Rachunek prawdopodobieństwa - Teoria - Przypomnienie

Podstawy

Definicja 1. *Schemat klasyczny - wszystkie zdarzenia elementarne są równo prawdopodobne, licząc prawdopodobieństwo liczymy stosunek liczby zdarzeń sprzyjających do wszystkich zdarzeń.*

Niech A, B zdarzenia losowe. Prawdopodobieństwo warunkowe: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
A i B są niezależne, gdy $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Fakt 1 (Wzór na prawdopodobieństwo całkowite)**.** *Niech $A, B_1, \ldots, B_n$ zdarzenia losowe, zdarzenia B_i dla $i = 1, \ldots, n$ tworzą podział Ω, czyli zawsze zachodzi dokładnie jedno z nich. Wówczas:*

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

Fakt 2 (Wzór Bayesa)**.** *Niech $A, B_1, \ldots, B_n$ zdarzenia losowe, zdarzenia B_i dla $i = 1, \ldots, n$ tworzą podział Ω, czyli zawsze zachodzi dokładnie jedno z nich. Wówczas:*

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum\limits_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Dowód. Po prostu rozpisujemy, najpierw z definicji prawdopodobieństwa warunkowego, potem korzystamy ze wzoru na prawdopodobieństwo całkowite.

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum\limits_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

□

Rozkłady dyskretne

Aby zdefiniować rozkład zmiennej losowej X o wartościach w liczbach naturalnych należy i wystarczy odpowiedzieć na pytanie ile wynosi $\mathbb{P}(X = k)$ dla $k \in \mathbb{N}$.
Przez $X \sim Y$ będziemy cały czas oznaczać (i zawsze się tak oznacza), że X i Y mają dokładnie ten sam rozkład, czyli tego samego typu i z takimi samymi parametrami.
Ważne rozkłady (dyskretne, czyli o skończonym zbiorze wartości):

Następujący schemat nazwiemy schematem Bernoulliego. Wykonujemy n niezależnych prób, każda odnosi sukces w prawdopodobieństwem p. To się często dzieje w życiu, więc z tym schematem związanych jest wiele użytecznych rozkładów.

Rozkład dwumianowy (patrzymy jakie jest prawdopodobieństwo uzyskania dokładnie k sukcesów w schemacie Bern.), oznaczamy $X \sim Bin(n, p)$.
Wówczas $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.

Rozkład geometryczny (patrzymy jaki jest czas oczekiwania na pierwszy sukces), oznaczamy powiedzmy $X \sim G(p)$.

Wówczas $\mathbb{P}(X=k) = p(1-p)^{k-1}$ dla $k \geqslant 1$.

Rozkład Poissona (jest to graniczny przypadek rozkładu dwumianowego, o czym mówi twierdzenie Poissona - później), oznaczamy $X \sim Poiss(\lambda)$.
Wówczas $\mathbb{P}(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}$.

Twierdzenie 1 (Twierdzenie Poissona)**.** *Niech $X_n \sim Bin(n, p_n)$, $Y \sim Poiss(\lambda)$ oraz $\lim_{n\to\infty} np_n = \lambda$. Wówczas*

$$\forall_{k\in\mathbb{N}} \lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k).$$

Dowód. Niech $\lambda_n = np_n$. Wówczas $\lambda_n \to \lambda$. Wówczas dla dowolnego $k \in \mathbb{N}$ zachodzi

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(X_n = k) &= \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} p_n^k \frac{(1-p_n)^n}{(1-p_n)^k} \\ &= \frac{1}{k!} \lim_{n\to\infty} n(n-1)\dots(n-k+1) p_n^k \Big(\frac{1}{1-p_n}\Big)^k \Big(1 - \frac{\lambda_n}{n}\Big)^n \\ &= \frac{1}{k!} \lim_{n\to\infty} n^k \frac{n}{n}\frac{n-1}{n} \dots \frac{n-k+1}{n} p_n^k \Big(\frac{1}{1-p_n}\Big)^k \Big(1 - \frac{\lambda_n}{n}\Big)^n \\ &= \frac{1}{k!} \lim_{n\to\infty} (\lambda_n)^k \frac{n}{n}\frac{n-1}{n} \dots \frac{n-k}{n} \Big(\frac{1}{1-p_n}\Big)^k \Big(1 - \frac{\lambda_n}{n}\Big)^n\end{aligned}$$

$$= \frac{1}{k!}\lambda^k e^{-\lambda} = \mathbb{P}(Y = k).$$

□

Momenty

Definicja 2. *Niech X zmienna losowa o wartościach w liczbach naturalnych. Wówczas definiujemy:*
wartość oczekiwana to $\mathbb{E}X = \sum_{k=1}^{\infty} k\mathbb{P}(X = k)$
wariancja to $VarX = \mathbb{E}((X - \mathbb{E}X)^2)$
odchylenie standardowe to $\sigma(X) = \sqrt{Var\ X}$
k-ty moment to $\mathbb{E}X^k$

