Uwagi do zadań z listy nr 3

Antoni Kościelski

1 Zadanie 1

W związku z problemami, które pojawiły się na ćwiczeniach, może jednak rozwiążę zadanie 1. W tym zadaniu (choć nie jest to jasno powiedziane) mamy wyliczyć prawdopodobieństwa jakiegoś zdarzenia mając dane prawdopodobieństwa innych. W związku z tym musimy wiedzieć, co oznacza słowo prawdopodobieństwo.

Słowo to jest także używane w języku potocznym. W matematyce posługujemy się jednak ściśle określonymi pojęciamy. Aby rozwiązać zadanie 1 musimy w pierwszym rzędzie znać matematyczną definicję tego pojęcia. Najlepsza (moim zdaniem) definicja, w rzeczywistości tzw. przestrzeni probabilistycznej, jest następująca:

1. przestrzeń probabilistyczna (Ω, Σ, P) składa się z trzech elementów takich, że

2. Ω jest nazywane zbiorem zdarzeń elementarnych i jest dowolnym niepustym zbiorem,

3. Σ jest niepustym σ-ciałem podzbiorów Ω, a więc $\Sigma \subseteq 2^\Omega$, elementy Σ nazywamy zdarzeniami, a przez σ-ciało rozumiemy dowolną rodzinę zbiorów zamkniętą ze względu na skończone działania mnogościowe (sumę i dopełnienie (do Ω), i w konsekwencji także na przekrój, różnicę i różnicę symetryczną) oraz na działania przeliczalne (sumę i przekrój nieskończonego ciągu zbiorów),

4. P jest funkcją, którą nazywamy prawdopodobieństwem, która dowolnemu elementowi $A \in \Sigma$ przyporządkowuje liczbę rzeczywistą $P(A)$ z przedziału $[0, 1]$ zwaną prawdopodobieństwem zdarzenia A, i która ma następujące własności

   (a) $P(\Omega) = 1$,
   (b) $P(A \cup B) = P(A) + P(B)$ dla dowolnych rozłącznych zdarzeń A i B,
   (c) $P(\bigcup_{i=0}^\infty A_i) = \lim_{i\to\infty} P(A_i)$ (ten warunek zwykle jest zastępowany innym, patrz Uwagi do listy 2).

Znamy już definicję pojęcia prawdopodobieństwa. Matematycy znają wiele mniej lub bardziej ciekawych konkretnych przykładów prawdopodobieństw. Rozwiązując zadanie 1 powinniśmy odwołać się do przytoczonej definicji.

Aby rozwiązać to zadanie wystarczy zauważyć, że ponieważ zdarzenie A i zdarzenie do niego przeciwne (czyli dopełnienie) A^c są rozłączne, więc

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) = P(A) + \frac{1}{3}.$$

Stąd $P(A) = \frac{2}{3}$. Podobnie

$$\frac{1}{2} = P(B) = P(B \setminus A \cup A \cap B) = P(B \setminus A) + P(A \cap B) = P(B \setminus A) + \frac{1}{4}.$$

Tak więc $P(B \setminus A) = \frac{1}{4}$. W końcu

$$P(A \cup B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) = \frac{2}{3} + \frac{1}{4} = \frac{11}{12}.$$

