Kapitel 1

Repetition från Grundkursen

Denna föreläsning kommer mest att vara en repetition av stoff från grundkursen. Längden på detta dokument kan tyckas vara oproportionerligt långt i jämförelse med längden på en föreläsning. Detta är också tanken; jag har försökt lägga till så många kommentarer jag kommit på för att på så sätt ge er en möjlighet att läsa in er och förhoppningsvis få en god grund att stå på när vi börjar med stokastiska processer. Mycket nöje!

- Utfall, ω : resultatet av ett slumpmässigt försök.
- Utfallsrum, Ω : mängden av alla möjliga utfall.
- Händelse: en delmängd A till Ω ($A \subseteq \Omega$) kallas för en händelse.

Exempel 1 (Kast med sexsidig tärning).

Det är här naturligt att låta utfallsrummet vara $\Omega = \{1, 2, 3, 4, 5, 6\}$, *där siffrorna självklart betecknar antalet prickar "upp".*

Exempel på händelser är:

$$\begin{aligned}
A =& \{1, 2, 3\} = \{\textit{antalet prickar mindre eller lika med 3}\} \\
B =& \{1, 3, 5\} = \{\textit{antalet prickar udda}\} \\
C =& \{4, 5, 6\} = \{\textit{antalet prickar fler eller lika med 4}\} \\
D =& \{2\}
\end{aligned}$$

Observera att en och samma händelse kan beskrivas på olika sätt. Ytterligare beskrivningar på händelsen A från exemplet ovan är

$$A = \left\{ \frac{\text{antalet prickar}}{3} \leq 1 \right\} = \{\text{antalet prickar ej strikt fler än 3}\}$$

Nedan går vi igenom de vanligaste operationerna på mängder (och därmed händelser). Om inget annat anges, betecknar stora bokstäver (A, B etc) mängder i allmänhet.

- Komplement: vi skriver A^c för komplementet till A, dvs händelsen att A inte inträffar. I exemplet ovan är

$$A^c = \{\text{antalet prickar strikt fler än 3}\} = \{4, 5, 6\} = C$$

- Tomma mängden, $\emptyset$: händelsen att inget inträffar, $\emptyset = \Omega^c$
- Snitt ($\cap$) och union ($\cup$):
  $A \cap B$ händelsen att både A och B inträffar
  $A \cup B$ händelsen att A eller B, eller både A och B, inträffar
- Disjunkta (oförenliga) händelser: händelserna A och B sägs vara disjunkta om $A \cap B = \emptyset$. I exemplet ovan är bland annat D och C disjunkta.
- Delmängd, $\subseteq$: Vi skriver $A \subseteq B$ om $\omega \in A$ medför att $\omega \in B$, dvs B inträffar om A inträffar. I exemplet ovan är $D \subseteq A$
- Mängddifferens $\setminus$: $A \setminus B = A \cap B^c$ är händelsen att A, men inte B, inträffar.
  I exemplet ovan är $A \setminus B = \{1,2,3\} \setminus \{1,3,5\} = \{2\}$.

Observera (och verifiera själv som övning) att följande gäller för alla händelser A och B:

$$\begin{aligned}
A \cap B &\subseteq A \\
A &\subseteq A \cup B \\
A^c \cup A &= \Omega \\
A^c \cap A &= \emptyset \\
A \setminus B &\subseteq A \\
A \cup B &= \{A \setminus B\} \cup \{B \setminus A\} \cup \{A \cap B\}
\end{aligned}$$

Nu fortsätter vi med sannolikheter. Referenserna till satser och definitioner är till kursboken Stokastiska Processer av Patrik Albin.

Def 0.1 (Sannolikhetsmått).

Ett sannolikhetsmått är en funktion $\mathbf{P}$, definierad på händelser $A \subseteq \Omega$ så att

$$0 \leq \mathbf{P}(A) \leq 1, \quad \mathbf{P}(\emptyset) = 0, \ \mathbf{P}(\Omega) = 1, \ \textit{samt}$$
$$\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i) \ \ \textit{då} \ \ A_1, A_2, \ldots \subseteq \Omega \ \ \textit{är disjunkta}$$

Exempel 1 (forts). *Om vi antar att den sexsidiga tärningen är rättvis så har vi $\mathbf{P}(E) = \frac{1}{6}\ \#\{i \in \Omega : i \in E\}$ för $E \subseteq \Omega$, dvs antal "godkända" element ($\{7\}$ är inte ett godkänt element) i E dividerat med det totala antalet element i Ω. Speciellt: $\mathbf{P}(k) = 1/6$ för $k \in \{1, \ldots, 6\}$.*

För våra händelser A, B och C ger detta $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 1/2$ och $\mathbf{P}(A \cap B) = \mathbf{P}(\{1,3\}) = 1/3$, $\mathbf{P}(A \cup B) = \mathbf{P}(\{1,2,3,5\}) = 2/3$, etc för kombinationer.

Egenskaperna hos $\mathbf{P}$ i nedanstående sats är en god övning att kontrollera på egen hand hed hjälp av definitionen ovan samt vad du lärt dig om mängdoperationer.

