SF1901: Sannolikhetslära och statistik
Föreläsning 9.
Statistik: Intervallskattning (konfidensintervall)

Jan Grandell & Timo Koski

21.02.2012

KTH Matematik

Idén med konfidensintervall

Det föreligger som vanligt ett slumpmässigt stickprov $x_1, \ldots, x_n$ från en fördelning som beror av den okända parametern θ (eventuellt flera stickprov från fördelningar som beror av θ).

En vag definition som strax skall preciseras: Ett intervall I_θ som med sannolikheten $1 - \alpha$ täcker över θ kallas ett *konfidensintervall* för θ med *konfidensgraden* $1 - \alpha$.

KTH Matematik

Definition

Låt $x_1, x_2, \ldots, x_n$ vara utfall av $X_1, X_2, \ldots, X_n$ vars fördelning beror av en okänd parameter θ. Intervallet

$$I_\theta = (a_1(x_1, \ldots, x_n), a_2(x_1, \ldots, x_n))$$

kallas ett konfidensintervall för θ med konfidensgrad $1 - \alpha$ om

$$P(a_1(X_1, \ldots, X_n) < \theta < a_2(X_1, \ldots, X_n)) = 1 - \alpha.$$

KTH VETENSKAP OCH KONST
KTH Matematik

Innehåll

- Konfidensintervall, definition, konfidensgrad
  - konfidensintervall för θ i $N(\theta, \sigma)$ med känt σ.
  - konfidensintervall för θ i $N(\theta, \sigma)$ med okänt σ.
  - konfidensintervall för σ^2 i $N(\theta, \sigma)$
- Nya sannolikhetsfördelningar dyker upp vid konfidensintervall
  - t-fördelning
  - $\chi^2(n)$-fördelning
- Konfidensintervall med approximativ konfidensgrad
  - konfidensintervall för p i $Bin(p)$
  - konfidensintervall för θ i $Po(\theta)$

KTH VETENSKAP OCH KONST
KTH Matematik

Exempel: Mätningar med fel

- $$\text{mätvärde} = \theta + \text{ slumpmässigt normalfördelat mätfel}$$

$$X_i = \theta + \sigma Z_i, \quad Z_i \in N(0,1), \quad i = 1, 2, \ldots, n$$

- $x_1, \ldots, x_n$, observerade utfall av $X_1 \in N(\theta, \sigma)$, $\ldots$ $X_n \in N(\theta, \sigma)$, respektive.
- Vi vill ha en uppfattning om precisionen i skattningen. Visserligen vet vi att

$$E(\overline{X}) = \theta \quad \text{och} \quad D(\overline{X}) = \frac{\sigma}{\sqrt{n}},$$

men vi vill ha en mera informativ och lättbegriplig beskrivning $\rightarrow$ konfidensintervall.

KTH VETENSKAP OCH KONST
KTH Matematik

Repetition: Kvantiler

α-kvantilen för $Z \in N(0,1)$ λ_α definieras av $P(Z > \lambda_\alpha) = \alpha$ eller

$$\alpha = 1 - \Phi(\lambda_\alpha).$$

Symmetri ger

$$\alpha = \Phi(\lambda_{1-\alpha})$$

$$\Leftrightarrow$$

$$\alpha = 1 - \Phi(-\lambda_{1-\alpha}),$$

vilket ger

$$\lambda_{1-\alpha} = -\lambda_\alpha.$$

(Slut på repetition.)

KTH Matematik

$N(\theta, \sigma)$, konfidensintervall för θ fall a) σ känt

Vi antar att $X_1, X_2, \ldots, X_n$ är oberoende och $N(\theta, \sigma)$-fördelade.
Detta innebär att

$$\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \text{ är } N(0,1)\text{-fördelad.}$$

Således gäller att

$$P\left(-\lambda_{\alpha/2} < \frac{\overline{X} - \theta}{\sigma/\sqrt{n}} < \lambda_{\alpha/2}\right) = 1 - \alpha.$$

