20180423 Kom ihåg

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Sats (STL)

$X_1, X_2, X_3, \ldots$ i.i.d. med $\mu = \mathbb{E}[X_1]$, $\sigma^2 = Var(X_1) < \infty$.
Då gäller att $\forall \varepsilon > 0$ $\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$.

Anm. uppenbarligen innebär detta att $\bar{X}_n \to \mu$ i ngn mening.

Def Vi säger att sekvensen $X_1, X_2, \ldots$ konv. mot X i sannolikhet om $\forall \varepsilon > 0$ $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.
Vi skriver $X_n \xrightarrow{p} X$

Anm: STL: $\bar{X}_n \xrightarrow{p} \mu$.

Ex Monte-Carlo simulering

Låt $X_1, X_2, \ldots$ vara iid & $X_k \sim U[0,1]$. Vi har $\mathbb{E}[e^{X_i^2}] = \int_0^1 e^{x^2}dx$.
Låt $Y_k = e^{X_k^2}$. STL ger då att

$$\bar{Y}_n = \frac{1}{n}\sum_{k=1}^{n} e^{X_k^2} \xrightarrow{p} \mathbb{E}[Y_1] = \int_0^1 e^{x^2}dx.$$

Använd MATLAB för att generera $X_1, X_2, \ldots, X_{10000}$

$$\frac{1}{10000}\sum_{k=1}^{10000} e^{X_k^2} \approx \int_0^1 e^{x^2}dx$$

Momentgenererande funktioner

Def den momentgenererande funktionen, mgf, av en s.v. X betecknas $M_X(t)$ & def av

$$M_X(t) = \mathbb{E}[e^{Xt}],$$

för de t s.a. väntevärdet är ändligt.

Anm 1) Vi får att $M_X(t) = \int_{-\infty}^{\infty} e^{xt} f_X(x)\,dx$, om X kont. s.v. medan för diskreta s.v. är $M_X(t) = \sum_{x_j} e^{x_j t} \mathbb{P}(X = x_j)$.

2) Antag att X är kont. Vi får att

$$M_X'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{xt} f_X(x)\,dx \overset{(?!)}{=} \int_{-\infty}^{\infty} \frac{d}{dt} e^{xt} f_X(x)\,dx = \int_{-\infty}^{\infty} x e^{xt} f_X(x)\,dx$$

$$\Rightarrow M_X'(0) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \mathbb{E}[X].$$

p.s.s. $M_X^{(n)}(0) = \mathbb{E}[X^n] \leftarrow$ n:te momentet

---

Sats Om $M_X(t)$ existerar $\forall t$ i ett öppet intervall runt 0, så bestämmer detta slf/tfkn/fördelningen av X.

Anm: I själva verket är mgf en transform.

$$\text{fördelningar} \xleftrightarrow{\text{mgf}} \text{funktioner } \mathbb{R} \to \mathbb{R}$$

Satsen säger att transformen är <u>inverterbar</u>

Tal hitta mgf för $X_1 \sim Poi(\lambda_1)$

L vi har att $M_{X_1}(t) = \mathbb{E}[e^{X_1 t}] = \sum_{k=0}^{\infty} e^{kt} \mathbb{P}(X_1 = k) = \sum_{k=0}^{\infty} e^{kt} \frac{\lambda_1^k}{k!} e^{-\lambda_1}$

$$= e^{-\lambda_1} \sum_{k=0}^{\infty} \frac{(\lambda_1 e^t)^k}{k!} \underset{\uparrow \text{Taylor}}{=} e^{-\lambda_1} e^{\lambda_1 e^t} = e^{\lambda_1 (e^t - 1)} \quad //$$

Sats Om $X_1, X_2, \ldots$ är ober. gäller att

$$M_{X_1 + X_2 + \ldots + X_n}(t) = \prod_{k=1}^{n} M_{X_k}(t).$$

Bevis Vi har att

$$M_{X_1+X_2+\dots+X_n}(t) = \mathbb{E}[e^{t(X_1+\dots+X_n)}] = \mathbb{E}[\prod_{k=1}^{n} e^{tX_k}] \underset{\substack{\uparrow \\ ober.}}{=}$$

$$= \prod_{k=1}^{n} \mathbb{E}[e^{tX_k}] = \prod_{k=1}^{n} M_{X_k}(t) \quad \blacksquare$$

Tal låt $X_1 \sim Poi(\lambda_1)$ & $X_2 \sim Poi(\lambda_2)$ vara ober. s.v.
Hitta fördelningen för $X_1 + X_2$

$\mathcal{L}$ **Metod 1:** beräkna $\mathbb{P}(X_1 + X_2 = k)$ $\forall k$.

