SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

En introduktion till stora avvikelser

av

Simon Blom

2023 - K3

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

En introduktion till stora avvikelser

Simon Blom

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Alan Sola

2023

1 Sammanfattning

Detta arbete ger en introduktion till stora avvikelser och några exempel på när vi kan använda denna kunskap. Detta arbete är riktat till dem som vill ha en grundläggande förståelse för satsen om stora avvikelser.
Vi börjar med att gå igenom två essentiella satser samt vilka begrepp som kommer att användas för att förstå sig på dessa satser, stora talens lag samt centrala gränsvärdessatsen. Vi kommer att gå igenom vad bägge satserna säger samt bevisa dem för att sedan utnyttja dessa kunskaper för att formulera en sats om stora avvikelser och sedan härleda ett bevis.

This work gives an introduction to large deviations and a few examples for when we can use this knowledge. This paper is written for those who want a basic knowledge for understanding the theroem of large deviations.
We begin with two essential theorems as well as the concepts we will use to understand these theroems, the law of large numbers and the central limit theorem. We will explore both theorems as well as proving them so that we then can then use them to formulate the theorem of large deviations as well a proving it.

Innehåll

1 Sammanfattning **1**

2 Introduktion **3**

3 Begrepp **3**

4 Stora avvikelser, en början (med exempel) **12**

5 Stora talens lag (STL) **13**

6 Centrala gränsvärdessatsen (CGS) **15**

7 Stora avvikelser **21**

8 Källor **27**

2 Introduktion

Tänk dig att du är på väg till jobbet. På vägen måste du passera fem övergångställen. Skulle man då kunna beräkna, med säkerhet, hur lång tid det tar för dig att komma fram? Ja det är möjligt, vi kan mäta din hastighet och avståndet till jobbet. Det knepiga införs med övergångställena. Växlar de alltid efter en viss tid, eller har de sensorer som känner att en bil väntar på grönt men just nu står ingen vid övergångstället? I så fall växlar rödljuset snabbare till grönt för bilen. Därefter återgår trafiksignalen till sitt intervall. Detta ger oss en osäkerhet i beräkningen som gör att det är näst intill omöjligt att beräkna, på sekunden, hur lång tid det tar för dig att komma till jobbet. Istället kan vi räkna ut sannolikheten att det tar S antal sekunder att ta sig till jobbet.
Men tänk om du har ett viktigt möte en dag men du bestämmer dig för att chansa och i alla fall ta den tid på dig som det brukar ta för att komma till jobbet. Men vad är sannolikheten att till exempel alla lysena är röda denna morgon? Dessa extrema sannolikheter ska vi utforska i denna text.

3 Begrepp

Innan vi börjar med texten så är det lämpligt att gå igenom några viktiga begrepp som kommer att förekomma.
Formuleringen av begreppen följer nära boken Stokastik [1] av Sven Erick Alm och Tom Britton.

Definition 1 (Slump)**.** (Stokastik Sid 1)
Vi tänker oss att vi har ett försök där utfallet inte är förutsägbart. Vi känner till vilka möjligheter som finns, men vi kan inte säga vilket av alternativen som kommer att ske innan vi utför försöket.

Exempel 1. Vi kan kasta ett mynt. Vi vet att vi kan få krona eller klave, men innan vi kastat myntet vet vi inte vilket av alternativen vi kommer att få med säkerhet.

Definition 2 (Utfallsrum)**.** (Stokastik Sid 5)
Resultatet av ett slumpförsök kallas ett $utfall$. Mängden av möjliga utfall från ett visst slumpförsök kallas $utfallsrum$.
En viss specificerad mängd utfall kallas för en *händelse* - således är enskilda utfall, liksom hela utfallsrummet, också händelser.
Enskilda utfall betecknas med $u_1, u_2, ...$, händelser betecknas med versaler $A, B, ...$ och utfallsrummet med Ω. Utfallsrum med ändligt eller uppräkneligt oändligt många (alltså lika stor som mängden av naturliga tal) utfall kallas *diskreta* utfallsrum medan övriga kallas *kontinuerliga* utfallsrum.

Definition 3 (Slumpexperiment/slumpförsök). (Stokastik Sid 5 och 10)
En situation där något kommer att inträffa men vi vet inte på förhand, med säkerhet, vad.
Med hjälp av definitionen för utfallsrum kan vi definiera slumpförsök på följande vis; ett slumpförsök på ett utfallsrum består av ett försök som resulterar i ett av utfallen i utfallsrummet, där man på förhand inte kan veta exakt vilket av utfallen i utfallsrummet som kommer att inträffa.

Definition 4 (Sannolikhet för händelse, eller Sannolikhetsfunktion). (Stokastik Sid 10)
Vi kan beskriva slumpförsöket genom att precisera sannolikheterna för alla händelser i utfallsrummet. Sannolikheten för händelsen A brukar skrivas $P(A)$. För att funktionen $P(A)$ ska få kallas en sannolikhetsfunktion ställer vi dock vissa krav.
1. $0 \leq P(A) \leq 1$ för alla händelser $A \subset \Omega$,
2. $P(\Omega) = 1$,
3. om $A \cap B = 0$ så gäller $P(A \cup B) = P(A) + P(B)$.
Dessa villkor brukar kallas för Kolmogorovs axiom.

Exempel 2. Den vanligaste tolkningen av en sannolikhet, till exempel, $P(A) = \frac{3}{10}$, är att om man upprepar slumpförsöket många gånger så kommer den relativa frekvensen för händelsen A, det vill säga andelen försök där A inträffar, att ligga nära $\frac{3}{10}$.

Definition 5 (Slumpvariabel). (Stokastik Sid 42)
En *slumpvariabel* (alternativt *stokastisk variabel*) $X(u)$ är en reellvärd funktion definierad på utfallsrummet, $X : \Omega \to \mathbf{R}$. När slumpförsöket genomförts och ett utfall erhållits sägs funktionens värde för utfallet vara en *observation* av slumpvariabeln.

Exempel 3. Vi kastar 2 rättvisa tärningar. Vi har då 36 utfall där varje utfall har sannolikheten $\frac{1}{36}$ att rullas. Däremot har slumpvariablernas värden inte samma sannolikhet. Vi kan, till exempel, få 7 ögon/prickar på tärningen på betydligt fler sätt än att få 3 ögon. Z står för hur många ögon/prickar de två tärningarna visar tillsammans.

$$P(Z = 7) = \frac{6}{36} = \frac{1}{6}$$

Tärningsparen ovan blir då: $(1 + 6), (2 + 5), (3 + 4), (4 + 3), (5 + 2), (6 + 1)$.

$$P(Z = 3) = \frac{2}{36} = \frac{1}{18}$$

Tärningsparen ovan blir då: $(1+2), (2+1)$.

Definition 6 (Diskret slumpvariabel). (Stokastik Sid 44)
En slumpvariabel X är diskret om den endast kan anta ändligt eller uppräkneligt oändligt antal värden $x_1, x_2, ...$

Exempel 4. Vi kan skriva upp en tabell på alla kombinationer av ögon två tärningar kan visa.