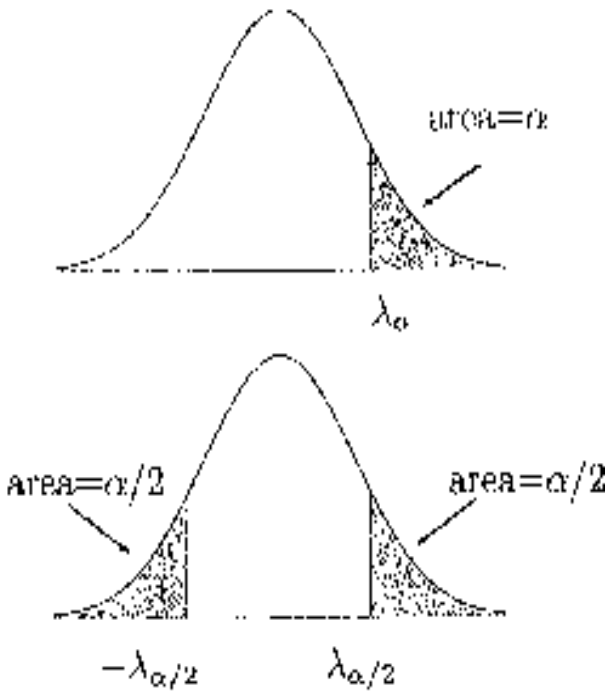
KTH Matematik

Kursens formelsamling: kvantiler för $N(0,1)$

Tab 2. Normalfördelningens kvantiler

$P(X > \lambda_\alpha) = \alpha$ där $X \subset N(0,1)$

α	λ_α	α	λ_α
0.10	1.2816	0.001	3.0902
0.05	1.6449	0.0005	3.2905
0.025	1.9600	0.0001	3.7190
0.010	2.3263	0.00005	3.8906
0.005	2.5758	0.00001	4.2649



KTH Matematik

$N(\theta,\sigma)$, konfidensintervall för θ a) σ känt

Detta ger

$$P\left(-\lambda_{\alpha/2} < \frac{\overline{X}-\theta}{\sigma/\sqrt{n}} < \lambda_{\alpha/2}\right) = 1-\alpha.$$

$$\Updownarrow$$

$$P\left(-\lambda_{\alpha/2}\sigma/\sqrt{n} < \overline{X}-\theta < \lambda_{\alpha/2}\sigma/\sqrt{n}\right) = 1-\alpha$$

$$\Updownarrow$$

KTH Matematik

$N(\theta,\sigma)$, konfidensintervall för θ fall a) σ känt

$$P\left(-\lambda_{\alpha/2}\sigma/\sqrt{n} < \theta - \overline{X} < \lambda_{\alpha/2}\sigma/\sqrt{n}\right) = 1-\alpha$$

$$\Updownarrow$$

$$P\left(\overline{X} - \lambda_{\alpha/2}\sigma/\sqrt{n} < \theta < \overline{X} + \lambda_{\alpha/2}\sigma/\sqrt{n}\right) = 1-\alpha.$$

Jmf. vi definitionen av konfidensintervall så inser vi att

$$\underline{I_\theta = \overline{x} \pm \lambda_{\alpha/2}\sigma/\sqrt{n}}$$

har konfidensgrad $1-\alpha$. En vanlig konfidensgrad är 95%. Då är $\lambda_{0.025} = 1.96$.