Metod 2: hitta mgf $X_1 + X_2$:

$$M_{X_1+X_2}(t) \overset{ober.}{=} M_{X_1}(t) + M_{X_2}(t) = e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

$$\Rightarrow X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2) \quad //$$

Sats Om X har mgf $M_X(t)$ & $Y = a + bX$

$$\Rightarrow M_Y(t) = e^{at} M_X(bt).$$

Bevis $M_Y(t) = \mathbb{E}[e^{Yt}] = \mathbb{E}[e^{(a+bX)t}] = \mathbb{E}[e^{at} e^{btX}] =$

$$= e^{at}\, \mathbb{E}[e^{X(bt)}] = e^{at} M_X(bt) \quad \blacksquare$$

Ex låt $X \sim N(0,1)$, $M_X(t) = \dots = e^{t^2/2}$.
Enl. tidigare $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$.

Vi får $M_Y(t) = e^{\mu t} M_X(\sigma t) = e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}$ //

Def för två s.v. (X, Y) def vi deras gemensamma mgf som

$$M_{X,Y}(s,t) = \mathbb{E}[e^{sX+tY}].$$

Vi får att $M_X(s) = M_{X,Y}(s,0)$

Anm: man kan visa att
X, Y ober $\iff M_{X,Y}(s,t) = M_X(s)\, M_Y(t)$.

Mgf i olikheter

Vi har att om $X \geq 0$ & $a > 0$

$$\mathbb{P}(X \geq a) \underset{t>0}{=} \mathbb{P}(e^{tX} \geq e^{ta}) \underset{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = e^{-at} M_X(t)$$

Tal hitta mgf $X \sim \text{Bin}(n,p)$

L låt $X_k = \begin{cases} 1 & \text{om försök } k \text{ lyckas} \\ 0 & \text{annars,} \end{cases}$

dvs. låt $X_1, \ldots, X_n$ vara iid $\text{Bern}(p)$, då har vi att $X = X_1 + \ldots + X_n$.

Vi har att $M_X(t) = M_{X_1+\ldots+X_n}(t) = \prod_{k=1}^{n} M_{X_k}(t) = [M_{X_1}(t)]^n$.

Vidare är $M_{X_1}(t) = \mathbb{E}[e^{X_1 t}] = e^t p + e^0(1-p) = pe^t + 1 - p$.

Vi får då att $M_X(t) = (pe^t + 1 - p)^n$ //

Tal låt $X \sim \text{Bin}(n,p)$. Hitta en <u>bra</u> övre begränsning på $\mathbb{P}(X \geq np(1+\varepsilon))$.

L $\mathbb{P}(X \geq np(1+\varepsilon)) \leq e^{-np(1+\varepsilon)t} M_X(t) = e^{-np(1+\varepsilon)t}(pe^t + 1 - p)^n =$

$$= e^{-np(1+\varepsilon)t}(1 + p(e^t - 1))^n \underset{1+x \leq e^x}{\leq} e^{-np(1+\varepsilon)t}(e^{p(e^t-1)})^n = e^{np(e^t - 1 - t(1+\varepsilon))}$$

Vidare är $e^t = 1 + t + t^2/2 + \ldots \underset{\text{om } t \leq \varepsilon}{\leq} 1 + (1 + \frac{2\varepsilon}{3})t$

$$\Rightarrow \mathbb{P}(X \geq np(1+\varepsilon)) \leq e^{np(e^t - 1 - t(1+\varepsilon))} \leq e^{np(1 + (1+\frac{2\varepsilon}{3})t - 1 - t(1+\varepsilon))} =$$

$$= e^{-np\frac{\varepsilon t}{2}} \quad \text{om } t \leq \varepsilon$$

dvs. $\mathbb{P}(X \geq np(1+\varepsilon)) \leq e^{-np\frac{\varepsilon^2}{3}}$

Anm: innan hade vi mha Cheb. & $p=\frac{1}{4}$ att
$\mathbb{P}(X \geq n/2) \leq 3/n$.