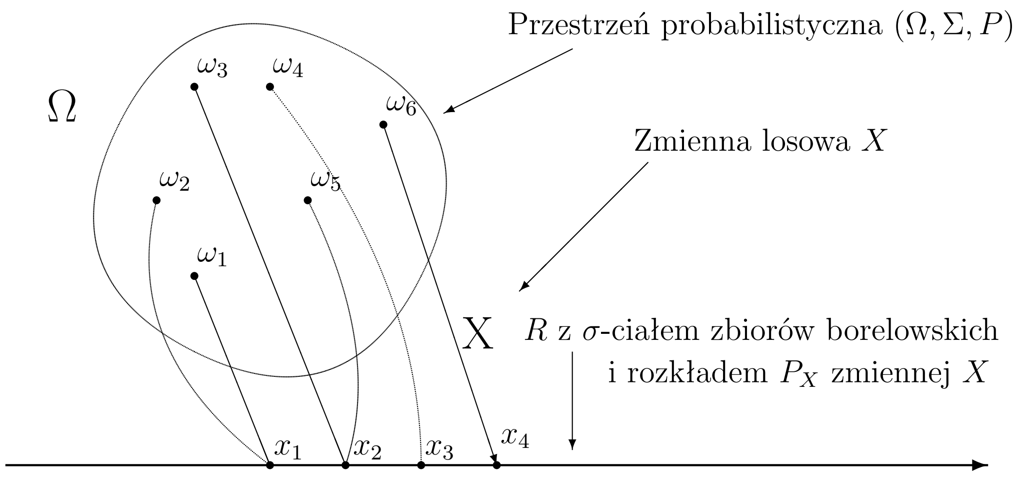
To samo rozumowanie pozwala na udowodnienie następującego wzoru

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

który mógł zostać wykorzystany w rozwiązaniu zadania.

2 Zadanie 3

Jeżeli mówimy o zmiennej losowej, to zwykle mamy sytucję taką, jak na rysunku:



O tej sytuacji można myśleć na dwa sposoby. Przestrzeń probabilistyczna (Ω, Σ, P) może być konkretnym obiektem matematycznym, ścisłe zdefiniowanym, lub nawet bliżej nieokreślonym obiektem matematycznym, spełniającym jednak warunki wymienione w definicji i – dodatkowo – jakieś przyjęte założenia (np. na Ω jest określona zmienna losowa o rozkładzie Bernouliego). Myśląc inaczej, zwłaszcza, gdy mamy praktyczny problem dający się rozwiązać statystycznie, możemy uważać tę przestrzeń za fragment rzeczywistości, o którym chcemy się coś dowiedzieć. Wtedy powinniśmy zadbać o to, aby były spełnione założenia leżące u podstaw stosowanych metod matematycznych. W tym drugim przypadku będziemy unikać z korzystania z przestrzeni (Ω, Σ, P), informacje o tej przestrzeni będziemy zdobywać w specyficzny sposób, wykonując np. serie doświadczeń. Jeżeli w końcu uznamy, że znamy rozkład interesującej nas zmiennej, to będziemy mogli wyliczyć interesujące nas prawdopodobieństwa.

Związek między prawdopodobieństwem P i rozkładem P_X zmiennej wyrażają wzory

$$P_X(A) = P(X \in A) = P(\{\omega \in \Omega | X(\omega) \in A\}),$$

gdzie środkowy wzór jest często stosowaną konwencją (typy się trochę nie zgadzają), albo w konkretnym przypadku

$$P_X((-\infty, a]) = P(X \leqslant a) = P(\{\omega \in \Omega | X(\omega) \leqslant a\}).$$

2.1 Rozwiązanie zadania

W zadaniu zmienna X ma rozkład $B(m, p)$. Oznacza to na przykład, że zmienna X przyjmuje jako wartości liczby naturalne $\leqslant m$ i nie przyjmuje innych wartości lub przynajmniej jest nieprawdopodobne, że przyjmuje inne. Znamy też rozkład tej zmiennej, a więc znamy prawdopodobieństwa

$$P(X = k) = \binom{m}{k} \cdot p^k (1-p)^{m-k} \quad \text{dla naturalnych} \quad k \leqslant m.$$

Analogiczne wzory zachodzą dla zmiennej losowej Y o rozkładzie $B(n, p)$. Mamy znaleźć rozkład zmiennej $X + Y$.

