TAMS79: Föreläsning 8
Konvergens, Stora talens lag, CGS

Johan Thim (johan.thim@liu.se)

29 november 2018

8.1 Konvergens

För att kunna få lite precision i argumenten i detta område behöver vi lite begrepp angående konvergens av stokastiska variabler. Eftersom vi har introducerat sannolikhet så uppstår nu en hel rad spännande möjligheter till olika typer av konvergens. Kom ihåg att en stokastisk variabel X är en funktion från utfallsrummet Ω till $\mathbf{R}$ (eller mer generellt $\mathbf{R}^n$). En följd stokastiska variabler X_n är alltså en följd funktioner, och som bekant från envariabelanalysen kan denna följd *konvergera* mot en funktion X om det är så att

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \text{för alla } \omega \in \Omega.$$

Detta brukar kallas *punktvis konvergens*, eller i sannolikhetstermer: **säker konvergens**. Nu är det sällan vi kommer att ha säker konvergens eftersom sannolikhet är inblandad, så låt oss börja med en annan typ av konvergens som kan vara värd att ha sett om inte annat än för att kunna säga saker som att något är "nästan säkert" och faktiskt mena något väldigt specifikt...

Nästan säker konvergens

Definition. Låt X_n, $n = 1, 2, \ldots$, vara en följd stokastiska variabler. Vi säger att X_n konvergerar till X nästan säkert (almost surely) om

$$P(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1.$$

Vi skriver i detta fall att $X_n \xrightarrow{\text{a.s.}} X$. Underförstått är att samtliga variabler är definierade på samma utfallsrum Ω.

Definitionen ovan säger att *följden* konvergerar punktvis: $X_n(\omega) \to X(\omega)$ för alla ω förutom på en delmängd av Ω som har sannolikhet noll.

Konvergens i sannolikhet

Definition. Låt X_n, $n = 1, 2, \ldots$, vara en följd stokastiska variabler. Vi säger att X_n konvergerar till en stokastisk variabel X i sannolikhet om för alla $\epsilon > 0$ så gäller att

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

och vi skriver i detta fall att $X_n \xrightarrow{\text{P}} X$. Generaliserar naturligt till högre dimensioner.

Vi kan notera att

$$X_n \xrightarrow{\text{a.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{\text{P}} X,$$

men inte omvänt. Detta är inte självklart utan hänger i princip på att vi kan byta ut ordningen på att beräkna sannolikhet och ta ett gränsvärde. Den intresserade kan slå upp Fatous lemma. Att något konvergerar i sannolikhet innebär inte heller att vi kan säga så mycket om väntevärde eller varians, något följande exempel visar.

Exempel

Låt X_n vara Bernoullifördelad enligt $X_n = n$ med sannolikhet $1/n$ och $X_n = 0$ med sannolikhet $1 - 1/n$. Visa att $X_n \xrightarrow{\text{P}} 0$ då $n \to \infty$ men att $E(X_n) = 1$ och $V(X_n) = n - 1 \to \infty$ då $n \to \infty$.

Lösning. Vi ser att

$$E(X_n) = 0 \cdot \left(1 - \frac{1}{n}\right) + n \cdot \frac{1}{n} = 1$$

och att

$$E(X_n^2) - E(X_n)^2 = 0^2 \cdot \left(1 - \frac{1}{n}\right) + n^2 \cdot \frac{1}{n} - 1^2 = n - 1.$$

Men för varje $n \geq \epsilon > 0$ (övre gränsen gör inget då $n \to \infty$) så gäller att

$$P(|X_n| \geq \epsilon) = P(X_n > 0) = \frac{1}{n} \to 0,$$

då $n \to \infty$, eftersom X_n endast antar värdena 0 och n och nollan prickar vi aldrig då $\epsilon > 0$. En naturlig fråga är nu om vi har konvergens nästan säkert, men då får vi problem eftersom det underliggande utfallsrummet inte är specificerat. Vi kan alltså inte svara på den frågan.

Den sista konvergenstypen vi betraktar är konvergens i fördelning. Vad detta innebär informellt är att fördelningsfunktionerna för X_n konvergerar punktvis mot fördelningsfunktionen för X.

Definition. Låt X_n, $n = 1, 2, \ldots$, vara en följd stokastiska variabler. Vi säger att X_n konvergerar till en stokastisk variabel X i fördelning om

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

för alla x (där F är kontinuerlig). Här är F_{X_n} och F_X respektive fördelningsfunktion, och vi skriver att $X_n \xrightarrow{\text{D}} X$. I högre dimensioner formuleras ofta kraven direkt i termer av sannolikhet enligt

$$\boldsymbol{X}_n \xrightarrow{\text{D}} \boldsymbol{X} \quad \Leftrightarrow \quad \lim_{n \to \infty} P(\boldsymbol{X_n} \in E) = P(\boldsymbol{X} \in E)$$

för alla rimliga mängder E sådana att $P(\partial E) = 0$.

