

Sannolikhetslära F

Urban Hjorth

April 1, 2003

1 Finns slump?

Mellan idealiserade fysikaliska lagar och vanliga observationer behövs ett “kitt” som klarar av att det inte stämmer exakt ens i vetenskapliga laboratorier och ännu mindre vid mätningar i fält. Den industriella situationen är kanske ännu mer approximativ än vetenskapliga laboratorie- och fältmätningar för där klarar man ofta inte att teoretiskt modellera den fysik och kemi som hanteras. Hur beror t.ex. papperets draghållfasthet på egenskaperna hos den ved och de tillsatser som man kokar? Istället mäter man experimentellt och ofta med betydande variationer som skall hanteras och tolkas.

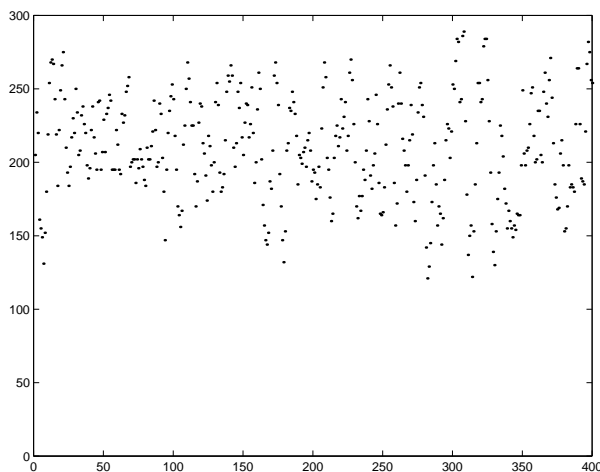


Figure 1: Variationer på kvalitetsmått i löpande papperstillverkning.

Genom att sluta klamra sig fast vid en förenklad deterministisk (ödesbestämd) världsbild och acceptera slumpmässig variation som ett naturligt fenomen kan vi både mäta och tolka den information som ges av störda data. Detta har man tagit fasta på i modern processtyrning där datorernas eller styrkretsarnas ekvationer baseras på slumpmodeller för kommande avvikelser från den ideala banan. Några situationer att fundera över:

Om man studerar jorden från satelliter eller studerar atmosfären genom att mäta mottaget solljus så får man kraftiga störningar att hantera. Dessa kan

inte förklaras av någon deterministisk modell eftersom vi inte har en chans att kartlägga temperaturskiktningar, småskalig turbulens och fuktvariationer som ljuset passerar utan resultaten modelleras bäst som slumpprocesser. Motsvarande gäller för de flesta typer av kvalificerade mätningar och för telekommunikation, radar, atmosfärfysik, miljömätningar etc.

En makroskopisk bild av materia ger t.ex. att gaser har tryck, temperatur och volym och att PV/T är konstant för en innesluten gas. Försök att mäta dig fram till denna lag så märker du snart att det finns variationer i data som inte ger konstant resultat ens om utrustningen är förstklassig (fast felet blir naturligtvis mindre då). Om gasen blir extremt tunn blir inte den fysikaliska tolkningen av relationen PV/T lika självklar. Trycket ges av molekylernas bombardemang och påverkas av deras hastigheter och massor. Molekylernas hastighetsfördelning blir intressant och man uppfattar till sist de individuella partiklarna eller åtminstone variationer som de ger upphov till och begreppet tryck blir en medelvärdesbildning liksom temperatur blir en egenskap i hastighetsfördelningen.

På ännu finare nivå än gasens molekyler kan man inte teoretiskt förklara det grundläggande spelet hos elementarpartiklarna utan att införa sannolikhetsmodeller och det har verkligen gjorts försök att undvika detta. Trots det kan man kanske inte utesluta möjligheten att en ny ickestokastisk teori kan formuleras för elementarpartiklar.

Begreppet slump är tydligen användbart till att beskriva effekterna av en massa okända detaljer och till modellering av partikelfysik liksom till modellering av störningar och dynamik i diverse industriella och andra sammanhang. Det är ett viktigt inslag i mycket av dagens ingenjörskonst. Ändå kan man naturligtvis diskutera om slumpen finns som en fysikalisk realitet. Genom kaosmodeller har man t.ex. visat att mycket oregelbundna och märkliga resultat kan fås ur enkla slumpfria uttryck och datorernas slumptal, som så väl illustrerar hur slump fungerar, är ju ingen riktig slump om inte slumpen kan föras in via startvärdet. Var och en har möjlighet att välja sin tolkning men det blir onekligen roligare att lära sig något som existerar än något som bara approximerar.

Det finns filosofiska övertoner i valet av teoretisk världsbild. Medan deterministens värld är ödesbestämd och din och min vilja därmed också given på förhand så är den stokastiska världsbilden betydligt rikare på möjligheter. Det hela kan börja i vilken enkel situation som helst, t.ex. tärningkastet eller lotten. Uppfattar vi slump i någon sådan situation så finns den i naturen och därav följer det hela.

2 Grundmodellen

Om det som vi observerar kan ge flera möjliga resultat trots samma (synbara) yttre förutsättningar så är det läge för en sannolikhetsmodell. Den består i grunden av två enkla delar:

- en beskrivning av alla möjliga resultat;
- ett mått på sannolikheten för de olika resultaten.

2.1 Utfall och utfallsrum

Vi kallar möjliga resultat för *utfall* ω och mängden av alla utfall kallar vi *utfallsrummet*

$$\Omega = \text{mängden av utfall } \omega. \quad (1)$$

Ett sådant utfallsrum kan vara allting från en lista med några få möjligheter till ett komplicerat matematiskt begrepp med oändligt många utfall. Det blir gott om exempel på både ändliga och oändliga utfallsrum längre fram men vi ger ett par inledande fall här. Ett tärningkast har $\Omega = \{1, 2, 3, 4, 5, 6\}$ när man studerar vilken sida som kommer upp. Ett annat enkelt försök är att gå ut på gatan och se hur många personer man behöver möta tills man möter någon som “ser värre ut” än man själv. Det försöket har (nästan) ingen övre gräns så vi väljer $\Omega = \{1, 2, 3, \dots\} = N$ som mängden av alla naturliga tal och det är ju ett numrerbart oändligt utfallsrum.

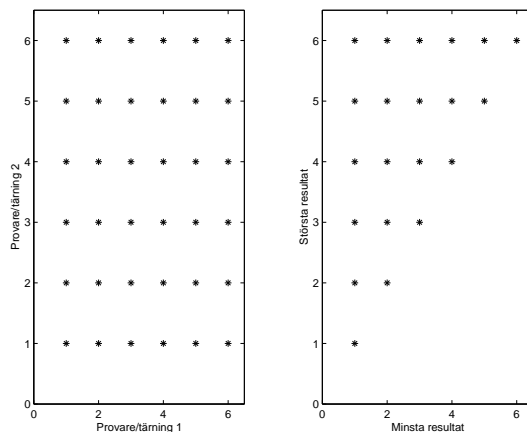


Figure 2: Utfallsrum för två smakprovare (tärningkast) som a) identifierar individerna (tärningarna), b) inte skiljer dem åt utan sorterar resultaten.

Exempel 2.1 Tärningkast eller vinprovning kan beskrivas med samma utfallsrum. Om vi vill jämföra ett nytt vin med fem välkända vinsorter så kan vi ge en försöksperson i uppdrag att rangordna smakupplevelsen utan att veta vilket

vin som är det nya. Det nya vinet kan då hamna på platserna 1 till 6 precis som tärningen. (Sannolikheterna kommer naturligtvis att bero på hur gott det nya vinet är.) Utfallsrummet för ett sådant försök med två vinprovare (eller kast av två tärningar) kan göras på åtminstone två sätt som vi kan åskådliggöra grafiskt i figur 2.

Exempel 2.2 Om vi närmar oss en riktig provsmakningssituation och i exempel ?? skaffar en större smakpanel bestående av t.ex. 24 försökspersoner som alla får rangordna det nya vinet jämfört med de fem gamla (eller om vi gör 24 kast med tärning) så kan vi inte längre rita utfallsrummet. Istället skapar vi då en vektorbeskrivning så att ett utfall ω antingen blir en vektor med 24 element i $\omega = (u_1, u_2, \dots, u_{24})$, där u_i är heltal mellan 1 och 6 svarande mot den i :te försökspersonens rangvärde på vinet eller också offrar vi kopplingen till individerna och beskriver utfall med vektorer $\omega = (v_1, v_2, \dots, v_6)$ där v_i anger hur många gånger rangvärdet i erhålls och därför är heltal mellan 0 och 24 som uppfyller $\sum v_i = 24$.

Anmärkning 2.3 Exemplet visar ett par viktiga saker. För det första finns det flera utfallsrum som kan passa på samma praktiska problem så vi har ingen entydighet utan försöker välja ett som gör beräkningar och sannolikheter så enkla som möjligt. Antalet utfall är i regel så stort att det inte lönar sig att sträva efter att skapa få utfall utan det är bättre att behålla symmetrier och en tydlig struktur. Dessutom behöver vi nästan aldrig summera eller på annat sätt peta i de enskilda utfallen utan är mest intresserade av att de finns och kanske kan grupperas i olika mängder. Ibland behöver vi kunna räkna dem. Till exempel är det enklare att arbeta med 24-dimensionella vektorer i utfallsrummet än med den kortare beskrivningen med 6-dimensionella vektorer med bivillkor. Detta trots att det blir enormt mycket fler utfall med de långa vektorerna. ($6^{24} = 4.7 \times 10^{18}$ mot $\binom{24+5}{5} = 118755$ dvs drygt tio "svenska" biljoner (10^{13}) gånger flera.)

Anmärkning 2.4 Exemplet visar också att när ett försök upprepas eller när man gör en kedja av delförsök (här 24 upprepningar) så kan vi låta hela resultatserien representeras av ett samansatt utfall som kan uttryckas med en vektor eller något likvärdigt (en funktion eller en bild t.ex.).

2.2 Händelser som grund för sannolikheter

Sannolikheter betecknas med \mathbf{P} (som i probabilité eller probability) och om Ω är ändligt eller numrerbart oändligt så kan man oftast ge varje utfall en egen sannolikhet $p_i = P(\omega_i)$. Varje utfall i tärningkastet (när man bara gör ett kast) bör t.ex. rimligen få sannolikheten $1/6$ eftersom alla sidor har samma chans. Det kan bli besvärligare när utfallen är fler än numrerbart många. Vi kan t.ex. ha reella axeln som utfallsrum, eller bara en liten del av axeln t.ex. intervallet $(0,1)$. Tänker vi då att utfall intill varandra bör ha ungefär samma sannolikhet och försöker ge denna ett värde större än noll så blir omedelbart totalsannolikheten

oändlig om vi vill tillåta summation. I det läget utgår vi istället från s.k. *händelser* som är sannolikhetslärans namn för (en lämplig klass av) mängder av utfall.

På reella talaxeln kan händelserna vara delintervall och sammansättningar av flera delintervall (ända upp till numrerbart många) och sannolikheter kan då ofta beskrivas som integraler av någon positiv funktion över det område som ingår i händelsen. Om sannolikheter ges av integraler så vet vi att vanliga integraler blir noll om man bara integrerar över en enda punkt. Eftersom ett helt intervall bör ha mer sannolikhet än en punkt inuti intervallet så gäller för utfallet $\omega = a$ (händelsen bestående av detta enda utfall)

$$P(a) \leq \int_{a-h}^{a+h} f(x)dx \rightarrow 0, h \rightarrow 0.$$

I en sådan beskrivning är det därför naturligt att enskilda utfall har sannolikheten 0 men att en händelse bestående av överuppräknligt många utfall ändå kan få positiv sannolikhet. En annan sida av saken är då att händelser som har sannolikheten noll faktiskt inträffar. Om vi t.ex. studerar tiden tills nästa strömavbrott inträffar så är det naturligt att använda positiva reella axeln som utfallsrum och se tider intill varandra som ungefär lika sannolika. Vi har ju ingen atomär bild av tiden utan ser den som en kontinuerlig storhet. Vi skall nu försöka hitta en funktion att integrera fram sannolikheter ur. Funktionen $f(x) = \alpha e^{-\alpha x}$ är ett naturligt val i den här situationen vilket vi motiverar längre fram. I något ögonblick händer ett avbrott men det ögonblicket hade ju liksom alla andra sannolikheten noll.

2.2.1 Spelregler för händelser

I oändliga utfallsrum behövde vi införa en klass av händelser som vi sedan kan definiera sannolikheter för. Det är då viktigt att den klassen blir så stor så att vi kan göra de beräkningar som behövs. Samma mängdoperationer går självklart att utföra på ändliga utfallsrum så det som följer gäller generellt.

Låt \mathbf{H} vara mängden av alla händelser som vi definierar och använd stora bokstäver A, B, C, \dots i alfabetets början som namn på enskilda händelser i \mathbf{H} . Vi kräver nu att \mathbf{H} inte är tom och att om A, B , och A_i tillhör \mathbf{H} så skall också följande mängder ligga i \mathbf{H} ¹

$$\begin{array}{ll} A \cup B & \text{union, utfall som ligger i minst en av } A \text{ och } B \\ \bigcup_1^\infty A_i & \text{numrerbar union} \\ A \cap B & \text{snittmängd, utfall som ligger i både } A \text{ och } B \\ \overline{A} & \text{komplement, utfall som inte tillhör } A. \end{array} \quad (2)$$

Kraven gör att \mathbf{H} blir en händelsealgebra (en sigmaalgebra) som är sluten under union, snitt och komplement bara vi begränsar oss till numrerbart många sådana operationer.

¹Egentligen har vi onödigt många villkor i (??). T.ex. innebär anmärkning ?? tillämpad på 2 händelser A och B att $A \cap B$ följer av union och komplement och parvis union följer av numrerbar union om man sätter alla händelser lika utom en. Av pedagogiska skäl och för att spara tid till viktigare saker har egenskaperna listats direkt.

Anmärkning 2.5 Om Ω är ändligt så kan vi definiera \mathbf{H} som klassen av alla delmängder till utfallsrummet. Då finns det bara ändligt många händelser och numererbara unioner och snitt kommer i så fall att upprepade gånger använda samma händelse. Självklart ligger då alla resultat av mängdoperationer på händelserna kvar i \mathbf{H} eftersom \mathbf{H} innehöll allt.

Anmärkning 2.6 Övertyga dig om att numererbara snitt är tillåtna utan att lämna \mathbf{H} om (??) gäller genom att verifiera att $\overline{\cup A_i} = \cap \overline{A_i}$, där strecket över mängduttryck betyder komplementhändelsen. Eftersom vänster led bara använder tillåtna operationer så ligger det i \mathbf{H} och därmed gör också höger led det.

Vad skall man ha dessa begrepp till? Jo dels visar det sig lämpligt att definiera sannolikheter för en sådan klass \mathbf{H} , dels behöver vi dem för en beskrivning av många vardagshändelser som kombinationer av mer elementära händelser. Vi antyder ett par sådana fall här.

Exempel 2.7 Antag att ett teledokument behöver gå över k länkar för att komma fram och att vi gör en stokastisk modell för när länkarna är hela eller trasiga. Den modellen får oändligt många utfall om vi vill göra en tidsbeskrivning och inte bara se på en enda tidpunkt. Låt $A_i =$ "länk i är trasig när meddelandet skall fram", d.v.s. A_i är mängden av sådana utfall i utfallsrummet. $\cup_1^k A_i$ betyder att meddelandet stoppas någonstans och $\cap_1^k \overline{A_i} = \overline{\cup_1^k A_i}$ betyder att det kommer fram.

Exempel 2.8 En lastbil med 3 hjulaxlar har 2 framhjul och 4 bakhjul. Den kan köras så länge båda framhjulen och minst ett bakhjul på vardera sidan är hela. I ett utfallsrum modelleras hjulens tillstånd. Låt A_1, A_2 vara händelsen att höger resp. vänster framhjul är sönder, B_1, B_2 detsamma för främre bakhjulen och C_1, C_2 samma för bakre bakhjulen. Låt D vara händelsen att den inte kan köras. Då är

$$D = A_1 \cup A_2 \cup (B_1 \cap C_1) \cup (B_2 \cap C_2)$$

$$\overline{D} = \overline{A_1} \cap \overline{A_2} \cap (\overline{B_1} \cup \overline{C_1}) \cap (\overline{B_2} \cup \overline{C_2})$$

Eftersom det i vissa modeller kan vara enklare att beräkna sannolikheter för snitt än för unioner och komplementhändelser ibland blir mycket enklare att studera så gör man rätt ofta omskrivningar av sina händelser.

3 Sannolikheter

Sannolikhetsmättet P definieras för klassen \mathbf{H} av händelser så att ett antal naturliga axiom uppfylls. Låt \emptyset beteckna den tomma mängden och kalla två mängder A och B *disjunkta* om $A \cap B = \emptyset$. De är då uppbyggda av helt olika utfall. Ryssen A. Kolmogorov införde på 30-talet ett axiomsystem för sannolikheter.

Kolmogorovs axiom

- a) $0 \leq P(A) \leq 1$
 - b) $P(A \cup B) = P(A) + P(B)$ om $A \cap B = \emptyset$ disjunkt union
 - c) $P(\cup A_i) = \sum P(A_i)$ om $A_i \cap A_j = \emptyset, i \neq j$. numrerbar union
 - d) $P(\Omega) = 1$.
- (3)

Ur axiomen kan alla grundläggande resultat härledas. T.ex. följer det självklara resultatet att $P(\bar{A}) = 1 - P(A)$. (Om sannolikheten för A är 0.3 så är sannolikheten att A inte inträffar 0.7.) Använd följande steg för att formellt visa detta. Eftersom \bar{A} innehåller precis de utfall som inte finns i A så gäller $A \cap \bar{A} = \emptyset$ och $A \cup \bar{A} = \Omega$ varför b och d ger $P(A) + P(\bar{A}) = P(\Omega) = 1$, eller $P(\bar{A}) = 1 - P(A)$.

En enkel figur gör många sannolikhetsresonemang självklara. Om vi bara betraktar 2 händelser så finns det i princip 4 möjligheter att tillhöra eller inte tillhöra händelserna. Vi kan antingen göra en fyrfältstablå eller rita ett s.k. Venndiagram:

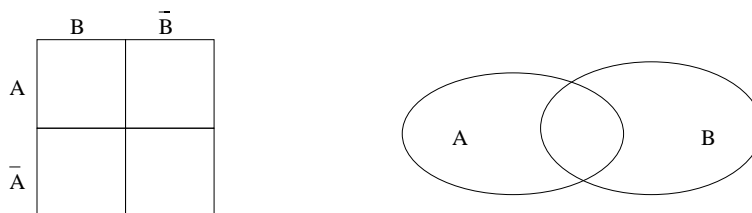


Figure 3: Fyrfältstablå till vänster och Venndiagram till höger. I Venndiagrammet ligger $\bar{A} \cap \bar{B}$ som yta utanför mängderna A, B .

Vi kan nu lätt se att det finns en generalisering av unionssannolikheten i axiomen

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{4}$$

I figuren ges unionen av de tre delområden som tillhör bara A eller bara B eller $A \cap B$ och vänster led ger summan av deras sannolikheter (enligt ?? punkt b eller c). Summan $P(A) + P(B)$ räknar in snittsannolikheten två gånger och det korrigerar vi genom att subtrahera den i höger led.

Förväxla inte mängdregler och sannolikhetsmått. I unioner av händelser finns det ingen dubbelräkning av utfallen – det som är med i en enda unionshändelse är lika mycket med som det som ingår i flera av unionshändelserna. I $P(A) + P(B)$ blir

däremot snittsannolikheten räknad dubbelt medan den i unions sannolikheten bara skall räknas en gång.

Exempel 3.1 Ett tvådagars konsertarrangemang skall regnförsäkras och försäkringsbolaget vill bedöma risken att få betala. Man beslutar att ersättning skall utgå om det någon av dagarna blir regn, definierat som minst 0.1 mm nederbörd i en närliggande mätstation under den aktuella tiden. Sannolikheten för mer än 0.1 mm nederbörd under de 8 timmar som är kritiska under en kväll är 0.08 enligt klimatdata för den aktuella månaden och orten och sannolikheten att det regnar två dagar i rad under dessa timmar är 0.04. Vi skall beräkna sannolikheten att det blir regn någon av dagarna. Inför händelserna $A_1 =$ regn dag 1, $A_2 =$ regn dag 2. Då är $P(A_1) = P(A_2) = 0.08$ och $P(A_1 \cap A_2) = 0.04$ och vi får $P(A_1 \cup A_2) = 0.08 + 0.08 - 0.04 = 0.12$ som är sannolikheten att få betala ut ersättning.

Unioner av flera händelser

En bra övning i unioner är att visa att

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Man kan antingen se det i ett Venndiagram genom att kontrollera vilka områden (utfall) som räknas flerdubbelt eller inte alls i olika steg av summan, eller också använder man unionsformeln för 2 händelser upprepade gånger. Därefter ser man troligen hur formeln kan byggas ut till unionen av n händelser.

Exempel 3.2 Det är fest och n stycken Emil går dit med varsin tjej och varsin tofsmössa. Efter ett tag ligger alla mössorna huller om buller och i det tillstånd som råder tar alla Emil med sig en flicka och en mössa helt på måfå när det är dags att gå hem. Flickorna vet nog vad de gör men vad är sannolikheten att ingen Emil får sin egen mössa med sig hem? Det är inget lätt problem att lösa men motsatsen, sannolikheten att någon får sin egen mössa, kan vi åtminstone teckna som en union av n händelser $A_i =$ Emil nr i tar sin egen mössa. Med den formel för sannolikheten för en sådan union som du eventuellt har kommit fram till, och eftersom så många sannolikheter blir lika ($P(A_i) = 1/n$, $P(A_i \cap A_j) = \frac{1}{n} \frac{1}{n-1}$ etc.) så får man ett vackert resultat som dessutom konvergerar när $n \rightarrow \infty$ och redan ett litet antal $n \geq 5$ räcker nästan för konvergens².

Exempel 3.3 Vid en statistisk slutsats som baseras på data tar man alltid en viss risk att göra fel. Konsten är att kunna bedöma risken och göra den rimligt liten. Låt α_i vara sannolikheten att slutsats nummer i blir fel och antag att vi behöver göra k olika sådana slutsatser. Vi skall visa att sannolikheten att alla blir korrekta är minst $1 - \sum_1^k \alpha_i$.

²Sannolikheten att någon får sin egen, $P(\cup A_i) \rightarrow 1 - \frac{1}{e}$ Sannolikheten att ingen får den $\rightarrow \frac{1}{e} = 0.368$. $n = 4$ ger $9/24=0.375$, $n = 5$ ger 0.3667 och $n = 6$ ger 0.368 .

Inför händelsen $B =$ alla slutsatser är korrekta, och händelserna $A_i =$ slutsats i blir fel. Låt $A = \cup A_i$ som innebär att någon slutsats blir fel. Då är $A = \overline{B}$ och $1 - P(B) = P(\cup A_i)$. Vi skall visa att den sista unionens sannolikhet begränsas av summan av de ingående händelsernas sannolikheter.

För två händelser gav (??) att $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$ så där stämmer det. Men stämmer det för två så stämmer det för tre eftersom $A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3$, och då blir $P((A_1 \cup A_2) \cup A_3) \leq P(A_1 \cup A_2) + P(A_3)$ enligt resultatet för två händelser (där unionen i parentes blir den ena händelsen). Men vi hade redan $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$ och får därför $P(A_1 \cup A_2 \cup A_3) \leq P(A_1) + P(A_2) + P(A_3)$. Samma metod kan användas rekursivt för att öka upp antalet händelser till ett godtyckligt k . Slutligen får vi $P(B) = 1 - P(\cup A_i) \geq 1 - \sum P(A_i)$ eller hur?.

Det vi kom fram till i beviset av exemplet är en enkel och välkänd olikhet (Booles olikhet eller Bonferronis olikhet beroende på sammanhanget).

$$P(\cup A_i) \leq \sum P(A_i); \tag{5}$$

där indexet i tar högst numrerbart många värden. Fundera gärna ut vad den betyder i ett Venndiagram.

3.1 Hur många möjligheter? Ett sätt att bestämma sannolikheter.

När vi vill se på hur många sätt något kan göras så är vi eventuellt tillbaka i utfallsrummet och räknar utfall. Det blir ofta stora tal. Antalsräkningen kan ofta kopplas till sannolikheter, särskilt när de olika utfallen är lika sannolika på grund av symmetri (och ändligt många). Då får vi sannolikheter för händelser om vi dividerar antalet utfall i händelserna med det totala antalet utfall. I oändliga utfallsrum kan man ibland ha nytta av att istället räkna antalet händelser av en viss typ så metoder för att systematiskt beräkna antal är nyttiga där också. Liknande antalsräkning kan vara bra i helt andra situationer också, t.ex. för att kontrollera hur många specialfall en programmerare tvingar datorn att löpa igenom.

3.1.1 Multiplikationsprincipen

Om matsedeln på krogen har 4 förrätter, 7 varmrätter, 3 deserter, 4 sorters kaffe och 12 dryckesalternativ och vi bestämmer oss för att äta en rejäl middag med alla de här ingredienserna, hur många olika beställningar kan vi då göra? Ja, för varje förrätt kan vi välja 7 huvudrätter så redan där har vi $4 \times 7 = 28$ möjligheter. Varje sådan kan sedan kombineras med vilken desert som helst. Ny produkt eller hur! Det totala antalet kombinationer ges av produkten $4 \cdot 7 \cdot 3 \cdot 4 \cdot 12 = 4032$ så kyparen har en del att hålla reda på. I det här fallet är det samma varmrätter som kan beställas oavsett vilken förrätt man tog, men det enda som är viktigt för multiplikationsprincipen är att *antalet* möjligheter är givet oavsett tidigare val.

Multiplikationsprincipen kan därför formuleras för en serie av k steg

Om det finns n_1 alternativ i steg 1 och för varje sådant alternativ finns n_2 alternativ i steg 2, och för varje sätt att utföra steg 1 och 2 finns n_3 alternativ i steg 3 o.s.v. ända fram till steg k så kan serien av steg utföras på $n_1 \times n_2 \times \dots \times n_k$ sätt.

3.1.2 Antalet permutationer

a	b	c
a	c	b
b	a	c
b	c	a
c	a	b
c	b	a

Antalet sätt att permutera (ordna) n föremål är $n!$. Detta kan vi se på åtminstone två sätt. Antingen tänker vi oss n fack placerade i rad där vi skall lägga en enhet i varje. Den första kan då hamna på n platser, för varje sådan plats kan sedan den andra hamna på $n - 1$ platser, sedan den 3:e på $n - 2$ ända ner till den sista som bara har en plats. Varje sådan utplacering blir unik och ger en permutatiom av föremålen. Här används multiplikationsprincipen med lika många, men inte samma, alternativ i de olika stegen.

Det andra sättet att tänka börjar från motsatt håll. Ett föremål kan bara ordnas på ett sätt. Nästa föremål kan antingen placeras framför eller bakom så det blir två sätt. För varje ordning av de båda första kan ett tredje föremål placeras in på tre platser (främst, i mitten, sist) så det blir tre gånger fler möjligheter än med 2 föremål. För varje sådan ordning av tre kan en fjärde placeras in på 4 platser, så antalet möjligheter blir då 4 gånger fler än med tre föremål. Vi tillämpar konsekvent multiplikationsprincipen och kommer fram till $n!$ efter n steg på båda sätten.

På en gata med 20 parkeringsplatser kan 20 bilar parkeras på $20!$ sätt.

På en bjudning med 5 par kan man para ihop kavaljer med dam på $5! = 120$ sätt om man inte tar hänsyn till vem som sällskapar med vem.

Kortleken med sina 52 kort kan blandas på $52!$ sätt.

Framför mig i bokhyllan står just nu 67 böcker. De kan ordnas på $67!$ sätt men det skulle ta ett tag att genomföra alla dessa omsorteringar. Fakulteter kan nämligen bli stora tal.

Det är tur att algoritmer som storleksordnar vektorer av tal är smarta (t.ex. rutinen `sort` i `matlab`). En slumpvis sortering skulle kunna passera ett rätt stort

antal av de möjliga permutationerna och bli omöjlig att genomföra under datorns livstid.

En uppfattning om fakulteternas mäktighet ges av följande tablå:

$1! = 1$	
$2! = 2$	
$3! = 6$	Fakulteter och andra tal
$4! = 24$	
$5! = 120$	
$6! = 720$	
$7! = 5040$	
$8! = 40320$	Växjö
$9! = 362880$	Göteborg
$10! = 3628800$	Sverige
$11! = 39916800$	
$12! = 479001600$	
$13! = 6.2270 \times 10^9$	Jordens befolkning
$15! = 1.3077 \times 10^{12}$	
$20! = 2.4329 \times 10^{18}$	
$23! = 2.5852 \times 10^{22}$	en liter luftmolekyler
$30! = 2.6525 \times 10^{32}$	
$38! = 5.2302 \times 10^{44}$	molekyler i atmosfären
$43! = 6.0415 \times 10^{52}$	atomer på jorden
$48! = 1.2414 \times 10^{61}$	
$52! = 8.0658 \times 10^{67}$	kortleken

Det sista talet är astronomiskt, 10^{15} jordar grovt räknat. Det börjar likna vintergatans atomer även om jag saknar en siffra på hur många dessa är. Talet är i själva verket så enormt stort att vi med mycket stor tillförsikt kan våga nacken på att en väl blandad kortlek, där alla ordningar är lika sannolika, ligger på ett sätt som ingen annan kortlek någonsin har legat.

