

Föreläsning 2

Thomas Önskog

31/10 2017

Sannolikheteusteoriens grunder

På förra föreläsningen definierade vi utfall, händelser och sannolikhetsmått. Vi avslutade med likformiga sannolikhetsmått där sannolikheten för en händelse kunde bestämmas som kvoten av antalet gynnsamma utfall för händelsen ifråga och antalet möjliga utfall. För att bestämma antalet gynnsamma och antalet möjliga utfall 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. Multiplikationsprincipen ger därmed att det totalt sätt finns n^k valmöjligheter.

2. **Utan återläggning, med hänsyn till ordning:** $n(n-1)(n-2)\cdots(n-k+1) = n!/(n-k)!$.

Vid första dragningen, finns n element att välja bland, vid andra dragningen $n-1$ element osv.

3. **Utan återläggning, utan hänsyn till ordning:** $\frac{n!}{(n-k)!k!} := \binom{n}{k}$ (utläses "n över k")

De k kulorna som valts ut i fall 2 kan ordnas på $k!$ olika sätt, så om hänsyn inte ska tas till ordningen, så måste antalet fall divideras med $k!$.

Vi tillämpar dessa resultat på en **urnmodell** som är användbar inom kvalitetskontroll.

Exempel. Betrakta 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. Antalet möjliga fall och gynnsamma fall är

$$\binom{s+v}{m} \quad \text{respektive} \quad \binom{v}{l} \binom{s}{m-l},$$

så sannolikheten att vi får l vita kulor är

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

I fallet med återläggning räknar vi med hänsyn till ordning. Antalet möjliga och gynnsamma fall är

$$(s + v)^m \quad \text{respektive} \quad \binom{m}{l} v^l s^{m-l},$$

där den första termen i uttrycket för antalet gynnsamma fall motsvarar antalet sätt att välja de l omgångar då vit kula väljs, andra termen motsvarar antalet sätt att välja l vita kulor med återläggning och tredje termen motsvarar antalet sätt att välja $m - l$ svarta kulor med återläggning. Sannolikheten att vi får l vita kulor är alltså

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

Betingade sannolikheter

Betingade sannolikheter används för att avgöra om vetenskapen att en händelse inträffat påverkar sannolikheten för andra händelser. Vi inleder med ett exempel.

Exempel. Av 10000 insamlade e-brev visar sig 6243 vara skräppost och resterande ej skräppost. Förekomsten av ordet "free" i breven undersöks också och vi får tabellen

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

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

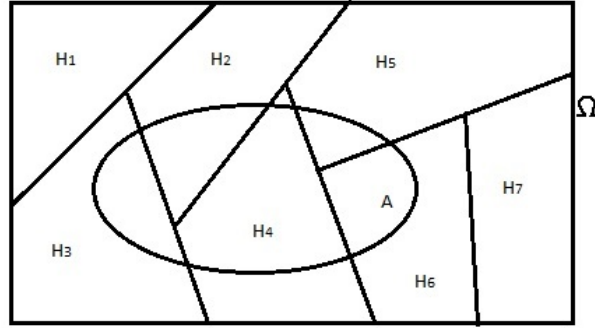
Definition. Låt A och B vara två händelser och antag att $\mathbb{P}(B) > 0$. Den **betingade sannolikheten** för A givet B betecknas $\mathbb{P}(A|B)$ och definieras som

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

$\mathbb{P}(A|B)$ 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 $H_1, H_2, \dots, H_n$ är parvis disjunkta och att $\bigcup_{i=1}^n H_i = \Omega$. Då gäller, för varje händelse A , att

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|H_i)\mathbb{P}(H_i).$$



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

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap H_i) = [\text{Definition av betingad sannolikhet}] = \sum_{i=1}^n \mathbb{P}(A|H_i)\mathbb{P}(H_i).$$

□

De betingade sannolikheterna $\mathbb{P}(A|B)$ och $\mathbb{P}(B|A)$ är relaterade till varandra enligt ett samband som kallas **Bayes sats**. Per definition gäller

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{och} \quad \mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}, \quad \text{så} \quad \mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}.$$

Sats. Under samma villkor som i lagen om total sannolikhet, så gäller

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

Vi illustrerar satsens användbarhet med ett exempel.

Exempel. 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 cirka 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 att $\mathbb{P}(C^+) = 0.004$, $\mathbb{P}(C^-) = 0.996$, $\mathbb{P}(T^+|C^+) = 0.995$ samt att $\mathbb{P}(T^+|C^-) = 0.01$ och vill bestämma $\mathbb{P}(C^+|T^+)$. Notera att C^+ och C^- är disjunkta händelser och att $C^+ \cup C^- = \Omega$. Bayes sats med $H_1 = C^+$, $H_2 = C^-$ och $A = T^+$ ger då att

$$\mathbb{P}(C^+|T^+) = \frac{\mathbb{P}(T^+|C^+)\mathbb{P}(C^+)}{\mathbb{P}(T^+|C^+)\mathbb{P}(C^+) + \mathbb{P}(T^+|C^-)\mathbb{P}(C^-)} = \frac{0.995 \cdot 0.004}{0.995 \cdot 0.004 + 0.01 \cdot 0.996} = 0.286.$$

Screeningen ger upphov till onödig behandling i 71% av fallen. Ett bröstcancer-test med samma precision ger bara onödig behandling i 4% av fallen. Vi drar slutsatsen att screening är ineffektivt för ovanliga sjukdomar.

Oberoende händelser

Intuitivt är två händelser A och B oberoende om inträffandet av A inte ger någon information om huruvida B inträffat eller ej. Detta kan uttryckas i formler som

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

Vidare gäller det, per definition, att $\mathbb{P}(B|A) = \mathbb{P}(A \cap B)/\mathbb{P}(A)$, så följande definition är naturlig

Definition. Två händelser A och B är **oberoende** om $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Om A och B är oberoende, så är även A^* och B , A och B^* samt A^* och B^* oberoende. Den första implikationen följer av att

$$\begin{aligned}\mathbb{P}(A^* \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) = [A \text{ och } B \text{ är oberoende}] = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^*).\end{aligned}$$