Egenskapen att mängden E uppfyller att $P(\partial E) = 0$ brukar kallas för att E är en kontinuitetsmängd (kommer från mått-teorin) för måttet P. Alternativt kan man betrakta fördelningsfunktionen, så låt

$$E(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbf{R}^k : y_1 \leq x_1,\ y_2 \leq x_2,\ \ldots,\ y_k \leq x_k\}$$

så att fördelningsfunktionen F ges av $F(x_1, x_2, \ldots, x_k) = P(\boldsymbol{X} \in E(\boldsymbol{x}))$. Detta följer direkt från att den flerdimensionella fördelningsfunktionen ges av

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k)$$

Randen ∂E till E kan vi uttrycka som

$$E(\boldsymbol{x}) \cap \{\boldsymbol{y} \in \mathbf{R}^k : y_i = x_i \text{ för något } i,\ 1 \leq i \leq k\},$$

så om $P(\boldsymbol{X} \in \partial E(\boldsymbol{x})) = 0$ är helt enkelt F kontinuerlig i punkten $\boldsymbol{x}$.

Exempel

Vad betyder det att $X_n \xrightarrow{\text{D}} c$ för någon konstant c?

Lösning. Vi vill alltså beskriva en stokastisk variabel som är konstant. Åtminstone två varianter finns. Den ena är att variabeln identiskt (för varje $\omega \in \Omega$) är lika med konstanten. Den andra är att variabeln sammanfaller med konstanten för alla $\omega \in \Omega$ förutom på någon mängd med mått noll. I båda fallen kommer fördelningsfunktionen ges av

$$F(x) = P(c \leq x) = \begin{cases} 0, & x < c, \\ 1, & x \geq c. \end{cases}$$

Alltså en stegfunktion. Om $X_n \xrightarrow{\text{D}} c$ innebär det alltså att $F_{X_n}(x) \to F(x)$ för alla $x \neq 0$ (eftersom F inte är kontinuerlig i 0:an).

Värt även att notera att

$$X_n \xrightarrow{\text{P}} X \quad \Rightarrow \quad X_n \xrightarrow{\text{D}} X.$$

Även här är beviset ganska tekniskt, men det är värt att komma ihåg att konvergens i fördelning är den svagaste typen av konvergens vi tagit upp.

Definition. Om $X_n \xrightarrow{\text{D}} X$ kallas fördelningen för X för den **asymptotiska** fördelningen för sekvensen X_n, $n = 1, 2, \ldots$.

Ibland krävs att denna fördelning inte är allt för degenererad för att kallas för en asymptotisk fördelning. Att sekvensen konvergerar mot en konstant till exempel brukar ofta kallas för ett degenererat fall (en konstant är en ganska skum stokastisk variabel).

Exempel

Låt $X_n \sim N(0, \sqrt{n})$. Visa att $F_{X_n}(x) \to \dfrac{1}{2}$ då $n \to \infty$. Konvergerar X_n i fördelning?

Lösning. Om $X_n \sim N(0, \sqrt{n})$ så ges fördelningsfunktionen av

$$F_{X_n}(x) = P(X_n \le x) = \frac{1}{\sqrt{2\pi n}} \int_{-\infty}^{x} e^{-t^2/(2n)}\, dt = \left/ \begin{array}{c} t = u\sqrt{n} \\ dt = \sqrt{n}\, du \end{array} \right/ = \frac{\sqrt{n}}{\sqrt{2\pi n}} \int_{-\infty}^{x/\sqrt{n}} e^{-u^2/2} du$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/\sqrt{n}} e^{-u^2/2} du \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-u^2/2} du = \frac{1}{2},$$

då $n \to \infty$ för alla $x \in \mathbf{R}$. Detta är uppenbarligen ingen giltig fördelningsfunktion (den går inte mot noll då $x \to -\infty$ eller mot ett då $x \to \infty$ till exempel). Alltså konvergerar *inte* följden mot något i fördelning.

Så även om fördelningsfunktionerna konvergerar mot *något*, så behöver vi inte ha konvergens i fördelning. Ett mer subtilt problem uppstår om vi funderar över vad som händer med täthetsfunktionerna. Då visar det sig att dessa i princip inte har något med saken att göra. Nå, det kanske är lite väl hårt, men det måste ställas extra krav för att kunna dra några slutsatser om eller från hur täthetsfunktionerna beter sig.

Exempel

Låt $X_1, X_2, \ldots$ vara en följd stokastiska variabler med $F_{X_n}(x) = x - \dfrac{\sin(2n\pi x)}{2n\pi}$ för $0 \le x \le 1$. Visa att $X_n \xrightarrow{\text{D}} X$, där $X \sim \text{Re}(0,1)$, men att täthetsfunktionerna $f_{X_n}(x)$ saknar gränsvärde då $n \to \infty$.

