Sannolikhet och statistik

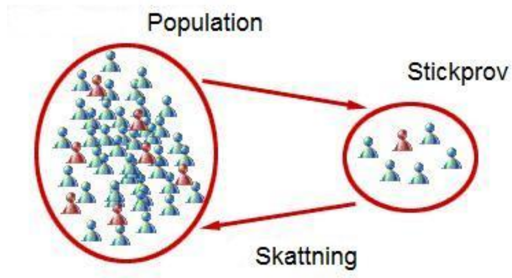
Punktskattning

VT 2009

Uwe.Menzel@math.uu.se

http://www.math.uu.se/∼uwe/

Skattning / Inferens



Figur: Skattning: att dra slutsatser för en hel population pga. ett stickprov

Hur skattas en parameter - intuitivt

Normalfördelning:

$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\left(\sigma^2\right)^*_{obs} = s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\left(\sigma^2\right)^*_{obs} = s^2_\mu = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$$

Binomialfördelning:

$$p^*_{obs} = x/n$$

Frågor

- Finns bättre skattningar än de intuitiva?
- Vad är en bättre skattning, hur bedömer jag det?
- Det enda som är säkert är att skattningen måste vara någon funktion av värdena i stickprovet ...

Olika skattningar

Skattningen måste vara en funktion av mätvärdena

$$\theta^*_{obs} = \theta^*\left(x_1, x_2, \ldots x_n\right)$$

Till exempel:

$$\theta^*_{obs} = \bar{x} = \frac{1}{x}\sum_{i=1}^{n} x_i$$

$$\theta^*_{obs} = max\left(x_1, x_2, \ldots x_n\right)$$

$$\theta^*_{obs} = \tilde{x} \quad \text{(median)}$$

$$\theta^*_{obs} = 1/\bar{x}$$

Från stickprovet till stickprovsvariabel

$$\bar{x} = \frac{1}{n}\sum x_i \quad \Longrightarrow \quad \bar{X} = \frac{1}{n}\sum X_i$$

$$p = \frac{x}{n} \quad \Longrightarrow \quad \hat{p} = \frac{X}{n}$$

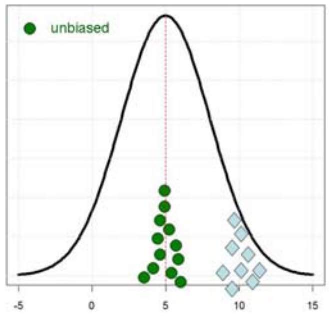
Observationerna $x_1, \ldots x_n$ ses som utfall av oberoende s.v. $X_1, \ldots X_n$ som antas har alla samma fördelning.

Fördel: Man kan (kanske) beräkna fördelningen för stickprovsvariabeln (sampling distribution).

Bedömning av en skattning

1. Väntevärdesriktighet
2. Konsistens
3. Effektivitet

1. Väntevärdesriktig skattning: $E(\theta^*) = \theta$



2. Konsistent skattning

Konsistens: $P(|\theta_n^* - \theta| > \epsilon) \to 0$ om $n \to \infty$

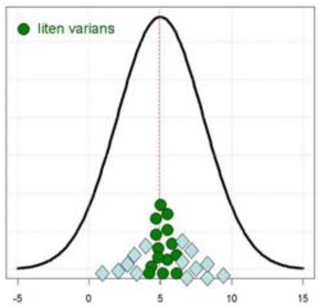
Ju större n, desto större sannolikheten att skattningen θ_n^* ligger i närheten av det sanna värdet θ.

En väntevärdesriktig skattning är konsistent om

$$\lim_{n\to\infty} V(\theta_n^*) = 0$$

3. Effektiv skattning

Effektivitet: $V(\theta^*) < V(\hat{\theta})$



Effektiv skattning

Effektivitet: $V(\theta^*) < V(\hat{\theta})$

En skattning med en mindre varians är bättre (risken är mindre att skattningen ligger långt ifrån det sanna värdet)

Två skattningar kan jämföras genom att definera effektiviteten:

$$\frac{V_1}{V_2} = \frac{\text{mindre varians}}{\text{större varians}} < 1$$

Effektivitet av två skattningarna

Två skattningar:

$$\theta^* = (X_1 + X_2 + X_3 + X_4 + X_5)/5 \quad \text{och} \quad \hat{\theta} = (X_1 + X_5)/2$$

$$\begin{aligned} V(\theta^*) &= \frac{1}{25} V(X_1 + X_2 + X_3 + X_4 + X_5) = \frac{1}{25} \cdot 5\sigma^2 = \frac{1}{5}\sigma^2 \\ V(\hat{\theta}) &= \frac{1}{4} V(X_1 + X_5) = \frac{1}{4} \cdot 2\sigma^2 = \frac{1}{2}\sigma^2 \end{aligned}$$

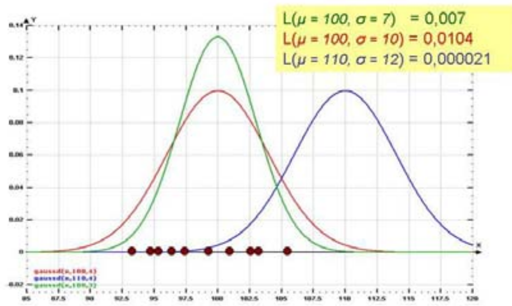
Hur får man ett recept för en skattning?

