Analiza Danych - Podstawy Statystyczne

Estymacja punktowa, przedziałowa i bayesowska

Marek Rupniewski

30 marca 2018

Model parametryczny

x_i to wynik i-tego pomiaru pewnej nieznanej wielkości x (pomiary obarczone są pewnym losowym błędem).

Jak oszacować wartość x?

Jak oszacować parametry błędu pomiaru (np. wariancję)?

Jak oszacować prawdopodobieństwo, że następny pomiar da wynik większy niż dotychczasowe?

itd.

Potrzebny model, np. $X_i \sim N(x, \sigma^2)$, X_i niezależne zmienne losowe, a x_i to ich realizacje, czyli $x_1, \dots, x_n$ to próba losowa z rozkładu $N(x, \sigma^2)$.

Model parametryczny

x_i to numer seryjny modułu, który trafia do serwisu ($i = 1, \ldots, n$).

Jak oszacować ilość modułów na rynku?

Jak oszacować ilość modułów, które trafią do serwisu w przeciągu kolejnego miesiąca?

itd.

Zaczynamy analizę od konstrukcji modelu, np. x_i to realizacje niezależnych zmiennych

$$X_i \sim \mathrm{Unif}\left([n_P, n_K]\right).$$

Model parametryczny

W modelu parametrycznym mamy do czynienia z rodziną rozkładów

$$\mathcal{F} = \{F_\theta(x) \colon \theta \in \Theta\}, \qquad \Theta \subset \mathbb{R}^k,$$

gdzie

$$\theta = (\theta_1, \dots, \theta_k)$$

to parametr (rozkładu).

Zamiast dystrybuant F_θ będziemy też rozważać funkcje prawdopodobieństwa lub funkcje gęstości prawdopodobieństwa f_θ.

Czasami będziemy stosowali oznaczenia typu:

$$F(x; \theta), f(x; \theta), \dots$$

Model parametryczny

Będziemy dalej zakładać, że mamy do dyspozycji próbę losową

$$X_1, \dots, X_n,$$

(ciąg niezależnych zmiennych o tym samym rozkładzie)
z pewnego rozkładu F_θ należącego do rodziny parametrycznej

$$\mathcal{F} = \{F_\theta(x) \colon \theta \in \Theta\}, \qquad \Theta \subset \mathbb{R}^k.$$

Wartość parametru $\theta \in \Theta$ nie jest znana!

Wszystkie zmienne $X_1, \dots, X_n$ odpowiadają tej samej wartości θ.

Model parametryczny

Typowymi zagadnieniami rozważanymi w kontekście modeli parametrycznych są:

- wyznaczenie, na podstawie obserwacji (próby): $X_1, \ldots, X_n$, nieznanego parametru θ, tzn. wyznaczenie takiej funkcji (estymatora)
$$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n),$$
która w jakimś sensie przybliża θ (teoria estymacji).
- sprawdzenie, na postawie obserwacji: $X_1, \ldots, X_n$, czy θ spełnia pewne warunki, np.
$$\theta = \theta_1, \qquad \theta > \theta_1, \qquad \theta \neq \theta_1$$
(testowanie hipotez statystycznych).

Model parametryczny — przykłady

$$\mathcal{F} = \{f_p(x) \colon p \in [0,1]\} \qquad (\theta = p).$$

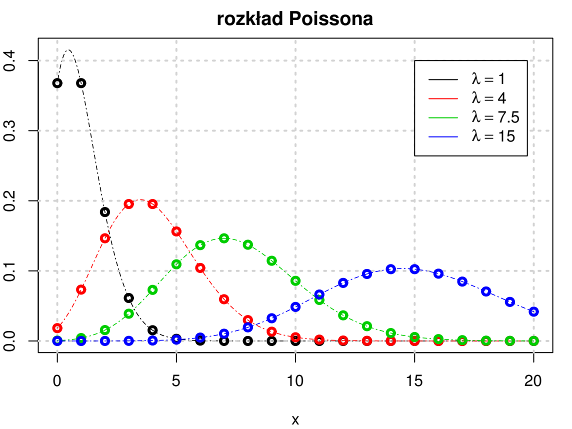
$$f_p(x) = \begin{cases} p & x = 1, \\ 1 - p & x = 0. \end{cases}$$

Próba losowa $X_1, \ldots, X_N$ z rozkładu z rodziny $\mathcal{F}$ może modelować N niezależnych rzutów tą samą monetą, o której nie wiemy jak bardzo jest niesymetryczna.