(Om 10 miljarder människor sitter i 100 år vardera och varje år är 10000 timmar och varje timme ger 1000 slumpvisa permutationer per person så har totalt 10^{19} permutationer uppkommit av de 8×10^{67} som är möjliga. Sannolikheten att alla dessa permutationer är olika kan beräknas till ca $1 - 10^{-30}$ och man kan räkna med att få den första upprepningen efter storleksordningen kvadratroten ur antalet möjligheter vilket blir ca 10^{34} dragna permutationer, vilket är 10^{15} gånger fler än min kortgalna population ovan har hunnit med.)

3.1.3 Antalet sätt att dra k bland n

Det finns olika sätt att beskriva dragningar. Om vi har en kortlek med 52 kort och delar ut 13 kort (en bridgehand) till en spelare så bryr man sig bara om vilka kort det blev utan att ta hänsyn till i vilken ordning de delades ut. Att

räkna antalet möjliga sådana mängder av 13 kort kallas att räkna utan hänsyn till ordningen. Dessutom lägger man inte tillbaka korten medan man drar så en fullständig formulering är att dra utan återläggning och utan hänsyn till ordningen. Låt för ett ögonblick C_k^n beteckna antalet sätt att på detta sätt dra k bland n .

Det andra sättet att räkna antal dragningar utan återläggning på är att skilja på olika ordningar. Multiplikationsprincipen ger att vi då kan dra k bland n ($k \leq n$) på $n(n-1)(n-2)\cdots(n-k+1)$ olika sätt. Det blir naturligtvis fler sätt när ordningen beaktas och inte bara vilka enheter man fick. De k dragna korten kan ordnas om på $1 \times 2 \times \cdots \times k = k!$ sätt. Eftersom mängden av ordnade dragningar kan erhållas genom att först dra unika "händer" och sedan skapa alla ordningar av dessa händer så måste $C_k^n k! = n(n-1)\cdots(n-k+1)$ vilket ger

$$C_k^n = \frac{n(n-1)\cdots(n-k+1)}{k!} = \binom{n}{k}. \quad (6)$$

I fortsättningen glömmet vi beteckningen C_k^n och minns istället $\binom{n}{k}$, där $\binom{n}{k}$ är den s.k. binomialkoefficienten. (Specialfall $\binom{n}{0} = \binom{n}{n} = 1$ och namnet binomial har den fått av att $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$ där $(a+b)$ är ett binom.)

Lätt konstateras nu ett det finns $\binom{52}{13}$ olika bridgehänder och $\binom{52}{5}$ pokerhänder liksom att man bland 90 teknologer kan plocka ut en styrelse med fem personer i på $\binom{90}{5}$ sätt om man inte väljer styrelsen till sina funktioner. En styrelse bestående av ordförande, kassör, sekreterare, övrig ledamot och sexmästare kan däremot väljas bland samma teknologer på $90 \times 89 \times 88 \times 87 \times 86$ sätt eftersom man nu kan se det som en ordnad mängd. (Denna typ av tolkningar får man ofta inse själv t.ex. genom att fantisera ihop ett valförfarande där man först väljer ordförande, sedan kassör och så vidare.)

I vissa sammanhang läggs dragna individer tillbaka och kan bli dragna på nytt. Om man då håller reda på ordningen så finns det n^k olika dragningar. Om man däremot bortser från ordningsföljden så brukar man ändå beakta hur många gånger varje enhet blev dragen och då kan man med ett finurligt men rätt svårfunnet knep visa att det finns $\binom{n+k-1}{k}$ möjliga sådana oordnade dragningar. (Det är sällan man har nytta av det resultatet eftersom dessa utfall inte blir lika sannolika vid en slumpmässig dragning.)

Sammanfattning: Antalet sätt att dra k bland n

		Hänsyn till dragningsordning	
		nej	ja
Återlägger dragna	nej	$\binom{n}{k}$	$n(n-1)\cdots(n-k+1)$
	ja	$\binom{n+k-1}{k}$	n^k

3.2 Hypotetiska sannolikheter och sura sjöar

I situationer där man inte vet så mycket om utfallens egentliga sannolikheter kan man ändå ha nytta av hypotetiska sannolikheter för att styrka (eller argumentera mot) vissa slutsatser. Detta illustreras bäst av ett exempel.

Man har mätt försurningen (PH) i fem orörda sjöar i Småland och i tre orörda sjöar i Dalarna. Sjöarna kan anses valda på måfå. Det visade sig att alla tre sjöarna i Dalarna var surare än Smålandssjöarna.

Det ligger nu nära till hands att säga att försurningen drabbat Dalarna värre än Småland, men faktaunderlaget är ju ganska blygsamt. Det kan faktiskt vara ett resultat av slumpen.

Den sista meningen är förrädisk eftersom vi inte har någon sannolikhetsmodell, men låt oss göra en (hypotetisk) sådan. Antag att det inte var någon skillnad alls på försurningsläget i de båda landskapen. Då är det bara slumpen som avgör ordningen mellan sjöarna vad gäller PH. (I så fall kunde vi lika gärna mätt 8 sjöar i samma landskap och slumpmässigt delat in dem i två grupper med 5 respektive 3 sjöar.) Vi skall beräkna sannolikheten att få ett lika extremt resultat som det observerade under detta hypotetiska antagande. På en surhetsskala skall vi nu placera in fem stycken S och tre D och se på chansen att alla D:na kommer i ena ytterkanten. Det kan vi beräkna på några olika sätt, t.ex. med följande metod.

Lösning: Utfallsrummet definieras som alla sätt att ordna tre D och fem S (när vi inte skiljer på D:na och S:en inbördes).

$$\Omega = \left\{ \begin{array}{l} DDDSSSSS \\ DDS DSSSS \\ DSSDSSSS \\ \vdots \\ SSSSDDD \end{array} \right. \quad (7)$$

Antalet utfall blir antalet sätt att välja 3 positioner bland 8, vilket ger $\binom{8}{3} = \frac{8 \times 7 \times 6}{1 \times 2 \times 3} = 56$ eftersom det kan jämföras med dragning av 3 bland 8 utan återläggning och utan avseende på dragningsordningen. Om bara slumpen avgör ordningen så är utfallen lika sannolika och sannolikheten att de tre dalasjöarna är surast blir då $1/56$ vilket är en ganska liten sannolikhet. Ju mindre denna sannolikhet är, desto starkare blir argumentet att Dalarna faktiskt är mer försurat än Småland.

Detta är ett specialfall av hypotestestning som är en klassisk vetenskaplig metod. När fakta (data) talar tillräckligt starkt emot en modell så förkastar man den och satsar på en troligare förklaring. Man brukar ofta anse sannolikheter över 0.05 som alltför stora att ta hänsyn till och sannolikheter under 0.01 som rätt starka bevis mot antagandet om ren slump (eller mot någon annan lämplig

hypotes om sannolikheterna). Vår sannolikhet är ca 0.02. Om man beaktar att vi hade häpnat precis lika mycket om alla dalasjöarna varit minst försurade så bör man egentligen addera chanserna för utfallen $DDSSSSS$ och $SSSSDDDD$ när man bedömer om resultatet talar mot den hypotetiska sannolikhetsfördelningen. Det ger $2\frac{1}{56} \approx 0.036$ vilket fortfarande understiger 0.05, så vi har ändå rätt bra stöd för påståendet att resultatet inte beror på slumpen enbart och att Dalarna är mer försurat.

Övning: Välj ett annat utfallsrum, t.ex. ett där alla sjöarna är individer och alla permutationer av de 8 sjöarna bildar utfall. Genomför sannolikhetsberäkningen och visa att den ger samma resultat.

Kommentar: Analyser av det här slaget blir naturligtvis svårare om man har ett mer blandat resultat. Antag att vi har två grupper av mätningar (vi kan fortfarande kalla dem S och D även om de kan betyda något annat) och att resultatet ordnar sig så här

$SSDSSSDDSSSSSDSDDSDDDSDDDDD$.

Här har vi 12 S och 13 D och det finns $\binom{25}{13}$ olika mönster av bokstäverna. Det är ett mycket stort tal (ca 226 miljoner) så varje mönster blir väldigt osannolikt om vi antar att ordningen är helt slumpmässig. Eftersom alla mönster i så fall har samma lilla sannolikhet $1/\binom{25}{13}$ så kan vi inte kassera antagandet om ren slump bara på den grunden. I teorin för hypotestestning lär man sig att istället dela in utfallen i två grupper. En grupp bildar en händelse A där man fortsätter att anse ren slump som en fullt möjlig förklaring och resten av utfallen blir en händelse $C = \bar{A}$ som får tala mot detta. Händelsen C skall ha liten sannolikhet i den hypotetiska, rent slumpmässiga, fördelningen men få större chans vid den alternativa förklaringen av typen "Dalarna är mer försurat". Wilcoxon's rangsummetest genererar en sådan uppdelning (med hjälp av summan av platsnumren för bokstäverna D) och ger ett enkelt sätt att avgöra vilken slutsats som bör dras, men vi går inte närmare in på det här.

3.3 Oberoende händelser – och beroende

Oberoende är en av statistikens viktigaste egenskaper och förenklar många sannolikhetsmodeller. Antag att du slår en tärning och kastar ett mynt. Då har resultatet av tärningkastet ingen som helst koppling till resultatet av myntkastet. De båda delresultaten är *oberoende* av varandra.

Sannolikheten att få 5 på tärningen och klave på myntet ges då av *produkten* $\frac{1}{6} \times \frac{1}{2}$ eftersom man om experimentet upprepas bör få klave ungefär varannan gång som tärningen visar 5:a. Om $A = \{5 \text{ på tärningen}\}$ och $B = \{\text{klave på myntet}\}$ så innebär vårt resonemang att $P(A \cap B) = P(A)P(B)$. På motsvarande sätt

kan man utföra mätningar så att slumpfel och variationer i en mätning inte har något samband med fel och variationer i en annan mätning. Då får man också en multiplikation av sannolikheter så att $P(A \cap B) = P(A)P(B)$ om A är en händelse som bestäms av mätning 1 och B en händelse som bestäms av mätning 2.

Vi tar nu bort den bakomliggande motiveringen och definierar rent allmänt oberoende på följande sätt.

Definition 3.4 *Två händelser A och B är oberoende om*

$$P(A \cap B) = P(A)P(B), \quad (8)$$

och fler än två händelser A_1, A_2, \dots är oberoende om

$$P(A_{i_1} \cap A_{i_2} \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})P(A_{i_k}) \quad (9)$$

för varje möjlig uppsättning av k och i_1, i_2, \dots, i_k där i na är olika.

Motsatsen till oberoende heter beroende. För att förstå oberoendet behöver man diskutera båda begreppen. Det visar sig att mätningar som ligger nära varandra i tiden, eller som görs intill varandra i rummet, ofta ger mer lika resultat än observationer tagna längre isär. De nära observationerna är då beroende genom att de påverkas av samma slumpvisa faktorer. Det är lätt att tro att små variationer betyder säkra resultat och att observationer tätt ihop därmed ger en fördel. Detta kan vara helt fel. En slumpvis avvikelse kan lätt ligga kvar ett tag när man mäter t.ex. i en industriell process. Ett visst väderläge ligger också kvar ett tag och har utbredning i rummet och kan påverka väderkänsliga mätningar. Ett felkalibrerat instrument kan störa flera observationer om man inte hinner kalibrera om mellan mätningarna. Mätning av föroreningar i jord eller vatten blir naturligtvis nästan identiska om prover tas alldeles intill varandra fast man vill uttala sig om ett större område. Sådana faktorer skapar beroenden och tätt tagna data ger då händelser som vi inte skall multiplicera sannolikheterna för. Medan oberoende felkällor sannolikt delvis motverkar varandra så brukar beroende felkällor av den typ vi angett verka åt samma håll. Om sådana beroenden inte upptäcks och man räknar som om observationer var oberoende så blir dels resultaten osäkrare, samtidigt blir tyvärr oftast måtten på osäkerhet mindre. Med andra ord kommer man att bli lurad att tro mera på de osäkrare resultaten. För att uppnå oberoende måste mätningar planeras väl och ofta slumpar man fram strategiska delar av planläggningen.

Exempel 3.5 Beroende eller oberoende väderhändelser. Sannolikheten för en mätbar mängd regn under en dag är 0.3 och sannolikheten att två på förhand valda dagar efter varandra (t.ex. tisdag och onsdag) båda får regn är 0.2. Sannolikheten att tisdag och torsdag båda får regn är 0.13 och sannolikheten för tisdag och fredag är 0.10 och för tisdag och lördag är sannolikheten ungefär 0.09 och stannar på det värdet för alla dagpar med ännu större tidsavstånd. Låt $A_0 =$ regn på tisdag (dag 0),

A_k = regn på dag k (onsdag=dag 1 etc.).

Då är

$P(A_k) = P(A_0) = 0.3$, samma chans alla dagar

$P(A_0 \cap A_1) = 0.2$,

$P(A_0 \cap A_2) = 0.13$,

$P(A_0 \cap A_3) = 0.10$,

$P(A_0 \cap A_4) = P(A_0 \cap A_5) = \dots = 0.09$.

Samtidigt är produkten alltid $P(A_0)P(A_k) = 0.09$. Slutsats: regnhändelserna är beroende på tidsavståndet 1,2,3 men blir (approximativt) oberoende på tidsavståndet 4 eller mera.

Siffrorna i exemplet är tillverkade men rätt realistiska. Verkliga regnsannolikheter beror på årstid och geografiska faktorer. Det regnar t.ex. oftare i Borås än i Nässjö och olika ofta på vår och höst. Däremot stämmer ungefär strukturen på beroendet/oberoendet. Ur klimatdata kan mer exakta bedömningar av sådana sannolikheter göras.

Exempel 3.6 Olika individers reaktion på medicinsk behandling, t.ex. händelserna $A_i = \{\text{individ } i \text{ tillfrisknar}\}$, brukar modelleras som oberoende. (Förutsätter att det inte finns någon gemensam storkälla i själva medicinen eller i datahanteringen och individerna bör inte vara nära släkt.)

Exempel 3.7 Händelserna att en älv svämmar över under vårfloden olika år är skapligt oberoende. Händelserna att olika svenska älvar svämmar över samma år däremot beroende. Försök gärna ange orsaker till detta.

Exempel 3.8 I en nyss genomförd Göteborgsstudie försågs ett antal (ca 50 tror jag) personer med en mätare att ha på bröstet. Mätaren observerade ackumulerade mängden luftföroreningar under ca en vecka. Personerna valdes ut *slumpvis* ur en lämplig population av individer och blev oberoende observationer av populationens exponering för sådana föroreningar under mätperioden.

Ett rätt vanligt missförstånd när man ser på händelser är att disjunkta händelser skulle vara oberoende. Om både A och B har positiv sannolikhet och är oberoende så blir $P(A \cap B) = P(A)P(B) > 0$. Då kan inte $A \cap B$ vara tom så oberoende händelser med sannolikhet måste delvis täcka över varandra.

3.4 Betingade sannolikheter

I ett spel kan du slå ut din motståndare om dina två tärningar visar 7 tillsammans. Tänk dig ett utfallsrum med 6 gånger 6 lika sannolika utfall (resultaten på tärning ett och tärning två). Summan blir 7 i ena diagonalen så sannolikheten är $6/36=1/6$ att lyckas. Om nu kastet utförs dolt och någon annan ser vad det blev så kan hon eller han tänkas rapportera att resultatet blev *udda*. Rimligen har du

nu större utsikter att slå ut motståndaren eftersom det är ett udda resultat som behövs. Betingade sannolikheter ger måttet på hur sannolikheten har förändrats med hänsyn till sådan information om resultatet. Principen är enkel. Låt $A = \{\text{udda summa}\}$, $B = \{\text{summan blir } 7\}$. Eftersom A inträffat så blir alla utfall utanför A omöjliga (d.v.s. utfall med jämna summor) och därför får de sannolikheten 0 efter informationen. Sannolikheter för utfall eller händelser innanför A får behålla sina gamla proportioner men normeras om så att A får sannolikheten 1. Det gör man genom att dividera deras sannolikheter med $P(A)$. I exemplet ger hälften av utfallen udda summa så $P(A) = 0.5$ och eftersom alla utfallen i B ligger i A så blir den nya sannolikheten $P(B|A) = P(B)/P(A) = (1/6)/0.5 = 1/3$.

Vi behöver en generellare definition som täcker händelser med utfall både innanför och utanför A . Av händelsen B ligger alltid snitthändelsen $A \cap B$ innanför A och resten är omöjlig om A har inträffat. Därför inför vi följande definition:

Definition 3.9 *Betingade sannolikheten för B givet A ges av*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0, \quad (10)$$

och lämnas tills vidare odefinierad om nämnaren är noll.

Andra uttryck för samma sak är t.ex. sannolikheten för B betingat A eller sannolikheten för B om A inträffat eller liknande. Ibland ger betingade sannolikheter ganska förvånande resultat, men de är logiska och fungerar både i teori och praktik.

Exempel 3.10 Sannolikheten för regn under en dag är 0.3 och om man tar två dagar i följd så är sannolikheten 0.2 att båda dagarna får regn. Den betingade sannolikheten för regn dag 2 givet att det kommer regn dag ett, $P(A_2|A_1) = P(A_1 \cap A_2)/P(A_1) = 0.2/0.3 = 2/3$ vilket är betydligt högre än $P(A_2) = 0.3$. Pröva gärna att beräkna $P(A_2|\bar{A}_1)$, d.v.s. sannolikheten för regn dag två om det inte regnar dag ett. (Det finns bara fyra fall i Venndiagrammet eller fyrfältstabellen.)

Om A och B är oberoende så bör inte information om den ena påverka sannolikheten för den andra. Oberoendet innebar att $P(A \cap B) = P(A)P(B)$ och mycket riktigt blir

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

3.4.1 Betingningskedjor

Vi kan vända på betingningsdefinitionen och lösa ut snittsannolikheten. Det ger

$$P(A \cap B) = P(A)P(B|A) \quad (11)$$

d.v.s. chansen för båda händelserna ges av sannolikheten för den första gånger sannolikheten för den andra betingat den första. Det kan man lätt skaffa sig en intuitiv bild av. Analogt blir

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

eftersom $P([A \cap B] \cap C) = P(A \cap B)P(C|[A \cap B]) = P(A)P(B|A)P(C|A \cap B)$ med dubbel användning av resultatet för två händelser. Man kan gå vidare och bygga hur långa sådana kedjor som helst.

Exempel 3.11 I en nötskål finns 7 friska och 3 kassa hasselnötter. Vad är chansen att vi kan knäcka 4 nötter och bara få friska? Lösning: Låt A_i = den i te nöten är frisk, $i = 1, 2, 3, 4$. Direkt ser vi att $P(A_1) = 7/10$. (Den obetingade sannolikheten är faktiskt lika stor för alla A_i eftersom varje nöt kan bli den i :te med samma chans.) Här är det enklare att se vad de betingade sannolikheterna blir, eftersom det är lätt att se vad som blir kvar i nötskålen. $P(A_2|A_1) = 6/9$, $P(A_3|A_1 \cap A_2) = 5/8$, $P(A_4|A_1 \cap A_2 \cap A_3) = 4/7$. Det ger med en kedja av typen (??) att $P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8} \times \frac{4}{7}$. Resultatet $\frac{7 \times 6 \times 5 \times 4}{10 \times 9 \times 8 \times 7}$ kan formas om till $\binom{7}{4} / \binom{10}{4}$ om man dividerar med $4!$ i täljare och nämnare. Det visar på att samma resultat kan beräknas i ett utfallsrum med alla (oordnade) dragningar av 4 nötter bland de 10 som utfall. På $\binom{7}{4}$ sätt kan vi bilda utfall med bara friska nötter genom att dra 4 bland dessa. Våra intuitiva betingade sannolikheter baserade direkt på nötterna i skålen (utan att vi preciserat utfallsrummet) ger alltså samma resultat via betingningskedjan som en alternativ beräkning i det andra utfallsrummet.

3.4.2 Bayesformler

Formler som vänder på betingningar tillskrivs ofta Bayes (en engelsk präst Reverend Thomas Bayes, 1723-91 som studerade logiken i slutsatsdragningar).

Bayes sats:

$$P(B|A) = P(A|B) \frac{P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} \quad (12)$$

Bevis: Första likheten följer direkt från definitionen av betingad sannolikhet. Eftersom $A = (A \cap B) \cup (A \cap \bar{B})$ där parenteserna är disjunkta eftersom $B \cap \bar{B} = \emptyset$ så blir $P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$.

Generalisering, partitioneringssatsen:

Om B_i är disjunkta ($B_i \cap B_j = \emptyset, i \neq j$) och $\cup B_i = \Omega$ så är B_i en partitionering av utfallsrummet och

$$P(A) = \sum P(A|B_i)P(B_i). \quad (13)$$

Detta kan tolkas så att B_i ses som olika möjliga fall som kan inträffa där man för varje sådant fall kan uttrycka chansen för A som $P(A|B_i)$. Beviset är helt analogt med Bayes sats.

Exempel 3.12 (Rekordslagning) Man observerar en slumpmässig storhet som har kontinuerlig fördelning (alla resultat blir då olika med sannolikheten 1). Vi intresserar oss för stora värden. Antag att alla observationer är oberoende med samma fördelning (ungefär som årets värsta väder vid stabilt klimat). Låt

$$A_i = \{\text{ite observationen är ett rekord}\}.$$

Händelsen inträffar om observation i är större än alla föregående observationer. Två frågor intresserar oss speciellt:

- a) Finn sannolikheten för A_i ;
- b) Är händelserna A_i , $i = 1, 2, \dots$ oberoende?

Pröva gärna att själv angripa detta innan du tar del av följande ansats.

a) $P(A_1) = 1$ självklart. Låt $i > 1$. A_i beror på de i första observationerna. Sannolikheten att den största av dessa kommer sist blir $1/i$ eftersom alla har samma chans att vara störst, $P(A_i) = 1/i$.

b) Studera först parvis oberoende. Låt $1 < i < j$. Händelsen $A_i \cap A_j$ avgörs av de j första observationerna och där skall dels den j te vara störst, dels den i te vara störst bland de i första.

$$P(A_i \cap A_j) = P(A_j)P(A_i|A_j) = \frac{1}{j}P(A_i|A_j).$$

Vi behöver inte beskriva hur stora observationerna är, bara deras inbördes storleksordning. Inför en rangordning med 1 för den minsta, 2 för den näst minsta ända till j för den största av de j första obsarna. Samma fördelning och oberoendet hos observationerna ger att alla permutationer blir lika sannolika när man skriver upp rangerna i den ordning som observationerna tagits. Betingning på A_j ger att rangsiffran j står sist i alla kvarvarande utfall men i övrigt har vi kvar alla permutationer av rangerna $1, 2, \dots, j-1$ som har samma chans. Vilka ranger som hamnar på de i första positionerna beror på slumpen. Det finns $\binom{j-1}{i}$ sätt att välja dessa ranger. För varje sådant val är fortfarande alla permutationer av rangvärdena på de i första positionerna lika sannolika (studera det för något litet fall t.ex $i = 3, j = 5$ genom att skriva upp alla permutationer med j sist om du tvekar) och därmed blir $P(A_i|A_j) = \frac{1}{i}$ och

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

vilket visar parvis oberoende.

Fler än 2 händelser: $P(A_{i_1} \cap A_{i_2} \dots \cap A_{i_k})$. (Alla sådana snitt skall ge produkten av händelsernas sannolikheter.) Utgå från högsta indexet (i_k säg) och alla rangordningar av så många variabler. Samma resonemang som ovan ger att högsta indexet skall få högsta rangen och att betingat detta blir alla permutationer av lägre ranger lika sannolika. För det näst högsta indexet (låt det vara i_{k-1}) gäller lösningen för två händelser som ger att den betingade sannolikheten att motsvarande rang är större än alla ranger för tidigare obsar är $1/\text{indexet}$. Betingat dessa båda händelser är nu alla permutationer för de kvarvarande rangerna fortfarande lika sannolika och resonemanget förs vidare på samma sätt. Det ger

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_k})P(A_{i_{k-1}}|A_{i_k})P(A_{i_{k-2}}|A_{i_k} \cap A_{i_{k-1}}) \dots P(A_{i_1}|\cap_2^k A_{i_r}) = \prod \frac{1}{i_r}.$$

Rekordhändelserna är alltså oberoende.

Exemplet kan ha betydelse för hur man ser på rekord och extrema händelser. Dessa

inträffar oberoende och med rätt långsamt avtagande sannolikhet under förutsättningarna ovan. Efter t.ex. 100 år av observationer är då chansen 1/100 vilket förefaller litet, men samtidigt är chansen 0 att det inte kommer något nytt rekord

$$\frac{99}{100} \frac{100}{101} \dots \frac{n-1}{n} \rightarrow 0, n \rightarrow \infty$$

.

En känd sats, Borel-Cantellis lemma, som ligger utanför denna kurs säger att om A_i är oberoende och $\sum P(A_i) = \infty$ så inträffar med sannolikheten 1 oändligt många av händelserna A_i . Eftersom $\sum \frac{1}{i} = \infty$ så gäller den här.

Slutligen behöver naturligtvis inte riktiga rekord typ extrema väderhändelser komma från oberoende observationer. Det kan finnas många skäl som ger beroenden, t.ex. astronomiska effekter, variationer i solens aktivitet där bl.a. en 11-årscykel är berömd och system som lagrar energi på jorden t.ex. haven och atmosfären. Sådana effekter har i allmänhet en tendens att öka chansen för serier av tätt kommande väderrekord.

4 Stokastiska variabler

Ett mätvärde eller en annan slumpmässig storhet som ger ett *tal* brukar modelleras som en stokastisk variabel. Upprepade mätningar eller andra serier av värden modelleras som flera stokastiska variabler och brukar skiljas åt med ett index. Mätningens praktiska eller fysikaliska innebörd ligger då inbakad i dess sannolikhetsfördelning, t.ex. som fördelningens mittpunkt. Som beteckning används normalt stora bokstäver i slutet av alfabetet X, Y, X_i etc. (medan bokstäverna i alfabetets början reserverats för händelser A, B, A_i). Givetvis är detta bara en praktisk konvention och som i all matematik har vi rätt att välja andra betydelser på vårt alfabet men i denna text liksom i flertalet böcker gäller den. Vi förutsätter här att observationer och variabler har *reella* värden och ibland begränsar vi oss till heltal eller andra diskreta (ändligt eller numrerbart många) värden. Vektorvärda eller komplexvärda stokastiska variabler anavänds i en del tillämpningar men behövs inte just nu.

Man kan göra utfallsrum direkt för variablerna men oftast föredrar man en generellare modell där man har ett bakomliggande utfallsrum (med händelser och sannolikheter) och de stokastiska variablerna blir då funktioner av utfallen, $X = X(\omega)$, $\omega \in \Omega$. Vi kan illustrera en ensam variabel som en avbildning (en pil) från utfallsrummet upp till en reell axel. Om man har flera stokastiska variabler i ett problem så är de alltid definierade på samma utfallsrum. Slumpen väljer ett av utfallen och alla variabler får då sina värden.

Sannolikhetsbeskrivningen för en stokastisk variabel ges i princip av sannolikheterna i utfallsrummet, men detta utnyttjas sällan. I stället flyttar man upp den till samma reella axel som där variabeln X antar sina värden (eller till R^n om man har n variabler X_1, \dots, X_n). Speciella satser finns formulerade som visar att det för varje rimlig sannolikhetsbeskrivning för stokastiska variabler existerar ett gemensamt utfallsrum med händelser och sannolikhetsmått så att modellen blir komplett.

När mätningar modelleras som stokastiska variabler så är det i början svårt att se dem på två olika sätt. Dels har vi mätningarnas observerade värden, t.ex. 2.21, 3.12, 2.87, 2.55, dels har vi deras modell X_1, X_2, X_3, X_4 där många, kanske alla, reella värden är möjliga. Den stokastiska modellen analyserar om man så vill värdena innan de blev mätta, men samma analys är intressant efteråt eftersom man inte vet det fysikaliskt sanna värdet som observationerna strävar mot. För något okänt tal μ är det kanske $2.21 - \mu, 3.12 - \mu, 2.87 - \mu, 2.55 - \mu$ som är det vi verkligen är intresserade av (som fel i observationerna) och det kan vi faktiskt studera teoretiskt med $X_1 - \mu, X_2 - \mu, X_3 - \mu, X_4 - \mu$ och få fram förvånandsvårt skarpa resultat för.

4.1 Sannolikhetsmått för variabler

Vi ger först ett generellt begrepp, fördelningsfunktionen, därefter måste vi dela in i diskreta och kontinuerliga variabler beroende på om varje värde kan ges en sannolikhet eller om vi skall integrera fram sannolikheter.