2	3	4	5	6	7	8	9	10	11	12
1+1	1+2	1+3	1+4	1+5	1+6	2+6	3+6	4+6	5+6	6+6
	2+1	3+1	4+1	5+1	6+1	6+2	6+3	6+4	6+5	
		2+2	2+3	2+4	2+5	3+5	4+5	5+5		
			3+2	4+2	5+2	5+3	5+4			
				3+3	3+4	4+4				
					4+3					

Definition 7 (Kontinuerlig slumpvariabel). (Stokastik Sid 52, 55)
En slumpvariabel X sägs vara kontinuerlig om det finns en funktion $f_X(x)$ sådan att det för "alla" mängder A gäller att

$$P(X \in A) = \int_A f_X(t)dt.$$

Viktigt att notera, sannolikheten att en kontinuerlig slumpvariabel antar ett bestämt enskilt värde är 0. Detta följer av att integralen över ett allt mindre intervall går mot 0.

$$P(X = x) \leq P(x-h < X \leq x+h) = \int_{x-h}^{x+h} f_X(t)dt.$$

Denna integral går mot 0 då h går mot 0, se Stokastik sid 55.

När vi säger "alla" mängder så menas åtminstone alla öppna och slutna mängder i $\mathbf{R}$. För en sluten mängd av reella värden är randen av intervallet inkluderad, vilket den inte är i en öppen mängd.

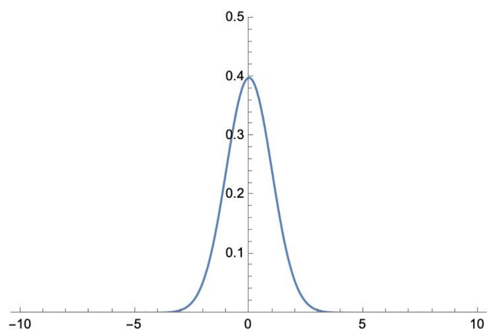
Definition 8 (Sannolikhetsfunktion). (Stokastik Sid 44)
Sannolikhetsfunktionen, p_X, för en diskret slumpvariabel X definieras av

$$p_X(x) := P(X = x) = P(X \text{ antar värdet } x),$$

där $x \in (x_1, x_2, ...)$ och x_j är de möjliga värdena X kan anta.

Definition 9 (Fördelning (Fördelningsfunktion)). (Stokastik Sid 48 och 53)
Ett sätt att uttrycka slumpstrukturen hos slumpvariabeln.
Fördelningsfunktionen $F_X(t)$ för en slumpvariabel X definieras av
$F_X(t) := P(X \leq t), \; -\infty < t < \infty$.

Definition 10 (Täthetsfunktion). (Stokastik Sid 52)
Funktionen $f_X(\cdot)$ (se kontinuerlig slumpvariabel ovan) kallas för slumpvariabelns täthetsfunktion.



(Figur X. Täthetsfunktion $f(x)$ för en kontinuerlig slumpvariabel.)

Definition 11 (Normalfördelning). (Stokastik Sid 100)
En kontinuerlig slumpvariabel X sägs vara normalfördelad med parametrar μ och σ^2 ($\sigma^2 > 0$) om täthetsfunktionen ges av

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

Beteckning: $X \sim N(\mu, \sigma^2)$.

Definition 12 (Medelvärde). (Stokastik Sid 243)
Medelvärde är ett lägesmått för att beskriva var och hur statistisk data är utspritt. I fallen vi är intresserade av är detta reella tal. Vi sätter

$$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + ... + x_n),$$

där $x_1, ..., x_n$ är våra observationer.

Definition 13 (Väntevärde). (Stokastik Sid 57)
Väntevärdet för en slumpvariabel X betecknas med $E(X)$, μ_X, eller bara μ om det inte kan förväxlas med andra väntevärden, och är ett reellt tal. För en diskret slumpvariabel definieras det som

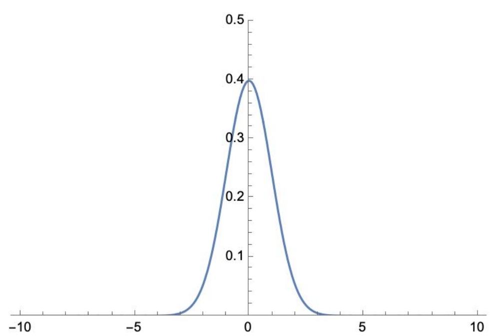
$$E(X) := \sum_{k=1}^{\infty} x_k P(X = x_k),$$

där $(x_j)_{j=1}^{\infty}$ är värdena slumpvariabeln X antar.

Och för en kontinuerlig som

$$E(X) := \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

Väntevärdet är ett mått på storleken av, eller läget för, en slumpvariabel. En slumpvariabel kan anta olika värden så väntevärdet ger oss ett sätt att väga in sannolikheten för olika värden för att sedan beskriva väntevärdet som en storlek.
Den fysikaliska tolkningen av väntevärde är tyngdpunkt. Om vi, till exempel, ritar upp täthetsfunktionen för en slumpvariabel X är $E(X)$ positionen som gör att figuren "balanserar", se bilden nedan.



För två slumpvariabler X och Y gäller: operationen att ta väntevärde är linjär

$E[X + Y] = E[X] + E[Y]$.

Definition 14 (Varians). (Stokastik Sid 63)
Variansen $V(X)$ för en slumpvariabel X med väntevärde μ definieras som $V(X) := E((X - \mu)^2)$ om detta väntevärde är ändligt. Variansen betecknas ofta σ^2 eller σ_x^2.

Varians är ett mått på spridningen, det ger oss ett sätt att representera om slumpen ofta ligger nära sitt väntevärde eller om den ofta avviker från väntevärdet.

Exempel 5. Antag att $V(X) = 0$.
Detta är samma som $E((X - \mu)^2) = 0$.
Det vill säga $X - \mu = 0$.
Alltså $X = \mu$, så då är slumpvariabeln konstant.

Exempel 6. $V(X)$ kan också skrivas som $E(X^2) - (E(X))^2$.
Vi börjar med beteckningen $m = E(X)$.

$V(X) = E((X - m)^2) = E(X^2 - 2mX + m^2) = E(X^2) - 2mE(X) + m^2 = E(X^2) - m^2 = E(X^2) - (E(X))^2$.

Definition 15 (Standardavvikelse)**.** (Stokastik Sid 65)
Standardavvikelsen $D(X)$ för en slumpvariabel X definieras som

$$D(X) := \sqrt{V(X)}.$$

(Observera att $V(X) \geq 0$.)

Standardavvikelsen betecknas ofta med σ eller σ_x.

Definition 16 (Standardiserad slumpvariabel)**.** (Stokastik Sid 66)
Låt Y vara en slumpvariabel med väntevärde $\mu = E(Y)$ och standardavvikelse $\sigma = D(Y)$. Då är $X = \dfrac{(Y - \mu)}{\sigma}$ motsvarande standardiserade slumpvariabel. Med denna transformation får vi en slumpvariabel X med väntevärde 0 och standardavvikelse 1.

Det är inte så självklart att väntevärdet för X blir 0 och standardavvikelsen för X blir 1, så vi verifierar detta. Vi börjar med väntevärdet för X.

$$E(X) = E(\frac{Y - \mu}{\sigma}) = E(\frac{Y}{\sigma} - \frac{\mu}{\sigma}) = E(\frac{Y}{\sigma}) + E(-\frac{\mu}{\sigma}).$$

Ovan har vi utnyttjat att väntevärdet är linjärt. Här ser vi att $E(-\frac{\mu}{\sigma})$ endast består av konstanter. Och väntevärdet av en konstant blir konstanten själv. Vi får då

$$E(\frac{Y}{\sigma}) - \frac{\mu}{\sigma} = \frac{1}{\sigma}E(Y) - \frac{\mu}{\sigma} = \frac{1}{\sigma} \cdot \mu - \frac{\mu}{\sigma} = 0.$$

Så vi ser då att $E(X) = 0$.
Då går vi vidare till standardavvikelsen för X. Vi använder oss av definitionen för varians.
Observera att $\mu_X = 0$, det vi precis verifierade, detta leder till att vi får
$E((X - \mu_X)^2) = E(X^2) = E((\frac{Y-\mu}{\sigma})^2) = E(\frac{(Y-\mu)^2}{\sigma^2}) = \frac{1}{\sigma^2}E((Y - \mu)^2)$.
Här kan vi se att detta är definitionen för variansen för Y
$\frac{1}{\sigma^2}E((Y - \mu)^2) = \frac{1}{\sigma^2}V(Y) = \frac{1}{\sigma^2}\sigma^2 = 1$.
Så vi ser då att $D(X) = 1$.