Lösning. För $0 \le x \le 1$ så ser vi att $F_{X_n}(x) \to x$ då $n \to \infty$ eftersom sin-termen är begränsad. Men $F_X(x) = x$ för $0 \le x \le 1$ är fördelningsfunktionen för en likformig fördelning på $[0,1]$. Däremot ser vi att

$$f_{X_n}(x) = \frac{d}{dx} F_{X_n}(x) = 1 - \cos(2\pi n x) \to ???$$

då $n \to \infty$ för alla $0 < x < 1$.

Exempel

Låt $X_1, X_2, \ldots$ vara en följd Laplace-fördelade stokastiska variabler så att $f_{X_n}(x) = \dfrac{n}{2} e^{-n|x|}$. Visa att $X_n \xrightarrow{\text{D}} 0$ då $n \to \infty$ men att $f_{X_n}(x) \to 0$.

Lösning. Efter att ha sett föregående exempel så bör vi vara försiktiga med att titta direkt på täthetsfunktionerna. Mycket riktigt så konvergerar dessa mot 0, men det är ingen giltig täthetsfunktion. Låt oss titta på fördelningen istället:

$$F_{X_n}(x) = \begin{cases} \frac{1}{2} e^{nx}, & x < 0, \\ 1 - \frac{1}{2} e^{-nx}, & x \ge 0, \end{cases} \to \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & x = 0, \\ 1, & x > 0. \end{cases}$$

Alltså kommer $X_n \to 0$ ty $F_{X_n}(x) \to 0$ om $x < 0$ och $F_{X_n}(x) \to 1$ om $x > 0$. När $x = 0$ har vi en diskontinuitetspunkt, så där behöver vi ej ha konvergens.

I det diskreta fallet beter sig dock saker trevligare.

Sats. Låt $X_1, X_2, \dots$ och X vara stokastiska variabler sådana att $X_i(\Omega) \subset \mathbf{N}$ och $X(\Omega) \subset \mathbf{N}$ (det vill säga samtliga variabler antar endast icke-negativa heltal). Då gäller att

$$X_n \xrightarrow{\text{D}} X \quad \Leftrightarrow \quad \lim_{n\to\infty} p_{X_n}(k) = p_X(k), \; k = 0, 1, 2, \dots$$

Konvergens i fördelning är alltså ekvivalent med att *sannolikhetsfunktionerna* konvergerar i det diskreta fallet.

Lösning. Antag att $X_n \xrightarrow{\text{D}} X$. Då gäller att $F_{X_n}(x) \to F_X(x)$ för alla $x \in \mathbf{R} \setminus \mathbf{N}$ (alla positiva reella tal förutom heltalen) eftersom F_X är kontinuerlig då vi håller oss borta från heltalen. Låt $k = 0, 1, 2, \dots$ Då gäller att

$$p_{X_n}(k) = F_{X_n}(k + 1/2) - F_{X_n}(k - 1/2) \to F_X(k + 1/2) - F_X(k - 1/2) = p_X(k).$$

Omvänt, antag att $\lim\limits_{n\to\infty} p_{X_n}(k) = p_X(k)$ för varje $k = 0, 1, 2, \dots$ Då är

$$\begin{aligned}
\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty} P(X_n \le x) = \lim_{n\to\infty} \sum_{k=0}^{\lfloor x \rfloor} p_{X_n}(k) \\
&= \sum_{k=0}^{\lfloor x \rfloor} \lim_{n\to\infty} p_{X_n}(k) = \sum_{k=0}^{\lfloor x \rfloor} p_X(k) = P(X \le x) = F_X(x),
\end{aligned}$$

där det är ok att byta ordning på summa och gränsvärde eftersom summan är ändlig.

8.2 Ordo i sannolikhet

Redo för något riktigt skoj? För att beskriva hastigheten hos konvergens används ibland varianter av ordo-notationen ni har stött på tidigare (envariabel del 2).
Vi skriver att $X_n = o_p(a_n)$ om $\dfrac{X_n}{a_n} = o_p(1)$, där

$$Y_n = o_p(1) \quad \Leftrightarrow \quad Y_n \to 0 \text{ i sannolikhet.}$$

Vi säger att $X_n = O_p(a_n)$ om det för varje $\epsilon > 0$ finns ett ändligt $M > 0$ och ett ändligt $N > 0$ så att

$$P\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon \text{ för alla } n > N.$$

8.3 Ett par användbara resultat (utan bevis)

Vi har nu introducerat några begrepp kring konvergens av följder av stokastiska variabler. Ofta är man intresserad av funktioner av stokastiska variabler på olika sätt, så vad kan man säga om konvergensen efter att ha gjort någon form av sammansättning?
Till exempel kan vi notera att $X_n \xrightarrow{\text{D}} X$ och $Y_n \xrightarrow{\text{D}} Y$ *inte* medför att $X_n + Y_n \xrightarrow{\text{D}} X + Y$ eller $X_n Y_n \xrightarrow{\text{D}} XY$ i det generella fallet. Men under vissa förutsättningar har vi resultat som ofta duger när vi vill åt ovanstående.