Definition 4.1 *Fördelningsfunktionen för en (reellvärd) stokastisk variabel X*

$$F(x) = P(X \leq x). \quad (14)$$

Vi använder begreppet P från den bakomliggande beskrivningen och dessutom finns i princip kravet att händelsen $A = \{\omega; X(\omega) \leq x\}$ skall vara en tillåten händelse i utfallsrummet för varje reellt x (kallas mätbarhet hos funktionen $X(\omega)$).

Några egenskaper för fördelningsfunktionen följer lätt t.ex.

$$0 \leq F(x) \leq 1;$$

$F(x)$ växande (inklusive konstant på delintervall);

$$F(x) \rightarrow 1, x \rightarrow \infty, \quad F(x) \rightarrow 0, x \rightarrow -\infty,$$

där det sista gäller om vi förutsätter att variabeln X är ändlig (med sannolikheten 1). En mer noggrann observation visar att om något enstaka värde x_0 har $P(x_0) > 0$ så får $F(x)$ ett språng av storleken $P(x_0)$ där och antar det *övre* värdet i punkten x_0 . $F(x)$ blir därför högerkontinuerlig men inte vänsterkontinuerlig vid punktsannolikheter. (Om vi använt sträng olikhet $X < x$ i definitionen av F så hade funktionen blivit vänsterkontinuerlig istället.)

De angivna egenskaperna räcker, d.v.s. *varje högerkontinuerlig funktion som växer från 0 till 1 och aldrig avtar är en möjlig fördelningsfunktion* för en stokastisk variabel som är ändlig (med sannolikheten 1).

Ur fördelningsfunktionen får vi bland annat sannolikheter för intervall som

$$\begin{aligned} P(a < X \leq b) &= F(b) - F(a), \text{ om } a < b; \\ P(a \leq X \leq b) &= F(b) - F(a) + P(X = a); \\ P(a < X < b) &= F(b) - F(a) - P(X = b). \end{aligned}$$

Fördelningsfunktioner för vissa ofta använda fördelningar finns tabellerade i böcker och programmerade i färdiga programpaket (t.ex. i matlabs statistiska toolbox), men i flera fall har man så pass snygga funktionsuttryck att de kan uttryckas med vanliga matematiska funktioner (se t.ex. radioaktivt sönderfall nedan). Just nu inför vi bara begreppet och återkommer till beräkningar och exempel.

Kärt barn har många namn och tyvärr är inte terminologin för nästa begrepp enhetlig varken i svensk eller engelsk litteratur.

Definition 4.2 *Frekvensfunktionen/sannolikhetstätheten/sannolikhetsfunktionen för en diskret stokastisk variabel X ges som*

$$f(x_i) = P(X = x_i) \quad (15)$$

för de möjliga värdena x_i . Vi kan formellt definiera $f(x) = 0$ för övriga x .

Definition 4.3 *Frekvensfunktionen/sannolikhetstätheten för en kontinuerlig stokastisk variabel X ges implicit som den funktion $f(x)$ man integrerar för att få sannolikheter. För varje integrerbart område A på reella axeln blir*

$$P(X \in A) = \int_A f(x) dx. \quad (16)$$

Frekvensfunktioner (diskreta eller kontinuerliga) är ickenegativa och måste summera eller integrera till 1. Det är det enda som krävs. Varje positiv funktion med ändlig summa/integral kan divideras med sin summa/integral och på så sätt normeras till en möjlig frekvensfunktion. Fundera gärna igenom vilka funktioner du känner till som är positiva på hela reella axeln och har ändlig integral.

Medan fördelningsfunktioner ger en direkt bild av sannolikheten att komma till vänster om en given punkt x på talaxeln, så ger frekvensfunktionen istället en bild av hur sannolika olika resultat är i förhållande till varandra. Eftersom alla resultat har sannolikheten 0 vid kontinuerlig fördelning måste detta då tolkas som sannolikheter för små och lika långa intervall omkring argumentet x .

4.1.1 Samband mellan fördelnings- och frekvensfunktion

Eftersom $F(x) = P(X \leq x)$ så måste

$$F(x) = \begin{cases} \sum_{x_i \leq x} f(x_i) \\ \int_{-\infty}^x f(x') dx' \end{cases} \quad (17)$$

i det diskreta respektive kontinuerliga fallet. Omvänt gäller om vi storleksordnar en diskret variabels värden så att $x_1 < x_2 < \dots$ att

$$f(x_i) = F(x_i) - F(x_{i-1}), \quad (18)$$

och om istället X är kontinuerlig och $F(x)$ deriverbar så blir

$$f(x) = \frac{dF}{dx}, \quad (19)$$

vilket man ser om man deriverar den kontinuerliga versionen av (??).

Anmärkning 4.4 Det kan förekomma blandade variabler som varken är rent diskreta eller rent kontinuerliga. I en modell för väntetiden i ett betjäningssystem har t.ex. tiden 0 sannolikhet eftersom kön kan vara tom när man anländer, men annars har väntetiden en kontinuerlig fördelning. Sannolikheten för händelser (intervall) får man då genom att både summera de diskreta sannolikheter som ingår och integrera en lämplig funktion som täcker resten av sannolikheten.

Exempel 4.5 Låt $x_i = i$, $i = 1, 2, 3, \dots$ vara de möjliga värdena och

$$f(x_i) = a^{i-1}(1-a)$$

där a är ett tal i intervallet $(0,1)$. Då blir

$$F(x_i) = \sum_1^i a^{j-1}(1-a) = \frac{1-a^i}{1-a}(1-a) = 1-a^i;$$

$$F(x) = 0, x < 1;$$

och

$$F(x) = F([x]) = 1 - a^{[x]}, x > 1,$$

där $[x]$ betyder heltalsdelen.

Märk att oavsett definitionsområdet för variabeln så kan fördelningsfunktionen framställas för hela reella axeln.

Exempel 4.6 Låt $f(x) = xe^{-x}$, $x > 0$, (underförstått $f = 0$, $x < 0$). Då blir

$$F(x) = \int_0^x ye^{-y} dy = [-ye^{-y}]_0^x + \int_0^x e^{-y} dy = 1 - e^{-x} - xe^{-x}.$$

Detta gäller positiva x så vi kompletterar med

$$F(x) = 0, x < 0.$$

Man kontrollerar lätt att $F(x) \rightarrow 1$, $x \rightarrow \infty$ och att dessutom $F(x) \rightarrow 0$, $x \rightarrow 0$. Funktionen är växande eftersom vi integrerat en positiv funktion (och konstant på negativa axeln eftersom vi definierat den så eller integrerat en nolla om vi så vill).

4.1.2 Enkla exempel på frekvensfunktioner och normering

Om alla värden i ett intervall är lika sannolika så bör frekvensfunktionen vara konstant på intervallet. Antag att vi drar ett tal på måfå mellan 2 och 10. Det dragna talet X tänker vi oss då har samma chans att ta vilket som helst av dessa tal. Då låter vi $f(x) = c$, $2 < x < 10$ och konstanten c bestäms av att totala sannolikheten $\int_2^{10} c dx = 8c = 1$, $c = 0.125$.

Om istället chansen för ett x i intervallet (vi kan använda samma intervall $(2,10)$) är proportionell mot x , så att stora värden har större chans än små, så blir $f(x) = Cx$, $2 < x < 10$. På nytt bestäms konstanten av att totala sannolikheten skall vara ett och vi får $\int_2^{10} f(x) dx = \int_2^{10} Cx dx = C(100 - 4)/2 = 1$, $C = 1/48$.

Det är inte alls ovanligt att man av fysikaliska eller andra skäl har sådana enkla fördelningar. Ofta kan man på grund av symmetri anse att olika riktningar är lika sannolika (t.ex. för strålningen av ljus eller partiklar från en atom) och om detta gäller i tre dimensioner så kan man övertyga sig om att projektionen

av en helt slumpmässigt riktad enhetsvektor på en given annan vektor blir sådan att varje värde mellan -1 och $+1$ får samma chans. (Resultatet är en konsekvens av att kalottytan av en boll är proportionell mot kalottens höjd i vertikal led). Likaså blir då vissa vinklar helt slumpmässiga mellan 0 och 2π eller motsvarande. Vi återkommer till detta i samband med olika sannolikhetsfördelningar i kapitel ??.

En vanlig diskret modell för partiklar utgår från att det finns ett antal möjliga tillstånd med olika energinivåer E_1, E_2, \dots, E_n och att sannolikheten att en partikel befinner sig i dessa ges av $f(k) = ce^{-\alpha E_k}$, $k = 1, \dots, n$. Om vi definierar X som numret på energinivån så blir detta en frekvensfunktion för X . För varje givet värde på α bestäms konstanten c av att sannolikheterna skall summera till 1 , men konstanten α måste skattas från data eller tas från annan fysikalisk teori.

Några fördelningar har intressanta bakomliggande härledningar som motiverar när de bör användas. Andra fördelningar saknar detta men kan ändå vara nyttiga därför att de anpassar sig väl till vissa datamaterial. Man klarar sig förvånadsvärt långt med ett 10-tal olika fördelningstyper så det är ingen ogenomtränglig massa men behöver behärska för att börja göra nyttiga statistiska analyser. I nästa avsnitt ger vi ett första exempel på hur man kan härleda fram en sannolikhetsfördelning från en grundläggande ide om konstant risk. En mer systematisk presentation av fördelningar ges i kapitel ??.

4.2 Radioaktivt sönderfall och livstidsfördelning

En radioaktiv atom sönderfaller så småningom och skickar då ut strålning som kan uppfångas i en räknare om strålningen träffar räknarens känsliga del. De inre orsakerna till själva sönderfallet är fördolda men utifrån sett är det alltid samma risk att atomen sönderfaller ända tills det verkligen inträffar. En av isotoperna av Amerikanum används i min brandvarnare och denna isotop har en halveringstid som är 475 år. Det betyder att hälften av Amerikanumatomerna har sönderfallit efter den tiden så i det fallet är det ingen jättesnabb process. För en ensam atom innebär detta att den med sannolikheten 0.5 överlever 475 år. Låt oss härleda resten av sannolikhetsfördelningen.

Låt atomen vara hel vid tiden 0 . Det måste finnas en mycket liten sannolikhet c (storleksordningen 10^{-10}) att den sönderfaller redan under den första sekunden. Om den överlever den sekunden så är (den betingade) sannolikheten att den sönderfaller i nästa sekund precis lika liten. Om vi istället tar ett tvåsekundersintervall blir sannolikheten fortfarande liten men i alla fall dubbelt så stor (egentligen $c + (1 - c)c = 2c - c^2 \approx 2c$). Litet generellare finner vi att för korta tider h blir sannolikheten att en dittills hel atom sönderfaller proportionell mot h och kan skrivas ch med ett fel av storleken $o(h)$ ³. Låt X vara atomens livstid, $F(x) = P(X \leq$

³Ordnotermen $o(h)$ definieras av att $o(h)/h \rightarrow 0$, $h \rightarrow 0$ vilket bl.a. gäller för h^2 . Symbolen brukar användas för att samla upp alla resttermer av den storleken och behöver inte betyda

x) fördelningsfunktionen för X och inför dessutom $G(x) = 1 - F(x) = P(X > x)$ som blir litet snyggare att räkna ut. (I tillförlitlighetsteorin kallas $G(x)$ ofta för överlevnadsfunktionen.) För att överleva tiden $x + h$ måste atomen först överleva tiden x och sedan klara ytterligare tiden h . Risken för en hel atom att sönderfalla är alltid lika stor vilket gör att

$$G(x + h) = P(X > x + h) = P(X > x)(1 - ch) + o(h) = G(x)(1 - ch) + o(h)$$

och det ger

$$\frac{G(x + h) - G(x)}{h} = -cG(x) + o(h)/h \rightarrow -cG, \quad h \rightarrow 0.$$

Differentialekvationen $G' = -cG$ eller $G'/G = -c$ ger

$$\frac{d}{dx} \ln(G) = -c$$

och vi löser

$$\ln(G(x)) = -cx + a;$$

$$G(x) = e^{a-cx}.$$

Eftersom atomen var hel vid tiden 0 så blir $X > 0$ med sannolikheten 1. Då skall $G(0) = 1$ vilket ger $a = 0$. Lösningen blir att

$$G(x) = e^{-cx}, \quad F(x) = 1 - e^{-cx}, \quad x > 0.$$

eller deriverat

$$f(x) = ce^{-cx}, \quad x > 0.$$

Detta är exponentialfördelningen som visar sig vara en av ämnets viktigaste fördelningar och uppträder i många olika sammanhang.

Sannolikheten att överleva halveringstiden måste vara 0.5. Detta bestämmer konstanten c . 475 år är $475 \times 365 \times 24 \times 3600 = s$ sekunder, och $G(s) = \exp(-cs) = 0.5$ ger $c = \ln(2)/s$, vilket kallas sönderfallsintensiteten. (Vi kan använda en större tidsenhet om vi vill och få behändigare tal men brandvarnarens angivna styrka ges som sönderfall per sekund så tidsenheten är kanske lämplig.) I lösningen har vi använt att $G(x) > 0$ eftersom det är en sannolikhet. Annars finns inte logaritmen som en reell funktion.

4.3 Fördelningars centrum och spridning

Efter att ha härlett den radioaktiva atomens livslängdsfördelning återgår vi till grundläggande egenskaper hos sannolikhetsfördelningar. Vi behöver något slags mittpunkt för en fördelning. I exemplet med den radioaktiva atomen kan vi fråga oss hur länge en atom i genomsnitt överlever eller vad som bör anses som

 exakt samma sak i olika led.

en typisk livslängd. Samma storheter är ju intressanta mått på bl.a. människors livstider. Det finns två särskilt naturliga sådana mått, väntevärdet och medianen. Väntevärdet ger en tyngdpunkt för fördelningen och i en livslängdsfördelning ger det medellivslängden. Där vägs nämligen varje möjligt resultat med sannolikheten att det inträffar (på samma sätt som man vid tyngdpunktsberäkning väger varje avstånd med den massa som ligger därute och dividerar med totalmassan, och massa genom totalmassa liknar sannolikhet). För matematisk statistik är en annan egenskap hos väntevärdet mycket viktigare, nämligen att det är en storhet som är nyttig när man summerar eller tar medelvärde av många variabler och vill bestämma fördelningen för resultatet.

Definition 4.7 *Väntevärdet för variabeln X ges av*

$$\mu = E[X] = \begin{cases} \sum_i x_i f(x_i), & \text{diskret} \\ \int x f(x) dx, & \text{kontinuerlig.} \end{cases} \quad (20)$$

om summan respektive integralen⁴ är absolutkonvergent ($\sum |x_i| f(x_i) < \infty$ eller motsvarande integralvillkor).

Vi inför här två olika beteckningar, den korta μ är bra i många formler medan den längre $E[X]$ blir tydlig inte minst när man stoppar in olika uttryck istället för X . För en funktion $h(X)$ kan väntevärdet erhållas som

$$E[h(X)] = \begin{cases} \sum_i h(x_i) f(x_i), & \text{diskret} \\ \int h(x) f(x) dx, & \text{kontinuerlig.} \end{cases} \quad (21)$$

om beräkningen är absolutkonvergent. Man kan visa att detta ger samma resultat som om man istället först härleder fördelningen för $h(X)$ och sedan tillämpar den första definitionen.

I de diskreta väntevärdesformlerna är det tydligt hur varje resultat vägs med sin sannolikhet och det sker också i integralen om vi tolkar $f(x)dx$ som en sannolikhet för ett litet område runt x med längden dx (kanske tydligast innan man går i limes med Riemannintegralens approximerande summor). Om vi kunde simulera nästan obegränsat många oberoende variabler med samma fördelning $f(x)$ så skulle de utfalla ungefär som sannolikhetsfördelningen och deras medelvärde blir då mycket nära μ . Man kan visa att följderna av medelvärdena för de n första variablerna konvergerar till μ med sannolikheten 1 när $n \rightarrow \infty$.

Exempel 4.8 Väntevärdet för atomens livslängd (exponentialfördelning).

Vi hade frekvensfunktionen $f(x) = ce^{-cx}$, $x > 0$ som beskrev en exponentialfördelning med intensiteten c . Härledningen gjordes för en radioaktiv atoms

⁴Vi skriver ofta inte ut integralgränserna men underförstår att integralen skall tas över hela reella axeln eller över det delområde där $f(x) > 0$.

livslängd men är giltig för alla livstider med en konstant intensitet att ta slut. Väntevärdet beräknas som

$$\mu = \int_0^{\infty} x c e^{-cx} dx = [-x e^{-cx}]_0^{\infty} + \int e^{-cx} dx = \frac{1}{c}.$$

Man ser att en hög sönderfallsintensitet eller dödsintensitet ger kort förväntad livslängd och omvänt vilket är naturligt.

Tekniska finesser: Kravet på absolutkonvergens innebär att t.ex. frekvensfunktionen $f(x) = c/(1+x^2)$, $-\infty < x < \infty$, $c = 1/\pi$, inte har något väntevärde trots att $x = 0$ är en tydlig symmetripunkt i fördelningen. Det innebär också att vi kan ha ett väntevärde för X men sakna det för $h(X)$ om h är tillräckligt stor i svansarna.

Vi kan utan problem generalisera definitionen av väntevärde till att väntevärdet $E[X]$ är $+\infty$ om summan/integralen över negativa värden är ändlig och summan/integralen över positiva värden blir oändlig. Givetvis gäller motsvarande för $-\infty$ så det enda riktigt förbjudna fallet är att en positiv och en negativ oändlighet står mot varandra.

Det är inte ofta som man får några problem med oändligheter även om det finns några intressanta sådana fall. I nästan alla situationer är väntevärden tacksamma att arbeta med eftersom de har snygga egenskaper vid linjära funktioner av variabler och många resultat grundas på sådana uttryck.

Medianen är ett okomplicerat mått i den meningen att den alltid existerar.

Definition 4.9 *Medianen är en centralpunkt som delar sannolikhetsmassan mitt itu. Om det finns flera punkter som gör detta så definieras medianen som mittvärdet av alla sådana punkter.*

Om det finns en entydig punkt m där $F(m) = 0.5$ så är m fördelningens median. Att flera punkter delar sannolikhetsmassan mitt itu kan bara förekomma om $F(x) = 0.5$ i ett intervall och då ger definitionen intervallets mittpunkt. För att täcka språng förbi värdet 0.5 i fördelningsfunktionen sätter vi i så fall m som det värde där språnget inträffar.

Ibland är det en fördel att eventuella konstigheter i fördelningens svansar inte påverkar medianen eftersom den bestäms av fördelningsfunktionen i dess centrala del. (På motsvarande sätt är fördelningen för medianen i ett datamaterial okänslig för om man har grova fel i ett antal data eftersom dessa data i så fall bör hamna i svansarna på datamaterialet.)

Visa så många som möjligt av följande påståenden

Påstående 1: Medianen och väntevärdet blir lika om fördelningen har en symmetripunkt (och väntevärdet existerar).

Påstående 2: Om alla värden i ett intervall har samma sannolikhet och värden utanför intervallet är omöjliga så är väntevärdet mitt i intervallet.

Påstående 3: Väntevärdet för ett tärningkast är 3.5.

Påstående 4: Väntevärdet för kvadraten av ett tärningkast är ...

Påstående 5: Väntevärdet för $(X - a)^2$ blir minst om $a = \mu$. (Problem 3.5).

Påstående 6: Väntevärdet för $X - \mu$ blir 0.

Påstående 7: Väntevärdet för $|X - a|$ blir minst om $a = m =$ medianen. (Problem 3.6).

När vi bestämt oss för ett lämpligt centralmått så är det dags att definiera ett motsvarande mått för spridning. Det vanligaste måttet på spridning utgår från kvadraten på avståndet till väntevärdet. För att få ett enda mått tar man medelvärdet, dvs väntevärdet, av dessa kvadratavstånd.

Definition 4.10 *Variansen för variabeln X ges av*

$$\sigma^2 = \text{Var}(X) = E[(X - \mu)^2] = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i), & \text{diskret} \\ \int (x - \mu)^2 f(x) dx, & \text{kontinuerlig.} \end{cases} \quad (22)$$

med de vanliga kraven på väntevärdets existens.

Vi inför två beteckningar, σ^2 och $\text{Var}(X)$, för att det är praktiskt i olika sammanhang att ta den lämpligaste formen.

För att underlätta tolkningen av måttet i samma skala som variabeln X kan man istället beräkna standardavvikelsen.

Definition 4.11 *Standardavvikelsen för variabeln X ges av*

$$\sigma = \sqrt{\text{Var}(X)} = \begin{cases} \sqrt{\sum_i (x_i - \mu)^2 f(x_i)}, & \text{diskret} \\ \sqrt{\int (x - \mu)^2 f(x) dx}, & \text{kontinuerlig.} \end{cases} \quad (23)$$

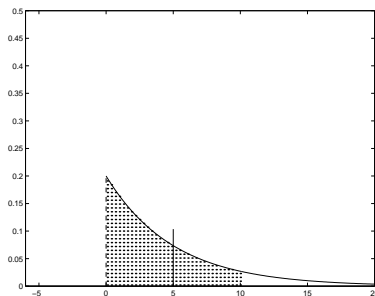


Figure 4: Frekvensfunktion, väntevärde och sannolikhet för $\mu \pm \sigma$ som punkterad yta under frekvensfunktionen för exponentialfördelning med intensitet 0.2.

Man kan skaffa sig en känsla för dessa mått genom att räkna ut dem för olika enkla fördelningar och rita in väntevärdet och väntevärdet $\pm \sigma$ tillsammans med frekvensfunktionen som i figur 4. I många situationer behöver man inte veta mer om en sannlikhetsfördelning än väntevärdet och standardavvikelsen. Måtten får senare en speciell innebörd i samband med normalfördelningar och

centrala gränsvärdesatsen där väntevärde och standardavvikelse bestämmer vilka normalfördelningar som medelvärdenas och summornas fördelningar kommer att närma sig om man har tillräckligt mycket data. Inga andra fördelningsparametrar har motsvarande generella betydelse.

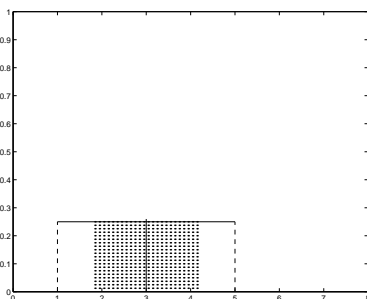


Figure 5: Frekvensfunktion, väntevärde markerat med vertikalt streck och sannolikhet för $\mu \pm \sigma$ som punkterad yta under frekvensfunktionen för likformig fördelning mellan 1 och 5.

Exempel 4.12 Varians och standardavvikelse i exponentialfördelning.

Väntevärdet i exponentialfördelningen beräknades i Exempel ?? till $1/c$, där c var intensiteten. Frekvensfunktionen var $f(x) = ce^{-cx}$, $x > 0$. Variansens definition ger då att

$$\sigma^2 = E[(X-\mu)^2] = \int_0^\infty (x-\frac{1}{c})^2 ce^{-cx} dx = [-(x-\frac{1}{c})^2 e^{-cx}]_0^\infty + \int 2(x-\frac{1}{c})e^{-cx} dx = \frac{1}{c^2},$$

där resultatet kommer ur klammerns undre gräns och den sista integralen ger $2(\mu-\mu)/c = 0$. (Både x gånger frekvensfunktionen och μ gånger frekvensfunktionen ger integralen μ .) Standardavvikelsen blir sedan

$$\sigma = \sqrt{Var(X)} = \frac{1}{c}.$$

Resultatet innebär att väntevärdet och standardavvikelsen är lika stora i exponentialfördelningen.

Exempel 4.13 Ett värde X dras på måfå mellan a och b , där $a < b$. Det innebär att alla värden i intervallet är lika sannolika och frekvensfunktionen blir $f(x) = \frac{1}{b-a}$, $a < x < b$. Väntevärdet ges av $\mu = \int_a^b x \frac{1}{b-a} dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$ som ligger mitt i intervallet. Variansen ges av

$$Var(X) = \sigma^2 = \int_a^b (x - \frac{a+b}{2})^2 \frac{1}{b-a} dx = \left[\frac{(x - \frac{a+b}{2})^3}{3(b-a)} \right]_a^b = \frac{(b-a)^2}{12},$$

och standardavvikelsen blir

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}}.$$

Som rimlighetskontroll ser vi att det beräknade väntevärdet hamnar centralt i fördelningen och att väntevärdet \pm en standardavvikelse också hamnar innanför fördelningens område och representerar en typisk avvikelse från väntevärdet. Kvadreringen gör att värden långt ut får litet större tyngd än om vi t.ex. hade tagit medelvärdet $(b - a)/4$ av absolutvärdena för avstånden till μ . Se ett exempel i figur 5.

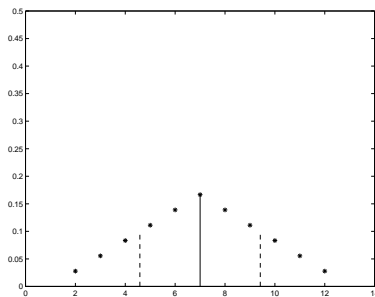


Figure 6: Diskret frekvensfunktion, väntevärde (heldraget vertikastreck) och väntevärde \pm en standardavvikelse för summan av två tärningkast.

4.3.1 Oändlig förväntad tid till återkomst i slumpvandring, kursivt

Låt $X_i = \pm 1$ med sannolikheten $1/2$ vardera (myntkast). Sätt $S_n = X_1 + X_2 + \dots + X_n$ där X_i är oberoende (nya myntkast). S_n blir slumpvandrarens position efter n steg. Låt $T = \min\{n : S_n = 0\}$ vara tiden till första återkomst till utgångspunkten. Vi skall visa att $\mu = E[T] = \infty$.

Bevis: Av symmetriskäl tar det lika lång tid att återvända från negativa sidan som från positiva sidan. Låt $\mu_k =$ väntevärdet för tiden från $\pm k$ till 0. Antag att μ_1 är ändligt. Vi skall visa att detta leder till en motsägelse. Observera att $\mu_2 = \mu_1 + \mu_1$ eftersom man först måste gå från avståndet 2 till avståndet 1 och därefter från 1 till 0 och båda stegen har samma sannolikhetsfördelning. Vi har

$$\begin{aligned}\mu &= 1 + \mu_1, \\ \mu_1 &= 1 + \frac{1}{2}0 + \frac{1}{2}\mu_2,\end{aligned}$$

eftersom man först tar ett steg och därefter i genomsnitt så många steg till som väntevärdet från den position man hamnat i. Vi får

$$\mu = 1 + \mu_1 = 1 + 1 + \frac{1}{2}(\mu_1 + \mu_1)$$

eller $1 = 2$ vilket ger motsägelse. Antagandet att $\mu_1 < \infty$ är därför felaktigt och eftersom $T > 0$ måste $\mu_1 = \infty$ och $\mu = \infty$. (Man kan bevisa att $P(T < \infty) = 1$ så man återvänder med sannolikheten 1 men den förväntade tiden är oändlig.)

En slumpvandrare behöver inte vara någon full gubbe eller så och inte heller bara en matematisk tankelek. En diffunderande partikel slumpvandrar i sitt medium och ett värmekvantum hos en molekyl gör detsamma genom att förmedla sin energi till en slumpvis utvald granne. Dessa diffusioner är tredimensionella istället för vår endimensionella beskrivning men exempelvis i en gaskromatograf är det en endimensionell projektion som utnyttjas (överlagrad på en systematisk drift). I en kontinuerlig beskrivning kan slumpvandrarens sannolikhetsfördelning och värmeledningen, diffusionen beskrivas med samma differentialekvation och sannolikhetsfördelningen blir intressantast när de enskilda partiklarna blir urskiljbara och som förklaring av resultatens "suddighet" orsakad av slumpvariation.

5 Sannolikhetsmått för flera variabler

Ofta behöver man mer än en stokastisk variabel på grund av att man antingen observerar flera olika saker eller upprepar en och samma typ av mätningar. I den stokastiska modellen representeras varje observation eller mått av en egen stokastisk variabel eftersom slumpen kan ge dem olika värden. Alla variabler som behandlas i en analys är definierade på samma utfallsrum vilket markeras av att vi kan uttrycka sannolikheter för olika händelser med ett och samma sannolikhetsmått P . I den här kursen blir variablerna nästan alltid ändligt många men i kurser i stokastiska processer måste ännu flera variabler hanteras för att kunna beskriva oändliga följder, kontinuerliga förlopp av typen resultat från ständigt registrerande instrument eller processer på ytor och i atmosfären mm.

För 2 eller flera variabler definieras fördelningsfunktionen

$$F(x_1, \dots, x_m) = P(X_1 \leq x_1, \dots, X_m \leq x_m), \quad (24)$$

där kommatecknet skall läsas som snitt dvs alla villkoren skall vara uppfyllda. Vi förutsätter att alla variablerna är ändliga med sannolikheten 1 (så att ingen sannolikhet ligger ute i $\pm\infty$)⁵.