Definition 17 (Standardiserad normalfördelning). (Stokastik Sid 103)
En normalfördelad slumpvariabel Z med parametrar $\mu = 0$ och $\sigma = 1$ sägs vara standardiserad normalfördelad, $Z \sim N(0, 1)$. Täthetsfunktionen $f_Z(z)$ och fördelningsfunktionen $F_Z(z)$ för en standardiserad normalfördelning har egna beteckningar $\varphi(z)$ respektive $\Phi(z)$. Vi kan nu skriva in värdena för μ och σ för definitionen vi redan gett för normalfördelning, vi får då följande

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, \quad -\infty < z < \infty,$$

och

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt.$$

Definition 18 (Betingad sannolikhet). (Stokastik Sid 22)
Antag att för händelsen A gäller $P(A) > 0$. Den betingade sannolikheten för händelsen B, givet att A har inträffat, skrivs $P(B|A)$ och definieras som

$$P(B|A) := \frac{P(B \cap A)}{P(A)}.$$

Definition 19 (Oberoende händelser). (Stokastik Sid 25-26)
Två händelser A och B är oberoende om $P(A|B) = P(A)$ förutsatt att $P(B) > 0$, och $P(B|A) = P(B)$ förutsatt att $P(A) > 0$.

Alternativ definition: Två händelser A och B är oberoende om

$$P(A \cap B) = P(A) \cdot P(B).$$

Om två händelser är oberoende betyder det informellt att vare sig händelse A sker eller ej påverkar det inte sannolikheten för händelse B.

Definition 20 (Oberoende slumpvariabler)**.** (Stokastik Sid 125,126,(131))
Två slumpvariabler X och Y sägs vara oberoende om deras simultana fördelning satisfierar $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, för alla (x, y), i det diskreta fallet, respektive att

$$f_{X,Y}(x, y) = f_X(x)f_Y(y),$$

för "alla" (x, y), i det kontinuerliga fallet.

Definition 21 (Gränsvärde)**.** (Analys i en variabel Sid 136,137)
I fallet då $x \to +\infty$ definieras gränsvärde på följande sätt.

Antag att $f(x)$ är en funktion vars definitionsmängd innehåller godtyckligt stora reella tal. Vi säger att $f(x)$ har gränsvärdet A då x går mot oändligheten om det för varje givet tal $\epsilon > 0$ finns ett tal ω (beroende av ϵ) sådant att

$$x > \omega \textbf{ och } x \in D_f \longrightarrow |f(x) - A| < \epsilon.$$

Detta skrivs som $\lim_{x\to+\infty} f(x) = A$.

Innebörden i definitionen är följande: funktionen har gränsvärdet A då $x \to \infty$ om funktionsvärdena $f(x)$ uppfyller varje givet toleranskrav av formen

$$A - \epsilon < f(x) < A + \epsilon$$

så snart x är tillräckligt stort, det vill säga för alla $x > \omega$. Ju större noggrannhet (dvs. ju mindre ϵ) som anges, desto större måste i allmänhet ω väljas för att toleranskravet ska vara uppfyllt för alla $x > \omega$.

Definition 22 (Konvergens av talföljd)**.** (Analys i en variabel Sid 137, Stokastik Sid 160)
Vi säger att talföljden $(a_n)_{n=1}^{\infty}$ konvergerar till A om det för varje $\epsilon > 0$ existerar ett $N \in \mathbf{N}$ sådant att

$$|a_n - A| < \epsilon$$

för alla $n \geq N$.

Definition 23 (Konvergens i sannolikhet)**.** (Analys i en variabel Sid 137, Stokastik Sid 160)
Vi säger att en följd $(X_k)_{n=1}^{\infty}$ av slumpvariabler konvergerar i sannolikhet mot slumpvariabeln X om

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

för varje $\epsilon > 0$.

Resultat från grundläggande analys

Definition 24 (Taylorutveckling)**.** (Analys i en variabel Sid 426)
Taylorutvecklingar handlar om att approximera funktioner med polynom som är lätta att hantera.
Metoden går ut på att hitta ungefär hur en funktion ser ut vid en viss punkt genom att utnyttja ett antal derivator till funktionen i den punkten. Informellt sagt, ju mer man deriverar, desto mer exakt blir approximationen!
Ett exempel på en Taylorutveckling är

$$e^x = 1 + x + \frac{x^2}{2} + ... + \frac{x^n}{n!} + \frac{f^{(n+1)}(\Theta x)}{(n+1)!}x^{n+1}.$$

Resttermen R_n kan uttryckas med hjälp av $f^{(n+1)}$.

Sats 1 (Maclaurins formel)**.** (Person Böiers Sid 425)
Antag att funktionen f och dess derivator till och med ordning $n+1$ är kontinuerliga i en omgivning av 0. Då gäller för alla x i denna omgivning

$$f(x) = f(0) + \tfrac{f'(0)}{1!}x + \tfrac{f''(0)}{2!}x^2 + ... + \tfrac{f^{(n)}(0)}{n!}x^n + \tfrac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1},$$

där talet θ beror av n och x och där $0 \leq \theta \leq 1$.

Definition 25 (Talet e)**.** $e = \lim_{n\to\infty}(1 + \frac{1}{n})^n$.

(Detta är ett viktigt resultat vi kommer använda senare.)

Sats 2. Antag $E(X) = 0$ och $E(Y) = 0$, samt att X och Y är oberoende.
Så $E(X + Y) = 0$.
Då är variansen

$$V(X + Y) = E((X + Y)^2) = V(X) + V(Y).$$

Bevis. Vi vet att väntevärdet av en summa är summan av väntevärdena om de är oberoende.
$E((X+Y)^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2)$.
Vi har då $E(X^2) = V(X)$
om $E(Y^2) = V(Y)$
samt $2E(XY) = 0$, om X och Y är okorrelerade.
Alltså följer $V(X+Y) = V(X) + V(Y)$.

□

4 Stora avvikelser, en början (med exempel)

Så vad är teorin om stora avvikelser? Den är till för att ge oss en ungefärlig sannolikhet för sällsynta situationer. För att få en enkel förståelse så kan vi tänka oss att vi har ett rättvist mynt, alltså att det är 50% chans att den visar krona (och i så fall självklart också 50% chans för klave). Så vi flippar detta mynt 100 gånger och antecknar hur många kronor vi får. Därefter upprepar vi experimentet tusentals gånger. När vi upprepat detta test ett stort antal gånger, och sedan ritat ut en graf för antalet gånger vi fick krona, så kommer denna graf med stor sannolikhet att likna en normalfördelad kurva. (Fortsätter vi med experimenten så kommer grafen troligtvis att likna denna graf nu mer och mer).

Vi ser då att vi med stor sannolikhet kommer att få krona 50 av kasten per experiment. Så vad är en sällsynt situation i myntkastning? Jo det skulle ju vara att vi säg får 99 kronor, eller 1 krona, av 100 kast.

