

På förra föreläsningen definierade vi utfall, händelser och sannolikhetsmått. Vi avslutade med likformiga sannolikhetsmått där sannolikheten för en händelse kunde bestämmas ur antalet gynnsamma respektive möjliga fall. För att bestämma antalet gynnsamma och möjliga fall är det ofta användbart att veta på hur många sätt man kan välja ut k element ur en mängd av n olika element. Observera att urvalet kan göras på flera olika sätt.

1. Med återläggning, med hänsyn till ordning: n^k

(Vid var och en av de k dragningarna finns n element att välja bland)

2. Utan återläggning, med hänsyn till ordning: $n(n-1)(n-2)\dots(n-k+1) = n!/(n-k)!$
(Vid första dragningen finns n element att välja bland, vid andra dragningen $n-1$ element osv.)

3. Utan återläggning, utan hänsyn till ordning: $\frac{n!}{(n-k)!} / k! := \binom{n}{k}$ (uttäses "n över k")
(De k kulorna som valts ut i 2. kan ordnas på $k!$ olika sätt, så om hänsyn inte ska tas till ordningen måste antalet fall divideras med $k!$).

Vi tillämpar dessa resultat på en urnmodell som är användbar inom kvalitetskontroll. Vi tänker oss en urna som innehåller s svarta och v vita kulor. Kulorna får representera hela respektive defekta enheter i ett parti av någon produkt. Vilken är sannolikheten att vi får l vita kulor vid dragning av m kulor (med respektive utan återläggning)?

I fallet utan återläggning räknar vi utan hänsyn till ordning.

möjliga fall:

$$\binom{s+v}{m}$$

gynnsamma fall:

$$\binom{v}{l} \binom{s}{m-l}$$

$$\Rightarrow p = \frac{\binom{v}{l} \binom{s}{m-l}}{\binom{s+v}{m}}$$

I fallet med återläggning räknar vi med hänsyn till ordning.

möjliga fall: $(s+v)^m$

gynnsamma fall:

$$\binom{m}{l} v^l s^{m-l}$$

sätt att välja $m-l$ svarta kulor (med återläggning)

sätt att välja l vita kulor (med återl.)

sätt att välja de l omgångar då vit kula väljs.

$$p = \frac{\binom{m}{l} v^l s^{m-l}}{(s+v)^m}$$

Vi ägnar resten av föreläsningen åt betingade sannolikheter. Vi använder dessa för att avgöra om information om att händelse inträffat påverkar sannolikheten för andra händelser. Vi inleder med ett exempel.

Ex: Av 10000 insamlade e-brev visar sig 6243 vara spam och resterande ej spam. Förekomsten av ordet "free" i breven undersöks också och tabellen till höger fås:

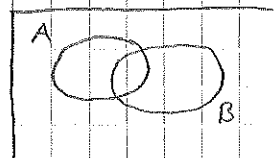
Låt S vara händelsen att ett e-brev är spam och F händelsen att ett e-brev innehåller ordet "free". Vi ser att $IP(S) = 62\%$, men att sannolikheten för S givet att F inträffat är $5494/5676 = 97\%$, dvs 97% av alla e-brev som innehåller "free" är spam.

	spam	ej spam	
"free"	5494	182	5676
ej "free"	749	3575	4324
	6243	3757	10000

Def: Låt A och B vara två händelser och antag att $IP(B) > 0$. Den betingade sannolikheten för A givet B betecknas $IP(A|B)$ och definieras som

$$IP(A|B) = IP(A \cap B) / IP(B)$$

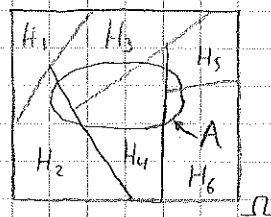
$IP(A|B)$ är sannolikheten för A givet att B inträffat. När vi betingar på B reduceras utfallsrummet till B .



