Högskolan i Halmstad

Punktskattning

Exempel En viss specialklass vid grundskolan innehåller alltid 5 elever. Från tidigare år har antalet som lärt sig busvissla innan nian (nionde klass) varit

1997	1998	1999	2000	2001	2002
4	3	3	5	2	4

av de 5 som går i klassen. Skatta m.h.a. dessa siffror sannolikheten att minst 4 av de 5 kan busvissla innan nian.

Lösnig: Låt X vara antalet som kan busvissla innan nian. Då är

$$\begin{aligned}
P(X \geq 4) &= 1 - P(X < 4) \\
&= 1 - P(X \leq 3) \\
&= 1 - \sum_{x=0}^{3} P(X = x) \\
& \left\{ \begin{array}{l} X \text{ binomialfördelad} \\ \text{med } n = 5 \text{ och } p = ? \end{array} \right\} \\
&= 1 - \sum_{x=0}^{3} \binom{5}{x} p^x (1-p)^{5-x}
\end{aligned}$$

Vi kan visserligen utveckla detta men mycket längre än så här kommer vi inte utan ett värde på p. Det har man dock sällan i praktiken. För en skattning av $P(X \geq 4)$ räcker det dock med en skattning av p. Vi vet att väntevärdet, $E(X)$, kan skattas med $\bar{x}$. Dessutom är för en binomialfördelad variabel $E(X) = np$ varmed p kan skattas med p^*_{obs} där

$$np^*_{\text{obs}} = \bar{x} \quad \text{dvs} \quad p^*_{\text{obs}} = \frac{1}{n} \cdot \bar{x} = \frac{1}{5n} \sum_{i=1}^{5} x_i = \frac{4+3+3+5+2+4}{5 \cdot 6} = 0.7$$

och därmed är

$$\begin{aligned}
P(X \geq 4)^*_{\text{obs}} &= 1 - \sum_{x=0}^{3} \binom{5}{x} 0.7^x (1-0.7)^{5-x} \\
&= \sum_{x=4}^{5} \binom{5}{x} 0.7^x 0.3^{5-x} \\
&= 5 \cdot 0.2401 \cdot 0.3 + 0.16807 \\
&\approx 0.528
\end{aligned}$$

□

De vanligaste punktskattningarna

Låt $X_1, \ldots, X_n$ vara ett stickprov på X där att $E(X) = \mu$ och $V(X) = \sigma^2$. Skattningarna μ^*, σ^* och S är funktioner av stokastiska variabler och därmed själva stokastiska variabler.

Beteckning	*Definition*	*Förklaring*	*Fördelning*
μ^*, $\bar{X}$	$\frac{1}{n}\sum_{i=1}^n X_i$	p-skattn. av μ	$\mu^* \in N(\mu, \frac{\sigma}{\sqrt{n}})$
$(\sigma^2)^*$	$\frac{1}{n}\sum_{i=1}^n X_i^2 - \mu^2$	p-skattn. av σ om μ känd	$\frac{(n-1)(\sigma^*)^2}{\sigma^2} \in \chi^2(n-1)$
S^2	$\frac{1}{n-1}(\sum_{i=1}^n X_i^2 - n\bar{X}^2)$	p-skattn. av σ om μ okänd	$\frac{(n-1)s^2}{\sigma^2} \overset{\text{appr.}}{\in} \chi^2(n-1)$

Väntevärdesriktighet och effektivitet

Definition 1 *Antag att θ^* är en punktskattning av parametern θ i fördelningen av en stokastisk variabel. θ^* kallas* **väntevärdesriktig (vvr)** *om $E(\theta^*) = \theta$.*

Exempel Låt $X_1, \ldots, X_n$ vara ett stickprov på $X \in N(\mu, \sigma)$ och $\mu^* = \frac{1}{n}\sum_{i=1}^n X_i$. Då är μ^* vvr skattning av μ ty

$$E(\mu^*) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$

För beviset av att σ^* är en vvr skattning av σ, se Vännman s. 184. $\square$

Definition 2 *Antag att θ_1^* och θ_2^* båda är vvr skattningar av θ.*
Då kallas θ_1^* **effektivare än** *θ_2^* om $V(\theta_1^*) < V(\theta_2^*)$.*

Exempel Låt $X_1, \ldots, X_n$ vara ett stickprov på X som är $\begin{cases} 0 & \text{med sannolikhet } p \\ 1 & \text{med sannolikhet } 1-p \end{cases}$.
Bilda skattningarna $p_1^* = \frac{n - \sum X_i}{a_n}$ och $p_2^* = b_n \min_i X_i$.

(a) Vilka värden måste a_n och b_n ha så p_1^* och p_2^* blir vvr skattningar av p?

(b) Vilken skattning är effektivast?

