SUPPLEMENT
til
Matematiske Grundbegreber

INDHOLD

SUPPLEMENT: 2A BEVISER VEDRØRENDE NORMALFORDELINGEN.

I definitionen af normalfordelingen indgår, at $f(x)=\frac{1}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$ er en tæthedsfunktion,

dvs. $f(x)\geq 0$ og $\int_{-\infty}^{\infty} f(x)dx=1$

Bevis: Da en eksponentialfunktion er positiv, er $f(x)\geq 0$.

Substitueres $t=\frac{x-\mu}{\sigma}$ fås $\int_{-\infty}^{\infty} f(x)dx=1 \Leftrightarrow \int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot t^2}dt=\sqrt{2\pi}$

Sættes $I=\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\cdot t^2}dt$ fås

$$I^2=\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\cdot t^2}dt\cdot\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\cdot t^2}dt=\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\cdot t^2}dt\cdot\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\cdot u^2}du=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(t^2+u^2)}dt\,du$$

Indføres nu polære koordinater $t=r\cdot\cos v,\quad u=r\cdot\sin v$ fås

$$I^2=\int_0^{2\pi} dv\int_0^{r\to\infty} \mathrm{e}^{-\frac{1}{2}\cdot r^2}rdr=-2\pi\left[\mathrm{e}^{-\frac{1}{2}\cdot r^2}\right]_0^{r\to\infty}=2\pi.$$

Heraf fås det ønskede.

✡

SÆTNING 2.1 (*Middelværdi og spredning for normalfordeling*). *Normalfordelingen* $n(\mu,\sigma)$ *har middelværdien* μ *og spredningen* σ.

Bevis: $E(X)=\int_{-\infty}^{\infty}\frac{x}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}dx$.

Substitueres $t=\frac{x-\mu}{\sigma}\Leftrightarrow x=\mu+\sigma\cdot t$ fås

$$E(X)=\mu\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}t^2}dt+\sigma\int_{-\infty}^{\infty}\frac{t}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}t^2}dt$$

Det første integral er 1, da integranden er en tæthedsfunktion, og det andet integral er 0, da integranden er en ulige funktion. Følgelig er $E(X)=\mu$.

3) $V(X)=\int_{-\infty}^{\infty}\frac{(x-\mu)^2}{\sqrt{2\cdot\pi\sigma}}\cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$. Substitueres $t=\frac{x-\mu}{\sigma}\Leftrightarrow x=\mu+\sigma\cdot t$ fås

$$V(X)=\sigma^2\int_{-\infty}^{\infty}\frac{t^2}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}t^2}dt=\sigma^2\left[-\frac{t}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}t^2}\right]_{-\infty}^{\infty}+\sigma^2\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{1}{2}t^2}dt=\sigma^2$$

(delvis integration , samt som før at sidste integral er 1) .

✡

SUPPLEMENT 3A χ^2 -FORDELINGEN

Definition af χ^2 -fordelingen. *Lad $U_1, U_2, \ldots, U_f$ være uafhængige normerede normalfordelte variable. Sandsynlighedsfordelingen for den statistiske variabel $\chi^2 = U_1^2 + U_2^2 + , \ldots , U_f^2$ kaldes χ^2 - fordelingen med frihedsgradstallet f og betegnes $\chi^2(f)$.*

Det kan vises, at tæthedsfunktionen for $\chi^2(f)$ er bestemt ved

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{f}{2}} \cdot \Gamma\left(\frac{f}{2}\right)} x^{\frac{f}{2}-1} e^{-\frac{x}{2}} & \text{for} \quad x > 0 \\ 0 & \text{ellers} \end{cases}$$

hvor Gammafunktionen $\Gamma(x)$ er defineret ved $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ [1]

SÆTNING 3A.1 *(middelværdi og varians for χ^2 -fordeling).*
χ^2 - fordelingen med f frihedsgrader har middelværdien $E(\chi^2) = f$ og variansen $V(\chi^2) = f$.

Bevis: I kapitel 1 blev det for en vilkårlig statistisk variabel X vist, at

$V(X) = E(X^2) - (E(X))^2$. (1)

Lad U være den normerede normalfordeling.

Da $E(U) = 0$ og $V(U) = 1$ fås af regel (1), at $E(U^2) = 1$.

Idet $\chi^2 = U_1^2 + U_2^2 + \ldots + U_f^2$ fås af linearitetsreglen, at

$E(\chi^2) = E(U_1^2) + E(U_2^2) + \ldots + E(U_f^2) = 1 + 1 + \ldots + 1 = f$

Benyttes regel (1) på den statistiske variabel U^2 fås $V(U^2) = E(U^4) - (E(U^2))^2$.

Af definitionen på middelværdi fås

$$E(U^4) = \int_{-\infty}^{\infty} u^4 \cdot f(u) du = \tfrac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} u^4 \cdot \mathrm{e}^{-\frac{1}{2} \cdot u^2} du = -\tfrac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} u^3 \cdot \mathrm{e}^{-\frac{1}{2} \cdot u^2} d\left(-\tfrac{1}{2} u^2\right)$$

Benyttes delvis integration fås:

$$E(U^4) = -\tfrac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} u^3 \cdot d\left(\mathrm{e}^{-\frac{1}{2} \cdot u^2}\right) = -\tfrac{1}{\sqrt{2 \cdot \pi}} \left(\left[u^3 \cdot \mathrm{e}^{-\frac{1}{2} \cdot u^2} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2} \cdot u^2} du^3 \right)$$

$$= -\tfrac{1}{\sqrt{2 \cdot \pi}} \left(0 - 3 \int_{-\infty}^{\infty} u^2 \cdot \mathrm{e}^{-\frac{1}{2} \cdot u^2} du \right) = 3 \tfrac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\infty} u^2 \cdot \mathrm{e}^{-\frac{1}{2} \cdot u^2} du = 3 \cdot E(U^2) = 3$$

Vi har nu $V(U^2) = E(U^4) - (E(U^2))^2 = 3 - 1 = 2$

Da $U_1, U_2, \ldots U_f$ er uafhængige fås $V(\chi^2) = V(U_1^2) + V(U_2^2) + \ldots + V(U_f^2) = 2 + 2 + \ldots + 2 = 2 \cdot f$

✸

---

[1] Specielt gælder $\Gamma(x) = \begin{cases} (x-1)! & \text{for} \quad x \in \{1,2,3,\ldots\} \\ \sqrt{\pi} & \text{for} \quad x = \frac{1}{2} \\ (x-1) \cdot (x-2) \cdot \ldots \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} & \text{for} \quad x \in \{\frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \ldots\} \end{cases}$

SÆTNING 3A.2 *(additionssætning for χ^2 - fordelte variable). Lad $\chi_1^2, \chi_2^2, \ldots, \chi_k^2$ være uafhængige χ^2 - fordelte variable med frihedsgradstallene henholdsvis $f_1, f_2, \ldots, f_k$.*
Den statistiske variabel $\chi^2 = \chi_1^2 + \chi_2^2 + \ldots + \chi_k^2$ er da χ^2 - fordelt med frihedsgradstallet $f = f_1 + f_2 + \ldots + f_k$

Bevis:

Da hver af de k χ^2 - fordelte variable er en sum af kvadraterne på en række normerede uafhængige noralfordelte variable, vil χ^2 naturligvis også være det. Det samlede antal led i χ^2 er f, hvormed sætningen er bevist.

✸

Vi vil i de følgende sætninger gå ud fra, at $X_1, X_2, \ldots, X_n$ er normalfordelte statistiske variable med middelværdi μ og spredning σ. Lad endvidere som sædvanlig

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}, \qquad S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \quad \text{og} \quad S_\mu^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}.$$

Uden bevis anføres følgende sætning:

SÆTNING 3A.3 (Uafhængighed af $\overline{X}$ og S^2). *De statistiske variable $\overline{X}$ og S^2 er uafhængige.*

SÆTNING 3A.4 ($\frac{(n-1)\cdot S^2}{\sigma^2}$ er χ^2 - fordelt).

