TAMS15: SS4
CGS i repris

Johan Thim (johan.thim@liu.se)

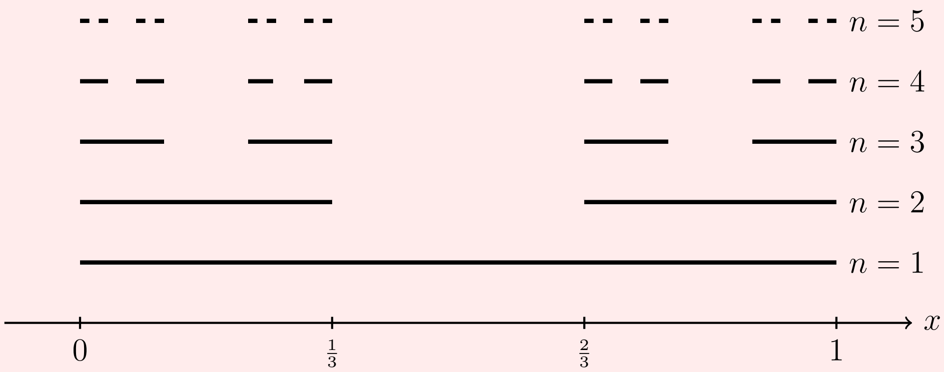
23 november 2018

14.1 Stokastiska variabler: diskreta, kontinuerliga, singulära?

Vi har delat upp klassen av stokastiska variabler i två delar. En del vi kallat diskret, där vi använt oss av en sannolikhetsfunktion $P(X = k)$, och en kontinuerlig där vi *krävt* att det finns en täthetsfunktion. Detta krav är emellertid inte alltid uppfyllt. Det finns icke-diskreta stokastiska variabler som inte har någon täthetsfunktion. Variabler av denna typ brukar sägas vara singulära. Ibland så kräver man till och med definitionsmässigt att en stokastisk variabel ska ha en täthetsfunktion för att kallas kontinuerlig. Så vad gör man åt situationen att det verkar finnas andra djur i djungeln? Dessa blir tyvärr svåra att hantera med den teori vi har tillgång till just nu, men låt oss titta på ett välkänt exempel som åtminstone visar på troligheten att singulära fördelningar finns.

Cantormängden

Låt oss börja med något kul, nämligen en tämligen sönderstyckad delmängd av intervallet $[0, 1]$. Processen fungerar enligt följande. Vi delar $C_1 = [0, 1]$ i tre delar och tar bort den öppna mittersta delen. Vi kallar de punkter som är kvar (dvs $[0, 1/3] \cup [2/3, 1]$) för C_2. Vi delar nu de två intervallen i C_2 i tre likadana delar vardera och tar bort mittensegmenten. Kalla den nya mängden C_3. Sen upprepar vi processen gång på gång och kallar mängden som är kvar vid steg k för C_k.



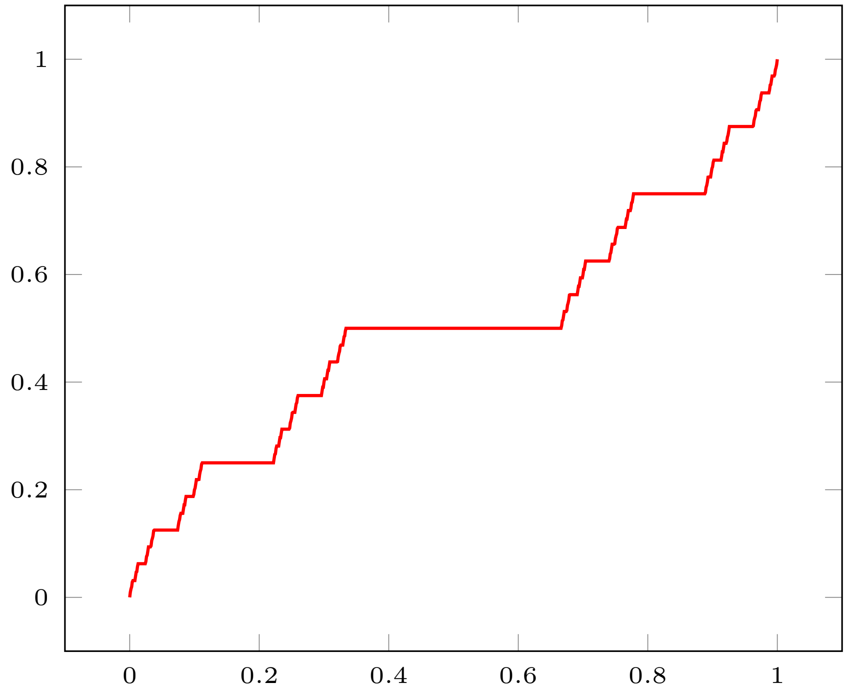
Vi noterar att $C_1 \supset C_2 \supset C_3 \supset \cdots$ och definierar $C = \cap_{k=1}^{\infty} C_k$. Detta är **Cantormängden**.

