

Analiza danych

Mariusz Przybycień

Wydział Fizyki i Informatyki Stosowanej
Akademia Górniczo-Hutnicza

Wykład 2

Rozkład energii produktu rozpadu

- Przykład: Rozważmy rozpad $a \rightarrow b + c$ cząstki a o masie spoczynkowej m_a , energii E_a i pędzie \vec{p}_a w LAB na dwie cząstki o masach spoczynkowych m_b i m_c .
 - wielkości oznaczone przez \star to wielkości mierzone w układzie własnym cząstki a ,
 - niech θ, θ^\star oznaczają kąt emisji cząstki b względem pędu \vec{p}_a ,
 - oznaczenia: $p_b^\star = p_c^\star \equiv p^\star$, $\beta = p_a/E_a$, $\gamma = E_a/m_a$, $B = p_b^\star/E_b^\star$

$$p_b \cos \theta = \gamma(p^\star \cos \theta^\star + \beta E_b^\star) = \gamma E_b^\star (B \cos \theta^\star + \beta)$$

$$p_b \sin \theta = p^\star \sin \theta^\star$$

$$E_b = \gamma(E_b^\star + \beta p^\star \cos \theta) = \gamma E_b^\star (1 + \beta B \cos \theta^\star)$$

Dozwolony zakres energii cząstki b :

$$\left. \begin{array}{l} E_- = \gamma E_b^\star - \gamma \beta p^\star \\ E_+ = \gamma E_b^\star + \gamma \beta p^\star \end{array} \right\} \Rightarrow \Delta E \equiv E_+ - E_- = 2\gamma \beta p^\star = 2 \frac{p_a p^\star}{m_a}$$

Ponieważ: $f(x = \cos \theta^\star, \varphi) = \frac{1}{4\pi}$, $-1 \leq x \leq 1$, $0 \leq \varphi \leq 2\pi$

więc:

$$g(E_b, \varphi) = f(x(E_b), \varphi) \left| \frac{dx}{dE_b} \right| = \frac{1}{2\pi \Delta E}, \quad E_- \leq E_b \leq E_+, \quad 0 \leq \varphi \leq 2\pi$$

- Rozkład energii cząstki b otrzymujemy jako rozkład brzegowy:

$$g_1(E_b) = \int_0^{2\pi} g(E_b, \varphi) d\varphi = \frac{1}{\Delta E} \quad E_- \leq E_b \leq E_+$$

$$f(E_b) = g(E_b|E_a^{(1)}) P_1 + g(E_b|E_a^{(2)}) P_2 + g(E_b|E_a^{(3)}) P_3$$

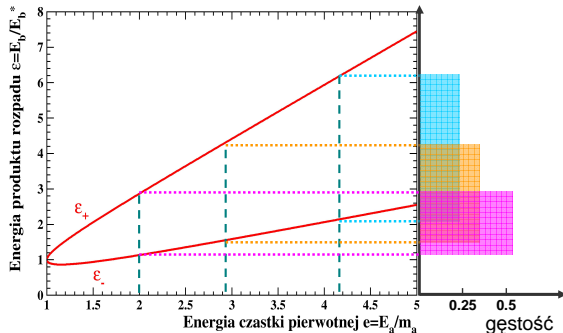
$$f(E_b) = \int_{E_a^-(E_b)}^{E_a^+(E_b)} g(E_b|E_a) h(E_a) dE_a = \frac{m_a}{2p^*} \int_{E_a^-(E_b)}^{E_a^+(E_b)} \frac{1}{E_a^2 - m_a^2} h(E_a) dE_a$$

$$\varepsilon_{\pm} \equiv \frac{E_{\pm}}{E_b^*} =$$

$$= \frac{E_a}{m_a} \pm B \sqrt{\left(\frac{E_a}{m_a}\right)^2 - 1}$$

- Ogólnie:

