Tema 2. Estimació puntual

1. Considera una variable aleatoria poblacional amb distribució $B(m,p)$, en mostres aleatories de mida n, s'estima el paràmetre p amb l'estimador:

$$\tilde{p} = \frac{\bar{X}}{m+1}$$

   Es tracta d'un estimador sense biaix?

2. El pes de les taronges de calibre 3/5 es pot considerar que es distribueix segons una llei normal. S'ha pres una mostra de n = 15 taronjes i se les ha pesat. El pes promig de les 15 taronges ha estat $\bar{x} = 150$g i desviació tipus $s = 10$g,

   a) Quin és el millor estimador puntual per la μ, mitjana dels pesos de les taronges de calibre 3/5?

   b) Demostra que $\bar{x}$ és un estimador no esbiaixat i consistent.

   c) Quin és el millor estimador puntual per la σ^2, variància dels pesos de les taronges de calibre 3/5?

   d) Demostra que S^2 és un estimador no esbiaixat i consistent.

3. Es vol estimar la taxa de fallades, θ, d'un ventilador diesel per a camions. Es disposa de dos estimadors independents sense biaix per a estimar θ: $\hat{\theta}_1$ i $\hat{\theta}_2$ amb variàncies conegudes $\sigma_1^2 > 0$ i $\sigma_2^2 > 0$ respectivament. En Martí, estudiant del grau d'estadística, proposa un nou estimador $\hat{\theta}_3 = (1-a)\hat{\theta}_1 + a\hat{\theta}_2$,

   a) Ajudeu al Martí a demostrar que $\hat{\theta}_3$ també és un estimador sense biaix de θ per qualsevol valor de a.

   b) Calculeu la variància de $\hat{\theta}_3$ en funció de a, σ_1^2 i σ_2^2.

4. Comprova que entre tots els estimadors lineals sense biaix $\hat{\theta} = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ de la mitjana poblacional, la mitjana mostral $\bar{X}$ és el de menor variància.

5. Considera una variable aleatòria poblacional amb distribució $N(\mu,\sigma)$ i dos estimadors de la mitjana poblacional μ, en mostres aleatòries de mida n,

$$\hat{\mu}_1 = \frac{X_1 + \cdots + X_n}{n-1} \qquad i \qquad \hat{\mu}_2 = \frac{X_1 + \cdots + X_n}{n+1}$$

   Selecciona el millor tenint en compte l'error quadràtic mitjà de cadascun.

6. En Pere està estudiant la velocitat de creixement d'uns fongs en el laboratori. En Pere sap que la distribució de la velocitat es pot considerar normal però en desconeix els seus paràmetres.De fet, està preocupat especialment en estimar la seva variància. Després de molt pensar dubta entre dos possibles estimadors de la variància poblacional

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

   a) Calculeu el biaix dels dos estimadors. Algún és no esbiaixat?

   b) Calculeu la variància de cadascun dels dos estimadors

   c) Quin és l'error quadràtic mig dels dos estimadors?

7. L'alçada d'una partícula es pot considerar que es distribueix segons una llei exponencial amb paràmetre desconegut λ. Estimeu λ segons el mètode dels moments.

8. En Jan, que està estudiant el grau d'estadística, està jugant amb una xinxeta i recorda que en l'última classe va aprendre que la variable X que compta el nombre de vegades que ha de llençar la xinxeta fins que queda amb la punxa enlaire es pot modelar fent servir una distribu ció Geomètrica amb paràmetre p. Ajuda al Jan a trobar l'estimador de p pel mètode dels moments.

9. El temps de vida d'una bombeta halògena es distribueix segons una llei lognormal amb paràmetres μ i σ. Es pren una mostra aleatòria de n bombetes i es posen a funcionar en un banc de proves tot apuntant el temps de funcionament fins que es fonen. Trobeu estimadors de μ i σ pel mètode dels moments.

10. A partir d'una mostra aleatòria de mida n, d'una distribució uniforme $(a, a+b)$, determineu els estimadors de a i b amb el mètode dels moments.