Jest oczywiste, że zmienna ta przyjmuje wartości naturalne $\leqslant m + n$. Mamy więc wyliczyć prawdopodobieństwa $P(X+Y = k)$ dla liczb naturalnych $k \leqslant m+n$. Zauważmy, że

$$X + Y = k \iff \exists l \leqslant k \;\; X = l \wedge Y = k - l$$

(kwantyfikator jest ograniczony do liczb naturalnych), a więc zdarzenie to jest sumą $k + 1$ rozłącznych zdarzeń (funkcje dla ustalonego argumentu nie przyjmują dwoch różnych wartości). Wobec tego

$$P(X + Y = k) = P(\exists l \leqslant k \;\; X = l \wedge Y = k - l) = \sum_{l=0}^{k} P(X = l \wedge Y = k - l) =$$

Korzystając teraz z założenia o niezależności X i Y otrzymujemy

$$= \sum_{l=0}^{k} P(X = l) \cdot P(Y = k - l) = \sum_{l=0}^{k} \binom{m}{l} \binom{n}{k-l} p^k (1-p)^{m+n-k}.$$

Aby dokończyć zadania należy dowieść jeszcze dla $m, n, k \in N$ słuszność wzoru

$$\sum_{l=0}^{k} \binom{m}{l} \binom{n}{k-l} = \binom{m+n}{k}.$$

Prawdę mówiąc, dojście do tego wzoru zawiera błąd. Liczba n może być duża, wtedy także k może być duże, w szczególności m może być mniejsze od k, a przyjęliśmy, że $P(X = k) = \binom{m}{k} p^k (1-p)^k$ w sytuacji, gdy jest to prawdopodobieństwo jest równe 0. Jest to drobny błąd, bo może zostać poprawiony przez dobór definicji symboli Newtona. Najlepiej je definiować rekurencyjnie, dla wszystkich naturalnych m i całkowitych k przyjmując

$$\binom{0}{0} = 1, \binom{0}{k} = 0 \quad \text{dla} \quad k \neq 0, \; \binom{m+1}{k} = \binom{m}{k-1} + \binom{m}{k}.$$

Tak zdefiniowane symbole $\binom{m}{k}$ są zwykłymi symbolami Newtona dla naturalnych $k \leqslant m$, a poza tym są równe 0. Tym samym dla takich symboli, licząc prawdopodobieństwo $P(X+Y=k)$ nie popełniliśmy błędu.

Potrzebny wzór łatwo dowodzi się przez indukcję po m dla wszystkich naturalnych n i $k>0$. Pierwszy krok jest oczywisty. Także jest oczywiste, że wzór zawsze zachodzi dla $k=0$.

$$\sum_{l=0}^{k}\binom{m+1}{l}\binom{n}{k-l}=\sum_{l=0}^{k}\left(\binom{m}{l-1}+\binom{m}{l}\right)\binom{n}{k-l}=$$

$$=\sum_{l=0}^{k}\binom{m}{l-1}\binom{n}{k-l}+\sum_{l=0}^{k}\binom{m}{l}\binom{n}{k-l}=\sum_{l=1}^{k}\binom{m}{l-1}\binom{n}{k-l}+\binom{m+n}{k}=$$

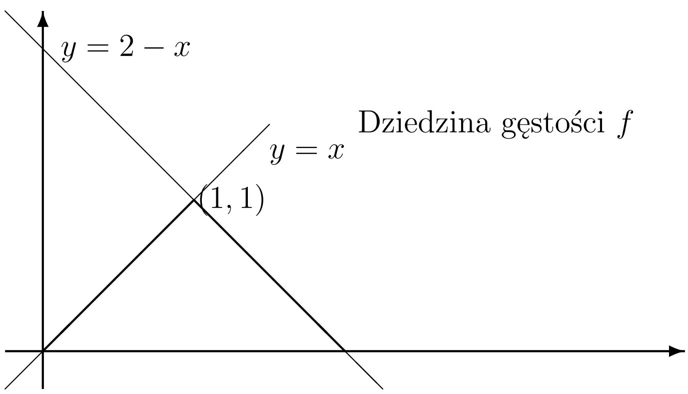
$$=\sum_{l=0}^{k-1}\binom{m}{l}\binom{n}{k-1-l}+\binom{m+n}{k}=\binom{m+n}{k-1}+\binom{m+n}{k}=\binom{m+n+1}{k}.$$

3 Zadanie 4

Analogiczne do poprzedniego.