$$f(x) = \int_{-\infty}^{+\infty} g(x|y) h(y) dy$$



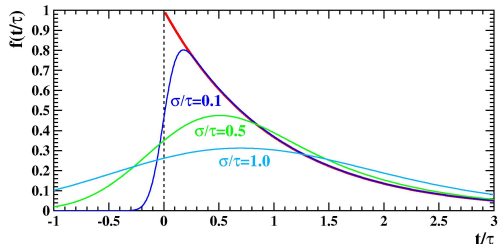
Splatanie rozkładu wykładniczego

Przykład: Mierzmy czas t oczekiwania na rozpad jądra promieniotwórczego za pomocą przyrządu którego zdolność rozdzielcza ma kształt gaussowski:

$$h(t) = \lambda \exp(-\lambda t) = \frac{1}{\tau} \exp\left(-\frac{t}{\tau}\right) \quad r(t'|t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t' - t)^2}{2\sigma^2}\right)$$

Splatamy teoretyczną funkcję rozkładu z funkcją zdolności rozdzielczej:

$$\begin{aligned} f(t') &= \int_0^{\infty} r(t'|t)h(t) dt = \frac{\lambda}{\sqrt{2\pi}\sigma} \int_0^{\infty} \exp\left(-\lambda t - \frac{(t' - t)^2}{2\sigma^2}\right) dt = \\ &= \frac{\lambda}{\sqrt{2\pi}} \exp\left(-\lambda t' + \frac{\lambda^2 \sigma^2}{2}\right) \int_{-\infty}^{t'/\sigma - \lambda\sigma} \exp\left(-\frac{y^2}{2}\right) dy \end{aligned}$$



- Splatanie prowadzi do wygładzania rozkładów, tym silniejszego im mniej precyzyjnie przeprowadzamy pomiar.
- Zdarza się, że w wyniku mało precyzyjnego pomiaru otrzymujemy wielkości mierzone w niefizycznym obszarze.

Relatywistyczny rozkład Breita-Wignera

- Relatywistyczny rozkład B-W opisuje p-two produkcji rezonansów (cząstek niestabilnych) w zależności od dostępnej energii E :

$$f(E) = \frac{k}{(E^2 - M^2)^2 + \Gamma^2 M^2}, \quad k = \frac{2\sqrt{2}M\Gamma\gamma}{\pi\sqrt{M^2 + \gamma}}, \quad \gamma = \sqrt{M^2(M^2 + \Gamma^2)}$$

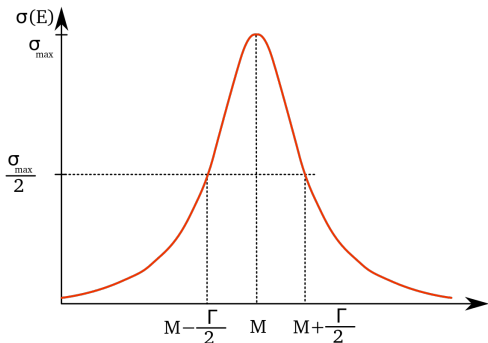
- Postać rozkładu wynika z propagatora cząstki niestabilnej, który w układzie spoczynkowym jest proporcjonalny do kwantowo-mechanicznej amplitudy opisującej rozpad rezonansu:

$$\Psi \propto \frac{\sqrt{k}}{(E^2 - M^2) + iM\Gamma}$$

- W przypadku reakcji



- $E = M_{cd}$ - energia w CMS,
- Γ - szerokość rozpadu związana ze średnim czasem życia $\tau = \hbar/\Gamma$.



Niech X_1, X_2, \dots będą zmiennymi losowymi o rozkładzie zadany dystrybuantą $F(x)$.

Definicja: Dla $k = 1, 2, \dots, n$ niech $X_{(k)}$ oznacza k -tą z kolei najmniejszą wartość spośród X_1, X_2, \dots, X_n . Wówczas niemalejący ciąg $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ nazywamy **statystyką porządku**, a $X_{(k)}$ k -tą zmienną w porządku.