Exercicis resolts

Exercici 2

a) Un estimador de la mitjana poblacional $\mu = E(X)$, és la mitjana mostral.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

En general, té bones propietats com veurem en els apartats següents.

b) Es verifica

$$E(\bar{X}) = E\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}E\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$$

Per tant, la mitjana mostral és un estimador sense biaix per a la mitjana poblacional μ.

Per estudiar la consistència de $\bar{X}$ ens serà interessant recordar la desigualtat de Txebyshev aplicada a la variable aleatòria $\bar{X}$. Es compleix la desigualtat

$$P\big(|\bar{X} - \mu| \geq \epsilon\big) \leq \frac{var\big(\bar{X}\big)}{\epsilon^2} \qquad \forall\epsilon > 0$$

D'altra banda, denotem per $\sigma^2 = var(X) < \infty$, en mostres aleatòries tenim

$$var\big(\bar{X}\big) = var\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}var\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

I la desigualtat de Txebyshev ens quedarà

$$P\big(|\bar{X} - \mu| \geq \epsilon\big) \leq \frac{\sigma^2}{n\epsilon^2} \qquad \forall\epsilon > 0$$

La consistència de $\bar{X}$ passa per comprovar si

$$\lim_{n\to\infty} P\big(|\bar{X}_n - \mu| \geq \epsilon\big) = 0 \qquad \forall\epsilon > 0$$

Hem utilitzat el subíndex n per indicar la depedència en n de $\bar{X}$. En efecte, tenint en compte la desigualtat de Txebyshev,

$$\lim_{n\to\infty} P\big(|\bar{X}_n - \mu| \geq \epsilon\big) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

c) Veurem que la variància mostral corregida, que anomenarem variància mostral,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

és un bon estimador de la variància poblacional $\sigma^2 = var(X)$.

d) El Teorema de Fisher afirma que en mostres aletòries de mida n de poblacions normals $N(\mu, \sigma)$ es compleix

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

D'altra banda, de les propietats de la distribució χ^2, tenim

$$E(\chi^2_{n-1}) = n-1, \qquad var(\chi^2_{n-1}) = 2(n-1)$$

Per tant,

$$E(\frac{(n-1)S^2}{\sigma^2}) = n-1$$

També

$$E(\frac{(n-1)S^2}{\sigma^2}) = \frac{(n-1)}{\sigma^2}E(S^2)$$

Igualant les dues expressions anteriors

$$\frac{(n-1)}{\sigma^2}E(S^2) = n-1 \Rightarrow E(S^2) = \sigma^2$$

I concluim que S^2 és un estimador sense biaix de σ^2.

Estudiem la variància de S^2. De les propietats de la χ^2 deduim que

$$var\Big((\frac{(n-1)S^2}{\sigma^2})\Big) = 2(n-1)$$

Però també

$$var\Big(\frac{(n-1)S^2}{\sigma^2}\Big) = \frac{(n-1)^2}{\sigma^4}var(S^2)$$

Igualant les dues expressions anteriors

$$\frac{(n-1)^2}{\sigma^4}var(S^2) = 2(n-1) \Rightarrow var(S^2) = \frac{2}{n-1}\sigma^4$$

Aplicant la desigualtat de Txebyshev a la variable aleatòria S^2 resulta que

$$\lim_{n\to\infty} P\big(|S^2_n - \sigma^2| \geq \epsilon\big) \leq \lim_{n\to\infty} \frac{\sigma^4}{(n-1)\epsilon^2} = 0$$

Per tant, la variància mostral S^2 és un estimador consistent de σ^2.

Exercici 3
Calculem l'esperança de $\hat{\theta}_3$.

$$E(\hat{\theta}_3) = E\big((1-a)\hat{\theta}_1 + a\hat{\theta}_2\big)$$

Tenint en compte la linealitat de l'esperança, queda

$$E(\hat{\theta}_3) = (1-a)E\big(\hat{\theta}_1\big) + aE\big(\hat{\theta}_2\big) = (1-a)\theta + a\theta = \theta$$

Certament, es tracta d'un estimador sense biaix per a θ.