Innan vi fortsätter med matematiken bakom teorin så kan det vara värt att nämna varför, och främst när, detta kan vara intressant att studera i vardagen. De som skulle kunna tänka sig använda denna teori är exempelvis försäkringsbolag. De beräknar hur mycket de vanligtvis behöver betala ut till de som köpt försäkringar hos bolaget över flera månaders tid. Men det finns alltid en chans att ett visst år är väldigt dåligt, inget är säkert. Så då vill bolagen veta, hur dålig är ett "dåligt år"? Skulle de kunna klara av att betala ut försäkringar då, och hur stor är sannolikheten för ett sådant år? Om sannolikheten är 25% så borde de ändra sina priser!

Nu har vi sett ett enkelt exempel på vad stora avvikelser är och när det används i vardagen. Så vi kan snart gå vidare till hur teorin fungerar rent matematiskt. Men först så måste vi bekanta oss med några andra lagar och satser som vi kommer att använda oss av för att förstå teorin om stora avvikelser. De lagarna och satserna är Lagen om stora tal och Centrala gränsvärdes-satsen.

5 Stora talens lag (STL)

Stora talens lag säger att om vi utför ett försök och upprepar det många gånger så bör den relativa andelen som resulterar i händelse A ligga nära sannolikheten för A, P(A). Vi visar detta med satsen nedan, citerad från boken Stokastik av Alm och Britton, sida 160. (Notera att satsen visas för medelvärden av slumpvariabler som inte nödvändigtvis har samma fördelning men har samma väntevärde och standardavvikelse.)

Sats 3 (Stora talens lag)**.** Låt $X_1, X_2, X_3, X_4, \ldots$ vara en följd av oberoende slumpvariabler, alla med samma väntevärde $E(X_i) = \mu$ och standardavvikelse $D(X_i) = \sigma < \infty$.
Låt $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ vara medelvärdet av de n första variablerna. Då gäller, för godtyckligt $\epsilon > 0$, att

$$P(|\bar{X}_n - \mu| > \epsilon) \to 0, \text{ då n} \to \infty.$$

Med andra ord konvergerar $\bar{X}_n$ i sannolikhet till μ.

Bevis. Nu ska vi bevisa denna sats. Vi börjar med att bevisa att $E(\bar{X}_n) = \mu$. Det gör vi på följande vis

$$E[\tfrac{1}{n}(X_1 + \ldots + X_n)] = E[\tfrac{1}{n}X_1] + \ldots + E[\tfrac{1}{n}X_n] = \tfrac{1}{n}E[X_1] + \ldots + \tfrac{1}{n}E[X_n].$$

Vi har redan som krav att slumpvariablerna har samma väntevärde, vilket vi kan använda oss av

$$\tfrac{1}{n}E[X_1] + \ldots + \tfrac{1}{n}E[X_1] = \tfrac{1}{n}(\textstyle\sum_1^N E[X_1]) = \tfrac{1}{n} \cdot nE[X_1] = E[X_1] = \mu.$$

Beviset för variansen följer likadant, men vi har ytterligare en faktor av $\frac{1}{n}$

$$V[\tfrac{1}{n}(X_1 + \ldots + X_n)] = V(\tfrac{1}{n}X_1 + \ldots + \tfrac{1}{n}X_n).$$

Eftersom $X_1, \ldots, X_n$ är oberoende (Se sats 2) leder det till följande

$$V(\tfrac{1}{n}X_1 + \ldots + \tfrac{1}{n}X_n) = V[\tfrac{1}{n}X_1] + \ldots + V[\tfrac{1}{n}X_n].$$

Nu kan vi använda oss av räknereglerna för varians (Se Stokastik sida 66). Vi får då

$$V[\tfrac{1}{n}X_1] + \ldots + V[\tfrac{1}{n}X_n] = \tfrac{1}{n^2}V(X_1) + \ldots + \tfrac{1}{n^2}V(X_n) = \tfrac{1}{n^2} \cdot nV[X_1] = \tfrac{1}{n}V[X_1].$$

Vi får då att $V(\bar{X}) = \frac{\sigma^2}{n}$ (ty $V(X_1) = \sigma^2$).

Sedan låter vi $\epsilon > 0$ vara fixt. Då kan vi använda oss av Chebyshevs olikhet för att bevisa satsen. Den är till för att ge oss en övre gräns för sannolikheten av slumpvariablerna förutsatt att vi känner till väntevärdet och standardavvikelsen för slumpvariablerna men inte hela deras fördelning. Följande sats finner vi i Stokastik-boken sida 68.

Sats 4 (Chebyshevs sats)**.** Låt X vara en slumpvariabel med ändligt väntevärde μ och standardavvikelse σ. Då gäller, för varje $a > 0$, att

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

(Beviset finner vi på sida 69 i Stokastik-boken). Observera att man i Stokastik-boken antar att slumpvariablerna är kontinuerliga, att de har en täthetsfunktion, för att utföra beviset.

Så vi låter $\epsilon > 0$ vara fixt och använder oss av Chebyshevs olikhet för att få följande

$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2}$.

Vi använder då oss av resultatet för variansen $V(\bar{X}) = \frac{\sigma^2}{n}$ för att få följande

$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$.

Här kan vi nu enkelt se att när $n \to \infty$ så går detta mot 0 vilket bevisar satsen.

$\square$

Så vad har vi för användning av stora talens lag? Den är användbar då vi gör många oberoende observationer av någon slumpvariabel. Då kommer observationernas medelvärde att ligga nära det sanna väntevärdet. Några exempel på sådana observationer kan vara blodtrycket hos patienter behandlade med en ny medicin eller livslängden hos personer i ett försäkringskollektiv. Dock så är oftast händelser i vardagen inte helt oberoende. Om vi tar exemplet med livslängden i ett försäkringskollektiv. Om livslängden ett år skulle sjunka drastiskt så kan det bero på en sjukdom som slår till i landet. Dessutom om en smittas så kommer sjukdomen troligtvis spridas i individens sociala grupp. Detta är ett tydligt exempel på att spridningen av sjukdommen i den gruppen inte längre är en oberoende händelse.

Så vad är viktigt att ta med sig av detta? Vi kanske inte kan ha helt oberoende observationer i vardagen, men det betyder inte att stora talens lag inte kan användas för att ge en bra approximation av verkligheten. Detta kallas för empirisk vetenskap (undersökningar av verkligheten) och stora talens lag är en av grundstenarna vi då använder oss av.

6 Centrala gränsvärdessatsen (CGS)

Vi har nu gått igenom stora talens lag, en av grundstenarna inom empirisk vetenskap. Nu ska vi gå vidare till vad de flesta skulle säga är sannolikhetslärans viktigaste sats, centrala gränsvärdessatsen.

Stora talens lag visade att medelvärdet $\bar{X}_n = (X_1 + ... + X_n)/n$ av en följd oberoende slumpvariabler, om alla slumpvariablerna har samma väntevärde μ och standardavvikelse σ, konvergerar i sannolikhet mot μ när $n \to \infty$.

Om vi antar att slumpvariablerna har samma väntevärde och standardavvikelse, samt att de också har samma fördelning, så kan vi studera fördelningen för $\bar{X}_n - \mu$ när n växer. Detta är centrala gränsvärdessatsen.

Vi kan dock se att $\bar{X}_n - \mu$ går mot 0 i sannolikhet, så vi måste skala upp $\bar{X}_n - \mu$. Så först måste vi standardisera $\bar{X}_n$. Vi har redan visat för stora talens lag att $E(\bar{X}_n) = \mu$. För att få ut den standardiserade slumpvariabeln behöver vi standardavvikelsen. Vi vet att standardavvikelsen är roten ur variansen, vilket vi fick ut från beviset av stora talens lag. Vi får då $D(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$. Detta ger oss den standardiserade slumpvariabeln $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$, som har väntevärde 0 och standardavvikelse 1 för alla n. Vi kommer alltså att behöva förstora avvikelserna från väntevärdet på $\sqrt{n}$-skala. Vi kommer senare se att denna standardiserade slumpvariabel kommer att konvergera mot en och samma fördelning, den standardiserade normalfördelningen, oavsett vilken fördelning de ursprungliga slumpvariablerna hade.