KTH Matematik

$N(\theta, \sigma)$, konfidensintervall för θ. Fall a) σ känt

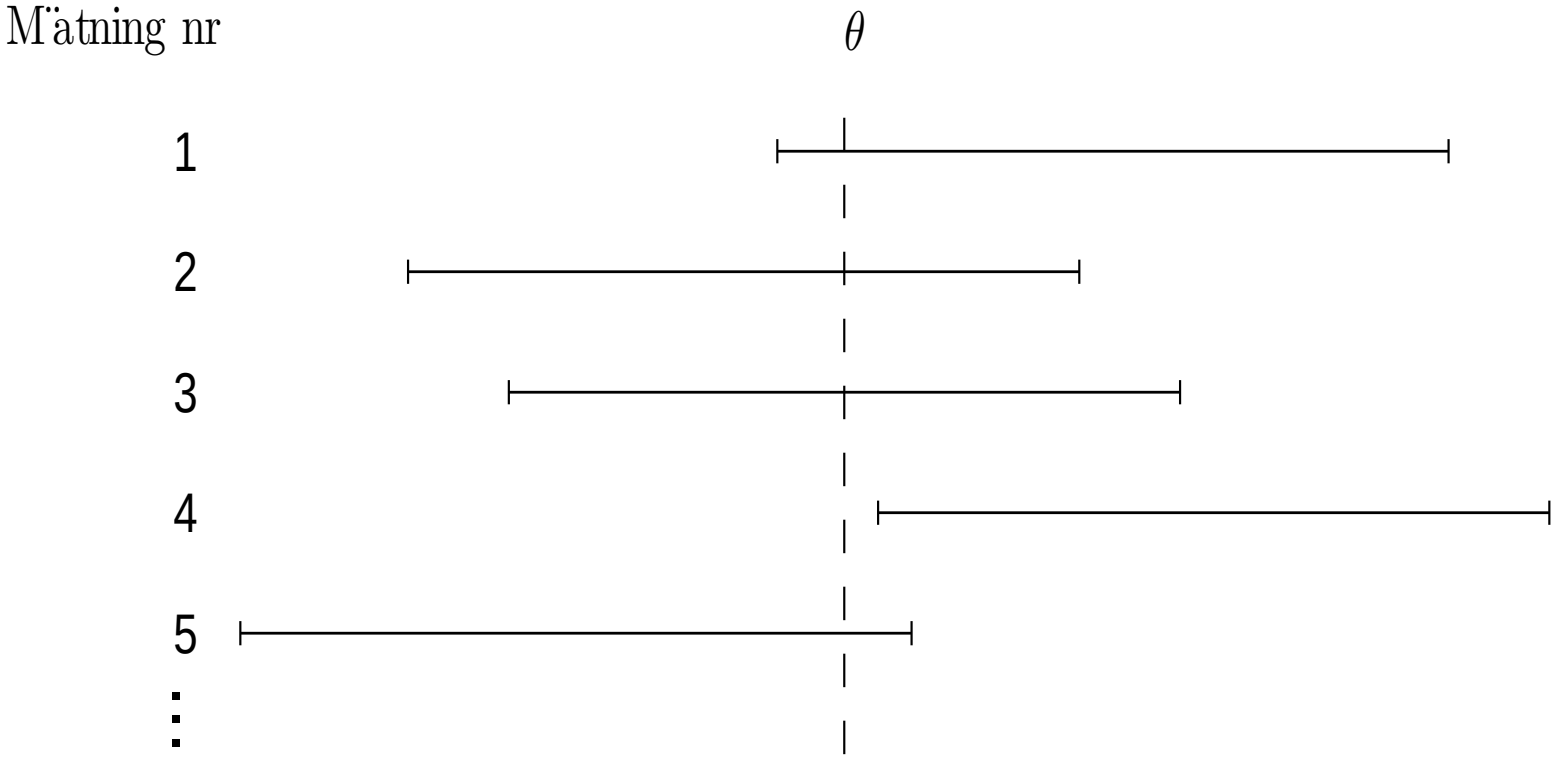
$$I_\theta = \overline{x} \pm \lambda_{\alpha/2}\sigma/\sqrt{n}$$

Intervallängden beror som synes av σ. Ju större precision mätningen har, desto mindre är σ och desto kortare blir intervallet. Längden är också beroende av den önskade konfidensgraden. Med 95 % konfidensgrad blir konstanten $\lambda_{\alpha/2} = 1.96$, 99 % konfidensgrad ger $\lambda_{\alpha/2} = 2.58$, och intervallet är ca 30 % längre.

KTH VETENSKAP OCH KONST
KTH Matematik

$N(\theta,\sigma)$, konfidensintervall för θ a) σ känt

Låt oss gång på gång upprepa insamlingen av data och varje gång bestämma ett tvåsidigt 95 % intervall. I det långa loppet skulle 95% av dem täcka över det okända värdet θ, medan återstoden skulle "missa" det.