Den statistiske variabel $\frac{(n-1)\cdot S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}$ *er fordelt* $\chi^2(n-1)$.

Bevis: Idet

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}((X_i - \mu) + (\mu - \overline{X}))^2 = \sum_{i=1}^{n}((X_i - \mu)^2 + (\mu - \overline{X})^2 + 2(X_i - \mu)(\mu - \overline{X}))$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{n}(\mu - \overline{X})^2 + 2(\mu - \overline{X})\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n}(X_i - \mu)^2 + n(\mu - \overline{X})^2 + 2(\mu - \overline{X})\left(\sum_{i=1}^{n}X_i - n\mu\right)$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 + n(\mu - \overline{X})^2 + 2(\mu - \overline{X})(n\overline{X} - n\mu) = \sum_{i=1}^{n}(X_i - \mu)^2 - n\cdot(\overline{X} - \mu)^2$$

fås
$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} - \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \qquad (1)$$

Idet $\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=\sum_{i=1}^{n}\left(\dfrac{X-\mu}{\sigma}\right)^2$ er en sum af kvadrerede normerede normalfordelte variable, er den χ^2 - fordelt med n frihedsgrader.

Endvidere ses, at $\left(\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$ er en kvadreret normeret normalfordelt variabel, så den er χ^2 - fordelt med 1 frihedsgrad.

De to led i (1) er ifølge sætning 3A.3 uafhængige χ^2 - fordelte variable statistiske variable.
Da en sum af uafhængige χ^2 - fordelte variable atter er χ^2 - fordelte ifølge sætning 3A.2, har vi derfor, at
$\dfrac{(n-1)\cdot S^2}{\sigma^2}=\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}-\left(\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$ er χ^2 - fordelt.

Det kan endvidere vises, at f = n - 1.

✸

SÆTNING 3A.5 (*varians af* S^2 *og* S_μ^2). $V(S^2)=\dfrac{2\cdot\sigma^4}{n-1}$ *og* $V(S_\mu^2)=\dfrac{2\cdot\sigma^4}{n}$.

Bevis:

a) Af omskrivningen $S^2=\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}=\dfrac{\sigma^2}{n-1}\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}=\dfrac{\sigma^2}{n-1}\chi^2$ hvor χ^2 er fordelt $\chi^2(n-1)$, fås

$$V(S^2)=\left(\frac{\sigma^2}{n-1}\right)^2V(\chi^2)=\left(\frac{\sigma^2}{n-1}\right)^2\cdot 2\cdot(n-1)=\frac{2\cdot\sigma^4}{n-1}$$

b) Af omskrivningen $S_\mu^2=\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}=\dfrac{\sigma^2}{n}\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}=\dfrac{\sigma^2}{n}\chi^2$ hvor χ^2 er fordelt $\chi^2(n)$, fås

$$V(S_\mu^2)=\left(\frac{\sigma^2}{n}\right)^2V(\chi^2)=\left(\frac{\sigma^2}{n}\right)^2\cdot 2\cdot n=\frac{2\cdot\sigma^4}{n}.$$

✸

SUPPLEMENT 3B *t* - FORDELINGEN

DEFINITION af *t* - fordelingen. *Lad U være normalfordelt n (0.1) og* χ^2 *være fordelt* $\chi^2(f)$. *Hvis U og* χ^2 *er uafhængige kaldes sandsynlighedsfordelingen for den statistiske variabel* $t = \dfrac{U}{\sqrt{\dfrac{\chi^2}{f}}}$ *for t-fordelt med frihedsgradstallet f og betegnes t(f).*

Det kan vises, at tæthedsfunktionen for *t*(*f*) er bestemt ved $f(x) = \dfrac{\Gamma\left(\dfrac{f+1}{2}\right)}{\sqrt{\pi \cdot f} \cdot \Gamma\left(\dfrac{f}{2}\right)} \cdot \left(1 + \dfrac{x^2}{f}\right)^{-\frac{f+1}{2}}$

Da $f(x) = f(-x)$ er *t* - fordelingerne alle symmetriske om y aksen , med $E(X) = 0$.

For de statistiske anvendelser er følgende sætning vigtig:

SÆTNING 3B.1 ($\dfrac{\overline{X}-\mu}{\dfrac{S}{\sqrt{n}}}$ er t-fordelt).

Lad $X_1, X_2, \ldots, X_n$ *være normalfordelte statistiske variable med middelværdi* μ *og spredning* σ. *Lad endvidere* $\overline{X}$ *og* S^2 *være de sædvanlige estimater for middelværdi og varians.*

Den statistiske variabel $T = \dfrac{\overline{X}-\mu}{\dfrac{S}{\sqrt{n}}}$ *er da t - fordelt* $t(n-1)$.

Bevis: Vi har $T = \dfrac{\overline{X}-\mu}{\dfrac{S}{\sqrt{n}}} = \dfrac{\dfrac{(\overline{X}-\mu)\cdot\sqrt{n}}{\sigma}}{\dfrac{S}{\sigma}} = \dfrac{\dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}}$

Her er tælleren $\dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$ normeret normalfordelt, mens $\dfrac{(n-1)S^2}{\sigma^2}$ er χ^2 - fordelt med $f = n - 1$ frihedsgrader. Ifølge sætning 3A.3 er $\overline{X}$ og S^2 uafhængige, hvilket gør, at betingelserne er opfyldt for, at *T* er *t* - fordelt.

✡

SUPPLEMENT 3C F-FORDELINGEN

DEFINITION af *F*-fordelingen. *Lad* χ_T^2 *og* χ_N^2 *være statistisk uafhængige og fordelt henholdsvis* $\chi^2(f_T)$ *og* $\chi^2(f_N)$. *Sandsynlighedsfordelingen for den statistiske variabel* $F = \dfrac{\frac{\chi_T^2}{f_T}}{\frac{\chi_N^2}{f_N}}$ *siges at være F-fordelt med tællerfrihedsgradstallet* f_T *og nævnerfrihedsgrads-tallet* f_N *og betegnes* $F(f_T, f_N)$.

Det kan vises, at tæthedsfunktionen for $F(f_T, f_N)$ er bestemt ved

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{f_T + f_N}{2}\right) \cdot f_T^{\frac{f_T}{2}} \cdot x^{\frac{f_T}{2}-1}}{\Gamma\left(\dfrac{f_T}{2}\right) \cdot \Gamma\left(\dfrac{f_N}{2}\right) \cdot f_N^{\frac{f_N}{2}} \cdot \left(1 + \dfrac{f_N}{f_T} x\right)^{\frac{f_T+f_N}{2}}} & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

og har middelværdien $E(F) = \dfrac{f_N}{f_N - 2}$ (for $f_N > 2$)

og spredningen $\sigma(F) = \dfrac{f_N}{f_N - 2} \cdot \sqrt{\dfrac{2 \cdot (f_T + f_N - 2)}{f_T \cdot (f_N - 4)}}$ (for $f_N > 4$)

SÆTNING 3C.1 . (*relationer vedrørende F-fordelingen*).

a) $F_p(f_T, f_N) = \dfrac{1}{F_{1-p}(f_N, f_T)}$ *b)* $F_{1-p}(1, f_N) = \left(t_{1-\frac{p}{2}}(f_N)\right)^2$

