Prawdopodobieństwo warunkowe

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \tag{1}$$

$$P(E_1 \cap \dots \cap E_n) = \prod_{i=1}^{n} P(X_i|X_1, \dots, X_{i-1}) \tag{2}$$

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i) \text{ dla } \bigcap_{i=1}^{n} F_i = \Omega \tag{3}$$

	zmienne dyskretne	zmienne ciągłe
definicja	$P(x) = P(X = x)$	$f(x) = F'(x)$
obliczanie prawdopodobieństwa	$P(X \in A) = \sum_{x \in A} P(x)$	$P(X \in A) = \int_A f(x)dx$
skumulowana f. rozkładu	$F(x) = P(X \leqslant x) = \sum_{y \leqslant x} P(y)$	$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(y)dy$
całkowite prawdopodobieństwo	$\sum_x P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$
wartość oczekiwana	$EX = \sum_x xP(x)$	$EX = \int xf(x)dx$
wariancja	$VarX = \sigma^2 = E[(X - \mu)^2]$	$VarX = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

Nierówność Czebyszewa

$$P(|X - \mu| > \epsilon) \leqslant (\frac{\sigma}{\epsilon})^2 \tag{4}$$

Rozkłady dyskretne

Rozkład	P(x)	EX	VarX	
Bernoulli(p)	$P(x) = \begin{cases} p, & \text{for } x = 1 \\ q = (1-p), & \text{for } x = 0 \end{cases}$	$EX = p$	$VarX = pq$	próba
Binomial(n, p)	$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \dots$	$EX = np$	$VarX = npq$	liczba sukcesów z n prób
Geometric(p)	$P(x) = (1-p)^{x-1}p$ for $x = 1, 2, \dots$ $P(X > k) = (1-p)^k$	$EX = \frac{1}{p}$	$VarX = \frac{1-p}{p^2}$	liczba prób do sukcesu
Poiss(λ)	$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ for $x = 0, 1, \dots$	$EX = \lambda$	$VarX = \lambda$	rozkład zdarzeń rzadkich

Rozkłady ciągłe

Rozkład	f(x), F(x)	EX	VarX	
Unif(a,b)	$f(x) = \frac{1}{b-a}$ for $a \leqslant x \leqslant b$ $F(x) = \begin{cases} 0, & \text{for } x < a \\ \frac{x-a}{b-a}, & \text{for } a \leqslant x < b \\ 1, & \text{for } x \geqslant b \end{cases}$	$EX = \frac{a+b}{2}$	$VarX = \frac{(b-a)^2}{12}$	
Exp(λ)	$f(x) = \lambda e^{-\lambda x}$ for $x \geqslant 0$ $F(x) = 1 - e^{-\lambda x}$	$EX = \frac{1}{\lambda}$	$VarX = \frac{1}{\lambda^2}$	modelowanie czasu, brak pamięci
Gamma(α, λ)	$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ $F(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-\lambda t} dt$	$EX = \frac{\alpha}{\lambda}$	$VarX = \frac{\alpha}{\lambda^2}$	łączny czas α niezależnych zdarzeń $\sim Exp(\lambda)$
N(μ, σ)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ $F(x) = \Phi(x)$ dla N(0,1)	$EX = \mu$	$VarX = \sigma^2$	

$$Bin(n,p) \approx Poiss(\lambda) \tag{5}$$

$$P(T \leqslant t) = P(X \geqslant \alpha) \tag{6}$$
$$T \sim Gamma(\alpha, \lambda), X \sim Poiss(\lambda t)$$

$$Binomial(n,p) = N(np, \sqrt{np(1-p)}) \tag{7}$$
$$X_i \sim Bernoulli(p), S_n = \sum_{i=1}^{n} X_i, 0.05 \leqslant p \leqslant -0.95$$

Rozkład łączny

$$F_{(X,Y)}(x,y) = P(X \leqslant x \cap Y \leqslant y)$$
$$f_{(X,Y)}(x,y) = \frac{\delta^2}{\delta x \delta y} F_{(X,Y)}(x,y)$$

	zmienne dyskretne	zmienne ciągłe
rozkłady brzegowe	$P(x) = \sum_y P(x,y)$ $P(y) = \sum_x P(x,y)$	$f(x) = \int_Y f(x,y)dy$ $f(y) = \int_X f(x,y)dx$
niezależność	$P(x,y) = P(x)P(y)$	$f(x,y) = f(x)f(y)$
obliczanie prawdopodobieństwa	$P((X,Y) \in A) = \sum_{(x,y)\in A} P(x,y)$	$\int\int_{(x,y)\in A} f(x,y)dxdy$

Centralne Twierdzenie Graniczne

$$Z_n = \frac{S_n - E(S_n)}{Std(S_n)} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \to N(0,1) \text{ for } n \text{ to } \infty \tag{8}$$
$$S_n = \sum_{i=1}^{n} X_i, E(X_i) = \mu, Std(X_i) = \sigma$$

Estymatory - Monte Carlo

dla X, $p = P(X \in A)$

$$\hat{p} = \hat{P}(X \in A) = \frac{\#(X_i \in A)}{n}$$

$$E(\hat{p}) = \frac{1}{n}(np) = p$$

$$Std(\hat{p}) = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}$$

Dokładność

$$P(|\hat{p} - p| > \epsilon) = P\left(\frac{|\hat{p} - p|}{\sqrt{\frac{p(1-p)}{n}}} > \frac{\epsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) \approx 2\Phi\left(\frac{-\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

Estymacja średniej $\bar{X}$ z $X_1, \ldots, X_n$ ze wspólnym μ i σ

$$E\bar{X} = \frac{1}{n}(EX_1 + \cdots + EX_n) = \frac{1}{n}n\mu = \mu$$

$$Var\bar{X} = \frac{1}{n^2}(Var(X_1) + \cdots + Var(X_n)) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Estymator wariancji

