

Fizyka statystyczna
Rachunek prawdopodobieństwa
(szkic wstępu do zarysu problematyki)

P. F. Góra

<http://th-www.if.uj.edu.pl/zfs/gora/>

2020

Niech Ω będzie pewnym zbiorem, którego elementy nazywamy *zdarzeniami elementarnymi*. Dalej, rozważmy *zbiór borelowski* \mathcal{B} , czyli taki zbiór podzbiorów Ω , że

- $\Omega \in \mathcal{B}$
- Jeżeli $A \in \mathcal{B}$, to $\bar{A} \in \mathcal{B}$. (\bar{A} oznacza dopełnienie zbioru.)
- Jeżeli $A, B \in \mathcal{B}$, to $A \cup B \in \mathcal{B}$. Co więcej, jeżeli $\{A_i\}$ oznacza dowolną *przeliczalną* rodzinę zbiorów należących do \mathcal{B} , to $\cup_i A_i \in \mathcal{B}$.

Elementy zbioru \mathcal{B} nazywamy zdarzeniami. Widać, że zdarzenia elementarne są zdarzeniami. Z powyższych aksjomatów wynika także, że $\emptyset \in \mathcal{B}$, oraz jeżeli $A, B \in \mathcal{B}$, to $A \cap B \in \mathcal{B}$.

Przykłady

Jeśli rzucamy kostką, zdarzeniem elementarnym jest poszczególna ściana kostki; ścianom tym można przypisać liczby do 1 do 6. W tym przypadku $\Omega = \{1, 2, 3, 4, 5, 6\}$, natomiast zbiór \mathcal{B} tworzą wszystkie podzbiory Ω .

Rozważmy zbiór wszystkich liczb rzeczywistych \mathbb{R} . Wszystkie przedziały domknięte $x_1 \leq x \leq x_2$ są borelowskie. Borelowskie są zbiory jednopunktowe (jednoliczbowe), przedziały otwarte i zbiory postaci $x \leq \lambda$ dla dowolnych $\lambda \in \mathbb{R}$. **Nieborelowskie** są zbiory niemierzalne, a także pewne szczególne zbiory mierzalne, na przykład zbiór Cantora.

Aksjomaty Kołmogorowa

Rozważamy pewien zbiór Ω i odpowiadającą mu borelowską przestrzeń zdarzeń \mathcal{B} . Każdemu zdarzeniu $A \in \mathcal{B}$ przyporządkowujemy pewną liczbę $P(A)$, zwaną **prawdopodobieństwem zdarzenia A** . Przyporządkowanie to musi spełniać następujące aksjomaty:

- $\forall A \in \mathcal{B}: P(A) \geq 0$.
- $P(\Omega) = 1$.
- Niech $\{A_i\}$ stanowi dowolną przeliczalną rodzinę zdarzeń wzajemnie rozłącznych z \mathcal{B} , to znaczy $A_i \cap A_j = \emptyset$ dla $i \neq j$. Wówczas

$$P(\cup_i A_i) = \sum_i P(A_i).$$

Uwagi

1. Ponieważ $P(A) + P(\bar{A}) = P(\Omega) = 1$, to $P(A) \leq 1$.
2. Jeżeli $A_1 \subseteq A_2$, to $P(A_1) \leq P(A_2)$.
3. W ogólności

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

“Praktyczna” miara prawdopodobieństwa

“Szkolna” definicja prawdopodobieństwa powiada, że

$$P(A) = \frac{\text{liczba zdarzeń elementarnych sprzyjających zdarzeniu } A}{\text{liczba wszystkich zdarzeń elementarnych}} \quad (1)$$

Tę definicję *tylko czasami* daje się zastosować w praktyce. Przykład: prawdopodobieństwo wyrzucenia parzystej liczby oczek wynosi

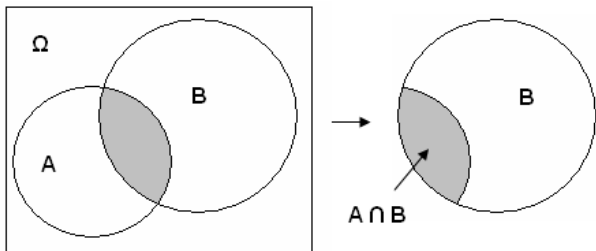
$$P(\text{parzysta}) = \frac{\#\{2, 4, 6\}}{\#\{1, 2, 3, 4, 5, 6\}} = \frac{3}{6} = \frac{1}{2}.$$