The Continuous Mapping Theorem

Sats. Låt g vara kontinuerlig. Då gäller att

(i) $X_n \xrightarrow{\text{P}} X \quad \Rightarrow \quad g(X_n) \xrightarrow{\text{P}} g(X)$;

(ii) $X_n \xrightarrow{\text{D}} X \quad \Rightarrow \quad g(X_n) \xrightarrow{\text{D}} g(X)$.

Generaliserar till vektorvärda stokastiska variabler.

Kontinuerliga operationer fungerar alltså precis som vi förväntar oss. Konvergens i fördelning eller sannolikhet bevaras av kontinuerliga avbildningar. Beviset är inte jättekomplicerat, men bygger på argument och definitioner vi inte har tillgång till i nuläget (åter igen denna mått- och integrationsteori).

Man kan ju tro att summor av följder borde fungera lika enkelt, men så är inte fallet om vi endast har konvergens i fördelning. Det är alltså inte självklart vad som händer med $X_n + Y_n$ då $n \to \infty$. Men följande specialfall kan visas.

Slutskys sats

Sats. Om $X_n \xrightarrow{\text{D}} X$ och $Y_n \xrightarrow{\text{P}} c$, där c är en konstant, så gäller att

(i) $X_n + Y_n \xrightarrow{\text{D}} X + c$;

(ii) $X_n Y_n \xrightarrow{\text{D}} cX$;

(iii) $\dfrac{X_n}{Y_n} \xrightarrow{\text{D}} \dfrac{X}{c}$ under förutsättning att $c \neq 0$.

Även detta generaliserar till vektorvärda stokastiska variabler.

8.4 Delta-metoden

En naturlig fråga är följande: om vi vet att X_n har en asymptotisk fördelning, vad kan man säga om $g(X_n)$? Ett angreppsätt är givetvis att helt enkelt ta en Taylorutveckling av g och visa att resttermen beter sig som $o_p(1)$ så att den inte stör. Det generella fallet blir lite bökigt, så vi koncentrerar oss på normalfördelningen.

Delta-metoden för asymptotisk normalfördelning

Sats. Antag att $X_n \xrightarrow{\text{P}} \theta$ och $\sqrt{n}(X_n - \theta) \xrightarrow{\text{D}} X \sim N(0, \sigma)$ då $n \to \infty$ (i princip resultatet av centrala gränsvärdessatsen) och låt $g \in C^1$ i en omgivning av θ samt antag att $g'(\theta) \neq 0$. Då gäller att

$$\sqrt{n}\,(g(X_n) - g(\theta)) \xrightarrow{\text{D}} X \sim N(0, \sigma\sqrt{(g'(\theta))^2}),$$

då $n \to \infty$.

Bevis. Enligt medelvärdessatsen så gäller att

$$g(X_n) - g(\theta) = g'(\xi)\,(X_n - \theta)\,,$$

där ξ ligger mellan X_n och θ. Eftersom $X_n \xrightarrow{\text{P}} \theta$ så måste även $\xi \xrightarrow{\text{P}} \theta$. Satsen ovan om kontinuerliga avbildningar medför då att $g'(\xi) \xrightarrow{\text{P}} g'(\theta)$. Alltså måste

$$\sqrt{n}\,(g(X_n) - g(\theta)) \xrightarrow{\text{D}} Z \sim N(0, \sigma|g'(\theta)|),$$

allt enligt Slutskys sats! Vi kan även formulera det hela asymptotiskt enligt

$$\sqrt{n}\,(g(X_n) - g(\theta)) = g'(\theta)\sqrt{n}\,(X_n - \theta) + o_p(1),$$

genom att helt enkelt skriva

$$\begin{aligned}\sqrt{n}\,(g(X_n) - g(\theta)) &= \sqrt{n} g'(\xi)\,(X_n - \theta)\\ &= \sqrt{n}(X_n - \theta)g'(\theta) + \underbrace{\sqrt{n}\,(X_n - \theta)}_{=O_p(1)}\underbrace{(g'(\xi) - g'(\theta))}_{=o_p(1)},\end{aligned}$$

där vi nyttjar att $\sqrt{n}(X_n - \theta)$ konvergerar i fördelning – vilket betyder stokastiskt begränsad så $O_p(1)$ – och att $g'(\xi) \xrightarrow{\text{P}} g'(\theta)$. □

8.5 De stora talens lag

Vi kommer nu betrakta en fundamental situation i sannolikhetslära. Vi låter $X_1, X_2, \ldots$ vara en oändlig följd av oberoende och likafördelade stokastiska variabler. Vi låter $E(X_i) = \mu$ och $V(X_i) = \sigma^2$ (så vi antar att variansen är ändlig just nu). I vanlig ordning definierar vi det aritmetiska medelvärdet $\overline{X}_n$ av de n första variablerna i följden som