Sats 0.1. *För ett sannolikhetsmått $\mathbf{P}$ gäller*

- $\mathbf{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbf{P}(A_i)$ *för disjunkta* $A_1, \ldots, A_n \subseteq \Omega$

- $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$ *för* $A, B \subseteq \Omega$
- $\mathbf{P}(A) \leq \mathbf{P}(B)$ *för* $A, B \subseteq \Omega$ *med* $A \subseteq B$
- $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$ *för* $A \subseteq \Omega$
- $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ *för* $A, B \subseteq \Omega$

Def 0.2 (Oberoende).

Händelserna A *och* B *är oberoende om* $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Definitionen av oberoende bör stämma överens med vad man vanligtvis menar med oberoende; sannolikheten för att båda händelserna skall inträffa är produkten av sannolikheterna för varje händelse för sig. Ett annat sätt att se det på är, huruvida den ena händelsen inträffar eller inte, påverkar inte möjligheten för den andra. Tolkningen blir kanske lättare dels efter nedanstående exempel, men kanske framförallt efter definitionen av betingad sannolikhet (se nedan).

Exempel 2. *Vi drar ett kort slumpmässigt från en kortlek (utan jokrar och med alla 52 korten tillgängliga). Låt* A=*{kortet är ett ess} och* B=*{kortet är svart}. Under antagandet om likformig slumpmässighet, dvs att alla de 52 korten i korleken har lika stor sannolikhet att dras, så får vi* $\mathbf{P}(A) = 4/52 = 1/13$ *och* $\mathbf{P}(B) = 26/52 = 1/2$*, eftersom det endast finns 4 ess, respektive 26 svarta kort, av 52 möjliga. Med samma resonemang kan vi räkna ut att snitthändelsen* $A \cap B$=*{kortet är svart och kortet är ett ess}* =*{kortet är antingen spader eller klöver ess} har sannolikhet* $\mathbf{P}(A \cap B) = 2/52 = 1/26$.

Vi ser att

$$\mathbf{P}(A \cap B) = \frac{1}{26} = \frac{1}{13 \cdot 2} = \mathbf{P}(A)\mathbf{P}(B),$$

vilket ger att händelserna A *och* B *är oberoende.*

Sats 0.2.

Om A *och* B *är oberoende, så är* A^c *och* B *också oberoende*

Def 0.3 (Betingad sannolikhet).

För $A, B \subseteq \Omega$ *med* $\mathbf{P}(A) > 0$*, så definieras den betingade sannolikheten för* <u>*B givet att A inträffar*</u>*, med beteckning* $\mathbf{P}(B|A)$*, enligt*

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

Sats 0.2.

För disjunkta $A_1, A_2, \ldots, A_n$ *så att* $\bigcup_{i=1}^n A_i = \Omega$ *och* $\mathbf{P}(A_i) > 0$*, så har vi*

1. $\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B|A_i)\mathbf{P}(A_i) \quad$ *för* $B \subseteq \Omega$
2. $\mathbf{P}(A_k|B) = \dfrac{\mathbf{P}(B|A_k)\mathbf{P}(A_k)}{\mathbf{P}(B)} \quad$ *för* $B \subseteq \Omega$ *med* $\mathbf{P}(B) > 0$

Första delen av Sats 0.2 är ett mycket viktigt verktyg för att beräkna sannolikheter av olika slag. För en komplicerad händelse B, så är det ibland möjligt att välja händelserna A_i (så att $\cup_i A_i = \Omega$) på ett sätt så att $\mathbf{P}(B|A_i)$ är mycket lätt att räkna ut. Vi kommer att använda detta flera gånger i denna kurs.

Betingad sannolikhet kan ses som om vi konstruerar ett nytt sannolikhetsmått $\mathbf{Q}(A) = \mathbf{P}(A|B)$ för $A \subseteq \Omega$, under villkoret att B har inträffat (eller "måste" inträffa), därav $\mathbf{P}(A \cap B)$ i täljaren (vilket ju står för sannolikheten att A och B inträffar), samt $\mathbf{P}(B)$ i nämnaren (vilket fungerar som en normalisering, så att $\mathbf{P}(\Omega|B) = 1$). Notera att händelser $E \subseteq B^c$ har sannolikheten noll betingat på att B har inträffat.

Som övning kan du verifiera att $\mathbf{Q}$ är ett sannoliketsmått på Ω (dvs kontrollera att villkoren i Def 0.1 är uppfyllda).

Notera:

- För A och B <u>oberoende</u>, så är

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

  vilket kan tolkas som att informationen om att B inträffat, inte är värdefull för oss då vi försöker bestämma sannolikheten för att A kommer att inträffa eller inte. Jämför med Exempel 2 där vi inte har någon nytta av att veta huruvida kortet var svart eller rött, då vi försöker ta reda på om det var ett ess.

- För <u>disjunkta</u> A och B, har vi

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(\emptyset)}{\mathbf{P}(B)} = 0,$$

  vilket bör vara uppenbart eftersom A och B ju inte kan inträffa "samtidigt". Notera att detta inte alls är icke-informativt; ekvationen säger ju faktiskt att händelse A inträffar med sannolikhet 0, om vi vet att B inträffat.