W praktyce, jako fizycy*, posługujemy się następującą metodą: Przeprowadzamy $N \gg 1$ powtórzeń jakiegoś eksperymentu. Zliczamy wystąpienia pewnego zdarzenia A . Niech zdarzenie to zachodzi w $n(A, N)$ przypadkach. Wówczas

$$\lim_{N \rightarrow \infty} \frac{n(A, N)}{N} = P(A). \quad (2)$$

*A raczej jako [frekwentyści](#).

Prawdopodobieństwo warunkowe



Przypuśćmy, że *wiemy*, iż zaszło zdarzenie A . Co, dysponując tą wiedzą, możemy powiedzieć o prawdopodobieństwie zdarzenia B ?

Prawdopodobieństwo warunkowe zdarzenia B , pod warunkiem, że zaszło zdarzenie A , wynosi:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (3)$$

Prawdopodobieństwo całkowite

Niech

$$\bigcup_{i=1}^N B_i = \Omega, \quad (4a)$$

przy czym $B_i \cap B_j = \emptyset$ dla $i \neq j$. Wówczas

$$P(A) = P(A \cap \Omega) = \sum_{i=1}^N P(A \cap B_i) = \sum_{i=1}^N P(A|B_i)P(B_i). \quad (4b)$$

Wzór (4b) nazywamy wzorem na prawdopodobieństwo całkowite.

Twierdzenie Bayesa

Łączne prawdopodobieństwo wystąpienia zdarzeń A, B

$$P(A \cap B) = P(A|B)P(B). \quad (5a)$$

Oczywiście

$$P(A \cap B) = P(B|A)P(A), \quad (5b)$$

a zatem

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (6)$$

Twierdzenie Bayesa pozwala obliczyć prawdopodobieństwo $P(B|A)$ jeśli znamy $P(A|B)$ oraz **prawdopodobieństwa a priori** $P(A), P(B)$. Prawdopodobieństwo $P(B|A)$ nazywa się *w tym kontekście* **prawdopodobieństwem a posteriori**.

Przykład 1

Firma produkuje pewne urządzenie w dwu fabrykach. Pierwsza dostarcza 60% całej produkcji, przy czym 35% wyrobów tej fabryki jest wadliwych. Druga dostarcza 40% całej produkcji, przy czym tylko 25% wyrobów drugiej fabryki jest wadliwych. Klient kupił losowe urządzenie i okazało się, że jest ono wadliwe. Jakie jest prawdopodobieństwo, że zostało ono wytworzone w pierwszej fabryce?

Niech A oznacza “urządzenie wyprodukowano w pierwszej fabryce”, B oznacza “urządzenie jest wadliwe”. Mamy

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} \\ &= \frac{0.6 \cdot 0.35}{0.6 \cdot 0.35 + 0.25 \cdot 0.4} \simeq 0.677 \end{aligned}$$

Przykład 2

Wiadomo, że na pewną chorobę choruje jedna osoba na 1000. Test wykrywający tę chorobę daje następujące wyniki:

- pozytywny z prawdopodobieństwem 95%, gdy badana osoba jest chora,
- negatywny z prawdopodobieństwem 5%, gdy badana osoba jest chora,
- pozytywny z prawdopodobieństwem 5%, gdy badana osoba jest zdrowa,
- negatywny z prawdopodobieństwem 95%, gdy badana osoba jest zdrowa.

Jakie jest prawdopodobieństwo, że osoba, u której test dał wynik pozytywny, rzeczywiście jest chora?

Oznaczmy C/Z fakt, że osoba jest chora/zdrowa oraz $+/-$, że test dał wynik pozytywny/negatywny. Z warunków zadania mamy

$P(+ C) = 0.95$	<i>true positive</i>
$P(+ Z) = 0.05$	<i>false positive</i>
$P(- C) = 0.05$	<i>false negative</i>
$P(- Z) = 0.95$	<i>true negative</i>
$P(C) = 0.001$	

Szukamy $P(C|+)$.

$$\begin{aligned}
 P(C|+) &= \frac{P(+|C)P(C)}{P(+)} \\
 &= \frac{P(+|C)P(C)}{P(+|C)P(C) + P(+|Z)P(Z)} \\
 &= \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.05 \cdot 0.999} \simeq 0.02.
 \end{aligned}$$

Niezależność statystyczna

Dwa zdarzenia A, B są *statystycznie niezależne*, jeżeli

$$P(A \cap B) = P(A) \cdot P(B). \quad (7)$$

W takim wypadku zachodzi

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \quad (8)$$

Ta obserwacja uzasadnia przyjętą nazwę tej kategorii zdarzeń.