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Procesy Markowa

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix}$$

Rozkład w czasie h: $P_h = P_0 * P^h$

Rozkład stacjonarny: $\pi P = \pi$, $\sum \pi_i = 1$

Estymacja: **metoda momentów**

k-ty moment z populacji	$\mu_k = E(X^k)$
k-ty centralny moment z populacji	$\mu'_k = E((X - \mu_1)^k)$
k-ty moment z próby	$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
k-ty centralny moment z próby	$m'_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$

$\mu_1 = EX, \mu'_2 = VarX, \mu_k = m_k$

Estymacja: **metoda największej wiarygodności**

$$P(X = (X_1, \ldots, X_n)) = P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i)$$

Szukamy ekstremum: $\frac{\delta P}{\delta \theta}(x) = 0$, używając logarytmu.

Przedziały ufności

$$\hat{\theta} \pm z_{\frac{\alpha}{2}}\sigma(\hat{\theta})$$

$$\bar{X} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - \bar{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

Rozkład t-studenta

$t = \frac{\hat{\theta} = \theta}{s(\hat{\theta})} \leftarrow$ zastępujemy $Std(\hat{\theta})$ przez $s(\hat{\theta})$, n-1 stopni swobody

$$\bar{X} \pm t_{\frac{\alpha}{2}}^{(n-1)} \frac{s(\hat{\theta})}{\sqrt{n}}$$

Z-testy

Hipoteza zerowa	Parametr, estymator	jeśli H_0 jest prawdziwa:		Statystyka
H_0	$\theta, \hat{\theta}$	$E(\hat{\theta})$	$Var(\hat{\theta})$	$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{Var(\hat{\theta})}}$
$\mu = \mu_0$	μ, $\bar{X}$	μ_0	$\frac{\sigma^2}{n}$	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
$p = p_0$	p, $\hat{p}$	p_0	$\frac{p_0(1-p_0)}{n}$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y$, $\bar{X} - \bar{Y}$	D	$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}$	$Z = \frac{\bar{X} + \bar{Y} - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2$, $\hat{p_1} - \hat{p_2}$	D	$\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$	$Z = \frac{\hat{p_1} - \hat{p_2} - D}{\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n} + \frac{\hat{p_2}(1-\hat{p_2})}{m}}}$
$p_1 = p_2$	$p_1 - p_2$, $\hat{p_1} - \hat{p_2}$	0	$p(1-p)(\frac{1}{n} + \frac{1}{m})$ gdzie $p = p_1 = p_2$	$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n} + \frac{1}{m})}}$ gdzie $\hat{p} = \frac{n\hat{p_1} + m\hat{p_2}}{n+m}$

T-testy

Hipoteza zerowa	Warunki	Statystyka	Stopnie swobody
$\mu = \mu_0$	Rozmiar próby n; nieznana σ	$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$	$n - 1$
$\mu_X - \mu_Y = D$	Rozmiary prób n, m nieznane, równe $\sigma_X = \sigma_Y$	$t = \frac{\bar{X} - \bar{Y} - D}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$	$n + m - 2$
$\mu_X - \mu_Y = D$	Rozmiary prób n, m; nieznane, różne $\sigma_X \neq \sigma_Y$	$t = \frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$	aproksymacja Satterthwaite

P-wartości - akceptacja H_0 jeśli $P \geqslant 0.1$, odrzucenie dla $P \leqslant 0.01$

Hipoteza zerowa	Hipoteza alternatywna	P-wartość	Wyliczenie
$\theta = \theta_0$	prawostronna $\theta > \theta_0$	$P\{Z \geqslant Z_{obs}\}$	$1 - \Phi(Z_{obs})$
	lewostronna $\theta < \theta_0$	$P\{Z \leqslant Z_{obs}\}$	$\Phi(Z_{obs})$
	dwustronna $\theta \neq \theta_0$	$P\{\lvert Z \rvert \geqslant \lvert Z_{obs} \rvert\}$	$2(1 - \Phi(\lvert Z_{obs} \rvert))$

Rozkład obserwacji o rozkładzie normalnym i wspólnej wariancji σ^2

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim Chi - square(n-1) \sim Gamma\left(\frac{n-1}{2}, \frac{1}{2} \right)$$

Zatem przedział ufności:

$$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \right]$$

Testy Chi kwadrat

H_0	H_A	Test statistic	Rejection region	P-value
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2_{obs} > \chi^2_{\alpha}$ $\chi^2_{obs} < \chi^2_{\alpha}$ $\chi^2_{obs} \geqslant \chi^2_{\frac{\alpha}{2}}$ or $\chi^2_{obs} \leqslant \chi^2_{\frac{\alpha}{2}}$	$P\chi^2 \geqslant \chi^2_{obs}$ $P\chi^2 \leqslant \chi^2_{obs}$ $2min(P\chi^2 \geqslant \chi^2_{obs}, P\chi^2 \leqslant \chi^2_{obs})$

Statystyka Chi-kwadrat

$$\chi^2 = \sum_{k=1}^{N} \frac{(Obs(k) - Exp(k))^2}{Exp(k)}, R = [\chi^2_{\alpha}, +\infty], P = P\chi^2 \geqslant \chi^2_{obs}$$

Rule of thumb: $Exp(k) \geqslant 5$ for all $k = 1, \ldots, N$.

Test Chi-kwadrat niezależności A i B

$$\chi^2_{obs} = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(Obs(i,j) - \hat{Exp}(i,j))^2}{\hat{Exp}(i,j)}, \hat{Exp}(i,j) = \frac{(n_{i.})(n_{.j})}{n}$$