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

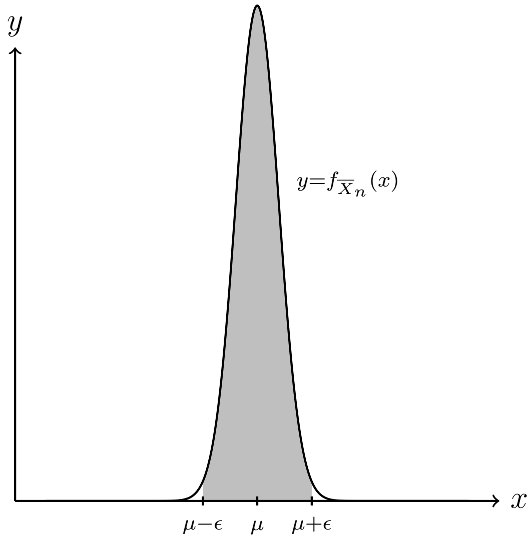
De stora talens lag (svag formulering)

Sats. För varje $\epsilon > 0$ gäller att

$$P(|\overline{X}_n - \mu| < \epsilon) \to 1 \text{ då } n \to \infty.$$

Med andra ord gäller att $\overline{X}_n \xrightarrow{\text{P}} \mu$ då $n \to \infty$.

En tolkning av satsen är att det aritmetiska medelvärdet av en följd oberoende och likafördelade variabler kommer att ha sin sannolikhetsmassa koncentrerad kring väntevärdet μ:



Variansen för $\overline{X}_n$ är som bekant

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

eftersom variablerna är oberoende (och vi antagit ändlig varians). Så då $n \to \infty$ ser vi att variansen för medelvärdet går mot noll. Konvergensen i de stora talens lag förefaller alltså rimlig. Ett mer ordentligt bevis följer från kända olikheter, så låt oss formulera dessa.

Markovs olikhet

Sats. Om X är en icke-negativ stokastisk variabel med ändligt väntevärde så gäller att

$$P(X \geq a) \leq \frac{E(X)}{a}, \quad a > 0.$$

Bevis. För det kontinuerliga fallet med täthetsfunktion, eftersom $a > 0$ och $f_X(x) \geq 0$,

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \geq \int_{a}^{\infty} x\, f(x)\, dx \geq a \int_{a}^{\infty} f_X(x)\, dx = aP(X \geq a),$$

så följer att $P(X \geq a) \leq \dfrac{E(X)}{a}$. Det diskreta fallet hanteras analogt (gör det!) $\square$

Tjebysjovs (Chebychevs) olikhet

Sats. Låt X vara en stokastisk variabel med ändligt väntevärde $E(X) = \mu$ och ändlig varians $V(X) = \sigma^2$, och låt $k > 0$. Då gäller att

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Bevis. Eftersom $(X-\mu)^2$ är en icke-negativ stokastisk variabel och $E(X-\mu)=0$, så gäller enligt Markovs olikhet att

$$
\begin{aligned}
P(|X-\mu| \geq k\sigma) = P((X-\mu)^2 \geq k^2\sigma^2) &\leq \frac{E((X-\mu)^2)}{k^2\sigma^2} \\
&= \frac{V(X-\mu)}{k^2\sigma^2} = \frac{1}{k^2},
\end{aligned}
$$

där den näst sista likheten är Steiners sats. □

En följd av denna olikhet är att vi får en direkt uppskattning av hur mycket sannolikhetsmassa som finns i intervall av typen $(\mu - k\sigma,\ \mu + k\sigma)$. Vi kan till exempel se att det finns minst 50% av sannolikhetsmassan om $k=\sqrt{2}$, minst 75% om $k=2$ och minst 96% om $k=5$.

Vanligt missförstånd

Låt oss (oberoende) kasta en sex-sidig balanserad tärning 1800 gånger. Vi förväntar oss att medelvärdet ligger nära 3.5. Antag att medelvärdet blev 4.0. Betyder detta enligt satsen ovan att vi kommer att få fler resultat $1, 2, 3$ än $4, 5, 6$ om vi kastar tärningen 1800 gånger till? Svaret är nej. De olika kasten anses oberoende, och kan därför inte påverkas av tidigare utfall. Så hur kan då satsen ovan gälla? Faktum är att det inte behöver vara fler låga resultat vid kommande upprepningar, det räcker med att medelvärdet av de nya resultaten är mindre än 4.0 för att vi ska hamna närmare 3.5 *totalt* sett.

Det lönar sig alltså inte att satsa mer pengar bara för att man förlorat så många gånger på rad (om händelserna är oberoende, annars kan lite vad som helst inträffa!).

Bevis av de stora talens lag. I princip följer detta direkt av olikheterna ovan. Låt $\epsilon > 0$. Vi ser direkt att

$$
P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{V(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0,
$$

då $n \to \infty$. □

Exempel

Låt X_n vara antalet krona vid n stycken oberoende slantsinglingar med ett ärligt mynt. Visa att $\dfrac{X_n}{n} \xrightarrow{\text{P}} \dfrac{1}{2}$.