$N(\theta,\sigma)$, konfidensintervall för θ. Fall b) σ okänt

Vi utgår nu från

$$\frac{\overline{X}-\theta}{S/\sqrt{n}},$$

där

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

Man kan bestämma fördelningen för $\frac{\bar{X}-\theta}{S/\sqrt{n}}$.

KTH VETENSKAP OCH KONST
KTH Matematik

$N(\theta, \sigma)$, konfidensintervall för θ. Fall b) σ okänt

Det gäller att

$$\frac{\overline{X} - \theta}{S/\sqrt{n}}$$

är t-fördelad med $n-1$ frihetsgrader, eller att den är $t(n-1)$-fördelad.
Vi återkommer till t-fördelningens matematik senare i denna föreläsning.

KTH Matematik

Tätheten för $t(n)$

$t(n)$-fördelningen kallas även Students t-fördelning. $n = 1, 2, 3, \ldots$

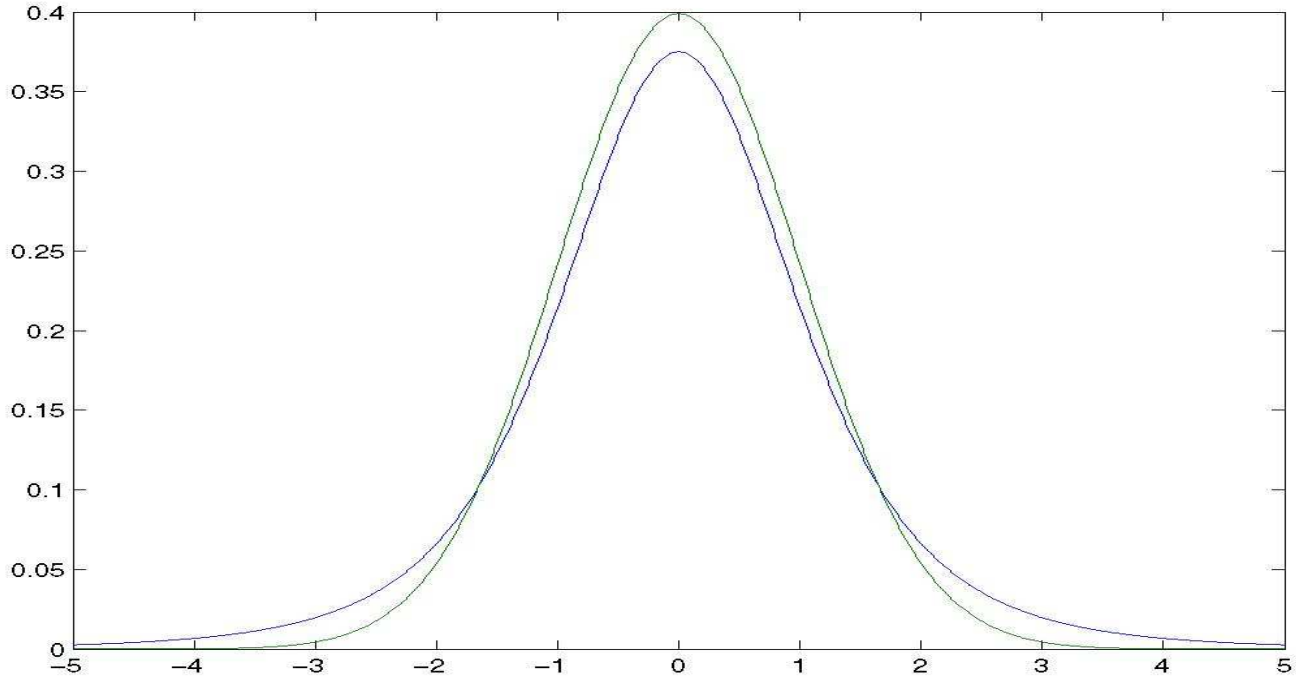
$$f_X(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

Vi kommer att använda t-fördelningens kvantiler men inte alls med denna formel. $\Gamma(x)$ finns definierad på sid. 64 av G. Blom m.fl..