Bevis: a) Lad F være fordelt $F(f_T, f_N)$. Der foretages omskrivningen

$$F = \frac{\frac{\chi_T^2}{f_T}}{\frac{\chi_N^2}{f_N}} = \frac{1}{\left(\dfrac{\frac{\chi_N^2}{f_N}}{\frac{\chi_T^2}{f_T}}\right)} = \frac{1}{F^*},$$ hvor F^* er fordelt $F(f_N, f_T)$. Heraf følger:

$$P(F \le x) = p \Leftrightarrow P\left(\frac{1}{F^*} \le x\right) = p \Leftrightarrow P\left(F^* \ge \frac{1}{x}\right) = p \Leftrightarrow P\left(F^* < \frac{1}{x}\right) = 1 - p.$$

Hermed er a) bevist.

b) Lad F være fordelt $F(1, f_N)$ og t være fordelt $t(f_N)$. Der foretages omskrivningen

$$F = \frac{\frac{\chi_1^2}{1}}{\frac{\chi_N^2}{f_N}} = \frac{U^2}{\frac{\chi_N^2}{f_N}} = \left(\frac{U}{\sqrt{\frac{\chi_N^2}{f_N}}}\right)^2 = t^2$$. Heraf følger:

$$P(F \le x) = 1 - p \Leftrightarrow P\left(t^2 \le x\right) = 1 - p \Leftrightarrow P\left(-\sqrt{x} \le t \le \sqrt{x}\right) = 1 - p \Leftrightarrow P\left(t \le \sqrt{x}\right) = 1 - \frac{p}{2}$$

Hermed er b) bevist.

✡

SÆTNING 3C.2 $\left(\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \text{ er } F - \text{fordelt}\right)$

Lad X_1 og X_2 være normalfordelte statistiske uafhængige normalfordelte variable ,lad n_1 og n_2 betegne stikprøvestørrelsen for de to variable, og lad endvidere $\overline{X}_1, \overline{X}_2, S_1^2$ og S_2^2 være de sædvanlige estimater for middelværdierne og varianserne μ_1, μ_2, σ_1^2 og σ_2^2

Den statistiske variabel $F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}$ *er da fordelt F- fordelt* $F(n_1 - 1, n_2 - 1)$.

Bevis

Ved anvendelse af sætning 3A.4 fås følgende omformning:

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{\frac{(n_1 - 1)S_1^2}{\sigma_1^2 (n_1 - 1)}}{\frac{(n_2 - 1)S_2^2}{\sigma_2^2 (n_2 - 1)}} = \frac{\frac{\chi_1^2}{(n_1 - 1)}}{\frac{\chi_2^2}{(n_2 - 1)}}$$

Af definitionen på F- fordelingen følger da, at sætningen er bevist.

✡

Som et vigtigt specialtilfælde ses, at hvis de to variable har samme varians, vil $F = \frac{S_1^2}{S_2^2}$ være F-fordelt $F(n_1 - 1, n_2 - 1)$.

Supplement 4A DEFINITIONER OG EKSEMPLER PÅ CENTRALE OG EFFEKTIVE ESTIMATORER

Central estimator.
Et rimeligt krav til en estimator Φ er, at Φ i middel skal "ramme" den ukendte parameter, dvs. ikke systematisk angive en for lille værdi eller en for stor værdi.

> **DEFINITION af central estimator.** *En punktestimator Φ siges at være central for φ hvis* $E(\Phi) = \varphi$.

Eksempelvis er $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ en central estimator, da vi i kapitel 1 afsnit 1.6 viste, at $E(\overline{X}) = \mu$.

Effektiv estimator.
Blandt de centrale estimatorer har estimatorer med en lille varians særlig interesse. Har man således 2 centrale estimatorer for den samme parameter, vil man naturligt vælge den med mindst varians. Man siger, at en sådan estimator er mere **effektiv** end den anden.

> **DEFINITION af effektiv estimator.**
> *En central estimator Φ kaldes en effektiv estimator for φ, hvis $V(\Phi) \to 0$ for $n \to \infty$.*

En effektiv estimator vil altså for stor stikprøvestørrelse n have både den rigtige middelværdi og lille varians.
Eksempelvis er $\overline{X}$ en effektiv estimator, da vi i kapitel 1 afsnit 1.6 viste, at $\sigma(\overline{X}) = \dfrac{\sigma(X)}{\sqrt{n}}$, og det deraf følger, at $\sigma(\overline{X}) \to 0$ for $n \to \infty$.

Der vil nu blive bevist, at de estimatorer vi tidligere har benyttet, er såvel centrale som effektive.

> **SÆTNING 4A.1.** (*centrale estimatorer*) *Lad X være en statistisk variabel med middelværdi μ og varians σ^2.*
> *Lad $X_1, X_2, \ldots, X_n$, være en stikprøve af størrelsen n fra X.*
>
> *a)* $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ *er en central estimator for middelværdien μ:*
>
> *b)* $S^2 = \displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$ *er en central estimator for variansen σ^2*
>
> *c)* $S_\mu^2 = \displaystyle\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{n}$ *er en central estimator for variansen σ^2.*
>
> *d)* $S_1^2 = \displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n}$ *er ikke en central estimator for variansen σ^2*

Bevis:

a) I kapitel 2 afsnit 2.3 viste vi, at $E(\overline{X}) = \mu$, dvs. $\overline{X}$ er central .

b) Som en del af beviset for sætning 3A.4 viste vi, at $\displaystyle\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n \cdot (\overline{X} - \mu)^2$

Heraf fås $E(S^2) = E\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X_i - \mu)^2 - n\cdot E(\overline{X} - \mu)^2\right)$

$= \frac{1}{n-1}\left(\sum_{i=1}^{n} V(X) - n\cdot V(\overline{X})\right) = \frac{1}{n-1}\left(n\cdot V(X) - n\frac{V(X)}{n}\right) = V(X) = \sigma^2$, dvs. S^2 er central.

c) $$E(S_\mu^2) = E\left(\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n} V(X) = \sigma^2$$

d) $$E(S_\mu^2) = E\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}\right) = \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i - \mu)^2 - n\cdot E(\overline{X} - \mu)^2\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} V(X) - n\cdot V(\overline{X})\right) = \frac{1}{n}\left(n\cdot V(X) - n\frac{V(X)}{n}\right) = \frac{n-1}{n}V(X) = \frac{n-1}{n}\sigma^2$$

✡

SÆTNING 4A.2. (effektive estimatorer) *Lad X være en statistisk variabel med middelværdi μ og varians σ^2.*
Lad $X_1, X_2, \ldots, X_n$, være en stikprøve af størrelsen n fra X.

a) $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ *er en effektiv estimator for middelværdien μ :*

b) $S^2 = \sum_{i=1}^{n}\frac{(X_i - \overline{X})^2}{n-1}$ er en central estimator for σ^2 *forudsat X er normalfordelt*

c) $S_\mu^2 = \sum_{i=1}^{n}\frac{(X_i - \mu)^2}{n}$ *er en effektiv estimator for σ^2 forudsat X er normalfordelt og μ er kendt.*

d) *Forudsat μ er kendt vil S_μ^2 være en mere effektiv estimator end S^2 for variansen σ^2.*

Bevis: a) I kapitel 1 afsnit 1.6 blev det bevist, at $\sigma(\overline{X}) = \frac{\sigma(X)}{\sqrt{n}}$.

Heraf følger at $\sigma(\overline{X}) \to 0$ for $n \to \infty$.

b) I kapitel 3 sætning 3A.5 bliver det bevist, at $V(S^2) = \frac{2\cdot\sigma^4}{n-1}$.

Heraf følger, at $\sigma(S^2) \to 0$ for $n \to \infty$.

c) I kapitel 39 sætning 3A.5 bliver det bevist, at $V(S_\mu^2) = \frac{2\cdot\sigma^4}{n}$.