Lösning. Låt $Y_k = 0$ om resultatet vid singling k är klave och $Y_k = 1$ om det är en krona. Då är $X_n = \sum_{k=1}^{n} Y_k$. Det är tydligt att

$$
E(Y_k) = \frac{1}{2}\cdot 0 + \frac{1}{2}\cdot 1 = \frac{1}{2},
$$

så enligt de stora talens lag gäller att

$$
\lim_{n\to\infty} P\left(\left| \frac{X_n}{n} - \frac{1}{2} \right| \geq \epsilon \right) = 0
$$

för varje $\epsilon > 0$. Detta är definitionen av konvergens i sannolikhet, så vi har visat att $\dfrac{X_n}{n} \xrightarrow{\text{P}} \dfrac{1}{2}$ då $n \to \infty$. Gå tillbaka till föreläsning 1 och betrakta frekvenstolkningen igen!

Det finns starkare formuleringar av de stora talens lag. Först och främst så behövs inte kravet på ändlig varians, men givetvis kan vi inte använda beviset ovan längre. Dessutom kan vi byta till nästan säker konvergens istället för i sannolikhet. Det sista brukar brukar kallas för den starka formuleringen av de stora talens lag. Så mycket starkare konvergens än så kan vi inte få.

De stora talens lag (stark formulering)

Sats. Låt $X_1, X_2, \dots$ vara en följd av oberoende och likafördelade stokastiska variabler. Då är

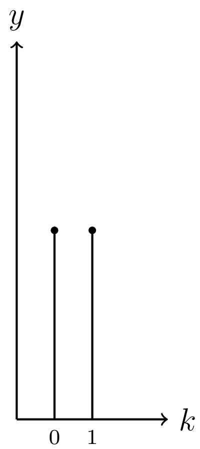
$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1.$$

Med andra ord gäller att $\overline{X}_n \xrightarrow{\text{a.s.}} \mu$ då $n \to \infty$.

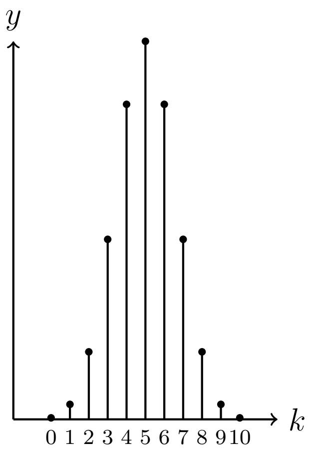
8.6 Centrala gränsvärdessatsen

Alla vägar leder till Rom. Eller åtminstone: alla fördelningar leder till normalfördelning? Faktum är att det är precis det den centrala gränsvärdessatsen säger: summan av ett stort antal oberoende och likafördelade stokastiska variabler är approximativt normalfördelad.

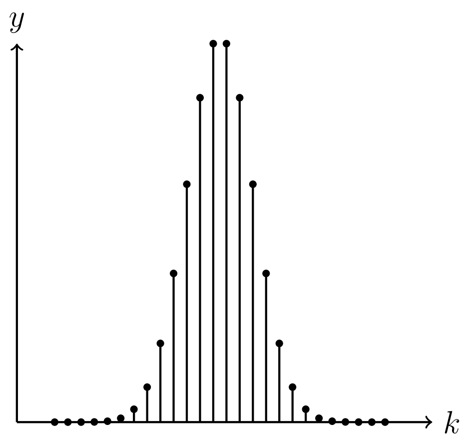
Vi betraktar ett exempel. Låt oss utföra det klassiska experimentet med slantsingling och räkna antalet X kronor vid ett visst antal, säg n, kast. Från tidigare exempel (inbrottstjuven) så vet vi att $X \sim \text{Bin}(n, p)$, där $p = 1/2$ om myntet är rättvist. En binomialfördelad variabel kan ses som en summa av oberoende Bernoulli-fördelade variabler X_k, en variabel för varje försök (slantsingling), där $X_k = 0$ om försök nr k "misslyckas" (klave), och $X_k = 1$ om försök k lyckas (krona). Alltså kan vi skriva $X = \displaystyle\sum_{k=1}^{n} X_k$. Varje X_k har sannolikhetsfunktionen $p_{X_k}(1) = p$ och $p_{X_k}(0) = 1 - p$. Med andra ord, binomialfördelningen kan ses som en summa av oberoende och likafördelade variabler. Om bara n är tillräckligt stort borde vi i så fall närma oss normalfördelningen. Hur stort? Vi skisserar några fall när n blir större och större och $p = 0.5$.



