\sqrt{n})} + \frac{1}{n+\sqrt{n}} \sum_{i=1}^{n} X_i$$

per al paràmetre desconegut $p \in (0, 1)$.

(a) És T un estimador no esbiaixat per a $p \in (0, 1)$? Justifica la teva resposta.

(b) Calcula el MSE de T i fes una representació de T com a funció del paràmetre desconegut $p \in (0, 1)$ considerant $n = 10$ i $n = 50$. Què observes?

(c) És T consistent? I asimptòticament no esbiaixat? Justifica les teves respostes.

2 Resolució del primer apartat

T serà un estimador de X_i si empíricament obtenim el valor teòric de l'esperança de $X_i, \forall i \in \{1, \cdots, n\}$; és a dir, T serà un estimador no esbiaixat si i només si

$$Bias(T) = 0 \Leftrightarrow E(X_i) - E(T) = 0 \Leftrightarrow E(X_i) = E(T) \;\; \forall i \in \{1, \cdots, n\}$$

Escollim per a exemple $i = 1$ (ja que dona igual quina variable aleatòria escollim perquè són independents i idènticament distribuïdes). Sabem que
$E\left(X \sim Binomial(1, p)\right) = E(Y \sim Bernoulli(p)) = p.$

Calculem el valor de *E(T)*:

$$\text{E(T)} = \text{E}\left(\tfrac{\sqrt{n}}{2(n+\sqrt{n})} + \tfrac{1}{n+\sqrt{n}} \textstyle\sum_{i=1}^{n} X_i\right) = E\left(\tfrac{\sqrt{n}}{2(n+\sqrt{n})}\right) + E\left(\tfrac{1}{n+\sqrt{n}} \textstyle\sum_{i=1}^{n} X_i\right) =$$

$$= \frac{\sqrt{n}}{2(n+\sqrt{n})} + \frac{1}{n+\sqrt{n}} E\left(\sum_{i=1}^{n} X_i\right) \underset{\underset{(I)}{\uparrow}}{=} \frac{\sqrt{n}}{2(n+\sqrt{n})} + \frac{n}{n+\sqrt{n}} E(X_1) =$$

$$= \frac{\sqrt{n}}{2(n+\sqrt{n})} + \frac{n}{n+\sqrt{n}} p = \frac{\sqrt{n}+2np}{2(n+\sqrt{n})} \Longrightarrow E(T) = \frac{\sqrt{n}+2np}{2(n+\sqrt{n})}$$

(I): $X_1, \cdots, X_n$ *independents i idènticament distribuïdes*

Calculem ara el *Bias(T)*

$$-Bias(T) = E(X_1) - E(T) = p - \frac{\sqrt{n}+2np}{2(n+\sqrt{n})} = \frac{2\sqrt{n}p + 2np - \sqrt{n} - 2np}{2(\sqrt{n}+n)} =$$

$$= \frac{(2p-1)\sqrt{n}}{2(\sqrt{n}+n)} = \frac{2p-1}{2(1+\sqrt{n})}$$

Veiem de l'expressió trobada que s'anul·larà el numerador (i per tant el $Bias(T) = 0$) quan $p = \frac{1}{2}$.
Observem, per tant, que:

$$\boxed{T \text{ és un estimador esbiaixat } \forall p \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1], \text{ perquè per aquests valors } Bias(T) \neq 0}$$

3 Resolució del segon Apartat

3.1 Càlcul del MSE

Per a calcular el MSE del estimador T cal saber abans la seva variància, ja que $MSE(T) = V(T) - Bias^2(T)$ i el valor del $Bias(T)$ ja el tenim conegut per l'apartat anterior.

Calculem el valor de *V(T)*:

$$V(T) = V\left(\frac{\sqrt{n}}{2(n+\sqrt{n})} + \frac{1}{n+\sqrt{n}}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n+\sqrt{n}}\right)^2 V\left(\sum_{i=1}^{n} X_i\right) \underset{\substack{\uparrow \\ (II)}}{=}$$

$$= \frac{1}{(n+\sqrt{n})^2} nV(X_1) \underset{\substack{\uparrow \\ (II)}}{=} \frac{np(1-p)}{(n+\sqrt{n})^2} = \frac{np(1-p)}{n^2+2n\sqrt{n}+n} = \frac{p(1-p)}{n+2\sqrt{n}+1}$$