4 Zadanie 5

Może najpierw rozwiązanie formalne. Warto przeczytać uwagi do zadania 6.



4.1 Rozwiązanie zadania

Mając gęstość f pary zmiennych X,Y, gęstość brzegową $f_1=f_X$ wyliczamy z wzoru

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy=\begin{cases}0 & \text{jeżeli } x<0 \text{ lub } x>2,\\ \int_0^x 3xy\,dy=\frac{3x^3}{2} & \text{jeżeli } 0\leqslant x\leqslant 1,\\ \int_0^{2-x}3xy\,dy=\frac{3x(x-2)^2}{2} & \text{jeżeli } 1<x\leqslant 2.\end{cases}$$

Analogicznie wyliczamy $f_2=f_Y$:

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,dx=\begin{cases}0 & \text{jeżeli } x<0 \text{ lub } x>1,\\ \int_y^{2-y}3xy\,dx=6y(1-y) & \text{jeżeli } 0\leqslant y\leqslant 1.\end{cases}$$

4.2 Kilka słów komentarza

Chyba należy dodać jeszcze parę słów o tym, co tak naprawdę liczyliśmy. Można oczywiście mówić, że mamy dany dwuwymiarowy rozkład na płaszczyźnie (lub jego gęstość f), i szukamy dwóch innych (jednowymiarowych) rozkładów z nim związanych (lub ich gęstości f_1 i f_2). Lepiej jednak mówić o tym w języku zmiennych losowych. Jeżeli mamy rozkład dwuwymiarowej zmiennej losowej X, Y, w tym jego gęstość $f = f_{X,Y}$, to oczywiście mogą nas interesować rozkłady jednowymiarowych zmiennych X i Y i ich gęstości f_X i f_Y. Pojęcie gęstości brzegowych wydaje się nieco ogólniejsze, ale w tym drugim przypadku gęstościami brzegowymi są właśnie gęstości f_X i f_Y.

Znając rozkład zmiennej dwuwymiarowej X, Y potrafimy odpowiadać na pytać na pytania o prawdopodobieństwa zdarzeń zdefiniowanych przez wskazanie własność tej pary. Przypuśćmy, że takiej sytuacji chcemy poznać prawdopodobieństwa zdefiniowanego jako własność X, na przykład interesuje nas prawdopodobieństwo tego, że zmienna X przyjmuje wartość $< t$. Wtedy mamy

$$P(X \in (-\infty, t)) = P((X, Y) \in (-\infty, t) \times R) = \iint_{(-\infty,t)\times R} f(x, y)\, dxy =$$

$$= \int_{-\infty}^{t} \int_{-\infty}^{\infty} f(x, y)\, dydx = \int_{-\infty}^{t} g(x)\, dx, \quad \text{gdzie } \; g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy.$$

Z przedstawionych rachunków właściwie wynika, że funkcja $g(x)$, oznaczana często jako $f_X = g$, jest gęstością rozkładu zmiennej X. Więcej podobnych szczegółów można znaleźć w rozwiązaniu zadania 9.

5 Zadanie 6

Rozkład zmiennej losowej, gęstość i dystrybuanta rozkładu są funkcjami i mają swoje dziedziny. Co gorsze, dziedziny te można wybierać na wiele sposobów. Podstawowy, najbardziej ogólny i niewygodny sposób jest następujący: rozkład zwykłej zmiennej losowej jest prawdopodobieństwem określonym na σ-ciele wszystkich podzbiorów borelowskich zbioru liczb rzeczywistych, a gęstości i dystrybuanty są określone dla wszystkich liczb rzeczywistych. W przypadku zmiennych dwuwymiarowych "wszystko" powinno być określone na R^2 zamiast na R.