Fördelningsfunktionen är växande (\geq) i alla sina argument, den avtar mot noll om $x_i \rightarrow -\infty$ för *något* x_i och den växer mot 1 om $x_i \rightarrow +\infty$ för *alla* x_i .

För diskreta variabler definieras frekvensfunktionen

$$f(x_1, \dots, x_m) = P(X_1 = x_1, \dots, X_m = x_m) \quad (25)$$

i alla punkter $(x_1, \dots, x_m)'$ som variabelvektorn kan anta. (Vi kan formellt definiera den som 0 i andra punkter.) Sannolikheter för händelser får man givetvis genom att summera över de punkter som ingår i händelserna.

För kontinuerliga variabler söker vi istället en funktion som kan integreras till sannolikhet och det räcker om vi kan integrera fram fördelningsfunktionens

⁵Om man dividerar två räknade värden från radioaktiva preparat så kan i princip antalet impulser bli 0 och kvoten oändlig så det finns fall som måste specialbehandlas

sannolikheter för då kan också sannolikheten för andra integrerbara områden i R^m fås med samma funktion. I alla punkter där $F(x_1, \dots, x_m)$ är deriverbar får vi då funktionen

$$f(x_1, \dots, x_m) = \frac{\partial^m F(x_1, \dots, x_m)}{\partial x_1 \dots \partial x_m}. \quad (26)$$

Eftersom en fullständigt deriverbar fördelningsfunktion uppfyller

$$F(x_1, \dots, x_m) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_m} \frac{\partial^m F(x'_1, \dots, x'_m)}{\partial x'_1 \dots \partial x'_m} dx'_1 \dots dx'_m$$

så fungerar f som vi önskar. Även om deriverbarheten inte är uppfylld i någon punkt eller i någon annan mängd som har sannolikheten 0 så fungerar f genom att man vid integrering helt enkelt kan bortse från punkter som tillsammans har sannolikheten 0 men det kräver egentligen ett mer avancerat integralbegrepp, Lebesques integral.

Om man utgår från frekvensfunktionen när man modellerar så gäller motsvarande krav som vid endimensionell fördelning: Frekvensfunktioner måste vara icke-negativa och integreras (i diskreta fallet summeras) till 1, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1$. Det är allt som krävs. Om man har en positiv funktion som ger att integralen är ändlig men inte 1 så kan man normera funktionen genom att dividera den med dess integral.

Exempel 5.1 Vid tillverkning av fordonskomponenter måste ett antal mått ligga inom vissa toleranser. En bildörr måste ju passa till motsvarande karm. I modern produktion kan man inte anpassa enskilda dörrar till enskilda karmar utan varje dörr bör fungera direkt vid monteringen och vid eventuella utbyten. Märk att detta förskjuter intresset från mått på de enskilda objekten till populationen av mått dvs till sannolikhetsfördelningen. Vektorn $X = (X_1, \dots, X_m)$ kan vara m sådana mått på en bildörr och eftersom måtten varierar något, några mm, mellan olika dörrar är det lämpligt att modellera X som stokastisk.

När man har beskrivit vektorns sannolikhetsfördelning så kan man sedan analysera den mot en vektor $Y = (Y_1, \dots, Y_m)'$, som beskriver motsvarande mått på dörrkarmen, för att bedöma risken att de inte passar ihop och för att se hur olika toleransgränser påverkar kassationsfrekvenser. Genom en bra modellering och skattning av fördelningen minskar man kraftigt behovet av provexemplar och kan teoretiskt studera olika delproblem.

Ofta är frekvensfunktionen mer användbar än fördelningsfunktionen så det gäller kanske i första hand att få tag på den. I vissa problem har vi då nytta av att först bestämma fördelningsfunktionen och sedan derivera och i andra fall kan vi få tag på frekvensfunktionen enklare genom andra resonemang, speciellt gäller det när man sätter samman oberoende variabler med kända frekvensfunktioner. I nästa avsnitt beskriver vi hur man kopplar endimensionella och flerdimensionella fördelningar.

5.1 Marginalfördelningar

Vad händer om vi sätter in oändligheten i några av fördelningsfunktionens argument? Då får vi enligt definitionen (låt oss ta fallet med 3 variabler som exempel och hänga variablerna som index på funktionerna)

$$F_{X_1, X_2, X_3}(x_1, \infty, x_3) = P(X_1 \leq x_1, X_2 \leq \infty, X_3 \leq x_3) = F_{X_1, X_3}(x_1, x_3)$$

och

$$F_{X_1, X_2, X_3}(x_1, \infty, \infty) = P(X_1 \leq x_1, X_2 \leq \infty, X_3 \leq \infty) = F_{X_1}(x_1),$$

eftersom villkoret $X_i \leq \infty$ alltid blir uppfyllt och därför kan strykas ur sannolikheten. Fördelningsfunktioner i lägre dimensioner får man tydligen genom att sätta in oändligheten för de variabler som inte skall vara med och eftersom ingen sannolikhet fick ligga i oändlighetspunkterna så kan vi också få dem genom gränsövergång när $x_i \rightarrow \infty$ för de variabler som skall bort ur fördelningsfunktionen. Om man utgår från m variabler så kallas fördelningar för färre variabler för *marginalfördelningar* och ofta avses de endimensionella fördelningarna.

Marginella frekvensfunktioner får man i det kontinuerliga fallet ur samband av typen $F_{X_1}(x_1) = F_{X_1, X_2, X_3}(x_1, \infty, \infty) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x'_1, x'_2, x'_3) dx'_3 dx'_2 dx'_1$, och derivering på x_1 ger sedan

$$f_{X_1}(x_1) = \frac{d}{dx_1} \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x'_1, x'_2, x'_3) dx'_3 dx'_2 dx'_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x'_2, x'_3) dx'_3 dx'_2. \quad (27)$$

Resultatet innebär att variabler som skall avlägsnas ur frekvensfunktionen integreras bort.

Den diskreta motsvarigheten är ännu enklare. För att få fram $f_{X_1}(x_1) = P(X_1 = x_1)$ så summerar man sannolikheten för alla sådana punkter vilket ger

$$f_{X_1}(x_1) = \sum_{x_2} \sum_{x_3} f(x_1, x_2, x_3).$$

Namnet marginalfördelning kommer egentligen från tvådimensionella tabeller över sannolikheter där man kan summera ut sannolikheterna i marginalerna och få de endimensionella fördelningarna på så sätt.

5.2 Beroende och oberoende variabler

Kom ihåg att händelser var oberoende om de inte hade någon koppling till varandra dvs om en händelse inträffade lika lätt alldeles oavsett om en annan inträffade eller ej. För oberoende händelser gällde $P(A \cap B) = P(A)P(B)$. Om det inte finns någon koppling mellan slumpvariationerna för variablerna X och Y så blir händelserna $A = \{X \leq x\}$ och $B = \{Y \leq y\}$ oberoende. Det innebär att $P(X \leq x, Y \leq y) = P(A \cap B) = P(X \leq x)P(Y \leq y)$. Vi definierar nu helt generellt variabler som oberoende om detta gäller för alla x och y .

Definition 5.2 *Två variabler X och Y är oberoende om*

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad (28)$$

och fler än två variabler är oberoende om deras fördelningsfunktion är produkten av deras endimensionella marginella fördelningsfunktioner.

Samma produktgenskap gäller för frekvensfunktionerna. I det kontinuerliga fallet följer det direkt genom att derivera (??) eller dess m -dimensionella motsvarighet och omvänt får vi tillbaka produkten av fördelningsfunktioner genom att integrera produkten av marginella frekvensfunktioner. I det diskreta fallet gäller givetvis att om variablerna är oberoende så blir $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$. Vi kan därför lika gärna definiera variablerna som oberoende om

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad (29)$$

och genom motsvarande produkt för fler än två variabler. Om funktionerna bara gäller i ett visst område så skall området tas med i produkten. Det gör att man för alla värden på x skall ha samma område för y när variablerna är oberoendet.

Den vanligaste oberoendesituationen är att man har utfört sina mätningar på ett sådant sätt att inga gemensamma slumpfaktorer kan påverka dem (åtminstone inte i någon betydande grad). Då utgår man helt enkelt från att produktgenskapen gäller och modellerar bara de endimensionella marginella fördelningarna, ofta genom sina frekvensfunktioner utan att passera fördelningsfunktionen. I några intressanta fall arbetar man med beroende variabler från början men lyckas skapa oberoende variabler som är funktioner av ursprungsvariablerna. (Dynamiska system som t.ex. uppstår i många industriella processer och reglersystem ger beroende data men i varje ögonblick finns det en oförutsägbar överraskning, "innovation", som i regel kan modelleras som oberoende och likafördelade variabler. Denna får man om man först förutsäger observationen (variabeln) på ett optimalt sätt ur gamla data och sedan tar variabelns verkliga värde minus prognosen. Det är alltså en ibland ganska komplicerad funktion av de beroende variablerna och där måste man härleda oberoendet.)

Exempel 5.3 Två bärningsföretag väntar på utryckningslarm och eftersom vi kan anse att intensiteten (risken) för ett larm är konstant blir tiderna X och Y tills de får sina larm fördelade likadant som radioaktiva atomers livslängder dvs exponentialfördelade. Om de bevakar helt olika distrikt blir tiderna till larmen oberoende med $f_X(x) = c_1 e^{-c_1 x}$ och $f_Y(y) = c_2 e^{-c_2 y}$, där intensiteterna $c_i = 1/\mu_i$. Det ger för $x > 0, y > 0$

$$f(x, y) = c_1 c_2 e^{-c_1 x - c_2 y}.$$

5.3 Väntevärde och varians för linjära uttryck

Låt X och Y vara kontinuerliga variabler med frekvensfunktionen f . Då kan väntevärdet för en funktion $Z = h(X, Y)$ av variablerna skrivas

$$E[Z] = E[h(X, Y)] = \int \int h(x, y)f(x, y)dydx, \quad (30)$$

där integreringen som vanligt är över hela planet eller över den delmängd där frekvensfunktionen inte är noll. (För att väntevärdet skall existera och vara ändligt krävs absolutkonvergens dvs samma dubbelintegral med $|h(x, y)|$ skall vara ändlig.) Man kan visa att resultatet blir samma som om man härledde frekvensfunktionen för Z och tog det vanliga väntevärdet i dess endimensionella fördelning.

Ett linjärt uttryck i variablerna har formen $Z = aX + bY + c$ där a, b, c är konstanter. Uttryckets väntevärde och varians ges av följande sats:

Sats 5.4 Låt X och Y ha väntevärden μ_x, μ_y och varianser σ_x^2, σ_y^2 . Då är

$$E[aX + bY + c] = \int \int (ax + by + c)f(x, y)dydx = a\mu_x + b\mu_y + c, \quad (31)$$

$$\text{Var}(aX + bY + c) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\text{Cov}(X, Y), \quad (32)$$

där $\text{Cov}(X, Y)$ definieras nedan. En viktig förenkling är att

$$\text{Var}(aX + bY + c) = a^2\sigma_x^2 + b^2\sigma_y^2, \quad \text{om } X, Y \text{ oberoende.} \quad (33)$$

Som specialfall blir

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \text{ vid oberoende;}$$

$$\text{Var}(X - Y) = ?, \text{ sätt } b = -1 \text{ i summaformeln;}$$

$$\text{Var}(aX) = a^2\text{Var}(X) \text{ och } \sigma_{aX} = |a|\sigma_x,$$

eftersom standardavvikelser definieras positiva men a kan vara negativ. Varianslagarna (??) och (??) är speciellt viktiga att kunna. Från dem härleder man lätt andra variansresultat för linjära uttryck.

Bevis av (??):

$$\int \int xf(x, y)dydx = \int x\left(\int f(x, y)dy\right)dx = \int xf_X(x)dx = E[X]$$

och motsvarande gäller för dubbelintegralen av $yf(x, y)$ om vi byter integrationsordningen. Konstanten c tas gånger totala integralen av frekvensfunktionen som är en etta. Motsvarande resultat visas på samma sätt för flera kontinuerliga variabler och med summor istället för integraler kommer man till samma resultat för diskreta variabler. Det betyder att väntevärdet för linjära uttryck i stokastiska variabler alltid kan fås som samma linjära uttryck i variablernas väntevärden. \square

Ett nyttigt produktresultat För oberoende variabler X och Y blir väntevärdet av *produkten* också enkel eftersom produktregeln (??) ger

$$E[XY] = \int \int xyf_X(x)f_Y(y)dydx = \int xf_X(x)dx \int yf_Y(y)dy = \mu_x\mu_y. \quad (34)$$

För oberoende diskreta variabler får vi samma resultat med hjälp av summor.

Det är lätt att se att produktregeln inte behöver gälla för beroende variabler. Låt t.ex. X ha väntevärde 0 och sätt $Y = X$ (eller $Y = X + Z$ där Z är oberoende av X) så blir $E[XY] = Var(X)$ och sambandet spricker om $Var(X) > 0$.

Bevis av varianslagarna (??), (??): Från definitionen av varians har vi att

$$\begin{aligned} Var(aX + bY + c) &= E[(aX + bY + c - a\mu_x - b\mu_y - c)^2] \\ &= E[a^2(X - \mu_x)^2 + b^2(Y - \mu_y)^2 + 2ab(X - \mu_x)(Y - \mu_y)]. \end{aligned}$$

Utvecklingen av kvadraten, där vi först fört ihop variabler och deras väntevärden, ger en summa av tre termer och vårt just visade resultat för väntevärden av summor (linjärkombinationer) gör att vi kan behandla dessa var för sig. Det ger $Var(aX+bY+c) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$, där vi infört *kovariansen*

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]. \quad (35)$$

Kovarians används ofta som ett grovt mått på hur variablerna samvarierar.

Övning: Visa att $E[X - \mu_x] = 0$. Resultatet behövs om ett par rader.

Om variablerna X och Y är oberoende så är också $X - \mu_x$ och $Y - \mu_y$ oberoende. Kolla det i fördelningsfunktionen om du inte ser det direkt. I så fall blir enligt (??) och övningen

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E[X - \mu_x]E[Y - \mu_y] = 0 \times 0 = 0. \quad (36)$$

Då följer satsens oberoenderesultat (??). □

5.4 Medelvärden och stora talens lag

Antag att vi gör flera mätningar av samma typ. Då är det naturligt att bilda deras medelvärde. Idealet är att mätningarna organiseras så att inte samma slumpfel och variationer kan påverka flera mätningar och om detta lyckas ansätter vi oberoende likafördelade variabler i modellen.

Låt därför X_1, \dots, X_n vara oberoende och likafördelade med $E[X_i] = \mu$, $Var(X_i) = \sigma^2$. Medelvärdet ges av

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

vilket är ett linjärt uttryck. Lagarna för väntevärde och varians (??), (??), generaliseras lätt till n termer och ger att

$$E[\bar{X}] = \frac{1}{n}n\mu = \mu; \quad (37)$$

$$Var(\bar{X}) = \frac{1}{n^2}Var(X_1 + \dots + X_n) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}; \quad (38)$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}. \quad (39)$$

Vi ser att osäkerheten mätt som standardavvikelse avtar som $1/\sqrt{n}$ medan väntevärdet blir oförändrat. Flera oberoende mätningar ger alltså säkrare resultat även om $1/\sqrt{n}$ avtar rätt sakta utom i början. Det krävs 4 gånger fler mätningar för att halvera osäkerheten och 100 gånger fler för att vinna en säker decimal till. Om man redan mätt ett stort antal gånger blir detta tungt men som regel finns det inget bättre att göra.

5.4.1 Stora talens lag

Först en uppskattning. Av variansens definition $\sigma^2 = E[X - \mu]^2$ ser man att om X med stor sannolikhet ligger långt från μ så kommer σ att bli stor. För ett givet värde på σ måste det därför finnas en begränsning på sannolikheten att komma långt ut.

Sats 5.5 (Chebyshevs olikhet)

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}. \quad (40)$$

De ryska bokstäverna i Chebyshev översätts med olika stavningar. Bli därför inte förvånad om det står Tjebytvov eller något liknande istället som namn på denne ryss.

Bevis: För en kontinuerlig variabel kan vi skriva

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \geq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 f(x) dx \\ &\geq k^2 \sigma^2 \left(\int_{-\infty}^{\mu - k\sigma} f(x) dx + \int_{\mu + k\sigma}^{\infty} f(x) dx \right) = k^2 \sigma^2 P(|X - \mu| > k\sigma). \end{aligned} \quad (41)$$

Olikheterna motiveras av att vi dels tar bort en del av integrationsområdet och har positiv integrand, dels ersätter integrandens parentes med sin undre gräns i det område som blir kvar. Dividera slutligen med $\sigma^2 k^2$ i första och sista led. \square

Med denna olikhet som verktyg skall vi nu ge en asymptotisk koppling mellan medelvärden och väntevärden. Resultatet kan ses som en motivering av varför man vill bilda medelvärdet av många observationer. En mer ingående asymptotisk motivering ges senare i form av den centrala gränsvärdessatsen.

Sats 5.6 (Stora talens lag) Låt \bar{X}_n vara medelvärdet av n oberoende likafördelade variabler med väntevärde μ . Då gäller

$$P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty, \quad (42)$$

för varje valt $\varepsilon > 0$.

Denna lag kallas för den svaga formen av stora talens lag. Det finns också en stark form när man har en oändlig sekvens av variabler X_i och bildar sekvensen av medelvärden $\bar{X}_n = \sum^n X_i/n$. Då gäller $P(\bar{X}_n \rightarrow \mu) = 1$, dvs med undantag av en mängd utfall som har sannolikhet 0 så konvergerar medelvärdena för alla andra utfall. Detta påstår vi inte i den svaga formuleringen och det kräver också ett mer avancerat bevis.

Bevis av stora talens lag: Vi bevisar den svaga formen av lagen för variabler med ändligt σ . Det täcker praktiskt taget alla fall av praktiskt intresse även om lagen är giltig utan krav på σ . Utnyttja att $Var(\bar{X}_n) = \sigma^2/n$ i Chebyshevs olikhet.

$$P(|\bar{X}_n - \mu| > \varepsilon) = P(|\bar{X}_n - \mu| > \frac{\varepsilon}{\sigma/\sqrt{n}} \frac{\sigma}{\sqrt{n}}) \quad (43)$$

Denna är nu på formen $|\bar{X}_n - \mu| > k\sigma_{\bar{X}_n}$ med $k = \sqrt{n}\varepsilon/\sigma$. Eftersom $k \rightarrow \infty$ när $n \rightarrow \infty$ så konvergerar begränsningen $\frac{1}{k^2} \rightarrow 0$ i Tjebytsjevs olikhet. \square

6 Några viktiga sannolikhetsfördelningar

6.1 Diskreta fördelningar

En första grupp av diskreta fördelningar grundar sig på oberoende försök med sannolikheten p att lyckas. Beroende på hur man samlar in resultaten kan antingen en binomialfördelning, en geometrisk fördelning eller en negativ binomialfördelning uppstå.

6.1.1 Bernoulli- och binomialfördelning

Om man bara gör ett försök och definierar $X = 1$ eller 0 beroende på om försöket lyckas eller ej så blir

$$X = \begin{cases} 1 & \text{med sannolikheten } p, \\ 0 & \text{med sannolikheten } 1 - p; \end{cases} \quad f(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases}.$$

Detta är en Bernoullifördelning med sannolikheten p som parameter (eller ett specialfall av binomialfördelningen). Situationen kan uppstå på oerhört många sätt. Kasta en tärning och se om ett visst resultat kommer upp (och sätt 1 eller 0 beroende på om det lyckades eller ej) gör en mätning och se om den kommer över eller under en angiven nivå etc. Trots att fördelningen är banalt enkel så kan det vara bra att beräkna variabelns väntevärde och varians.

$$E[X] = 1 \times p + 0 \times (1 - p) = p,$$

$$\text{Var}(X) = E[(X - p)^2] = (1 - p)^2 p + (0 - p)^2 (1 - p) = (1 - p)p[1 - p + p] = (1 - p)p.$$

Antag att vi istället från början bestämmer oss för att göra n oberoende försök där vart och ett lyckas med sannolikheten p . Definiera X som antalet försök som lyckas. Då är X en binomialfördelad variabel som kan anta heltalsvärdena $0, 1, \dots, n$. Vad blir sannolikheten för de olika resultaten?

Sannolikheten att de k första försöken lyckas och att sedan alla följande misslyckas är en produkt $pp \dots p(1 - p) \dots (1 - p) = p^k(1 - p)^{n - k}$ eftersom försöken är oberoende. Samma sannolikhet har varje mönster med k lyckade försök. Antalet möjliga sätt att få k lyckade på n försök är samma som antalet sätt att dra k positioner bland n (dragning utan avseende på ordningen och utan återläggning). Det härledde vi tidigare till $\binom{n}{k}$. Alltså blir

$$f(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}, \quad k = 0, 1, \dots, n. \quad (44)$$

Variabelns sannolikhetsfördelning bestäms av 2 parametrar, n och p och vi skriver $Bi(n, p)$ respektive $X \in Bi(n, p)$ som beteckning och för att ange vilken fördelning X har.

Väntevärdet och variansen kan beräknas direkt ur definitionen men en annan metod är elegantare. Låt $X_i = 1$ om det i te försöket lyckas och annars noll. Då är $X = X_1 + X_2 + \dots + X_n$ där X_i är oberoende Bernoullivariabler med väntevärde p och varians $p(1 - p)$ enligt ovan. Eftersom väntevärdena alltid kan summeras och varianserna kan summeras vid oberoende så blir

$$E[X] = np; \quad \text{Var}(X) = np(1 - p). \quad (45)$$

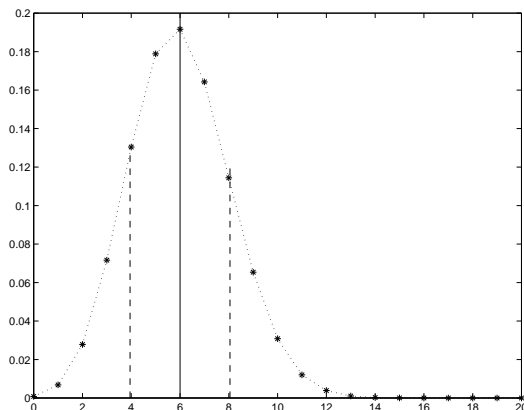


Figure 7: Frekvensfunktion, med inritat väntevärde och väntevärde \pm en standardavvikelse för binomialfördelning $Bi(20, 0.2)$.

6.1.2 Geometrisk fördelning och negativ binomial

Om vi inte fixerar antalet försök från början utan bestämmer oss för att göra oberoende försök med samma sannolikhet p tills vi lyckas så kan vi räkna antalet försök som går åt och kalla det för Y . Variabeln Y har då en geometrisk fördelning som vi kort betecknar $Ge(p)$ och som styrs av en enda parameter nämligen sannolikheten p . Det är lätt att härleda frekvensfunktionen

$$f(k) = P(Y = k) = (1 - p)^{k-1}p, \quad k = 1, \dots, n, \quad (46)$$

eftersom vi först måste få $k - 1$ misslyckade och sedan ett lyckat försök om Y skall bli k . Kontrollera gärna att summan av sannolikheterna blir 1.

För varje positivt heltal k kan vi också uttrycka fördelningsfunktionen.

$$F(k) = P(Y \leq k) = 1 - P(Y > k).$$

Sannolikheten att $Y > k$ innebär att alla de k första försöken misslyckas och det ger $(1 - p)^k$. Alltså är $F(k) = 1 - (1 - p)^k$. Pröva som övning att komma till samma resultat genom direkt summering av frekvensfunktionen. Med hjälp av högerkontinuiteten bestäms fördelningsfunktionen $F(x)$ för hela talaxeln.

Försök visa att

$$\mu = 1/p, \quad \sigma^2 = \frac{1-p}{p^2}.$$

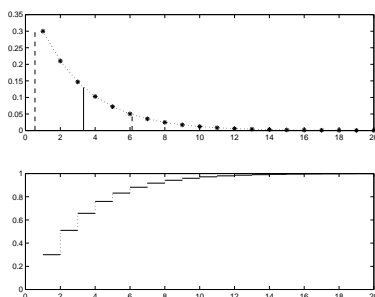


Figure 8: Frekvensfunktion för geometrisk fördelning med $p = 0.3$ överst och motsvarande fördelningsfunktion nederst, Väntevärde och väntevärde \pm en standardavvikelse har markerats.

Den negativa binomialfördelningen uppstår om man istället räknar antalet försök tills man får r lyckade, $r > 1$. Låt oss kalla antalet för Z . Fortfarande är försöken oberoende med sannolikheten p att lyckas. För att få r lyckade så måste man först hålla på tills man får ett lyckat, därefter startar man på sätt och vis om och inväntar nästa lyckade och detta upprepas exakt r gånger. Det innebär att $Z = Y_1 + Y_2 + \dots + Y_r$ där Y_i är oberoende och geometriskt fördelade. Väntevärde och varians kan vi därför plocka ur den geometriska fördelningen och summera.

Sannolikheterna får tas med ett annat resonemang. Om $k \geq r$ så måste följande inträffa om $Z = k$.

Försök k måste lyckas. Sannolikheten för det är p .

På de $k-1$ första försöken måste man dessutom ha $r-1$ lyckade. Sannolikheten för det senare beskrivs av binomialfördelningen eftersom man på $n = k-1$ försök skall ha ett givet antal lyckade. Det ger

$$f_Z(k) = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} \times p, \quad k = r, r+1, \dots$$

(Ibland definieras istället den negativa binomialfördelningen för $Z-r$ vilket betyder antalet misslyckade som måste till för att få r lyckade. Samma sannolikheter flyttas då ner till värdena $0, 1, 2, \dots$). Om $U = Z-r$ blir då

$$f_U(k) = \binom{k+r-1}{r-1} p^{r-1} (1-p)^k \times p, \quad k = 0, 1, \dots$$

Samma typ av fördelning men generaliserad så r inte nödvändigtvis är ett heltal kan också fås på andra sätt, t.ex. om man har en slumpmässig parameter i

en Poissonfördelning (se i nästa avsnitt hur Poissonfördelningen uppkommer och definieras). Fördelningen har av det skälet ofta använts när man blandar data från olyckor eller andra Poissonfördelade antal där inte olika data har exakt samma Poissonfördelning.

6.1.3 Radioaktivt preparat – antal sönderfall – Poissonfördelning

Den sista diskreta fördelningen vi tar upp i detta avsnitt är oerhört användbar. Vi härleder den för en fysikalisk situation (radioaktivt sönderfall) där vi nu inte ser på en ensam atom utan på ett helt preparat med sådana atomer. Sedan generaliseras resonemanget till helt andra situationer med samma sannolikhetsegenskaper.

Ett radioaktivt preparat har ett stort antal atomer (kanske långt över 10^{10} stycken). Tillsammans ger de en strålningsintensitet för hela preparatet. Till skillnad från en ensam atom så fortsätter preparatet att ha intensitet så länge det finns radioaktiva atomer kvar, och även om styrkan avklingar efter hand som sönderfallen sker så kan vi anse den konstant under tider som är korta jämfört med halveringstiden för preparatet. Har man 10^{10} stycken så betyder det ingenting för intensiteten om någon miljon av dem sönderfaller.

Kursivt: Intensiteten kan definieras som det förväntade antalet sönderfall per tidsenhet och man kan lätt övertyga sig om att detta blir (det förväntade) antalet radioaktiva atomer gånger sönderfallsintensiteten för en atom. Eftersom vi tidigare beräknat att en atom överlever tiden x med sannolikheten $G(x) = e^{-cx}$ så får preparatet intensiteten $A_0 e^{-cx} c$ efter tiden x . Här är A_0 antalet atomer när man börjar och c är atomernas sönderfallsintensitet. (Atomerna kan anses sönderfalla oberoende när vi inte har nukleära kedjereaktioner – för kedjereaktioner finns det istället andra stokastiska modeller i form av förgreningsprocesser.)

För det alfastrålande preparatet Am-241 i brandvarnaren tar det ca 475 år (halveringstiden) innan hälften av atomerna sönderfallit och efter 7 år har bara ca 1% av atomerna hunnit förbrukas. Det gör att intensiteten kan anses konstant åtminstone under så lång tid, vilket täcker de flesta mätsituationer.