Stokastiska variabler

Från grundkursen kommer du kanske ihåg att en $\mathbb{R}$-värd stokastisk variabel ξ, är en funktion som till varje $\omega \in \Omega$ tilldelas ett tal i $\mathbb{R}$ dvs,

$$\xi : \Omega \mapsto \mathbb{R} \qquad \text{dvs} \quad \xi(\omega) \in \mathbb{R} \quad \text{för alla } \omega \in \Omega$$

Ofta upplever jag att många elever har svårt att förstå vad en stokastisk variabel är. (Kanske beror det på att det är ett nytt, krångligt ord?) I själva verket så borde stokastiska variabler vara mer greppbara än det underliggande utfallsrummet Ω. Det senare är ju bara något som existerar i tankevärlden i vår modell av hur slumpförsöket gick till; Ω blir en symbol för något slumpmässigt eller stokastiskt. Den stokastiska variabeln däremot, kan ses som det faktiska resultatet av slumpförsöket och det vi är intresserade av.

Exempel 1 (forts). *Om vi låter ξ vara antalet prickar uppåt på tärningen, så har vi*

$$\xi(\omega) = \omega \quad \forall \omega \in \Omega$$

för vårt val av Ω. Vi ser här att vi modellerade vårt utfallsrum direkt för vad vi var intresserade av; utfallet ω blir samma sak som värdet på den stokastiska variablen ξ.

Exempel 3 (Indikatorvariabel). *Om vi låter A vara en händelse om sätter*

$$I_A(\omega) = \begin{cases} 1 & \text{om } \omega \in A \\ 0 & \text{annars, dvs om } \omega \notin A \end{cases} \tag{1.1}$$

så ser vi att I_A är en stokastisk variabel, ty en funktion från Ω till $\mathbb{R}$. Den kallas för indikatorvariabeln för händelsen A.

Mer allmänt, kommer vi att betrakta $\mathbb{R}^n$-värda stokastiska variabler, dvs funktioner

$$\xi : \Omega \mapsto \mathbb{R}^n. \tag{1.2}$$

I princip finns det inget som hindrar oss från att definiera stokastiska variabler från Ω till ett valfritt rum. Ett direkt exempel är t ex att tillåta $\mathbb{C}$-värda stokastiska variabler (dvs komplex-värda). Ett annat, kanske mindre uppenbart, får vi om vi först låter $T = [0, \infty)$, och sedan definierar

$$\xi : \Omega \mapsto \mathbb{R}^T$$

så är ξ en oändligtdimensionell stokastisk variabel eller, om man så vill, en stokastisk process: för varje $\omega \in \Omega$ så har vi en $\mathbb{R}$-värd funktion $\xi(t)$ i variabeln $t \in T$.

Eftersom stokastiska processer generellt sätt är oändligtdimensionella stokastiska variabler, så det kan verka rimligt att man först lär sig behärska stokastiska variabler av ändlig dimension. Man har dessutom stor direkt nytta av dessa grundläggande kunskaper i analysen av stokastiska prcocesser vilket vi kommer att se senare.

Det bör påpekas att det normala språkbruket är att man med en stokastiska variabel menar ett stokastist objekt av ändlig dimension, typiskt enligt ekvation (1.2) ovan, men det är nyttigt att förstå att en stokastisk process endast är en generalisering av detta koncept.

Notera att händelser uttryckta för en stokastiska variabel ξ, t ex $\xi \in B$ för en delmängd $B \in \mathbb{R}^n$, överförs till händelser uttryckta som en delmängd av Ω [1] enligt

$$\{\xi \in B\} = \{\omega \in \Omega : \xi(\omega) \in B\} \subseteq \Omega$$

där uttrycket till höger uttalas: "de $\omega \in \Omega$ så att $\xi(\omega)$ tillhör B". ω-beroendet markeras sällan explicit, utan vi nöjer oss med att händelser uttrycks i ξ. Vi kommer sålunda att nöja oss med att skriva $\{\xi \leq 2\}$ då vi menar $\{\omega \in \Omega : \xi(\omega) \leq 2\}$, för en $\mathbb{R}$-värd stokastisk variabel ξ.

I fortsättningen kommer vi oftast att skriva s.v. för stokastisk variabel.

[1] Detta är egentligen ett krav på den stokastiska variabeln, att "lämpliga" delmängder av $\mathbb{R}^n$ (i detta fall), överförs till "lämpliga" delmängder av Ω, som vi kan beräkna sannolikheten för. Vi säger att den stokastiska variabeln skall vara mätbar. Detta är dock tekniska deltaljer som tillhör en fortsättningskurs i sannolikhetsteori.

För en $\mathbb{R}^n$-värd s.v. $\xi = (\xi_1, \ldots, \xi_n)$ och en vektor $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, låter vi $\{\xi \leq x\}$ beteckna

$$\{\xi \leq x\} = \{\xi_1 \leq x_1, \ldots, \xi_n \leq x_n\}$$

Def 0.4 (Fördelningsfunktion).

Fördelningsfunktionen F_ξ för en $\mathbb{R}^n$-värd s.v. ξ defineras enligt:

$$F_\xi(x) = \mathbf{P}(\xi \leq x) = \mathbf{P}(\xi_1 \leq x_1, \ldots, \xi_n \leq x_n)$$

för $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Exempel 1 (forts).*Den stokastiska variabeln $\xi = \{antalet\ prickar\ upp\}$ enligt tidigare har fördelningsfunktion*

$$F_\xi(x) = \begin{cases} 0 & \text{för } x < 1 \\ \frac{k}{n} & \text{för } k \leq x < k+1 \text{ och } k \in \{1, \ldots, 5\} \\ 1 & \text{för } x \geq 6 \end{cases}$$

Exempel 3 (forts).*Indikatorvariabeln I_A för händelsen A har fördelningsfunktionen*

$$F_{I_A}(x) = \begin{cases} 0 & \text{för } x < 0 \\ 1 - \mathbf{P}(A) & \text{för } 0 \leq x < 1 \\ 1 & \text{för } x \geq 1 \end{cases}$$

Följande enkla resultat är nyttigt att ha i åtanke när man jobbar med fördelningsfunktioner:

Sats 0.5 (ena riktningen).