Zmienne losowe

Dyskretna zmienna losowa to kolekcja wszystkich możliwych zdarzeń elementarnych wraz z ich prawdopodobieństwami. **Realizacja** dyskretnej zmiennej losowej daje jedno z możliwych zdarzeń elementarnych i czyni to z odpowiednim prawdopodobieństwem. Inna realizacja może dać inne zdarzenie losowe, ale gdybyśmy mieli odpowiednio dużo realizacji, częstości wystąpienia poszczególnych zdarzeń elementarnych **powinny** odpowiadać ich prawdopodobieństwom; patrz uwaga o interpretacji frekwencystycznej wyżej.

Ciągła zmienna losowa nie może zawierać prawdopodobieństw wszystkich możliwych zdarzeń elementarnych, a jedynie pewien **rozkład prawdopodobieństwa**. Jak to można zrobić?

Rozważmy rodzinę zbiorów borelowskich na \mathbb{R} . Niech $P_x(\lambda) = P(x \leq \lambda)$ oznacza *jakiś* sposób[†] przyporządkowania prawdopodobieństwa zdarzeniu “wylosowano liczbę nie większą od λ ”. $P_x(\lambda)$ nazywa się *dystrybuantą zmiennej losowej*.

W każdym możliwym przypadku musi zachodzić

$$\lambda \rightarrow +\infty : P_x(\lambda) \rightarrow 1 \quad (9a)$$

$$\lambda \rightarrow -\infty : P_x(\lambda) \rightarrow 0 \quad (9b)$$

Jeżeli $P_x(\lambda)$ jest różniczkowalne, to

$$\rho(\lambda) = \frac{dP_x(\lambda)}{d\lambda}. \quad (10)$$

[†]Sposobów tych jest nieskończenie wiele!

W tej sytuacji

$$P_x(\lambda) = \int_{-\infty}^{\lambda} \rho(x) dx, \quad (11)$$

przy czym *musi* zachodzić

$$\rho(x) \geq 0. \quad (12)$$

Funkcję $\rho(x)$ nazywamy **gęstością rozkładu prawdopodobieństwa**. Mówiąc niezbyt precyzyjnie, $\rho(x)dx$ jest prawdopodobieństwem wylosowania liczby z infinytezymalnie małego przedziału $(x, x + dx)$. Bardzo często ciągłe rozkłady zmiennej losowej określa się przez podanie ich gęstości.

Zachodzi

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} \rho(x) dx . \quad (13a)$$

W szczególności

$$\int_{-\infty}^{+\infty} \rho(x) dx = 1 . \quad (13b)$$

Wszystkie powyższe wyrażenia (np prawdopodobieństwo całkowite itd.) można łatwo przepisać w języku ciągłych rozkładów prawdopodobieństwa.

Przykład: Jednorodny rozkład prawdopodobieństwa

$$\rho(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases} \quad (14)$$

Jest to **ciągły** rozkład prawdopodobieństwa.

Przykład: Rozkład Gaussa

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (15)$$

Rozkład ten nazywany jest także *rozkładem normalnym*. Niekiedy jest on oznaczany $N(\mu, \sigma^2)$. Jest to **ciągły** rozkład prawdopodobieństwa.

Przykład: Rozkład dwumienny

Rozważmy zmienną losową, która może przyjmować tylko dwie wartości, x_1 z prawdopodobieństwem p oraz x_2 z prawdopodobieństwem $1-p$. Rozpatrujemy n realizacji tej zmiennej losowej. Jakie jest prawdopodobieństwo, że zdarzenie x_1 wystąpiło dokładnie k razy?

$$B(n, p; k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (16)$$

dla $0 \leq k \leq n$ i 0 poza tym. Rozkład ten nazywany jest także rozkładem Bernoullego. Jest to **dyskretny** rozkład prawdopodobieństwa.