Często (np. w zadaniu 5) rozkład definiujemy przez wskazanie gęstości określonej na mniejszej dziedzinie: zamiast na R^2 gęstość f została określona na mniejszej dziedzinie D, którą jest trójkąt uwidoczniony na rysunku (powiedzmy, że z brzegiem). Wtedy mamy dwie możliwości. Możemy myśleć, że co prawda gęstość została określona na D, ale w rzeczywistości jest określona na R^2 i poza D jest stale równa 0. Możemy też myśleć, że gęstość jednak nie jest określona poza D i wtedy "wszystko" musimy zrelatywizować do D. Ten drugi sposób jest wygodniejszy, ale trzeba pamiętać, że "wszystko" jest zrelatywizowane.

5.1 Nieco długie rozwiązanie

Podstawowe twierdzenie pozwalające na badanie niezależności zmiennych losowych w takiej sytuacji stwierdza, że

Twierdzenie 5.1 *Jeżeli X i Y są zmiennymi losowymi, ich rozkłady mają gęstości f_X i f_Y odpowiednio, a $f_{X,Y}$ jest gęstością łącznego rozkładu tych zmiennych, to zmienne te są niezależne wtedy i tylko wtedy, gdy*

$$\forall\, x \in R\, \forall\, y \in R \;\; f_{X,Y}(x, y) = f_X(x) f_Y(y). \;\; \square$$

Twierdzenie to jest dostosowane sposobu myślenia, który zakłada, że gęstości są określone dla wszystkich liczb rzeczywistych i wymaga testowania odpowiedniej równości dla wszystkich par liczb rzeczywistych. Wynika z niego natychmiast, że

Lemat 5.2 *Jeżeli gęstości f_X i f_Y niezależnych zmiennych losowych są dodatnie odpowiednio na odcinkach $[a,b]$ i $[c,d]$, to gęstość łączna $f_{X,Y}$ pary zmiennych X i Y jest dodatnia na prostokącie $[a,b]\times[c,d]$.* □

Stąd otrzymujemy, że dziedzina pary niezależnych zmiennych losowych przypomina prostąt (w prostych przypadkach jest wręcz prostokątem) i raczej nie powinna być trójkątem (chyba, że jest sztucznie określona na niepotrzebnie dużym trójkącie). Jest to zresztą zgodne z intuicjami. Jeżeli zmienne losowe mają gęstości określone na odcinkach, to dziedzina gęstości pary tych zmiennych powinna zależeć raczej od tych odcinków, być określona na iloczynie kartezjańskim tych odcinków i nie powinna w jakiś szczególny sposób zależeć od samych zmiennych.

W szczególności, samo zadanie 6 jest z rodzaju "głupich" pytań. Dobrym rozwiązania tego zadania jest stwierdzenie, że przecież zmienne z zadania nie mają prawa być niezależne: dziedzina gęstości łącznej zależy od jakiś związków między X i Y. Nie znaczy to, że zadanie jest niedobre, wszyscy powinniśmy umieć odpowiadać też na takie pytania.

Formalne rozwiązanie oczywiście wynika z przytoczonego twierdzenia. Mamy

$$\frac{27}{128}=\frac{3}{16}\cdot\frac{9}{8}=f_X(\frac{1}{2})f_Y(\frac{3}{4})\neq f_{X,Y}(\frac{1}{2},\frac{3}{4})\ \text{(wart. nieokr., ew. } =0),$$

a także

$$\frac{9}{4}=\frac{3}{2}\cdot\frac{3}{2}=f_X(1)f_Y(\frac{1}{2})\neq f_{X,Y}(1,\frac{1}{2})=\frac{3}{2}.$$