Model parametryczny — przykłady

$$\mathcal{F} = \{f_\lambda(x) \colon \lambda > 0\} \qquad (\theta = \lambda).$$

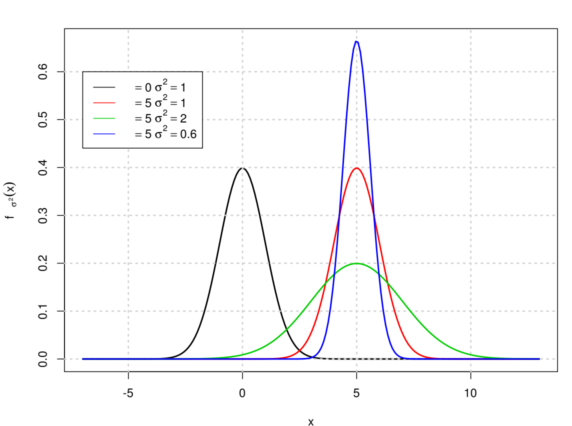
$$f_\lambda(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x \in \mathbb{N} = \{0, 1, 2, \dots\}.$$



Rodzina rozkładów normalnych

$$\mathcal{F} = \left\{ f_{(\mu,\sigma^2)}(x) \colon \mu \in \mathbb{R}, \sigma^2 > 0 \right\} \qquad \theta = (\mu, \sigma^2).$$

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$



Rodzina rozkładów gamma

$$f_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \qquad x > 0,\ \alpha, \beta > 0,$$

gdzie

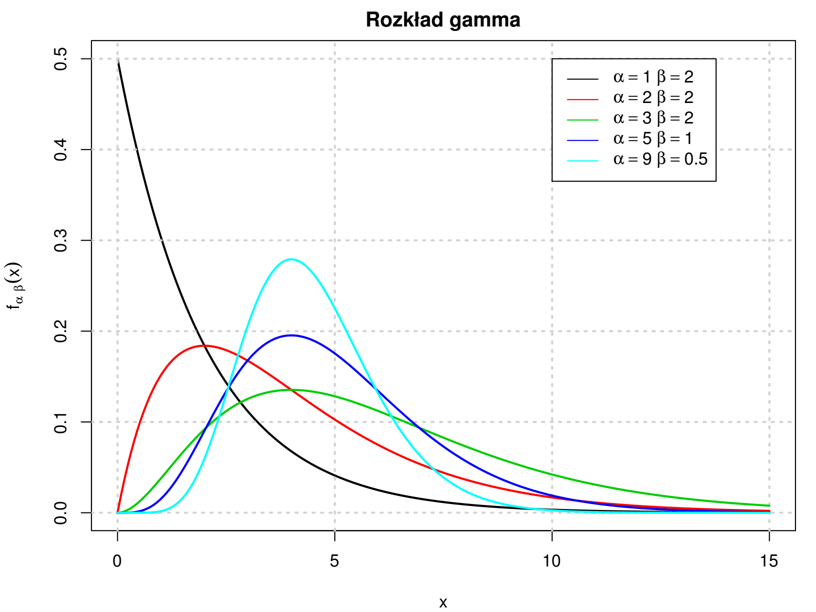
$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} \mathrm{d}u, \qquad \alpha > 0.$$

$$\Gamma(x+1) = x\Gamma(x) \quad x > 0, \quad \Gamma(1) = 1, \quad \Gamma(n+1) = n! \quad n \in \mathbb{N}.$$

Dla $\alpha = 1$ otrzymujemy rozkład Exp($\frac{1}{\beta}$).

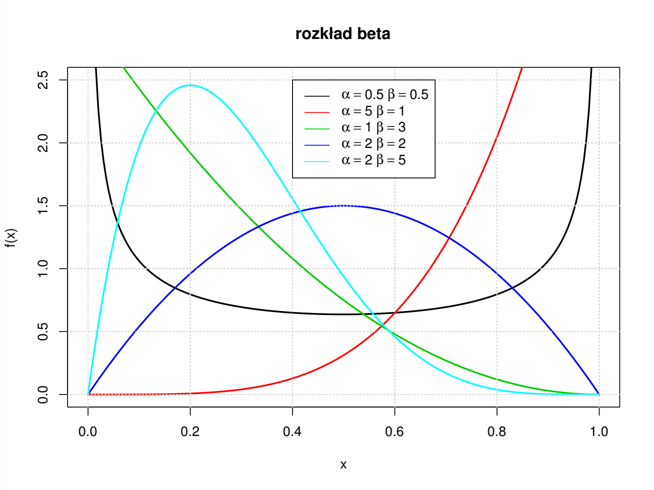
Uwaga! Czasami zamiast parametru skali β używa się parametru $\frac{1}{\beta}$.

