Rozwiązanie zad. 2 (kolokwium z RPiS, 6 grudnia 2008)

Zadanie 2. (10 punktów) Każdego dnia deszcz pada z prawdopodobieństwem $\frac{1}{2}$, w przeciwnym przypadku mamy słoneczną pogodę. Telewizyjna prognoza pogody myli się z prawdopodobieństwem $\frac{1}{3}$, t.j. jeśli prognoza mówi, że pogoda będzie deszczowa, to z prawdopodobieństwem $\frac{2}{3}$ faktycznie będzie, podobnie, jeśli prognoza mówi, że pogoda będzie słoneczna, to z prawdopodobieństwem $\frac{2}{3}$ faktycznie będzie.

Profesor zawsze zabiera ze sobą parasol, jeśli prognoza zapowiada pogodę deszczową. Jeśli natomiast zapowiadana jest pogoda słoneczna, profesor zabiera parasol z prawdopodobieństwem $\frac{1}{3}$.

1. Oblicz prawdopodobieństwo wystąpienia prognozy deszczowej (3 punkty).
2. Zakładając, że pada, oblicz prawdopodobieństwo tego, że profesor nie ma ze sobą parasola (3 punkty).
3. Zakładając, że profesor ma ze sobą parasol, oblicz prawdopodobieństwo tego, że nie pada (4 punkty).

Rozwiązanie Będziemy używać następujących oznaczeń dla zdarzeń:

- D - pada deszcz,
- S - świeci słońce,
- P_D - prognoza twierdzi, że będzie padał deszcz,
- P_S - prognoza twierdzi, że będzie świeciło słońce,
- U_+ - profesor bierze parasol,
- U_- - profesor nie bierze parasola.

1. Wiemy, że $P(D|P_D) = P(S|P_S) = 2/3$ oraz $P(S|P_D) = P(D|P_S) = 1/3$. A zatem, ze wzoru na prawdopodobieństwo całkowite, mamy

   $$P(S) = P(P_D) \cdot P(S|P_D) + P(P_S) \cdot P(S|P_S) = 1/3 \cdot P(P_D) + 2/3 \cdot P(P_S).$$

   Podobnie

   $$P(D) = P(P_D) \cdot P(D|P_D) + P(P_S) \cdot P(D|P_S) = 2/3 \cdot P(P_D) + 1/3 \cdot P(P_S).$$

   Ponieważ $P(S) = 1/2 = P(D)$, prawe strony tych wyrażeń są równe i dostajemy $P(P_S) = P(P_D)$, a zatem oba są równe $1/2$.

2. Chcemy obliczyć $P(U_-|D)$. Mamy $P(U_-|D) = P(U_- \cap D)/P(D) = 2 * P(U_- \cap D)$. Rozbijając ze względu na to jaka jest prognoza dostajemy $P(U_- \cap D) = P(U_- \cap P_D \cap D) + P(U_- \cap P_S \cap D)$, ale z założeń zadania wynika, że $P(U_- \cap P_D) = 0$, więc $P(U_- \cap D) = P(U_- \cap P_S \cap D)$. Ze wzoru "na mnożenie" dostajemy:

   $$P(U_- \cap P_S \cap D) = P(P_S) \cdot P(D|P_S) \cdot P(U_-|D \cap P_S) = 1/2 \cdot 1/3 \cdot 2/3 = 1/9.$$

   Łącząc ten wynik z wcześniej uzyskanym $P(U_-|D) = 2 * P(U_- \cap D)$ dostajemy $P(U_-|D) = 2/9$.

3. Chcemy obliczyć $P(S|U_+)$. Korzystając ze wzoru Bayesa dostajemy

$$P(S|U_+) = \frac{P(S) \cdot P(U_+|S)}{P(S) \cdot P(U_+|S) + P(D) \cdot P(U_+|D)}.$$

Z poprzedniego punktu mamy $P(U_+|D) = 1 - P(U_-|D) = 7/9$.

$P(U_+|S)$ obliczamy analogicznie, t.j.

$$P(U_+|S) = 2 \cdot P(U_+ \cap S) = 2 \cdot (P(P_D \cap S \cap U_+) + P(P_S \cap S \cap U_+) =$$

i dalej rozpisując ze wzoru "na mnożenie"

$$= 2 \cdot (P(P_D) \cdot P(S|P_D) \cdot P(U_+|S \cap P_D) + P(P_S) \cdot P(S|P_S) \cdot P(U_+|S \cap P_S)) = 2 \cdot (1/6 + 1/9) = 5/9.$$

A zatem $P(S|U_+) = \frac{5/9}{5/9+7/9} = 5/12$.

Dyskusja Bardzo popularnym błędem było stwierdzenie (w pierwszym punkcie), że

$$P(P_D) = P(D) \cdot P(P_D|D) + P(S) \cdot P(P_D|S),$$

i uznanie, że skoro prognoza myli się z prawdopodobieństwem 1/3 zarówno wtedy, gdy jest (prognoza) deszczowa, jak i wtedy, kiedy jest słoneczna, to musi także mylić się z prawdopodobieństwem 1/3, gdy jest deszcz i 1/3, gdy jest słońce, t.j. $P(P_D|D) = 2/3$ i $P(P_D|S) = 1/3$. Mimo, że akurat te dwie równości są prawdziwe, to jest to zbieżność przypadkowa!!!

O tym, że samo rozumowanie jest błędne można się przekonać wstawiając do zadania inne wartości $P(S)$ i $P(D)$. Czytelnik może sprawdzić, że jeśli $P(D) = 1/3$ i $P(S) = 2/3$, to pozostawiając resztę założeń bez zmian dostajemy $P(P_S) = 1$ i $P(P_D) = 0$ (wynik ten można uzyskać stosując rozumowanie przedstawione w rozwiązaniu wzorcowym punktu 1). Oczywiście w tej sytuacji $P(P_D|D) = 0 = P(P_D|S) = 0$. Widać, że błędne rozumowanie przypadkowo prowadzi do dobrego wyniku, tylko dlatego, że $P(S) = P(D)$.

Warto przy tym zwrócić uwagę na to, że opisany błąd dużo trudniej jest popełnić jeśli oznaczy się wszystkie zdarzenia i przeprowadza cały dowód formalnie, a nie, jak to zrobiła duża grupa osób, po prostu wypisuje iloczyn kilku ułamków lub rysuje drzewko bez dokładnego opisu.

Innym niepoprawnym (ale prawie poprawnym) rozwiązaniem było stwierdzenie, że $P(P_D)$ musi być równe $P(P_S)$ ze względu na symetrie. To rozwiązanie jest niekompletne, ponieważ nie jest wykluczone, że dane zawarte w treści zadania są niewystarczające do wyznaczenia wartości $P(P_D)$ i $P(P_S)$ (rozwiązaniem mogłyby być np. przedziały wartości).