Rachunek prawdopodobieństwa i statystyka

Wykład 1

Rachunek prawdopodobieństwa i statystyka matematyczna

Dr hab. inż. Mariusz Przybycień

Literatura:

- Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, tom I i II, W. Krysicki i in., PWN 2005.
- Wstęp do rachunku prawdopodobieństwa, W. Feller, PWN 2006.
- Zadania z probabilistyki, A. Plucińska, E. Pluciński, PWN 1983.
- Statystyka dla fizyków, R.N. Nowak, PWN, 2002.
- Statystyka dla fizyków. Ćwiczenia, R.N. Nowak, PWN, 2002.
- Analiza danych, S. Brandt, PWN, 1998.
- Wstęp do analizy błędu pomiarowego, R.J. Taylor, PWN, 2001.
- Jak analizować wyniki pomiarów, H. Abramowicz, PWN 1992.
- Statistical Data Analysis, G. Cowan, Oxford Univ. Press, 1998.

- http://home.agh.edu.pl/mariuszp

Statystyka - podstawowe pojęcia

Statystyka – nauka zajmująca się planowaniem badań, a także zbieraniem, organizacją, prezentacją i analizą danych, oraz wyciąganiem wniosków i podejmowaniem decyzji na ich podstawie.

Słowo „statystyka" jest także używane do określenia samych danych i wielkości pochodnych.

Populacja – zbiór wszystkich przedstawicieli posiadających badaną cechę.

Przykład: Badania demograficzne - spis powszechny. Kontrola jakości – zbiór wszystkich urządzeń danego typu produkowanych przez fabrykę.

Próbka losowa – reprezentatywna próbka całej populacji, tzn. taka, która odzwierciedla wszystkie cechy i związki w niej występujące.

Przykład: Próbkami losowymi nie są np. sondaże wśród czytelników dowolnego z czasopism, wśród przechodniów na ulicy, głosowania telewidzów w programach

Mówimy, że próbka jest prosta jeśli rezultat wyboru jednego elementu nie ma wpływu na rezultat wyboru innego elementu.

Przykład: Losując bez zwracania kule z urny, która jest wypełniona skończoną liczbą kul białych i czarnych, mamy do czynienia z próbką, która nie jest prosta.

Zdarzenia elementarne

Zachowanie układu, którego nie możemy przewidzieć z całkowitą pewnością, nazywamy przypadkowym. Miarą „przypadkowości" jest prawdopodobieństwo.

Pojęcia pierwotne rachunku prawdopodobieństwa to zdarzenie elementarne i przestrzeń zdarzeń elementarnych Ω. Dowolny podzbiór $A \subset \Omega$ nazywamy zdarzeniem losowym.

Uwaga: Zdarzenia elementarne muszą być ekskluzywne – dane zdarzenie elementarne nie zawiera w sobie innych zdarzeń elementarnych.

Przykłady:

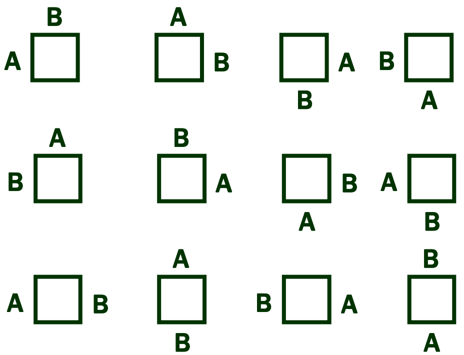
- jednokrotny rzut monetą: „orzeł" (O) i „reszka" (R) to dwa zdarzenia elementarne, które budują całą przestrzeń $\Omega = \{O, R\}$
- rzut dwoma monetami: $\Omega = \{(O,O), (O,R), (R,O), (R,R)\}$
- miesiąc urodzin: Ω = {Sty, Lut, Mar, Kwi, Maj, Cze, Lip, Sie, Wrz, Paź, Lis, Gru}
- czas życia żarówki – zdarzenia elementarne są dowolnymi liczbami dodatnimi, $t>0$, a przestrzeń jest nieskończona; ale może w konkretnym problemie lepiej wybrać $\Omega = \{1, 2, 3, \ldots\}$ godziny/dni/miesiące.
- liczba wypadków na skrzyżowaniu w ciągu miesiąca: $\Omega = \{1,2,3,\ldots\}$ a może $\Omega = \{1, 2, 3, \ldots, 1000\}$ – często prościej rozważyć nieskończony ciąg.
- jądro promieniotwórcze w kolejnych odstępach 1s może się rozpaść (R) lub nie (B). Przestrzeń zdarzeń elementarnych jest nieskończona: R, BR, BBR, BBBR, ...