Fördelningen för antalet sönderfall

Låt ν vara preparatets (konstanta) intensitet och låt t vara tiden sedan vi började studera preparatet. Låt dessutom

$$N(t) = \text{antalet sönderfall fram till tiden } t.$$

Vi är framför allt intresserade av sannolikhetsfördelningen för $N(t)$ och betecknar den

$$p_k(t) = P(N(t) = k).$$

Vid tiden 0 är $N(0) = 0$ och vi får därför kända startvillkor för sannolikheterna,

$$p_0(0) = 1, p_k(0) = 0, k \geq 1.$$

Även om preparatet strålar häftigt och tusentals atomer sönderfaller varje sekund så kan vi välja ett så kort tidsintervall (av längden h) att vi troligen inte får något sönderfall alls i intervallet eller möjligen ett enda. För så små intervall blir sannolikheten att få ett sönderfall inom intervallet proportionell mot h .⁶

Mellan tiden t och $t + h$ inträffar

0 sönderfall med sannolikheten $1 - \nu h$,

1 sönderfall med sannolikheten νh ,

2 eller fler med sannolikheten 0;

med ett fel av storleken $o(h)$ (där $o(h)/h \rightarrow 0$, $h \rightarrow 0$). Jämför vi $N(t + h)$ med $N(t)$ så blir de antingen lika eller också ett högre om inte en $o(h)$ -händelse inträffar. Det ger

$$p_0(t + h) = p_0(t)(1 - \nu h) + o(h); \quad (47)$$

$$p_k(t + h) = p_{k-1}(t)\nu h + p_k(t)(1 - \nu h) + o(h), \quad k \geq 1. \quad (48)$$

Detta kan vi forma om till differentialekvationer

$$\frac{dp_0}{dt} = \lim_{h \rightarrow 0} \frac{p_0(t + h) - p_0(t)}{h} = \lim_{h \rightarrow 0} -\nu p_0(t) + \frac{o(h)}{h} = -\nu p_0(t); \quad (49)$$

$$\frac{dp_k}{dt} = \text{analogt} = \nu p_{k-1} - \nu p_k(t), \quad k \geq 1. \quad (50)$$

Ekvationerna kan lösas på flera sätt, t.ex. följande

$$\frac{1}{p_0} \frac{dp_0}{dt} = -\nu; \quad \frac{d \ln p_0}{dt} = -\nu;$$

$$\ln p_0 = -\nu t + C;$$

där C är en integrationskonstant. Startvillkoret $p_0(0) = 1$, $p_k(0) = 0$, $k \geq 1$, ger $C = 0$ och

$$p_0(t) = e^{-\nu t}.$$

Multipluera ekvationen för $k > 0$ med faktorn $e^{\nu t}$ så får vi följande lösning:

$$e^{\nu t} \frac{dp_k}{dt} + e^{\nu t} \nu p_k(t) = e^{\nu t} \nu p_{k-1}(t);$$

⁶Se kursiv text längre fram som styrker proportionaliteten, men den är också rätt intuitiv. Jämfört med en ensam atom måste vi göra intervallet mycket kortare, men det betyder inget eftersom vi ändå vill ha limes när intervallet går mot 0.

$$\frac{d}{dt}(e^{\nu t} p_k(t)) = e^{\nu t} \nu p_{k-1}(t);$$

$$e^{\nu t} p_k(t) = \int e^{\nu t} \nu p_{k-1}(t) dt;$$

$k = 1$:

$$e^{\nu t} p_1(t) = \int e^{\nu t} \nu e^{-\nu t} dt = \nu t + C.$$

Startvillkoret $p_1(0) = 0$ ger $C = 0$ och vi får

$$p_1(t) = \nu t e^{-\nu t}.$$

$k = 2$:

$$e^{\nu t} p_2(t) = \int e^{\nu t} \nu \nu t e^{-\nu t} dt = \frac{\nu^2 t^2}{2} + C;$$

och vi får på nytt $C = 0$ pga startvillkoret. Det ger

$$P_2(t) = \frac{(\nu t)^2}{2} e^{-\nu t}.$$

På samma sätt fortsätter vi och nystar upp sannolikheterna och får

$$p_k(t) = \frac{(\nu t)^k}{k!} e^{-\nu t}. \quad (51)$$

Detta är en Poissonfördelning som passar in på många olika fenomen där man kan se en (konstant) intensitet. Detta återkommer vi till. Prova gärna att genomföra lösningen ovan i fallet $k = 3$ eller gör ett induktionsbevis. En litet tråkigare koll är att sätta in den föreslagna lösningen i differkvationen och verifiera att den stämmer.

6.1.4 Poissonfördelning

Vi har kommit fram till den berömda Poissonfördelningen med parameter $\lambda = \nu t$.

Definition 6.1 *En variabel X som antar värdena $0, 1, 2, \dots$ med sannolikheterna*

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (52)$$

för något positivt parametervärde λ sägs vara Poissonfördelad, $X \in Po(\lambda)$.

Eftersom fördelningen bara har ett parametervärde så måste dess väntevärde och varians bero på parametern, d.v.s. båda är funktioner av λ . Vi beräknar dem på följande sätt

$$\mu = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \dots = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda, \quad (53)$$

eftersom den sista summan kan skrivas som $\sum_0^\infty \frac{\lambda^r}{r!}$ vilket är serieutvecklingen av e^λ .

Variansberäkningen förenklas en aning om vi utnyttjar att $\sigma^2 = E[(X - \mu)^2] = E[X^2] - \mu^2$.

$$E[X^2] = \sum_0^\infty k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_0^\infty (k(k-1) + k) \frac{\lambda^k}{k!} e^{-\lambda} = \sum_2^\infty \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \lambda = \lambda^2 + \lambda.$$

Vi utnyttjade väntevärdesumman (??) och motsvarande omskrivningar som där efter att ha brutit ut λ^2 . Vi får följande enkla variansuttryck

$$\sigma^2 = Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad (54)$$

Varians och väntevärde är alltså lika stora och standardavvikelsen blir då $\sigma = \sqrt{\lambda} = \sqrt{\mu}$. Om man förväntas räkna många sönderfall (eller andra poissonfördelade händelser) så blir därför osäkerheten (mätt som standardavvikelse) liten jämfört med μ men vid mycket små väntevärden blir osäkerheten betydande jämfört med väntevärdet.

Antalet trafikolyckor under ett år är nära Poissonfördelat och det gäller också för en given klass av trafikolyckor, t.ex. dödsolyckor. Under 1998 dog ca 540 i trafiken och 1999 var det mycket pressdiskussioner om att riskerna ökat eftersom det dog 570 stycken. Nu dör det ibland flera stycken i samma olycka vilket ökar variansen mer än det ökar väntevärdet, men även om vi tänker oss att varje dödsolycka har ett offer och att $\mu = \lambda \approx 540$ så blir $\sigma \approx 23$ vilket visar att skillnaden mellan dödsantalen är obetydligt över en standardavvikelse (och troligen inte ens det om man korrigerar för flera offer i samma olycka). Nu kan det finnas andra goda skäl att inte acceptera antalet fatala trafikolyckor men underlaget i form av antalet dödade ger ingen saklig grund för påståendet att faran ökat i trafiken.

6.1.5 Poissonprocessen

Själva tidsförloppet $N(t)$ är en Poissonprocess, vilket är den kanske mest grundläggande stokastiska processen i sannolikhetsläran. Det finns en stor och spännande teori för stokastiska processer som ger modeller för olika slumpmässiga tidsförlopp och modeller för stokastiska fält och punkthändelser i rum och rum-tid.

Ser man på det radioaktiva preparatets egenskaper så bör processen $N(t)$ (med konstant intensitet) ha följande egenskaper.

- 1°. $E[N(t)] = \nu t$, intensitet gånger tid;
- 2°. Oberoende antal impulser i olika (disjunkta) tidsintervall;
- 3°. På tiden t blir antalet händelser (sönderfall) Poissonfördelat med parameter $\lambda = \nu t$, oavsett var man börjar (så länge intensiteten är konstant).
- 4°. Dessutom kan vi se tillbaka på härledningen av en atoms livslängdsfördelning, där vi också hade konstant intensitet, och finna att tiden mellan sönderfall bör vara

exponentialfördelad med en intensiteten ν och med $f(x) = \nu e^{-\nu x}$. Det blir alltså fördelningen för tiden mellan ökningarna i processen $N(t)$. Om vi har A_0 oberoende atomer med intensiteten c vardera så blir preparatets intensitet $\nu = A_0 c$.

5°. Om vi *mäter* strålningen från vårt preparat så kommer bara den strålning som träffar mätinstrumentet att registreras. Det innebär att vi uppfattar bara strålningen i en viss rymdvinkel. Eftersom strålningen sker i helt slumpmässiga riktningar så uppfattar instrumentet varje sönderfall med en sannolikhet p som beror på instrumentets andel av totala rymdvinkeln⁷, och detta sker oberoende för varje sönderfall. Det ger också en poissonprocess med intensiteten νp .

Anmärkning 6.2 I punkt 2 utnyttjar vi den stora mängden atomer som gör att även om det finns ett stort antal sönderfall i ett intervall så finns det ändå praktiskt sett lika många radioaktiva atomer kvar. Rent principiellt ger annars preparatet ett extremt svagt negativt beroende, så att mycket i det ena intervallet tenderar att ge ett lägre antal i det andra eftersom det finns färre atomer att "smälla" där. Modellen $N(t)$ och differentialekvationerna bör ge (och ger) exakt oberoende eftersom de baseras på konstant ν .

Vad finns det för andra fenomen som också ger intensiteter av liknande slag? Samtal till en telefonväxel, anrop till internet, larm till brandstationen, olyckor i trafiken, strömavbrott (diskuterades redan i kapitel 2), kundankomster till butiken, materialfel i produktionen etc. Alla dessa har en ständig intensitet att inträffa och intensiteten påverkas knappast av vad som redan har inträffat. Däremot är inte alltid intensiteten lika stor. Under begränsade tidsintervall stämmer därför den härledning vi gjort ovan och om man vill täcka varierande intensitet så finns en generaliserad lösning som också leder till Poissonfördelningen (om intensitetsfunktionen kan anses deterministisk). Generaliseringen får man genom att antingen definiera om tidsbegreppet så att intensiteten blir konstant (tiden får gå snabbare när intensiteten är hög) eller genom att inse att parametern νt egentligen skall tolkas som integrerad intensitet (i det radioaktiva exemplet $\nu t = \int_0^t \nu dt$), och att den tolkningen lätt generaliseras till tidsvarierande intensitet.

Kursivt: Radioaktiva atomers livslängd var exponentialfördelad. Exakt gäller att sannolikheten för att en ensam atom sönderfaller mellan t och $t + h$ blir

$$\begin{aligned} \int_t^{t+h} \nu e^{-\nu t} dt &= e^{-\nu t} - e^{-\nu(t+h)} = e^{-\nu t}(1 - e^{-\nu h}) \\ &= e^{-\nu t}(\nu h - \frac{(\nu h)^2}{2!} + \frac{(\nu h)^3}{3!}) = e^{-\nu t}(\nu h + o(h)). \end{aligned}$$

Faktorn $e^{-\nu t}$ är nära ett om t är mycket mindre än halveringstiden (t.ex. om vi har kortare tid än 7 år vid Am-241) så den approximerar vi med en etta. Låt $p = \nu h + o(h)$

⁷ p är instrumentets andel av den totala rymdvinkeln om varje sönderfall ger upphov till en partikel som kan registreras och riktningen är slumpmässig, men den blir större om fler registrerbara partiklar sänds ut. Varje sönderfall ger högst en räknad impuls pga instrumentets tröghet.

vara sannolikheten för sönderfall för *en* atom. För A_0 atomer blir antalet sönderfall $U = N(t+h) - N(t)$ en variabel som är $Bi(A_0, p)$ och vi väljer h så litet att det förväntade antalet A_0p också blir litet. Då är $P(U = 0) = (1 - p)^{A_0} = 1 - A_0p + \binom{A_0}{2}p^2 + \dots = 1 - A_0p + o(A_0p)$, där A_0p liten gör resttermen liten. $P(U = 1) = A_0p(1 - p)^{A_0-1} = A_0p(1 - (A_0 - 1)p + o(A_0p)) = A_0p + o(A_0p)$. Övriga sannolikheter måste då rymmas inom resttermen $o(A_0p)$ eftersom summan av de båda första blir $1 + o(h)$, så för korta tidsintervall ($h \rightarrow 0$) behöver vi bara räkna med 0 eller ett sönderfall.

Uppskattningarna kan förenklas om man utnyttjar det matematiska resultatet $(1 + \frac{x}{n})^n \rightarrow e^x$, $n \rightarrow \infty$ till approximationen $(1 - p)^{A_0} \approx e^{-A_0p}$. Uppskattningen förenklas ännu mer om man vet att binomialfördelningen för små p kan approximeras med en Poissonfördelning. Jag har inte förutsatt att något av dessa resultat är kända i detta skede för läsaren.

Övning: Visa genom direkt uträkning att om $N_1 \in Po(\lambda_1)$ och $N_2 \in Po(\lambda_2)$ är oberoende så får summan $N_1 + N_2$ fördelningen $Po(\lambda_1 + \lambda_2)$. Detta är nödvändigt om det skall gå att klyva ett intervall i Poissonprocessen i två delar med oberoende Poissonfördelade antal i varje, vilket vi förmodade i punkterna 2^o, 3^o.

Övning: Vad blir den betingade fördelningen för N_1 givet att $N_1 + N_2 = n$ om $N_1 \in Po(\lambda_1)$ och $N_2 \in Po(\lambda_2)$ är oberoende. Bestäm $P(N_1 = k | N_1 + N_2 = n)$.

Övning: Visa att om N_0 är $Po(\lambda)$ och varje räknad händelse i N_0 oberoende och med sannolikheten p räknas i N_1 så blir den senare variabeln poissonfördelad med parameter λp . Nödvändigt om den räknade radioaktiviteten skall vara en poissonprocess när sönderfallet är det.

6.2 Kontinuerliga fördelningar

Stokastiska variabler som kan anta alla värden på reella axeln eller inom ett eller flera delintervall av den reella axeln kallas kontinuerliga förutsatt att varje enskild punkt har sannolikheten noll. (Om det finns enstaka punkter med positiv sannolikhet men variabeln i övrigt kan anta alla värden på (delintervall av) reella axeln så är fördelningen av blandad typ.) Typiskt för kontinuerliga sannolikhetsfördelningar är att sannolikheter integreras fram ur frekvensfunktionen om de inte kan läsas av direkt ur fördelningsfunktionen.

6.2.1 Likformig fördelning – Rektangelfördelning

Låt a, b , $a < b$, vara två tal och antag att en variabel X kan anta alla värden i (a, b) med samma chans. Detta uttrycker vi ofta som att man drar ett tal på måfå mellan a och b . Vi får då en likformig fördelning vars frekvensfunktion blir konstant på intervallet. Fördelningen kallas lika ofta rektangelfördelning på grund av att en figur av frekvensfunktionen ser ut som en rektangel med basen (a, b) . Beteckning: $X \in Re(a, b)$. Variabeln X är $Re(a, b)$ om

$$f_X(x) = \frac{1}{b - a}, \quad a < x < b \quad (55)$$

och är noll utanför (a, b) . Man ser lätt att totala sannolikheten blir 1 och att fördelningsfunktionen blir $F(x) = \int_a^x \frac{dx}{b-a} = \frac{x-a}{b-a}$, om $a < x < b$. Det ger

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a} & \text{om } a < x < b \\ 1 & \text{om } x > b \end{cases}$$

Väntevärdet beräknas som

$$\mu = \int_a^b x \frac{dx}{b-a} = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

och ligger helt logiskt mitt i intervallet. Variansen beräknas bekvämast som $E[X^2] - \mu^2$ och $E[X^2] = \int_a^b x^2 \frac{dx}{b-a} = \frac{b^3 - a^3}{3(b-a)}$ ger

$$\text{Var}(X) = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Slumptal i datorer är vanligtvis likformigt fördelade mellan 0 och 1. Genom att bilda rätt funktioner av sådana slumptal kan sedan alla andra fördelningar skapas. Se t.ex. kapitel 7 längre fram eller kapitel 10 i *Statistisk slutledning ...*, eller den mer omfattande texten i *Lewis-Orav: Simulation methods for statisticians, operation analysts and engineers*, Wadsworth & Brooks/Cole (1989).

Slumpmässiga riktningar är rätt vanliga bland annat i fysik. I två dimensioner kan man då definiera en vinkel V som är $Re(-\pi, \pi)$. Fragmenten från radioaktiva sönderfall kan få slumpmässiga riktningar och om man i något sammanhang vill visa att riktningarna inte är helt slumpmässiga utan att vissa är mer sannolika så måste man först räkna ut sannolikhetsfördelningen för vad data från helt slumpmässiga riktningar skulle kunna ge.

Exempel 6.3 Slumpmässiga riktningar i tre dimensioner. Slumpmässiga riktningar i tre dimensioner hanteras med en likformig sannolikhetsfördelning över ytan av en enhetsfär med centrum i origo. Sannolikheten att riktningen tillhör en riktningsmängd kan då ges som arean av motsvarande yta på enhetsfären dividerad med sfärens totala area. (Detta förutsätter att riktningsmängden är mätbar, så att arean kan bestämmas.) Om man arbetar med polära koordinater och vinklarna θ och ϕ så kan koordinaterna för en punkt (x, y, z) på enhetsfärens yta skrivas

$$\begin{aligned} x &= \cos \phi \cos \theta \\ y &= \cos \phi \sin \theta \\ z &= \sin \phi \end{aligned}$$

En likformig fördelning över ytan erhålls när θ och ϕ är oberoende, θ är likformigt fördelad över en sträcka av längden 2π , t.ex. $\theta \in Re(-\pi, \pi)$ och ϕ har frekvensfunktionen $f_\phi(u) = \frac{1}{2} \cos u$, $-\frac{\pi}{2} < u < \frac{\pi}{2}$.

Vinkeln ϕ är då definierad som vinkeln mellan Ortsvektorn (x, y, z) och ekvatorialplanet och räknad positiv i norra halvklotet med en bild från vårt jordklot. Skälet till att ϕ inte är likformig är att vägen runt ekvatorn är längre än vägen

runt jorden på en mer nordlig eller sydlig latitud och väglängden är proportionell mot $\cos \phi$. Ett vinkelområde $\phi \in (u, u + \delta)$ tar därför upp areor som motsvarar δ gånger vägen runt sfären på olika sådana latituder.

Om vi slumpar ut vinklarna så blir de tre koordinaterna (X, Y, Z) för motsvarande punkt på enhetssfären tre beroende variabler som knyts ihop av villkoret $X^2 + Y^2 + Z^2 = 1$. Av symmetriskäl har de samma marginella fördelningar och det är lättast att beräkna den för variabeln $Z = \sin \phi$.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(\sin \phi \leq z) = P(\phi \leq \arcsin z) \\ &= \int_{-\pi/2}^{\arcsin z} \frac{1}{2} \cos u \, du = \frac{1}{2}(\sin(\arcsin z) + 1) = \frac{z+1}{2}, \quad -1 < z < 1. \end{aligned}$$

Fördelningsfunktionen växer linjärt med z och frekvensfunktionen blir $f(z) = \frac{dF}{dz} = \frac{1}{2}$, $-1 < z < 1$. Detta känner vi igen som en likformig fördelning, $Re(-1, 1)$. Resultatet är inte alls självklart innan man har räknat igenom det men motsvarande geometriska resultat är att arean mellan två latituder alltid är proportionell mot skillnaden mellan de båda latitudernas z -koordinater vilket man kan kontrollera genom integrering. Sammanfattningsvis innebär en likformig fördelning över sfärens yta att fördelningen för en av vinklarna i den polära framställningen och de tre marginella fördelningarna för ortens koordinater är likformiga.

Simulering av slumpmässig riktning: I tre dimensioner kan vi simulera vinklarna θ och ϕ enligt de givna frekvensfunktionerna. Ett bra sätt att simulera slumpmässiga riktningar i en godtycklig dimension d (även $d = 3$) är att först dra en vektor med d oberoende $N(0, 1)$ -fördelade variabler och normera den genom att dividera med roten ur kvadratsumman. I Matlab fungerar t.ex. kommandona $u = randn(d, 1)$; $u = u/\text{sqrt}(u' \times u)$; Det vi erhåller är då koordinaterna ut till punkten och det är bara när $d = 3$ som vi kan förvänta oss likformiga marginella fördelningar för dessa.

Likformig fördelning kan också definieras på andra ändliga områden i högre dimensioner, t.ex. på olika tvådimensionella ytor, i volymen av en kropp etc. Typiskt är då att frekvensfunktionen är konstant i det aktuella området och är noll utanför. Om man byter koordinatsystem och inte har linjärt samband mellan koordinatsystemen så kan man normalt vänta sig att likformighet i ett koordinatsystem inte motsvaras av likformighet i det andra.

6.2.2 Exponentialfördelning

Denna behandlades i avsnittet om radioaktiva atomens sönderfall där vi hade en situation med konstant risk (intensitet) för sönderfall och beskrev fördelningen för tiden till sönderfallet. Fördelningen passar in på en mängd situationer med konstant risk. Man kan antingen parametrisera fördelningen med intensiteten c eller med väntevärdet $\mu = 1/c$. Om man sätter in siffror på parametern är det därför viktigt att ange om siffran gäller c eller μ . X är exponentialfördelad med

intensiteten c om

$$f(x) = ce^{-cx}, \quad x > 0, \quad F(x) = 1 - e^{-cx}, \quad x > 0.$$

Med parametreringen μ gäller istället

$$f(x) = \frac{1}{\mu}e^{-x/\mu}, \quad x > 0, \quad F(x) = 1 - e^{-x/\mu}, \quad x > 0.$$

Vi fick tidigare $\mu = 1/c$, $\sigma = 1/c$ ($= \mu$). Många elektroniska komponenter har med god approximation en exponentialfördelad livslängd. En intressant egenskap är i så fall att begagnade men fungerande sådana komponenter har en kvarstående livstid som är likadan som den ursprungliga. För en komponent med den stokastiska livslängden X som fortfarande håller efter tiden s gäller nämligen $P(X > s+t | X > s) = e^{-c(s+t)}/e^{-cs} = e^{-ct}$, vilket är sannolikheten för en ny enhet att hålla tiden t . (Att komponenten fortfarande är hel vid s blir en information om att $X > s$ som vi skall betinga på.)

6.2.3 Gammafördelning

Gammafördelningen är ofta användbar för att beskriva positiva variabler. Fördelningen har 2 parametrar och täcker bland annat exponentialfördelningen som ett specialfall. Den kan uppstå ur andra fördelningar på några olika sätt. Om man adderar oberoende exponentialfördelade variabler med samma intensitet c så blir summan gammafördelad och om man adderar kvadraterna på oberoende normalfördelade variabler med samma varians och väntevärde 0 så blir också resultatet gammafördelat.

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$
$$\mu = \frac{\alpha}{\beta}, \quad \sigma = \frac{\sqrt{\alpha}}{\beta}.$$

Tyvärr är det lika vanligt att fördelningen anges med β utbytt mot $1/\beta$ så även här får man vara försiktig och definiera vilken parameter som avses.

6.2.4 Normalfördelning

Normalfördelningen är den mest studerade fördelningen i statistiska sammanhang. Den motiveras ofta genom centrala gränsvärdessatsen (se längre fram) i situationer där slumpvariationen kan ses som summan av många "små" slumpmässiga bidrag. Det är en ofta använd fördelning för mätfel och för naturliga fysikaliska variationer vid makroskopiska mätningar där resultatet ofta kan ses just som summor (medelvärden) av många bidrag. Den kan också uppstå genom att man gör många oberoende observationer på något som inte alls är normalfördelat och bildar summan/medelvärdet av dessa. Centrala gränsvärdessatsen som vi bevisar längre fram ger att resultatet får en fördelning som approximeras bra med normalfördelningen.

Normalfördelningens frekvensfunktion:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty. \quad (56)$$

Som parametrar har vi direkt satt in väntevärdet μ och standardavvikelsen σ och för en variabel X med denna fördelning skriver vi att

$$X \text{ är } N(\mu, \sigma).^8$$

Vi verifierar längre fram att det verkligen är en frekvensfunktion med avsett väntevärde och standardavvikelse.

Standardiserad normalfördelning:

Om $\mu = 0$ och $\sigma = 1$ har vi den standardiserade normalfördelningen $N(0, 1)$. Dess frekvensfunktion betecknar vi

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty. \quad (57)$$

Motsvarande fördelningsfunktion skriver vi

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (58)$$

Fördelningsfunktionen kan bara beräknas numeriskt och är därför tabellerad eller tillgänglig som standardfunktion i vissa program, bl.a. matlabs statistiska toolbox. På grund av enkla samband mellan olika normalfördelningar behöver inte fördelningsfunktionen tabelleras för några andra parametervärden.

Sats 6.4 Om X är $N(\mu, \sigma)$ så har $Z = (X - \mu)/\sigma$ den standardiserade normalfördelningen, $N(0, 1)$.

Sats 6.5 Om X är $N(\mu, \sigma)$ så är varje linjär transform $aX + b$ också normalfördelat $N(a\mu + b, |a|\sigma)$. Speciellt gäller då att om Z är $N(0, 1)$ så blir $\mu + \sigma Z$ $N(\mu, \sigma)$, ($\sigma > 0$).

Kommandot $Z = randn$ ger i matlab ett $N(0, 1)$ -fördelat slumpstal och exempelvis $3 + 5 \times Z$ ger då ett $N(\mu = 3, \sigma = 5)$ -fördelat slumpstal.

Sats 6.6 Om X och Y är oberoende $N(\mu_x, \sigma_x)$ resp. $N(\mu_y, \sigma_y)$ så är varje linjärt uttryck $U = aX + bY + c$ normalfördelat $N(\mu_u, \sigma_u)$ där $\mu_u = a\mu_x + b\mu_y + c$, $\sigma_u = \sqrt{a^2\sigma_x^2 + b^2\sigma_y^2}$.

⁸Lika ofta används variansen som parameter och beteckningen $N(\mu, \sigma^2)$. För att undvika misstag när siffror sätts in kan man skriva $N(\mu = 3, \sigma = 2)$ i vår beteckning eller $N(\mu = 3, \sigma^2 = 4)$ i den alternativa.

Bevis: Uttrycken för väntevärde och standardavvikelse följer i alla satserna de allmänna lagarna för väntevärde och varians/standardavvikelse. Se (??, ??). Bevis för satserna ??, ?? följer lätt ur (??) där vi studerar frekvensfunktioner för linjära funktioner av variabler. Den tredje satsen kan förenklas till ett bevis av att $X + Y$ blir normalfördelad eftersom enligt de föregående satserna $X' = aX$ och $Y' = bY + c$ båda är normalfördelade och oberoende. Ett bevis med transformering (momentgenererande funktioner) ges senare, se (??) och texten som följer, och ett bevis med direkt integrering till sammans frekvensfunktion kan genomföras på ett par A4-sidor med metoden i exempel ?? där man om $U = X + Y$ bildar

$$f_U(u) = \int_{-\infty}^{\infty} f_X(x)f_Y(u-x)dx$$

och räknar fram normalfördelningens frekvensfunktion.

Kontroll 1: Frekvensfunktionen (??) ger verkligen väntevärdet μ och standardavvikelsen σ .

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

substitution $\frac{x-\mu}{\sigma} = t$

$$= \int (\mu + \sigma t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

eftersom $te^{-t^2/2}$ är udda och ger konvergent integral. Att den sista integralen blir 1 är ett av matematikens paradexempel på nyttan av polära koordinater och visas nedan. Alltså följer $E[X] = \mu$.

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

samma substitution och partiell integration

$$\begin{aligned} &= \int \sigma^2 t^2 \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &= \sigma^2 \left(\frac{1}{\sqrt{2\pi}} [t(-e^{-t^2/2})]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \right) = \sigma^2 \end{aligned}$$

eftersom den utintegrerade biten blir 0 i båda gränserna och sista integralen blir 1 som tidigare.

Kontroll 2: Funktionen (??) är en frekvensfunktion.

(Krav: positiv funktion med integralen 1).

a) $f \geq 0$ inses lätt;

b) $\int_{-\infty}^{\infty} f(x)dx =$ subst. $t = (x - \mu)/\sigma = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} I$.

Kvadrera integralen och skriv

$$I^2 = \int_{-\infty}^{\infty} e^{-t^2/2} dt \int_{-\infty}^{\infty} e^{-u^2/2} du$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{t^2+u^2}{2}} dt du =$$

subst. polära koordinater $t = r \cos \alpha$, $u = r \sin \alpha$, $dt du = r dr d\alpha$ som integrerar över hela (t, u) -planet om vi bildar

$$\int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\alpha = \int_0^{2\pi} 1 d\alpha = 2\pi.$$

Alltså är $I = \sqrt{2\pi}$ och $\int f(x) dx = 1$.

6.3 Hastighetsfördelning i en gas

Tänk dig att vi drar en molekyl på måfå i luften i vår närhet. Låt oss också specialisera så att alla molekyler har samma vikt, eller betinga på att vi t.ex. drar en O_2 -molekyl. Hur fort rör sig denna molekyl? Vi söker sannolikhetsfördelningen för hastighetsvektorn.

Förutsättning 1: Molekylens hastighetsfördelning är densamma i alla riktningar, d.v.s. om vi projicerar dess hastighetsvektor på en riktning så är sannolikhetsfördelningen för projektionen alltid likadan.

(Man kan argumentera för att tyngdkraften skapar en skillnad i höjddled vilket måste kompenseras av att trycket avtar med höjden, men detta är en mycket svag effekt jämfört med molekylhastigheterna och försumbar vid de tryckgradienter som gäller.)

Förutsättning 2: Hastighetsvektorns projektion på ortogonala riktningar ger oberoende resultat.