KTH Matematik

Kursens formelsamling: kvantiler för $t(n)$

Fördelning är symmetrisk, och för stora värden på n, lik $N(0,1)$-fördelningen. I bilden ser vi $t(4)$ (funktionsgraf i blått) och $N(0,1)$ (funktionsgraf i grönt).

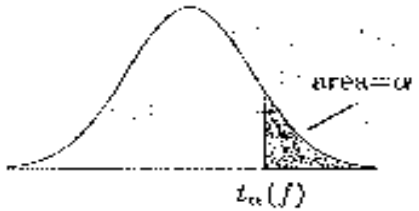


KTH Matematik

Kursens formelsamling: kvantiler för $t(n)$

I tabellen är antalet frihetsgrader betecknat med f

Tabell 3. t-fördelningen.
$P(X > t_\alpha(f)) = \alpha$, där $X \in t(f)$.



f \ α	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	10.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59

KTH Matematik

$N(\theta,\sigma)$, konfidensintervall för θ b) σ okänt

På samma sätt som i a) fås nu att

$$\underline{I_\theta = \overline{x} \pm t_{\alpha/2}(n-1)s/\sqrt{n}}$$

är ett konfidensintervall för θ med konfidensgrad $1-\alpha$.
I fallet med $n = 10$ gäller $t_{0.025}(9) = 2.26$, vilket kan jämföras med $\lambda_{0.025} = 1.96$.

KTH Matematik

Sammanfattning

Sats

Låt $x_1, \ldots, x_n$ vara ett slumpmässigt stickprov från $N(\theta, \sigma)$ där θ är okänt. Då är

$$I_\theta = (\bar{x} - \lambda_{\alpha/2} D, \bar{x} + \lambda_{\alpha/2} D)$$

om σ är känt ($D = \sigma/\sqrt{n}$)

samt

$$I_\theta = \left(\bar{x} - t_{\alpha/2}(f)d, \bar{x} + t_{\alpha/2}(f)d\right)$$

om σ är okänt ($d = s/\sqrt{n},\ f = n-1$)

ett tvåsidigt konfidensintervall för θ med konfidensgraden $1 - \alpha$.

KTH Matematik

Pausbild

The derivation of the t-distribution was first published in 1908 by William Sealy Gosset, while he worked at a Guinness Brewery in Dublin. He was prohibited from publishing under his own name, so the paper was written under the pseudonym Student.

KTH VETENSKAP OCH KONST
KTH Matematik

$\chi^2(n)$-fördelning

Vi ska börja med ett par sannolikhetsteoretiska resultat.

Sats

Om $Z_1, \ldots, Z_n$ är oberoende och $N(0,1)$-fördelade, så är

$$\sum_{i=1}^{n} Z_i^2$$

$\chi^2(n)$-fördelad (uttal: tji-två eller ki-två) med n frihetsgrader.

$\chi^2(n)$-fördelning: täthet

Definition

Om den s.v. X har täthetsfunktionen

$$f_X(x) = \begin{cases} \dfrac{x^{\frac{f}{2}-1}e^{-x/2}}{\Gamma(f/2)2^{f/2}} & \text{om } x > 0 \\ 0 & \text{om } x \leq 0, \end{cases}$$

säges X vara χ^2-fördelad med f frihetsgrader.

Vi använder kvantilerna för $\chi^2(n)$-fördelning men inte alls formeln för tätheten. $\Gamma(f/2)$ finns definierad på sid. 64 av G. Blom .fl..

KTH Matematik

$\chi^2(n)$-fördelning

$X \in \chi^2(n)$.

$$P\left(X \geq \chi^2_{\alpha/2}(n)\right) = \frac{\alpha}{2}$$

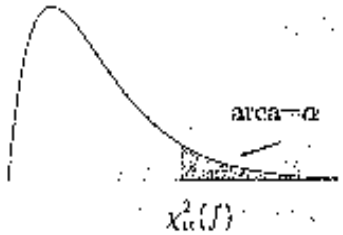
$$P\left(X \geq \chi^2_{1-\alpha/2}(n)\right) = 1 - \frac{\alpha}{2}$$

Vi är nu i en lite besvärligare situation än för μ, eftersom χ^2-fördelningen inte är symmetrisk. I normal- respektive t-fallet utnyttjade vi att symmetrin medförde att $\lambda_{1-\alpha} = -\lambda_\alpha$ resp. $t_{1-\alpha}(n-1) = -t_\alpha(n-1)$.