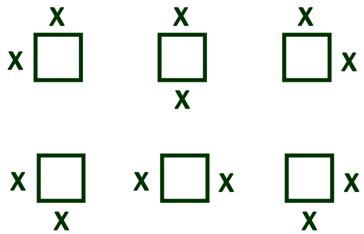
Zdarzenia elementarne - przykłady

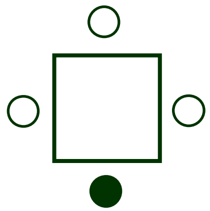
- rzut dwoma kostkami do gry: $\Omega = \{x, y\}$, gdzie $x = 1,\ldots,6$ oraz $y = 1,\ldots,6$

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- suma oczek wyrzuconych na dwóch kostkach $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
- dwie osoby przy kwadratowym stole – interesuje nas zdarzenie A: „siedzą w rogu”. przykłady różnych wyborów przestrzeni zdarzeń elementarnych:







Zdarzenia elementarne - przykłady

Przykłady przestrzeni zdarzeń elementarnych, w których możliwa jest tylko częściowa obserwacja zdarzeń elementarnych:

- **Kandydat do pracy powinien charakteryzować się pewnymi zdolnościami na poziomie z. Jednak dysponujemy jedynie wynikiem x testu oceniajacego tę zdolność.**

  Dysponujemy jedynie ograniczoną znajomością przestrzeni $\Omega = \{z, x\}$

  Chcemy znaleźć kryteria selekcji z_0, x_0, takie aby stosując regułę „akceptuj kandydatów z wynikiem $x > x_0$" jednocześnie maksymalizować liczbę akceptowanych osób, które spełniają warunek $z > z_0$ ...

  Uwaga: konieczna jest znajomość statystycznej relacji pomiędzy x i z.

- **„odpowiedzi randomizowane" – dotyczy ankiet na tematy, na które pytani z reguły nie udzielają szczerej odpowiedzi lub nie chcą w ogóle odpowiadać (seks, narotyki, łamanie prawa, podatki, ...)**

  Ankietowany rzuca dwoma kostkmi: zieloną i białą, i tylko on zna wynik rzutów. Następnie odpowiada na jedno z dwóch pytań „Czy jesteś osobą typu Q?" lub „Czy na białej kostce wypadła 1?" w zależności od tego czy na zielonej kostce wypadła nieparzysta czy parzysta liczba oczek.

Pojęcie prawdopodobieństwa

<u>Pewniki rachunku prawdopodobieństwa (Kolmogorov 1933):</u>

1) Każdemu zdarzeniu losowemu $A \subset \Omega$ przypisujemy liczbę P(A), zwaną prawdopodobieństwem tego zdarzenia, taką że $0 \leqslant P(A) \leqslant 1$.
2) P-two zdarzenia pewnego jest równe jedności $P(\Omega) = 1$.
3) P-two sumy ekskluzywnych zdarzeń losowych A i B, czyli takich że $A \cap B = \emptyset$, jest równe sumie p-tw tych zdarzeń $P(A \cup B) = P(A) + P(B)$

<u>Klasyczna definicja prawdopodobieństwa:</u>

$$P(A) \equiv \frac{n_A}{n} \equiv \frac{\text{liczba zdarzen elementarnych sprzyjajacych zdarzeniu A}}{\text{liczba wszystkich zdarzen elementarnych}}$$

<u>Definicja częstościowa prawdopodobieństwa:</u> Miarą p-twa jest względna częstości występowania zdarzenia A kiedy liczba obserwacji dąży do nieskończoności.