(c) Antag att man observerat $0, 1, 1, 0, 1, 1, 1$ av X. Vad blir då observationerna $(p_1^*)_{\text{obs}}$ och $(p_2^*)_{\text{obs}}$?

Lösning:

(a) p_1^* : $E(X_i) = \sum_x xp(x) = 0 \cdot p(0) + 1 \cdot p(1) = 1 - p.$

För att p_1^* ska vara vvr måste $p = E(p_1^*) = E(\frac{n - \sum X_i}{a_n}) = \frac{1}{a_n} E(n - \sum X_i) =$
$= \frac{1}{a_n}(n - \sum E(X_i)) = \frac{1}{a_n}(n - n(1-p)) = \frac{1}{a_n} np$ dvs a_n måste vara $= n$
varmed $p_1^* = \frac{n - \sum X_i}{n} = 1 - \frac{1}{n} \sum X_i$.

p_2^* : För att kunna beräkna $E(\min_i X_i)$ härleder vi först fördelningen av $\min_i X_i$:
$P(\min_i X_i \leq x) = 1 - P(\min_i X_i > x) = 1 - P(X_1 > x) \cdots P(X_n > x) =$
$= 1 - (1 - P(X_1 \leq x)) \cdots (1 - P(X_n \leq x)) = \begin{cases} 1 - (1-p)^n & x = 0 \\ 1 & x \geq 1 \end{cases}$

så $P(\min_i X_i = x) = \begin{cases} 1 - (1-p)^n & x = 0 \\ (1-p)^n & x = 1 \\ 0 & \text{annars} \end{cases} \qquad (*)$

Därmed är $E(\min_i X_i) = \sum_x xp(x) = 0 \cdot p(0) + 1 \cdot p(1) = (1-p)^n \quad (**)$
varmed $p = E(p_2^*) = b_n E(\min_i X_i) = b_n (1-p)^n$
så $b_n = \frac{p}{(1-p)^n}$ och $p_2^* = \frac{p}{(1-p)^n} \min_i X_i$.

(b) Vilken skattning är effektivast (dvs har minst varians)?

p_1^* : $V(p_1^*) = V(1 - \frac{1}{n} \sum X_i) = \frac{1}{n^2} \sum V(X_i)$.
Eftersom $E(X_i^2) = 0^2 p + 1^2 (1-p) = 1 - p$ är
$V(X_i) = E(X_i^2) - E(X_i)^2 = (1-p) - (1-p)^2 = (1-p)p$.
Därmed är $V(p_1^*) = \frac{1}{n^2} \sum (1-p)p = \frac{1}{n}(1-p)p$.

p_2^* : $V(p_2^*) = V(\frac{p}{(1-p)^n} \min_i X_i) = \frac{p^2}{(1-p)^{2n}} V(\min_i X_i)$.
Låt $Z = \min_i X_i$.
Då är $E(Z^2) = \sum_z z^2 p(z) = 0^2 P(\min_i X_i = 0) + 1^2 P(\min_i X_i = 1) =$
$= P(\min_i X_i = 1) \overset{(*)}{=} (1-p)^n$ och $E(Z) \overset{(**)}{=} (1-p)^n$ varmed
$V(Z) = E(Z^2) - E(Z)^2 = (1-p)^n - (1-p)^{2n} = (1-p)(1-(1-p)^n)$.
Därmed kan variansen för p_2^* fås som
$V(p_2^*) = \frac{p^2}{(1-p)^{2n}} V(\min_i X_i) = \frac{p^2}{(1-p)^{2n}} (1-p)^n (1-(1-p)^n) = \frac{p^2(1-(1-p)^n)}{(1-p)^n}$.
Då man jämför $V(p_1^*) = \frac{1}{n}(1-p)p$ och $V(p_2^*) = \frac{p^2(1-(1-p)^n)}{(1-p)^n}$ ser man att
$V(p_1^*) \searrow 0$ medan $V(p_2^*) \nearrow \infty$ då $n \to \infty$ ty
$p \in (0,1) \;\Rightarrow\; 1 - p \in (0,1) \;\Rightarrow\; (1-p)^n \searrow 0 \;\Rightarrow\; \frac{1}{(1-p)^n} \nearrow \infty$

så $V(p_2^*) = p^2 \underbrace{(1-(1-p)^n)}_{\to 1} \underbrace{\frac{1}{(1-p)^n}}_{\to \infty}$.

Alltså är p_1^* effektivare än p_2^*, åtminstone för tillräckligt stora n.

(c) $(p_1^*)_{\text{obs}} = 1 - \frac{1}{7} \sum x_i = 1 - \frac{1}{7}(0 + 1 + 1 + 0 + 1 + 1 + 1) = \frac{2}{7}$
$(p_2^*)_{\text{obs}} = \frac{p}{(1-p)^7} \min_i x_i = \frac{p}{(1-p)^7} \cdot 0 = 0$ om $p < 1$, annars odefinierad.

□