Calculem la variància de $\hat{\theta}_3$:

$$\mathrm{var}\big(\hat{\theta}_3\big) = \mathrm{var}\big((1-a)\hat{\theta}_1 + a\hat{\theta}_2\big)$$

Tenint en compte que $\hat{\theta}_1$ i $\hat{\theta}_2$ són independents, queda

$$\mathrm{var}\big(\hat{\theta}_3\big) = (1-a)^2\mathrm{var}\big(\hat{\theta}_1\big) + a^2\mathrm{var}\big(\hat{\theta}_2\big) = (1-a)^2\sigma_1^2 + a^2\sigma_2^2$$

Exercici 5
Necessitem calcular:

$$\begin{aligned} EQM(\hat{\mu_1}) = E((\hat{\mu_1}-\mu)^2) = V(\hat{\mu_1}) + [E(\hat{\mu_1})-\mu]^2 \\ EQM(\hat{\mu_2}) = E((\hat{\mu_2}-\mu)^2) = V(\hat{\mu_2}) + [E(\hat{\mu_2})-\mu]^2 \end{aligned}$$

Les expressions de les esperances i variàncies per cada estimador són:

$$\begin{aligned} E(\hat{\mu_1}) &= \frac{n}{n-1}\mu \qquad & V(\hat{\mu_1}) = \frac{n}{(n-1)^2}\sigma^2 \\ E(\hat{\mu_2}) &= \frac{n}{n+1}\mu \qquad & V(\hat{\mu_2}) = \frac{n}{(n+1)^2}\sigma^2 \end{aligned}$$

Així,

$$\begin{aligned} EQM(\hat{\mu_1}) = \frac{n\sigma^2}{(n-1)^2} + \left[\frac{n}{n-1}\mu - \mu\right]^2 = \frac{n\sigma^2+\mu^2}{(n-1)^2} \\ EQM(\hat{\mu_2}) = \frac{n\sigma^2}{(n+1)^2} + \left[\frac{n}{n+1}\mu - \mu\right]^2 = \frac{n\sigma^2+\mu^2}{(n+1)^2} \end{aligned}$$

Per decidir quin dels dos estimadors és preferible, suposarem que

$$EQM(\hat{\mu_1}) > EQM(\hat{\mu_2})$$

i si fós així, s'ha de verificar que

$$\frac{n\sigma^2 + \mu^2}{(n-1)^2} > \frac{n\sigma^2 + \mu^2}{(n+1)^2}$$

Veiem que la desigualtat es compleix, ja que, operant, arribem a veure que:

$$(n+1)^2 > (n-1)^2$$

I per tant, el millor estimador tenint en compte l'error quadràtic mitjà és $\hat{\mu_2}$.

Exercici 6

a) En l'apartat d) del problema 2 hem comprovat que S^2 és un estimador sense biaix de σ^2. D'altra banda

$$\hat{\sigma}^2 = \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2$$

Aleshores

$$E(\hat{\sigma}^2) = E\big(\frac{n-1}{n}S^2\big) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$$

Veiem que $\hat{\sigma}^2$ és un estimador amb biaix de σ^2.

A l'apartat d) del problema 2 hem comprovat que

$$var(S^2) = \frac{2\sigma^4}{n-1}$$

Aleshores

$$var(\hat{\sigma}^2) = var\Big(\frac{n-1}{n}S^2\Big) = \frac{(n-1)^2}{n^2}var(S^2) = \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} = \frac{2(n-1)}{n^2}\sigma^4$$

Estudiem els errors quadràtics mitjans (EQM) de cada estimador.

$$\text{EQM}(S^2) = (\text{biaix}(S^2))^2 + var(S^2) = 0 + \frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n-1}$$