Wartość oczekiwana to średnia, odchylenie standardowe to średnie odchylenie od tej średniej. Wariancja to kwadrat tego odchylenia.

Własności:

$$\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$$

$$Var\ tX = \mathbb{E}(tX)^2 - (\mathbb{E}tX)^2 = t^2\mathbb{E}X^2 - t^2(\mathbb{E}X)^2 = t^2VarX$$

Dla X, Y niezależnych

$$\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$$

$$Var(X + Y) = VarX + VarY$$

Inne wyrażenie $\mathbb{E}$ (dla zmiennych o wartościach w $\mathbb{N}$)

$$\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geqslant k)$$

Można łatwo pokazać jego równoważność definicji licząc ile razy w każdej z sum pojawił się składnik $\mathbb{P}(X = k)$ dla określonego k.

Łatwiejszy wzór na wariancję

$$\mathbb{E}((X - \mathbb{E}X)^2) = \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

Funkcja tworząca prawdopodobieństwa to dobre narzędzie, ułatwia zrobienie wielu zadań.

Definicja 3. *Funkcją tworzącą prawdopodobieństwa zmiennej losowej X nazwiemy funkcję $f : \mathbb{R} \to \mathbb{R}$ określoną następująco:*

$$f_X(t) = \mathbb{E}t^X.$$

Rozwijając definicję dostajemy

$$f_X(t) = \mathbb{E}t^X = \sum_{k=0}^{\infty} \mathbb{P}(X = k)t^k.$$

Funkcja tworząca ma wiele dobrych własności, najważniejsze to

- znając funkcję tworzącą zmiennej losowej X możemy odtworzyć jej rozkład, stosuje się to często mówiąc, że skoro $f_X = f_Y$, to $X \sim Y$
- z funkcji tworzącej f_X łatwo odzyskujemy $\mathbb{E}X$ oraz $VarX$, mianowicie $\mathbb{E}X = f'_X(1)$, $VarX = f''_X(1) + f'_X(1) - f'_X(1)^2$
- jeśli zmienne losowe X i Y są niezależne, to $f_{X+Y} = f_X \cdot f_Y$, z ten sposób można łatwo uzyskiwać własności sumy zmiennych losowych

Uzasadnienie:
Z definicji mamy

$$f_X(t) = \sum_{k=0}^{\infty} t^k \mathbb{P}(X = k).$$

Po jednokrotnym zróżniczkowaniu stronami otrzymujemy (należałoby formalnie uzasadnić dlaczego możemy różniczkować szereg wyraz po wyrazie, ale to na razie odpuśćmy)

$$f'_X(t) = \sum_{k=0}^{\infty} kt^{k-1}\mathbb{P}(X = k) = \sum_{k=1}^{\infty} kt^{k-1}\mathbb{P}(X = k)$$

Analogicznie po n krotnym zróżniczkowaniu mamy

$$f_X^{(n)}(t) = \sum_{k=0}^{\infty} k^{\underline{n}} t^{k-n}\mathbb{P}(X = k) = \sum_{k=n}^{\infty} k^{\underline{n}} t^{k-n}\mathbb{P}(X = k),$$

gdzie przez $k^{\underline{n}}$ oznaczamy $k(k-1)\dots(k-n+1)$.

Aby odtworzyć rozkład X z f_X wystarczy zauważyć, że

$$f_X^{(n)}(0) = \sum_{k=n}^{\infty} k^{\underline{n}} 0^{k-n} \mathbb{P}(X = k) = n^{\underline{n}} \cdot \mathbb{P}(X = n) = n! \cdot \mathbb{P}(X = n),$$

gdyż wszystkie wyrazy poza $k = n$ wynoszą 0, czyli mamy wzór $\mathbb{P}(X = n) = \frac{f_X^{(n)}(0)}{n!}$.

