Sannolikhet och statistik III

Johan Jonasson

Mars 2019

Oberoende av flera händelser

Man säger att händelserna A, B och C är oberoende om de är parvis oberoende och

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

Mer generellt,

Definition: Man säger att händelserna $A_1, A_2, A_3, \ldots$ är oberoende om det för alla n och alla indexmängder $i_1, \ldots, i_n$ gäller att

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} \mathbb{P}(A_{i_k}).$$

Att vissa händelser är oberoende är oftast något som man antar i sin modell snarare än att man räknar ut det. Exempelvis, om vi talar om upprepade tärningskast, så är det rimligt att helt enkelt anta att händelser som har med olika kast att göra är oberoende. T.ex. om A_n är händelsen att man får sexa i kast nr n, så antar vi att $A_1, A_2, \ldots$ är oberoende. Vi får till exempel

$$\mathbb{P}(\text{första sexan i kast nr } n) = \mathbb{P}(A_1^c \cap A_2^c \cap \ldots \cap A_{n-1}^c \cap A_n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}.$$

Vi har också, enligt kontinuitet hos sannolikheter,

$$\mathbb{P}(\text{man får aldrig någon sexa}) = \mathbb{P}(A_1^c \cap A_2^c \cap \ldots)$$

$$= \lim_{n\to\infty} \mathbb{P}(A_1^c \cap \ldots \cap A_n^c) = \lim_{n\to\infty} \left(\frac{5}{6}\right)^n = 0.$$

Generellt: om en händelse har en positiv sannolikhet och försöket upprepas igen och igen, kommer denna händelse garanterat att inträffa förr eller senare.

I många sammanhang är det lättare att förstå vad $\mathbb{P}(B|A)$ är snarare än $\mathbb{P}(A \cap B)$. Därför är formeln

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

lika användbar som definitionen av betingad sannolikhet.
Antag att $A_1, A_2, \ldots, A_n$ är en partition av S, dvs parvis disjunkta mängder sådana att $A_1 \cup A_2 \ldots \cup A_n = S$. Då gäller att

$$\begin{aligned} \mathbb{P}(B) &= \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) \\ &= \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i). \end{aligned}$$

Detta är den *totala sannolikhetslagen*.

Specialfall: $\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)$.

Exempel: Tre procent av Sveriges befokning drabbas någon gång av lungcancer. Bland rökare drabbas tio procent någon gång. Rökarna utgör 20 procent av befolkningen. Vad är risken för en ickerökare att drabbas av lungcancer någon gång i livet?

Välj en svensk person, NN, på måfå och låt

$$A = \{\text{NN är rökare}\},\ B = \{\text{NN får lungcancer}\}.$$

Vi söker $\mathbb{P}(B|A^c)$. Givet i uppgiften är att $\mathbb{P}(A) = 0.2$, $\mathbb{P}(B|A) = 0.1$, $\mathbb{P}(B) = 0.03$. Enligt TSL:

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

Detta ger

$$\mathbb{P}(B|A^c) = \frac{\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A^c)} = \frac{0.03 - 0.2 \cdot 0.1}{0.8} = 0.0125.$$

Hur stor andel av dem som drabbas av lungcancer är rökare? Här söker vi $\mathbb{P}(A|B)$.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{0.1 \cdot 0.2}{0.03} = \frac{2}{3}.$$

I slutet på exemplet vände vi på betingning. Enligt TSL kan vi skriva om nämnaren och få

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}.$$

Detta är *Bayes formel.*

Mer generellt gäller att om $A_1, A_2, \ldots, A_n$ är en partition så är

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{k=1}^{n} \mathbb{P}(B|A_k)\mathbb{P}(A_k)}.$$

Exempel: I en hatt låg tre mynt: mynt nummer 1 som ger klave med sannolikhet 0.3, mynt nummer 2 som ger klave med sannolikhet 0.5 och mynt nummer 3 som ger klave med sannolikhet 0.8. Ett mynt valdes på måfå, kastades tre gånger och det blev klave exakt en gång. Vad är den betingade sannolikheten att det var mynt nummer i som valdes?