Nu har vi att $p=\frac{1}{4}$, $\varepsilon=1$ $\Rightarrow$ $\mathbb{P}(X \geq n/2) \leq e^{-n\frac{1}{12}}$

Anm: $X \sim \mathrm{Bin}(10^{20}, 1/2)$

$$\mathbb{P}\left(\left|\frac{X}{n} - \frac{1}{2}\right| > 10^{-9}\right) \leq \ldots \leq 2e^{-\frac{n}{12}\varepsilon^2} = 2e^{-\frac{100}{12}}$$

2018 04 24 Kom ihåg

1) mgf $M_X(t) = \mathbb{E}[e^{Xt}]$ för de t s.a. väntevärdet är ändligt

2) $(1+a_n)^{b_n} \to e^c$, $n\to\infty$ ifall $a_n \to 0$ & $a_n b_n \to c$ då $n\to\infty$

Centrala gränsvärdessatsen, CGS, CLT

Def låt $X_1, X_2, \ldots$ vara en sekvens av s.v. med ffkn $F_1, F_2, \ldots$ & 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 X är kontinuerlig.

Anm: 1) $\lim_{n\to\infty} F_n(x) = F(x) \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$

2) kom ihåg: $X_n \sim Bin(n, p_n)$ där $np_n \to \lambda$ då $n\to\infty$.
Vi fick då att $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)$ där $X \sim Poi(\lambda)$.
Vi får då att $F_n(x) = \mathbb{P}(X_n \leq x) = \sum_{k\leq x} \mathbb{P}(X_n = k) \longrightarrow$
$\longrightarrow \sum_{k\leq x} \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x) = F(x)$.

dvs. $X_n \to X$ i fördelning.

3) Vi skriver ofta $X_n \xrightarrow{d} X$ d for distribution

---

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

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

Bevis svårt. ■

Anm: extremt användbart verktyg.

Tal använd mgf för att visa $X_n \xrightarrow{d} X$ då
$X_n \sim Bin(n, p_n)$, $X \sim Poi(\lambda)$ & $np_n \to \lambda$.

Kom ihåg: $X \sim Bin(n,p) \Rightarrow M_X(t) = (pe^t + 1 - p)^n$,
$X \sim Poi(\lambda)$, $M_X(t) = e^{\lambda(e^t - 1)}$

Bevis vi får $M_{X_n} = (p_n e^t + 1 - p_n)^n = (1 + p_n(e^t - 1))^n \to e^{\lambda(e^t-1)} = M_X(t)$

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

Anm: detta konv. resultat kallas ibland för små talens lag.

Sats (CGS/CLT)

Låt $X_1, X_2, \ldots$ vara en iid sekvens med $\mu = \mathbb{E}[X_1]$,
$\sigma^2 = Var(X_1) < \infty$.

Låt $S_n = \sum_{k=1}^{n} X_k$. Vi har då att $\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) \to \Phi(x) \quad \forall x \in \mathbb{R}$

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

Bevis

Kom ihåg: $M_{a+Xb}(t) = e^{at} M_X(bt)$
$M_Z(t) = e^{t^2/2}$ om $Z \sim N(0,1)$

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

Vi får att $M_{Z_n}(t) = M_{\frac{T_n}{\sigma\sqrt{n}}} = M_{T_n}\left(\frac{t}{\sigma\sqrt{n}}\right) \overset{\text{ober.}}{=} \prod_{k=1}^{n} M_{Y_k}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left[M_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$.