Przykład: Rozkład Poissona

W rozkładzie dwumiennym dokonajmy przejścia granicznego

$$\begin{cases} p \rightarrow 0 \\ n \rightarrow \infty \\ pn = \lambda = \text{const} \end{cases} \quad (17)$$

Otrzymujemy

$$\rho(\lambda; k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots \quad (18)$$

Jest to rozkład **dyskretny**. k jest liczbą zdarzeń w przedziale $(0, T)$, jeżeli prawdopodobieństwo dokładnie jednego zdarzenia w czasie dt wynosi $\lambda dt/T$; bierzemy teraz $dt = T/n$, $p = \lambda/n$ i przechodzimy do granicy.

Klasyczne przykłady rozkładu Poissona to liczba rozpadów promieniotwórczych w jednostce czasu lub liczba rozmów telefonicznych zainicjowanych w jednostce czasu. Można pokazać, że rozkład Poissona pojawia się zawsze w sytuacji, w której mamy nieskończenie wiele “kontenerów”, przy czym do każdego z nich może wpaść całkowita liczba elementów i zawartość poszczególnych “kontenerów” jest niezależna od zawartości innych kontenerów.

Momenty zmiennej losowej

Niech x będzie zmienną losową w \mathbb{R} o gęstości $\rho(x)$ i niech $H(x)$ będzie jakąś funkcją tej zmiennej losowej. **Wartość oczekiwaną** tej funkcji definiujemy jako

$$\langle H(x) \rangle = \int_{-\infty}^{\infty} H(x)\rho(x)dx . \quad (19)$$

Uwaga: Całka w (19) nie musi istnieć! Jeśli całka w (19) nie istnieje, mówimy, że dana wielość H **nie posiada wartości średniej**.

Szczególnie ważnymi wielkościami są (jeśli istnieją)

wartość oczekiwana zmiennej losowej $\langle x \rangle = \int_{-\infty}^{\infty} x \rho(x) dx$

drugi moment zmiennej losowej $\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 \rho(x) dx$

wyższe momenty zmiennej losowej $\langle x^n \rangle = \int_{-\infty}^{\infty} x^n \rho(x) dx$

Wariancja zmiennej losowej, jeśli istnieje, zdefiniowana jest przez

$$\text{Var}(x) = \langle (x - \langle x \rangle)^2 \rangle \quad (20)$$

Funkcja charakterystyczna

Niech $\rho(x)$ będzie jakąś gęstością rozkładu prawdopodobieństwa. Definiujemy **funkcję charakterystyczną**:

$$G(k) = \langle e^{ikx} \rangle = \int_{-\infty}^{\infty} e^{ikx} \rho(x) dx \quad (21)$$

Funkcja charakterystyczna jest, z dokładnością do konwencji normalizacyjnej, równa transformacie Fouriera gęstości rozkładu prawdopodobieństwa.

Rozwinięcie kumulantów

Jeżeli wszystkie momenty rozkładu istnieją, funkcję charakterystyczną możemy przedstawić w postaci szeregu

$$G(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle \quad (22)$$

Uwaga: Funkcja charakterystyczna istnieje nawet gdy rozwinięcie (22) nie istnieje, to znaczy, gdy nie istnieją wyższe momenty.

Logarytm funkcji charakterystycznej możemy z kolei przedstawić jako

$$\ln G(k) = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n, \quad (23)$$

gdzie κ_n to kolejne *kumulanty*:

$$\kappa_1 = \langle x \rangle \quad (24a)$$

$$\kappa_2 = \text{Var}(x) = \langle x^2 \rangle - \langle x \rangle^2 \quad (24b)$$

$$\kappa_3 = \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \quad (24c)$$

itd.

Momenty rozkładu normalnego

Dla rozkładu normalnego (15)

- $\langle x \rangle = \kappa_1 = \mu$
- $\text{Var}(x) = \kappa_2 = \sigma^2$
- $G(k) = \exp\left(i\mu k - \frac{1}{2}\sigma^2 k^2\right)$ (transformata Fouriera zmiennej Gaussowskiej jest Gaussowska)
- wszystkie wyższe momenty nieparzyste znikają, wszystkie wyższe momenty parzyste mają postać $1 \cdot 3 \cdot \dots \cdot (n-1) \cdot \sigma^2$, wobec czego
- wszystkie wyższe kumulanty, począwszy od trzeciego, znikają.