Heraf følger, at $\sigma(S_\mu^2) \to 0$ for $n \to \infty$.

d) Da $V(S_\mu^2) = \frac{2\cdot\sigma^4}{n} < \frac{2\cdot\sigma^4}{n-1} = V(S^2)$ har S_μ^2 en mindre varians end S^2 .

✡

Supplement 4B Beviser for formler for konfidensintervaller

Vi vil i dette supplement bevise de 4 første formler for konfidensintervaller der findes i appendix 4.1

Lad i den følgende sætning X være en normalfordelt statistisk variable med middelværdi μ og spredning σ, og lad $X_1, X_2, \ldots, X_n$ være en stikprøve af størrelsen n fra X. Lad endvidere som sædvanlig

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}, \quad S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1} \text{ og } S_\mu^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{n}.$$

SÆTNING 4 B.1 ($100 \cdot \beta$ *% konfidensintervaller for normalfordelte variable*). *Lad* $\alpha = 1 - \beta$

1) *Konfidensinterval for* μ, *hvis* σ ***er kendt***: $\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$,

   hvor $u_{1-\frac{\alpha}{2}}$ *er* $1 - \frac{1}{2}\alpha$ *fraktilen i den normerede normalfordeling.*

2) *Konfidensinterval for* μ, *hvis* σ ***er ukendt***: $\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s}{\sqrt{n}}$,

   hvor $t_{1-\frac{\alpha}{2}}(n-1)$ *er* $1 - \frac{1}{2}\alpha$ *fraktilen i t-fordelingen med n - 1 frihedsgrader.*

3) *Konfidensinterval for* σ^2, *hvis* μ ***er kendt***:

$$\frac{(n-1)s^2 + n \cdot (\bar{x} - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \le \sigma^2 \le \frac{(n-1)s^2 + n \cdot (\bar{x} - \mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)},$$

   hvor $\chi^2_{\frac{\alpha}{2}}(n)$ *og* $\chi^2_{1-\frac{\alpha}{2}}(n)$ *er henholdsvis* $\frac{1}{2}\alpha$ *og* $1 - \frac{1}{2}\alpha$ *fraktilen i* χ^2 *- fordelingen med n frihedsgrader.*

4) *Konfidensinterval for* σ^2, *hvis* μ ***er ukendt***: $\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$,

   hvor $\chi^2_{\frac{\alpha}{2}}(n-1)$ *og* $\chi^2_{1-\frac{\alpha}{2}}(n-1)$ *er henholdsvis* $\frac{1}{2}\alpha$ *og* $1 - \frac{1}{2}\alpha$ *fraktilen i* χ^2 *- fordelingen med n - 1 frihedsgrader.*

Bevis:

1) Er σ kendt vil $U = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ være normeret normalfordelt $n(\mu, \sigma)$. Da U - fordelingen er symmetrisk omkring 0,

har vi følgelig, at $P\left(-u_{1-\frac{\alpha}{2}} \le U \le u_{1-\frac{\alpha}{2}}\right) = \beta$ eller $P\left(-u_{1-\frac{\alpha}{2}} \le \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le u_{1-\frac{\alpha}{2}}\right) = \beta$

Idet $-u_{1-\frac{\alpha}{2}} \le \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le u_{1-\frac{\alpha}{2}} \Leftrightarrow -u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \overline{X} - \mu \le u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$\Leftrightarrow \overline{X} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ er punkt 1 bevist.

2) I kapitel 3 anførte vi, at $T = \dfrac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}}$ er t - fordelt med $f = n$ - 1 frihedsgrader. De samme regninger som under punkt 1 kan nu gennemføres, hvilket er sket ved beviset for sætning 4.1. Hermed er punkt 2 bevist.

3) Ifølge beviset for sætning 3A.4 er $S_\mu^2 = \sum_{i=1}^{n} \dfrac{(X_i - \mu)^2}{n}$ χ^2 - fordelt med $f = n$ frihedsgrader.

Vi har følgelig at $P\left(\chi^2_{\frac{\alpha}{2}}(n) \le S_\mu^2 \le \chi^2_{1-\frac{\alpha}{2}}(n)\right) = \beta$ eller $P\left(\chi^2_{\frac{\alpha}{2}}(n) \le \dfrac{\sum_{i=1}^{n}(X_1 - \mu)^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}(n)\right) = \beta$.

Igen ifølge beviset for sætning 3A.4 er $\sum_{i=1}^{n}(X_i - \mu)^2 = (n-1)\cdot S^2 + n(\overline{X} - \mu)^2$.

Indsættes dette i ovenstående ulighed fås: $\chi^2_{\frac{\alpha}{2}}(n) \le \dfrac{(n-1)S^2 + n(\overline{X} - \mu)^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}(n)$

$\Leftrightarrow \sigma^2 \cdot \chi^2_{\frac{\alpha}{2}}(n) \le (n-1)S^2 + n(\overline{X} - \mu)^2 \le \sigma^2 \cdot \chi^2_{1-\frac{\alpha}{2}}(n)$

$\Leftrightarrow \dfrac{(n-1)S^2 + n(\overline{X} - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \cdot \le \sigma^2 \le \dfrac{(n-1)S^2 + n(\overline{X} - \mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}$.

Hermed er formel 3 bevist.

4) Ifølge beviset for sætning 3A.4 er $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$ χ^2 - fordelt med $f = n$ - 1 frihedsgrader.

Vi har følgelig at

$P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \le \chi^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = \beta$ eller $P\left(\chi^2_{\frac{\alpha}{2}}(n-1) \le \dfrac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) = \beta$.

Idet $\chi^2_{\frac{\alpha}{2}}(n-1) \le \dfrac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}(n-1)$

$\Leftrightarrow \sigma^2 \cdot \chi^2_{\frac{\alpha}{2}}(n-1) \le (n-1)S^2 \le \sigma^2 \cdot \chi^2_{1-\frac{\alpha}{2}}(n-1)$

$\Leftrightarrow \dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \cdot \le \sigma^2 \le \dfrac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$ er formel 4 bevist.

SUPPLEMENT 5A. BEVISER FOR FORMLER I HYPOTESETEST

Lad X være en normalfordelt statistisk variable med middelværdi μ og spredning σ, og lad $X_1, X_2, \ldots, X_n$ være en stikprøve af størrelsen n fra X. Lad endvidere som sædvanlig $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ og $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$.

Lad testene alle have signifikansniveauet α.

Testformler i appendix 5.1 (σ ukendt og erstattes af s).

I sætning 3B viste vi, at såfremt $H_0: \mu = \mu_0$ er sand, er $T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ t- fordelt med $f = n - 1$ frihedsgrader. Dette benyttes i det følgende

Række 1: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu > \mu_0$.
Af ovenstående og af definitionen på signifikansniveau fremgår, at H_0 forkastes, såfremt $p = P(T \geq t) < \alpha$, hvor $T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$, og T er t - fordelt med $f = n - 1$ frihedsgrader.

Række 2: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu < \mu_0$.
H_0 forkastes, såfremt $p = P(T \leq t) < \alpha$, hvor $T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$, og T er t-fordelt med $f = n - 1$ frihedsgrader.

Række 3: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu \neq \mu_0$.
Ifølge definitionen på signifikansniveau, forkastes H_0 både når $\overline{x}$ er så stor, at $P(T \geq t) < \frac{1}{2}\alpha$, og når $\overline{x}$ er så lille, at $P(T \leq t) < \frac{1}{2}\alpha$, hvor $T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ og T er t -fordelt med $f = n - 1$ frihedsgrader.

Række 4: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu > \mu_0$.
Da $P(T \geq t) < \alpha \Leftrightarrow 1 - P(T < t) < \alpha \Leftrightarrow P(T < t) > 1 - \alpha \Leftrightarrow t > t_{1-\alpha}(n-1)$ er påstanden bevist.