Vi kan Taylorutv. $M_{Y_1}(s)$.
$M_{Y_1}(s) = M_{Y_1}(0) + sM'_{Y_1}(0) + s^2 M''_{Y_1}(0) + \mathcal{O}(s^3)$.

Vi har

- $M_{Y_1}(0) = \mathbb{E}[e^{Y_1 0}] = \mathbb{E}[1] = 1$

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

För fixt t så fås att $M_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t}{\sigma\sqrt{n}} \cdot 0 + \frac{1}{2}\frac{t^2}{\sigma^2 n}\sigma^2 + \mathcal{O}(n^{-3/2})$

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

$$\lim_{n\to\infty} M_{Z_{1,n}}(t) = \lim_{n\to\infty}\left[M_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2})\right]^n = e^{t^2/2},$$

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

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

Hur använder vi CGS i praktiken?

För "stora" n är

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

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

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

Kom ihåg: $X \sim Bin(n,p)$ där n stort & np litet (säg ≤ 10)
$\Rightarrow X \approx Poi(\lambda)$ där $\lambda = np$.

Om ist n stort & np stort? Vi har att
$X = X_1 + X_2 + \ldots + X_n$, där $X_1, \ldots, X_n$ är iid. $Be(p)$.
CGS kan appliceras på denna summa. Vi får att
$X \approx N(\mathbb{E}[X], Var(X)) = N(np, np(1-p))$

Ex Spel med 3 val

a) du får 4500 SEK

b) singla slant, om
 1) H får du 10 000 SEK
 2) T får du 0 SEK

c) singla 10 000 slantar, du får alla som visar H.

Vilket ska man välja? Vi analyserar c).

Sätt $X = \#$ enkronor som vinns $\sim Bin(10000, 1/2)$.

CGS säger $X \approx N(\mathbb{E}[X], Var(X)) = N(5000, 2500)$.

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

$$\mathbb{P}(X \leq 4500) = \mathbb{P}\left(\frac{X-5000}{\sqrt{2500}} \leq \frac{4500-5000}{\sqrt{2500}}\right) \approx \mathbb{P}(Z_1 \leq -10) = \Phi(-10) \lesssim 10^{-10}$$

där $Z_1 \sim N(0,1)$.

2018 04 25 Parametrisk punktskattning

STATISTIK!

Vår statistiska modell

1) Vi har en population av storlek N, ur vilken vi drar ett slumpvis stickprov $X_1, X_2, \dots, X_n$ (av storlek n).

2) Vi kommer (nästan) alltid anta att $N = \infty$. Då blir $X_1, X_2, \dots, X_n$ iid.

3) Vi antar att fördelningen av X_i är känd sånär som på värdet av en eller flera parametrar.

Ex: $X_i \sim Poi(\lambda)$ med (värdet på) λ okänt.

Ex: $X_i \sim N(\mu, \sigma^2)$ med μ och/eller σ^2 okänt.

Vi vill skatta (=gissa) parametrarna & vi låter θ utgöra den generiska beteckningen för dessa.

← hatt = gissning

Def en skattning är en funk. $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ av stickprovet $(X_1, \dots, X_n) = \underline{X}$.

Ex: 1) Om $\theta = \mu = \mathbb{E}[X_i]$, då använder vi skattningen

$$\hat{\theta}(\underline{X}) = \hat{\mu} = (X_1 + \dots + X_n)/n = \overline{X}$$

2) Om $\theta = \sigma^2 = Var(X_i)$, då används ofta

$$\hat{\theta}(\underline{X}) = \widehat{\sigma^2} = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

om $\mu = \mathbb{E}[X_i]$ är okänd. Alternativt använder vi

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2,$$ om μ är känd.

3) Om $X_i \sim Exp(\lambda)$ så är $\mu = \mathbb{E}[X_i] = \lambda^{-1}$. Därför är $\hat{\lambda} = \frac{1}{\overline{X}}$ naturligt (ty $\hat{\lambda} = \frac{1}{\hat{\mu}} = \frac{1}{\overline{X}}$)

Def 1) $\hat{\theta}$ är väntevärdesriktig (VVR) om
$\mathbb{E}[\hat{\theta}(\underline{X})] = \theta$ (unbiased)

2) Effektiviteten (av VVR-skattningar) bedöms av variansen $Var(\hat{\theta}(\underline{X}))$ (liten = bra).

