Ett *slumpmässigt försök* är den ram vi lägger runt våra sannolikhetsberäkningar. Vi observerar *variabler* men det vi räknar ut sannolikheter för är händelser. En *händelse* ("event") är något som antingen inträffar eller inte inträffar då vi utför ett slumpmässigt försök. Ofta använder vi ordet *utfall* ("sample point, outcome") för resultatet av ett slumpmässigt försök. Mängden av alla möjliga utfall kallar vi för försökets *utfallsrum* ("sample space"). Det är inget annat än variabelns värdemängd.

Tänk dig att du utför ett slumpmässigt försök som består i att observera en variabel $x$. När du gjort försöket har du erhållit ett utfall som vi också kallar $x$. Tänker du göra fler försök, så är det naturligt att låta $x_1$ beteckna det första försöksresultatet, $x_2$ det andra, etc. Låt $X$ vara motsvarande stokastiska variabel.

Variabeln $x$ kan egentligen vara av vilken matematisk typ som helst, men vi ska tänka oss den reellvärd. Utfallsrummet är då en delmängd av eller hela $R$. Låt $A \subseteq R$. Utsagan $X \in A$ är en händelse, ty efter det att försöket har utförts vet vi om försöksresultatet (dess utfall) $x$ tillhör $A$ eller ej. Alla händelser kan kopplas ihop med en mängd $A$ på detta sätt. Ofta studeras endast en variabel. Då är den ofta underförstådd och man skriver $A$ för händelsen och säger att $A$ inträffar istället för $X \in A$.



Tänk dig nu att vi ska utföra ett slumpmässigt försök. Det kan vara allt ifrån att man observerar hur många ägg ett visst svalpar lägger i boet till en avancerad mätning av mängden ozon i luften vi andas..

*Utfallsrummet* $S$ är mängden av alla möjliga försöksresultat. Elementen $x$ i $S$ kallas för *utfall*. Så $x$ är variabeln vi mäter och till den hör en stokastisk variabel $X$. Delmängderna $A, B, \ldots$ till $S$ kallas för *händelser*. Händelsen $A$ utläses "$A$ inträffar". Den kan också skrivas $X \in A$ eller $x \in A$.

Sannolikheten att $A$ inträffar skrives

$$P(A) \text{ eller } P(X \in A) \text{ eller } P(x \in A).$$

Låt $A, B$ vara händelser. Då betecknar

$A \cap B$ händelsen "$A$ och $B$"

$A \cup B$ händelsen "$A$ eller $B$"

$A^c$ eller $A'$ händelsen "icke-$A$"

Dessutom definierar vi

$A \setminus B = A \cap B^c$ (utläses $A$ minus $B$ eller $A$ men ej $B$)

$A \triangle B = A \setminus B \cup B \setminus A$ (symmetriska differensen)

Observera att $A \triangle B$ inträffar om, och endast om, exakt en av händelserna $A$ och $B$ inträffar.



Händelserna $A$ och $B$ sägs vara *ömsesidigt uteslutande* eller *disjunkta*, om

$$A \cap B = \emptyset$$

Händelserna $A_1, \ldots, A_n$ *partitionerar* en händelse $B$, om

$$B \subseteq \bigcup_{i=1}^{n} A_i$$

och

$$A_i \cap A_j = \emptyset \quad \text{då} \quad i \neq j$$

Istället för partitionerar kan man säga *delar in i disjunkta eller ömsesidigt uteslutande delar.*



**Exempel 1** Förorenad mark. Låt $y$ vara koncentrationen av någon giftig tungmetall. Låt $x$ vara resultatet av en mätning av $y$.

Marken betraktas som förorenad om $y \geq c$, där $c$ är ett av myndigheterna specificerat kritiskt värde. Denna händelse betecknas $C$.

Låt $d$ vara den s k detekteringsnivån (obs att typiskt är $d \neq c$). Händelsen $x \geq d$ betecknas $D$.

**Sannolikhetsaxiomen (s 190-191)**

Låt $A, B$ vara två godtyckliga händelser i utfallsrummet $S$. Då

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. $P(A \cup B) = P(A) + P(B)$ om $A, B$ är ömsesidigt uteslutande

**Sats 1:** $P(A \setminus B) = P(A) - P(A \cap B)$

**Följd 1:** $P(\emptyset) = 0$

**Följd 2:** $P(A^c) = 1 - P(A)$

**Följd 3:** $A \subseteq B \Rightarrow P(A) \leq P(B)$

**Sats 2:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

**Följd 1:** $P(A \cup B) \leq P(A) + P(B)$



**Exempel 2** Förorenad mark (forts)

Antag $P(C) = 0.8$, $P(D) = 0.74$ och $P(C \cap D) = 0.72$. Beräkna

1. $P(C \cup D)$
2. $P(C \setminus D)$
3. $P(D \setminus C)$
4. $P(C \triangle D)$

**Hemövning 1** Antag istället att $P(C) = 0.9$, $P(D) = 0.7$ och $P(C \cap D) = 0.65$. Beräkna $P(C \cup D)$, $P(C \setminus D)$, $P(D \setminus C)$ och $P(C \triangle D)$.