<u>Prawdopodobieństwa subiektywne:</u> (przydatne gdy mamy do czynienia ze zdarzeniami które są niepowtarzalne, np. niepewności systematyczne, p-two że istnieje bozon Higgs'a)

- A, B, ... traktujemy jako hipotezy (tzn. stwierdzenia które albo są prawdziwe albo fałszywe); P(A) = miara naszej wiary w to, że A jest prawdziwe

Przykład: wybór pomiędzy otrzymaniem gotówki a udziałem w loterii

Zdarzenia elementarne - przykłady

<u>Przykład:</u> Rozmieszczenie r = 3 kul (rozróżnialnych) w n = 3 komórkach:

1. $\{abc\|-\|-\}$	10. $\{a\|bc\|-\}$	19. $\{-\|a\|bc\}$
2. $\{-\|abc\|-\}$	11. $\{b\|ac\|-\}$	20. $\{-\|b\|ac\}$
3. $\{-\|-\|abc\}$	12. $\{c\|ab\|-\}$	21. $\{-\|c\|ab\}$
4. $\{ab\|c\|-\}$	13. $\{a\|-\|bc\}$	22. $\{a\|b\|c\}$
5. $\{ac\|b\|-\}$	14. $\{b\|-\|ac\}$	23. $\{a\|c\|b\}$
6. $\{bc\|a\|-\}$	15. $\{c\|-\|ab\}$	24. $\{b\|a\|c\}$
7. $\{ab\|-\|c\}$	16. $\{-\|ab\|c\}$	25. $\{b\|c\|a\}$
8. $\{ac\|-\|b\}$	17. $\{-\|ac\|b\}$	26. $\{c\|a\|b\}$
9. $\{bc\|-\|a\}$	18. $\{-\|bc\|a\}$	27. $\{c\|b\|a\}$

$$P(\mathrm{E}_i) = \frac{1}{27}$$

$$i = 1, \ldots, 27$$

- Zdarzenie A: „w jednej z komórek są co najmniej dwie kule” – zdarzenia el. 1–21
- Zdarzenie B: „pierwsza komórka nie jest pusta” – zdarzenia el. 1, 4–15, 22–27
- Zdarzenie C: „zachodzi zarówno A jak i B” – zdarzenia el. 1, 4–15
- Zdarzenie D: „zachodzi A lub B” – zdarzenia el. 1–27 (cała przestrzeń)

Zdarzenia elementarne - przykłady

Przykłady zagadnień typu „r kul w n komórkach”:

- Dni urodzin: r – osób, n = 365 dni
- Wypadki: r – wypadków, n = 7 dni tygodnia
- Strzelanie do celu: r – trafień, n – celów
- Klasyfikacja elementów (np. ludzi) wg. kategorii (wiek): r – elementów, n – kategorii
- Opuszczanie windy przez ludzi: r – pasażerów, n – pięter
- Rzuty kośćmi do gry: r – kości (rzutów), n = 6 możliwych wyników danego rzutu

Przypadek kul nierozróżnialnych:

$$1.\ \{\bullet\bullet\bullet|-|-\} \quad 4.\ \{\bullet\bullet|\bullet|-\} \quad 7.\ \{\bullet|-|\bullet\bullet\} \quad 10.\ \{\bullet|\bullet|\bullet\}$$

$$2.\ \{-|\bullet\bullet\bullet|-\} \quad 5.\ \{\bullet\bullet|-|\bullet\} \quad 8.\ \{-|\bullet\bullet|\bullet\}$$

$$3.\ \{-|-|\bullet\bullet\bullet\} \quad 6.\ \{\bullet|\bullet\bullet|-\} \quad 9.\ \{-|\bullet|\bullet\bullet\}$$

- $P(\mathrm{E}_1) = P(\mathrm{E}_2) = P(\mathrm{E}_3) = \frac{1}{27}$ $\quad P(\mathrm{E}_4) = \ldots = P(\mathrm{E}_9) = \frac{1}{9}$ $\quad P(\mathrm{E}_{10}) = \frac{2}{9}$
- $P(\mathrm{E}_i) = \frac{1}{10} \qquad i = 1, \ldots, 10$