Så vad ska vi med denna mängd till? Givetvis att konstruera något intressant motexempel. Man kan se att om vi representerar ett tal $x \in [0, 1]$ i basen 3, dvs vi skriver $x = \displaystyle\sum_{k=1}^{\infty} x_k 3^{-k}$

där varje $x_k = 0, 1$ eller 2, så kommer $x_k \neq 1$ för alla k om $x \in C$. Vi tar hela tiden bort mittendelen, vilket gör att vi aldrig får motsvarande siffra i expansionen i basen 3. Den så kallade Cantorfördelningen erhåller vi nu om vi definierar fördelningsfunktionen enligt följande:

$$F(x) = \frac{1}{2} \sum_{k=1}^{\infty} \frac{x_k}{2^k}, \quad x \in C.$$

För $x < 0$ definierar vi $F(x) = 0$ och för $x > 1$ definierar vi $F(x) = 1$. Vi täpper till hålen vi lämnat i $[0, 1]$ genom att observera att $F(x_1) \leq F(x_2)$ för $x_1 \leq x_2$, så F är växande. Vi låter $F(x) = \sup\{F(y) : y \in C,\ y < x\}$ för $x \in [0, 1] \setminus C$, vilket definierar $F(x)$ för alla $x \in \mathbf{R}$ så att F är kontinuerlig. Observera att denna definition innebär att $F(x)$ är konstant på de linjesegment som vi tog bort när vi konstruerade C. Vi ser också att $F(x) \to 1$ då $x \to 1$ och $F(x) \to 0$ då $x \to 0$. Denna funktion brukar gå under namnet *the Devil's staircase* – Djävulens trappa – just för att den har ett par roliga egenskaper som går lite mot intuitionen.



Detta är alltså en fördelningsfunktion. Så vad var poängen med detta? Jo, poängen är den att en stokastisk variabel X med fördelningsfunktionen F *inte* är diskret men *saknar* täthetsfunktion! Yikes. Att X inte är diskret följer av att F är kontinuerlig (det blir alltid hopp i F om variabeln är diskret). Så varför finns ingen täthetsfunktion? Det följer av att $F(x)$ är konstant nästan överallt. När vi konstruerade C så tog vi hela tiden bort tredjedelar av linjesegment, så om vi tittar på figuren ovan där vi konstruerade Cantormängden så ser vi att vi tar bort längderna

$$\frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \cdots = \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1.$$

Huh? Då är ju inget kvar? Nja, inte direkt. Dels är argumentet ovan lite handviftande och dels så kan vi fortfarande har kvar en synnerligen sönderstyckad mängd där varje punkt är hyfsat isolerad så att det inte blir några intervall. Faktum är att Cantormängden faktiskt inte är uppräkningsbar (så betydligt större än till exempel $[0, 1] \cap \mathbf{Q}$). Men viktigast för oss, derivatan av fördelningsfunktionen måste vara noll nästan överallt eftersom funktionen är konstant där. Därför kan vi inte ha någon täthetsfunktion.

Så vad var poängen med allt det där? En poäng är att vi egentligen borde sträva efter att formulera saker utan att explicit välja summa eller integral av täthetsfunktion. Kan vi formulera och bevisa resultat där vi endast använder sannolikhetsmåttet (och fördelningsfunktionen) så får vi resultat som gäller generellt.

14.2 Momentgenererande funktioner

Låt oss skifta riktning en aning. Vi börjar med en definition.

Moment

Definition. För $k = 1, 2, 3, \ldots$ så definieras det k:te **momentet** av X enligt $E(X^k)$ om detta väntevärde är ändligt. De **centrerade momenten** definieras enligt $E((X - E(X))^k)$.