6 Kilka przydatnych wzorów

Zauważmy, że

$$(1-q)\sum_{k=0}^{\infty}q^k=\sum_{k=0}^{\infty}q^k-\sum_{k=0}^{\infty}q^{k+1}=\sum_{k=0}^{\infty}q^k-\sum_{i=1}^{\infty}q^k=1. \tag{1}$$

$$(1-q)\sum_{k=1}^{\infty}kq^{k-1}=\sum_{k=1}^{\infty}kq^{k-1}-\sum_{k=1}^{\infty}kq^k=\sum_{k=1}^{\infty}kq^{k-1}-\sum_{k=2}^{\infty}(k-1)q^{k-1}=\sum_{k=0}^{\infty}q^k. \tag{2}$$

$$(1-q)\sum_{k=2}^{\infty}k(k-1)q^{k-2}=\sum_{k=2}^{\infty}k(k-1)q^{k-2}-\sum_{k=2}^{\infty}k(k-1)q^{k-1}= \tag{3}$$

$$=\sum_{k=2}^{\infty}k(k-1)q^{k-2}-\sum_{k=3}^{\infty}(k-1)(k-2)q^{k-2}=2+\sum_{k=3}^{\infty}2(k-1)q^{k-2}=2\sum_{k=1}^{\infty}kq^{k-1}.$$

Albo inaczej, korzystając z twierdzenia o tym, że pochodna sumy szeregu potęgowego jest sumą szeregu pochodnych

$$\sum_{k=2}^{\infty}k(k-1)(k-2)q^{k-3}=\sum_{k=2}^{\infty}(k(k-1)q^{k-2})'=\left(\sum_{k=2}^{\infty}k(k-1)q^{k-2}\right)'=\left(\frac{2}{(1-q)^3}\right)'=\frac{6}{(1-q)^4}.$$

7 Zadanie 7

Przypuśćmy, że rozważamy jakiś doświaczenie Bernouliego i wynik doświadczenia jest wartością zmiennej losowej D. Przypuśćmy, że w serii k takich doświadczeń przeprowadzanych w sposób niezależny otrzymaliśmy wyniki $w_1, w_2, \ldots w_k$ (wynik to sukces lub porażka). Prawdopodobieństwo uzyskania takiego wyniku wyraża się wzorem

$$P(D_1 = w_1 \wedge \ldots D_k = w_k) = P(D_1 = w_1) \cdot \ldots \cdot P(D_k = w_k) = p^s q^{k-s},$$

gdzie s jest liczbą sukcesów w ciągu $w_1, w_2, \ldots, w_k$. Jest to konsekwencja założenia o niezależności poszczególnych doświadczeń.

Wartością zmiennej losowej X jest długość seriii doświadczeń prowadzonej do wystąpienia trzeciego sukcesu.

Zdarzenie polegające na tym, że trzeci sukces wystąpił w k-tym doświadczeniu jest sumą rozłącznych zdarzeń polegających uzyskaniu w serii k doświadczeń ciągu wyników odpowiedniej postaci. Ciągów wyników długości k, odpowiedniej postaci, a więc kończących się sukcesem, z dokładnie dwoma wcześniejszymi sukcesami, jest $\binom{k-1}{2}$. Prawdopodobieństwo uzyskania konkretnego, ciągu wyników takiej postaci, długości k, zgodnie z wcześniejszymi ustaleniami, jest równe $p^3 q^{k-3}$

Tak więc mamy

$$P(X = k) = \binom{k-1}{2} p^3 q^{k-3} = \frac{p^3}{2}(k-1)(k-2)q^{k-3} \quad \text{dla} \quad k \geqslant 3.$$