Sats: (Lagen om total sannolikhet)

Antag att händelserna $H_1, \dots, H_n$ är parvis disjunkta och att $\bigcup_{i=1}^n H_i = \Omega$. Då gäller, för varje händelse A , att

$$IP(A) = \sum_{i=1}^n IP(A|H_i)IP(H_i)$$



Bevis: Satsen kan bevisas från axiomen (se boken), men vi motiverar satsen från Venn-diagrammet istället. Händelsen A kan delas upp i de parvis disjunkta händelserna $A \cap H_1, A \cap H_2, \dots, A \cap H_n$. Vi har

$$IP(A) = \sum_{i=1}^n IP(A \cap H_i) = [\text{definition av betingad slh}] = \sum_{i=1}^n IP(A|H_i)IP(H_i)$$

De betingade sannolikheterna $IP(A|B)$ och $IP(B|A)$ är relaterade till varandra enligt ett samband som kallas Bayes sats. Per definition gäller

$$IP(A|B) = \frac{IP(A \cap B)}{IP(B)} \quad \text{och} \quad IP(B|A) = \frac{IP(A \cap B)}{IP(A)} \quad \text{så} \quad IP(B|A) = \frac{IP(A|B)IP(B)}{IP(A)}$$

Sats: Under samma villkor som i satsen om total sannolikhet, så gäller det att

$$IP(H_j|A) = \frac{IP(A|H_j)IP(H_j)}{\sum_{i=1}^n IP(A|H_i)IP(H_i)}$$

Vi illustrerar satsens användbarhet med ett exempel.

Ex: Antag att 0,4% av befolkningen drabbas av en viss form av cancer. Ett diagnostiskt test ger positivt resultat i 99,5% av fallen om den som testas har cancerformen och i 1% av fallen om den som testas inte har cancerformen (jfr screening mot bröstcancer som dock är en mycket vanligare cancerform och som drabbar ca 20% av alla svenska kvinnor). Vi vill beräkna sannolikheten att en slumpmässigt vald person som testar positivt verkligen har cancerformen. Låt C^+ och C^- beteckna händelserna att personen har respektive inte har cancerformen och låt T^+ och T^- beteckna händelserna att personen testar positivt respektive negativt. Vi vet $IP(C^+) = 0,004$, $IP(C^-) = 0,996$, $IP(T^+|C^+) = 0,995$ och $IP(T^+|C^-) = 0,01$ och vill bestämma $IP(C^+|T^+)$. Bayes sats ger (eftersom C^+ och C^- är disjunkta händelser och $C^+ \cup C^- = \Omega$)

$$IP(C^+|T^+) = \frac{IP(T^+|C^+)IP(C^+)}{IP(T^+|C^+)IP(C^+) + IP(T^+|C^-)IP(C^-)} = \frac{0,995 \cdot 0,004}{0,995 \cdot 0,004 + 0,01 \cdot 0,996} = 0,286$$

Screeningen ger upphov till onödigt behandling i 71% av fallen. Ett bröstcancer-test med samma precision ger bara onödigt behandling i 4% av fallen. Screening är ineffektivt för ovanliga sjukdomar.

Vi avslutar med att diskutera oberoende händelser. Intuitivt är två händelser A och B oberoende om inträffandet av A inte ger någon information om hurvida B inträffat eller ej. I formen innebär detta att

$$IP(B|A) = IP(B) \quad \text{Men} \quad IP(B|A) = IP(A \cap B)/IP(A), \text{ så följande definition är naturlig.}$$

Def: Två händelser A och B är oberoende om $IP(A \cap B) = IP(A)IP(B)$

Om A och B är oberoende, så är även A^* och B , A och B^* samt A^* och B^* oberoende. Det första påståendet följer av

$$IP(A^* \cap B) = IP(B) - IP(A \cap B) = IP(B) - IP(A)IP(B) = IP(B)(1 - IP(A)) = IP(A^*)IP(B)$$