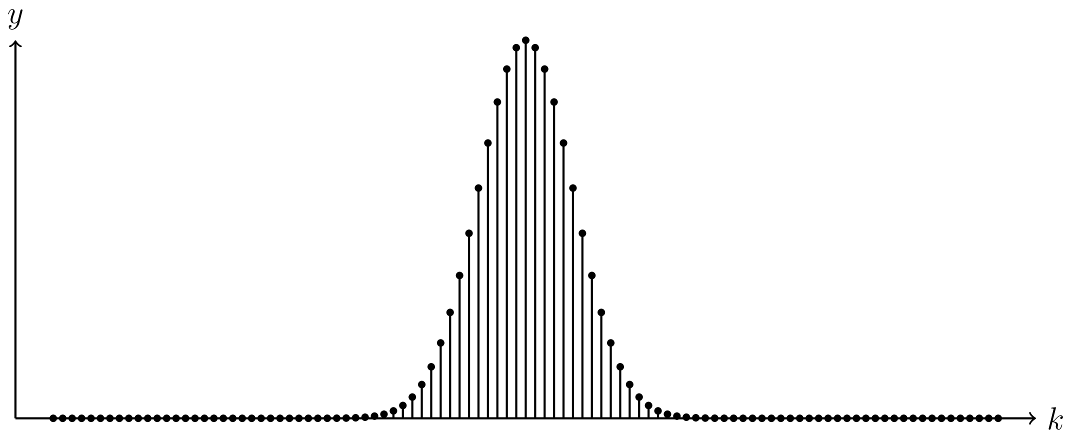
Med $n = 1$.



Med $n = 10$.



Med $n = 25$.



Med $n = 100$.

Här ser vi ganska tydligt att ju större n blir, desto mer lik blir sannolikhetsfördelning en normalfördelningskurva. Följande sats verkar alltså rimlig (åtminstone i Binomialfallet).

Centrala gränsvärdessatsen (CGS)

Sats. Låt $X_1, X_2, \ldots$ vara en oändlig följd av likafördelade och oberoende stokastiska variabler. Vidare, låt $E(X_k) = \mu$ och $V(X_k) = \sigma^2$ för $k = 1, 2, \ldots$. Då gäller att $Y_n = \sum_{k=1}^{n} X_k$ konvergerar i fördelning enligt $\frac{Y_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{D}} Z \sim N(0,1)$. Även medelvärdet konvergerar i fördelning enligt $\sqrt{n}(\overline{X} - \mu) \xrightarrow{\text{D}} Z$, där $Z \sim N(0,1)$.

Vi noterar att resultaten i satsen kan formuleras enligt följande (kanske mer lättanvänt).

CGS: Alternativ formulering

(i) summan $X = \sum_{k=1}^{n} X_k$ uppfyller

$$P\left(a < \frac{X - n\mu}{\sigma\sqrt{n}} < b\right) \to \Phi(b) - \Phi(a) \text{ då } n \to \infty$$

för alla $a, b \in \mathbf{R}$ med $a < b$. Vi säger att X är **asymptotiskt** normalfördelad.

(ii) medelvärdet $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ uppfyller

$$P\left(a < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < b\right) \to \Phi(b) - \Phi(a) \text{ då } n \to \infty$$

för alla $a, b \in \mathbf{R}$ med $a < b$.

Beviset för satsen faller utanför ramen för denna kurs. Se, till exempel, Rick Durret: *Probability: Theory and Examples* eller Allan Gut: *An Intermediate Course in Probability*. För er som läser TAMS15 återkommer vi till detta.
Så hur använder vi CGS?

Approximation via CGS

Med samma beteckningar och förutsättningar som ovan så är

$$P(X \leq x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right) \quad \text{och} \quad P(\overline{X} \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right), \quad x \in \mathbf{R},$$

om n är stort. Oftast brukar $n \geq 30$ duga, men skeva fördelningar kräver större n. Vi skriver $X \overset{\text{appr.}}{\sim} N(n\mu, \sqrt{n}\sigma)$ och $\overline{X} \overset{\text{appr.}}{\sim} N(\mu, \sigma/\sqrt{n})$; variablerna är **approximativt** normalfördelade.

Exempel

Vad är sannolikheten att summan av 50 stycken slumptal mellan 0 och 2 överstiger 53?

Lösning: Vi antar att slumptalen är likformigt fördelade, så varje slumptal $X_k \sim \text{Re}(0, 2)$, och att slumptalen är oberoende av varandra. Det råder likformig fördelning, så $E(X_k) = 1$ och $V(X_k) = 1/3$. Varför? Enkelt att se från definitionen:

$$E(X_k) = \int_0^2 x\frac{1}{2}\,dx = \left[\frac{x^2}{4}\right]_0^2 = 1$$

och

$$V(X_k) = \int_0^2 x^2\frac{1}{2}\,dx - 1^2 = \left[\frac{x^3}{6}\right]_0^2 - 1 = 1/3.$$

Så vi har en summa av 50 stycken likformigt fördelade variabler X_k med samma väntevärde och varians. Låt $X = \sum_{k=1}^{50} X_k$. CGS implicerar att $X \overset{\text{appr.}}{\sim} \text{N}(50, \sqrt{50/3})$. Alltså erhåller vi

$$P(X > 53) = 1 - P(X \leq 53) \approx 1 - \Phi(3/\sqrt{50/3}) = 1 - \Phi(0.7348) \approx 0.2312.$$

Det är alltså ca 23% chans att summan överstiger 53.