Wartość oczekiwana zmiennej X wyraża się wzorem

$$E(X) = \sum_{k=3}^{\infty} \frac{p^3}{2} k(k-1)(k-2)q^{k-3}.$$

Zgodnie z wcześniej wykazanym wzorem

$$\sum_{k=3}^{\infty} k(k-1)(k-2)q^{k-3} = \frac{6}{(1-q)^4}.$$

Stąd

$$E(X) = \frac{p^3}{2} \frac{6}{(1-q)^4} = \frac{3}{p}.$$

8 Zadanie 9

Rozwiążemy dokładniej część a). Mamy więc zmienną losową X. Co o niej wiemy? No, że właśnie mamy problem "dziedzin". Trzeba go jakoś rozwiązać. A więc zakładam, że zmienna X przyjmuje wartości z przedziału $[-2, 2]$, ma rozkład jednostajny określony na tym przedziale, to znaczy jego gęstością f_X jest funkcja określona jedynie na tym odcinku, przyjmująca stale wartość $\frac{1}{4}$. Zajmijmy się więc zmienną losową $Y = |\, X \,|$. Łatwo ustalić, że przyjmuje ona wartości w przedziale $[0, 2]$. Będziemy szukać dystrybuanty F_Y (lub podobnej funkcji) dla zmiennej Y. Wtedy dla $t \in [0, 2]$ mamy

$$F_Y(t) = P(Y < t) = P(|\, X \,| < t) = P(-t < X < t) = \int_{-t}^{t} \frac{1}{4}\, dt = \frac{t}{2}$$

(a dla innych t – niekoniecznie, zwłaszcza zawodzi trzecia równość). Gdyby zmienna losowa Y miała ciągłą gęstość f, to także

$$P(Y < t) = \int_0^t f(x)\,dx = \frac{t}{2}.$$

Główne twierdzenie z podstawowego wykładu analizy matematycznej dla informatyków, dotyczące całki Riemanna stwierdza, że dla ciągłych funkcji podcałkowych, pochodna po t

$$\left(\int_0^t f(x)\,dx\right)' = f(t), \quad \text{a więc} \quad f(t) = \left(\frac{t}{2}\right)' = \frac{1}{2}.$$

W ten sposób pokazaliśmy dwie rzeczy:

1. dystrybuanta $F_Y : [0,2] \to [0,1]$ jest dana wzorem $F_Y(t) = \frac{t}{2}$,
2. jeżeli rozkład zmiennej Y ma na odcinku $[0,2]$ ciągłą gęstość f, to $f(x) = \frac{1}{2}$.

Aby pozbyć się niepotrzebnego założenia zauważmy, że oba rozkłady: zmiennej Y i ten o gęstości f, mają tę samą dystrybuantę. Trzeba jeszcze wiedzieć, że dystrybuanta jednoznacznie wyznacza rozkład, a więc rozkład zmiennej Y jest rozkładem o gęstości f.

Może jeszcze drugi przykład bez szczegółów. Mając zmienną X o rozkładzie jednostajnym na odcinku $[-1,1]$ szukamy rozkładu zmiennej $Y = X^3$. Oczywiście Y przyjmuje wartości z przedziału $[-1,1]$ i dla t z tego przedziału

$$F_Y(t) = P(Y < t) = P(X^3 < t) = P(X < \sqrt[3]{t}) = \int_{-1}^{\sqrt[3]{t}} \frac{1}{2}\,dx = \frac{\sqrt[3]{t}+1}{2}.$$

Stąd dla $t \in [-1,1]$

$$f_Y(t) = \left(\frac{\sqrt[3]{t}+1}{2}\right)' = \frac{1}{6\sqrt[3]{t^2}}.$$

9 Zadanie 10

Zmienna X ma rozkład geometryczny, a więc przyjmuje dodatnie naturalne wartości k oraz

$$P(X = k) = pq^{k-1}.$$

W związku z tym wartość oczekiwana X wyraża się wzorem (patrz wzór (2))

$$E(X) = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

W oparciu o rachunki z przeprowadzone w części 6 łatwo liczymy różnicę

$$E(X^2) - E(X) = \sum_{k=1}^{\infty} k^2pq^{k-1} - \sum_{k=1}^{\infty} kpq^{k-1} = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} = pq\sum_{k=2}^{\infty} k(k-1)q^{k-2}$$

czyli

$$E(X^2) = E(X) + \frac{2pq}{(1-q)^3} = \frac{1}{p} + \frac{2q}{p^2} = \frac{1+q}{p^2}.$$