Rozkład Gamma — przykłady



Rozkład beta

$$f_{\alpha,\beta}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad x \in [0,1],\ \alpha,\beta > 0$$



Model parametryczny

Powody mogą być dwojakie:

- wiemy skądinąd, że nasz model jest właściwy, lecz ta „zewnętrzna” wiedza nie daje nam informacji jakie jest θ,
- nie wiemy, że model jest właściwy, ale chcemy w jakiś zwarty sposób opisać pewną „rzeczywistość”.

Metoda momentów

Rozważmy rodzinę

$$\left\{N(\mu,\sigma^2)\colon \mu \in \mathbb{R},\ \sigma > 0\right\}, \qquad \theta = (\mu,\sigma^2).$$

oraz n-elementową próbę $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$.

Chcemy na podstawie tej próby wyznaczyć nieznane μ, σ.

Zgodnie z prawem wielkich liczb mamy

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}X_1 = \mu$$

oraz

$$\overline{(X^2)}_n = \frac{X_1^2 + \cdots + X_n^2}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1^2 = \mathbb{V}(X_1) + (\mathbb{E}X_1)^2 = \sigma^2 + \mu^2.$$

Za estymatory parametrów μ, σ możemy zatem przyjąć, odpowiednio,

$$\hat{\mu} = \overline{X}_n, \qquad \hat{\sigma} = \sqrt{\overline{(X^2)}_n - (\overline{X}_n)^2}.$$

Momenty

Definicja

Momentem k-tego rzędu zmiennej losowej X nazywamy liczbę

$$\mu_k = \mathbb{E}X^k.$$

Definicja

Momentem k-tego rzędu z próby $X_1, \ldots, X_n$ nazywamy zmienną losową

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k = \overline{(X^k)}_n.$$

Metoda momentów

Definicja

Estymator metody momentów dla parametru $\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$ i próby losowej $X_1, \ldots, X_n$, to taka statystyka (funkcja próby, a więc zmienna losowa!) $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$, która „podstawiona w miejsce parametru θ" daje wybrane momenty rozkładu równe odpowiednim momentom z próby (np. $\mu_1 = m_1$, $\mu_2 = m_2$ i $\mu_4 = m_4$).

Próba losowa $X_1, \ldots, X_n$ — estymator $\hat{\theta}(X_1, \ldots, X_n)$ (zmienna losowa).
Wartości próby $x_1, \ldots, x_n$ — wartość estymatora $\hat{\theta}(x_1, \ldots, x_n)$ (liczba, wektor liczbowy).

Metoda momentów

$$X_1, \dots, X_n \sim \text{Bern}(p).$$

$$\mu_1 = \mathbb{E}X = p.$$

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Zatem estymatorem m. m. parametru p jest

$$\hat{p} = m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n.$$

Metoda momentów

$$X_1, \ldots, X_n \sim \mathrm{Pois}(\lambda).$$

$$\mu_1 = \lambda.$$

Zatem estymatorem m. m. parametru λ jest średnia z próby:

$$\hat{\lambda} = m_1.$$

Metoda momentów — przykład (Bortkiewicz, 1898)

Analizowana była liczba zgonów od kopnięcia konia dla 10 korpusów pruskiej kawalerii w przeciągu 20-letniego okresu (mamy 200 „korpuso-lat”).

liczba zgonów/rok	liczba „korpuso-lat”
0	109
1	65
2	22
3	3
4	1

Próbujemy „dopasować” rozkład Pois(λ).

$$\hat{\lambda} = (109 \times 0 + 65 \times 1 + 22 \times 2 + 3 \times 3 + 1 \times 4)/200 = 0.61$$

Metoda momentów

zgonów/rok	„korpuso-lat”	częstość	$\mathbb{P}_{\hat{\lambda}}$	„k-l” wg $\hat{\lambda}$
0	109	0.545	0.543	109
1	65	0.325	0.331	66
2	22	0.110	0.101	20
3	3	0.015	0.021	4
4	1	0.005	0.003	1

Metoda momentów

$$X_1, \ldots, X_n \sim \text{Gamma}(\alpha, \beta).$$

$$\mu_1 = \alpha\beta, \qquad \mu_2 = \alpha(\alpha+1)\beta^2.$$

Wyznaczamy parametry w funkcji momentów:

$$\mu_2 = \mu_1^2 + \mu_1\beta \implies \beta = \frac{\mu_2 - \mu_1^2}{\mu_1}, \qquad \alpha = \frac{\mu_1}{\beta} = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.$$

Estymatorami m. m. parametrów α, β są

$$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}, \qquad \hat{\beta} = \frac{m_2 - m_1^2}{m_1}.$$

Zgodność estymatorów metody momentów

$X_1, \ldots, X_n$ próba losowa z rozkładu z parametrem $\theta = (\theta_1, \ldots, \theta_k)$.

