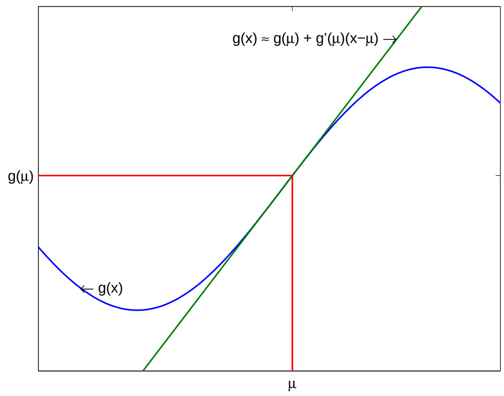
Matematisk statistik för B, K, N, BME och Kemister

Föreläsning 5

Johan Lindström

11 september 2017

Linjärisering av $g(x)$ kring punkten $\mu = E(X)$



Gauss approximationsformler i en variabel (Kap. 5.2)

$Y = g(X)$. Taylorutveckla funktionen g kring $\mu = E(X)$

$$g(X) \approx g(\mu) + (X - \mu)g'(\mu) \quad \Longrightarrow$$

- $E(Y) \approx g(E(X))$
- $V(Y) \approx g'[E(X)]^2 V(X)$

För en funktion av n variabler fås på samma sätt
$Y = g(X_1, \ldots, X_n)$

$$E(Y) \approx g(E(X_1), \ldots, E(X_n))$$
$$V(Y) \approx \sum_{i=1}^{n} c_i^2\, V(X_i) \qquad \text{om } X_i \text{ oberoende}$$

där $c_i = \dfrac{\partial g}{\partial x_i}\Big(E(X_1), \ldots, E(X_n)\Big)$

Delta metoden (CGS + Gaussapproximation)

Om $X_1, X_2, \ldots, X_n$ är oberoende lika fördelade variabler med $E(X_i) = \mu,\ V(X_i) = \sigma^2$ så gäller att

$$g(\overline{X}_n) \underset{\approx}{\in} \mathrm{N}\left(g(\mu), |g'(\mu)|^2 \cdot \frac{\sigma^2}{n}\right) \qquad \text{då } n \text{ stort.}$$

Repetition
- Gauss approximation
- Delta metoden

Statistik
- Översikt
- Grundläggande begrepp
- Exempel

Konfidensintervall
- Chi 2-fördelning
- t-fördelning
- Intervall för N(mu,sigma2)
- Ensidiga konfidensintervall

Exempel: Kvalitetskontroll

Vi kontrollerar n st slumpmässigt utvalda komponenter från ett stort parti och ser om de fungerar.

Modell: $X =$antalet trasiga komponenter
$X \underset{\sim}{\in} \mathrm{Bin}(n, p)$, där p är andelen trasiga kommponenter. p är okänd en parameter i fördelningen.

Möjliga frågeställlningar:

1. Vad är en bra uppskattning av p?
2. Hur stor är osäkerheten i uppskattningen?
3. Vilket intervall tror vi p ligger inom?
4. Hur stort måste n vara för att uppnå en "tillräckligt liten" osäkerhet?

Statistikteori – översikt

Punktskattning
: Hur gör man en bra gissning av en okänd storhet? Hur vet man att den är bra?

Intervallskattning
: Hitta istället ett intervall som täcker den okända storheten med en given (stor) sannolikhet.

Hypotestest
: Om gissningen blev 0.013, kan rätt värde på den okända storheten ändå vara 0.01?

Regression
: Hur vet vi om två variabler påverkar varandra?

Försöksplanering & Faktorförsök
: Hur konstruerar man studier som på bäst sätt (minst antal mätningar) undersöker effekten av olika faktorer (behandlingar)?

Statistikteori: Grundläggande begrepp (Kap. 7.1)

Stickprov

Ett **stickprov**, $x_1, x_2, \ldots, x_n$, är **observationer** av s.v. $X_1, \ldots, X_n$ från någon fördelning $X_i \in F(\vartheta)$ där ϑ är en okänd **parameter**.