Med detta kan vi formulera satsen (Stokastik Sida 163).

Sats 5 (Centrala gränsvärdesatsen)**.** Låt $X_1, X_2, ...$ vara oberoende och likafördelade slumpvariabler med $E(X_i) = \mu$ och $D(X_i) = \sigma$, där $0 < \sigma < \infty$, och låt $\bar{X}_n := \sum_{i=1}^{n} \frac{X_i}{n}$. För godtyckliga $a < b$ gäller då att

$P(a < \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < b) \to \Phi(b) - \Phi(a)$, då $n \to \infty$,

där Φ är fördelningsfunktionen för $N(0, 1)$.

För att bevisa satsen behöver vi några begrepp som inte står med i begreppdelen, ty de är mer komplicerade och kan själva kräva bevis. Vi börjar med att förstå vad en momentgenerarande funktion (förkortad mgf) och moment av en slumpvariabel är.

Definition 26 (Moment av en slumpvariabel). (Stokastik Sid 149)
Låt X vara en slumpvariabel. Dess k:te moment (eller nollpunktsmoment, $k \geq 0$) definieras som $a_k := E(X^k)$. Det k:te centralmomentet definieras som $\mu_k := E((X-\mu)^k)$, där $\mu = E(X)$ är väntevärdet.

Momentgenererande funktioner

Definition 27 (Momentgenererande funktion (mgf)). (Stokastik Sid 150)
Låt X vara en slumpvariabel.
Dess momentgenererande funktion $\Psi_X(t)$ ($t \in \mathbf{R}$) definieras som

$$\Psi_X(t) := E(e^{tX}),$$

för sådana t där väntevärdet till höger är definierat.

Det kan vara svårt att se hur vi kan få nytta av den momentgenererande funktionen, så vi går igenom några exempel.

Exempel 7. Den momentgenererande funktionen för slumpvariabler har några kända egenskaper, bland annat att $\Psi(0) = 1$. Vi kan visa detta på följande vis

$$\Psi_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx.$$

Nu kan vi få fram väntevärdet då $t = 0$

$$\Psi_X(0) = E(e^{0\cdot x}) = E(1) = 1.$$

Vi fortsäter med momentet för den momentgenererande funktionens första derivata

$$\Psi_X'(t) = \frac{d}{dt}E(e^{tx}) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)dx.$$

Derivatan kan vi flytta in under integraltecknet

$$\frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{dt}e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} xe^{tx} f_X(x)dx.$$

Sedan sätter vi att $t = 0$

$$\Psi_X'(0) = \int_{-\infty}^{\infty} xe^{0\cdot x} f_X(x)dx = \int_{-\infty}^{\infty} x \cdot f_X(x)dx.$$

Detta är då en standardintegral för täthetsfunktionen. Detta ger oss alltså att $\Psi_X'(0) = E(X)$.

På liknande sätt kan vi beräkna andra derivatan. Vi kommer då få $\Psi''_X(0) = E(X^2)$. Med detta kan vi även formulera följande sats.

Sats 6 (Moment från mgf)**.** (Stokastik Sid 151)
Om $\Psi_X(t)$ är momentgenererande funktion till slumpvariabeln X och $\Psi_X(t)$ är definierad för alla $|t| < h$ för något $h > 0$, så gäller det att $\Psi_X^{(k)}(0) = E(X^k)$. (Observera att vi visade detta för $k = 1$ i exempel 6)

Vi kan göra några intressanta observationer om vi antar att slumpvariablerna i exempel 6 är standardiserat normalfördelade. Detta leder till att $\mu = 0$ samt att $\sigma = 1$, på grund av definitionen för den standardiserade normalfördelningen. (Anledningen till att vi valde just dessa som exempel är för att få en bättre bild över hur momentgenererande funktioner kan användas. Dessutom så kommer dessa resultat att hjälpa oss i beviset för centrala gränsvärdessatsen senare. Detta är kanske inte så tydligt nu men det kommer förhoppningsvis klarna om ett tag.)
I exempel 6 har vi beräknat några moment för slumpvariabler med hjälp av mgf. Vi kan nu fortsätta med att beräkna den momentgenererande funktionen för normalfördelade slumpvariabler.

Exempel 8. $\Psi_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$.

Nu behöver vi påminna oss om vad normalfördelningen hade för täthetsfunktion. Den är följande

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Så nu kan vi sätta in värdena för de standardiserade variablerna i täthetsfunktionen ($\mu = 0$, $\sigma = 1$) och fortsätta

$$\int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} e^{tx} e^{\frac{-x^2}{2}} \frac{dx}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{\frac{-x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx-\frac{x^2}{2}} dx.$$

Nu vill vi försöka kvadratkomplettera, så vi börjar med att försöka skriva om exponenten i integralen

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(\frac{x^2}{2}-tx)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx)} dx.$$

Här ser vi att detta kan vi kvadratkomplettera. Det gör vi på följande vis

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}((x-t)^2-t^2)} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2+\frac{t^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} \cdot e^{\frac{t^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} dx \cdot e^{\frac{t^2}{2}}.$$

I integralen ($\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} dx$) gör vi ett variabelbyte. Vi sätter $y = x - t$ och $dy = dx$. Gränserna ser vi kommer att fortsätta vara $-\infty$ samt ∞. Vi flyttar fram $e^{\frac{t^2}{2}}$ framför integralen och får då följande

$e^{\frac{t^2}{2}} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \frac{dy}{\sqrt{2x}}$.

Här ser vi att integranden är en typisk täthetsfunktion för standardiserad normalfördelade slumpvariabler vars värde är lika med 1. Så vi har alltså visat att $\Psi_X(t)$ är $e^{\frac{t^2}{2}}$.

Exempel 9. Vi betraktar en slumpvariabel X med
$P(X = 1) = \frac{1}{2}$,
och $P(X = -1) = \frac{1}{2}$.

Då får vi följande

$\Psi_X(t) = E(e^{tx})$
$= \frac{1}{2} \cdot e^t + \frac{1}{2} \cdot e^{-t} = \frac{1}{2}(e^t + e^{-t})$ $(= \cosh(t))$.

Sats 7 (Entydighetssatsen för mgf)**.** (Stokastik Sid 152)
Låt X och Y vara slumpvariabler. Om $\Psi_X(t) = \Psi_Y(t)$ för alla $|t| < h$, för något $h > 0$, så har X och Y samma fördelning.

(Entydighetssatsen är extra viktig att lägga på minnet.)

Sats 8 (Konvergenssatsen för mgf)**.** (Stokastik Sid 153)
Låt $X, X_1, X_2, X_3, ...$ vara slumpvariabler med momentgenererande funktioner $\Psi_X(t), \Psi_{X_1}(t), \Psi_{X_2}(t), ...,$ alla definierade för $|t| < h$, för något $h > 0$. Om $\Psi_{X_n}(t) \to \Psi_X(t)$ då $n \to \infty$ för $|t| < h$, så gäller det att $F_{X_n}(x) \to F_X(x)$, för alla x, då $n \to \infty$.

Nu har vi allt vi behöver för att bevisa centrala gränsvärdessatsen.