Række 5: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu < \mu_0$.
Da $P(T \leq t) < \alpha \Leftrightarrow t < t_{\alpha}(n-1) \Leftrightarrow t < -t_{1-\alpha}(n-1)$
(da t-fordelingen er symmetrisk om 0, se figur 5.1) er formlen bevist.

Række 6: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu \neq \mu_0$.
Hvis $\overline{x} \geq \mu_0$ kan regningerne under række 4 gentages, og vi får $t > t_{1-\frac{\alpha}{2}}(n-1)$.
Hvis $\overline{x} < \mu_0$ kan regningerne under række 5 gentages, og vi får $t < -t_{1-\frac{\alpha}{2}}(n-1)$.
Generelt fås derfor, at $|t| > t_{1-\frac{\alpha}{2}}(n-1)$ og dermed er formlen bevist.

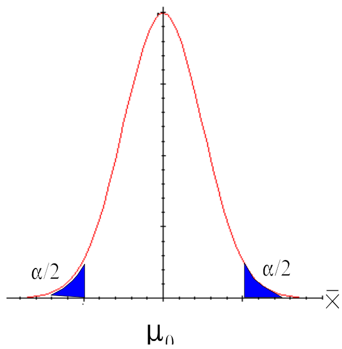

Fig 5.1. t- fordeling

Testformler i appendix 5.2 (σ antages kendt). Lad Y være fordelt $n\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$.

Række 1: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu > \mu_0$.
Ifølge definitionen på signifikansniveau, forkastes H_0 såfremt $p = P(Y \geq \bar{x}) < \alpha$

Række 2: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu < \mu_0$.
Ifølge definitionen på signifikansniveau, forkastes H_0 såfremt $p = P(Y \leq \bar{x}) < \alpha$

Række 3: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu \neq \mu_0$.
Ifølge definitionen på signifikansniveau, forkastes H_0 både når $\bar{x}$ er så stor, at $P(Y \geq \bar{x}) < \frac{1}{2}\alpha$ (se figur 5.1), og når $\bar{x}$ er så lille, at $P(Y \leq \bar{x}) < \frac{1}{2}\alpha$.

Række 4: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu > \mu_0$.

$$P(Y \geq \bar{x}) < \alpha \Leftrightarrow 1 - P(Y < \bar{x}) < \alpha \Leftrightarrow 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) < \alpha \Leftrightarrow \Phi\left(\frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma}\right) > 1 - \alpha \Leftrightarrow u > u_{1-\alpha}$$

hvor $u = \dfrac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma}$.

Række 5: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu < \mu_0$.

$$p = P(Y \leq \bar{x}) < \alpha \Leftrightarrow \Phi\left(\frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma}\right) < \alpha \Leftrightarrow u < u_\alpha \text{ hvor } u = \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma}.$$

Da $u_\alpha = -u_{1-\alpha}$ er påstanden bevist.

Række 6: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu \neq \mu_0$.
Hvis $\bar{x} > \mu_0$ kan regningerne under række 4 gentages, og vi får $u > u_{1-\frac{\alpha}{2}}$

Hvis $\bar{x} < \mu_0$ kan regningerne under række 5 gentages, og vi får $u < -u_{1-\frac{\alpha}{2}}$

Generelt fås derfor, at $|u| > u_{1-\frac{\alpha}{2}}$ og deraf formlen.

Testformler i appendix 5.3 for varians σ^2 (μ ukendt)

I kapitel 3 sætning 3A viste vi, at såfremt $H_0: \sigma^2 = \sigma_0^2$ er sand, er $\chi^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \dfrac{(n-1)\cdot S^2}{\sigma^2}$

χ^2 - fordelt med $f = n$ - 1 frihedsgrader.

I det følgende vil vi endvidere antage, at Q er χ^2 - fordelt med n - 1 frihedsgrader.

Række 1: Nulhypotese $H_0: \sigma^2 = \sigma_0^2$. Alternativ hypotese $H: \sigma^2 > \sigma_0^2$.
Af ovenstående og af definitionen på signifikansnveau fås, at H_0 forkastes, såfremt

$p = P(Q \geq \chi^2) < \alpha$, hvor $\chi^2 = \dfrac{(n-1)\cdot s^2}{\sigma_0^2}$. Heraf følger formlen.

Række 2: Nulhypotese $H_0: \sigma^2 = \sigma_0^2$. Alternativ hypotese $H: \sigma^2 < \sigma_0^2$.
Af ovenstående og af definitionen på signifikansnveau fås, at H_0 forkastes, såfremt $p = P(Q \leq \chi^2) < \alpha$, hvor $\chi^2 = \dfrac{(n-1)\cdot s^2}{\sigma_0^2}$. Heraf følger formlen.

Række 3: Nulhypotese $H_0: \sigma^2 = \sigma_0^2$. Alternativ hypotese $H: \sigma^2 \neq \sigma_0^2$.
Ifølge definitionen på signifikansniveau, forkastes H_0 både når χ^2 er så stor, at $P(Q \geq \chi^2) < \frac{1}{2}\alpha$, og når χ^2 er så lille, at $P(Q \leq \chi^2) < \frac{1}{2}\alpha$, hvor $\chi^2 = \dfrac{(n-1)\cdot s^2}{\sigma_0^2}$. Heraf følger formlen.

Række 4: Nulhypotese $H_0: \sigma^2 = \sigma_0^2$. Alternativ hypotese $H: \sigma^2 > \sigma_0^2$.

$$P(Q \geq \chi^2) < \alpha \Leftrightarrow 1 - P(Q < \chi^2) < \alpha \Leftrightarrow P(Q < \chi^2) > 1 - \alpha \Leftrightarrow \chi^2 > \chi^2_{1-\alpha}(n-1).$$

Række 5: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 < \sigma_0^2$.
Da $P(Q \le \chi^2) < \alpha \Leftrightarrow \chi^2 < \chi^2_{\alpha}(n-1)$ er formlen bevist.

Række 6: Da $E(\chi^2) = f$, hvor $f = n$ - 1 er frihedsgradstallet, fås:

Hvis $\chi^2 > n$ - 1 kan regningerne under række 4 gentages, og vi får $\chi^2 > \chi^2_{1-\frac{\alpha}{2}}(n-1)$.

Hvis $\chi^2 < n$ - 1 kan regningerne under række 5 gentages, og vi får $\chi^2 < \chi^2_{1-\frac{\alpha}{2}}(n-1)$

Hermed fås formlen.

Testformler i appendix 5.4 for varians σ^2 (μ kendt)

I forbindelse med beviset for sætning 3A.4 blev vist, at såfremt $H_0\!:\sigma^2 = \sigma_0^2$ er sand, er $\chi^2 = \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$ χ^2 - fordelt med $f = n$ frihedsgrader. Endvidere fandt vi, at $\sum_{i=1}^{n}(X_i-\mu)^2 = (n-1)\cdot S^2 + n(\overline{X}-\mu)^2$.

I det følgende vil vi endvidere antage, at Q er χ^2 - fordelt med n frihedsgrader.

Række 1: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 > \sigma_0^2$.
Af ovenstående og af definitionen på signifikansnveau fås, at H_0 forkastes, såfremt $p = P(Q \ge \chi^2) < \alpha$,
hvor $\chi^2 = \dfrac{(n-1)\cdot s^2 + n(\overline{x}-\mu)^2}{\sigma_0^2}$. Heraf følger formlen.

Række 2: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 < \sigma_0^2$.
Af ovenstående og af definitionen på signifikansnveau fås, at H_0 forkastes, såfremt
$p = P(Q \le \chi^2) < \alpha$, hvor $\chi^2 = \dfrac{(n-1)\cdot s^2 + n(\overline{x}-\mu)^2}{\sigma_0^2}$. Heraf følger formlen.