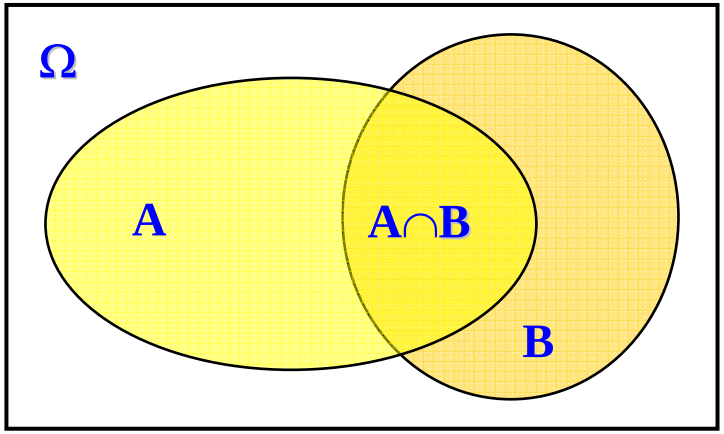
Przypadek nierozróżnialnych zarówno kul jak i komórek:

$$1.\ \{\bullet\bullet\bullet|-|-\} \quad 2.\ \{\bullet\bullet|\bullet|-\} \quad 3.\ \{\bullet|\bullet|\bullet\}$$

Działania na zbiorach

Prawa de Morgana:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$$



Prawa rozdzielności dla dodawania i mnożenia:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Wniosek z pewnika (3):

Mamy $A \cup B = A \cup (B - A \cap B)$ oraz $B = (A \cap B) \cup (B - A \cap B)$

$$P(A \cup B) = P(A) + P(B - A \cap B) \qquad P(B) = P(A \cap B) + P(B - A \cap B)$$

Odejmując stronami dostajemy: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Przykład:

$$P(A \cup B \cup C) = P(A \cup (B \cup C)) = P(A) + P(B \cup C) - P(A \cap (B \cup C))$$

$$= P(A) + P(B) + P(C) - P(B \cap C) - P((A \cap B) \cup (A \cap C)) =$$

$$= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

P-two zdarzenia przeciwnego do zdarzenia A: $P(\overline{A}) = 1 - P(A)$

Elementy kombinatoryki

Jeżeli przestrzeń zdarzeń elementarnych jest skończona, to obliczanie prawdopodobieństw zdarzeń będących podzbiorami tej przestrzeni ułatwiają pojęcia i twierdzenia z kombinatoryki.

Reguła iloczynu: jeśli pewną czynność wykonuje się w k etapach, przy czym etap 1 można wykonać na n_1 sposobów, etap 2 na n_2 sposobów, ... , wreszcie etap k-ty na n_k sposobów, to liczba N sposobów jakimi można wykonać tę czynność wynosi:

$$N = n_1 n_2 \cdots n_k$$

Rozróżniamy dwa typy losowań:

- bez powtórzeń – raz wylosowany element nie wraca do populacji,
- z powtórzeniami – wylosowany element wraca do populacji przed kolejnym losowaniem.

Rozróżniamy dwa typy uporządkowania:

- kolejność losowanych elementów jest istotna (wariacje, permutacje),
- kolejność losowanych elementów nie jest istotna (kombinacje).

Wariacje z powtórzeniami

Losujemy k elementów spośród n elementów przy czym wylosowany element za każdym razem zwracamy do populacji (losowanie ze zwracaniem).
Każdy z elementów możemy wybrać na n sposobów. Oznacza to, że liczba k-wyrazowych wariacji z powtórzeniami ze zbioru n-elementowego wynosi:

$$\mathrm{W}_n^k = n^k$$

Przykłady:

- Rzucając k kostek do gry uzyskamy jedną z 6^k możliwych konfiguracji. Wśród nich w 5^k przypadkach nie będzie ani jednej jedynki, natomiast w $k{\cdot}5^{k-1}$ przypadkach wypadnie dokładnie jedna jedynka.
- Z cyfr 1, 2, 3, 4, 5 można utworzyć 5^3 trzycyfrowych liczb naturalnych, w których każda z cyfr może się powtarzać dowolną ilość razy.
- Jeśli w każdej komórce może znaleźć się dowolna liczba cząstek, to k roz-różnialnych cząstek możemy rozmieścić w n komórkach na n^k sposobów (rozkład Maxwella-Boltzmanna).