Exempel

Antag att samtalstiderna till *1177* är oberoende och exponentialfördelade med väntevärde 15 minuter. Om en sjuksköterska förväntas svara på 28 samtal under ett åtta-timmars pass, vad är sannolikheten att hon lyckas?

Lösning: Låt $X_k \sim \text{Exp}(1/15)$ vara tiden för samtal k, $k = 1, 2, \ldots, 28$. Den totala tiden för 28 samtal ges av $X = \sum_{k=1}^{28} X_k$. Faktum är att man kan visa att X blir gamma-fördelad (se Blom et al.), men den fördelningen är ganska bökig att arbeta med. Vad säger CGS? Vi har kring 30 stycken samtal, så $X \overset{\text{appr.}}{\sim} \text{N}(28 \cdot 15, \sqrt{28} \cdot 15) = \text{N}(420, \sqrt{6300})$. Alltså är

$$P(X \leq 8 \cdot 60) \approx \Phi\left(\frac{480 - 420}{\sqrt{6300}}\right) \approx \Phi(0.76) = 0.7764.$$

Nästan 80% chans alltså! Hur bra stämmer då detta? Man kan härleda att X i själva verket har fördelningen $X \sim \Gamma(28, 1/15)$, så $P(X \leq 480) = 0.7838$ (MATLAB, `gamcdf`).

8.7 Flerdimensionella centrala gränsvärdessatsen

Det finns motsvarande satser som gäller i högre dimensioner, men i vanlig ordning har vi nu en hel kovariansmatris att hålla ordning på istället för endast variansen. En variant av den multivariata CGS kan formuleras enligt följande.

Flerdimensionella centrala gränsvärdessatsen

Sats. Låt $\boldsymbol{X} = (X_1, \ldots, X_k)$ vara en vektorvärd stokastisk variabel med kovariansmatris $\boldsymbol{\Sigma}$. Implicit här är att $V(X_j) < \infty$ för $j = 1, 2, \ldots, k$. Låt $\boldsymbol{X}_n$ vara en följd av oberoende vektorer med samma fördelning som $\boldsymbol{X}$. Då gäller att

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{X}_i - E(\boldsymbol{X})) \xrightarrow{\text{D}} N(0, \boldsymbol{\Sigma}).$$

Notera även att

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{X}_i - E(\boldsymbol{X})) = \sqrt{n}\left(\overline{\boldsymbol{X}}_n - E(\boldsymbol{X})\right),$$

så vi kan mer kompakt skriva $\sqrt{n}\left(\overline{\boldsymbol{X}}_n - E(\boldsymbol{X})\right) \to N(0, \boldsymbol{\Sigma})$.

8.7.1 Delta-metoden i flera variabler

Vad händer i flera dimensioner? I princip är det helt analogt med envariabelfallet. Låt $\widehat{\boldsymbol{\theta}}$ vara en konsistent skattning av $\boldsymbol{\theta}$ (vilket innebär att $\widehat{\boldsymbol{\theta}} \xrightarrow{\text{P}} \boldsymbol{\theta}$) så att

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{\text{D}} Z \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Om vi för enkelhetens skull antar att $g \in C^2$ i en omgivning av $\boldsymbol{\theta}$, så är som bekant

$$g(\widehat{\boldsymbol{\theta}}) = g(\boldsymbol{\theta}) + \nabla g(\boldsymbol{\theta})^T \cdot \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) + R(\widehat{\boldsymbol{\theta}}).$$

Om vi betraktar variansen för vår approximation (där vi bortser från resttermen) så ser vi att

$$\begin{aligned} V\left(g(\boldsymbol{\theta}) + \nabla g(\boldsymbol{\theta})^T \cdot \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\right) &= V\left(\nabla g(\boldsymbol{\theta})^T \cdot \widehat{\boldsymbol{\theta}}\right) = \text{Cov}\left(\nabla g(\boldsymbol{\theta})^T \cdot \widehat{\boldsymbol{\theta}}\right) \\ &= \nabla g(\boldsymbol{\theta})^T \text{Cov}(\widehat{\boldsymbol{\theta}}) \nabla g(\boldsymbol{\theta}) = \nabla g(\boldsymbol{\theta})^T \frac{1}{n} \boldsymbol{\Sigma} \, \nabla g(\boldsymbol{\theta}). \end{aligned}$$

Genom att likt i envariabelfallet använda medelvärdessatsen kan vi nu visa att

$$\sqrt{n}\left(g(\widehat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})\right) \xrightarrow{\text{D}} Z \sim N\left(\mathbf{0}, \nabla g(\boldsymbol{\theta})^T \boldsymbol{\Sigma} \nabla g(\boldsymbol{\theta})\right).$$