KTH Matematik

Kursens formelsamling: kvantiler för $\chi^2(n)$-fördelning

Tabell 4. χ^2**-fördelningen.**
$P(X > \chi^2_\alpha(f)) = \alpha$, där $X \in \chi^2(f)$.



f \ α	0.9995	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.00	0.00	0.00	0.00	0.00	0.00	3.84	5.02	6.63	7.88	10.8	12.1
2	0.00	0.00	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.6	13.8	15.2
3	0.02	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.3	12.8	16.3	17.7
4	0.06	0.09	0.21	0.30	0.48	0.71	9.49	11.1	13.3	14.9	18.5	20.0
5	0.16	0.21	0.41	0.55	0.83	1.15	11.1	12.8	15.1	16.7	20.5	22.1
6	0.30	0.38	0.68	0.87	1.24	1.64	12.6	14.4	16.8	18.5	22.5	24.1
7	0.48	0.60	0.99	1.24	1.69	2.17	14.1	16.0	18.5	20.3	24.3	26.0
8	0.71	0.86	1.34	1.65	2.18	2.73	15.5	17.5	20.1	22.0	26.1	27.9

KTH Matematik

$\chi^2(n)$-fördelning

Sats

Om $X_1, X_2, \ldots, X_n$ är oberoende och $N(\theta, \sigma)$-fördelade så är

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

$\chi^2(n-1)$-fördelad.

Detta är den "riktiga" motiveringen till att man i s^2 dividerar med $n-1$.

KTH VETENSKAP OCH KONST
KTH Matematik

konfidensintervall för σ^2 i $N(\theta, \sigma)$

Vi kommer att använda $\frac{(n-1)S^2}{\sigma^2}$ för konfidensintervall lite på samma sätt som vi använde oss av $\bar{X}$ för konfidensintervall för θ.

KTH Matematik

Konfidensintervall för σ^2

Låt nu $\chi^2_\alpha(n-1)$ vara α-kvantilen i $\chi^2(n-1)$-fördelningen. Då gäller

$$P\left(\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}(n-1)\right) = 1-\alpha$$

$$\Updownarrow$$

$$P\left(\frac{\chi^2_{1-\alpha/2}(n-1)}{n-1} < \frac{S^2}{\sigma^2} < \frac{\chi^2_{\alpha/2}(n-1)}{n-1}\right) = 1-\alpha$$

$$\Updownarrow$$

KTH Matematik

Konfidensintervall för σ^2

$$P\left(\frac{n-1}{\chi^2_{\alpha/2}(n-1)} < \frac{\sigma^2}{S^2} < \frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}\right) = 1-\alpha$$

$$\Updownarrow$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) = 1-\alpha$$

KTH VETENSKAP OCH KONST
KTH Matematik

Ett stickprov, konfidensintervall för σ^2

$$\Updownarrow$$

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right) = 1-\alpha.$$

KTH Matematik

Ett stickprov, konfidensintervall för σ^2

Detta ger att

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

resp.

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

är konfidensintervall för σ^2 resp. σ med konfidensgrad $1-\alpha$.

KTH VETENSKAP OCH KONST
KTH Matematik

t-fördelningen

Vi återgår nu lite till t-fördelningen.

Sats

Om X är $N(0,1)$-fördelad, Y är $\chi^2(f)$-fördelad, och X och Y är oberoende, så är

$$\frac{X}{\sqrt{Y/f}}$$

$t(f)$-fördelad.

t-fördelningen

Sats

Om $X_1, X_2, \ldots, X_n$ är oberoende och $N(\theta, \sigma)$-fördelade så är $\overline{X}$ och S^2 oberoende.

Denna sats karakteriserar normalfördelningen! Med detta avses att satsen inte är sann för någon annan fördelning.