Rozkład Cauchy'ego

Funkcja gęstości rozkładu Cauchy'ego

$$\rho(x) = \frac{1}{\pi} \cdot \frac{1}{x^2 + \gamma^2} \quad (25)$$

tak wolno zmierza do zera w $\pm\infty$, że wartość oczekiwana zmiennej losowej o takim rozkładzie istnieje tylko w sensie wartości głównej, natomiast wyższe momenty nie istnieją. Funkcja charakterystyczna ma postać

$$G(k) = e^{-\gamma|k|}, \quad (\gamma > 0) \quad (26)$$

Wielowymiarowe zmienne losowe

Pojęcie ciągłej zmiennej losowej można uogólnić na przypadek wielowymiarowy: Rozpatruje się d -wymiarowe wektory \mathbf{x} z przestrzeni \mathbb{R}^d . Podobnie jak w przypadku jednowymiarowym, wprowadza się wielowymiarową gęstość rozkładu prawdopodobieństwa $\rho(\mathbf{x}) = \rho(x_1, x_2, \dots, x_d)$. Na przypadek wielowymiarowy można też uogólnić pojęcia wartości oczekiwanej, momentów, funkcji charakterystycznej itd.

Przypuśćmy, że mamy dwuwymiarową zmienną losową o gęstości rozkładu $\rho(x, y)$. Można wówczas zdefiniować **brzegowy rozkład prawdopodobieństwa**

$$\rho_x(x) = \int_{-\infty}^{\infty} \rho(x, y) dy \quad (27)$$

Jest to efekt “rzutowania” rozkładu dwuwymiarowego na jedną ze zmiennych.

Przykład: Wielowymiarowy rozkład Gaussa

Niech $\mathbf{x}, \mathbf{a} \in \mathbb{R}^N$, $\mathbf{B} \in \mathbb{R}^{N \times N}$ będzie macierzą symetryczną i dodatnio określoną. \mathbf{a} jest ustalonym wektorem, \mathbf{x} jest N -wymiarową zmienną losową. Wówczas

$$\rho(\mathbf{x}) = \sqrt{\frac{\det \mathbf{B}}{(2\pi)^N}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{a})^T \mathbf{B}(\mathbf{x} - \mathbf{a})\right) \quad (28)$$

jest N -wymiarowym rozkładem Gaussa. Jego rozkłady brzegowe są jednowymiarowymi rozkładami Gaussa, a $\langle \mathbf{x} \rangle = \mathbf{a}$.

Zmiana zmiennych w rozkładach prawdopodobieństwa

Niech $x \in \mathbb{R}$ będzie jakąś zmienną losową, o rozkładzie $\rho_x(x)$. Niech $z = f(x)$ będzie jakąś funkcją zmiennej losowej x . z jest samo zmienną losową. Jaki jest jej rozkład prawdopodobieństwa? (Ściślej: Jaka jest gęstość rozkładu prawdopodobieństwa zmiennej losowej z ?)

$$\rho_z(z) = \int_{-\infty}^{\infty} \delta(z - f(x)) \rho_x(x) dx = \sum_i \frac{1}{|f'(x_i(z))|} \rho_x(x_i(z)) \quad (29)$$

gdzie $\delta(\cdot)$ oznacza deltę Diraca, zaś $x_i(z)$ oznaczają rozwiązania równania $z = f(x)$ ze względu na x ; sumujemy po wszystkich możliwych rozwiązaniach.

Przykład

Przypuśćmy, że zmienna losowa x ma rozkład jednostajny na przedziale $[0, 1]$. Niech $z = -\ln x$. $z \in [0, +\infty)$. Jaki jest rozkład zmiennej z ?

$$\rho_z(z) = \int_0^1 \delta(z + \ln x) \cdot 1 dx$$

Podstawiam $\ln x = -u$, $dx/x = -du$, $dx = -x du = -e^{-u} du$. $0 \rightarrow +\infty$, $1 \rightarrow 0$. Zatem

$$\rho_z(z) = \int_{+\infty}^0 \left(-e^{-u} \delta(z - u) \right) du = \int_0^{\infty} e^{-u} \delta(z - u) du = e^{-z}.$$

Centralne Twierdzenie Graniczne

Niech $\{X_i\}_{i=1}^{\infty}$ będą *niezależnymi* zmiennymi losowymi pochodzącymi z tego samego rozkładu (ang. Independent, Identically Distributed random variables, *i.i.d.*), przy czym rozkład ten posiada dwa pierwsze momenty[‡]. Wówczas rozkład zmiennej losowej będącej sumą zmiennych X_i dąży do rozkładu Gaussa.