Mamy również

$$f'_X(1) = \sum_{k=1}^{\infty} k 1^{k-1} \mathbb{P}(X = k) = \mathbb{E}X.$$

oraz

$$f''_X(1) = \sum_{k=2}^{\infty} k(k-1) 1^{k-2} \mathbb{P}(X = k),$$

czyli

$$f''_X(1) + f'_X(1) = \sum_{k=1}^{\infty} \mathbb{P}(X = k)(k(k-1) + k) = \sum_{k=2}^{\infty} \mathbb{P}(X = k) k^2 = \mathbb{E}X^2.$$

Zatem $Var\ X = f''_X(1) + f'_X(1) - f'_X(1)^2$.

Własność zmiennych niezależnych wynika z tego, że $f_{X+Y}(t) = \mathbb{E}t^{X+Y} = \mathbb{E}t^X t^Y = \mathbb{E}t^X \cdot \mathbb{E}t^Y = f_X(t) \cdot f_Y(t)$, gdzie trzecia równość wynika z niezależności.

Jest jeszcze jedna przydatna własność funkcji tworzącej prawdopodobieństwa, o której mówi

Twierdzenie 2. *Jeśli $X, X_1, X_2, \ldots$ niezależne zmienne losowe o jednakowym rozkładzie (X wprowadzam tylko po to, żeby miał ten sam rozkład, co X_i i napisy wyglądały ładnie), N zmienna losowa niezależna od wszystkich pozostałych oraz $S = \sum_{i=1}^{N} X_i$, to $f_S = f_N \circ f_X$.*

Rozkłady ciągłe

Rozkłady ciągłe na tym przedmiocie traktowane są bardzo, bardzo skrótowo. Rozkłady ciągłe określamy podając nie prawdopodobieństwo przyjęcia konkretnej wartości (gdyż prawdopodobieństwo przyjęcia konkretnej wartości wynosi zawsze 0 w przypadku tych rozkładów), lecz gęstość rozkładu w danym miejscu. Przyjmujemy potem, że $\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b g_X(x)dx$. Najważniejsze rozkłady ciągłe to:

Rozkład jednostajny (losujemy liczbę z odcinka $[a, b]$ z taką samą gęstością na całym odcinku), oznaczamy $X \sim U([a, b])$.
Wówczas $g_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}$, gdzie przez $\mathbb{1}_S$ oznaczamy funkcję, która na zbiorze S przyjmuje wartość 1, a na reszcie 0.

Rozkład wykładniczy (możemy o nim myśleć jako o uciągleniu rozkładu geometrycznego, czyli o czasie czekania na pewne wydarzenie, które ma prawd. zajścia zawsze takie samo), oznaczamy $X \sim \mathcal{E}xp(\lambda)$.
Wówczas $g_X(x) = \lambda e^{-x\lambda} \mathbb{1}_{\{x \geqslant 0\}}$.

Rozkład normalny (inaczej gaussowski, najważniejszy rozkład wszechczasów :)). Występuje często w przyrodzie i nie tylko, a to dlatego, że można go traktować jako rozkład, który jest

uciągleniem sumy wielu podobnych niezależnych zmiennych losowych. Sformalizowaniem tego faktu jest Centralne Twierdzenie Graniczne. Oznaczamy go $X \sim \mathcal{N}(a, \sigma^2)$, gdzie a to wartość oczekiwana X, a σ^2 to wariancja X.
Wówczas $g_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-a)^2}{2\sigma^2}}$.
Często stosowany jest tak zwany kanoniczny rozkład normalny, $\mathcal{N}(0, 1)$.

Wartość oczekiwaną rozkładu ciągłego definiuje się następująco

$$\mathbb{E}X = \int\limits_{-\infty}^{\infty} g_X(x)\ x\ dx.$$

Nierówności probabilistyczne

Nierówność Markowa
Niech X nieujemna zmienna losowa, $t > 0$, wówczas

$$\mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}X}{t}.$$

Dowód: Gdyby $\mathbb{P}(X \geqslant t) > \frac{\mathbb{E}X}{t}$, to

$$\mathbb{E}X \geqslant \mathbb{E}X\mathbb{1}_{\{X \geqslant t\}} \geqslant \mathbb{E}t\mathbb{1}_{\{X \geqslant t\}} = t\mathbb{E}\mathbb{1}_{\{X \geqslant t\}} = t\mathbb{P}(X \geqslant t) > t\frac{\mathbb{E}X}{t} = \mathbb{E}X,$$

sprzeczność.