En fördelningsfunktion F_ξ för en $\mathbb{R}$-värd s.v. är

- *växande, dvs $F_\xi(x) \leq F_\xi(y)$ för $x \leq y$*
- *högerkontinuerlig, dvs $\lim_{y \downarrow x} F_\xi(y) = F_\xi(x)$, samt har följande gränsvärden*
- $\lim_{x \to -\infty} F_\xi(x) = 0$ *och* $\lim_{x \to \infty} F_\xi(x) = 1$

Diskreta stokastiska variabler

En stokastisk variabel sägs ξ sägs vara diskret om mängden av möjliga värden Ω_ξ är diskret, t ex $\mathbb{N}$ eller $\mathbb{Z}$. I så fall definieras dess frekvensfunktion enligt

$$f_\xi(x) = \mathbf{P}(\xi = x) = \mathbf{P}(\xi_1 = x_1, \ldots, \xi_n = x_n)$$

För $x \in \mathbb{R}^n \setminus \Omega_\xi$ låter vi $f_\xi(x) = 0$.

Exempel 1 (forts).*I vårt tärningsexempel så är givetvis ξ diskret ty $\Omega_\xi = \{1, \ldots, 6\}$. Dess frekvensfunktion är*

$$f_\xi(x) = \mathbf{P}(\xi = x) = \frac{1}{6} \qquad \text{för } x \in \Omega_\xi$$

Notera att

$$f_\xi(x) \geq 0 \qquad \forall x \in \mathbb{R}^n$$

samt att

$$\sum_{x\in\mathbb{R}^n} f_\xi(x) = \sum_{x\in\Omega_\xi} f_\xi(x) = 1$$

pga definitionen av sannolikhetsmått.

Kontinuerliga stokastiska variabler

En stokastisk variabel sägs ξ sägs vara kontinuerlig[2] om dess fördelningsfunktion kan skrivas

$$F_\xi(x) = \mathbf{P}(\xi \le x) = \int_{y\le x} f_\xi(y)\,dy$$

för någon icke-negativ funktion f_ξ (där integralen skall tolkas som en n-dimensionell integral). I så fall sägs f_ξ vara dess frekvensfunktion (eller täthetsfunktion).

Notera att vi även här har

$$f_\xi(x) \ge 0 \qquad \forall x \in \mathbb{R}^n$$

samt det motsvarande

$$\int_{x\in\mathbb{R}^n} f_\xi(x) = \mathbf{P}(\xi \in \mathbb{R}^n) = 1$$

Sats 0.6 och 0.9.

För en mängd $A \subseteq \mathbb{R}^n$ och en $\mathbb{R}^n$-värd s.v. ξ kan vi beräkna $\mathbf{P}(\xi \in A)$ genom

$$\mathbf{P}(\xi \in A) = \begin{cases} \displaystyle\sum_{x\in A} f_\xi(x) = \sum_{x\in A} \mathbf{P}(\xi = x) \\ \\ \displaystyle\int_{x\in A} f_\xi(x)\,dx \end{cases}$$

beroende på om ξ är diskret eller kontinuerlig.

Vi har dessutom:

Sats 0.8

För en kontinuerlig $\mathbb{R}^n$-värd s.v. är $\mathbf{P}(\xi = x) = 0$ för $x \in \mathbb{R}^n$.

Sannolikheten för att ξ skall anta ett specifikt värde x säger således ingenting för en kontinuerlig stokastisk variabel (ty den är noll för alla x). Däremot kan man fråga sig vad sannolikheten är att den skall anta värden i ett interval kring x:

$$\mathbf{P}(\xi \in (x - \Delta x, x + \Delta x)) = \int_{x-\Delta x}^{x+\Delta x} f_\xi(y)\,dy$$

Varning: Det finns stokastiska variabler (fördelningar) som varken är kontinuerliga eller diskreta.

[2] Denna definition skiljer sig från kursbokens. Denna är dock den brukliga och de är faktiskt ekvivalenta, så vi kan lika gärna använda denna.

Väntevärde, varians och kovarians

Def 0.7 (Väntevärde)

Väntevärdet $\mathbf{E}\{\xi\}$ *för en* $\mathbb{R}$*-värd s.v. definieras som*

$$\mathbf{E}\{\xi\} = \begin{cases} \displaystyle\sum_{x \in \mathbb{R}} x\, f_\xi(x) \\ \\ \displaystyle\int_{x \in \mathbb{R}} x\, f_\xi(x)\, dx \end{cases}$$

beroende på om ξ *är diskret eller kontinuerlig. Ovanstående gäller förutsatt att summan respektive integralen är absolutkonvergent. Om detta inte gäller, existerar inte väntevärdet och vi sätter* $\mathbf{E}\{\xi\} = \infty$*.*

Observera att väntevärdet av en $\mathbb{R}$-värd s.v. är ett reellt tal; inget stokastisk över detta alltså, vi har "integrerat bort" allt stokastiskt.