Den *betingade sannolikheten för A givet B*, är

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

Observera att

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

Händelserna $A$ och $B$ sägs vara *oberoende*, om

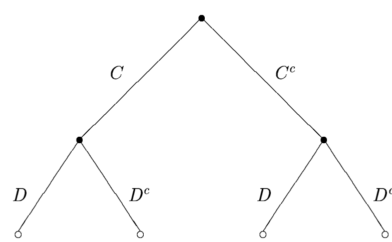
$$P(A \cap B) = P(A) \cdot P(B)$$

Observera att $A, B$ är oberoende om, och endast om,

$$P(A|B) = P(A)$$



**Exempel 3-5** Förorenad mark (forts)




Beräkna och skriv ut i figuren ovan sannolikheterna

$$P(D|C),\ P(D'|C),\ P(D|C') \text{ och } P(D'|C')$$

Skriv även in i figuren sannolikheterna

$$P(C) \text{ och } P(C')$$

Beräkna sedan sannolikheterna

$$P(C|D) \text{ och } P(C|D')$$

**Hemövning 2** Antag att $P(C) = 0.9$, $P(D) = 0.7$ och $P(C \cap D) = 0.65$ som i föregående hemövning. Beräkna $P(D|C')$ och $P(D'|C)$. Gäller

$$P(D|C') + P(D'|C) \leq 1$$

eller ej?

**Lagen om total sannolikhet:** Låt $A_1, \ldots, A_n$ partitionera utfallsrummet $S$. Då gäller

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

**Bayes formel:** Under samma förutsättningar som i lagen om total sannolikhet, gäller

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

I exemplet med förorenad mark, noterar vi att händelserna

$$C \text{ och } C'$$

partitionerar utfallsrummet. Bayes formel ger att

$$P(C|D) = \frac{P(C)P(D|C)}{P(C)P(D|C) + P(C')P(D|C')}$$
$$= \frac{0.8 \cdot 0.9}{0.8 \cdot 0.9 + 0.2 \cdot 0.1} = \frac{0.72}{0.72 + 0.02} = 0.973$$

**Hemövning 3** Antag att $P(C) = 0.9$, samt att $P(D|C') = 1/2$ och $P(D'|C) = 5/18$. Beräkna, analogt med ovan, $P(C|D)$ och $P(C|D')$. Gäller

$$P(C|D') \leq P(C) \leq P(C|D)$$

eller ej?



**Exempel 5.10** (s 207) Låt $T$ vara livstiden för en produkt. Funktionen

$$R(t) = P(T > t)$$

kallas för produktens *överlevnadsfunktion*. Väntevärdet

$$\mu_T = E[T]$$

kallas ofta för "mean time to failure" (MTTF) eller för"mean time between failures" (MTBF). För en exponentialfördelad variabel gäller att överlevnadsfunktionen är

$$R(t) = \int_t^{\infty} \lambda e^{-\lambda x}\, dx = \cdots = e^{-\lambda t}$$

och att MTTF är

$$E[T] = \int_0^{\infty} x \lambda e^{-\lambda x}\, dx = \cdots = \frac{1}{\lambda}$$

**Exempel 6** Gränsvärdet

$$h(t) = \lim_{h \to 0} \frac{P(T \leq t + h | T > t)}{h} = \frac{f(t)}{R(t)}$$

kallas för *felintensiteten*. För en exponentialfördelad variabel gäller att felintensiteten är konstant, ty

$$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Man kan visa att om felintensiteten är konstant, så måste variabeln vara exponentialfördelad.

**Hemövning 4** Leta reda på Weibullfördelningen i läroboken eller Beta eller…Skriv sedan ner Weibullfördelningens felintensitet.



Låt $x_1, \ldots, x_n$ vara ett *stickprov* på (den stokastiska) variabeln $x$. Då är alltså $x_1, \ldots, x_n$ oberoende observationer av $x$.

**Definition 1** (s 214) Reellvärda funktioner av ett stickprov kallar vi *stickprovsvariabler* (statistikor).

Stickprovsvariabler används för att dra slutsatser om $x$:s fördelning.

**Exempel 7** Proportionen observationer i ett intervall, medelvärdet, medianen, standardavvikelsen, IQR, undre och övre kvartilen, alla övriga kvantiler (percentiler).