Bevis. Vi kommer att visa att den momentgenererande funktionen (mgf) för $Y_n := \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$ konvergerar mot den mgf för en standardiserad normalfördelning när $n \to \infty$. Dess mgf är $\Psi(t) = e^{t^2/2}$, vilket vi redan visat i del 5.1. Vi använder oss dessutom av konvergenssatsen för mgf, av denna följer då att fördelningen för Y_n konvergerar mot den standardiserade normalfördelningen.

Om vi standardiserar de ursprungliga variablerna $X_1, X_2, ...$ blir dessa $Z_i := (X_i - \mu)/\sigma$ (enligt definitionen för standardiserad slumpvariabel). Då gäller

$Y_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) = \frac{\sqrt{n}}{\sigma}\sum_{i=1}^{n}(\frac{X_i}{n} - \frac{\mu}{n})$.

Vi använder oss av medelvärdet i satsen för centrala gränsvärdessatsen för att få fram höger likhet ovan. Vi fortsätter med att få in allt under summatecknet

$\frac{\sqrt{n}}{\sigma}\sum_{i=1}^{n}(\frac{X_i}{n} - \frac{\mu}{n}) = \sum_{i=1}^{n}(\frac{X_i-\mu}{n} \cdot \frac{\sqrt{n}}{\sigma}) = \sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma\sqrt{n}})$.

Nu kan vi använda oss av Z_i som vi definierade tidigare. Då får vi

$\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma\sqrt{n}}) = \sum_{i=1}^{n}(\frac{Z_i}{\sqrt{n}})$.

Nu ska vi visa att mgf för $Y_n = \sum_{i=1}^{n}(\frac{Z_i}{\sqrt{n}})$ konvergerar mot mgf för en standardiserad normalfördelning $e^{t^2/2}$. Men eftersom de ursprungliga variablerna $X_1, X_2, ...$ är oberoende så är även de standardiserade variablerna $Z_1, Z_2, ...$ oberoende. Då gäller följande

$\Psi_{Y_n}(t) = E(e^{\sum_{i=1}^{n}(\frac{tZ_i}{\sqrt{n}})})$,

enligt definitionen för momentgenererande funktioner. Vi fortsätter

$E(e^{\sum_{i=1}^{n}(\frac{tZ_i}{\sqrt{n}})}) = E(e^{(tZ_1/\sqrt{n})+(tZ_2/\sqrt{n})+...+(tZ_n/\sqrt{n})})$.

Vi utnyttjar att variablerna är oberoende för att skriva om ovanstående som en produkt av väntevärden. Väntevärdet av en produkt är produkten av väntevärdena när vi har oberoende variabler. Dessutom använder vi följande räkneregel för exponentialfunktioner

$a^x a^y = a^{x+y}$.

Vi kan då fortsätta med beviset

$E(e^{(tZ_1/\sqrt{n})+(tZ_2/\sqrt{n})+\cdots+(tZ_n/\sqrt{n})}) = \prod_{i=1}^{n} E(e^{tZ_i/\sqrt{n}})$.

Nu kan vi återigen använda den momentgenererande funktionen för Z för att få följande

$\prod_{i=1}^{n} E(e^{tZ_i/\sqrt{n}}) = (\Psi_Z(t/\sqrt{n}))^n$.

Anledningen till att vi får upphöjt med n är på grund av att vi hade en produkt i vänsterled.

Vi tar en titt på den momentgenererande funktionen $\Psi_Z(t/\sqrt{n})$. För ett fixt t har vi $\frac{t}{\sqrt{n}}$ som går mot noll då $n \to \infty$. Detta leder till att $\Psi_Z(0) = 1$.

Detta antyder att vi har följande

$(\Psi_Z(t/\sqrt{n}))^n = (1 + ...)^n.$

Detta påminner väldigt starkt om en Taylorutveckling, så vi Taylorutvecklar den momentgenererande funktionen $\Psi_Z(t/\sqrt{n})$ i $t = 0$. Vi börjar med att skriva upp denna

$\Psi_Z(s) = \Psi_Z(0) + s\Psi_Z'(0) + \frac{s^2}{2}\Psi''(0) + O(s^3)$ då $s \to 0$,

där $O(s^k)$ betecknar en funktion som dividerad med s^k är begränsad då $s \to 0$. Frågan är då hur många termer som måste tas med. Detta undersöker vi på följande vis. Av definition för momentgenererande funktioner följer att $\Psi(0) = 1$, och från satsen om moment för mgf vet vi att $\Psi_Z^{(k)}(0) = E(Z^k)$. Eftersom Z är standardiserad gäller det också att $E(Z) = \Psi'(0) = 0$ och $V(Z) = E(Z^2) = \Psi''(0) = 1$. Detta kan vi använda oss av i Taylorutvecklingen och får då att $\Psi_Z(s) = 1 + s^2/2 + O(s^3)$. Om vi väljer $s = t/\sqrt{n}$, för givet t, kan vi studera funktionens första derivata

$\Psi_Z(0) + \Psi_Z'(0) \cdot \frac{t}{\sqrt{n}} + ...$

Här ser vi då att vi får följande

$1 + 0 \cdot \frac{t}{\sqrt{n}} + ... = 1 + 0 + ...$

Så vi har inte fått ett till bidrag ännu. Vi fortsätter med andraderivatan

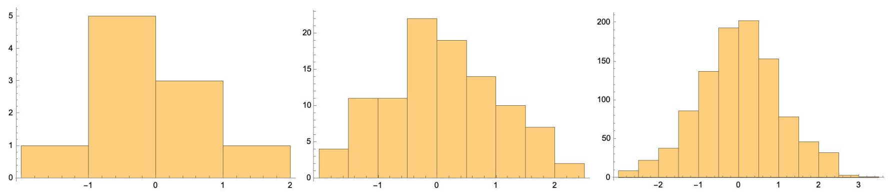
$1 + 0 + \Psi_Z''(0) \cdot \frac{t^2}{2n} + ... = 1 + \frac{t^2}{2n} + ...$

Om vi skriver ut resttermen så ser vi att detta är exakt samma Taylorutveckling som för $\Psi_Z(s)$ ovan. Vi får då följande

$\Psi_{Y_n}(t) = (\Psi_Z(t/\sqrt{n}))^n = (1 + \frac{t^2}{2n} + O(\frac{t^3}{n^{3/2}}))^n.$

Om vi låter $n \to \infty$ får vi därför att $\Psi_{Y_n}(t) \to e^{t^2/2}$ (som vi kan se i del 5.1) vilket är momentgenererande funktionen för den standardiserade normalfördelningen. Enligt entydighetssatsen för mgf så gäller det att om två momentgenererande funktioner är lika med varandra så har de motsvarande slumpvariablerna samma fördelning.

$\square$



Bilderna ovan går från $n = 10$, $n = 100$ samt $n = 1000$.

Centrala gränsvärdessatsen visar hur fördelningen kommer se ut när antalet observationer n går mot oändligheten.

När n ökar så ser man i bilden ovan att sannolikhetsfunktionen alltmer liknar den standardiserade normalfördelningstätheten.

Så hur är centrala gränsvärdessatsen så användbar?
Som vi sa för stora talens lag så är händelser i vardagen sällan helt oberoende. Så hur kan vi försöka studera dessa slumpvariabler i vardagen? Här är centrala gränsvärdessatsen väldigt användbar och viktig.
Den viktigaste orsaken för detta är att många slumpvariabler som dyker upp i verkligheten i sin tur påverkas av ett flertal andra slumpvariabler. Till exempel en persons längd. Det kan bero på individens gener, näringsintag under uppväxt med mera. Centrala gränsvärdessatsen säger då att om inflytandet från de ingående slumpvariablerna är någorlunda linjärt så blir den studerade slumpvariabeln normalfördelad.