Följande sats är mycket viktig och kan sägas vara en generalisering av definitionen:

Sats 0.12

Om g *är en funktion* $g : \mathbb{R} \mapsto \mathbb{R}$ *så har vi*

$$\mathbf{E}\{g(\xi)\} = \begin{cases} \displaystyle\sum_{x \in \mathbb{R}} g(x)\, f_\xi(x) \\ \\ \displaystyle\int_{x \in \mathbb{R}} g(x)\, f_\xi(x)\, dx \end{cases}$$

återigen förutsatt att vi har absolutkonvergens.

Exempel på vanliga $g(x)$ är x^2 och $|x|$.

Exempel 1 (forts.). *Väntevärdet av* $\xi = \{$*antal prickar upp*$\}$ *är*

$$\mathbf{E}\{\xi\} = \sum_{k=1}^{6} k\, \mathbf{P}(\xi = k) = \sum_{k=1}^{6} k\, \frac{1}{6} = 3\, \frac{1}{2}$$

vilket de flesta lär sig redan som barn även om man inte kallade det för väntevärde då...

Exempel 3 (forts.). *Väntevärdet för indikatorn* I_A *är*

$$\mathbf{E}\{I_A\} = 0 \cdot \mathbf{P}(I_A = 0) + 1 \cdot \mathbf{P}(I_A = 1) = \mathbf{P}(I_A = 1) = \mathbf{P}(A) \tag{1.3}$$

eftersom I_a *endast antar värdena 0 och 1, samt att* $\{I_A = 1\} = \{\omega \in A\}$ *enligt definitionen av* I_A*, se ekvation* (1.1)*.*

Notera att för diskreta s.v. så sammanfaller väntevärdet med medelvärdet av värdena som ξ kan anta om sannolikheten att anta ett värde är lika stor för alla värden. I allmänhet för diskreta s.v., kan man tolka väntevärdet som ett viktat medelvärde, där vikten för varje värde x är $\mathbf{P}(\xi = x)$. För indikatorvariabeln ser vi att väntevärdet skiljer sig från medelvärdet 1/2 så fort $\mathbf{P}(A) \neq 1/2$.

Def 0.8 (Varians och kovarians).

Kovariansen mellan två $\mathbb{R}$-värda s.v. ξ och η definieras

$$\mathbf{Cov}\{\xi,\eta\} = \mathbf{E}\big\{(\xi - \mathbf{E}\{\xi\})(\eta - \mathbf{E}\{\eta\})\big\}.$$

Variansen är för en s.v. ξ definieras

$$\mathbf{Var}\{\xi\} = \mathbf{Cov}\{\xi,\xi\} = \mathbf{E}\big\{(\xi - \mathbf{E}\{\xi\})^2\big\}$$

Definitionen av variansen kan läsas som "den förväntade kvadratavvikelsen från väntevärdet", och är således ett mått på spridningen av ξ från dess väntevärde. Kovariansen mellan två stokastiska variabler är ett mått på (linjärt) beroende mellan de två.

Genom att utveckla produkterna i Def 0.8, får vi Sats 0.14

$$\begin{aligned}\mathbf{Cov}\{\xi,\eta\} =&\mathbf{E}\big\{(\xi - \mathbf{E}\{\xi\})(\eta - \mathbf{E}\{\eta\})\big\}\\ =&\mathbf{E}\big\{\xi\eta - \mathbf{E}\{\xi\}\eta - \mathbf{E}\{\eta\}\xi + \mathbf{E}\{\eta\}\mathbf{E}\{\xi\}\big\}\\ =&\mathbf{E}\{\xi\eta\} - \mathbf{E}\{\xi\}\mathbf{E}\{\eta\}\end{aligned} \tag{1.4}$$

och således

$$\mathbf{Var}\{\xi\} = \mathbf{E}\{\xi^2\} - (\mathbf{E}\{\xi\})^2 \tag{1.5}$$

där vi i (1.4) först utvecklade produkten innanför det yttersta väntevärdet och därefter utnyttjade linjäriteten hos väntevärdet (Sats 0.13):

$$\mathbf{E}\{a_1\xi_1 + \ldots + a_n\xi_n\} = a_1\mathbf{E}\{\xi_1\} + \ldots + a_n\mathbf{E}\{\xi_n\} \tag{1.6}$$

för $\mathbb{R}$-värda stokastiska variabler $\xi_1,\ldots,\xi_n$ och tal $a_1,\ldots,a_n \in \mathbb{R}$.

Ekvation (1.5) används ofta för att beräkna variansen för stokastiska variabler där $\mathbf{E}\{\xi^2\}$ beräknas mha Sats 0.12.

Exempel på fördelningar

Här kommer en repetition av några fördelningar som du kanske kommer ihåg från grundkursen (även om namnen på några kan vara skiljda från de du lärt dig).

Bernoullifördelningen

Om vi låter ξ vara 1 med sannolikhet p, och 0 med sannolikhet $1-p$, säger vi att ξ är Bernoullifördelad med parameter p. För detta skriver vi $\xi \sim \text{Bernoulli}(p)$.

Exempel 3 (forts.).