Mówiąc bardziej precyzyjnie, rozkład zmiennej losowej

$$\sqrt{n} \left(\left(\frac{1}{n} \sum_{i=1}^n X_i \right) - \mu \right) \quad (30)$$

dąży do rozkładu normalnego $N(0, \sigma^2)$ gdy $n \rightarrow \infty$. μ oznacza średnią wyjściowego rozkładu a σ^2 jego wariancję.

[‡]To nie jest oczywiste założenie!

Uogólnienie centralnego twierdzenia granicznego

Jeśli jakiś rozkład prawdopodobieństwa ma tę własność, że suma dwóch zmiennych losowych o takim rozkładzie ma taki sam rozkład prawdopodobieństwa, co najwyżej przesunięty i przeskalowany, rozkład taki nazywamy *rozkładem stabilnym*.

Formalnie, niech X oraz $\{X_i\}_{i=1}^{\infty}$ będą i.i.d., pochodzącymi z pewnego rozkładu. Rozkład ten jest stabilny, jeśli dla każdego n istnieją takie stałe c_n, d_n , że zmienne

$$X_1 + X_2 + \cdots + X_n$$

oraz

$$c_n X + d_n$$

mają ten sam rozkład.

Istnieje nieskończenie wiele rozkładów stabilnych, przy czym tylko dla trzech z nich — rozkładu Gaussa, rozkładu Cauchy'ego i rozkładu Levy'ego-Smirnowa — można podać ich funkcje gęstości *explicite*. We *wszystkich* wypadkach znane są funkcje charakterystyczne rozkładów stabilnych. Spośród rozkładów stabilnych jedynie rozkład Gaussa ma dwa skończone momenty.

Uogólnienie centralnego twierdzenia granicznego głosi, że rozkład sumy i.i.d. zmiennych losowych, pochodzących z rozkładu, który nie ma skończonego drugiego lub pierwszego i drugiego momentu, jest zbieżny do jakiegoś rozkładu stabilnego.

Pojęcie entropii informacyjnej

Niech pewien układ fizyczny może znajdować się w jednym z N stanów $\{\omega_i\}_{i=1}^N$. Każdy z tych stanów może być przybrany z prawdopodobieństwem p_i . $\forall i: p_i \geq 0, \sum_i p_i = 1$.

Kiedy mamy *pełną* informację o układzie? Kiedy wiemy, że układ *z całą pewnością* znajduje się w konkretnym, znanym nam stanie ω_{i_0} : $p_{i_0} = 1, p_{i \neq i_0} = 0$.

Kiedy mamy *najmniejszą* informację o układzie? Gdy każdy stan układu jest obsadzony *z jednakowym prawdopodobieństwem*: $p_1 = p_2 = \dots = p_N (= 1/N)$.

Im *mniej* prawdopodobny jest stan, w którym znajduje się układ, tym *bardziej* uporządkowany nam się wydaje.

Jest to uogólnienie potocznej obserwacji, w której stany uznawane za uporządkowane są, w jakimś sensie, wyróżnione spośród wielu innych możliwych stanów.

Miara informacji

Przykład: w urnie jest sto ponumerowanych kul. Kule o numerach 1,2 są **czerwone**. Kule o numerach 3,4,...,100 są **niebieskie**. Ktoś wylosował jedną kulę i mówi nam

- **Wylosowałem kulę czerwoną** lub też
- **Wylosowałem kulę niebieską**

W którym przypadku dowiemy się więcej o szczegółowym (“mikroskopowym”) wyniku losowania?

Dowiemy się **więcej**, gdy uzyskany wynik będzie **mniej** prawdopodobny.

Miara informacji powinna być

1. Ciągłą funkcją prawdopodobieństwa zajścia zdarzenia,
2. Musi być tym większa, im prawdopodobieństwo jest mniejsze,
3. Jeżeli zdarzenie jest sumą dwu zdarzeń *niezależnych*, miara informacyjna zdarzenia łącznego musi być równa sumie miar zdarzeń składowych.