Wariacje bez powtórzeń

Losujemy k elementów spośród n elementów przy czym wylosowanego elementu nie zwracamy do populacji (losowanie bez zwracania). Pierwszy element można wybrać na n sposobów, drugi już tylko na $n-1$, trzeci na $n-2$, natomiast k-ty tylko na $n-k+1$ sposobów. Oznacza to, że liczba k-wyrazowych wariacji bez powtórzeń ze zbioru n-elementowego wynosi:

$$\mathrm{V}_n^k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

Przykłady:

- Mając do dyspozycji sześć pionowych pasów o różnych barwach możemy utworzyć 6!/(6−3)! trójkolorowych flag.
- Z cyfr 1, 2, 3, 4, 5 można utworzyć 5!/(5−3)! trzycyfrowych liczb naturalnych w których każda z cyfr może wystąpić co najwyżej jeden raz.
- Jeżeli w każdej komórce może znaleźć się tylko jedna kula, to k kul (kule są makroskopowe a więc rozróżnialne) można rozmieścić w nich na $n!/(n-k)!$ sposobów. (Analogia: komórki – piętra budynku, kule – osoby w windzie)

Permutacje bez i z powtórzeniami

Losujemy n elementów spośród n elementów bez zwracania. Pierwszy element można wybrać na n sposobów, drugi już tylko na $n-1$, trzeci na $n-2$, natomiast przedostatni tylko na 2 sposoby. Oznacza to, że liczba permutacji bez powtórzeń zbioru n-elementowego wynosi:

$$P_n = n(n-1)(n-2)\cdots 2\cdot 1 = n!$$

<u>Przykład:</u> (statystyka Maxwella-Boltzmanna) k cząstek można rozmieścić po jednej w k ustalonych komórkach na k! sposobów.

Jeśli wśród n elementów mamy k różnych elementów, z których pierwszy powtarza się n_1 razy, drugi n_2 razy, ... , k-ty n_k razy ($n_1+n_2+\ldots+n_k=n$), to liczba rozróżnialnych losowań bez zwracania, czyli liczba permutacji z powtórzeniami zbioru n-elementowego, w którym poszczególne elementy powtarzają się odpowiednio n_1, n_2, ..., n_k razy wynosi:

$$P_n(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!\,n_2!\cdots n_k!}$$

<u>Przykład:</u> (statystyka M-B) w pierwszej komórce k_1, w drugiej k_2, ..., w n-tej k_n cząstek można rozmieścić na $P_k(k_1, k_2, \ldots, k_n)$ sposobów.

Kombinacje bez powtórzeń

Losujemy k elementów spośród n elementów przy czym wylosowanego elementu nie zwracamy do populacji (losowanie bez zwracania). Nie interesuje nas również kolejność wylosowanych elementów. Mamy więc do czynienia z k-elementowymi podzbiorami zbioru n-elementowego.

Liczba k-elementowych kombinacji bez powtórzeń z n-elementowego zbioru wynosi:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Przykłady:

- Z 10 osób możemy utworzyć trzy zespoły liczące odpowiednio po 5, 3 i 2 osoby na $\binom{10}{5}\binom{5}{3} = \frac{10!}{5!(10-5)!}\frac{5!}{3!(5-3)!} = 2520$ sposobów.
- Na ile sposobów można rozmieścić 20 kul w trzech komórkach, tak aby w pierwszej było ich 10, w drugiej 6, a w trzeciej 4? Odpowiedź: $\binom{20}{10}\binom{10}{6}$
- Jeżeli w każdej komórce może znaleźć się tylko jedna cząstka, to k nierozróżnialnych cząstek można rozmieścić w $n \geqslant k$ komórkach na $\binom{n}{k}$ sposobów (rozkład Fermiego-Diraca).

Kombinacje z powtórzeniami

Rozważmy elementy n różnych rodzajów. Elementy tego samego rodzaju traktujemy jako nierozróżnialne. Każdy zbiór k elementowy ($k \lesseqgtr n$) gdy każdy element należy do jednego z tych n rodzajów nazywamy k-elementową kombinacją z powtórzeniami z n rodzajów elementów.