(II): *Sabem que la variància d'una* $Binomial(1,p)$ *és la suma de les variàncies que té de* $Bernoulli(p)$*; és a dir, en aquest cas* $V(X_1) = p(1-p)$

Ja amb tota la informació necessària, calculem el valor del $MSE(T)$

Càlcul del valor de *MSE(T)*:

$$MSE(T) = V(T) + Bias^2(T) = \frac{p(1-p)}{n+2\sqrt{n}+1} + \left(\frac{2p-1}{2(1+\sqrt{n})}\right)^2 =$$

$$= \frac{p(1-p)}{n+2\sqrt{n}+1} + \frac{4p^2-4p+1}{4(1+2\sqrt{n}+n)} = \frac{4p(1-p)+4p^2-4p+1}{4(1+2\sqrt{n}+n)} =$$

$$= \frac{4p-4p^2+4p^2-4p+1}{4(1+2\sqrt{n}+n)} = \frac{1}{4(1+2\sqrt{n}+n)} \Longrightarrow \boxed{MSE(T) = \frac{1}{4(1+2\sqrt{n}+n)}}$$

3.2 Representació de T

Per tal d'observar amb més detall com és de bo l'estimador T, hem decidit fer un seguit de representacions mitjançant *RStudio* i representar els resultats obtinguts. Així, podem fer-nos una idea de com interactua T i si els nostres resultats teòrics obtinguts són correctes.

3.2.1 Simulació de l'estimador T

El programa que simula el comportament del nostre estimador T està dividit en 4 parts:

- **Inicialització de les variables**
  Declara i assigna valors a constants que utilitzarem al llarg del programa.
  - $\rightarrow$ $n = \{10, 50\}$ nombre de mostres aleatòries.
  - $\rightarrow$ $M = 500$ nombre de simulacions a fer per cada probabilitat.
  - $\rightarrow$ $p = [0, 0.1, 0.2, 0.3, \cdots, 0.9, 1]$ llista de probabilitats per a fer les simulacions.
  - $\rightarrow$ $sq = \sqrt{n}$ variable que farem servir per a calcular els valors de l'estimador.
  - $\rightarrow$ $den = sq + n$ variable que farem servir per a calcular els valors de l'estimador.
  - $\rightarrow$ $coef\_T1 = \frac{sq}{2 \cdot den}$ variable que farem servir per a calcular els valors de l'estimador.
  - $\rightarrow$ $coef\_T2 = \frac{1}{den}$ variable que farem servir per a calcular els valors de l'estimador.
- **Declaració de les funcions**
  Declarem la funció T que farem servir durant la simulació, inserim aquí el codi de la funció.

```
T = function( x ){
return (coef_T_1 + coef_T_2*x)
}
```

Funció de l'estimador T

- **Main del programa simulador**
  Part principal del programa on és dur a terme les simulacions de l'estimador T. Inserim aquí el Main del programa

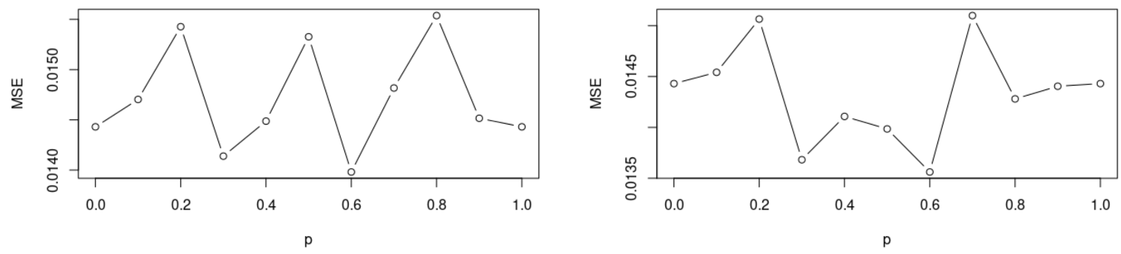
```
That = matrix(NA, length(p), M)
for (i in 1:length(p)) {
for (j in 1:M) {
That [i,j] = T(rbinom(1,n,p[i]))
}
}

MSE = (apply(That, MARGIN = 1, FUN = mean) -p)^2
MSE = MSE + apply(That, MARGIN = 1, FUN = var)

#OUTPUT
plot(p, MSE, type = "b")
```