Funkcją, która spełnia te wymagania, jest

$$I(\omega_i) = \log_2 \left(\frac{1}{P(\omega_i)} \right) = -\log_2 P(\omega_i) \quad (= -\log_2 p_i)$$

Można uzasadnić, że logarytm (przy podstawie większej od 1) jest jedyną “cywilizowaną” funkcją spełniającą powyższe wymagania.

Entropia Gibbsa-Shannona

Claude Shannon, 1948: **Entropia informacyjna** jest **średnią miarą informacji** zawartej w danym rozkładzie.

$$S = - \sum_{i=1}^N p_i \log_2 p_i$$

Jeśli dopuścimy podstawy inne, niż 2, a także (potencjalnie) pozwolimy na wyrażanie entropii w jakichś jednostkach, otrzymamy **entropię Gibbsa-Shannona**:

$$S = -k \sum_{i=1}^N p_i \ln p_i \quad (31)$$

gdzie k jest (na razie niezidentyfikowaną) stałą dodatnią. W przypadku rozkładów ciągłych zamiast (31) mamy

$$S = -k \int_{-\infty}^{\infty} \rho(x) \ln \rho(x) dx . \quad (32)$$

Gdybyśmy mieli do czynienia z układem kwantowym znajdującym się w stanie opisanym operatorem gęstości ρ , odpowiednikiem entropii Gibbsa-Shannona byłoby

$$S = -k \text{Tr} \{ \rho \ln \rho \} . \quad (33)$$

Zwróćmy uwagę, że ponieważ operator ρ jest hermitowski i dodatnio określony, jego operatorowy logarytm jest dobrze zdefiniowany.

Rozkład maksymalizujący entropię

Rozpatrujemy skończone, dyskretne rozkłady prawdopodobieństwa. Czy można wskazać rozkład, którego entropia jest możliwie największa? Okazuje się, że tak. W tym celu należy znaleźć maksimum warunkowe entropii (31) przy warunku $\sum_i p_i = 1$. Maksymalizujemy zatem

$$\mathcal{L} = -k \sum_{i=1}^N p_i \ln p_i - \lambda \left(\sum_{i=1}^N p_i - 1 \right), \quad (34)$$

gdzie λ jest *mnożnikiem Lagrange'a*. Otrzymujemy

$$\frac{\partial \mathcal{L}}{\partial p_i} = -k \ln p_i - k - \lambda = 0 \quad (35a)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^N p_i - 1 = 0 \quad (35b)$$

Pierwsze z równań (35) musi być spełnione *dla wszystkich i* . Wynika z tego, że *wszystkie prawdopodobieństwa muszą być równe*. Zatem $\forall i: p_i = \alpha$. Z drugiego z równań (35) wynika wówczas, że $\forall i: p_i = 1/N$. Łatwo teraz obliczyć, że wartość entropii wynosi wówczas $k \ln N$.

Rekapitulując, rozkład jednostajny, w którym wszystkie stany są jednakowo prawdopodobne, maksymalizuje entropię Gibbsa-Shannona. Maksymalna wartość entropii wynosi

$$S = k \ln N, \quad (36)$$

gdzie N jest liczbą dostępnych stanów. Równość prawdopodobieństw wszystkich stanów odpowiada równości prawdopodobieństw *a priori* w zespole mikrokanonicznym, a maksymalna wartość entropii — entropii Boltzmana.

Rozkład maksymalizujący entropię przy dodatkowym warunku

Założmy, że każdemu stanowi i odpowiada pewna wielkość E_i . Żądamy, aby wartość oczekiwana tych wielkości była z góry ustalona:

$$\sum_{i=1}^N E_i p_i = \bar{E}. \quad (37)$$

Jaki rozkład prawdopodobieństwa maksymalizuje entropię przy takim warunku dodatkowym?