3) $\hat{\theta}$ är konsistent om $\hat{\theta}(\underline{X}) \xrightarrow{P} \theta$, $n \to \infty$, dvs. om $\forall \varepsilon > 0$ $\mathbb{P}(|\hat{\theta}(\underline{X}) - \theta| > \varepsilon) \to 0$, $n \to \infty$.

Anm: 3) kräver $N = \infty$. Variant finns då $N < \infty$, detta kräver mycket extra räkning med relativt liten vinst.

Observera att $\hat{\theta}(\underline{X})$ är en s.v. Fördelningen för $\hat{\theta}(\underline{X})$ är av central betydelse.

Ex 1: betrakta $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Fördelningen av $\bar{X}$ är i allmänhet svåråtkomlig men:

i) $\mathbb{E}[\bar{X}] = \frac{1}{n}\mathbb{E}[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n} n\mu = \mu$ så VVR

ii) $Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) \underset{\text{ober.}}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

iii) Vi har att $\mathbb{P}(|\bar{X} - \mu| > \varepsilon) \to 0$, $n \to \infty$ (STL), så $\bar{X}$ är konstant.

Ex 2: betrakta $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Lång räkning ger (s. 211-212) $\mathbb{E}[s^2] = \ldots = \sigma^2$, så VVR (därav $\frac{1}{n-1}$ istf $\frac{1}{n}$).

Ex 3: Vi vill skatta $p = \mathbb{P}(A)$ där A är ngn händelse (utvald person har cancer). Vi vill skatta p & låter

$$X_i = \begin{cases} 1 & \text{om } A \text{ inträffar för individ } i \\ 0 & \text{annars.} \end{cases}$$

Vi tar $\hat{p} = \bar{X}$. Vi har att $X_1 + \ldots + X_n \sim Bin(n,p)$ s.a.
$\mathbb{E}[\bar{X}] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = p$ & $Var(\hat{p}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) =$

$$= \frac{1}{n^2} Var(X_1 + \dots + X_n) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

Vi är ofta intresserade av $Var(\hat{\theta}(X))$.

Def. standardfelet är $\sqrt{Var(\hat{\theta}(X))}$ & betecknas ibland $D(\hat{\theta})$.

Anm: Ofta beror variansen $Var(\hat{\theta}(X))$ på den okända parametern θ. Därför måste $D(\hat{\theta})$ skattas. Vi skattar en egenskap hos en skattare!

Ex: om $\bar{X}$ skattar μ har vi att $Var(\bar{X}) = \sigma^2/n$.
Vi kan skatta $\sqrt{Var(\bar{X})} = \sigma/\sqrt{n}$ med $s_{\bar{X}} = s/\sqrt{n}$, där
$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Ex: om $\hat{p}$ skattar p låter vi $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
Skatta $D(\hat{p}) = \sqrt{Var(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$.

Men hur ska vi hitta en bra skattning/skattare?

Momentmetoden

Om $X_1, X_2, \dots$ är iid med parameter θ är det ofta så att $E[X_i] = f(\theta)$, för ngn (snäll) funk. f.

Gör följande:

1) observera att STL säger att för n stort är $\bar{X} \approx E[X_1]$
2) lös ut $\hat{\theta}$ ur ekvationen $\bar{X} = f(\hat{\theta})$

Vi ser att $f(\Theta) = \mathbb{E}[X_1] \approx \overline{X} = f(\hat{\Theta})$ & därmed bör också $\hat{\Theta} \approx \Theta$, s.a. $\hat{\Theta}$ är en bra skattare.

Resultatet $(\hat{\Theta})$ kallas för en MME-skattare (method of moments estimator).

Tal väntetiden på en buss anses vara $U[0, \Theta]$

a) hitta MME-skattare för Θ.

b) om vi fick mätvärdena $x_1 = 2$, $x_2 = 7$, $x_3 = 4$, $x_4 = 15$, $x_5 = 11$ minuter, vilket blir det numeriska värdet på $\hat{\Theta} = \hat{\Theta}(x_1, x_2, x_3, x_4, x_5)$?