Nierówność Czebyszewa

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant t) \leqslant \frac{Var\ X}{t^2}.$$

Dowód: $\mathbb{P}(|X - \mathbb{E}X| \geqslant t) = \mathbb{P}(|X - \mathbb{E}X|^2 \geqslant t^2) \leqslant \frac{\mathbb{E}|X-\mathbb{E}X|^2}{t^2} = \frac{Var\ X}{t^2}$, gdzie nierówność wynika z nierówności Markowa dla zmiennej $Y = |X - \mathbb{E}X|^2$.

Nierówność Chernoffa
Zdefiniujmy funkcję tworzącą momenty $M_X(s) = \mathbb{E}e^{sX}$. Wówczas

$$\mathbb{P}(X \geqslant t) \leqslant \min_{s>0} \frac{M_X(s)}{e^{st}}$$

oraz

$$\mathbb{P}(X \leqslant t) \leqslant \min_{s<0} \frac{M_X(s)}{e^{st}}.$$

Dowód:
Niech $s > 0$, wówczas $\mathbb{P}(X \geqslant t) = \mathbb{P}(e^{sX} \geqslant e^{st}) \leqslant \frac{\mathbb{E}e^{sX}}{e^{st}} = \frac{M_X(s)}{e^{st}}$,gdzie pierwsza równość wynika z tego, że funkcja $x \to e^{sx}$ jest rosnąca, a jedyna nierówność jest zastosowaniem nierówności Markowa. Taka nierówność zachodzi dla dowolnego $s > 0$, więc jak prawą stronę zminimalizujemy po $s > 0$, to będzie też prawda. Podobnie druga nierówność.

Poszczególne przypadki zastosowania nierówności Chernoffa sprowadzają się zazwyczaj do policzenia funkcji $M_X(s)$ dla pewnej konkretnej zmiennej X i wstawienia. Na przykład, gdy X jest zmienną Bernoulliego (czyli przyjmuje 1 z pr. p, 0 wpp.), to $M_X(s) = (1 - p) + pe^s$, obliczamy minimum funkcji po prawej stronie i zapewne można otrzymać coś interesującego.

Łańcuchy Markowa

Teorię łańcuchów Markowa stosuje się standardowo do sytuacji, w których stan w chwili n zależy tylko i wyłącznie od tego, jaki był stan w chwili $n-1$ oraz od reguł przejścia. Można patrzeć na taką sytuację jako na pewien układ bez pamięci.

Zakładamy więc, że w układzie jest dokładnie n stanów. Prawdopodobieństwa przejścia między nimi nie zależą od numeru chwili, więc oznaczmy przez p_{ij} prawdopodobieństwo przejścia ze stanu nr i do stanu numer j. Możemy wówczas cała sytuację przedstawić w postaci macierzy, tak zwanej macierzy przejścia:

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

Zauważmy, że jeśli teraz przez $s = (s_1, s_2, \ldots, s_n)$ oznaczymy wektor prawdopodobieństw przebywania w chwili k w stanach $1, 2, \ldots, n$, to wówczas wektor prawdopodobieństw przebywania w chwili $k+1$ w odpowiednich stanach równy jest $s \cdot P$ (warto przyjrzeć się dlaczego tak naprawdę tak jest, co znajduje się na poszczególnych miejscach w wektorze $s \cdot P$).

Wprowadźmy teraz terminologię do mówienia o ł.M. Niech $p_{ij}(n)$ - prawdopodobieństwo dojścia ze stanu i do stanu j w dokładnie n ruchach. Stan j jest osiągalny ze stanu i jeśli $\exists_{n\in\mathbb{N}} p_{ij}(n) > 0$, oznaczamy $i \to j$. Jeśli $i \to j$ oraz $j \to i$, to mówimy, że stany i oraz j komunikują się ($i \leftrightarrow j$).

Jeśli $\forall_{i,j} i \leftrightarrow j$, to mówimy, że ł.M. jest nieprzywiedlny, intuicja tego pojęcia powinna być taka, że skoro wszystkie stany się ze sobą komunikują, to nie ma w tym ł.M. dziwnych rzeczy, a wszystkie stany, dlatego, że się komunikują okażą się być w pewnym sensie podobne.