Vi känner igen moment av ordning 1 och 2 som väntevärdet och något som vi brukar använda för att räkna ut variansen. Väntevärdet beskriver var fördelningen har sitt masscentrum och variansen hur utspridd den är. Moment av ordning 3 och högre beskriver skevhet hos fördelningen. En intressant fråga är om man kan återskapa en stokastisk variabel om man känner *alla* moment? Svaret är tyvärr nej, om man inte vet lite mer information. Dessutom kan det vara så att moment av vissa ordningar inte existerar för vissa stokastiska variabler. Vi kan till exempel titta på Cauchy-fördelningen.

Cauchyfördelningen

Låt $f_X(x) = \dfrac{1}{\pi(1+x^2)}$, $x \in \mathbf{R}$, vara täthetsfunktionen för en stokastisk variabel X. Vi säger att X är Cauchyfördelad. Visa att alla moment för X är oändliga.

Lösning. Vi ser att

$$\int_0^\infty \frac{x^k}{\pi(1+x^2)}\,dx$$

är generaliserad i oändligheten och att om vi betraktar integranden för stora x så ser det ut som $x^k/x^2 = x^{k-2}$, så integralen kommer vara divergent för alla $k \geq 1$. Om k är jämn så är integranden positiv, så integralen är divergent även när vi integrerar över hela $\mathbf{R}$. Om k är udda växlar visserligen integranden tecken i nollan, men då integralen från 0 till oändligheten är divergent hjälper inte teckenväxlingen (om vi inte blandar in principalvärden). Alltså kan inte momenten existera. Inte ens väntevärdet blir ändligt!

Momentgenererande funktion

Definition. Den **momentgenererande funktionen** $M_X(t)$ för en stokastisk variabel definieras som $M_X(t) = E(e^{tX})$ för de t där uttrycket existerar.

I det diskreta fallet så reduceras definitionen till

$$M_X(t) = \sum_k e^{tk} p_X(k)$$

och i det kontinuerliga (med täthetsfunktion) till

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

För singulära fördelningar blir det lite konstigare, så vi hoppar över det. I det diskreta fallet så pratar man ibland om den *sannolikhetsgenererande funktionen*

$$G_X(s) = E(s^X) = \sum_k s^k p_X(k).$$

Då gäller alltså att $M_X(t) = G_X(e^t)$, så dessa funktioner är nära besläktade. Dessutom ser vi direkt att $G_X(s)$ alltid konvergerar åtminstone för $|s| < 1$ (då koefficienterna $p_X(k)$ är begränsade). Den observante läsaren med lite transformteori i bagaget känner säkert igen detta som z-transformen av p_X. Snarlikt så kan vi även se att i det kontinuerliga fallet (åtminstone när det finns en täthetsfunktion) så gäller att $M_X(t) = \mathcal{L}(f_X)(-t)$ är Laplace-transformen av tätheten med $-t$ som argument. Vi kan alltså använda teori från z- och Laplacetransformerna för att härleda egenskaper för momentgenererande funktioner.

Exempel

Låt $X \sim N(0,1)$ och $Y \sim N(\mu, \sigma)$. Beräkna de momentgenererande funktionerna.

Lösning. Vi ser att

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

där vi bytte variabel $u = x - t$ i integralen och utnyttjade att vi arbetar med en täthetsfunktion. För att komma åt Y gör vi följande omskrivning. Låt $Y = \mu + \sigma X$, där $X \sim N(0,1)$. Då blir

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E(e^{t\sigma X} e^{\mu t}) = e^{\mu t} E(e^{t\sigma X}) = e^{\mu t} M_X(t\sigma) \\ &= e^{\mu t + t^2 \sigma^2 / 2}. \end{aligned}$$

Så varför kallar vi detta objekt för momentgenererande funktion? Vad har den med momenten att göra? Svaret kommer ganska direkt om vi tänker på hur Maclaurinutvecklingen av e^{tx} ser ut:

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2} + \frac{(tx)^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}.$$

Formellt skulle då alltså

$$E(e^{tX}) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k),$$

under förutsättningen att $E(X^k)$ existerar för $k = 0, 1, 2, \ldots$ och att vi kan byta plats på serien och väntevärdesoperatorn. Genom att derivera potensserien k gånger och sätta $t = 0$, så erhåller vi m_k:

$$m_k = E(X^k) = M_X^{(k)}(0).$$

Momenten finns alltså inbakade i den momentgenererande funktionen (under förutsättningen att den existerar i en omgivning av noll). Så om då $M_X(t)$ existerar i en omgivning av noll, innebär det att denna momentgenererande funktion definierar en entydigt bestämd fördelning? Svaret är ja.