**Definition 2** (s 215) Stickprovsvariabler är stokastiska variabler, så de har fördelningar. På engelska säger man ofta "sampling distribution" för att göra tydligt att det är fördelningen för en stickprovsvariabel som avses.

När vi räknar teoretiskt på stickprov är det praktiskt att använda stora bokstäver för de stokastiska variablerna och små för deras utfall. Vi skriver alltså i fortsättningen $X_1, \ldots, X_n$ för stickprovet och medelvärdet betecknar vi $\bar{X}$. Vi låter $S^2$ beteckna stickprovets varians och $s^2$ motsv. empiriska (uppmätta, experimentella) värde.



**Sats 1** (s 220) Låt $X_1, \ldots, X_n$ vara ett stickprov av den stokastiska variabeln $X$ med väntevärde $\mu$ och varians $\sigma^2$. Då gäller för medelvärdet $\bar{X}$ att

$$E[\bar{X}] = \mu$$
$$\mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n}$$

Således är

$$\sigma_{\bar{X}} = \sqrt{\mathrm{Var}[\bar{X}]} = \frac{\sigma}{\sqrt{n}}$$

Denna storhet kallas i engelskspråkig litteratur för "standard error". På svenska skulle man kunna säga standardfel eller standardosäkerhet.

**Sats 2** (s 221) Om $X$ är normalfördelad med parametrar $\mu$ och $\sigma$, så är $\bar{X}$ normalfördelad med parametrar $\mu$ och $\sigma/\sqrt{n}$.

**Sats 3 Centrala gränsvärdessatsen** (s 222) Om $X$ ej är normalfördelad, så är ändå fördelningen för medelvärdet $\bar{X}$ är approximativt normalfördelad med parametrar $\mu$ och $\sigma/\sqrt{n}$. Approximationen blir bättre ju fler variabler som ingår i stickprovet.

**Hemövning 5** Antag att du har ett stickprov på en variabel $x$ som är $\mathrm{Exp}(10^{-5})$. Du tänker simulera 1000 observationer av $x$. Ungefär hur stor är sannolikheten att medelvärdet blir minst 100 010.

**Sats 4** (s 225) Om $X$ antar värdet 1 med sannolikheten $p$ och 0 med sannolikheten $1 - p$, så gäller

$$\mu = E[X] = p$$
$$\sigma^2 = \mathrm{Var}[X] = p(1-p)$$

**Sats 5** (s 226) För relativa frekvensen $f/n$ gäller att väntevärdet är

$$E[f/n] = p$$

och att variansen är

$$\mathrm{Var}[f/n] = \frac{p(1-p)}{n}$$

**Sats 6** Då antalet observationer $n$ är stort, så är relativa frekvensen $f/n$ approximativt normalfördelad med parametrar $p$ och $\sqrt{p(1-p)/n}$.

**Hemövning 6** I en undersökning av väljarsympatier tänker man tillfråga $n = 1000$ slumpmässigt utvalda individer med rösträtt om huruvida de skulle rösta på socialdemokraterna om det vore val idag. Men förväntar sig att proportionen ja-svarare är ungefär $p = 0.3$. Ungefär hur stor är då sannolikheten att minst 330 av de tillfrågade svarar ja?



**Exempel 8** EC inbjuder till vadslagning och säger: Jag ska singla det här myntet 100 gånger. Det är helt symmetriskt. Titta gärna på det. Om du vill att vi ska singla ett annat mynt, så är det helt ok för mig. Jag satsar 10 kr om du satsar 2 kr emot, på att jag får 40 - 60 klave. Antar du vadet?

**Hemövning 7** Om du satsar 2 kr, hur mycket tycker du att EC ska satsa för att vadet ska bli rättvist?



**Nyckelord i kursdelen**

Utfallsrum för ett slumpmässigt försök, händelse, utfall

Variabler och stokastiska variabler

Sannolikheter

Ömsesidigt uteslutande (disjunkta) händelser, partition

Betingad sannolikhet

Oberoende

Venn- och träddiagram

Lagen om total sannolikhet och Bayes formel

Stickprovsvariabler

$X_1, \ldots, X_n$ oberoende $\mathrm{N}(\mu, \sigma)$-variabler $\Rightarrow \bar{X} \sim \mathrm{N}(\mu, \sigma/\sqrt{n})$

$X_1, \ldots, X_n$ oberoende och likafördelade med väntevärde $\mu$ och standardavvikelse $\sigma \Rightarrow \bar{X} \overset{\text{ap}}{\sim} \mathrm{N}(\mu, \sigma/\sqrt{n})$ (cgs)