Liczba k-elementowych kombinacji z powtórzeniami z elementów n rodzajów jest równa:

$$\bar{C}_n^k = C_{n+k-1}^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

Przykłady:

- Rozpatrzmy rozmieszczenie $k = 8$ nierozróżnialnych kul w $n = 6$ komórkach |●●●|●||||●●●●| - liczba rozróżnialnych rozmieszczeń wynosi $\bar{C}_6^8 = \binom{13}{8} = 1287$
- Rzucając r nierozróżnialnych kostek do gry, otrzymamy $\bar{C}_6^r = \binom{r+5}{r} = \binom{r+5}{5}$ nierozróżnialnych konfiguracji (komórki to liczby oczek, kule to kostki).
- Jeśli w każdej komórce może znaleźć się dowolna liczba cząstek, to k nierozróżnialnych cząstek możemy rozmieścić w n komórkach na $\bar{C}_n^k$ sposobów (rozkład Bosego-Einsteina). Wszystkie komórki będą zajęte w $\bar{C}_n^{k-n}$ przypadkach.

Przykład: mechanika statystyczna

Każdy możliwy stan układu to punkt w przestrzeni fazowej.

- Statystyka M-B(cząstki rozróżnialne, dowolna liczba cząstek w komórce): $W_n^k = n^k$
- Statystyka B-E (cząstki nierozróżnialne, dowolna liczba cząstek w komórce):

$$\bar{C}_n^k = C_{n+k-1}^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

- Statystyka F-D (cząstki nierozróżnialne, co najwyżej jedna cząstka w komórce):

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- p-two rozmieszczenia k cząstek po jednej w k ustalonych komórkach:

M-B: $p = \frac{k!}{n^k}$ B-E: $p = \frac{1}{\bar{C}_n^k} = \frac{k!(n-1)!}{(n+k-1)!}$ F-D: $p = \frac{1}{C_n^k} = \frac{k!(n-k)!}{n!}$

- Statystyka B-E:

- wszystkie komórki zajęte: $\bar{C}_n^{k-n} = \binom{n+k-n-1}{k-n} = \frac{(k-1)!}{(k-n)!(n-1)!} = \binom{k-1}{n-1}$

- dokładnie m cząstek w ustalonej komórce: $\bar{C}_{n-1}^{k-m} = \binom{n+k-m-2}{k-m}$

- dokładnie m cząstek w jednej z komórek: $n\bar{C}_{n-1}^{k-m}$

Prawdopodobieństwo - przykłady

Przykład 1: Winda z 7 pasażerami jedzie przez 10 pięter. Jaka jest szansa, że ludzie będą wysiadali pojedynczo na piętrach?

Analogia: k=7 kul które mamy rozmieścić w n=10 komórkach. Każdy przypadek oddzielnego wysiadania odpowiada losowaniu bez zwracania. Sytuację gdy po kilku pasażerów może wysiąść na jednym z pięter możemy opisać jako losowanie ze zwracaniem. Oczywiście pasażerowie są rozróżnialni i jest istotne to kto wysiądzie na którym piętrze. Szukane prawdopodobieństwo jest więc równe:

$$P = \frac{V_{10}^{7}}{W_{10}^{7}} \cong 0.06$$

Przykład 2: Wyciągamy 5 kart z talii 52 dobrze potasowanych kart. Jakie jest p-two, że są wśród nich a) 4 asy, b) 4 asy i jeden król, c) trzy 10-tki i dwa walety, d) 9-tka, 10-tka, walet, królowa i król, e) trzy są tego samego koloru i dwie inne, f) co najmniej jeden as?

a) $P = \frac{C_4^4 \cdot C_{48}^1}{C_{52}^5} = \frac{1}{54145}$

b) $P = \frac{C_4^4 \cdot C_4^1}{C_{52}^5} = \frac{1}{649740}$

d) $P = \frac{C_4^1 \cdot C_4^1 \cdot C_4^1 \cdot C_4^1 \cdot C_4^1}{C_{52}^5} = \frac{64}{162435}$

c) $P = \frac{C_4^3 \cdot C_4^2}{C_{52}^5} = \frac{1}{108290}$

e) $P = \frac{4C_{13}^3 \cdot 3C_{13}^2}{C_{52}^5} = \frac{429}{4165}$

f) $P = 1 - \frac{C_{48}^5}{C_{52}^5} = 1 - \frac{35673}{54145} = \frac{18482}{54145}$