Stąd i ze znanego wzoru wyrażającego wariancję

$$V(X) = E(X^2) - E^2(X) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

10 Zadanie 11

To zadanie ma niezbyt ładne sformułowanie i ma chyba świadczyć o tym, że zadań tak nie należy formułować.

Tak naprawdę zwrot "losowo wybieramy pewien element" jest pozbawiony jakiejkolwiek treści. Aby wyliczyć wartość oczekiwaną musimy jednak coś wiedzieć o prawdopodobieństwie. W treści zadania, żadnych istotnych informacji na ten temat nie ma.

Prawdę mówiąc, rzecz ma się podobnie do zadań z pociągami jadącymi z miasta A do miasta B. Trudno sobie wyobrazić, aby w tych zadaniach pociągi jeździły inaczej niż jednostajnie. Podobnie, jeżeli w zadaniu "losowo wybieramy pewien element" i rozkład prawdopodobieńswa nie jest wskazany, to najprawdopodobniej losujemy to zgodnie z rozkładem jednostajnym.

Zajmijmy się zmienną X. Pojawia się kolejny problem: ile elementów ma "całość"? Poprawność definicji zmiennej X wymaga, aby wybrany element umożliwiał jednoznaczne ustalenie, z jakiego zbioru został wybrany (a przynajmniej, jaka jest jego moc, gdybyśmy wybierali elementy ze zbiorów, wśród których są równoliczne, sytuacja byłaby bardziej skomplikowana). Aby zadanie miało sens, musimy założyć, że zbiory A_1, A_2, A_3 i A_4 są rozłączne. Wtedy definiując zmienną X możemy przyjąć, że przestrzeń zdarzeń elementarnych składa się z elementów sumy $A_1 \cup A_2 \cup A_3 \cup A_4$ i ma 142 elementy. Jeżeli w takiej przestrzeni losujemy elementy zgodnie z rozkładem jednostajnym, to pojedyńczy element losujemy z prawdopodobieństwem $\frac{1}{142}$, a zdarzenie złożone z k zdarzeń elementarnych zachodzi z prawdopodobieństwem $\frac{k}{142}$. Wobec tego

$$P(X = 40) = P(A_1) = \frac{40}{142}, \quad P(X = 32) = \frac{32}{142}, \quad P(X = 20) = \frac{20}{142}, \quad P(X = 50) = \frac{50}{142}.$$

Stąd

$$E(X) = 40\frac{40}{142} + 32\frac{32}{142} + 20\frac{20}{142} + 50\frac{50}{142} = \frac{5524}{142}38\frac{63}{71}.$$

Analogicznie

$$E(Y) = 40\frac{1}{4} + 32\frac{1}{4} + 20\frac{1}{4} + 50\frac{1}{4} = \frac{142}{4} = 35\frac{1}{2}.$$

Być może to zadanie pojawiło się, by zwrócić uwagę, że jeżeli coś losujemy, to wynik istotnie zależy od sposobu losowanie (wydaje się to oczywiste) i przestrzeń probabilistyczna opisująca losowanie powinna być określona. Tu na dwa sposoby losowaliśmy jeden spośród czterech zbiorów: bezpośrednio (gdy korzystaliśmy ze zmiennej Y) i pośrednio, poprzez losowanie elementów (gdy posługiwaliśmy się zmienną X). Pierwszy sposób losowania jest uniwersalny, dobierając odpowiednio liczność zbiorów otrzymamy losowanie z zadanym rozkładem z wymiernymi prawdopodobieństwami.