D'altra banda

$$
\begin{aligned}
\mathrm{EQM}(\hat{\sigma}^2) &= (\mathrm{biaix}(\hat{\sigma}^2))^2 + var(\hat{\sigma}^2) \\
&= \Big(\frac{n-1}{n}\sigma^2 - \sigma^2\Big)^2 + \frac{2(n-1)}{n^2}\sigma^4 \\
&= \Big((\frac{n-1}{n} - 1)\sigma^2\Big)^2 + \frac{2(n-1)}{n^2}\sigma^4 \\
&= \Big((\frac{n-1-n}{n})\sigma^2\Big)^2 + \frac{2(n-1)}{n^2}\sigma^4 \\
&= \frac{1}{n^2}\sigma^4 + \frac{2(n-1)}{n^2}\sigma^4 \\
&= \Big(\frac{1}{n^2} + \frac{2(n-1)}{n^2}\Big)\sigma^4 = \frac{2n-1}{n^2}\sigma^4
\end{aligned}
$$

Per comparar els errors quadràtics mitjans de cada estimador, comparem els coeficients que acompanyen a σ^4. Tenim,

$$
\frac{2n-1}{n^2} < \frac{2n}{n^2} = \frac{2}{n} < \frac{2}{n-1}
$$

D'on deduim que

$$
\mathrm{EQM}(\hat{\sigma}^2) < \mathrm{EQM}(S^2)
$$

Exercici 9

Per trobar els estimadors de μ i σ^2 amb el mètode dels moments, utilitzarem els moments de primer i segon ordre de la v.a. amb distribució lognormal. Tenim

$$
E(X) = e^{\mu + \frac{\sigma^2}{2}}
$$

i

$$
E(X^2) = e^{2\mu + 2\sigma^2}
$$

Identificant aquests moments poblacionals amb els seus equivalents empírics, resulta el sistema d'eqüacions següent:

$$
\begin{aligned}
e^{\mu + \frac{\sigma^2}{2}} &= \bar{X} \\
e^{2\mu + 2\sigma^2} &= \bar{X^2}
\end{aligned}
$$

Es tracta d'aillar els paràmetres en funció dels moments empírics. Tenim

$$
\begin{aligned}
\mu + \frac{\sigma^2}{2} &= \log \bar{X} \\
2\mu + 2\sigma^2 &= \log \bar{X^2}
\end{aligned}
$$

Si a la segona equació li restem dues vegades la primera, resulta l'estimador de σ^2

$$\hat{\sigma}^2 = \log \bar{X^2} - 2\log \bar{X} = \log \bar{X^2} - \log \bar{X}^2$$

Ara, substituint en la primera eqüació resulta:

$$\hat{\mu} = \log \bar{X} - \frac{\hat{\sigma}^2}{2} = \log \bar{X} - \frac{\log \bar{X^2} - \log \bar{X}^2}{2} = 2\log \bar{X} - \frac{1}{2}\log \bar{X^2}$$

Exercici 10

En funció dels paràmetres indicats, els primers moments de la distribució uniforme són:

$$\begin{aligned} E[X] &= \frac{1}{b}\int_a^{a+b} x dx = a + \frac{b}{2} \\ E[X^2] &= \frac{1}{b}\int_a^{a+b} x^2 dx = a^2 + ab + \frac{b^2}{3} \end{aligned}$$

De manera que els estimadors de a i b per el mètode dels moments s'obtenen del sistema d'ecuacions següent:

$$\begin{cases} a + \frac{b}{2} = \bar{X} \\ a^2 + ab + \frac{b^2}{3} = \frac{1}{n}\sum_{i=1}^n x_i^2 \end{cases}$$

Si restem el quadrat de la primera equació a la segona, resulta:

$$\frac{b^2}{12} = S^2$$

I per tant, els estimadors són:

$$\begin{aligned} \hat{b} &= 2\sqrt{3}S \\ \hat{a} &= \bar{X} - \sqrt{3}S \end{aligned}$$