Række 3: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 \ne \sigma_0^2$.

Ifølge definitionen på signifikansniveau, forkastes H_0 både når χ^2 er så stor, at $P(Q \ge \chi^2) < \frac{1}{2}\alpha$, og når χ^2 er så lille, at $P(Q \le \chi^2) < \frac{1}{2}\alpha$, hvor $\chi^2 = \dfrac{(n-1)\cdot s^2 + n(\overline{x}-\mu)^2}{\sigma_0^2}$. Heraf følger formlen.

Række 4: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 > \sigma_0^2$.
$P(Q \ge \chi^2) < \alpha \Leftrightarrow 1 - P(Q < \chi^2) < \alpha \Leftrightarrow P(Q < \chi^2) > 1-\alpha \Leftrightarrow \chi^2 > \chi^2_{1-\alpha}(n)$.

Række 5: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 < \sigma_0^2$.
Da $P(Q \le \chi^2) < \alpha \Leftrightarrow \chi^2 > \chi^2_{\alpha}(n)$ er formlen bevist.

Række 6: Nulhypotese $H_0\!:\sigma^2 = \sigma_0^2$. Alternativ hypotese $H\!:\sigma^2 \ne \sigma_0^2$.
Da $E(\chi^2) = f$, hvor $f = n$ er frihedsgradstallet, fås

Hvis $\chi^2 > n$ kan regningerne under række 4 gentages, og vi får $\chi^2 > \chi^2_{1-\frac{\alpha}{2}}(n)$.

Hvis $\chi^2 < n$ kan regningerne under række 5 gentages, og vi får $\chi^2 < \chi^2_{1-\frac{\alpha}{2}}(n)$

Hermed fås formlen.

SUPPLEMENT 5B. OC-KURVER OG DIMENSIONERING.

Forudsætning som appendix 5.2

Række 1: Alternativ hypotese $H: \mu > \mu_0$ (σ antages kendt)

I sætning 5.1 og 5.2 udledes formler for OC-kurve og antal gentagelser *n* i tilfældet, der svarer til dette tilfælde. Som det ses i det følgende forløber de tilsvarende beviser i de øvrige storrækker ganske analogt.

Række 2: Alternativ hypotese $H: \mu < \mu_0$ (σ antages kendt)

OC-kurve:

H_0 accepteres på signifikansniveau α , hvis $\dfrac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \le u_{1-\alpha} \Leftrightarrow \bar{x} \le \mu_0 + u_{1-\alpha} \cdot \dfrac{\sigma}{\sqrt{n}}$.

Er den sande middelværdi μ , er $U = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ normalfordelt med middelværdi 0 og spredning 1.

Vi har derfor, at $P(\text{type II fejl}) = P\left(\bar{X} \ge \mu_0 + u_\alpha \cdot \dfrac{\sigma}{\sqrt{n}}\right) = 1 - P\left(\bar{X} < \mu_0 + u_\alpha \cdot \dfrac{\sigma}{\sqrt{n}}\right) = 1 - \Phi\left(\dfrac{\mu_0 + u_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$

$$= 1 - \Phi\left(u_\alpha + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \Phi\left(-u_{1-\alpha} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right) \quad \text{d.v.s. } a(\mu) = 1 - \Phi\left(-u_{1-\alpha} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right).$$

Dimensionering: Af formlen for OC kurven fås, idet $\Delta = |\mu - \mu_0| = \mu_0 - \mu$ da $\mu < \mu_0$:

$$a(\mu_0 + \Delta) = 1 - \Phi\left(-u_{1-\alpha} + \frac{\Delta}{\frac{\sigma}{\sqrt{n}}}\right) \le \beta, \text{ dvs. } \Phi\left(-u_{1-\alpha} + \frac{\Delta\sqrt{n}}{\sigma}\right) \ge 1-\beta \Leftrightarrow -u_{1-\alpha} + \frac{\Delta\sqrt{n}}{\sigma} \ge u_{1-\beta} \Leftrightarrow \frac{\Delta \cdot \sqrt{n}}{\sigma} \ge u_{1-\alpha} + u_{1-\beta}$$

$$\Leftrightarrow \sqrt{n} \ge \frac{(u_{1-\beta} + u_{1-\alpha}) \cdot \sigma}{\Delta} \Leftrightarrow n \ge \left(\frac{(u_{1-\beta} + u_{1-\alpha}) \cdot \sigma}{\Delta}\right)^2 .$$

Række 3: Nulhypotese $H_0: \mu = \mu_0$ Alternativ hypotese $H_0: \mu \ne \mu_0$ (σ antages kendt).

OC-kurve:

H_0 accepteres på signifikansniveau α , hvis

$$u_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \le u_{1-\frac{\alpha}{2}} \Leftrightarrow \mu_0 + u_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} .$$

Er den sande middelværdi μ , er $U = \dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ normalfordelt med middelværdi 0 og spred-ning 1.Vi har derfor,

$$P(\text{type II fejl}) = P\left(\mu_0 + u_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \le \bar{X} \le \mu_0 + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = P\left(\bar{X} < \mu_0 + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) - P\left(\bar{X} < \mu_0 + u_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$= \Phi\left(\frac{\mu_0 + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{\mu_0 + u_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi\left(u_{1-\frac{\alpha}{2}} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(-u_{1-\frac{\alpha}{2}} + \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

d.v.s. $a(\mu) = \Phi\left(u_{1-\frac{\alpha}{2}} + \dfrac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(-u_{1-\frac{\alpha}{2}} + \dfrac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$.

Dimensionering:

Antages $\mu > \mu_0$ bliver sidste led i formlen for OC-kurven ca 0, og vi får samme formel som under storrække 1.
Antages $\mu < \mu_0$ bliver første led i formlen for OC-kurven ca 1, og vi får samme formel som under storrække 2.
I begge tilfælde er α dog erstattet af $\frac{1}{2}\alpha$
Dimensioneringsformlen bliver derfor den samme som før, hvor blot α er erstattet af $\frac{1}{2}\alpha$.

Dimensionering svarende til forudsætning i appendix 5.2. (spredning ukendt)

En dimensionering af forsøget er en del af forsøgsplanlægningen. Man har derfor endnu ikke foretaget nogle forsøg, og kender derfor ikke stikprøvens spredning s. Kan man ud fra tidligere lignende forsøg vurdere, at spredningen ikke bliver større end σ, må dette bruges. Da dimensioneringen kun afhænger af forholdet $d = \frac{\Delta}{\sigma}$, kan man også nøjes med at angive et skøn for dette forhold.
Ønsker man eksempelvis at kunne opdage selv relativt små forskelle i middelværdien, kan d sættes til et tal mindre end 1, mens ønskes kun at finde relativt store forskelle i middelværdien kan d eksempelvis sættes til 2.
Når man skal finde sandsynligheden for en type II fejl, får man brug for at kende fordelingen af størrelsen

$T = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$ når $H_0 : \mu = \mu_0$ er falsk

Hvis den sande værdi af middelværdien er $\mu = \mu_0 + \Delta$, kan T skrives

$$T = \frac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}} = \frac{\dfrac{(\overline{X} - (\mu_0 + \Delta))\sqrt{n}}{\sigma}}{\dfrac{S}{\sigma}} = \frac{U + \dfrac{\Delta\sqrt{n}}{\sigma}}{W}$$

Da U er normeret normalfordelt, og $\frac{\Delta\sqrt{n}}{\sigma}$ er en konstant forskellig fra nul, er tælleren normalfordelt med middelværdi $\frac{\Delta\sqrt{n}}{\sigma}$ og spredning 1. $\frac{(n-1)\cdot S^2}{\sigma^2} = (n-1)\cdot W^2$ er χ^2 fordelt med n - 1 frihedsgrader.