Det finns olika metoder för att hitta ett uttryck för en skattning:

- Momentmetoden
- Minsta-kvadrat-metoden
- Maximum-likelihood-metoden

Metoderna behandlas i kursen *Inferens*

⋆ Maximum-likelihood metoden



Figur: Maximum-likelihood-skattning: Välj parametrerna så att mätvärdena blir mest plausibla

⋆ Maximum-Likelihood metoden

$$L(\theta) = \begin{cases} p_X(x_1) \cdot p_X(x_2) \cdot \ldots \cdot p_X(x_n) & \text{diskret} \\ f_X(x_1) \cdot f_X(x_2) \cdot \ldots \cdot f_X(x_n) & \text{kontinuerlig} \end{cases}$$

Exempel normalfördelning:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu)^2/2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum(x_i-\mu)^2/2\sigma^2}$$

Hitta μ, σ som gör L (eller $\ln(L)$) maximalt:

$$\frac{\partial L}{\partial \mu} = 0 \quad \frac{\partial L}{\partial \sigma^2} = 0$$

Parametriska punktskattningar

Fördelning	Par.	Skattning	Metod	Referens
$N(\mu, \sigma)$	μ	$\mu^* = \bar{X}$	ML	Blom, s.265
$N(\mu, \sigma)$	σ^2	$S^2 = \frac{1}{n-1} S_{XX}$	ML	Blom, s.265
$N(\mu, \sigma)$	σ	$S = \sqrt{S^2}$	ML	Blom, s.252
$Po(\mu)$	μ	$\mu^* = \bar{X}$	ML	Blom, s.270
$Bin(n, p)$	p	$p^* = X/n$	ML	Blom, s.268
$Exp(\lambda)$	λ	$\lambda^* = 1/\bar{X}$	ML	Blom, s.257
$U(0, \theta)$	θ	$\theta^* = \frac{n+1}{n} \max_i(X_i)$	ML	Blom, s.258
$Weibull(b, c)$	b	$b^* = \frac{n}{\sum X_i^c}$	ML	Blom, s.258
$X \in N(\mu_1, \sigma)$	μ_1	$\mu_1^* = \bar{X}$	ML	Blom, s.266
$Y \in N(\mu_2, \sigma)$	μ_2	$\mu_2^* = \bar{Y}$	ML	Blom, s.266
	σ^2	$S^2 = \frac{S_{XX}+S_{YY}}{(n_1-1)+(n_2-1)}$	ML	Blom, s.266

$\bar{X} = \frac{1}{n}\sum X_i \qquad S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2$

Medelfelet: Normalfördelning, skattning för μ

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

$$V(\bar{X}) = \frac{\sigma^2}{n}$$

$$D(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Istället för σ används dess skattning s som fås från stickprovet.

$$d(\bar{X}) = \frac{s}{\sqrt{n}}$$

Medelfelet: Binomialfördelning, skattning för p

$$X \in Bin(n, p)$$

$$V(X) = n \cdot p \cdot (1-p)$$

$$\hat{p} = \frac{X}{n}$$

$$V(\hat{p}) = \frac{p \cdot (1-p)}{n}$$

$$D(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$$

$$d(\hat{p}) = \sqrt{\frac{p_{obs} \cdot (1-p_{obs})}{n}}$$

Exempel: Opinionsundersökning

$n = 1000$ personer, $x = 350$ svarade *Ja* (observation).
Sökes: andelen *Ja*-svarare i hela befolkningen (population) ?

Stickprov: $p^*_{obs} = x/n = 35\%$

Stickprovsvariablen: $p^* = X/n$ där $X \in Bin(n, p)$

$$d(\hat{p}) = \sqrt{\frac{p_{obs} \cdot (1 - p_{obs})}{n}} = \sqrt{\frac{0.35 \cdot (1 - 0.35)}{1000}} = 1.5\%$$

Andel *Ja*-svarare: 35 ± 1.5 %

Icke-parametriska skattningar

Det kan hända att vi har ingen uppfattning om den underliggande fördelningen som vara mätvärden kommer ifrån.

Vi kan ändå skatta medelvärdet och variansen.

Skattningen är robust - beror inte på någon (förmodad) fördelning.

Icke-parametriska skattningar

Allmän fördelning:

$$\mu^*_{obs} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$(\sigma^2)^*_{obs} = s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Sammanfattning

- Stickprov $\Longrightarrow$ stickprovsvariabel
- Väntevärdesriktigt, konsistent och effektiv skattning
- Parametriska skattningar (tabell)
- Medelfelet
- Icke-parametriska skattningar