Låt M_i vara händelsen att mynt nummer i valdes, $i = 1, 2, 3$, och låt A vara händelsen att det blir klave exakt en gång på tre kast. För alla i gäller $\mathbb{P}(M_i) = 1/3$. Vidare gäller (om vi förutsätter att de olika kastens resultat är oberoende) att

$$\mathbb{P}(A|M_1) = 3 \cdot 0.3 \cdot 0.7^2 = 0.441$$

$$\mathbb{P}(A|M_2) = 3 \cdot 0.5 \cdot 0.5^2 = 0.375$$

$$\mathbb{P}(A|M_3) = 3 \cdot 0.8 \cdot 0.2^2 = 0.096.$$

Enligt Bayes formel är

$$\mathbb{P}(M_i|A) = \frac{\mathbb{P}(A|M_i)\mathbb{P}(M_i)}{\sum_{j=1}^{3} \mathbb{P}(A|M_j)\mathbb{P}(M_j)}.$$

Nämnaren är

$$\sum_{j=1}^{3} \mathbb{P}(A|M_j)\mathbb{P}(M_j) = \frac{1}{3}(0.441 + 0.375 + 0.096) = 0.304.$$

Detta ger

$$\mathbb{P}(M_1|A) = \frac{\frac{1}{3} \cdot 0.441}{0.304} \approx 0.48$$

$$\mathbb{P}(M_2|A) = \frac{\frac{1}{3} \cdot 0.375}{0.304} \approx 0.41$$

$$\mathbb{P}(M_3|A) = \frac{\frac{1}{3} \cdot 0.096}{0.304} \approx 0.11.$$

Exempel: Sjukdomen S. Erfarenheten visar att bland de patienter som skickas för test för den ovanliga sjukdomen S är det 1 % av dem som verkligen har S. Testet är sådant att om en patient verkligen har S så visar testet positivt utslag med sannolikhet 0.95 (sensitivitet). Om patienten inte har S så visar testet ett negativt utslag med sannolikhet 0.9 (specificitet). Om en patient får ett positivt utslag på sitt test, vad är den betingade sannolikheten att han/hon verkligen har S?

Låt A vara händelsen att patienten har S och låt T vara händelsen att testet ger ett positivt utslag. Enligt uppgift gäller

$$\mathbb{P}(A) = 0.01,\ \mathbb{P}(T|A) = 0.95,\ \mathbb{P}(T|A^c) = 0.1.$$

Vi söker $\mathbb{P}(A|T)$.

$$\begin{aligned}\mathbb{P}(A|T) &= \frac{\mathbb{P}(T|A)\mathbb{P}(A)}{\mathbb{P}(T|A)\mathbb{P}(A) + \mathbb{P}(T|A^c)\mathbb{P}(A^c)} \\ &= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.1 \cdot 0.99} \\ &= 0.095.\end{aligned}$$

Säg nu att man testar en person från gatan (screening) och att S förekommer hos 0.01 % av befolkningen. I så fall får vi istället $\mathbb{P}(A) = 0.0001$ och

$$\mathbb{P}(A|T) = \frac{0.95 \cdot 0.0001}{0.95 \cdot 0.0001 + 0.1 \cdot 0.9999} = 0.00095.$$

Exempel: Gambler's ruin. Ann och Bo spelar på upprepade oberoende slantsinglar enligt regeln att när det blir klave ger Ann 1 kr till Bo och när det blir krona ger Bo en krona till Ann. Detta upprepas till någon av dem blir utan pengar. Om Ann vid starten har k kr och Bo har $n-k$ kr, vad är då sannolikheten att Ann slutligen vinner?

Låt A vara händelsen att Ann vinner. Låt $\mathbb{P}_k$ stå för sannolikhetsmåttet vid start från k kr för Ann. (Varje k ger noga taget en egen modell.) Låt H vara händelsen att första kastet ger klave.
Skriv $p_k = \mathbb{P}_k(A)$. Vi har $p_0 = 0$, $p_n = 1$ och för övriga k ger TSL

$$p_k = \mathbb{P}_k(A) = \mathbb{P}_k(A|H)\mathbb{P}_k(H) + \mathbb{P}_k(A|H^c)\mathbb{P}_k(H^c) = \frac{1}{2}\left(p_{k-1} + p_{k+1}\right).$$

Detta ger $p_{k+1} = 2p_k - p_{k-1}$. Lösningen är $p_k = k/n$.