1 Analiza rozkładu

1.1 Miary położenia

1. Średnia arytmetyczna
$$\bar{x} = \frac{1}{N}\sum_i x_i = \frac{1}{N}\sum_i x_i n_i = \sum_i x_i w_i$$
gdzie n_i to liczba wystąpień x_i, a w_i - częstość wystąpień (waga).
2. Średnia geometryczna
$$\bar{x} = \sqrt[n]{x_1 x_2 ... x_n}^n$$
3. średnia harmoniczna
$$\bar{x} = \frac{1}{\frac{1}{N}\sum_i \frac{1}{x_i}}$$
4. moda (dominanta): wartość występująca najczęściej (o największej wartości funkcji gęstości prawdopodobieństwa)
5. mediana: wartość, od której połowa obserwacji jest mniejsza i połowa - większa v
6. kwartyle i kwantyle: kwantyl rzędu $\frac{i}{j}$ - taka liczba, że $\frac{i}{j}$ wyników jest mniejsza od tej liczby. Kwartyle 1 (q_1) i 3 (q_3) to odpowiednio kwantyle rzędu 1/4 i 3/4.

1.2 Miary zróżnicowania

1. wariancja (moment centralny rzędu 2):
$$\sigma = \frac{1}{N}\sum_i (x_i - \bar{x})^2 = \frac{1}{N}\sum_i n_i (x_i - \bar{x})^2$$
lub
$$\sigma = \sum_i w_i (x_i - \bar{x})^2$$
2. odchylenie standardowe średniej:
$$s = \sqrt{(\sigma)}$$
3. rozstęp: różnica między największą i najmniejszą wartością wyników
4. rozstęp ćwiartkowy:
$$IQR = q_3 - q_1$$
5. odchylenie ćwiartkowe:
$$o = \frac{IQR}{2}$$

1.3 Miary asymetrii i koncentracji

1. moment centralny rzędu 3:
$$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^3 n_i$$
2. współczynniki skośności:
   (a)
   $$A_d = \frac{\mu - d}{s}$$
   (b)
   $$A_m = 3\frac{\mu - d}{s} = 3A_d$$
   (c)
   $$A_Q = \frac{Q_1 + Q_3 - 2m}{2_Q}$$
3. Współczynnik asymetrii
$$A = \frac{M_3}{s^3}$$

1.4 Rozkłady

1. Gęstość $f(t)$ i dystrybuanta $F(x)$
   (a)
   $$f(t) \geq 0$$
   ,
   $$F(x) \geq 0$$
   ,
   (b) dystrybuanta
   $$F(x) = \int_{-\infty}^{x} f(t)dt$$
   , w $-\infty = 0$ a w $infty = 1$,
   (c)
   $$f(x) = F'(x)$$
   gdy pochodna istnieje i 0 w przeciwnym przypadku,
   (d)
   $$\int_{-\infty}^{\infty} f(t)dt = 1$$
   - normalizacja gęstości, prawdopodobieństwo że zajdzie jakiekolwiek zdarzenie jest równe 1,
   (e) prawdopodobieństwo, że X jest większe od a i mniejsze od b:
   $$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$$
   ,
2. funkcje zmiennej losowej ciągłej: jeśli funkcja $g(x)$ ma odwrotną funkcję $h(y)$ taką, że $h(g(X)) = X$ (np $g(x) = \log x$ ma odwrotną e^y, bo $e^{\log x} = x$, to gęstość zmiennej losowej $Y = g(X)$ wyraża się wzorem
$$f[h(y)]|h'(y)|$$
w przedziale $(g(a), g(b))$ (zakładamy, że $g(a) < g(b)$),
3. entropia
$$H(X) = -\sum_i p_i \log(p_i)$$
, przy czym w zależności od podstawy logarytmu są różne jednostki entropii. Dla podstawy 2 jednostką jest bit. Entropia jest równa zeru, gdy rozkład jest jednopunktowy (zmienna przyjmuje tylko jedną wartość) i przyjmuje maksymalną wartość $\log n$ gdy rozkład jest jednostanjny (każdy wynik jest jednakowo prawdopodobny),

1.5 zmienne dwuwymiarowe

1. gęstość $f(x,y)$ jest nieujemna,
2. dystrybuanta $F(x,y) = P(X < x, Y < y)$,
3. dystrybuanty rozkładów brzegowych $F_x(x,\infty)$ i $F_y(\infty, y)$,
4. dystrybuanta zmiennej skokowej
$$F(x,y) = \sum_i \sum_j p_{ij}$$
, gdzie dla każdego i,j zachodzi $x_i < x$, $y_j < y$,
5. dystrybuanta zmiennej ciągłej
$$(F(x,y) = \int_{-\infty}^{x}\left(\int_{-\infty}^{y} f(t,u)du\right)dt$$
6. całka z gęstości po całej przestrzeni = 1, tj
$$(F(\infty,\infty) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(t,u)du\right)dt = 1$$
,
7. $f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y)$ o ile taka pochodna istnieje lub 0,
8. $P(a < X < b, c < Y < d) = \int_a^b \left(\int_c^d f(x,y)dy\right)dx$,

1.6 rozkłady brzegowe i warunkowe

1. $P(X = x_i, Y = y_j) = p_{ij}$,
2. $P(X = x_i) = p_{i.} = \sum_j p_{ij}$,
3. $P(Y = y_i) = p_{.j} = \sum_i p_{ij}$,
4. $f_X(x) = \int_{-\infty}^{\infty}(x,y)dy$ - gęstość rozkł. brzeg. zm. X,
5. $f_X(x) = \int_{-\infty}^{\infty}(x,y)dx$ - gęstość rozkł. brzeg. zm. Y,
6. $P(X = x_i, Y = y_j) = p_{ij}$,
7. prawdopodobieństwo wystąpienia x_i jeżeli wystąpiło y_j wynosi:
$$P(X = x_i|_{Y=y_j}) = \frac{PX = x_i, Y = y_j}{P(Y = y_j} = \frac{p_{ij}}{p_{.j}}$$
i odpowiednio, symetrycznie w odwrotnym przypadku. Analogicznie dla rozkładów ciągłych:
$$f_X(x|_{y_0} = \frac{F(x, y_0}{f_Y(y_0)}$$
i symetrycznie - odwrotnie,

1.7 zmienne losowe niezależne

Jeżeli
$$F(x,y) = F_X(x)F_Y(y)$$
to zmienne X,Y nazywamy niezależnymi. Wystarczy nawet, że
$$p_{ij} = p_i.p_{.j}$$
dla każdej pary (x_i, y_j). Dla zmiennej ciągłej analogicznie $f(x,y) = f_X(x)f_Y(y)$ w każdym punkcie $f(x,y)$.