Enligt ovanstående definition är en indikatorvariabel I_A för en händelse A, Bernoullifördelad med parameter $p = \mathbf{P}(A)$.

Enligt beräkningen (1.3) så är $\mathbf{E}\{\xi\} = p$. Om vi använder Sats 0.12 för att beräkna $\mathbf{E}\{\xi^2\} = p$ så får vi

$$\mathbf{Var}\{\xi\} = \mathbf{E}\{\xi^2\} - (\mathbf{E}\{\xi\})^2 = p - p^2 = p(1-p)$$

med hjälp av (1.5) ovan.

Binomialfördelningen

Om vi gör n stycken oberoende försök, där vart och ett har sannolikheten p att lyckas, och låter ξ vara summan av de lyckade försöken, så får ξ frekvensfunktion

$$f_\xi(x) = \mathbf{P}(\xi = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{för } x \in \{0, \dots, n\}$$

där $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, den så kallade binomialkoefficient. Uttrycket kommer sig av att $p^x(1-p)^{n-x}$ är sannolikheten för en specifik konfiguration av av x lyckade och $n-x$ misslyckade försök. Eftersom det finns $\binom{n}{x}$ olika sådana konfigurationer så får vi resultatet.

Vi säger att ξ är binomialfördelad och skriver $\xi \sim \text{Bin}(n, p)$.

Fördelningsfunktionen skriver vi

$$F_\xi(x) = \mathbf{P}(\xi \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$$

för $x \in [0, n]$, där vi med $\lfloor x \rfloor$ menar heltalsdelen av x. För $x < 0$ och $x > n$ är $F_\xi(x) = 0$ respektive 1.

Observera att vi kan skriva ξ som en summa av n stycken oberoende[3] Bernoulli(p)-variabler $\xi_1, \dots, \xi_n$

$$\xi = \xi_1 + \ldots + \xi_n$$

Härigenom blir väntevärdes beräkningen mycket lätt med hjälp av ekvation (1.6):

$$\mathbf{E}\{\xi\} = \mathbf{E}\{\xi_1 + \ldots + \xi_n\} = \mathbf{E}\{\xi_1\} + \ldots + \mathbf{E}\{\xi_n\} = np$$

För varianser behöver vi "låna" ett resultat av Sats 0.21 som säger att variansen av en summa av oberoende s.v. är summan av dess varianser. Sålunda har vi

$$\mathbf{Var}\{\xi\} = \mathbf{Var}\{\xi_1 + \ldots + \xi_n\} = \mathbf{Var}\{\xi_1\} + \ldots + \mathbf{Var}\{\xi_n\} = np(1-p)$$

med hjälp av variansen för en Bernoullifördelad s.v.

Exempel 1 (forts.). *Om vi kastar en sexsidig tärning 3 gånger och låter η vara antalet gånger vi får en sexa, så är $\eta \sim$ Bin$(3, \frac{1}{6})$ med frekvensfunktion*

$$\mathbf{P}(\xi = x) = \begin{cases} \binom{3}{0} p^0 (1-p)^3 = (1-p)^3 = \frac{5^3}{6^3} & \text{för } x = 0 \\ \binom{3}{1} p^1 (1-p)^2 = 3p(1-p)^2 = \frac{3 \cdot 5^2}{6^3} & \text{för } x = 1 \\ \binom{3}{2} p^2 (1-p)^1 = 3p^2(1-p) = \frac{3 \cdot 5}{6^3} & \text{för } x = 2 \\ \binom{3}{3} p^3 (1-p)^0 = p^3 = \frac{1}{6^3} & \text{för } x = 3 \end{cases}$$

Det förväntade antalet sexor på tre kast är

$$\mathbf{E}\{\xi\} = np = 3 \cdot \frac{1}{6} = \frac{1}{2}$$

[3] Notera att vi har inte gått igenom vad oberoende s.v. är ännu, bara vad oberoende händelser är. Definitionen är dock väldigt lik den för händelser och jag tror nog att du begriper (intuitivt) vad som menas med det...

Geometrisk fördelning

Betrakta en händelse som har sannolikhet p att lyckas och att vi upprepar försöken tills händelsen inträffar (och att vi antar att varje försök är oberoende av de föregående försöken).

Låt ξ beteckna antalet försök tills händelsen inträffar (dvs tills vi "lyckas" första gången). Då sägs ξ vara geometriskt fördelad med parameter p, beteckning $\xi \sim \text{Geo}(p)$. Frekvensfunktionen blir

$$f_\xi(x) = \mathbf{P}(\xi = x) = (1-p)^{x-1}p \qquad \text{för } x \in 1, 2, \ldots$$

där sannolikheten kommer sig av att vi har $x-1$ misslyckade försök, med sannolikhet $1-p$ vardera, samt ett lyckat.

Fördelningsfunktionen blir

$$F_\xi(x) = \sum_{k=1}^{\lfloor x \rfloor}(1-p)^{k-1}p = p\sum_{k=1}^{\lfloor x \rfloor}(1-p)^{k-1}$$

för $x \geq 0$. Vi ser att summan är en geometrisk summa, vilket är förklaringen till fördelningens namn.

Dess väntevärde är $\mathbf{E}\{\xi\} = \frac{1}{p}$, vilket är vad man hade sagt utan någon som helst sannolikhetsteori; om något har sannolikhet p att lyckas så får man i genomsnitt vänta $1/p$ innan det inträffar första gången (testa med $p = 1/6$, första gången man får en sex med en sexsidig tärning).