(Detta är ett mer tekniskt villkor, men hastigheten är resultatet av ständiga krockar med andra partiklar och det är inte alls orimligt att detta ger oberoende resultat i ortogonala riktningar (av symmetriskäl måste kollisionerna ge okorrelerade hastighetsbidrag i ortogonala riktningar). Ett principiellt motargument är att ljushastigheten ger en begränsning, så om en hastighetskomponent är mycket stor kan inte de andra också vara mycket stora. Vid normala temperaturer är sådana hastigheter så orimligt osannolika att begränsningen saknar betydelse.) Däremot är det viktigt att vi begränsar oss till partiklar med samma massa eftersom tunga partiklar rör sig långsammare (i genomsnitt) än lätta i en blandgas, och det kan skapa ett beroende.

Låt V_x, V_y, V_z vara de stokastiska hastigheterna i de ortogonala (x, y, z) -riktningarna. Inför frekvensfunktionen

$$f(v_x, v_y, v_z),$$

som bör vara kontinuerlig och troligen också deriverbar. Villkoret 1 att alla riktningar är likvärdiga ger att enbart hastighetsvektorns längd (eller längden i kvadrat) kan påverka funktionen f

$$f(v_x, v_y, v_z) = g(v_x^2 + v_y^2 + v_z^2). \quad (59)$$

Sätt $A = g(0)$ och inför $h(\cdot) = \ln(g(\cdot)/g(0))$ så får vi

$$f(v_x, v_y, v_z) = Ae^{h(v_x^2 + v_y^2 + v_z^2)}, \quad (60)$$

där $h(0) = 0$.

Oberoendet (förutsättning 2) ger sedan att

$$f(v_x, v_y, v_z) = f_1(v_x)f_1(v_y)f_1(v_z) = B^3 e^{h_1(v_x^2)} e^{h_1(v_y^2)} e^{h_1(v_z^2)}, \quad (61)$$

där $h_1(0) = 0$ om vi väljer $B = f_1(0)$.

Sambandet $(??) = (??)$ ger då att

$$f(v_x, v_y, v_z) = Ae^{h(v_x^2 + v_y^2 + v_z^2)} = B^3 e^{h_1(v_x^2) + h_1(v_y^2) + h_1(v_z^2)}. \quad (62)$$

Genom insättning av speciella värden får vi följande samband

$$v_x = v_y = v_z = 0 \text{ ger } A = B^3,$$

eftersom $h(0) = h_1(0) = 0$;

$$v_y = v_z = 0 \text{ ger } Ae^{h(v_x^2)} = B^3 e^{h_1(v_x^2)}$$

och $A = B^3$ gör att $h(v_x^2) = h_1(v_x^2)$, d.v.s. $h(\cdot) = h_1(\cdot)$; Därefter ser vi att

$$v_z = 0 \text{ ger } h(v_x^2 + v_y^2) = h(v_x^2) + h(v_y^2). \quad (63)$$

Hur ser då funktionen h ut? Vi har om vi ersätter v_x^2 med a , v_y^2 med b att

$$h(a + b) = h(a) + h(b).$$

Om h är deriverbar blir $h(a + \delta a) = h(a) + h(\delta a)$.

$$\frac{h(a + \delta a) - h(a)}{\delta a} = \frac{h(\delta a)}{\delta a}$$

och högerderivatan ($\delta a \rightarrow 0^+$) ger då $h'(a) = h'(0) = \text{konstant}$. Alltså blir $h(a) = \alpha a + \beta$ och $h(0) = 0$ ger $\beta = 0$. Det ger att $h(a) = \alpha a$, $a \geq 0$ blir den enda deriverbara lösningen. Även om vi nöjer oss med kontinuitet hos funktionen h så är detta den enda lösningen.⁹

Återgå till $(??)$ med $h(a) = \alpha a$ så blir

⁹Utnyttja att $h(a + b) = h(a) + h(b)$ ger $h(\frac{k}{n}) = kh(\frac{1}{n})$ och $h(\frac{1}{n}) = \frac{1}{n}h(1)$, så $h(\frac{k}{n}) = \frac{k}{n}h(1)$. Det ger linjariteten för alla rationella tal och kontinuiteten ger resten

$$f(v_x, v_y, v_z) = Ae^{\alpha(v_x^2 + v_y^2 + v_z^2)},$$

och $\alpha < 0$ krävs för ändlig integral (ändlig sannolikhet). Skriv $\alpha = -\frac{1}{2\sigma^2}$ och normera med A så blir

$$f(v_x, v_y, v_z) = Ae^{-\frac{1}{2}\frac{v_x^2 + v_y^2 + v_z^2}{\sigma^2}}, \quad (64)$$

med $A = \left(\frac{1}{2\pi\sigma^2}\right)^{3/2}$.

Detta är en 3-dimensionell normalfördelning med oberoende $N(0, \sigma)$ -fördelade komponenter. Maxwell-Boltzmanns hastighetsfördelning för en ideal gas.

Samma hastighetsfördelning är också användbar som modell för de fria elektronernas rörelse i ett elektriskt motstånd eller en ledare.

6.3.1 Fördelningsgodis

Vi har bara en parameter (standardavvikelsen σ) i fördelningen. Nu finns det ett sätt att få mera fart på partiklarna i gasen (hur då?) och det enda som kan ändra sig i hastighetsfördelningen är parametern σ , som därför måste bli en funktion av (vad då?). Dessutom blir standardavvikelsen en funktion av massan om olika gaskomponenter jämförs.

Givet massan och hastighetsfördelningen plus partikelmängden per volymenhet så kan man räkna ut trycket p.g.a. att partiklarna studsar¹⁰ mot en begränsande vägg. Det är då halva normalfördelningen (vinkelrätt mot väggen) som blir rätt riktad och träffar ytan och dessutom får man ta hänsyn till att snabba partiklar rör sig längre sträckor. Vid en blandgas erhålls ett partialtryck från varje partikeltyp. I kvicksilverbarometern bildar kvicksilvret en sådan vägg och på andra sidan väggen är trycket 0 (vakuum). Partikelbombardemanget är precis så kraftigt att det står emot ca 760 mm kvicksilver och samma bombardemang träffar givetvis alla ytor i omgivningen.

Partikelmassan multiplicerad med $(v_x^2 + v_y^2 + v_z^2)/2$ blir rörelseenergi och det är den som fångar upp tillförd energi i form av värme när gasen är innesluten i en konstant volym. Man kan rätt lätt räkna ut en sannolikhetsfördelning för partikelns rörelseenergi med hjälp av normalfördelningen och får då en konstant gånger en s.k. Chi2-fördelning med 3 frihetsgrader.

En annan intressant beräkning är fördelningen för hastighetens belopp (farten = längden av vektorn (v_x, v_y, v_z)). Typiska värden på farten hos luftmolekyler vid rumstemperatur ligger runt ljudhastigheten 100–500 m/s och kanske är det därför

¹⁰Studsar är bara en bild eftersom partiklar kan tänkas tränga in i väggen eller tillfälligt fastna på ytan. Vid jämvikt avges lika mycket partiklar som det upptas och med samma rörelsemoment och därför fungerar bilden.

ljudet har den fart det har (?).

Frågeställningar av denna typ visar att man behöver kunna gå från en fördelning till nya fördelningar för olika funktioner av variabeln. Exempel på enkla sådana räkningar tas upp på annat håll. Här ger vi bara en motivering för att de är intressanta.

6.3.2 Intensitet och Weibull

Se "Statistisk slutledning" avsnitt 1.3 och kap. 9.

6.4 Transform av fördelningen – den momentgenererande funktionen

Det finns flera olika sätt att transformera en funktion $f(x)$ till en annan funktion $m(t)$ så att den nya funktionen har bekväma egenskaper i beräkningar där den ursprungliga funktionen ger krångliga uttryck. Klassiska exempel är transformeringar av tidsfunktioner till frekvensfunktioner (sinus och kosinusuttryck eller motsvarande komplexa funktioner) där de senare är enkla att hantera i linjära uttryck som förekommer bl.a. i elektroniska kretsar och filter. Den momentgenererande funktionen som vi tar upp här ger enkla bevis av vissa fördelningsresultat och kan framför allt användas för att bevisa centrala gränsvärdesatsen för en stor klass av fördelningar. En komplex variant av liknande typ (den karaktäristiska funktionen) fungerar på ytterligare några fördelningar men kräver komplex matematik vilket vi inte förutsätter här.

Definition 6.7 För en variabel X med frekvensfunktionen $f(x)$ definieras den momentgenererande funktionen som

$$m(t) = E[e^{Xt}] = \begin{cases} \int e^{xt} f(x) dx & \text{kontinuerliga fallet} \\ \sum e^{x_i t} f(x_i) & \text{diskreta fallet} \end{cases} \quad (65)$$

om väntevärdet existerar (ändligt) för alla t i en öppen omgivning av $t = 0$.

Funktionen $m(t)$ är en transform av frekvensfunktionen. Man kan visa att transformen är unik i betydelsen att olika sannolikhetsfördelningar ger olika transformers och varje transform kan transformeras tillbaka till en unik fördelningsfunktion. Det betyder att man kan känna igen sannolikhetsfördelningar med hjälp av deras transformers.¹¹

Några exempel: En binomialfördelad variabel ($Bi(n, p)$ med $q = 1 - p$) har

$$m(t) = \sum e^{kt} \binom{n}{k} p^k q^{n-k} = \sum \binom{n}{k} (pe^t)^k q^{n-k} = (pe^t + q)^n.$$

¹¹Väntevärdet $E[X^n]$ kallas n te momentet för X medan $E[(X - \mu)^n]$ kallas n te centralmomentet. Om man deriverar $m(t)$ så får man $m'(t) = \frac{d}{dt} E[e^{Xt}] = E[\frac{d}{dt} e^{Xt}] = E[Xe^{Xt}]$ och $m'(0) = E[X]$. På samma sätt ger högre derivator högre moment. Den fräcka omkastningen av E och derivata kan visas om t.ex. $f(x) < Ae^{-B|x|}$ för något $A, B > 0$.

En Poissonfördelad variabel med parameter λ har

$$m(t) = \sum_0^{\infty} e^{kt} \frac{\lambda^k}{k!} e^{-\lambda} = \frac{(e^t \lambda)^k}{k!} e^{-\lambda} = e^{e^t \lambda - \lambda}.$$

En exponentialfördelad variabel med intensiteten c (och väntevärde $1/c$) har

$$m(t) = \int_0^{\infty} e^{tx} c e^{-cx} dx = c/(c-t)$$

om $t < c$ men är divergent om $t \geq c$. Eftersom den finns i en omgivning av $t = 0$ så fungerar den.

En standardiserat normalfördelad ($N(0, 1)$) variabel har

$$\begin{aligned} m(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int \frac{e^{-\frac{1}{2}(x^2 - 2tx + t^2 - t^2)}}{\sqrt{2\pi}} dx = e^{\frac{t^2}{2}} \int \frac{e^{-\frac{1}{2}(x-t)^2}}{\sqrt{2\pi}} dx = e^{\frac{t^2}{2}}, \end{aligned}$$

där vi efter komplettering av kvadraten i exponenten har brutit ut en faktor och fått integralen av en ny frekvensfunktion för normalfördelningen $N(t, 1)$ som ger 1.

En $N(\mu, \sigma)$ -fördelad variabel har

$$m(t) = \int e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

subst. $y = (x - \mu)/\sigma$

$$\int e^{t(\mu+\sigma y)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = e^{t\mu + \frac{t^2\sigma^2}{2}},$$

där vi utnyttjat att integralen blir samma som för den standardiserade normalfördelningen om $t\sigma$ får ersätta t och $e^{\mu t}$ bryts ut..

Transformbevis av satserna ??, ?? och ??.

Om Z är $N(0, 1)$ så får $Y = \mu + \sigma Z$ den momentgenererande funktionen

$$m_Y(t) = E[e^{t(\mu+\sigma Z)}] = e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} m_Z(\sigma t). \quad (66)$$

Vi använde tricket att se σt som ett argument i den momentgenererande funktionen för Z som vi beräknat ovan. Det ger

$$m_Y(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}},$$

vilket visar att $Y \in N(\mu, \sigma)$. På samma sätt visar vi omvändningen att om Y är $N(\mu, \sigma)$ så blir $(Y - \mu)/\sigma$ standardiserad normal, dvs $N(0, 1)$.

Om vi sedan tar $V = aY + b$ där Y fortfarande har sin normalfördelning så blir

$$m_V(t) = E[e^{t(aY+b)}] = e^{tb} m_Y(ta) = e^{tb} e^{ta\mu + t^2 a^2 \sigma^2 / 2}$$

$$= e^{t(a\mu+b)+t^2 a^2 \sigma^2 / 2}$$

vilket överensstämmer med MGF för $N(a\mu + b, |a|\sigma)$.

Detta visar de båda första satserna.

Summor av oberoende variabler ger produkter av deras momentgenererande funktioner. Det ser man direkt från definitionen. Låt $Y = \sum_1^n X_i$. Då är

$$m_Y(t) = E[e^{t \sum X_i}] = E[e^{tX_1} e^{tX_2} \dots e^{tX_n}] = \prod E[e^{tX_i}] = \prod m_{X_i}(t).$$

För normalfördelade variabler innebär detta att om $X_i \in N(\mu_i, \sigma_i)$ så blir

$$m_Y(t) = \prod e^{t\mu_i + \frac{\sigma_i^2 t^2}{2}} = e^{t \sum \mu_i + t^2 \sum \sigma_i^2 / 2},$$

vilket stämmer med MGF för en $N(\sum \mu_i, \sqrt{\sum \sigma_i^2})$. Det bevisar satsen ?? □

Beviset utvidgas lätt till en godtycklig linjärkombination av oberoende normalfördelade variabler, $\sum a_i X_i + b$ eftersom varje term $a_i X_i$ är normalfördelad och konstanten b kan adderas till någon av dem eller behandlas som $N(b, 0)$.

6.5 Centrala gränsvärdessatsen—approximativ normal

Fördelningen för medelvärdet (eller summan) av n oberoende variabler med väntevärde μ och varians $\sigma^2 < \infty$ närmar sig normalfördelningen när n växer. Konvergensten gäller oberoende av vilken fördelning (med ändlig varians) som variablerna har. Den exakta formuleringen för medelvärdet är

Sats 6.8 (Centrala gränsvärdessatsen) För oberoende likafördelade variabler X_i med väntevärde μ och varians $\sigma^2 < \infty$ gäller

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq y\right) \rightarrow \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \tag{67}$$

Bevis: Vi kommer att utnyttja att

$$\left(1 + \frac{x}{n}\right)^n \rightarrow e^x;$$

$$(1 + t)^n \approx e^{nt}, \quad t \text{ liten, } n \text{ stor};$$

$$N(0, 1) \text{ har } m(t) = e^{t^2/2}.$$

Dessutom gäller för momentgenererande funktionen av en variabel U att $m(0) = 1$, $m'(0) = E[U]$, $m''(0) = E[U^2]$, etc. och om $\mu_U = 0$ så blir $\sigma^2 = E[X^2] = m''(0)$. Detta använder vi för att serieutveckla funktionen. Låt

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum \frac{X_i - \mu}{\sigma}$$

Nu är $E[Y] = 0$, $Var(Y) = 1$ och samma gäller för de enskilda termerna $(X - \mu)/\sigma$ i summan. Definitionen av momentgenererande funktion ger att

$$m_Y(t) = E[\exp(t \frac{1}{\sqrt{n}} \sum \frac{X_i - \mu}{\sigma})]$$

och eftersom termerna i summan är oberoende ger detta en produkt av väntevärden

$$m_Y(t) = \left(E[e^{\frac{t}{\sqrt{n}} \frac{X - \mu}{\sigma}}] \right)^n.$$

Termerna i parentesen är nu momentgenererande funktioner tagna i punkten t/\sqrt{n} enligt

$$m_{\frac{X - \mu}{\sigma}}\left(\frac{t}{\sqrt{n}}\right).$$

Här kan X tolkas som vilken som helst av X_i .

En funktion $m(t)$ som är tillräckligt deriverbar i en omgivning av $t = 0$ kan ju serieutvecklas i omgivningen så att

$$m(t) = m(0) + tm'(0) + \frac{t^2}{2}m''(0) + \frac{t^3}{6}m'''(\theta)$$

där $\theta \in (0, t)$. Det ger om vi ersätter t med t/\sqrt{n} och utnyttjar kopplingen mellan derivatorna och väntevärde, varians att

$$m_{\frac{X - \mu}{\sigma}}\left(\frac{t}{\sqrt{n}}\right) = 1 + 0 + \frac{t^2}{2n} \times 1 + \frac{t^3}{6n\sqrt{n}}m'''(\theta)$$

där $\theta \in (0, t/\sqrt{n})$. Under den extra förutsättningen att tredjederivatans uppfyller $|m'''(\cdot)| < A < \infty$ i en omgivning av $t = 0$ kan vi nu få resultatet

$$m_Y(t) = \left(1 + \frac{t^2}{2n} + \frac{t^3}{6n\sqrt{n}}m'''(\theta) \right)^n \approx e^{\frac{t^2}{2} + \frac{t^3}{6\sqrt{n}}m'''} \rightarrow e^{\frac{t^2}{2}},$$

$n \rightarrow \infty$.

Det innebär att momentgenererande funktionen för $Y \rightarrow$ momentgenererande funktionen för $N(0, 1)$. En konvergenssats för MGF (som vi inte visat) ger att fördelningsfunktionen för Y också konvergerar till fördelningsfunktionen för $N(0, 1)$. Beviset gäller för fördelningar med MGF-funktioner som är deriverbara. En komplex transform generaliserar resultatet till alla fördelningar med ändlig varians och det finns en hel litteratur om ytterligare generaliseringar under mycket skiftande förutsättningar. Se t.ex. Feller II för en avancerad beskrivning av flera varianter. \square

Centrala gränsvärdesatsen ger mycket skarpare uppskattningar än vad Chebyshevs olikhet (??) ger i stora talens lag även om de samtidigt är mer approximativa. Som exempel säger Chebyshevs olikhet att $P(|\bar{X} - \mu| > 3\sigma/\sqrt{n}) \leq 1/9$ medan den asymptotiska normalfördelningen ger sannolikheten 0.0026 (se tabell).

6.6 Approximation av fördelningar

Med stöd av centrala gränsvärdessatsen kan många fördelningar approximeras med normalfördelningar. Centrala gränsvärdessatsen säger ingenting om hur snabbt approximationen gäller men med simulering och på andra sätt kan man avgöra när approximationen blir tillräckligt god för praktiskt bruk. Typiskt är att den exakta fördelningens väntevärde och standardavvikelse blir parametrar i den approximerande normalfördelningen.

6.6.1 Binomial \rightarrow Normal

En $Bi(n, p)$ -fördelad variabel X kan uppfattas som en summa $X = X_1 + \dots + X_n$ av oberoende $X_i = 1$ med sannolikheten p , 0 med sannolikheten $1 - p$. Därmed gäller asymptotiskt centrala gränsvärdessatsen och för stora n fungerar approximationen

$$Bi(n, p) \approx N(np, \sqrt{np(1-p)}). \quad (68)$$

Approximationen är rimligt god om $np(1-p) > 10$, åtminstone om man inte studerar fördelningen långt ut i svansarna eller har p mycket nära 0 eller 1.

6.6.2 Poisson \rightarrow Normal

En $Po(\lambda)$ -fördelad variabel X kan ses som en summa $X = X_1 + X_2 + \dots + X_n$, X_i oberoende $Po(\lambda/n)$. Vi kan inse detta genom att hacka en poissonprocess i lagom långa bitar. Fördelningen är oändligt delbar eftersom detta gäller för godtyckligt n . Centrala gränsvärdessatsen fungerar om man tolkar den rätt. Om delarnas parameter hålls fix och man ökar antalet delar, vilket kräver att λ ökar, så konvergerar fördelningen mot normal. Eftersom $E[X] = \lambda$ och $\sigma_X = \sqrt{\lambda}$ så gäller för stora λ att

$$Po(\lambda) \approx N(\lambda, \sqrt{\lambda}) \quad (69)$$

Approximationen är hygglig om $\lambda > 16$.

6.6.3 Gamma \rightarrow Normal

Också för gammafördelningen finns det en summatolkning. Om

$$f(x) = c^\alpha x^{\alpha-1} e^{-cx} / \Gamma(\alpha), \quad x > 0,$$

och parametern α blir stor så uppstår denna fördelning när man summerar α oberoende exponentialfördelade variabler med intensiteten c . Detta gäller för heltal α men fördelningen varierar mjukt med α så approximationen fungerar för alla stora α . Tiden till den n te händelsen i en poissonprocess har denna fördelning med $\alpha = n$. $\alpha \geq 10$ räcker för en hygglig approximation.

Vinsten med normalapproximation Bi-, Po- och Gamma-fördelningarna m.fl. är ju enkla i sig själva så varför består vitsen med att normalapproximera? Jo fördelen kommer när man räknar med variablerna eftersom normalfördelningen

bevaras i alla linjära uttryck medan de andra fördelningarna förstörs utom i några specialfall. För normalapproximationen gäller istället att den förbättras ju fler (dåligt normalapproximerade) variabler man adderar eller subtraherar i sitt uttryck. I vissa fall kan man därför sänka kravet på parametrarna när man normalapproximerar enskilda termer i ett större uttryck.

6.6.4 Binomial \rightarrow Poisson

Det finns också nyttiga approximationer till andra fördelningar än normal. Som exempel tar vi approximationen från den diskreta binomialfördelningen till den diskreta poissonfördelningen. Eftersom poissonfördelningen bara har en parameter och enklare uttryck för sannolikheterna så blir detta ibland praktiskt. Approximationen ökar också vår förståelse av när poissonfördelning är en bra modell. Om $p \rightarrow 0$ och $np \rightarrow \lambda$ så gäller att $Bi(n, p) \rightarrow Po(\lambda)$. På vägen dit uppnår man approximationen

$$Bi(n, p) \approx Po(np). \quad (70)$$

Approximationen används t.ex. ofta för antalet fel vid kvalitetskontroll när felsannolikheten p är liten.

Bevis: Låt X vara $Bi(n, p)$. Binomialsannolikheten kan då skrivas

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{np(n-1)p \dots (n-k+1)p}{k!} (1-p)^{n-k}.$$

k är fixt, $n \rightarrow \infty$, $np \rightarrow \lambda$, $p \rightarrow 0$.

$$(1-p)^{n-k} = \frac{(1-p)^n}{(1-p)^k} \approx \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \rightarrow \frac{e^{-\lambda}}{1};$$

$$(n-r)p \rightarrow \lambda, \quad r = 0, 1, \dots, k-1.$$

Tillsammans ger det konvergens mot sannolikheten $\frac{\lambda^k}{k!} e^{-\lambda}$.

För $p < 0.1$ blir approximation $Po(np)$ hygglig nästan oavsett n . Som vanligt skall vi inte pressa approximationen långt ut i fördelningens svans. Binomialsannolikheterna är t.ex. 0 ovanför $X = n$ medan poissonfördelningen fortsätter att ge sannolikheter för alla positiva k .

7 Fördelningen för funktioner av variabler

Vi har tidigare behandlat *linjära* funktioner av variabler och sett hur väntevärde och varians kan hanteras. I specialfallet med normalfördelade variabler har vi dessutom sett att normalfördelningen bevaras vid linjära uttryck. Nu ser vi mer generellt på funktioner av en eller flera variabler och deras sannolikhetsfördelningar.

7.1 Funktioner av en enda variabel

Så fort man börjar räkna med stokastiska variabler så skapar man uttryck där dessa ingår. Varje sådant uttryck får en sannolikhetsfördelning och åtminstone för viktiga slutresultat och andra intressanta konstruktioner vill vi komma åt egenskaper i fördelningen. Vi skall här ge litet ideer om hur dessa fördelningar beräknas exakt. (Det finns också approximationsmetoder som kan utnyttjas och ibland nöjer man sig med väntevärde och varians). En annan sorts tillämpning har man vid simulering där man skapar funktioner av datorernas slumpantal så att man får en önskad sannolikhetsfördelning och hur detta går till antyds i exempel längre fram.

Vi har en variabel X med en känd fördelningsfunktion $F_X(x)$ och frekvensfunktion $f_X(x)$. Vi behöver nu sannolikhetsfördelningen för ett uttryck i (en funktion $g(\cdot)$ av) variabeln. Låt Y vara detta uttryck, $Y = g(X)$. Givetvis blir Y en stokastisk variabel med en egen fördelningsfunktion och om inte g är urartad får Y också en frekvensfunktion. För enkelhets skull begränsar vi oss till kontinuerliga variabler. Ofta är det enklast att angripa fördelningsfunktionen $F_Y(y)$ och sedan derivera denna för att få frekvensfunktionen $f_Y(y)$. Definitionen av F_Y ger

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in g^{-1}((-\infty, y])), \quad (71)$$

där $g^{-1}((-\infty, y])$ är mängden av X -värden som ger resultat $g(X)$ i det aktuella intervallet. För att lösa denna måste g vara så snäll att området $g^{-1}((-\infty, y])$ blir hanterligt så att vi kan bestämma sannolikheten för X i det området. Det hela blir enklare med några exempel.

Exempel 7.1 Linjär avbildning X är kontinuerlig med känd fördelningsfunktion $F_X(x)$ och frekvensfunktion $f_X(x)$. Låt $a > 0$ och b vara konstanter och $Y = aX + b$. Då blir

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a}). \quad (72)$$

Derivering ger då

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\frac{y-b}{a}) = f_X(\frac{y-b}{a}) \frac{1}{a}. \quad (73)$$

Som konkret tillämpning kan du nu låta X vara $N(\mu, \sigma)$. Frekvensfunktionen för X har vi behandlat tidigare. Använd resultatet i exemplet på den linjära avbildningen $Y = (X - \mu)/\sigma$ och verifiera att den får $N(0, 1)$ -fördelningens frekvensfunktion.

Exempel 7.2 Kvadratisk funktion. Låt $Y = X^2$. Fördelningsfunktionen blir

$$F_Y(y) = P(X^2 \leq y) = /om y > 0/ = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}). \quad (74)$$

Frekvensfunktionen deriveras fram liksom tidigare.

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}, y > 0.$$

Specialfall: Om X är $N(0, 1)$ så är $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. $Y = X^2$ ger

$$f_Y(y) = 2 \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, y > 0.$$

(Detta är en Ch12fördelning med en frihetsgrad, en fördelningstyp som vi bekantar oss med senare, den är i sin tur ett specialfall av gammalfördelningar.)

Exempel 7.3 Logaritmerat slumpstal: Slumpstal i datorer är $Re(0, 1)$ -fördelade, dvs likformigt över intervallet $(0, 1)$ (åtminstone i originalutförande). Det innebär att $F_X(x) = x$, $0 < x < 1$, och $f_X(x) = 1$, $0 < x < 1$. (Utanför intervallet $(0, 1)$ är $F = 0$ till vänster och 1 till höger, medan $f = 0$ på båda sidor.)

Låt $Y = -a \ln X$, där $X \in Re(0, 1)$ och $a > 0$. Sambandet är monotont (avtagande). Området för Y ges av att $Y \rightarrow \infty$ om $X \rightarrow 0$ och $Y = 0$ om $X = 1$ så $0 \leq Y < \infty$. Låt $y > 0$.

$$F_Y(y) = P(-a \ln X \leq y) = P(\ln X \geq -y/a) = P(X \geq e^{-y/a}) = 1 - F_X(e^{-y/a}).$$

Men $e^{-y/a}$ ligger i $(0, 1)$ -intervallet där $F_X(x) = x$ och det ger $F_Y(y) = 1 - e^{-y/a}$. Derivering ger så

$$f_Y(y) = \frac{1}{a} e^{-y/a}.$$

Detta är exponentialfördelningen.

En principiell skillnad på exemplen ovan är att i det första linjära fallet har vi en 1:1-avbildning (bijektiv avbildning) mellan X och Y . I det kvadratiske exemplet har vi däremot en situation där samma positiva Y kan fås från två olika X och därför saknar vi en entydig invers. Eftersom sambandet är så enkelt och det alltid (utom vid $Y = 0$) finns exakt 2 X -värden svarande mot Y så kunde vi enkelt summera) bidragen till $f_Y(y)$ från de båda X -värdena. I det tredje fallet har vi på nytt 1:1-avbildning men på begränsade områden av talaxeln och det blir då viktigt att studera vilka områden som gäller.