Den resulterende fordeling for T kaldes den ikke-centrale t - fordeling med n - 1 frihedsgrader. Denne fordeling har en yderst besværlig tæthedsfunktion, så en nærmere beregning af P(type II fejl) og dimensionering af forsøg må sædvanligvis foretages ved benyttelse af edb. I nedennævnte notat er i et appendix er der således gengivet et Maple-program, som kan foretage disse beregninger. For ofte forekommende værdier af parametrene er der i tabel 8 udarbejdet en "dimensioneringstabel".

Dimensionering svarende til forudsætning i appendix 5.3 til 5.6

Disse begreber vil ikke blive gennemgået her, men man kan eksempelvis i Statgraphics under menupunktet "Describe", og derefter "Sample Size Determination" få et forslag til dimensionering af forsøget.

SUPPLEMENT: 7A Bevis for middelværdi og spredning for hypergeometrisk variabel.

Beviset for sætning 7.3 forudsætter, at man først læser kapitel 9.

SÆTNING 7.3. *(middelværdi og spredning for den hypergeometriske fordeling)*

Den hypergeometriske fordeling h (N, M, n) har

$E(X) = n \cdot p$ og $V(X) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}$, *hvor* $p = \frac{M}{N}$

Bevis: Lad os betragte n statistiske variable $X_1, X_2, \ldots, X_n$,

hvor $X_i = \begin{cases} 1 & \text{hvis i'te udtrækning giver sort kugle.} \\ 0 & \text{ellers} \end{cases}$

$X = X_1 + X_2 + \ldots + X_n$ er hypergeometrisk fordelt h (N, M, n) .

Vi har $P(X_i) = \frac{\binom{M}{1}}{\binom{N}{1}} = \frac{M}{N} = p$ for alle i. Heraf følger $E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$, og

$V(X_i) = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p - p^2 = p \cdot (1-p)$.

Af linearitetsreglen i kapitel 1afsnit 6, fås

$E(X) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n) = p + p + p + \ldots + p = n \cdot p$.

Som det fremgår af afsnit 9.2 gælder $V(X_i, X_j) = E(X_i \cdot X_j) - E(X_i) \cdot E(X_j)$

$= 1 \cdot 1 \cdot P((X_i = 1) \wedge (X_j = 1)) - p \cdot p = P(X_i) \cdot P(X_j | X_i) - p^2$

$= \frac{M}{N} \cdot \frac{M-1}{N-1} - p^2 = p \cdot \left(\frac{M-1}{N-1} - p \right) = p \cdot \left(\frac{M-1}{N-1} - \frac{M}{N} \right) = p \cdot \frac{M-N}{N \cdot (N-1)} = p \cdot \frac{p-1}{N-1}$

Af kvadratreglen (afsnit 9.3):

$$V(X) = \sum_{i=1}^{n} V(X_i) + 2 \cdot \sum_{i<j}^{n} V(X_i, X_j) = n \cdot p \cdot (1-p) + 2 \cdot \binom{n}{2} \cdot p \cdot \frac{p-1}{N-1}$$

$$= n \cdot p \cdot (1-p) + 2 \cdot \frac{n \cdot (n-1)}{2} \cdot p \cdot \frac{1-p}{N-1} = n \cdot p \cdot (1-p) \cdot \left(1 - \frac{n-1}{N-1} \right) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}$$

✡

SUPPLEMENT: 7B BEVISER FOR FORMLER I BINOMIAL- OG POISSON-TEST

Testformler i appendix 5.5 for parameter p i binomialfordeling.

Lad X være en binomialfordelt variabel $b($ n, $p)$, hvor n er kendt og p ukendt. Lad x være en stikprøveværdi på X.
Lad Y være fordelt b (n, p_0), hvor p_0 er en given konstant. Det erindres om, (se eventuelt appendix 7.1), at for $\frac{1}{10} \le p_0 \le \frac{9}{10}$ og $5 \le n \cdot p_0 \le n-5$ kan man approksimere binomialfordelingen $b(n, p_0)$ med normalfordelingen $n(\mu, \sigma)$,hvor $\mu = n \cdot p_0$ og $\sigma = \sqrt{n \cdot p_0 \cdot (1-p_0)}$.

Række 1: Nulhypotese $H_0 : p = p_0$. Alternativ hypotese $H : p > p_0$,

Ifølge definitionen på signifikansniveau, forkastes H_0 , såfremt $p = P(Y \ge x) < \alpha$

Række 2: Nulhypotese $H_0 : p = p_0$. Alternativ hypotese $H : p < p_0$,

Ifølge definitionen på signifikansniveau, forkastes H_0 såfremt $p = P(Y < x) < \alpha$

Række 3: Nulhypotese $H_0 : p = p_0$. Alternativ hypotese $H : p \ne p_0$.

Ifølge definitionen på signifikansniveau, forkastes H_0 både når x er så stor, at $P(Y \ge x) < \frac{\alpha}{2}$, og når x er så lille, at $P(Y \le x) < \frac{\alpha}{2}$ .Da $E(Y) = n \cdot p_0$ følger heraf formlen.

Række 4: Nulhypotese $H_0 : p = p_0$. Alternativ hypotese $H : p > p_0$,
Er betingelserne for approksimation med normalfordelingen opfyldt, gælder

$$P(Y \ge x) < \alpha \Leftrightarrow 1 - P(Y < x) < \alpha \Leftrightarrow 1 - \Phi\left(\frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1-p_0)}}\right) < \alpha$$

$$\Leftrightarrow \Phi\left(\frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1-p_0)}}\right) > 1 - \alpha \Leftrightarrow u > u_{1-\alpha}, \quad \text{hvor } u = \frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1-p_0)}}$$

Række 5: Nulhypotese $H_0 : p = p_0$. Alternativ hypotese $H : p < p_0$,
Er betingelserne for approksimation med normalfordelingen opfyldt, gælder

$$P(Y \le x) < \alpha \Leftrightarrow \Phi\left(\frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1-p_0)}}\right) < \alpha \Leftrightarrow u < u_\alpha \Leftrightarrow u < -u_{1-\alpha},$$

$$\text{hvor } u = \frac{x - \frac{1}{2} - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1-p_0)}}$$

Række 6: Nulhypotese $H_0 : p = p_0$. Alternativ hypotese $H : p \ne p_0$.

Er betingelserne for approksimation med normalfordelingen opfyldt, haves
Hvis $x \ge n \cdot p_0$ kan regningerne under række 4 gentages, og vi får $u > u_{1-\frac{\alpha}{2}}$.

Hvis $x < n \cdot p_0$ kan regningerne under række 5 gentages, og vi får $u < -u_{1-\frac{\alpha}{2}}$

Generelt fås derfor, at $|u| > u_{1-\frac{\alpha}{2}}$ og dermed formlen.

Testformler i appendix 5.6 for parameter μ i Poissonfordeling.

Lad X være en Poissonfordelt variabel $p(\mu)$, hvor μ er ukendt. Lad der foreligge en stikprøve på X af størrelsen n med gennemsnit $\bar{x}$. Lad Y være fordelt $p(n \cdot \mu_0)$, hvor μ_0 er en given konstant. Det erindres om, (se eventuelt appendix 7.1), at for $n \cdot \mu_0 \geq 10$ kan man approksimere Poissonfordelingen $p(n \cdot \mu_0)$ med normalfordelingen $n(n \cdot \mu_0, \sqrt{n \cdot \mu_0})$.