Sats. Om $M_X(t)$ existerar ändligt i en omgivning av noll så finns en entydigt bestämd fördelning med den momentgenererande funktionen $M_X(t)$. Vidare gäller att alla moment $E(X^k)$ existerar och att $M_X(t) = \displaystyle\sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$ i denna omgivning av noll.

För att bevisa detta brukar man använda satsen om inversion av Laplacetransformen, men det faller tyvärr lite utanför denna kurs. En direkt följd av satsen är följande korollarium.

Entydighet

Sats. Om $M_X(t) = M_Y(t)$ för t i en omgivning av origo, så har X och Y samma fördelning.

Notera att detta *inte* innebär att två stokastiska variabler har samma fördelning om alla deras moment sammanfaller. Detta på grund av att serien $\displaystyle\sum_{k=1}^{\infty} \frac{m_k}{k!} t^k$ kan vara divergent för $t \neq 0$. Det finns alltså en hel uppsjö av problem som kan uppstå. Exempelvis Cauchy-fördelningen saknar momentgenererande funktion för $t \neq 0$ och en exponentialfördelning har endast momentgenererande funktion $\lambda(\lambda - t)^{-1}$ för $t < \lambda$ (annars oändlig). Exempel finns även där alla moment existerar men den momentgenererande funktionen existerar bara för $t = 0$ (till exempel en variant baserad på lognormal-fördelningen), vilket visar att momentens existens inte är tillräckligt.

14.3 Karakteristisk funktion

För att komma runt problemet med att momentgenererande funktionen inte alltid existerar så använder man sig ibland av den *karakteristiska funktionen*. För en stokastisk variabel X definieras denna som $\phi_X(t) = E(e^{itX})$. I det kontinuerliga fallet med täthetsfunktion så är detta Fouriertransformen av täthetsfunktionen. En trevlig egenskap med den karakteristiska funktionen är att den alltid existerar. Nackdelen är att den är komplexvärd och kräver en del ordentlig analys för att få ordning på. I fallet med täthetsfunktion så följer det ganska omgående att ϕ_X är likformigt kontinuerlig till och med (via egenskaper för Fouriertransformen). I fallet då den momentgenererande funktionen också existerar så gäller sambandet $M_X(it) = \phi_X(t)$, så egenskaper för momentgenererande funktioner generaliserar oftast ganska direkt.

14.4 Summor av stokastiska variabler

Nu ska vi se på en av de stora styrkorna för momentgenererande funktioner (eller karakteristiska för den delen).

Momentgenererande funktion för summa

Sats. Låt $X_1, X_2, \dots, X_n$ vara oberoende stokastiska variabler med momentgenererande funktioner $M_{X_1}, \dots, M_{X_n}$. Om $Y = X_1 + X_2 + \cdots + X_n$ så är

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

Bevis. Eftersom variablerna är oberoende så kommer

$$\begin{aligned} E(e^{tY}) &= E(e^{t(X_1+X_2+\cdots+X_n)}) = E(e^{tX_1}e^{tX_2}\cdots e^{tX_n}) \\ &= E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n}) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t). \end{aligned}$$

Vad detta innebär rent praktiskt är att vi kan hitta fördelningen för en summa av stokastiska variabler genom att multiplicera ihop deras momentgenererande funktioner och sedan identifiera vilken fördelningen denna produkt hör ihop med (där vi utnyttjar entydigheten). Tänk tillbaka på faltningssatsen. Eftersom momentgenererande funktioner i princip är Laplacetransformer så är det så att Laplacetransformen av faltningen av två funktioner ges av produkten av dess Laplacetransformer (motsvarande gäller Fouriertransformen och karakteristiska funktioner). Motsvarande samband gäller för karakteristiska funktioner.

14.5 Centrala gränsvärdessatsen

Så anledningen till att vi tog upp allt detta om momentgenererande funktioner var för att – åtminstone i ett specialfall där den momentgenererande funktionen existerar nära noll – kunna bevisa centrala gränsvärdessatsen. Det är nämligen så trevligt att om $U_1, U_2, \dots$ är en följd stokastiska variabler med ändliga momentgenererande funktioner $M_{U_1}, M_{U_2}, \dots$ sådana att