7 Stora avvikelser

Nu har vi kommit till stora avvikelser. Vi börjar med att sammanfatta de två väsentliga satserna vi behöver för att härleda beviset och satsen för stora avvikelser.

Låt $X_1, X_2, X_3, ...$ vara oberoende likafördelade slumpvariabler med $E(X_1) = 0, E(X_1^2) = 1$. Låt dessa bilda $S_N = \sum_{k=1}^{N} X_k$. Då säger stora talens lag (STL) följande

$STL : \lim_{n\to\infty} \frac{1}{n} \sum_1^N X_k = \lim_{N\to\infty} \frac{1}{N} S_N = 0$,
där värdet är lika med $E(X_1)$.

Vi kan även skriva upp centrala gränsvärdessatsen (CGS)

$CGS : P(S_N \geq b\sqrt{n}) = \lim_{N\to\infty} P(\frac{S_N}{\sqrt{N}} \geq b) = P(Y \geq b)$,
Y får vi när vi normaliserat, $N(0, 1)$.

Då vill vi undersöka följande

$$P(S_N \geq aN) \ (= P(\frac{S_N}{N} \geq a)), \tag{1}$$

där $a \geq 0$. Detta betyder då att a mäter sannolikheten för att vi ligger "långt ifrån" väntevärdet.
Här kan vi se att om N går mot oändligheten så kommer sannolikheten att gå mot noll. Men vi vet än så länge inte hur snabbt detta sker, vilket är det vi vill undersöka.

Innan vi fortsätter är det lämpligt att införa satser som kommer att hjälpa oss med undersökningen.

Sats 9 (Markovs olikhet). (Stokastik Sid 68)
Låt X vara en icke-negativ slumpvariabel ($X \geq 0$) med ändligt väntevärde, $E(X)$. Då gäller, för varje $a > 0$, att

$P(X \geq a) \leq \frac{E(X)}{a}$.

Vi kan skriva om uttrycket i vänsterledet på följande vis

$P(S_N \geq aN) = P(e^{\lambda S_N} \geq e^{\lambda aN})$.

Vi kan göra denna omskrivning på grund av följande resultat

$x \mapsto e^{\lambda x}$ är växande, om $\lambda > 0$.

Vi kan sedan använda Markovs olikhet för att få

$P(e^{\lambda S_N} \geq e^{\lambda aN}) \leq e^{\frac{1}{\lambda aN}} E(e^{\lambda S_N})$.

Vi fortsätter och får följande

$e^{\frac{1}{\lambda aN}} E(e^{\lambda S_N}) = e^{\frac{1}{\lambda aN}} E(e^{\lambda(X_1+...+X_N)})$.

Sedan kan vi utnyttja att slumpvariablerna är oberoende av varandra

$e^{\frac{1}{\lambda aN}} E(e^{\lambda(X_1+...+X_N)}) = e^{\frac{1}{\lambda aN}} E(e^{\lambda X_1} \cdot e^{\lambda X_2} \cdot ... e^{\lambda X_N}) = e^{\frac{1}{\lambda aN}} (E(e^{\lambda X_1})...E(e^{\lambda X_N}))$.

Här kan vi observera att vi redan sett $E(e^{\lambda X_1})$ tidigare, detta är definitionen för en momentgenererande funktion. Då kan vi skriva om och får

$e^{\frac{1}{\lambda aN}} (E(e^{\lambda X_1})...E(e^{\lambda X_N})) = e^{\frac{1}{\lambda aN}} (\Psi_{X_1}(\lambda))^N$.

Nu vill vi försöka få den momentgenererande funktionen på exponentialform. Det kan vi få genom att använda logaritmen. Vi skriver dessutom om potensen för den vänstra funktionen i högerledet och får följande

$$e^{\frac{1}{\lambda a N}}(\Psi_{X_1}(\lambda))^N = e^{-\lambda a N} e^{N \text{ log } \Psi_{X_1}(\lambda)}.$$

Sedan kan vi använda oss av potensreglerna och få följande

$$e^{-\lambda a N} e^{N \text{ log } \Psi_{X_1}(\lambda)} = e^{-N(\lambda a - \text{log } \Psi_{X_1}(\lambda))}.$$

Tillsammans fås

Sats 10. Låt $X_1, X_2, X_3, ...$ vara en följd av oberoende slumpvariabler alla med väntevärdet $\mu = E[X_k] = 0$ och momentgenererande funktion $\Psi(\lambda) < \infty$ för något $\lambda > 0$. Då kommer sannolikheten för S_N bete sig på följande vis

$$P(S_N \geq aN) \leq e^{-N(\lambda a - \text{log } \Psi_{X_1}(\lambda))}.$$

Vi kan observera att om $\lambda a - \text{log } \Psi_{X_1}$ är större än 0 så kommer högerledet att gå mot 0, när $N \to \infty$.
Innan vi fortsätter så gör vi en omskrivning av sammanfattningen ovan

$$\begin{aligned} &\tfrac{1}{N}\text{log } P(S_N \geq aN) \leq \tfrac{1}{N}\text{log } e^{-N(\lambda a - \text{log } \Psi_{X_1}(\lambda))} \\ &= \tfrac{1}{N} \cdot (-N(\lambda a - \text{log } \Psi_{X_1}(\lambda))) = -(\lambda a - \text{log } \Psi_{X_1}(\lambda)). \end{aligned}$$

Nu vill vi försöka välja parametrar λ sådana att uttrycket i e i högerledet blir större än noll.
Det finns inte endast ett sätt att göra detta men vi behöver bara undersöka för ett sådant, när vi maximerar $\lambda a - \log \Psi_{X_1}$. Varför väljer vi just detta? Det blir ganska tydligt när vi ser på en verklig situation, försäkringsbolag. Om $P(S_N \geq aN)$ är kontrollerad innebär detta att vi uppskattar sannolikheten att det absolut värsta för försäkringsbolaget sker. Detta är precis den vinkel vi är intresserade av, sannolikheten för att de behöver betala tillbaka många försäkringar. När man har fått fram sannolikheten så kan man välja hur man vill driva företaget. Till exempel så vore det säkrast att ha så mycket pengar redo att betala om alla skulle kräva försäkringar. Men med detta tillvägagångsätt är det näst intill omöjligt att gå med vinst. Den andra sidan av spektrummet är att riskera allt och gå i konkurs så fort utbetalning krävs. Så helst vill man ligga någonstans mitt emellan för att gå med vinst. Detta är endast ett försök att skapa ett enkelt exempel för att lättare föreställa sig varför denna teori är intressant i vardagen.

Vad innebär det att maximera $\lambda a - \log \Psi_{X_1}$ rent matematiskt? Det innebär att vi försöker hitta det snabbaste sättet att gå mot noll. Vi har visat, i flera steg, att $P(S_N \geq aN) \leq \; ... \; \leq e^{-N(\lambda a - \text{log } \Psi_{X_1}(\lambda))}$. Här ser vi att högerledet är litet när $(\lambda a - \log\, \Psi_{X_1}(\lambda))$ är stort och positivt. Värt att notera är att när vi maximerar funktionen på detta vis så har vi gjort en serie överskattningar på vägen. Vi har nu maximerat $P(S_N \geq aN)$ ovanifrån. Vi skulle också kunna maximera denna sannolikhet nedifrån för att se hur stora överskattningarna är. Detta kommer vi dock inte att göra, ty det är för tekniskt för denna text.