Niech $p_{ij} = \sum_{n=1}^{\infty} p_{ij}(n)$, czyli teraz p_{ij} jest czymś innym, niż to stare, będące tak naprawdę $p_{ij}(1)$. Jeśli $p_{ii} = 1$, to mówimy, że stan i jest powracający, w przeciwnym przypadku - chwilowy. Jeśli $i \leftrightarrow j$ oraz i powracający, to j powracający, analogicznie chwilowy. Zatem jeśli ł.M. jest nieprzywiedlny, to możemy mówić, że cały jest powracający, lub chwilowy.

Okres stanu j to $o(j) = NWD\{n : p_{jj}(n) > 0\}$. Jeśli stan ma okres równy k, to znaczy, że tylko co k ruchów możemy się w nim pojawić. Okres różny od 1 mają tylko patologiczne stany w ł.M.. Przykładowo w ł.M., który jest cyklem o długości m wszystkie stany będą miały okres m, a w ł.M. takim, że $\forall_{i,j} p_{ij}(1) > 0$ wszystkie stany będą miały okres równy 1.

Okazuje się, że w nieprzywiedlnym ł.M. wszystkie stany mają ten sam okres (to nie jest trudne do wykazania). Jeśli ten okres jest równy 1, to mówimy, że łańcuch Markowa jest nieokresowy.

Stanem stacjonarnym w ł.M. nazwiemy taki rozkład prawdopodobieństw (nazywany tu stanem) na wierzchołkach, czyli $s = (s_1, s_2, \ldots, s_n)$, że po ruchu rozkład prawdopodobieństw będzie taki sam. Zauważmy, że wówczas s musi spełniać $sP = s$.

Większość tych definicji służyło do tego, żeby sformułować bardzo ważne twierdzenie w teorii łańcuchów Markowa, mianowicie twierdzenie ergodyczne. Intuicyjnie mówi ono, że jeśli mamy taki nieudziwniony łańcuch Markowa, czyli wszystko się z wszystkim komunikuje i nie ma dziwacznych okresów, to po pewnym czasie rozkład prawdopodobieństw na jego stanach się z grubsza ustabilizuje i to w dodatku będzie blisko stanu stacjonarnego (jedynego). Formalnie

Twierdzenie 3 (ergodyczne)**.** *Jeśli łańcuch Markowa jest nieprzywiedlny i nieokresowy, to istnieje dokładnie jeden stan stacjonarny $s = (s_1, \ldots, s_k)$, czyli taki, że $sP = s$ oraz zachodzi*

$$\forall_{i,j} \lim_{n\to\infty} p_{ij}(n) = s_j.$$

Oprócz tego teoria łańcuchów Markowa pozwala łatwo liczyć prawdopodobieństwa oraz wartości oczekiwane pewnych zdarzeń, nie z samej definicji, lecz używając innych metod - rozwiązywania równań.

Przedstawimy te metody na przykładzie. Przykład ten nie jest tylko jednym, losowym zadaniem, większość prostych zadań na ł.M. robi się właśnie tymi metodami.

Zadanie Dwóch graczy rzuca symetryczną monetą. Założyli się, pierwszy twierdzi, że w ciągu wyników najpierw pojawi się ciąg OOR (orzeł, orzeł, reszka), a drugi, że najpierw pojawi się ciąg ORR. Jakie są prawdopodobieństwa wygranej odpowiednich graczy oraz jaki jest oczekiwany czas gry, jeśli zakładamy, że gra skończy się przy wygranej jednego z graczy?

Rozwiązanie Jak zwykle w zadaniach na ł.M. nie widać ł.M. w treści, jednak nie przejmujmy się tym, stworzymy go - jak zresztą zazwyczaj się robi.

Zastanówmy się jakie są możliwe rozróżnialne sytuacje podczas gry. Istotne w grze jest to, jak dużo na swój rachunek każdy z graczy już uzbierał. Gracz pierwszy mógł zatem niczego jeszcze nie uzbierać, mógł zebrać jednego orła (O), mógł zebrać dwa orły (OO), bądź mógł wygrać już całą grę (OOR). Podobnie sytuacja drugiego gracza może przestawiać się jako: $\emptyset$, (O), (OR) lub (ORR).