$$\lim_{n\to\infty} M_{U_n}(t) = e^{t^2/2}, \quad t \in \mathbf{R},$$

så gäller att

$$\lim_{n\to\infty} P(U_n \le z) = \Phi(z), \quad z \in \mathbf{R}.$$

Detta är inte på något sätt självklart, utan behöver egentligen bevisas. Tyvärr kräver beviset en del saker som gör att det blir lite för ingående, så vi tar detta som givet. Lite handviftande kan man tänka att konvergens innan vi gör en Laplace- eller z-transform är detsamma som efter (vilket som sagt inte är helt självklart). Det är heller inget speciellt med just normalfördelningen här, utan konvergens av de momentgenererande funktionerna implicerar konvergens i fördelning. Motsvarande kontinuitetsegenskap finns även för karakteristiska funktioner: om $\phi_{X_n}(t) \to \phi_X(t)$ så kommer $X_n \xrightarrow{\text{D}} X$.

Centrala gränsvärdessatsen (CGS)

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

Bevis. Vi vill visa konvergens i fördelning. Låt $Z_n = \dfrac{Y_n - n\mu}{\sigma\sqrt{n}}$. Vi visar att

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z), \quad z \in \mathbf{R},$$

där

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

Definiera $U_k = \dfrac{X_k - \mu}{\sigma}$ så att $E(U_k) = 0$ och $V(U_k) = 1$. Eftersom $X_1, X_2, \ldots$ var oberoende och likafördelade så gäller även det för $U_1, U_2, \ldots$ Vi kan nu se att

$$M_{U_k}(t) = M_{X_k}\left(\frac{t}{\sigma}\right) e^{-\mu t},$$

så givet att M_{X_k} existerar så gör även M_{U_k} det. Eftersom M_{X_k} existerar nära noll kan vi Maclaurinutveckla enligt

$$M_{U_k}(u) = M_{U_k}(0) + M'_{U_k}(0)u + \frac{M''_{U_k}(0)}{2}u^2 + O(u^3) = 1 + \frac{u^2}{2} + O(u^3),$$

för u nära noll, då $M'_{U_k}(0) = E(U_k) = 0$ och $M''_{U_k}(0) = E(U_k^2) = V(U_k) + E(U_k)^2 = 1$. Vidare gäller att

$$Z_n = \frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} U_k,$$

så eftersom följden $U_1, U_2, \ldots$ är oberoende erhåller vi

$$M_{Z_n}(t) = E(e^{tZ_n}) = E\left(\exp\left(\frac{t}{\sqrt{n}}\sum_{k=1}^{n} U_k\right)\right) = \left(M_U\left(\frac{t}{\sqrt{n}}\right)\right)^n,$$

där M_U är den gemensamma momentgenererande funktionen för följden $U_1, U_2, \ldots$ Låt nu t vara ett fixerat tal. Om n är stort så gäller då att

$$\begin{aligned}
M_{Z_n}(t) &= \left(M_U\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{t^2}{2n} + O\left(\frac{t^3}{n\sqrt{n}}\right)\right)^n = \left(1 + \frac{t^2}{2n} + O\left(n^{-3/2}\right)\right)^n \\
&= \exp\left(n \ln\left(1 + \frac{t^2}{2n} + O\left(n^{-3/2}\right)\right)\right) = \exp\left(n\left(\frac{t^2}{2n} + O\left(n^{-3/2}\right)\right)\right) \\
&= e^{t^2/2} \exp\left(O\left(n^{-1/2}\right)\right) \to e^{t^2/2},
\end{aligned}$$

då $n \to \infty$. Men detta är den momentgenererande funktionen för $N(0,1)$-fördelningen, och enligt utsagan ovan så medför konvergens av momentgenererande funktioner konvergens i fördelning. Detta visar alltså centrala gränsvärdessatsen i fallet då momentgenererande funktioner existerar. □

14.6 Om de momentgenererande funktionerna inte finns då?

Man kan genomföra motsvarande bevis med karakteristiska funktioner istället. Dessa existerar alltid, vilket kräver lite mer teori för att visa ordentligt, men vi ser att e^{itx} är begränsad, vilket inte gäller e^{tx}, så resultatet är inte helt förvånande. Analogt med resultatet att konvergens av momentgenererande funktioner innebär konvergens i fördelning så gäller motsvarande resultat för karakteristiska funktioner. Samtliga steg i beviset ovan fungerar alltså om vi ersätter momentgenererande funktioner med karakteristiska (och lever med att saker och ting blir komplexvärda). Så varför gå omvägen via momentgenererande funktioner istället för att direkt nyttja karakteristiska funktioner? Bra fråga.