L a) om $X_1 \sim U[0, \Theta] \Rightarrow \mathbb{E}[X_1] = \Theta/2$.

MM: $\overline{X} = \hat{\Theta}/2 \Rightarrow \hat{\Theta} = 2\overline{X}$ är vår MME-skattare.

b) vi får att $\hat{\Theta}(x) = 2\overline{x} = 2 + \dfrac{x_1 + \ldots + x_5}{5} = 15.6$ minuter //

Om vi har två okända parametrar Θ_1, Θ_2 beräknar vi

$\mathbb{E}[X] = f_1(\Theta_1, \Theta_2)$, $\mathbb{E}[X^2] = f_2(\Theta_1, \Theta_2)$.

Vi ersätter $\overline{X} = f_1(\hat{\Theta}_1, \hat{\Theta}_2)$, $\overline{X^2} = f_2(\hat{\Theta}_1, \hat{\Theta}_2)$ & löser ut $\hat{\Theta}_1, \hat{\Theta}_2$.

Här är $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

20180425 Räkneövning 5
Konvergens

5.7 Visa att om $X_n \to c$ i sannolikhet & g är en kont. funk så $g(X_n) \to g(c)$ i sannolikhet.

Def konvergens i sannolikhet.
$X_n \to c$ i sanno. betyder att för varje $\varepsilon > 0$ gäller att
$\mathbb{P}(|X_n - c| > \varepsilon) \to 0$ då $n \to \infty$.

Def kontinuerlig funktion
Att g är en kont. funk. betyder att $\forall\ \varepsilon > 0$ finns ett $\delta > 0$
s.a. $|x-y| < \delta \Rightarrow |g(x) - g(y)| < \varepsilon$

$\mathcal{L}$ Välj godt. $\varepsilon > 0$. g kont. $\Rightarrow \exists\ \delta > 0$ s.a. $|X_n - c| < \delta$
$\Rightarrow |g(X_n) - g(c)| < \varepsilon$.
$X_n \to c$ i sanno. $\Rightarrow \mathbb{P}(|X_n - c| < \delta) \to 1$, då $n \to \infty$.
Eftersom $|X_n - c| < \delta$ implicerar händelsen $|g(X_n) - g(c)| < \varepsilon$,
gäller $\mathbb{P}(|g(X_n) - g(c)| < \varepsilon) \to 1$ då $n \to \infty$.

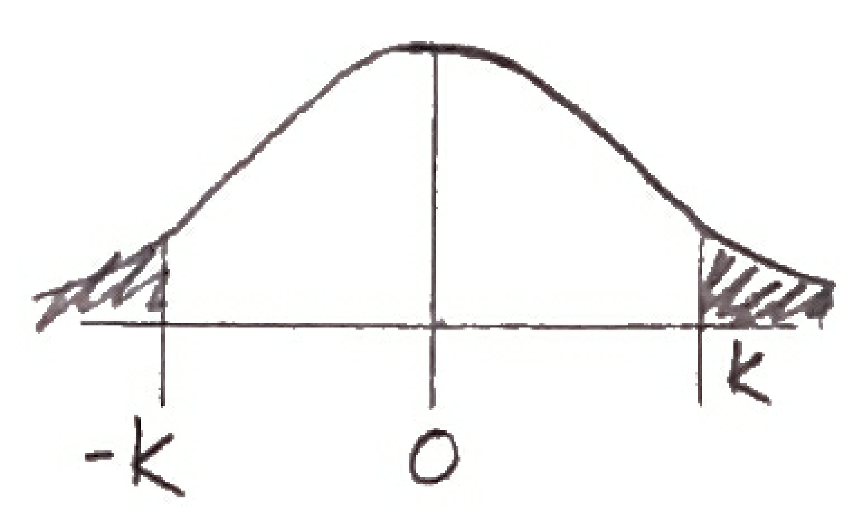
5.12 Anta att värden avrundas & att avrundningsfelet är uniformt fördelat på $[-1/2, 1/2]$. 100 värden summeras, approximera sanno. att det totala avrund. felet blir större än

a) 1 b) 2 c) 10

$\mathcal{L}$ Y_i : avrund felet för värde i, $Y_i \sim U[-1/2, 1/2]$
X : totala felet, $X = \sum_{i=1}^{100} Y_i$