Eftersom vi vill beräkna en sannolikhet så vet vi att $P(S_N \geq aN) \leq 1$. Om vi får ett tal som är större än 1 så har vi gjort något fel på vägen, enligt definitionen av sannolikhet.
Innan vi går vidare kan vi kort nämna några fall vi inte kommer att undersöka och varför. Vad skulle hända om $\lambda a - \log\, \Psi_{X_1}(\lambda)$ var negativt? Det skulle betyda att $e^{-N(\lambda a - \log\, \Psi_{X_1}(\lambda))} > 1$ vilket inte säger oss någonting om dess sannolikhet. Om $\lambda a - \log\, \Psi_{X_1}(\lambda) = 0$ istället så får vi att $e^{-N(\lambda a - \log\, \Psi_{X_1}(\lambda))} \leq 1$ vilket inte heller är intressant. Med detta går vi vidare.

För att maximera $\lambda a - \log \Psi_{X_1}$ inför vi följande

$\Phi(\lambda) = a\lambda - \log\, \Psi_{X_1}(\lambda)$.

För att maximera det ovanstående vill vi lösa följande

$\Phi'(\lambda) = 0$.

Vi börjar med att derivera $\Phi(\lambda)$, vilket blir

$\Phi'(\lambda) = a - \frac{1}{\Psi_{X_1}(\lambda)} \cdot \Psi'_{X_1}(\lambda) = a - \frac{\Psi'_{X_1}(\lambda)}{\Psi_{X_1}(\lambda)}$.

Då vill vi lösa följande

$0 = a - \frac{\Psi'_{X_1}(\lambda)}{\Psi_{X_1}(\lambda)}$.

Detta ger oss värde(n) på $\lambda : \lambda_0$.
Vi sätter sedan in λ_0 i satsen för att maximera funktionen.

Nu kan vi gå igenom några exempel.

Exempel 10. Låt X vara normalfördelad, då har vi följande mgf

$\Psi_{X_1}(\lambda) = e^{\frac{\lambda^2}{2}}$.

Vi har även följande

$\Psi'_{X_1}(\lambda) = \lambda e^{\frac{\lambda^2}{2}}$,
$\log \Psi_{X_1}(\lambda) = \frac{\lambda^2}{2}$.

Då kan vi beräkna a

$0 = a - \frac{\Psi'_{X_1}(\lambda)}{\Psi_{X_1}(\lambda)}$,

$0 = a - \frac{\lambda e^{\frac{\lambda^2}{2}}}{e^{\frac{\lambda^2}{2}}}$.

Här vet vi att $e^t > 0$ om $t > 0$, samt att $e^{\frac{\lambda^2}{2}}$ alltid är positivt, så de bägge $e^{\frac{\lambda^2}{2}}$ vi har i täljaren och nämnaren tar ut varandra och vi får följande

$0 = a - \lambda \rightarrow \lambda = a.$

Detta ger oss följande

$e^{-N(a^2 - \frac{a^2}{2})} = e^{-N\frac{a^2}{2}}.$

Detta betyder att $P(S_N \geq aN)$ avtar exponentiellt i N med kvadratiskt beroende av a.

Exempel 11. Betrakta slumpvariabeln X med sannolikheter $P(X_1 = 1) = P(X_1 = -1) = 1/2$ där $S_n = X_1 + ... + X_n$. Vi har också att $E(S_n) = 0$ i detta fall, ty väntevärdet är linjärt.
Vi börjar då med att skriva ut den momentgenererande funktionen

$\Psi_{X_1}(\lambda) = \frac{e^{\lambda} + e^{-\lambda}}{2} = \cosh \lambda,$

Vi vill fortfarande maximera $\Phi(\lambda) = \lambda a - \log \Psi_{X_1}$, vilket vi gör genom att lösa $\Phi'(\lambda) = 0$. Vi får då följande

$\log \Psi_{X_1}(\lambda) = \log(\cosh\ \lambda).$

Dess derivata blir då

$\Psi'_{X_1}(\lambda) = \frac{\sinh\ \lambda}{\cosh\ \lambda}$

och från $\Phi(\lambda)$ som vi nämnde tidigare får vi villkoret $\Psi'_{X_1}(\lambda) = a$. Vi kan lösa ut a genom att införa $t = e^{\lambda} \geq 0$. Detta ger följande

$\frac{sinh\ \lambda}{cosh\ \lambda} = a,$

$\frac{\frac{e^{\lambda} - e^{-\lambda}}{2}}{\frac{e^{\lambda} + e^{-\lambda}}{2}} = \frac{t - \frac{1}{t}}{t + \frac{1}{t}} = \frac{t^2 - 1}{t^2 + 1} = a.$

Vi får då

$\frac{t^2-1}{t^2+1} = a \to t^2-1 = t^2a+a \to t^2-t^2a = 1+a \to t^2(1-a) = 1+a \to t^2 = \frac{1+a}{1-a}$.

Eller för att skriva det lite tydligare och kortare

$t_a = e^{\lambda_a} = \sqrt{\frac{1+a}{1-a}}$.

Vi är endast intresserade av $0 < a < 1$ eftersom $\mathbf{P}(X > 1) = 0$. Dessutom noterar vi att

$e^{-\lambda_a} = \sqrt{\frac{1-a}{1+a}}$,

samt följande

$\lambda_a = \log[\frac{(1+a)^{\frac{1}{2}}}{(1-a)^{\frac{1}{2}}}]$,

$\lambda_a = \frac{1}{2}\log(1+a) - \frac{1}{2}\log(1-a)$.

Nu kan vi skriva ut $cosh\lambda_a$ och får följande

$\frac{1}{2}(e^{\lambda_a}+e^{-\lambda_a}) = \frac{1}{2}\sqrt{\frac{1+a}{1-a}}+\frac{1}{2}\sqrt{\frac{1-a}{1+a}} = \frac{1}{2}\frac{\sqrt{1+a}\cdot\sqrt{1+a}+\sqrt{1-a}\cdot\sqrt{1-a}}{\sqrt{1-a}\cdot\sqrt{1+a}} = \frac{1}{2}\frac{(1+a)+(1-a)}{\sqrt{1-a}\cdot\sqrt{1+a}}$

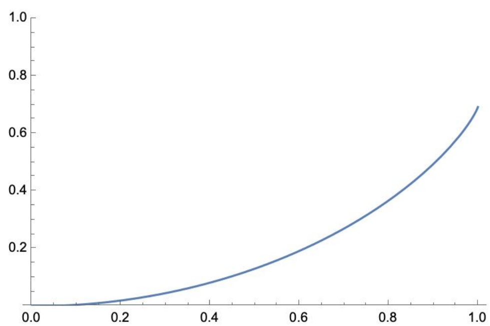
$= \frac{1}{2}\frac{2}{\sqrt{(1-a)\cdot(1+a)}} = \frac{1}{\sqrt{(1-a)\cdot(1+a)}}$.

Alltså vi får följande

$cosh\ \lambda_a = \frac{1}{\sqrt{(1+a)(1-a)}}$,

och om vi lägger ihop allt får vi följande Cramérfunktion

$\Phi(a) = \frac{1}{2}[(1+a)\log(1+a) - (a-1)\log(1-a)]$.



8 Källor

References

[1] Sven-Erik Alm och Tom Britton, Stokastik, 1:ta upplagan, Liber AB, Stockholm, 2008.

[2] Arne Persson och Lars-Christer Böiers, Analys i en variabel, 3:e upplagan, Studentlitteratur AB, Lund, 2010.

[3] Alan Sola, Advanced Probability, Lecture Notes, Department of Pure Mathematical Statistics, University of Cambridge, 2014

[4] ludo.co, Taylor- och Maclaurinutveckling, läst 16-12-2022.