Skattning

En **skattning** av ϑ, $\vartheta^*(x_1, \ldots, x_n)$ är en observation av den s.v. $\vartheta^*(X_1, \ldots, X_n)$. Båda betecknas oftast bara med ϑ^*.

Bra egenskaper för en skattning är

Väntevärdesriktig: $E(\vartheta^*) = \vartheta$, inget systematiskt fel.

Effektiv: liten varians (osäkerhet) $V(\vartheta^*)$.

Exempel: Mätning med slumpmässigt mätfel

Antag att vi vill mäta en storhet μ. Om man tar upp n st mätvärden, $x_1, \ldots, x_n$ är dessa observationer av

$$X_i = \mu + \varepsilon_i = \text{"Rätt värde"} + \text{"Mätfel"}$$

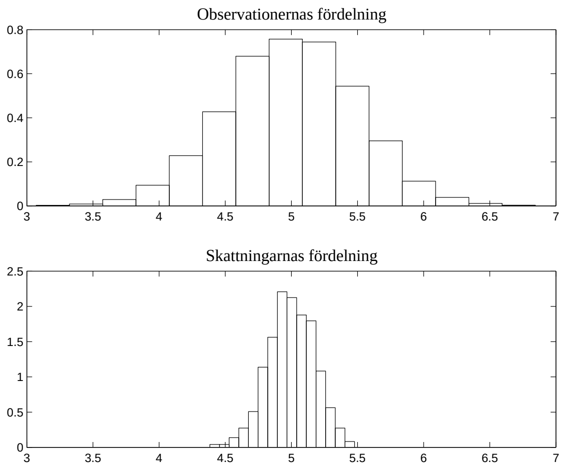
där ε_i är ett slumpmässigt mätfel.

Bestäm skattningar av

1. Medelvärdet μ_n^*.
2. Variansen $\left(\sigma^2\right)^*$.

Variation i observationer ger variation i skattningen

	Observationer, x_{jk}								$\mu^* = \bar{x}_j$
1	4.83	4.93	5.24	5.12	5.10	4.69	5.62	4.73	5.03
2	5.09	5.13	4.53	4.59	4.70	4.10	4.96	5.26	4.79
3	5.53	5.10	4.34	5.05	5.21	4.43	4.30	4.56	4.82
4	4.48	5.10	4.75	5.17	4.98	5.01	5.82	5.12	5.05
5	5.14	5.10	4.79	5.48	4.70	5.89	5.22	5.91	5.28
6	4.80	5.33	5.22	5.26	4.45	4.12	5.29	5.09	4.95
7	5.20	5.26	5.49	5.60	4.83	5.28	4.38	5.18	5.15
8	4.48	4.81	4.62	4.61	5.04	4.81	4.32	4.41	4.64
⋮									



Konfidensintervall (Kap. 7.3)

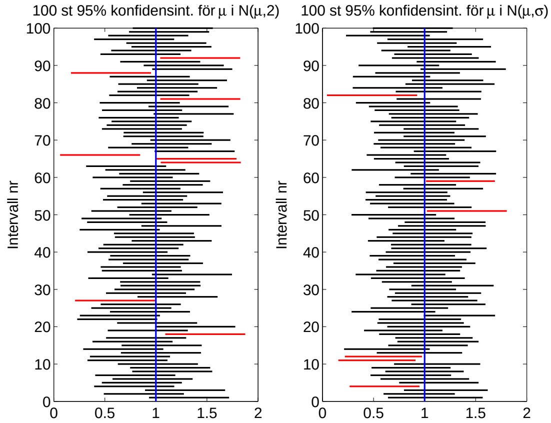
Ett **konfidensintervall** för en parameter ϑ täcker rätt värde på ϑ med sannolikheten $1 - \alpha$.

$1 - \alpha$ kallas **konfidensgrad**. Vanliga värden är 0.95, 0.99 och 0.999.