Række 1: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu > \mu_0$
Ifølge definitionen på signifikansniveau, forkastes H_0, såfremt $p = P(Y \geq n \cdot \bar{x}) < \alpha$

Række 2: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu < \mu_0$,
Ifølge definitionen på signifikansniveau, forkastes H_0 såfremt $p = P(Y \leq n \cdot \bar{x}) < \alpha$

Række 3: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu \neq \mu_0$.
Da $E(Y) = n \cdot \mu_0$ følger heraf formlerne i denne række.

Række 4: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu > \mu_0$
Er betingelserne for approksimation med normalfordelingen opfyldt, gælder

$$P(Y \geq n \cdot \bar{x}) < \alpha \Leftrightarrow 1 - P(Y < n \cdot \bar{x}) < \alpha \Leftrightarrow 1 - \Phi\left(\frac{n \cdot \bar{x} - \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}}\right) < \alpha$$

$$\Leftrightarrow \Phi\left(\frac{n \cdot \bar{x} - \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}}\right) > 1 - \alpha \Leftrightarrow u > u_{1-\alpha}, \quad \text{hvor } u = \frac{n \cdot \bar{x} - \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}}.$$

Række 5: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu < \mu_0$,

Er betingelserne for approksimation med normalfordelingen opfyldt, gælder

$$P(Y \leq n \cdot \bar{x}) < \alpha \Leftrightarrow \Phi\left(\frac{n \cdot \bar{x} + \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}}\right) < \alpha \Leftrightarrow u < u_\alpha \Leftrightarrow u < -u_{1-\alpha} \text{ hvor } u = \frac{n \cdot \bar{x} + \frac{1}{2} - n \cdot \mu_0}{\sqrt{n \cdot \mu_0}}.$$

Række 3: Nulhypotese $H_0: \mu = \mu_0$. Alternativ hypotese $H: \mu \neq \mu_0$.
Er betingelserne for approksimation med normalfordelingen opfyldt, haves
Hvis $\bar{x} \geq \mu_0$ kan regningerne under række 4 gentages, og vi får $u > u_{1-\frac{\alpha}{2}}$.

Hvis $\bar{x} < \mu_0$ kan regningerne under række 5 gentages, og vi får $u < -u_{1-\frac{\alpha}{2}}$

Generelt fås derfor, at $|u| > u_{1-\frac{\alpha}{2}}$ og dermed formlen.

SUPPLEMENT: 7C BEVISSKITSE FOR TÆTHEDSFUNKTIONEN FOR POISSONFORDELINGEN.

SÆTNING 7.2 (*Poissonfordeling*). *Lad X være en stokastisk variabel, som angiver antallet af impulser i et givet tidsrum (eller areal, volumen, produktionsenhed osv.), idet ethvert tidspunkt i tidsrummet har samme mulighed for at være impulstidspunkt som ethvert andet tidspunkt. Endvidere skal impulserne indtræffe tilfældigt og uafhængigt af hinanden *).*

Hvis det gennemsnitlige antal impulser i tidsrummet er $\mu > 0$ *, så siges X at være Poissonfordelt p (* μ *) med sandsynlighedsfordelingen (tæthedsfunktionen) f(x) = P(X = x) bestemt ved*

$$f(x) = \begin{cases} \dfrac{\mu^x}{x!} \cdot \mathrm{e}^{-\mu} & \text{for} \quad x \in \{0,1,2,\dots\} \\ 0 & \text{ellers} \end{cases}$$

Middelværdien for p(μ *) er E (X) =* μ *og spredningen er* $\sigma(X) = \sqrt{\mu}$.

I formuleringen af de ovennævnte betingelser kan efter behov "et lille tidsrum Δt*" erstattes med "en lille længde* $\Delta \ell$*", "et lille areal* ΔA*" eller "et lille volumen* ΔV*".*

*) Præcis formulering: Følgende 3 betingelser skal være opfyldt:

1) Sandsynligheden for netop én impuls i et meget lille tidsrum Δt er med tilnærmelse proportional med Δt .
   (Matematisk formulering $\lim\limits_{\Delta t \to 0} \dfrac{P(X=1)}{\Delta t} = \lambda$ (λ er en positiv konstant)
2) Sandsynligheden for 2 eller flere impulser i det meget lille tidsrum Δt er lille sammenlignet med Δt .
   (Matematisk formulering $\lim\limits_{\Delta t \to 0} \dfrac{P(X>1)}{\Delta t} = 0$)
3) Antal impulser i forskellige, ikke overlappende tidsrum er statistisk uafhængige.

Bevisskitse:

Lad μ være det gennemsnitlige antal impulser i tidsrummet [0 ; *T*], og lad *X* være det aktuelle antal impulser i samme tidsrum.

Intervallet [0 ;*T*] opdeles i *n* delintervaller $\Delta t = \dfrac{T}{n}$. Ifølge forudsætning 2 er det muligt at vælge Δt så lille, at sandsynligheden for at mere end 1 impuls indtræffer i Δt sekunder er praktisk taget 0. I Δt sekunder kan derfor kun ske 2 ting:

A: netop 1 impuls , eller $\overline{A}$: netop 0 impulser.

X er derfor binomialfordelt *b* (*n*, *P*(*A*)).

Da der i *T* sekunder i gennemsnit er μ impulser, vil der i Δt sekunder være $\dfrac{\mu}{T} \cdot \Delta t = \dfrac{\mu}{T} \cdot \dfrac{T}{n} = \dfrac{\mu}{n}$ impulser.

Vi har derfor $P(A) = \dfrac{\mu}{n}$. Vi foretager nu følgende omformning:

$$P(X=x) = \binom{n}{x}\left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x} = \frac{n\cdot(n-1)\cdot\ldots\cdot(n-x+1)}{x!}\,\frac{\mu^x}{n^x}\cdot\frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^x}$$

$$= \frac{\mu^x}{x!} \frac{n\cdot(n-1)\cdot\ldots\cdot(n-x+1)}{n^x} \cdot \frac{1}{\left(1-\frac{\mu}{n}\right)^x} \cdot e^{n\cdot\ln\left(1-\frac{\mu}{n}\right)}$$

$$= \frac{\mu^x}{x!} \cdot \left(1-\frac{1}{n}\right) \cdot \left(1-\frac{2}{n}\right) \cdot \left(1-\frac{3}{n}\right) \cdot\ldots\cdot \left(1-\frac{x-1}{n}\right) \cdot \frac{1}{\left(1-\frac{\mu}{n}\right)^x} \cdot \mathrm{e}^{\frac{\ln\left(1-\frac{\mu}{n}\right)}{\frac{1}{n}}}$$

For $n \to \infty$ vil tæller og nævner i brøken $\frac{\ln\left(1-\frac{\mu}{n}\right)}{\frac{1}{n}}$ gå mod 0. Man kan derfor bruge l'Hospitals regel.

Vi får ved differentiation af tæller og nævner: $\frac{1}{1-\frac{\mu}{n}} \cdot \frac{\frac{\mu}{n^2}}{\left(-\frac{1}{n^2}\right)} = \frac{1}{1-\frac{\mu}{n}} \cdot (-\mu) \to -\mu$ for $n \to \infty$

Vi har derfor, at $P(X=x) \to \frac{\mu^x}{x!} \cdot 1 \cdot 1 \cdot\ldots\cdot 1 \cdot \frac{1}{1}\mathrm{e}^{-\mu} = \frac{\mu^x}{x!} \cdot \mathrm{e}^{-\mu}$ for $n \to \infty$.
Idet binomialfordelingen har middelværdien $E(X) = n \cdot p = \mu$ og spredningen $\sigma(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{\mu \cdot \left(1-\frac{\mu}{n}\right)}$, fås for $n \to \infty$, at Poissonfordelingen har middelværdien μ og spredningen $\sqrt{\mu}$.