Centrala gränsv. satsen ger $\dfrac{X}{\sqrt{n\,Var(Y_i)}} \overset{approx.}{\sim} N(0,1)$

$Var(Y_i) = \frac{1}{12}$, $n = 100$



$\mathbb{P}(|X|>k) = 2\,\mathbb{P}(X>k) = 2(1-\mathbb{P}(X\le k))$

$\mathbb{P}(X\le k) = \mathbb{P}\left(\frac{X}{\sqrt{\frac{100}{12}}} \le \frac{k}{\sqrt{\frac{100}{12}}}\right), \quad \frac{k}{\sqrt{\frac{100}{12}}} = \frac{2\sqrt{3}k}{10} = \frac{\sqrt{3}k}{5}$

tabell i boken

a) $k=1$, $\frac{\sqrt{3}}{5} \approx 0{,}3464$, $\mathbb{P}(Z_1 \le \frac{\sqrt{3}}{5}) \approx 0{,}6368$

$\mathbb{P}(|X|>1) = 2(1-0{,}6368) \approx 0{,}726$

b) $k=2$, $\frac{2\sqrt{3}}{5} \approx 0{,}6928$, $\mathbb{P}(Z_1 \le 2\sqrt{3}/5) \approx 0{,}7549$

$\mathbb{P}(|X|>2) \approx 0{,}490$

c) $k=5$, $\sqrt{3}=1{,}732$, $\mathbb{P}(Z_1 \le \sqrt{3}) \approx 0{,}9582$

$\mathbb{P}(|X|>5) \approx 0{,}0836$

5.29 Låt $U_i \sim U[0,1]$, $i=1,\dots,n$ ober.
Låt $U_{(n)} = \max(U_1,\dots,U_n)$. Ange fördelningsfunktionen för $U_{(n)}$. Hitta en standardisering av $U_{(n)}$ samt visa att ffkn för den standardiserade variabeln går mot ett värde som ej beror på n, då $n\to\infty$.

L $F_{U_{(n)}}(x) = P(U_{(n)} \le x) = P(\max(U_1,\dots,U_n)\le x) = P(U_1\le x \cap \dots \cap U_n \le x) =$
$= P(U_1 \le x)^n = x^n$

Låt X_n vara en standardiserad $U_{(n)}$. Vi sätter $X_n = \frac{U_{(n)}-a_n}{b_n}$ & vill hitta a_n & b_n.

$F_{X_n}(x) = P(X_n \le x) = P\left(\frac{U_{(n)}-a_n}{b_n} \le x\right) = P(U_{(n)} \le xb_n + a_n) = (xb_n+a_n)^n$

Välj $b_n = \frac{1}{n}$ & $a_n = 1 \Rightarrow F_{X_n}(x) = \left(\frac{x}{n}+1\right)^n \to e^x$ då $n\to\infty$
för $x \le 0$ & 1 annars.

7.4 Medelvärdet skattas i två populationer genom slumpvis urval. Population 1 har sann standardavvikelse σ & urvalets storlek är n. Population 2 har sann

stand. avvik. 2σ & storlek $2n$ på urvalet.
Vilken skattning förväntas vara bäst?

2 $Var(\bar{X}_1) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_{1,i}\right) = \frac{1}{n^2}\sum Var(x_{1,i}) = \frac{\sigma^2}{n}$

$Var(\bar{X}_2) = \frac{(2\sigma)^2}{2n} = \frac{2\sigma^2}{n} = 2\,Var(\bar{X}_1)$

$\Rightarrow$ skattningen för pop 1 förväntas vara bäst.
(Båda är väntevärdesriktiga, $E(\bar{X}_1) = E(X_{1,i})$
$E(\bar{X}_2) = E(X_{2,i})$)