Zatem w naszym łańcuchu warto rozróżnić 6 stanów: $\emptyset$, O, OO, OR, OOR oraz ORR, oznaczają one co w danym momencie gry zostało już uzbierane, na przykład po ciągu $RROROROROR RO$ jesteśmy w stanie O.

Prawdopodobieństwa przejść między tymi stanami przedstawiają się następująco:
$\mathbb{P}(\emptyset \to O) = \frac{1}{2}$ (po wyrzuceniu O)
$\mathbb{P}(\emptyset \to \emptyset) = \frac{1}{2}$ (po wyrzuceniu R)
$\mathbb{P}(O \to OO) = \frac{1}{2}$ (po wyrzuceniu O)
$\mathbb{P}(O \to OR) = \frac{1}{2}$ (po wyrzuceniu R)
$\mathbb{P}(OO \to OO) = \frac{1}{2}$ (po wyrzuceniu O)
$\mathbb{P}(OO \to OOR) = \frac{1}{2}$ (po wyrzuceniu R)
$\mathbb{P}(OR \to O) = \frac{1}{2}$ (po wyrzuceniu O)
$\mathbb{P}(OR \to ORR) = \frac{1}{2}$ (po wyrzuceniu R)

Po trafieniu do OOR oraz ORR już nigdy stamtąd nie wyjdziemy, bo oznaczają one wygraną jednego z graczy.

Oznaczmy teraz dla dowolnego stanu v przez p_v prawdopodobieństwo wygrania gracza obstawiającego ciąg OOR pod warunkiem przebywania w stanie v. Jasne jest, że $p_{OOR} = 1$ oraz $p_{ORR} = 0$. Zauważmy, że patrząc na to, gdzie z danego stanu możemy wyjść możemy napisać równania: $p_{OO} = \frac{1}{2}p_{OOR} + \frac{1}{2}p_{OO} = \frac{1}{2} + \frac{1}{2}p_{OO}$ (ze stanu OO pójdziemy z prawd. $\frac{1}{2}$ do OO oraz z prawd. $\frac{1}{2}$ do OOR). Analogicznie
$p_{OR} = \frac{1}{2}p_O + \frac{1}{2}p_{ORR} = \frac{1}{2}p_O$
$p_O = \frac{1}{2}p_{OO} + \frac{1}{2}p_{OR}$
$p_\emptyset = \frac{1}{2}p_\emptyset + \frac{1}{2}p_O$
Z równań tych możemy obliczyć, że $p_{OO} = 1$, $p_{OR} = \frac{1}{3}$, $p_O = \frac{2}{3}$, $p_\emptyset = \frac{2}{3}$. Zatem prawdopodobieństwo wygranej gracza pierwszego wynosi $p_\emptyset = \frac{2}{3}$, drugiego $1 - \frac{2}{3} = \frac{1}{3}$.

Obliczmy teraz oczekiwaną długość gry. Oznaczmy przez $\mathbb{E}_v$ oczekiwaną długość gry ze stanu v. Wiemy, że $\mathbb{E}_{OOR} = \mathbb{E}_{ORR} = 0$. Możemy, podobnie jak dla prawdopodobieństwa napisać równania (można traktować ja jako równania rekurencyjne):
$\mathbb{E}_{OO} = \frac{1}{2}(1 + \mathbb{E}_{OOR}) + \frac{1}{2}(1 + \mathbb{E}_{OO})$

$\mathbb{E}_{OR} = \frac{1}{2}(1 + \mathbb{E}_{ORR}) + \frac{1}{2}(1 + \mathbb{E}_{O})$
$\mathbb{E}_{O} = \frac{1}{2}(1 + \mathbb{E}_{OO}) + \frac{1}{2}(1 + \mathbb{E}_{OR})$
$\mathbb{E}_{\emptyset} = \frac{1}{2}(1 + \mathbb{E}_{O}) + \frac{1}{2}(1 + \mathbb{E}_{\emptyset})$
Rozwiązując te równania otrzymujemy $\mathbb{E}_{OO} = 2$, $\mathbb{E}_{OR} = \frac{8}{3}$, $\mathbb{E}_{O} = \frac{10}{3}$, $\mathbb{E}_{\emptyset} = \frac{16}{3}$. Zatem oczekiwana długość trwania gry to $\mathbb{E}_{\emptyset} = 5\frac{1}{3}$.