TMA 321 $\boxed{F13}$ ①

Centrala gränsvärdessatsen (CGS, CLT)

Def: Låt $X_1, X_2, \ldots$ vara en sekvens av s.v. med ffkn $F_1, F_2, \ldots$ och låt X vara en s.v. med ffkn F. Vi säger att X_n konvergerar mot X i fördelning om

$$\lim_{n\to\infty} F_n(x) = F(x)$$

$\forall\ x$ där F är kontinuerlig.

Anm: 1) Vi har $\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x)$

2) Kom ihåg: $X_n \sim Bin(n, p_n)$ där $np_n \to \lambda$ då $n \to \infty$. Vi fick att

$$\mathbb{P}(X_n = k) \to \mathbb{P}(X = k) \quad \text{där } X \sim Poi(\lambda).$$

Vi får då att

$$F_n(x) = \mathbb{P}(X_n \le x) = \sum_{k \le x} \mathbb{P}(X_n = k) \to \sum_{k\le x} \mathbb{P}(X = k)$$

$$= \mathbb{P}(X \le x) = F_X(x).$$

Dvs $X_n \to X$ i fördelnins

3) Vi skriver ibland $X_n \xrightarrow{d} X$

<u>Sats:</u> Låt $X_1, X_2, X_3, \ldots$ vara s.v. där X_n har mgf M_n. Låt sedan X vara en s.v. med mgf M. Vi har att

$$X_n \xrightarrow{(d)} X \iff \lim_{n\to\infty} M_n(t) = M(t) \quad \forall\, t \text{ i ett öppet intervall runt } 0.$$

<u>B:</u> svårt $\square$

<u>Anm:</u> Extremt användbart verktyg.

<u>Tal:</u> Använd mgf för att visa $X_n \xrightarrow{(d)} X$, $X_n \sim Bin(n, p_n)$, och $X \sim Poi(\lambda)$, och $np_n \to \lambda$.

<u>B:</u> Vi har att

$$M_n(t) = E[e^{X_n t}] = (p_n e^t + 1 - p_n)^n = (1 + p_n(e^t - 1))^n$$

$$\longrightarrow e^{\lambda(e^t - 1)} = M_X(t) \quad \square$$

$$\uparrow$$
$$(1+a_n)^{b_n} \to e^c$$

<u>Anm:</u> ~~Detta~~ Dessa konvergensresultat kallas ibland för små talens lag.

Sats (CLT): Låt $X_1, X_2, \ldots$ vara en iid sekvens med $\mu = \mathbb{E}[X_1]$, $\sigma^2 = Var(X_1) < \infty$ och mgf $M(t)$. Låt $S_n = \sum_{k=1}^{n} X_k$. Vi har att

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right) \to \Phi(x) \quad \forall x \in \mathbb{R}.$$

Dvs

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to Z \sim N(0,1).$$

B: Kom ihåg: $(1+a_n)^{b_n} \to e^c$ om $a_n b_n \to c$

$$M_{a+bX}(t) = e^{at} M_X(bt)$$

$$M_Z(t) = e^{t^2/2} \quad \text{om} \quad Z \sim N(0,1)$$

Låt $Y_n = X_n - \mu$ s.a. $\sum_{k=1}^{n} Y_k = \left(\sum_{k=1}^{n} X_k\right) - n\mu = S_n - n\mu$.

Sätt $T_n = \sum_{k=1}^{n} Y_k$ och $Z_n = \frac{T_n}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$.

Vi får att

$$M_{Z_n}(t) = M_{\frac{T_n}{\sigma\sqrt{n}}}(t) = M_{T_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \prod_{k=1}^{n} M_{Y_k}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

$$= \left[M_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Vi kan Taylorutveckla $M_{Y_1}(s)$:

$$M_{Y_1}(s) = M_{Y_1}(0) + s M_{Y_1}'(0) + \frac{1}{2} s^2 M_{Y_1}''(0) + O(s^3)$$