KTH VETENSKAP OCH KONST
KTH Matematik

t-fördelningen

Av detta följer nu att

$$\frac{\overline{X}-\theta}{S/\sqrt{n}} = \frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{S^2}{\sigma^2}}$$

är $t(n-1)$-fördelad.

$$\frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{S^2}{\sigma^2}} = \frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}$$

$\frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \in N(0,1)$, $\frac{(n-1)S^2}{\sigma^2} \in \chi^2(n-1)$ och oberoende.

KTH Matematik

Användning av normalapproximation, approximativ konfidensgrad

- Fördelningen beror av en okänd parameter θ.
- en punktskattning θ^* som är ungefär normalfördelad med väntevärdet θ och standardavvikelsen $D = D(\theta^*)$.
- Då gäller approximativt

$$\frac{\theta^* - \theta}{D} \in N(0,1).$$

KTH VETENSKAP OCH KONST
KTH Matematik

Medelfel för en skattning

Vi använder variansen $V(\theta^*)$ eller, vilket i princip är samma sak, standardavvikelsen $D(\theta^*)$ som precisionsmått för en skattning θ^*. Ju mindre varians (större effektivitet), desto belåtnare är vi med skattningen. Ibland hamnar man då i en besvärlig situation: Variansen och standardavvikelsen är själva okända, emedan de beror av just den parameter som man vill skatta (och kanske av ytterligare andra okända parametrar).

KTH Matematik

Medelfel för en skattning

Om man vill få information om $D(\theta^*)$ får man försöka hitta på en skattning även av denna storhet (parameter). Konsekvensen blir att man inte får ett exakt precisionsmått utan bara ett ungefärligt. Vi skulle kunna beteckna denna numeriska skattning av osäkerheten med $D(\theta^*)^*_{\text{obs}}$ men skriver den i stället $d(\theta^*)$.

Definition

En skattning av $D(\theta^*)$ kallas medelfelet för θ^* och betecknas $d(\theta^*)$.

Hur medelfelet skall väljas får avgöras från fall till fall. Man borde tillse att $d(\theta^*)$ är en konsistent skattning av $D(\theta^*)$.
Detta kan nog verka förbryllande men man skall hålla isär begreppen: θ^*_{obs} är en skattning av θ och $d(\theta^*)$ är en skattning av $D(\theta^*)$. Det var detta vi gjorde i det inledande exemplet i i samband med analys av en opinionsundersökning.

KTH VETENSKAP OCH KONST
KTH Matematik

Konfidensintervall med en approximativ konfidensgrad

En punktskattning θ^* av en okänd parameter θ är ungefär normalfördelad med väntevärdet θ och standardavvikelsen D. Då är

$$I_\theta = (\theta^* - \lambda_{\alpha/2} D, \theta^* + \lambda_{\alpha/2} D) \qquad \text{om } D \text{ ej beror av } \theta$$
$$I_\theta = (\theta^* - \lambda_{\alpha/2} d, \theta^* + \lambda_{\alpha/2} d) \qquad \text{om } D \text{ beror av } \theta$$

(och d väljs lämpligt) ett konfidensintervall för θ med den approximativa konfidensgraden $1 - \alpha$.

KTH Matematik

Konfidensintervall med en approximativ konfidensgrad

Låt $x_1, \ldots, x_n$ vara ett stort slumpmässigt stickprov från en fördelning, där väntevärdet är θ och standardavvikelsen σ. Då är

$$I_\mu = (\bar{x} - \lambda_{\alpha/2} D, \bar{x} + \lambda_{\alpha/2} D) \qquad \text{om } \sigma \text{ känt } (D = \sigma/\sqrt{n})$$
$$I_\mu = (\bar{x} - \lambda_{\alpha/2} d, \bar{x} + \lambda_{\alpha/2} d) \qquad \text{om } \sigma \text{ okänt } (d = s/\sqrt{n})$$

konfidensintervall för θ med den approximativa konfidensgraden $1 - \alpha$.