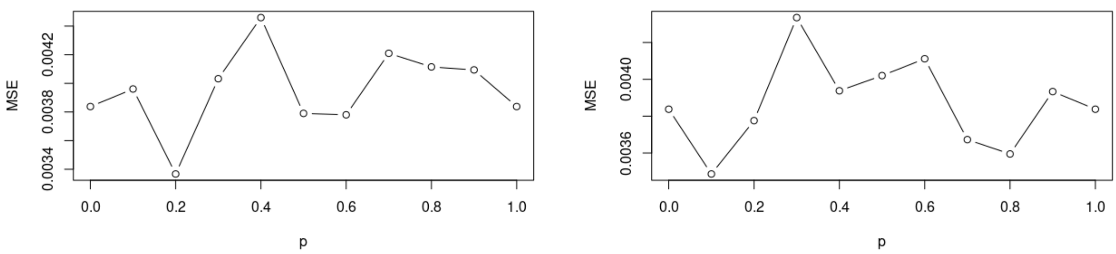
  Observació:
  L'output del programa és un gràfic amb el comportament de l'estimador

- **Output del programa: anàlisis dels gràfics**
  Inserim aquí una petita recopilació dels outputs obtinguts.



((a)) Simulació 1 ((b)) Simulació 2

Figura 1: Gràfic per les simulacions amb $n = 10$



((a)) Simulació 1 ((b)) Simulació 2

Figura 2: Gràfic per les simulacions amb $n = 50$

3.2.2 Conclusions extretes dels resultats obtinguts

Dels gràfics del comportament de l'estimador T, deduïm que aquest no segueix cap relació amb el paràmetre p, ja que no dur a terme cap comportament que no sembli atzarós a mesura que augmentem el valor del paràmetre.
Aquest comportament confirma els càlculs obtinguts teòrics obtinguts anteriorment que deien que l'MSE de l'estimador T no depenia en cap moment del valor de p.
Per altra banda observem que a mesura que augmentem el valor de l'n, els rangs de valors que obté l'MSE disminueix, apropant-se més al zero.
Aquest comportament també ha estat previst pels càlculs teòrics que mostraven relació entre el valor de l'n i l'MSE (ja que aquest paràmetre està en el denominador de la funció MSE obtinguda).
Deduïm per tant, que l'únic paràmetre rellevant per a disminuir l'MSE és el nombre de Bernouillis que conformen la Binomial X; és a dir, el valor de n.

4 Resolució del tercer Apartat

A partir del càlcul del $MSE(T)$ de l'apartat anterior. Calculem la consistència de l'estimador T i si és o no asimptòticament esbiaixat.

4.1 Consistencia de T

Sabem que l'estimador T és consistent $\iff \lim\limits_{n \to \infty} MSE(T) = 0$.
Calculem ara el límit:

$$\lim_{n \to \infty} MSE(T) = \lim_{n \to \infty} \frac{1}{4(1 + 2\sqrt{n} + n)} = \frac{1}{4(\infty + 2 \cdot \infty)} = \frac{1}{\infty} = 0 \Longrightarrow$$

$\Longrightarrow$ $\boxed{\text{l'estimador } T \text{ és consistent}}$

La consistència de l'estimador T també s'observa a partir de les gràfiques mostrades en la pàgina anterior, on a mesura que augmenta el valor de n el valor de l'MSE decreix cap a 0

4.2 Esbiaixat asimptòtic de T

Eventualment, calculem si $\lim\limits_{n \to \infty} Bias(T) = 0$, ja que T serà asimptòticament no esbiaixat si compleix la condició anterior.
Calculem ara el límit:

$$\lim_{n \to \infty} Bias(T) = \lim_{n \to \infty} \frac{2p - 1}{2(1 + \sqrt{n})} = \frac{2p - 1}{2(1 + \sqrt{\infty})} = \frac{2p - 1}{\infty} = 0 \Longrightarrow$$

$\Longrightarrow$ $\boxed{\text{l'estimador } T \text{ és asimptòticament no esbiaixat}}$

5 Conclusió

Observem per tant que l'estimador T és:

- És esbiaxat, ja que $\neg\forall p \in [0,1], Bias(T) = 0$ (demostrat a l'apartat 2)
- Té $MSE(T) = \frac{1}{4(1+2\sqrt{n}+n)}$ (demostrat a l'apartat 3)
- És Consistent, ja que $\underset{n\to\infty}{lim} MSE(T) = 0$ (demostrat a l'apartat 4.1)
- És asimptòticament esbiaxat, ja que $\underset{n\to\infty}{lim} Bias(T) = 0$ (demostrat a l'apartat 4.2)