Vi har att $M_{Y_1}(0) = \mathbb{E}[e^{Y_1 \cdot 0}] = 1$

$$M'_{Y_1}(0) = \mathbb{E}\{Y_1\} = \mathbb{E}[X_1 - \mu] = 0$$

$$M''_{Y_1}(0) = \mathbb{E}\{Y_1^2\} = \mathbb{E}[(X_1-\mu)^2] = \sigma^2.$$

För fixt t får vi att

$$M_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t}{\sigma\sqrt{n}} \cdot 0 + \frac{1}{2}\frac{t^2}{\sigma^2 n} \cdot \sigma^2 + \mathcal{O}\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^3\right)$$

$$= 1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2}).$$

Därmed får vi att

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left[M_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

$$= \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2})\right]^n = e^{t^2/2}$$

ty $\left(\frac{t^2}{2n} + \mathcal{O}(n^{-3/2})\right)n = \frac{t^2}{2} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \to \frac{t^2}{2}.$

Då $e^{t^2/2}$ är mgf för $Z \sim N(0,1)$

är beviset klart. $\square$

Hur använder vi CLT i praktiken? $\boxed{F13}$ 5

För stora n är

1) $\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) \approx N(0,1)$

2) $X_1 + X_2 + \dots + X_n \approx N\left(\mathbb{E}\left[\sum_{k=1}^{n} X_k\right], \mathrm{Var}\left(\sum_{k=1}^{n} X_k\right)\right)$

3) $\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \approx N\left(\mathbb{E}[\overline{X}_n], \mathrm{Var}(\overline{X}_n)\right)$

F13 6 här!

Ex: Spel med 3 val

a) Du får 4500 SEK

b) Singla slant: Om 1) H får du 10000 SEK
2) T —"— 0 —"—

c) Singla 10000 slantar, du får alla som visar H.

Vad väljer ni?

" $\mathbb{E}[b)] = \mathbb{E}[c)]$ men $\mathrm{Var}(b)) > \mathrm{Var}(c))$ "

Låt oss analysera c).

Sätt $X = $ # slantar som vinns $\sim \mathrm{Bin}(10000, \frac{1}{2})$

CGS säger att

$$\bar{X} \approx N(\mathbb{E}[\bar{X}], \mathrm{Var}(\bar{X})) = N(5000, 2500)$$

Vad är $\mathbb{P}(\bar{X} \leq 4500)$? (dvs c) sämre än a))

Jo $\mathbb{P}(\bar{X} \leq 4500) = \mathbb{P}\left(\frac{\bar{X} - 5000}{\sqrt{2500}} \leq \frac{4500 - 5000}{\sqrt{2500}}\right)$

$$\approx \mathbb{P}(Z \leq -10) = \Phi(-10) \leq 10^{-10}$$

Så a) är sämre än c).

Valet mellan b) och c) är kanske inte helt självklart.

Kom ihåg: $X \sim \mathrm{Bin}(n,p)$ där np litet (säg ≤ 10) och n stort $\Rightarrow X \approx \mathrm{Poi}(np)$.

Om istället n stort och np stort?

Vi har att

$$X = X_1 + \cdots + X_n$$

där $X_1, \ldots, X_n$ iid $\mathrm{Be}(p)$. CGS kan appliceras på denna summa. Vi får att

$$X \approx N(\mathbb{E}[X], \mathrm{Var}(X))$$

TMA 321

F13 ⑦

Tal (i mån av tid):

Sara jobbar på posten. Breven hon sorterar väger i genomsnitt 4.7 g med en varians av 25 g. Hon lägger 1000 brev i en kasse. Uppskatta sann. att kassen väger mer än 5 kg.

L: Låt X_i = vikt för brev i där $i=1,2,\dots,1000$.

Antag att $X_1, \dots$ är en iid sekvens.

Vi efterfrågar $\mathbb{P}(X_1+\dots+X_{1000} \geq 5000)$

och får $(X_1+\dots+X_{1000} \approx N(\mathbb{E}[X_1+\dots+X_{1000}], \mathrm{Var}(X_1+\dots+X_{1000}))$
$\approx N(4700, 25000)$

$$\mathbb{P}(X_1+\dots+X_{1000} \leq 5000) =$$

$$= \mathbb{P}\left(\frac{X_1+\dots+X_{1000}-4700}{\sqrt{25000}} \leq \frac{5000-4700}{\sqrt{25000}}\right)$$

$$\approx \mathbb{P}\left(Z \leq \frac{300}{\sqrt{25000}}\right) \approx \mathbb{P}(Z \leq 1.90) \approx 0.9713$$

CLT

$\Rightarrow$ den sökta sann. är ca $1-0.9713 \approx 0.029$