Znów użyjemy mnożników Lagrange'a, ale tym razem będą potrzebne dwa mnożniki:

$$\mathcal{L} = -k \sum_{i=1}^N p_i \ln p_i - \lambda_1 \left(\sum_{i=1}^N p_i - 1 \right) - \lambda_2 \left(\sum_{i=1}^N p_i E_i - \bar{E} \right). \quad (38)$$

Stąd

$$\frac{\partial \mathcal{L}}{\partial p_i} = -k \ln p_i - k - \lambda_1 - \lambda_2 E_i = 0, \quad (39a)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = -\sum_{i=1}^N p_i + 1 = 0, \quad (39b)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = -\sum_{i=1}^N E_i p_i + \bar{E} = 0. \quad (39c)$$

Z pierwszego z równań (39) otrzymujemy

$$\ln p_i = 1 - \frac{\lambda_1}{k} - \frac{\lambda_2}{k} E_i, \quad (40a)$$

$$p_i = \exp\left(1 - \frac{\lambda_1}{k}\right) \exp\left(-\frac{\lambda_2}{k} E_i\right) = C \exp\left(-\frac{\lambda_2}{k} E_i\right), \quad (40b)$$

gdzie C jest stałą w tym sensie, że nie zależy od numeru stanu, i . Drugie z równań (39) daje

$$\sum_{i=1}^N \exp\left(-\frac{\lambda_2}{k} E_i\right) = \frac{1}{C}, \quad (41)$$

a ostatnie

$$\sum_{i=1}^N E_i \exp\left(-\frac{\lambda_2}{k} E_i\right) = \frac{\bar{E}}{C}. \quad (42)$$

Różniczkując równanie (41) po λ_2 otrzymujemy

$$\sum_{i=1}^N \left(-\frac{E_i}{k}\right) \exp\left(-\frac{\lambda_2}{k} E_i\right) = -\frac{1}{C^2} \frac{dC}{d\lambda_2}, \quad (43)$$

a po skorzystaniu z (42)

$$\frac{\bar{E}}{kC} = \frac{1}{C^2} \frac{dC}{d\lambda_2}, \quad (44a)$$

$$\frac{\bar{E}}{k} C = \frac{dC}{d\lambda_2}. \quad (44b)$$

Rozwiązaniem równania różniczkowego (44b) jest

$$C = C_0 e^{\lambda_2 \bar{E}/k}. \quad (45)$$

Po skorzystaniu z (41)

$$C_0 = \frac{1}{e^{\lambda_2 \bar{E}/k} \sum_{i=1}^N e^{-\lambda_2 E_i/k}} \quad (46)$$

i ostatecznie, łącząc (40b), (45) i (46), otrzymujemy

$$\begin{aligned}
 p_i &= \frac{1}{\underbrace{e^{\lambda_2 \bar{E}/k} \sum_{i=1}^N e^{-\lambda_2 E_i/k}}_{C_0}} e^{\lambda_2 \bar{E}/k} e^{-\lambda_2 E_i/k} \\
 &= \frac{e^{-\lambda_2 E_i/k}}{\sum_{j=1}^N e^{-\lambda_2 E_j/k}}. \tag{47}
 \end{aligned}$$

Jeśli teraz *oznaczymy* ☺ $\lambda_2 = 1/T$,

$$p_i = Z^{-1} e^{-\frac{E_i}{kT}}, \tag{48a}$$

$$Z = \sum_{j=1}^N e^{-\frac{E_j}{kT}}. \tag{48b}$$

Maksymalna wartość entropii, odpowiadająca rozkładowi (48), wynosi

$$\begin{aligned}
 S &= -k \sum_{i=1}^N p_i \ln p_i = -k \sum_{i=1}^N p_i \left(-\frac{E_i}{kT} - \ln Z \right) \\
 &= \frac{1}{T} \sum_{i=1}^N E_i p_i + k \ln Z \sum_{i=1}^N p_i = \frac{1}{T} \bar{E} + k \ln Z \\
 &= \frac{1}{T} \bar{E} + k \left(-\frac{1}{kT} F \right) = \frac{1}{T} \bar{E} - \frac{1}{T} F, \tag{49}
 \end{aligned}$$

$$F = \bar{E} - TS, \tag{50}$$

gdzie *oznaczyliśmy* ☺ $Z = \exp(-F/kT)$. Również $\partial S/\partial \bar{E} = 1/T$. Jak widzimy, wyrażenia (48), (50) odpowiadają zespołowi kanonicznemu. **Przy narzuconym więzie na wartość oczekiwaną pewnej wielkości, entropia informacyjna Gibbsa-Shannona odpowiada zespołowi kanonicznemu.**