Variansen är $\mathbf{Var}\{\xi\} = \frac{1-p}{p^2}$, vilket kanske inte är lika lätt att tolka...

Notera skillnaden mot binomialfördelningen där vi först fixerar antalet försök n, och sedan räknar hur många av dessa som lyckades. I den geometriska fördelningen fortsätter vi försöken endast tills vi lyckas första gången.

Diskret likformig

Om vi, för givet n, låter

$$\mathbf{P}(\xi = x) = \frac{1}{n} \qquad \text{för } x \in \{1, \ldots, n\},$$

dvs drar ett element från n möjliga med lika stor sannolikhet för varje val, så säger vi att ξ är likformigt fördelad över de n elementen. Detta är således en generalisering av ξ i tärningsexemplet, där $n = 6$, till kast med en n-sidig tärning.

Ofta uttrycks detta "välj ett av de n elementen på måfå", där "på måfå" menas likformigt.

Väntevärdet:

$$\mathbf{E}\{\xi\} = \sum_{k=1}^{n} k\,\mathbf{P}(\xi = k) = \sum_{k=1}^{n} k\,\frac{1}{n} = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$$

Dessutom har vi

$$\mathbf{E}\{\xi^2\} = \frac{1}{n}\sum_{k=1}^{n} k^2 = \Big[\text{ använd Beta! }\Big] = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

så att

$$\mathbf{Var}\{\xi\} = \frac{(n+1)(2n+1)}{6} - \Big(\frac{n+1}{2}\Big)^2 = \frac{n^2-1}{12}$$

Exponentialfördelningen

Om ξ är en kontinuerlig s.v. med frekvensfunktion

$$f_\xi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{för } x \geq 0 \\ 0 & \text{för } x < 0 \end{cases}$$

för något $\lambda > 0$, så sägs ξ vara exponentialfördelad med parameter (eller intensitet) λ. Detta betecknas $\xi \sim \exp(\lambda)$.

Dess fördelningsfunktion är

$$F_\xi(x) = \mathbf{P}(\xi \leq x) = \int_{-\infty}^{x} f_\xi(y)\,dy = \int_0^x \lambda e^{-\lambda y}\,dy = \Big[-e^{-\lambda x}\Big]_0^x = 1 - e^{-\lambda x}$$

för $x \geq 0$ och $F_\xi(x) = 0$ för $x < 0$, där första likheten är definitionen av fördelningsfunktion och andra likheten av Sats 0.9 (egentligen följer de två första likheterna av vår definition på en kontinuerlig s.v.).

Väntevärde är

$$\mathbf{E}\{\xi\} = \int_{-\infty}^{\infty} x\, f_\xi(x)\,dx = \int_{-\infty}^{\infty} x\,\lambda e^{-\lambda x}\,dx = \Big[\text{Partiell int.}\Big] = \frac{1}{\lambda}.$$

Dessutom:

$$\mathbf{E}\{\xi^2\} = \int_{-\infty}^{\infty} x^2\, f_\xi(x)\,dx = \int_{-\infty}^{\infty} x^2\,\lambda e^{-\lambda x}\,dx = \Big[\text{Partiell int. två ggr}\Big] = \frac{2}{\lambda^2}$$

så

$$\mathbf{Var}\{\xi\} = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

återigen med hjälp av ekvation (1.5).

Exponentialfördelningen används ofta för att modellera betjäningstider (t ex den tid det tar för att få prata med en telefonoperatör när man ringer till CSN), samt livlängder för mekaniska och elektriska komponenter.

Kontinuerligt likformig

Låt $a < b$. Vi säger att ξ är likformigt fördelad över intervallet $[a, b]$ om

$$f_\xi(x) = \begin{cases} \frac{1}{b-a} & \text{för } x \in [a, b] \\ 0 & \text{för övriga} \end{cases}$$

Vi kommer att använda beteckningen $\mathrm{U}[a, b]$ för denna fördelning (som i *uniform* på engelska).

Även om $\mathbf{P}(\xi = x)$ saknar innebörd för en kontinuerlig s.v. (eftersom $\mathbf{P}(\xi = x) = 0$ för alla x enligt Sats 0.8), så kan man tänka på en likformig s.v. som att vi väljer ett reellt tal mellan a och b med lika stor sannolikhet, eller "på måfå".

Väntevärde och variansen beräknas lätt, mha ekvation (1.5), till

$$\mathbf{E}\{\xi\} = \frac{a+b}{2}, \quad \text{samt} \quad \mathbf{Var}\{\xi\} = \frac{(b-a)^2}{12}.$$

Poissonfördelningen

Låt $\lambda > 0$. Om ξ är en $\mathbb{R}$-värd dikret s.v. med frekvensfunktion

$$f_\xi(x) = \mathbf{P}(\xi = x) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{för } x \in \{0, 1, 2, \ldots\}$$

så säger vi att ξ är Poissonfördelad med intensitet λ. Beteckning $\xi \sim \mathrm{Po}(\lambda)$.