För 1:1-avbildningar i form av monotona deriverbara funktioner $Y = g(X)$ av kontinuerliga variabler finns det genvägar om man vill härleda samband mellan

frekvensfunktionerna. I princip gäller för sammanhörande värden x och y ($y = g(x)$) det enkla sambandet

$$f_X(x)dx = f_Y(y)dy$$

men för att täcka både monotont avtagande och monotont växande funktioner måste differentialerna tolkas positiva dvs.

$$f_X(x) \left| \frac{dx}{dy} \right| = f_Y(y), \quad f_X(x) = f_Y(y) \left| \frac{dy}{dx} \right| \quad (75)$$

gäller. Vi bevisar det rätt lätt på följande sätt. Funktionen g var monoton och deriverbar och X kontinuerlig.

$$F_Y(y) = P(g(X) \leq y) = \begin{cases} P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)), & g \text{ växande} \\ P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)), & g \text{ avtagande.} \end{cases}$$

För sammanhörande värden x och y i fekvnsfunktionen gäller nu $y = g(x)$ (eftersom variablerna är kopplade som $Y = g(X)$). Då är $g^{-1}(y) = x$ och vi kan skriva

$$F_Y(y) = \begin{cases} F_X(x), & \text{växande} \\ 1 - F_X(x), & \text{avtagande} \end{cases}, x = g^{-1}(y).$$

Derivering på y ger då

$$f_Y(y) = \begin{cases} f_X(x) \frac{dx}{dy}, & \text{växande} \\ -f_X(x) \frac{dx}{dy}, & \text{avtagande} \end{cases} = f_X(x) \left| \frac{dx}{dy} \right|,$$

för om g är växande (avtagande) så är också g^{-1} växande (avtagande). Derivatan $\frac{dx}{dy}$ blir positiv för växande g och negativ för avtagande g och i båda fallen gäller $\left| \frac{dx}{dy} \right|$. Dessutom är i sammanhörande punkter $\frac{dx}{dy} = 1/\frac{dy}{dx}$ så relationen gäller åt båda hållen. Observera tolkningen av $f_X(x)dx$ respektive $f_Y(y)dy$ som sannolikheter för små områden vid x respektive y . Väljer man sammanhörande x, y och dx, dy så är det naturligt att man får lika mycket sannolikhet för värden som avbildas på varandra.

Exempel 7.4 X är $N(\mu, \sigma)$. Vilken frekvensfunktion har X^3 ?

Lösning: Funktionen g , $Y = g(X) = X^3$, är monoton (och växande) över hela talaxeln. Sammanhörande argument ges av $y = x^3$. Från (??) får vi

$$f_Y(y) = f_X(x) \frac{dx}{dy} = f_X(x) \frac{1}{\frac{dy}{dx}} = f_X(x) \frac{1}{3x^2}$$

med $x = g^{-1}(y) = y^{1/3}$, dvs

$$f_Y(y) = f_X(y^{1/3}) \frac{1}{3y^{2/3}} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2} \frac{(y^{1/3}-\mu)^2}{\sigma^2}} \frac{1}{3y^{2/3}}, \quad -\infty < y < \infty.$$

Denna frekvensfunktion får en oändlighetspunkt i $y = 0$ och ett lokalt maximum eller en förhöjning nära $y = \mu$ och trots sin singularitet är den integrerbar med integralen 1.

Exempel 7.5 Fördelningsfunktionen som transformfunktion

Begreppsmässigt svårt brukar det vara om fördelningsfunktionen används i flera roller. Låt X ha fördelningsfunktionen $F_X(x)$. Bilda nu $Y = g(X)$ där $g(\cdot) = F_X(\cdot)$, dvs fördelningsfunktionen används också som avbildningsfunktion. (Det skulle vi kanske undvika om inte resultatet varit så intressant.) Vi söker nu $F_Y(\cdot)$ och $f_Y(\cdot)$. På nytt förutsätts X vara kontinuerlig och g (dvs F) deriverbar och monoton. Självklart är F växande. Området för Y är begränsat till $0 \leq Y \leq 1$. Låt $y \in (0, 1)$. Enligt (??) får vi

$$f_Y(y) \frac{dy}{dx} = f_X(x),$$

för sammanhörande punkter där $y = F_X(x)$. Eftersom

$$\frac{dy}{dx} = \frac{d}{dx} F_X(x) = f_X(x)$$

har vi då $f_Y(y)f_X(x) = f_X(x)$ vilket ger som enda lösning att $f_Y(y) = 1$ för alla y som hör samman med x där $f_X(x) > 0$. Det visar att vi har en rektangulärfördelning över intervallet $0 < y < 1$ (och det är en oviktig petitesse om ändpunkterna på intervallet skall räknas in eller ej eftersom dessa ändå får sannolikheten 0). Motsvarande fördelningsfunktion blir $F_Y(y) = y$, $0 < y < 1$, med de sedvanliga kompletteringarna för $y < 0$ och $y > 1$.

Man kan också härleda resultatet via fördelningsfunktionen på följande sätt. $F_Y(y) = P(F_X(X) \leq y)$. Här är det enklast att införa sammanhörande värden $y = F_X(x)$, ($x = F_X^{-1}(y)$) och på nytt observera begränsningen av y -värdena till $(0,1)$.

$$P(F_X(X) \leq y) = P(F_X(X) \leq F_X(x)) = \text{obs} = P(X \leq x) = F_X(x) = y.$$

Resultatet ger att $F_Y(y) = y$, $0 < y < 1$, vilket är samma som ovan.

Genom att använda fördelningsfunktionen som transform kom vi just fram till att man från varje fördelning som uppfyller villkoren på kontinuitet (F deriverbar) kan transformera till likformig (rektangulär) fördelning över $(0,1)$. Nu är datorernas slumpantal redan likformigt fördelade över detta intervall och den intressanta frågan är därför hur man kan vända på transformen och skapa en annan godtycklig fördelning.

Vi hade $Y = F_X(X)$, där Y blev $Re(0,1)$. Eftersom det är samma rektangelfördelning som datorernas slumpantal har så kan vi ersätta Y med ett slumpantal utan att ändra några fördelningsgenskaper. Om F_X är strängt monoton kan vi ta den inversa funktionen och skriva $X = F_X^{-1}(Y)$ och förvänta oss att X då har fördelningen $F_X(x)$. Detta fungerar utmärkt och kan sedan kompletteras med metoder att hantera diskreta fördelningar där fördelningsfunktionen har språng och fördelningsfunktioner som inte är strängt monotona, men i princip

klaras inversa fördelningsfunktionen att transformera likformiga slumpetal till alla fördelningar som har aptitliga fördelningsfunktioner.

Vi kan kombinera ihop våra resultat och se att konstruktionen $F_Z^{-1}(F_X(X))$ transformerar en stokastisk variabel X med fördelningsfunktionen F_X till en ny variabel med fördelningsfunktionen F_Z förutsatt att villkoren på deriverbarhet och sträng monotonitet är uppfyllda (eller kompletterade på lämpligt sätt). Till exempel kan normalfördelade variabler transformeras till godtyckliga andra kontinuerliga fördelningar och omvänt kan diverse kontinuerliga fördelningar transformeras till normalfördelade variabler.

7.2 Fördelningen för uttryck i mer än en variabel

Metod 1: Den mest direkta och råaste metoden att härleda fördelningen (F eller f) för en funktion $Y = g(X_1, \dots, X_n)$ av flera variabler är att bilda

$$F_Y(y) = P(g(X_1, \dots, X_n) \leq y) = \int_A \dots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

där $A = g^{-1}((-\infty, y])$ är det område för X -variablerna som gör $Y \leq y$. Resultatet kan sedan deriveras till f_Y . Ibland kan det bli enklare räkningar om man deriverar $F_Y(y)$ och dess integral innan man integrerar och får $f_Y(y)$. Se summaexemplet nedan.

Exempel 7.6 Låt $Y = X_1 + X_2$. Fördelningsfunktionen blir då

$$F_Y(y) = P(X_1 + X_2 \leq y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y-x_1} f(x_1, x_2) dx_2 \right) dx_1.$$

Här arbetar vi som om (X_1, X_2) kan ta alla värden i planet R^2 , men om detta inte gäller så är $f = 0$ i omöjliga punkter och det är först när vi sätter in speciella funktionsuttryck som vi måste begränsa integrationsområdet. Innan vi integrerar så kan vi istället derivera på y och söka frekvensfunktionen. Derivering på en övre integrationsgräns ger (vid enkelintegraler) integranden i den övre gränsen. Om vi helt fräckt deriverar uttrycket ovan genom att flytta in derivatan till det andra integrationstecknet så får vi

$$f_Y(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1) dx_1. \quad (76)$$

Resultatet är korrekt men härledningen tveksam. Ett formellt starkare bevis får vi om vi substituerar variabler i dubbelintegralen och sätter $u_1 = x_1$, $u_2 = x_1 + x_2$ ($x_1 = u_1$, $x_2 = u_2 - u_1$). Differentialerna $dx_2 dx_1$ byts mot

$$\left| \frac{\partial(x_1, x_2)}{\partial(u_1, u_2)} \right| du_1 du_2.$$

Här blir $\frac{\partial(x_1, x_2)}{\partial(u_1, u_2)} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ som får determinanten 1. Integralen ovan transformeras nu till

$$F_Y(y) = \int_{-\infty}^y \left(\int_{-\infty}^{\infty} f(u_1, u_2 - u_1) 1 du_1 du_2, \right.$$

och eftersom gränsen y nu ligger i den första integralen så kan vi direkt derivera på den och få integranden (dvs nästa integraluttryck) med övre integrationsgränsen y insatt istället för u_2 . Om vi går tillbaka till beteckningen x_1 istället för u_1 så har vi samma resultat.

Resultatet av beräkningen är rätt logiskt och lätt att minnas. För att få frekvensfunktionen i punkten y för summan integreras den tvådimensionella frekvensfunktionen över de kombinationer av argumenten x_1, x_2 (dvs $x_1, y - x_1$) som ger just summan y . (Det enkla resonemanget kan inte generaliseras till andra funktioner än summan).

Exempel 7.7 (Kvot) Låt X_1, X_2 vara oberoende och exponentialfördelade med samma parameter. Bilda $Y = X_1/X_2$. Vilken fördelning får kvoten? För enkelhets skull har vi valt variabler som bara kan bli positiva. Det gör att vi slipper indela efter tecken. Låt $y > 0$.

$$F_Y(y) = P(X_1/X_2 \leq y) = \int_0^{\infty} \left(\int_{x_1/y}^{\infty} \alpha e^{-\alpha x_1} \alpha e^{-\alpha x_2} dx_2 \right) dx_1.$$

Om vi väljer att lösa integralen så tar vi först det inre integraltecknet och får

$$F_Y(y) = \int_0^{\infty} \alpha e^{-\alpha x_1} \left[e^{-\alpha x_2} \right]_{x_1/y}^{\infty} dx_1 = \int_0^{\infty} \alpha e^{-\alpha x_1 - \alpha x_1/y} dx_1 = \frac{1}{1 + 1/y}, \quad y > 0.$$

Resultatet visar sig bli en fördelning som har oändligt väntevärde och som därför också saknar varians men fördelningsfunktionen och frekvensfunktionen existerar utan problem liksom medianen och andra percentiler i fördelningen. Man kan som vanligt derivera fördelningsfunktionen och få frekvensfunktionen.

Metod 2, transformer: Ibland har man nytta av transformer, t.ex. den momentgenererande funktionen där man bildar

$$m_Y(t) = E[e^{tY}] = E[e^{tg(X_1, \dots, X_n)}]$$

som beräknas antingen genom integrering eller genom att ta vara på kända egenskaper för X_i . Metoden fungerar framför allt på linjära uttryck. Om $m_Y(t)$ känns igen som momentgenererande funktion för någon fördelning så har Y denna fördelning. Annars måste en ofta komplicerad inverstransformering genomföras och det är inte säkert att man då vinner arbete jämfört med en direkt beräkning av fördelningen.

Exempel 7.8 En Poissonfördelad variabel X har momentgenererande funktionen $m(t) = E[e^{tX}] = \sum_0^\infty e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda e^t - \lambda}$ vilket följer av e -funktionens serieutveckling om man utnyttjar att $e^{tk} = (e^t)^k$.

Om X_i , $i = 1, \dots, n$ är oberoende och Poissonfördelade med parametrar λ_i , så blir momentgenererande funktionen för $Y = \sum X_i$

$$m_Y(t) = E(e^{tY}) = E[e^{\sum tX_i}] = \prod E[e^{tX_i}] = \prod e^{\lambda_i e^t - \lambda_i} = e^{(\sum \lambda_i) e^t - (\sum \lambda_i)},$$

vilket vi känner igen som mgf för en Poissonfördelad variabel med parametern $\sum \lambda_i$. Y har alltså en sådan Poissonfördelning. I det tredje likhetstecknet använder vi att funktioner av oberoende variabler blir oberoende och att väntevärdet av deras produkt blir produkten av väntevärdena. Försök gärna bevisa att en ren oviktad summa är den enda linjärkombinationen av oberoende Poissonfördelade variabler som ger Poissonfördelat resultat.

Metod 3, vid 1:1-avbildning: Om variabeln Y är en vektor med samma dimension som vektorn $X = (X_1, \dots, X_n)$ (eller om man kompletterar variabeln Y med ytterligare variabler så att detta gäller) och om man dessutom har deriverbara 1:1-samband mellan vektorerna X och Y så gäller en generalisering av frekvensfunktionssambandet (??)

$$f_Y(y_1, \dots, y_n) = f_X(x_1, \dots, x_n) \left\| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right\|, \quad (77)$$

där vi som vanligt låter sambandet mellan variabelvektorerna X och Y också definiera sambandet mellan argumenten när vi deriverar i determinanten. Se Statistisk slutledning där detta visas och används på normalfördelade vektorer. Om man bara vill ha fördelningen för en komponent i Y -vektorn så får man antingen ha sådan tur att den framgår direkt av vektorns fördelning (inträffar vid normalfördelning) eller också får man reducera bort övriga variabler genom integration. Ett möjligt val av kompletterande Y -variabler är att helt enkelt ta ett lagom antal av X -variablerna.

SLUT. Dags för statistisk slutledning.

References

- [1] Feller, W. (1971) An introduction to probability theory and its applications, Volume II. Wiley. (Avancerad text.)
- [2] Hjorth, U. (1998) Statistisk slutledning i ekonomi och teknik, Studentlitteratur.
- [3] Lewis-Orav, (1989) Simulation methods for statisticians, operation analysts and engineers, Wadsworth & Brooks/Cole.
- [4] Milton, J.S., Arnold, J.C. (1995) Introduction to probability and statistics, principles and applications for engineering and the computing sciences. McGraw-Hill.

Problem, egna och andras, i sannolikhetslära. Tagna från tentor, LiTH:s problemsamling mm.

R1: Utfall, sannolikhet, kombinatorik.

1.1 Låt utfallsrummet vara $\Omega = \{1, 2, 3, \dots, 12\}$ och $A = \{2, 4, 6, 8\}$, $B = \{3, 4, 5, 10, 11\}$. Ange händelserna $A \cup B$, $A \cap B$, $\bar{A} \cup \bar{B}$, $\bar{A} \cap \bar{B}$.

1.2 Visa att $\bar{A} \cap \bar{B} = \overline{A \cup B}$ och $\bar{A} \cup \bar{B} = \overline{A \cap B}$.

1.3 Familj I har 2 barn som antas bli pojke eller flicka med samma sannolikhet och oberoende av varandra. Gör ett utfallsrum för könen på dessa barn så att alla utfall får samma sannolikhet. Hur stor andel av utfallen ger lika många flickor som pojkar?

Familj II har 4 barn. Lös samma uppgift.

Familj k har $2k$ barn . . .

1.4 En partikel utslungas i en slumpmässig riktning vars tvådimensionella projektion observeras på en fotografisk plåt. Ange utfallsrum för den slumpvisa riktningen på den 2-dimensionella plåten. Gör sedan detsamma om 2 partiklar slungas ut.

1.5 Ett försök utförs tills ett lyckat resultat inträffar. Vi intresserar oss för antalet försök. Ge ett utfallsrum och gör gärna en figur som ett träd över de inledande utfallen.

1.6 Man räknar impulser under en minut vardera från två radioaktiva preparat och bildar kvoten av de räknade värdena. Ge ett utfallsrum för kvoten.

1.7 För en botanist är följande problem aktuellt när man studerar artrikedom (biodiversitet): a) Kasta ut en försöksyta på måfå på en äng ooh räkna antalet arter som finns innanför. Ange utfallsrummet för detta försök. b) Försöket i a) upprepas 16 gånger. Ge utfallsrum för det kombinerade försöket.

1.8 Skjut tre skott mot en tavla och observera träffbilden. Ge ett möjligt utfallsrum.

1.9 Mellan busshållplatsen och hemmet finns det 20 gatlyktor. Efter en tid har 3 av dessa gått sönder.

a) Vad är sannolikheten att alla tre sitter intill varandra?

b) (Svårare). Vad är sannolikheten att de sitter en och en med minst en fungerande lampa emellan. (Utgå exempelvis från händelser av typen "trasig lampa 1 och 2 är grannar".)

1.10 Precis samma sorts mätning görs 8 gånger. Alla resultaten blir olika trots att förutsättningarna är desamma varje gång. Låt A = den sista observationen blir störst, B = den näst sista blir näst störst. Vad blir sannolikheterna $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$?

1.11 Samma 8 mätningar igen. Vad är sannolikheten att den första blir minst och den sista blir störst?

1.12 På hur många sätt kan en kortlek (52 kort) delas ut till 4 spelare om alla får 13 kort och man inte beaktar ordningsföljden på de utdelade korten (däremot beaktas vem som fick vilka kort).

1.13 Tänk dig en oändlig serie av tärningkastresultat. Hur mäktigt (ändligt, numrerbart, kontinuerligt som R^n) blir utfallsrummet?

1.14 Man rör runt en sjö och tar sammanlagt 12 vattenprover vars fosforhalt analyseras. Det visar sig att de fyra högsta värdena kom från prover tagna intill varandra, vilket kanske skulle kunna tyda på utsläpp av fosfor i närheten. Vad är sannolikheten att få ett sådant resultat av en slump?

1.15 Genetikproblem. Varje individ har två anlag (ett från pappa och ett från mamma) på varje plats i sin genkedja eller mera exakt, inom varje kromosompar. Ett visst utseende eller symptom tänker vi oss kan kopplas till två genplatser. Använd symbolerna A och a för genetiska varianter på plats 1 och symbolerna B och b för genetiska varianter på plats 2 och på varje plats skall individen ha två anlag som kan vara lika eller olika.

a) Dra en individ på måfå och ange de möjliga utfallen av gensymboler som kan förekomma på dessa båda platser. Detta blir ett utfallsrum.

Anlaget B är dominant men inte A och för det speciella symptomet krävs dubbel A och minst ett B . b) En mamma har symptomet. Markera vilka utfall hon kan tillhöra. c) Pappan till samma barn har inte symptomet. Markera vilka utfall han kan tillhöra.

d) Gör ett kombinerat utfallsrum för mammans och pappans möjliga genkombinationer utan begränsning till om de visar symptomet eller ej. Här bör du ha så få utfall som möjligt och skapa ett rutnät av punkter. Ett tänkt barn till föräldrarna får slumpvis ett av pappans och ett av mammans anlag på respektive plats men vi avstår från att beskriva detta i utfallsrummet. Markera följande händelser $D =$ "mamman har dubbel B " ,

$E =$ "pappan har dubbel A " ,

$D \cap E$ och $D \cup E$, (ange i ord vad de betyder),

$F =$ "Ett barn till dessa föräldrar kan inte få symptomet",

$G =$ "Ett barn till dessa föräldrar måste få symptomet".

R2: Oberoende, beroende, betingning, Bayes.

2.1 Om 20% av en åldersgrupp har förhöjt blodtryck och 5% har diabetes, hur stor andel måste då ha båda symptomen för att dessa symptom skall vara oberoende i åldersgruppen?

2.2 Hur stora kan sannolikheten för A och B vara om de är oberoende och $P(A \cup B) = 0.76$ medan $P(A \cap B) = 0.14$.

2.3 Händelsen A är dubbelt så sannolik som B . Hur förhåller sig $P(A|B)$ och $P(B|A)$ till varandra?

2.4 Tre kort varav ett är svart på båda sidor, ett är rött på båda och ett är svart och rött stoppas i en hatt och blandas häftigt. Ett kort dras på måfå och läggs på bordet. Vad är sannolikheten att det är rött under om det visar röd översida? (Gissa först och räkna sedan).

2.5 Låt $p_1 = 0.01, p_2 = 0.05, p_3 = 0.007, p_4 = 0.002$ vara felsannolikheter på 4 olika delsystem som vart och ett kan stoppa en motor. Felen uppträder oberoende vilket innebär att mer än ett fel kan uppstå. Vad är sannolikheten för fel på de

olika delsystemen givet att motorn stannat och inga andra orsaker till stoppet är möjliga?

2.6 Varning! Detta problem är betydligt svårare än det föregående. Vi har samma oberoende felsannolikheter på delsystemen men nu har apparaten gjorts feltolerantare så att det blir stopp om minst två fel har inträffat men inte om bara ett. Vad är sannolikheten för de olika feltyperna givet att det har blivit stopp?

2.7 Vid en bakterieodling gör man för säkerhets skull 3 odlingar som oberoende av varandra tillväxer med sannolikheten p . a) Vad är sannolikheten p_k att k av dem tillväxer för $k = 0, 1, 2, 3$. b) En forskare genomför många sådana odlingar men kastar bort alla noteringar från fall som inte alls tillväxer i någon av de tre odlingarna och efter en tid minns han inte ens hur många gånger detta inträffade. Kvar är bara resultaten 1, 2, 3. Finn sannolikheterna för dessa resultat betingat att minst en odling tillväxer. Skattningar behandlas senare men tror du att man kan skatta värdet på p från data när man som forskaren slängt bort en grupp resultat?

2.8 Återgå till genetikproblemet 1.15. Antag att 25% av alla gener på genplats 1 i en population är A (resten a) och att 10% på genplats 2 är B (resten b). Låt alla mammans och pappans gener oberoende av varandra ha dessa sannolikheter. a) Finn sannolikheten att mamman har det aktuella symptomet. Finn därefter sannolikheten för kombinationen att mamman har symptomet och att pappan inte har det.

b) Vad är sannolikheten att barnet får symptomet givet (betingat) att mamman har det men pappan inte har det.

2.9 TV-leken med 3 dörrar.

2.10 I en bultfabrik svarar maskinerna a , b , och c för 25, 35 resp. 40% av produktionen. Av de tillverkade bultarna är 5, 4 resp. 2% defekta. En bult väljs ut på måfå och visar sig vara defekt. Hur stora är sannolikheterna att den kom från maskin a , b resp. c ?

2.11 Två tärningar kastas. Givet att de båda tärningarna visar olika resultat, vad är sannolikheten att någon av dem visar 4.

2.12 Vid tillverkning av reläer blir felfrekvensen 1%. För att minska denna före försäljning låter man alla reläer gå igenom en kontroll. Vid denna kasseras felaktiga reläer med sannolikheten 0.9 och felfria med sannolikheten 0.02. Vilken felfrekvens har man bland dem som går till försäljning? Hur stor blir felfrekvensen i den kasserade högen?

2.13 Klöver knekt tas bort ur en kortlek (som normalt innehåller 26 svarta och 26 röda kort). Därefter dras 13 kort som alla visar sig ha samma färg. Hur stor är sannolikheten att de är röda?

2.14 En tärning kastas 10 gånger. Beräkna sannolikheten att både 1 och 6 kommer upp.

2.15 Fem teknologer gör oberoende av varandra en försöksserie bestående av tre oberoende försök. Varje försök misslyckas med sannolikheten $1/4$. Beräkna

sannolikheten att ingen av de 5 teknologerna får en försöksserie med tre lyckade försök.

2.16 70% av alla män som söker vård för en besvärlig sjukdom är narkomaner. Samtidigt visar en studie att 2% av mänen i motsvarande åldrar använder narkotika. Slutligen är det 120 av 100000 som söker vård för sjukdomen. Hur stor andel av narkomanerna behöver söka vård för sjukdomen.

2.17 I 10 identiska kuvert läggs olika stora summor och vi har ingen aning om beloppen. Det tredje kuvertet som öppnas innehåller mer än de båda föregående. Vad är sannolikheten att det är det största beloppet av alla tio?

(Antag att man får behålla innehållet i det sist öppnade kuvertet. Finns det någon bra strategi om man vill maximera sannolikheten att få det högsta beloppet?)

R3: Stok. variabler, F , f , μ , σ .

3.1 Vid tillverkning kan två sorters fel uppstå oberoende av varandra och med sannolikheterna 0.08 respektive 0.05. Bestäm fördelningsfunktionen och frekvensfunktionen för $X =$ antalet fel på en tillverkad enhet.

3.2 Beräkna väntevärde, standardavvikelse och median för den stokastiska variabeln X med frekvensfunktionen $f(x) = 2x$, $0 < x < 1$.

3.3 Ett register består av N block och genomletas i ordning från block 1 till block N . Antag att 4 olika uppgifter var och en ligger i ett slumpmässigt valt block helt oberoende av varandra. Dessa söks i samma genomletning. Låt X vara antalet block som man måste leta i innan alla uppgifterna är funna. Bestäm frekvensfunktion och fördelningsfunktion för X . Beräkna dessutom väntevärdet i specialfallet $N = 5$.

3.4 Bestäm k så att $f(x) = kxe^{-x^2/2\beta^2}$, $x > 0$, blir en möjlig frekvensfunktion för en variabel X . Beräkna sedan $P(X > \beta)$. (Konstanten β är positiv.)

3.5 Visa att $a = \mu$ minimerar $E[(X - a)^2]$.

3.6 Visa att $b =$ medianen minimerar $E[|X - b|]$. Låt X vara kontinuerlig och $f(x) > 0$ vid medianen så blir det enklast.

3.7 Den kontinuerliga variabeln X har frekvensfunktionen $f(x) = \frac{2}{\pi\sqrt{1-x^2}}$, $0 < x < 1$. a) Bestäm $P(0.5 < X < 0.75)$. b) Bestäm väntevärdet $E[X]$.

3.8 Beräkna variansen för ett tärningkastresultat.

3.9 Livslängden hos en radioaktiv atom är exponentialfördelad. Atomens halveringstid definieras av att atomen med sannolikheten 0.5 överlever den tiden. Beräkna sannolikheten att atomen överlever 5 halveringstider.

3.10 En punkt väljs på måfå innanför en liksidig triangel med höjden 3. Låt X vara avståndet från punkten till närmast liggande sida. Bestäm fördelningsfunktionen och frekvensfunktionen för X och beräkna μ och σ .

3.11 Ett tal X dras på måfå mellan 0 och 2. Låt $Y = e^X$ och beräkna väntevärde och varians för Y .

3.12 Sju svensktillverkade och elva tysktillverkade kullager används på exakt samma sätt. Efter fyra år är alla utom fem nerslitna. Antag att kullagren är

exakt lika bra och bestäm i så fall frekvensfunktionen samt μ och σ för antalet svenska lager som fortfarande håller.

3.13 Drag en punkt slumpvis på en cirkelskiva med radien R . Bestäm fördelningsfunktion och frekvensfunktion för avståndet X från centrum till punkten. $P(X > R/2)$? $E[X]$?

3.14 Ett skogsföretag äger ett cirkelformat skogsområde med 50 km radie i vars centrum en fabrik är belägen. Beräkna den genomsnittliga transportsträckan om det avverkade timret är likformigt fördelat över skogsområdet och alla transporter sker fågelvägen mellan avverkningsplatsen och fabriken.

R4: Flera variabler, oberoende/beroende, linjära uttryck.

4.1 Variablerna X och Y har frekvensfunktionen $f(x, y) = e^{-y}$, $0 < x < y$, ($f(x, y) = 0$ i resten av planet).

- Bestäm de marginella frekvensfunktionerna för X och Y .
- Beräkna $P(X > 2|Y < 4)$.
- Är variablerna oberoende?

4.2 En punkt dras på måfå på (omkretsen av) en cirkel med radien R . Låt X och Y vara punktens koordinater. Bestäm väntevärde, standardavvikelse och kovarians för X och Y .

4.3 En vägningsmetod är sådan att mätfelet har väntevärde 0 och variansen σ^2 . Man väger med denna metod dels var och en av två guldklimpar, dels båda guldklimparna tillsammans. Låt dessa mätvärden vara X_1, X_2, X_3 .

- Visa att för varje α är $E[(\alpha(X_1 + X_2) + (1 - \alpha)X_3)] = \mu_1 + \mu_2$ om μ_1, μ_2 är de båda guldklimparnas verkliga vikter.
- Bestäm α så att uttryckets varians blir minimal om de tre mätfelen är oberoende.

4.4 Den tvådimensionella variabeln (X, Y) är likformigt fördelad på enhetscirkeln, dvs den har frekvensfunktionen $f(x, y) = 1/\pi$, $x^2 + y^2 \leq 1$. Beräkna kovariansen $Cov(X, Y)$. Avgör dessutom om variablerna är oberoende. (Beräkna gärna marginalfördelningen för X även om oberoendet kan studeras på annat sätt.)

4.5 Hur många tal med 5 korrekt avrundade decimaler kan man addera om det summerade avrundningsfelet skall ha en standardavvikelse som är högst 0.001?

4.6 För att bestämma svavelhalten i ett parti svavelolja tar man 4 prover av oljan. Varje prov analyseras 2 gånger varefter man tar medelvärdet av alla 8 analysvärdena. Svavelhalten varierar något mellan de olika provtagningarna och är oberoende normalfördelade $N(\mu, \sigma_1 = 0.1)$ (där μ är medelhalten i hela oljepartiet). Felen i analyserna är oberoende med väntevärde 0 och standardavvikelse $\sigma_2 = 0.2$. Bestäm standardavvikelsen för medelvärdet av de 8 analyserna. (Obs, det finns 12 oberoende variabler i problemet!)