Wniosek: Statystykę porządku otrzymuje się z nieuporządkowanej próbki poprzez permutację taką w wyniku której mamy: $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$

Uwaga: Statystyka porządku zależy także od n : $X_{(k)}$ jest k -tą najmniejszą wartością spośród n zmiennych X_1, X_2, \dots, X_n .

Czasem stosuje się bardziej precyzyjną notację $X_{1:n}, X_{2:n}, \dots, X_{n:n}$.

Wartości ekstremalne i ich rozkłady:

$$X_{(1)} = \min \{X_1, X_2, \dots, X_n\} \quad \text{oraz} \quad X_{(n)} = \max \{X_1, X_2, \dots, X_n\}$$

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, X_2 > x, \dots, X_n > x) = \\ &= 1 - \prod_{k=1}^n P(X_k > x) = 1 - (1 - F(x))^n \end{aligned}$$

$$F_{X_{(n)}}(x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = \prod_{k=1}^n P(X_k \leq x) = (F(x))^n$$

Dla rozkładów ciągłych znajdujemy funkcje gęstości prawdopodobieństwa:

$$f_{X_{(1)}}(x) = n(1 - F(x))^{n-1} f(x) \quad \text{oraz} \quad f_{X_{(n)}}(x) = n(F(x))^{n-1} f(x)$$

Przykład: W biegu na 100 metrów czasy uzyskane przez zawodników mają rozkład płaski na przedziale (9.6, 10) sekund. W finale uczestniczy 8 zawodników. Ile wynosi prawdopodobieństwo, że zwycięzca pobije rekord 9.69 s?

$$F(x) = 2.5x - 24 \Rightarrow F_{X_{(1)}}(x) = 1 - (25 - 2.5x)^8 \Rightarrow p = F_{X_{(1)}}(9.69) \approx 0.8699$$

Twierdzenie: Dla $k = 1, 2, \dots, n$ mamy:

$$F_{X_{(k)}}(x) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)} \int_0^{F(x)} y^{k-1} (1-y)^{n-k} dy$$

czyli

$$F_{X_{(k)}}(x) = F_{\beta(k, n+1-k)}(F(x)) \quad \text{gdzie} \quad \beta(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

Wniosek: W szczególności dla rozkładu płaskiego $u(0, 1)$ mamy

$$X_{(k)} \in \beta(k, n+1-k), \quad k = 1, 2, \dots, n$$

Przykład: Wylosowano 100 liczb z rozkładu płaskiego $u(0, 1)$. Znajdź prawdopodobieństwo, że druga z kolei najmniejsza liczba nie jest mniejsza niż 0.002.

$$F(x) = x \Rightarrow p = 1 - F_{X_{(2)}}(0.002) \approx 0.9826$$

Twierdzenie: Dla rozkładu ciągłego o funkcji gęstości $f(x)$, funkcja gęstości zmiennej $X_{(k)}$ dla $k = 1, 2, \dots, n$ dana jest przez pochodną $F_{X_{(k)}}(x)$:

$$\begin{aligned} f_{X_{(k)}}(x) &= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)} (F(x))^{k-1} (1-F(x))^{n-k} f(x) = \\ &= f_{\beta(k, n+1-k)}(F(x)) \cdot f(x) \end{aligned}$$

Dowód: Stosujemy regułę całkową Leibniza do $F_{X_{(k)}}(x)$:

$$\frac{d}{dx} \left(\int_{a(x)}^{b(x)} f(x, t) dt \right) = f(x, b(x)) \frac{db}{dx} - f(x, a(x)) \frac{da}{dx} + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x, t) dt$$

Przykład:

Łączny rozkład ekstremów

Twierdzenie: Łączna gęstość prawdopodobieństw zmiennych $X_{(1)}$ i $X_{(n)}$ dana jest przez:

$$f_{X_{(1)}, X_{(n)}}(x, y) = \begin{cases} n(n-1)(F(y) - F(x))^{n-2} f(y)f(x), & \text{dla } x < y \\ 0 & \text{dla pozostałych } (x, y) \end{cases}$$

Dowód: Korzystamy z niezależności zmiennych $X_{(k)}$ oraz rozłączności zdarzeń $(X_{(1)} \leq x)$ i $(X_{(1)} > x)$:

$$\begin{aligned} P(X_{(1)} > x, X_{(n)} \leq y) &= P(x < X_k \leq y, k = 1, 2, \dots, n) = \\ &= \prod_{k=1}^n P(x < X_k \leq y) = (F(y) - F(x))^n, \quad \text{dla } x < y \end{aligned}$$

$$P(X_{(1)} \leq x, X_{(n)} \leq y) + P(X_{(1)} > x, X_{(n)} \leq y) = P(X_{(n)} \leq y)$$

$$\begin{aligned} F_{X_{(1)}, X_{(n)}}(x, y) &= F_{X_{(n)}}(y) - P(X_{(1)} > x, X_{(n)} \leq y) = \\ &= \begin{cases} (F(y))^n - (F(y) - F(x))^n, & \text{dla } x < y \\ (F(y))^n, & \text{dla } x \geq y \end{cases} \end{aligned}$$

Aby otrzymać funkcję gęstości należy obliczyć pochodną $\frac{\partial^2}{\partial x \partial y} F_{X_{(1)}, X_{(n)}}(x, y)$.

Łączny rozkład ekstremów, zakres

Gęstość prawdopodobieństwa zmiennej losowej $R_n = X_{(n)} - X_{(1)}$ znajdujemy jako rozkład brzegowy dwuwymiarowej zmiennej losowej $(R_n, U = X_{(1)})$:

$$f_{R_n}(r) = n(n-1) \int_{-\infty}^{\infty} (F(u+r) - F(u))^{n-2} f(u+r)f(u)du, \quad \text{dla } r > 0$$

Przykład: Dla zmiennej losowej $X \in U(0,1)$ mamy:

$$f_{R_n}(r) = n(n-1) \int_0^{1-r} (u+r-u)^{n-2} \cdot 1 \cdot 1 \cdot du = n(n-1)r^{n-2}(1-r), \quad 0 < r < 1$$

Wartość oczekiwana: $\mathcal{E}[R_n] = \frac{n-1}{n+1}$.

Przykład: Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego $Exp(\lambda)$. Znajdź (a) $f_{X_{(1)}, X_{(n)}}(x, y)$, (b) $f_{R_n}(r)$.

$$\begin{aligned} f_{X_{(1)}, X_{(n)}}(x, y) &= n(n-1) (1 - e^{-\lambda y} - (1 - e^{-\lambda x}))^{n-2} \cdot \lambda e^{-\lambda y} \cdot \lambda e^{-\lambda x} = \\ &= n(n-1) (e^{-\lambda x} - e^{-\lambda y})^{n-2} \lambda^2 e^{-\lambda(x+y)}, \quad \text{dla } 0 < x < y \end{aligned}$$

$$\begin{aligned} f_{R_n}(r) &= n(n-1) \int_0^{\infty} (e^{-\lambda u} - e^{-\lambda(u+r)})^{n-2} \lambda^2 e^{-\lambda(2u+r)} du = \\ &= (n-1)\lambda (1 - e^{-\lambda r})^{n-2} e^{-\lambda r}, \quad \text{dla } r > 0 \end{aligned}$$

Warunkowe rozkłady statystyki porządku

Przykład cd: Dystrybuanta zakresu dla zmiennych z rozkładu wykładniczego:

$$F_{R_n}(r) = (1 - e^{-\lambda r})^{n-1} = (F(r))^{n-1}$$

Rezultat zgodny z "brakiem pamięci" rozkładu wykładniczego - $(X_{(2)} - X_{(1)}, X_{(3)} - X_{(1)}, \dots, X_{(n)} - X_{(1)})$ można interpretować jako statystykę porządku dla próbki $n - 1$ elementowej, gdzie $R_n = X_{(n)} - X_{(1)}$ jest jej wartością maksymalną.