Ett **tvåsidigt** konfidensintervall är alltså **två** skattningar a_1^*, a_2^* så att

$$P\Big(a_1^*(X_1, \ldots, X_n) < \vartheta < a_2^*(X_1, \ldots, X_n)\Big) = 1 - \alpha$$

Andelen $1-\alpha$ av intervallen täcker rätt värde i långa loppet



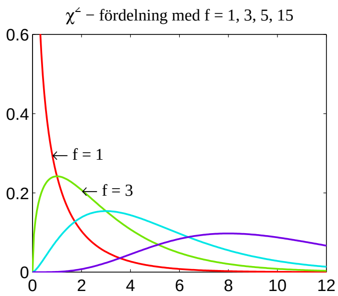
χ^2-fördelning (chi-två) (Kap. 7.2.1)

- $Y \in \chi^2(f)$. f kallas antal frihetsgrader.
- α-kvantil: $\chi^2_\alpha(f)$. Tabell 4.

Om $X_1, \ldots, X_n \in \mathrm{N}(\mu, \sigma^2)$ och oberoende så gäller

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \in \chi^2(n)$$

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



Student's t-fördelning (Kap. 7.2.2)

- $X \in t(f)$. f kallas antal frihetsgrader.
- α-kvantil: $t_\alpha(f)$. Tabell 3.

Om $X \in \mathrm{N}(0,1)$ och $Y \in \chi^2(f)$ är oberoende gäller

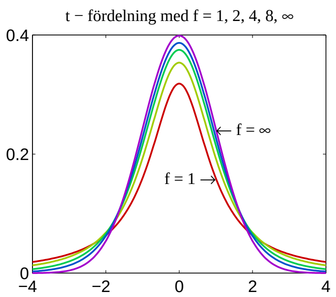
$$\frac{X}{\sqrt{Y/f}} \in t(f)$$

och speciellt för $X_i \in \mathrm{N}(\mu, \sigma^2)$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \in t(n-1)$$

där

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



Konfidensintervall för μ i $\mathsf{N}(\mu, \sigma^2)$ (Kap. 7.3)

$x_1, \dots, x_n$ observationer av $X_i \in \mathsf{N}(\mu, \sigma^2)$

σ^2 känd:

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \mu^* \pm \lambda_{\alpha/2} D(\mu^*)$$

σ^2 okänd:

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} = \mu^* \pm t_{\alpha/2}(f) d(\mu^*)$$

Exempel: Sockerinnehåll i betor

Sockerbetor har i regel ett sockerinnehåll på $16-18\%$ (enligt Dansukkers hemsida). Anta att sockerinnehållet i en godtycklig beta beskrivas av $X_i \in \mathsf{N}(\mu, \sigma^2)$ med σ^2 okänd. I ett visst betlass undersökte man sockerhalten hos 25 slumpmässigt utvalda betor.

$$\frac{1}{25} \sum_{i=1}^{25} x_i = 16.8 \qquad \sum_{i=1}^{25} (x_i - \bar{x})^2 = 4.8$$

Gör ett 95%-konfidensintervall för den förväntade sockerhalten i betlasset.

Konfidensintervall för σ^2 i $\mathsf{N}(\mu, \sigma^2)$ (Kap. 8.1.2)

$x_1, \dots, x_n$ observationer av $X_i \in \mathsf{N}(\mu, \sigma^2)$
Ett $1-\alpha$ konfidensintervall för σ^2 ges av

$$I_\sigma^2 = \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)} \right)$$

Där

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

och $\chi^2_{\alpha/2}(n-1)$ är χ^2-fördelningens kvantiler.

Ensidiga konfidensintervall (Kap. 7.3.4)

Konfidensintervall kan även vara uppåt- eller nedåt begränsade.

1. Ta ena gränsen i ett tvåsidigt konfidensintervall
2. Byt ut $\alpha/2 \longrightarrow \alpha$ för att få rätt konfidensgrad
3. Låt den andra gränsen bli så stor/liten som möjligt

Ex. Om det tvåsidiga intervallet ges av $\bar{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ är

- Nedåt begränsat intervall: $(\bar{x} - \lambda_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty)$
- Uppåt begränsat intervall: $(-\infty, \bar{x} + \lambda_{\alpha}\frac{\sigma}{\sqrt{n}})$