4.7 Ett bussbolag har 121 bussar och skall dimensionera en reparationsverkstad. Man önskar få en uppfattning om antalet reparationer som behöver göras under en månad. Av erfarenhet anser man sig veta en buss behöver genomgå 0, 1 eller 2 reparationer under en månad med sannolikheterna 0.3, 0.5, 0.2. Beräkna

väntevärde och standardavvikelse för sammanlagda antalet reparationer på de 121 bussarna under en månad.

4.8 Man kalibrerar ett instrument genom att ta medelvärdet \bar{x} av n_1 mätningar på en känd storhet. Därefter utför man n_2 mätningar i fält med det kalibrerade instrumentet och tar medelvärdet \bar{y} av dessa. Antag att alla dessa mätningar är oberoende med varianser σ_1^2 vid kalibreringen och σ_2^2 i fält. Om kalibreringsfelet antas vara konstant så blir slutresultatet att värdet i fält uppskattas till $\bar{y} - \bar{x} + K$, där K är det kända värdet vid kalibreringen. Hur bör man fördela mätresurserna om varje kalibreringsmätning kostar c_1 kronor och varje fältmätning kostar c_2 kronor och man för en tillåten kostnad C vill minimera variansen i slutresultatet.

4.9 En avdelning skall välja en styrelse med 7 ledamöter men ingen vill ta uppdraget så man beslutar att lotta ut posterna. Avdelningen består av 4 socialistiska kvinnor, 12 socialistiska män, 8 borgerliga kvinnor och 7 borgerliga män. Bestäm väntevärdena för antalet borgerliga ledamöter, antalet kvinnliga ledamöter, antalet borgerliga kvinnliga ledamöter.

R5: Olika fördelningar.

5.1 En dator genererar en slumpmässig följd av nollor och ettor. Varje tecken är en etta med sannolikheten p och oberoende av varandra.

a) Vilken sannolikhetsfördelning har antalet ettor bland de 100 första tecknen?

b) Efter varje etta kan det uppstå ett slumpmässigt antal nollor innan det blir en ny etta. Vad är sannolikheten att det blir k sådana nollor.

5.2 Beräkna väntevärdet i Poissonfördelningen direkt från definitionen.

5.3 I vissa länder är det viktigt för föräldrarnas försörjning att få en pojke. Antag att detta påverkade beteendet så att alla familjer skaffade sig barn tills de fick sin första pojke och sedan slutade. Vilken proportion pojkar och flickor skulle detta ge om sannolikheten för pojke var p ($p = 0.514$ ungefär). För en stor population avgörs proportionerna av väntevärdet för antalet flickor resp pojkar i en familj.

5.4 En partikel hoppar mellan intill varandra liggande heltal på reella axeln. Sannolikheten för hopp åt höger är 0.3 och för hopp åt vänster 0.7 och varje hopp är oberoende av tidigare hopp. Vad är sannolikheten att partikeln efter 20 hopp befinner sig 6 steg från utgångspunkte?

5.5 Vilken sorts sannolikhetsfördelning får man i följande situationer för X , Y , Z , U ?

a) En satellit löper ständigt samma risk att träffas av en meteorit. Satelliten färdas med konstant hastighet. Låt X vara den tillryggalagda sträckan till första meteoritträffen.

b) Efter ett år har satelliten fått Y träffar.

c) Efter 100 varv runt jorden har satelliten fått träffar under Z av varven.

d) Vi får information om att under 20 av de första 100 varven har satelliten blivit träffad. Man tar på måfå ut data från 10 av de hundra varven och får då U varv utan någon meteorträff.

5.6 Antalet radioaktiva partiklar som registreras under en viss tid är en Poissonfördelad variabel. Hur stor är sannolikheten att två eller flera registreras om

sannolikheten för 0 är $1/3$.

5.7 En stav med längden L kapas i en punkt som väljs helt på måfå. Bestäm fördelningsfunktionen och väntevärdet för den längsta biten.

5.8 Två larmcentraler bevakar varsitt distrikt och larmcentral 1 har larm dubbelt så ofta som larmcentral 2, dvs intensiteten är dubbel. Vad är sannolikheten att larmcentral 2 får larm först av de två?

5.9 Vid tillverkning av cylindrar blir innerdiametern $N(\mu = 20, \sigma = 0.1)$. Tillverkade axlar får en diameter som är $N(19.2, 0.1)$. Beräkna sannolikheten att en axel och en cylinder passar ihop, vilket inträffar om axelns diameter är mellan $1/2$ och 1 enhet mindre än innerdiametern.

5.10 Vid tillverkning av motstånd blir resistansen normalfördelad med $\mu = 100$ och $\sigma = 5$. Vad är sannolikheten att fyra seriekopplade sådana motstånd skall få en resistans mellan 385 och 415Ω ?

5.11 Vid tillverkning av motorer har man en rätt stor slumpmässig variation i utsläppet av olika ämnen. Ett kritiskt utsläpp kan man justera väntevärdet för men tappar då samtidigt litet av motoreffekten. Antag att värdena är normalfördelade med standardavvikelse σ . Hur långt under miljömyndighetens gränsvärde skall man ställa μ om bara 2% av fordonen skall få överskrida gränsvärdet?

R6: Fördelningar (forts), Poissonprocess, MGF.

6.1 Till en telefonstation sker anrop enligt en poissonprocess med intensiteten 40 samtal per minut. Samtalen varar i genomsnitt 2 minuter och 30 sekunder. Antag att alla samtal är exakt så långa och bestäm under den förutsättningen sannolikheten att det i ett visst ögonblick skall vara uppkopplat minst 125 samtal.

6.2 Antalet trafikolyckor under en dag i ett visst område kan anses vara en Poissonfördelad variabel med $\lambda = 0.1$. Beräkna sannolikheten att det under 10 dagar inträffar högst 3 olyckor.

6.3 Anrop till en telefonväxel beskrivs med en Poissonprocess som har intensiteten 0.6. Om man vet att det i tidsintervallet $[2,5]$ har kommit minst ett anrop, vad är den betingade sannolikheten att det har kommit exakt ett?

6.4 Antalet trafikolyckor under ett år är nära Poissonfördelat och det gäller också för en given klass av trafikolyckor, t.ex. dödsolyckor. Under 1998 dog ca 540 i trafiken och 1999 var det mycket pressdiskussioner om att riskerna ökat eftersom det dog 570 stycken. Nu dör det ibland flera stycken i samma olycka vilket ökar variansen mer än det ökar väntevärdet. Tänk dig för enkelhets skull att varje dödsolycka har exakt ett offer och att $\mu = \lambda \approx 540$. Visa att skillnaden mellan dödsantalen ligger nära en standardavvikelse (och troligen inte ens det om man korregerar för flera offer i samma olycka).

Det kan finnas andra goda skäl att inte acceptera antalet fatala trafikolyckor men underlaget i form av antalet dödade ger ingen saklig grund för påståendet att faran ökat i trafiken.

6.5 X_1 och X_2 är oberoende $Exp(\mu)$. Finn fördelningen för X_1 givet att $X_1 < s, X_1 + X_2 > s$. Det finns en koppling till Poissonprocessen eftersom tiderna

mellan händelser bestäms av den konstanta intensiteten.

6.6 En bilfärja skall byggas för ett färjeställe och dimensioneras för ett flöde av 0.5 bilar per minut som anländer till den ena sidan. Det tar 10 minuter för färjan att gå fram och tillbaka. Hur många bilar måste färjan byggas för om sannolikheten att den skall bli full skall vara mindre än 5%.

6.7 I en liter vatten finns det 2000 mikroorganismer utspridda. Man tar ett vattenprov på 1 ml. Beräkna sannolikheten att det inte innehåller någon mikroorganism. Hur stort vattenprov måste man ta för att med sannolikheten 0.9 erhålla minst en mikroorganism?

6.8 Vi har tre oberoende variabler där X_1 är $N(1, 2)$, X_2 är $N(2, 3)$, X_3 är $N(0, 2)$. (Vi använder skrivsättet $N(\mu, \sigma)$.) Beräkna $P(6X_1 - X_2 - 2X_3 > 17)$.

6.9 Bestäm den momentgenererande funktionen (MGF) för en Bernoullifördelad variabel. Verifiera att första- och andraderivatorna i punkten 0 ger $E[V]$ och $E[X^2]$.

6.10 Använd MGF från föregående problem och bestäm med hjälp av den MGF för en $Bi(n, p)$ -fördelad variabel. Visa att derivatorna i punkten 0 ger motsvarande resultat för den nya variabeln.

6.11 Man utför n oberoende försök där försök i lyckas med sannolikheten p_i . Visa att detta inte ger en binomialfördelning för antalet lyckade försök om p_i är olika. (Vid oberoende likafördelade variabler blir sannolikheterna för rekord $p_i = 1/i$ och rekordhändelserna oberoende. Antalet rekord fram till variabel n får därför denna typ av fördelning.)

R7: Centrala gränsvärdessatsen, approximation av fördelningar.

7.1 Det står 50 kunder i kön vid stängningsdags på en bank. Betjäningstiderna i minuter för dessa är oberoende med $f(x) = 0.25xe^{-x/2}$, $x > 0$. Uppskatta sannolikheten att den sammanlagda betjäningstiden understiger 4 timmar.

7.2 Hur länge måste man observera en Poissonprocess med intensiteten ν för att med 90% sannolikhet ha fått minst 100 händelser.

7.3 Ur ett parti med 5000 enheter dras slumpvis 100 som inspekteras. Antag att partiet har 15% som är defekta. Vad är sannolikheten att få högst 10 defekta bland de dragna?

7.4 Man kan illustrera centrala gränsvärdessatsen med simulering. Matlabkommandot `x=rand(k,1000)`; ger en matris med oberoende slumpstal som är likformigt förd. på $(0,1)$. Kommandot `y=sum(x)`; summerar kolumnerna och ger 1000 sådana summer. Till sist ger kommandot `hist(y)`; ett histogram som illustrerar fördelningen för kolumnsummorna. Pröva detta med $k=1, 2, 4, 8$ så ser du hur centrala gränsvärdessatsen arbetar vid likformig fördelning.

Sedan kan du pröva andra fördelningar genom att ta funktioner av x innan du summerar till y . Till exempel ger `x=-log(rand(k,100))`; exponentialfördelade data och man kan då behöva ta ett större k t.ex. 16 eller 32 för att få ett vackert resultat.

7.5 För närvarande föds det ungefär 80000 barn om året i Sverige. Varje födsel ger en pojke med sannolikheten 0.514 och vi bortser från enäggstvillingar. Uppskatta

sannolikheten att det kan födas fler flickor än pojkar nästa år.

7.6 Slumptal i dator beter sig som oberoende och likformigt fördelade på intervallet $(0,1)$. I en beräkning multipliceras 100 slumptal och det kan ge problem om produkten blir alltför liten, så därför tar man logaritmen av produkten. Låt Y vara produktens logaritm. Bestäm väntevärde och varians för Y och ange en approximativ fördelning.

7.7 Handlarna A och B säljer kaffeburkar. A tar 31.60 kronor burken och B tar 38 kronor. Till A anländer kaffekunder enligt en Poissonprocess med intensiteten 12 kunder per timme och till B kommer de med intensiteten 8 kunder per timme. Varje kund köper en burk. Vad är sannolikheten att A efter 7 timmar har sålt för mer pengar än B?

7.8 Fyra sorters öl testas av en smakpanel bestående av 40 provsmakare. Provs-makarna avgör oberoende av varandra och okunniga om vad de får i glaset vilket öl som är bäst, näst bäst, näst sämst och sämst. Det bästa får 4 poäng, och övriga får 3, 2, och 1 poäng. För varje ölsort summeras de 40 poängsiffrorna från olika provsmakare. Ölsort A är ny och om den får minst 115 poäng så kommer den att introduceras på en ny marknad. Beräkna sannolikheten att få en så hög poängsiffra om alla provsmakarna sätter poängen helt på måfå. (Detta är en helt hypotetisk fördelning som bara utnyttjas som referens men den skulle kunna inträffa om ölsorterna var exakt likvärdiga eller om provsmakarnas smak varierar slumpmässigt.)

R8: Funktioner av variabler.

8.1 Variabeln X är exponentialfördelad med intensiteten 3. Finn frekvensfunktionen för $2X - 5$.

8.2 Låt X vara $N(0,1)$. Finn frekvensfunktionen för $Y = X^n$ för olika heltal $n > 1$.

8.3 Variabeln Y , $Y > 0$, är lognormalfördelad om $X = \log(Y)$ är normalfördelad $N(\mu, \sigma)$. Finn fördelningsfunktionen (uttryckt i den standardiserade normalfördelningens $\Phi(x)$) och frekvensfunktionen för Y och rita gärna den senare för något värde på parametrarna om du har tillgång till lämpligt datorprogram.

8.4 Två datordragna slumptal multipliceras. Finn frekvensfunktionen för produkten.

8.5 Tre oberoende exponentialfördelade variabler med $\mu = 2$ adderas. Vilken fördelning får resultatet?

8.6 Variablerna X_i är oberoende med samma fördelningsfunktion $F(x)$ och frekvensfunktion $f(x)$. a) Finn frekvensfunktionerna för $Y = \min X_i$ och $Z = \max X_i$. b) Finn den tvådimensionella frekvensfunktionen för (Y, Z) .

8.7 Vi behöver kunna simulera en variabel som har fördelningsfunktionen $F(x) = 1 - .5e^{-x} - .5e^{-2x}$, $x > 0$ och $F(x) = 0$, $x < 0$. Finn en lämplig transform av datorns $Re(0,1)$ -fördelade slumptal.

8.8 En partikels hastighetskomponenter i vinkelräta riktningar är oberoende $N(0, \sigma)$ -fördelade. Finn fördelningsfunktionen för kvadraten på hastighetsvektorn i ett plan, dvs $V_x^2 + V_y^2$.

8.9 Frekvensfunktionen f växer linjärt på intervallet $0 < x < b$ och är noll för övrigt. Finn sannolikhetsfördelningen för $Y = \ln f(X)$ om den stokastiska variabeln X har frekvensfunktionen f . Bestäm också väntevärdet för Y .

Svar och kommentarer till problem.

R1: Utfall, sannolikhet, kombinatorik

1.1 $A \cup B = \{2, 3, 4, 5, 6, 8, 10, 11\}$, $A \cap B = \{4\}$,
 $\overline{A} \cup \overline{B} = \{1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12\}$, $\overline{A} \cap \overline{B} = \{1, 7, 9, 12\}$,

1.2 —

1.3 Familj I: $\frac{1}{2}$, Familj II: $\frac{3}{8}$, Familj k: $\binom{2k}{k} \left(\frac{1}{2}\right)^{2k}$

1.4 1 partikel: $\Omega = \{\theta : 0 \leq \theta < 2\pi\}$,

2 partiklar: $\Omega = \{(\theta_1, \theta_2) : 0 \leq \theta_1, \theta_2 < 2\pi\}$.

1.5 $\Omega = N = \{1, 2, 3, \dots\}$.

1.6 $S = \{\text{Alla positiva rationella tal, } \infty, 0, 0/0\}$.

1.7 a) $S_a = \{0, 1, 2, 3, \dots\}$. b) $S_b = \{\omega; \omega = (u_1, u_2, \dots, u_{16}), u_i \in S_a\}$.

1.8 Flera möjligheter, t.ex. i polära koord. $S = \{(r_1, \phi_1, r_2, \phi_2, r_3, \phi_3), 0 \leq r_1 \leq r_2 \leq r_3 \leq R, 0 \leq \phi_i < 2\pi\}$, samt $(r_1, \phi_1, r_2, \phi_2)$, (r_1, ϕ_1) , \emptyset där de senare markerar 1, 2 eller 3 bommar. Villkoren på radierna behövs om varje träffbild skall kopplas till ett enda utfall. Om $r_i = r_{i+1}$ får vi dessutom kräva t.ex. $\phi_i \leq \phi_{i+1}$ för entydighet.

1.9 a) $\frac{3}{190} \approx 0.016$. b) —

1.10 $P(A) = \frac{1}{8}$, $P(B) = \frac{1}{8}$, $P(A \cap B) = \frac{1}{56}$, $P(A \cup B) = \frac{13}{56}$.

1.11 $6!/8! = 1/56$

1.12 $\frac{52!}{(13!)^4}$.

1.13 Varje tal i $(0,1)$ kan utvecklas som "decimaltal" fast i basen 6. 1:1-avbildning är möjlig, alltså lika mäktigt. Även $(-\infty, \infty)$ och R^n har samma mäktighet (vart n:e kast kan utnyttjas för avbildning till R^n).

1.14 $12/\binom{12}{4} = \frac{4}{165}$.

1.15 a) Utfall

AABB AABb (AAbB) AAbb
AaBB AaBb (AabB) Aabb
(aABB aABb aAbB aAbb)
aaBB aaBb (aabB) aabb

Parenteserna runt rad 3 och kolumn 3 markerar att aA och Aa resp. bB och Bb kan slås ihop till samma symbol som fdubbel chans när bara en av dem förenas och 4-dubbel när både aA och bB ingår. b) AaBB AaBb (AabB). c) Resten av utfallen. d) Kräver rätt stor matris och ges ej. Gör rutnät med alla utfallen ovan som kolumn (pappans anlag) och som rader (mammans anlag). Det ger 16 gånger 16 utfall (9 gånger 9 om man slår ihop tillstånd).

R2: Oberoende, beroende, betingning, Bayes

2.1 1%

2.2 0.2 och 0.7 eller omvänt.

2.3 $P(A|B) = 2P(B|A)$

2.4 $2/3$

2.5 0.1472, 0.7358, 0.1030, 0.0294

2.6 Starta med sannol. för 0 och för 1 fel. Svar: 0.5611, 0.9056, 0.4118, 0.1267

2.7 a) $(1-p)^3, 3p(1-p)^2, 3p^2(1-p), p^3$. b) $\frac{(3p(1-p)^2, 3p^2(1-p), p^3)}{1-(1-p)^3}$.

2.8 a) 0.0119, 0.0117. b) Här måste man pilla igenom alla sådana fall. Svar 0.1358.

2.9 Dubbelt så bra att byta dörr.

2.10 0.3623, 0.4058 resp. 0.2319.

2.11 $1/3$

2.12 $P(\text{"fel"}|\text{"försäljning"}) = \frac{10}{9712} \approx 0.00103$, $P(\text{"fel"}|\text{"kasserad"}) = \frac{90}{288} \approx 0.3125$.

2.13 $\frac{\binom{26}{13}}{\binom{26}{13} + \binom{25}{13}} = \frac{2}{3}$

2.14 $1 - \left(2\left(\frac{5}{6}\right)^{10} - \left(\frac{2}{3}\right)^{10}\right) \approx 0.69$

2.15 $\left(1 - \left(\frac{3}{4}\right)^3\right)^5 \approx 0.065$

2.16 Definiera händelserna: $S =$ söker vård, $N =$ narkoman. $P(S|N) = 0.042$.

2.17 $3/10$

R3: Stokastiska variabler

3.1 $f_X(x) = \begin{cases} 0.874, & x = 0 \\ 0.122, & x = 1 \\ 0.004, & x = 2 \\ 0, & \text{annars} \end{cases} \quad F_X(x) = \begin{cases} 0, & x < 0 \\ 0.874, & 0 \leq x < 1 \\ 0.996, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$

3.2 $E[X] = 2/3$, $D[X] = 1/\sqrt{18}$, median = $1/\sqrt{18}$.

3.3 $F_X(x) = \left(\frac{x}{N}\right)^4$; $f_X(x) = F_X(x) - F_X(x-1) = \frac{x^4 - (x-1)^4}{N^4}$, $x = 1, 2, 3, \dots, N$;
 $E[X] = \frac{2771}{5^4} \approx 4.43$

3.4 $\int f(x)dx = 1$ ger $k = 1/\beta^2$, $P(X > \beta) = e^{-0.5} \approx 0.61$

3.5 Bevis: —

3.6 Bevis: —

3.7 a) $P(0.5 < X < 0.75) = \frac{2}{\pi} \left(\arcsin \frac{3}{4} - \arcsin \frac{1}{2}\right) \approx 0.2066$; b) $E[X] = \frac{2}{\pi}$

3.8 Varians 2.92. (Standardavvikelse 1.708).

3.9 $\frac{1}{32}$

3.10 $F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - (1-x)^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases} \quad f_X(x) = \begin{cases} 2(1-x), & 0 \leq x \leq 1 \\ 0, & \text{f.ö.} \end{cases}$

3.11 —

3.12 $f(k) = \binom{7}{k} \binom{11}{5-k} / \binom{18}{5}$, $k = 0, 1, 2, 3, 4, 5$.

3.13 $F(x) = x^2/R^2$, $0 < x < R$, $F = 0$, $x \leq 0$, $F = 1$, $x \geq R$; $f(x) = 2x/R^2$, $0 < x < R$;
 $P(X > R/2) = 3/4$; $E[X] = 2R/3$.

3.14 $33.3 = 2 * 50/3$.

R4: Flera variabler, oberoende/beroende, linjära uttryck

4.1 a) $f_X(x) = e^{-x}$, $x > 0$, $f_Y(y) = ye^{-y}$, $y > 0$.

b) $P(X > 2|Y < 4) = \frac{e^2-3}{e^4-5} \approx 0.0885$.

c) Nej! ($f_X(x)f_Y(y) = e^{-x}ye^{-y} \neq e^{-y} = f_{XY}(x, y)$)

4.2 Origo i cirkelns mitt ger $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = R/\sqrt{2}$, $Cov(X, Y) = 0$, men variablerna är beroende.

4.3 a) Bevis: —. b) $\alpha = 1/3$ (ger variansen $\frac{2}{3}\sigma^2$).

4.4 $C[X, Y] = 0$. X och Y är beroende. Variablerna är alltså okorrelerade, men ej oberoende.

4.5 120 000

4.6 $\sigma = \sqrt{\frac{\sigma_1^2}{4} + \frac{\sigma_2^2}{8}} \approx 0.0866$

4.7 Väntevärde $121 * 0.9 = 108.9$. Standardavvikelse $0.7 * \sqrt{121} = 7.7$.

4.8 $n_1 = ac_1C \pm \sqrt{(ac_1C)^2 - aC}$, $n_2 = \frac{C-n_1c_1}{c_2}$, med $a = \frac{\sigma_1^2}{c_1^2\sigma_2^2 - c_1c_2\sigma_2^2}$.

4.9 $\frac{7 \cdot 15}{31} \approx 3.4$, $\frac{7 \cdot 12}{31} \approx 2.7$ resp. $\frac{7 \cdot 8}{31} \approx 1.8$.

R5: Olika fördelningar

5.1 a) $Bi(100, p)$; b) $(1-p)^k p$, $k = 0, 1, 2, \dots$ c) $\frac{\binom{30}{10}\binom{70}{10}}{\binom{100}{20}}$.

5.2 —

5.3 Strategin påverkar inte. Proportionen blir 51.4 % pojkar om alla har samma överlevnadschans.

5.4 $\binom{20}{13}0.3^{13}0.7^7 + \binom{20}{7}0.3^70.7^{13} \approx 0.1653$

5.5 Exponentialfördelning, poissonfördelning, binomialfördelning, hypergeometrisk fördelning.

5.6 $P(X \geq 2) = \frac{2}{3} - \frac{\ln 3}{3} \approx 0.30$

5.7 $X \in \text{Re}(L/2, L)$, $E[X] = \frac{3}{4}L$.

5.8 Ansätt Poissonprocesser ger exponentialfördelade tider $T_1 \in \text{Exp}(1/\lambda_1)$ och $T_2 \in \text{Exp}(1/\lambda_2)$ med $\lambda_1 = 2\lambda_2$. $P(T_2 < T_1) = 1/3$.

5.9 $\Phi\left(\frac{1-0.8}{\sqrt{0.02}}\right) - \Phi\left(\frac{0.5-0.8}{\sqrt{0.02}}\right) \approx 0.90$

5.10 0.8664

5.11 Väntevärdet μ skall vara $z_{0.98}\sigma = 2.05\sigma$ under gränsvärdet.

R6: Fördelningar (forts), Poissonprocess, MGF

6.1 $1 - F(2.45) = 0.007$, $F \in N(0, 1)$.

6.2 $P(\text{'Högst 3 olyckor'}) = 0.981$.

6.3 $P(X = 1 | X \geq 1) = 0.356$. (Jfr. $P(X = 1) = 0.298$.)

6.4 Differensen mellan 2 år får $\sigma \approx 32$.

6.5 $f(x_1|s) = 1/s$, $0 < x_1 < s$.

6.6 Bortse från ev restkö från föregående resa. $X =$ antal anlända under 10 minuter är $Po(5)$. Tabell eller $0.95 < \sum_0^K \frac{5^i}{i!}e^{-5}$, ger $K = 9$ minimum.

6.7 $P(\text{'Ingen mikroorganism'}) = e^{-2} \approx 0.1353$, Volymen $V \geq \frac{\ln 10}{2} \text{ml} \approx 1.15 \text{ml}$.

6.8 0.16

6.9 $m_X(t) = pe^t + 1 - p$, $E[X] = \frac{dm_X}{dt}(0) = p$, $E[X^2] = \frac{d^2m_X}{dt^2}(0) = p$.

6.10 Sätt $Y = \sum_{i=1}^n X_i \in \text{Bi}(n, p)$. MGF för Y blir $m_Y(t) = \prod_{i=1}^n m_X(t) = (pe^t + 1 - p)^n$

6.11 Ta t ex rekordhändelser med $n = 2$, $p_1 = 1$ och $p_2 = 1/2$. Då blir $P(X = 0) = 0$, $P(X = 1) = P(X = 2) = 1/2$. Detta är inte en binomialfördelning!

R7: Centrala gränsvärdesatsen, approximation av fördelningar

7.1 $P(\text{'sammanlagda betjäningstiden understiger 4 timmar'}) \approx \Phi(2) \approx 0.977$.

7.2 Tiden $t \approx \frac{114}{\nu}$.

7.3 Egentligen hypergeometrisk fördelning. Eftersom man drar så liten del av partiet blir de dragna nästan oberoende. $X \approx \text{Bi}(100, 0.15) \approx N(15, 3.57)$. $F((10.5 - 15)/3.57) = 0.1037$.

7.4 —

7.5 $F(-7.92) \approx 1.16 \cdot 10^{-15}$ (anropet normcdf(-7.92) i matlab). Samma storleksordning får man med en grov uppskattning av integralen av frekvensfunktionen.

7.6 $\mu_Y = -100$, $\sigma_Y^2 = 100$, approx. $N(-100, 10)$.

7.7 Ca 0.90. Normalapprox. av poissonprocessvariabler.

7.8 Summan Y får $\mu = 100$, $\sigma^2 = 50$. $P(Y \geq 115) \approx 1 - F((114.5 - 100)/\sqrt{50}) = 0.020$. F ur $N(0, 1)$.

R8: Funktioner av variabler

8.1 $Y = 2X - 5$, $f_Y(y) = \frac{2}{3}e^{-\frac{2}{3}(y+5)}$, $y > -5$.

8.2 $f_Y(y) = \begin{cases} \frac{1}{n\sqrt{2\pi}}y^{\frac{1}{n}-1}e^{-\frac{1}{2}y^{\frac{2}{n}}} & n \text{ udda} \\ \frac{2}{n\sqrt{2\pi}}y^{\frac{1}{n}-1}e^{-\frac{1}{2}y^{\frac{2}{n}}} & n \text{ jämn, } y \geq 0 \end{cases}$

8.3 $F_Y(y) = \Phi(\log y)$, $f_Y(y) = \frac{1}{y}\phi(\log y)$.

8.4 $U = XY$ med $X, Y \in \text{Re}(0, 1)$ ger $f_U(u) = -\ln u$, $0 < u \leq 1$.

8.5 $f_V(v) = \frac{v^2}{2\mu^3}e^{-v/\mu}$, $v > 0$

8.6 $f_Y(y) = nf(y)(1 - F(y))^{n-1}$, $f_Z(z) = nf(z)(F(z))^{n-1}$,
 $f_{Y,Z}(y, z) = n(n-1)f(y)f(z)(F(z) - F(y))^{n-2}$.

8.7 $X = -\ln\left(\frac{-1 + \sqrt{1-8(U-1)}}{2}\right)$ har fördelningen $F(x)$ om $U \in \text{Re}(0, 1)$.

Använd inversmetoden: $u = F(x) \Rightarrow x = F^{-1}(u) = -\ln\left(\frac{-1 + \sqrt{1-8(u-1)}}{2}\right)$.

8.8 $Z = V_x^2 + V_y^2$, $F_Z(z) = \frac{\sqrt{2\pi}}{\sigma}\Phi\left(\frac{z}{\sigma}\right)$.

8.9 $f_Y(y) = \frac{e^y(1+m)-m}{k}$, $k > 0$, $m \geq 0$, $\frac{kb^2}{2} + mb = 1$.