Przykład: Niech X_1, X_2, X_3 będą niezależnymi zmiennymi losowymi z rozkładu wykładniczego $Exp(\lambda = 1)$. Znajdź $\mathcal{E} [X_{(3)} | X_{(1)} = x]$.

$$f_{X_{(1)}, X_{(3)}}(x, y) = 3 \cdot 2 (e^{-x} - e^{-y}) e^{-(x+y)}, \quad \text{dla } 0 < x < y$$

$$\begin{aligned} f_{X_{(3)} | X_{(1)} = x}(y) &= \frac{f_{X_{(1)}, X_{(3)}}(x, y)}{f_{X_{(1)}}(x)} = \frac{6 (e^{-x} - e^{-y}) e^{-(x+y)}}{3e^{-3x}} = \\ &= 2 (e^{-x} - e^{-y}) e^{2x-y} \quad \text{dla } 0 < x < y \end{aligned}$$

Warunkowa wartość oczekiwana jest więc równa:

$$\begin{aligned} \mathcal{E} [X_{(3)} | X_{(1)} = x] &= \int_x^\infty 2y (e^{-x} - e^{-y})^{n-2} e^{2x-y} dy = \{y = u + x\} = \\ &= \int_0^\infty 2(u + x) (e^{-x} - e^{-(u+x)}) e^{2x-u-x} du = x + \frac{3}{2} \end{aligned}$$

Łączny rozkład statystyki porządku

Twierdzenie: Łączna gęstość prawdopodobieństwazmiennych $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ dana jest przez:

$$f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(y_1, y_2, \dots, y_n) = \begin{cases} n! \prod_{k=1}^n f(y_k) & \text{dla } y_1 < y_2 < \dots < y_n \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

Przykład: Niech X_1, X_2, X_3 będą niezależnymi zmiennymi losowymi z rozkładu płaskiego $U(0, 1)$. Znajdź wszystkie rozkłady brzegowe statystyki porządku otrzymanej z tych zmiennych.

Łączny rozkład p-twa: $f_{X_{(1)}, X_{(2)}, X_{(3)}}(y_1, y_2, y_3) = 6$ dla $0 < y_1 < y_2 < y_3 < 1$

Rozkłady brzegowe:

$$f_{X_{(1)}, X_{(2)}}(y_1, y_2) = \int_{y_2}^1 6 dy_3 = 6(1 - y_2) \quad 0 < y_1 < y_2 < 1$$

$$f_{X_{(1)}, X_{(3)}}(y_1, y_3) = \int_{y_1}^{y_3} 6 dy_2 = 6(y_3 - y_1) \quad 0 < y_1 < y_3 < 1$$

$$f_{X_{(2)}, X_{(3)}}(y_2, y_3) = \int_0^{y_2} 6 dy_1 = 6y_2 \quad 0 < y_2 < y_3 < 1$$

$$f_{X_{(1)}}(y_1) = \int_{y_1}^1 6(1 - y_2) dy_2 = 3(1 - y_1)^2 \quad 0 < y_1 < 1$$

$$f_{X_{(2)}}(y_2) = \int_0^{y_2} 6(1 - y_2) dy_1 = 6y_2(1 - y_2) \quad 0 < y_2 < 1$$

$$f_{X_{(1)}}(y_1) = \int_{y_1}^1 6(y_3 - y_1) dy_3 = 3(1 - y_1)^2 \quad 0 < y_1 < 1$$

$$f_{X_{(3)}}(y_3) = \int_0^{y_3} 6(y_3 - y_1) dy_1 = 3y_3^2 \quad 0 < y_3 < 1$$