Väntevärdet är

$$\mathbf{E}\{\xi\} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \Big[n = k-1\Big] == e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^{n+1}}{n!} = \lambda$$

ty $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$ är Taylorutveckling för e^λ. Dessutom har vi

$$\mathbf{E}\{\xi^2\} = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{n=0}^{\infty} (n+1) \frac{\lambda^{n+1}}{n!} = \lambda(\lambda+1)$$

så att $\mathbf{Var}\{\xi\} = \lambda(\lambda+1) - \lambda^2 = \lambda$.

Sist men inte minst betydelsefull:

Normalfördelningen

Om ξ är en kontinuerlig $\mathbb{R}$-värd s.v. med frekvensfunktion

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{för } x \in \mathbb{R}$$

för något $\mu \in \mathbb{R}$ och $\sigma > 0$, så sägs ξ vara normalfördelad (eller Gaussiskt fördelad) med väntevärde μ och varians σ^2, beteckning $\mathrm{N}(\mu, \sigma^2)$.

En normalfördelad s.v. med $\mu = 0$ och $\sigma = 1$ sägs vara <u>standard normalfördelad</u>.

Föga förvånande, med tanke på namnen på parametrarna till fördelningen, så är $\mathbf{E}\{\xi\} = \mu$ och $\mathbf{Var}\{\xi\} = \sigma^2$. Detta är dock inget självklart, utan kräver sin beräkning (med bl a lösningen av en standardintegral från Matematisk Analys).

Det finns inget slutet uttryck för fördelningsfunktionen, utan man får förlita sig till datorn eller tabeller för detta (det senare fick du säkert öva näst intill förbannelse i grundkursen...).

Att normalfördelningen är viktig kommer sig av <u>Centrala Gränsvärdessatsen</u> (CGS), som säger lite löst att summan av n stycken oberoende stokastiska variabler $\xi_1, \ldots, \xi_n$ är approximativt normalfördelad för stora värden på n, under det enda villkoret att $\mathbf{Var}\{\xi_i\} = \sigma^2 < \infty$. (Variablerna ξ_i var för sig behöver alltså inte vara likafördelade, dvs ha samma fördelning.)

Kan man således motivera att "något slumpmässigt" tillkommit genom addition av en mängd olika, oberoende "slumpmässigheter" (och som beter sig enligt förutsättningarna för CGS), så har man motiverat användandet av normalfördelningen för detta.

Att normalfördelade variabler och processer dessutom ofta är lätta att räkna på (i alla fall i förhållande till mycket annat), bidrar ytterligare till normalfördelningens popularitet. Just på grund av detta skall man också kanske vara lite extra kritisk till en eventuell motivering enligt föregående stycke.

Som ett sista exempel skall jag visa användbarheten av Sats 0.2.

Exempel 4 .

Antag att vi är med i ett lotteri där man varje insats får ett Po(λ)-fördelat antal lotter. Varje lott är, oberoende av de andra lotterna, en vinstlott med sannolikheten $p \in (0,1)$. Låt η beteckna antalet vinstlotter och beräkna sannolikheten $\mathbf{P}(\eta = k)$.

Inför beteckningen ξ för antalet lotter. (Skilj på lotter och vinstlotter, en lott är endast vinstlott med sannolikhet p.)

Om ξ inte vore stokastiskt utan ett fixt tal, säg n, så hade det varit lätt att beräkna sannolikheten för k vinstlotter; i så fall är ju $\eta \sim$ Bin(n, p). Eller hur? (Titta på definitionen av binomialfördelningen.)

Vi har således kommit fram till att

$$\mathbf{P}(\eta = k \,|\, \xi = n) = \binom{n}{k} p^k (1-p)^{n-k} \quad \textit{för } k \in \{0, 1, \ldots, n\}.$$

Eftersom ξ är Po(λ)-fördelad, så $\mathbf{P}(\xi = n) = \frac{\lambda^n}{n!} e^{-\lambda}$, och eftersom $\bigcup_{n=0}^{\infty} \{\xi = n\} = \Omega$ (vi måste ju få något antal lotter!), kan vi applicera Sats 0.2:

För $k \in \{0, 1, \ldots\}$ gäller

$$\begin{aligned}
\mathbf{P}(\eta = k) &= \sum_{n=0}^{\infty} \mathbf{P}(\eta = k \,|\, \xi = n)\mathbf{P}(\xi = n) = \Big[\mathbf{P}(\eta = k \,|\, \xi = n) = 0 \textit{ för } n < k\Big] \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \frac{(p\lambda)^k}{k!} \sum_{n=k}^{\infty} \frac{((1-p)\lambda)^{n-k}}{(n-k)!} = \Big[m = n-k\Big] = \\
&= e^{-\lambda} \frac{(p\lambda)^k}{k!} \sum_{m=0}^{\infty} \frac{((1-p)\lambda)^m}{m!} = \Big[\textit{Summan är Taylorutvecklingen av } e^{(1-p)\lambda}\Big] \\
&= e^{-\lambda} \frac{(p\lambda)^k}{k!} e^{(1-p)\lambda} = \frac{(p\lambda)^k}{k!} e^{-p\lambda}
\end{aligned}$$

vilket identifieras som frekvensfunktionen för en Po($p\lambda$)-fördelad s.v.

Detta kallas för en uttunnad Poissonfördelning.

Räkningarna ovan tillhör kanske inte de lättaste, men Sats 0.2 gjorde det i alla fall möjligt att beräkna den sökta sannolikheten och detta är ett typiskt exempel på dess tillämpning.