KTH VETENSKAP OCH KONST
KTH Matematik

Konfidensintervall för p i $\mathrm{Bin}(n, p)$ med en approximativ konfidensgrad

Låt x vara en observation av X, där $X \in \mathrm{Bin}(n, p)$ och p är okänt. Maximum-likelihood skattningen är

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

Man kan inte enkelt konstruera ett intervall I_p som exakt har given konfidensgrad.

KTH VETENSKAP OCH KONST
KTH Matematik

Konfidensintervall för p i Bin(n, p) med en approximativ konfidensgrad

För stora n gäller approximativt att $X/n \in N(p, D)$ där $D = \sqrt{p(1-p)/n}$. Som medelfel tar vi $d = \sqrt{p^*_{\text{obs}}(1-p^*_{\text{obs}})/n}$. och får intervallet

$$I_p = (p^*_{\text{obs}} - \lambda_{\alpha/2} d, p^*_{\text{obs}} + \lambda_{\alpha/2} d) \qquad \left(d = \sqrt{p^*_{\text{obs}}(1-p^*_{\text{obs}})/n}\right). \quad (1)$$

KTH Matematik

Konfidensintervall för p i Bin(n, p) med en approximativ konfidensgrad

Vid intervjuer med $n = 250$ personer, uttagna slumpmässigt ur en mycket stor population, visade sig 42 ha en viss åsikt H. Som punktskattning av relativa antalet p i hela populationen med åsikten H tar vi $p^*_{\text{obs}} = 42/250 = 0.168$. Alltså blir

$$d = \sqrt{0.168 \cdot 0.832/250} = 0.023.$$

Ett approximativt 95 % konfidensintervall för p blir

$$I_p = (0.168 \pm 1.96 \cdot 0.023) = (0.168 \pm 0.046) = (0.12, 0.21).$$

KTH Matematik

Approximativa konfidensintervall

Man önskar planera undersökningen så, att man efteråt får ett ungefär 95 % konfidensintervall för p med högst längden $2 \cdot 0.05$. Hur många skall tillfrågas ?

KTH Matematik

Approximativa konfidensintervall

Man får

$$2 \cdot 1.96 \sqrt{\frac{p^*_{\text{obs}}(1 - p^*_{\text{obs}})}{n}} < 2 \cdot 0.05$$

varav, eftersom alltid $p(1-p) \leq 1/4$,

$$n \gtrsim \frac{1}{4}\left(\frac{1.96}{0.05}\right)^2 = 384.2.$$

385 personer bör alltså tillfrågas.

Approximativa konfidensintervall: populationsundersökning

ML-skattningen av p blir, som vi redan vet $p^*_{\text{obs}} = x/n$. Den approximativa formeln för konfidensintervallet blir

$$I_p = (p^*_{\text{obs}} - \lambda_{\alpha/2} d, p^*_{\text{obs}} + \lambda_{\alpha/2} d) \qquad (d = d_n \sqrt{p^*_{\text{obs}}(1 - p^*_{\text{obs}})/n}). \quad (2)$$

där $d_n = \sqrt{(N-n)/(N-1)}$ är en s.k. korrektionsfaktor för ändlig population.

KTH VETENSKAP OCH KONST
KTH Matematik

Approximativa konfidensintervall: poissonfördelning

Vi har ett slumpmässigt stickprov x av $X \in \mathrm{Po}(\theta)$ och punktskattar μ.
Under antagande att det är känt att $\theta >$ ca 15 kan normalapproximationen användas
Som skattning av θ tar vi $\theta^*_{\text{obs}} = x$. Tillhörande standardavvikelse är $D(X) = \sqrt{\theta}$ och vi tar därför som medelfel $d = \sqrt{x}$.

KTH VETENSKAP OCH KONST
KTH Matematik

Approximativa konfidensintervall: poissonfördelning

Konfidensintervallet för θ blir enligt den approximativa metoden

$$I_\theta = \left(x - \lambda_{\alpha/2}\sqrt{x}, x + \lambda_{\alpha/2}\sqrt{x}\right). \tag{3}$$

Dess konfidensgrad är ungefär $1 - \alpha$, och approximationen är bättre ju större θ är.

KTH Matematik