$$(\mu_1, \ldots, \mu_k) = g(\theta_1, \ldots, \theta_k) = g(\theta) = (g_1(\theta), \ldots, g_k(\theta)) .$$

$$\hat{\theta} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) = g^{-1}(m_1, \ldots, m_k).$$

Zmieniając liczność próby $n = 1, 2, \ldots$ dostajemy ciąg estymatorów:

$$\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(n)}, \ldots$$

Twierdzenie

Estymator m. m. dla parametru θ wyznaczony j.w. jest zgodny, tzn.

$$\hat{\theta}_{(n)} \xrightarrow[n\to\infty]{\mathbb{P}} \theta.$$

Zgodność estymatora — przykład

$$\mathcal{F} = \{\mathrm{Exp}(\lambda) \colon \lambda > 0\}$$

$$h = \mathrm{id}, \quad m(\lambda) = 1/\lambda = \mu_1(\lambda), \quad \hat{\lambda}_n = 1/\overline{X}_n$$

np. $\lambda = 2$, $\forall_\epsilon \mathbb{P}(|\hat{\lambda}_n - 2| > \epsilon) \xrightarrow{n\to\infty} 0$, weźmy np. $\epsilon_0 = 0.01$

$$\mathbb{P}(|\hat{\lambda}_{10} - 2| > \epsilon_0) \approx 0.99,$$

$$\mathbb{P}(|\hat{\lambda}_{100} - 2| > \epsilon_0) \approx 0.96,$$

$$\mathbb{P}(|\hat{\lambda}_{1000} - 2| > \epsilon_0) \approx 0.87,$$

$$\mathbb{P}(|\hat{\lambda}_{10000} - 2| > \epsilon_0) \approx 0.62,$$

$$\mathbb{P}(|\hat{\lambda}_{100000} - 2| > \epsilon_0) \approx 0.11,$$

$$\mathbb{P}(|\hat{\lambda}_{200000} - 2| > \epsilon_0) \approx 0.03,$$

Asymptotyczna normalność estymatorów m. m.

$\theta = (\theta_1, \ldots, \theta_k)$, $\hat{\theta}_{(n)} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) = g^{-1}(m_1, \ldots, m_k)$.

Twierdzenie

Estymator m. m. dla parametru θ wyznaczony j.w. jest asymptotycznie normalny, tzn.

$$\sqrt{n}\left(\hat{\theta}_{(n)} - \theta\right) \xrightarrow[n\to\infty]{d} N(0, \Sigma),$$

W przypadku skalarnym mamy

$$\sqrt{n}\left(\hat{\theta}_{(n)} - \theta\right) \xrightarrow[n\to\infty]{d} N(0, \sigma^2),$$

$$\sigma^2 = \frac{g_2(\theta)}{(g_1'(\theta))^2} = \frac{\mu_2(\theta)}{\left(\frac{\partial\mu_1(\theta)}{\partial\theta}\right)^2}.$$

Asymptotyczna normalność — przykład

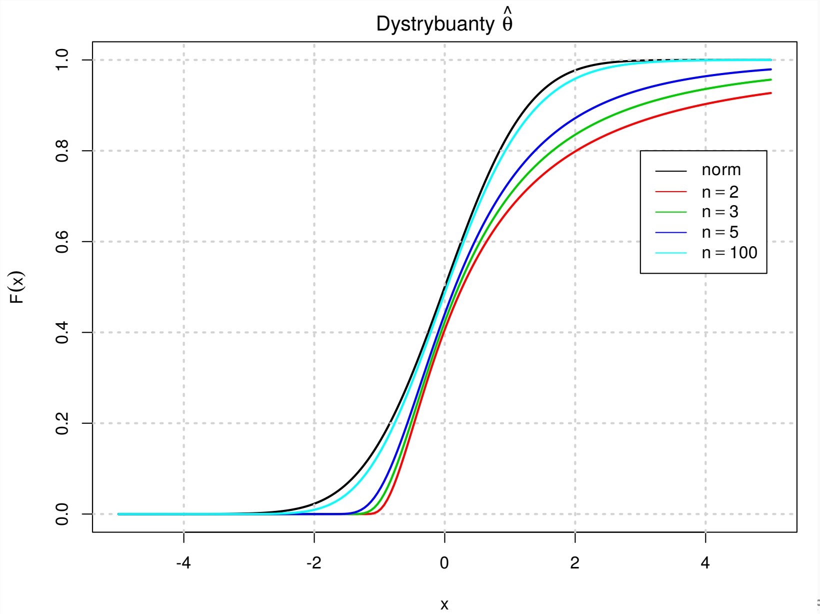
$$\mathcal{F} = \{\text{Exp}(\lambda)\colon \lambda > 0\}$$

$$h = \text{id}, \quad m(\lambda) = 1/\lambda = \mu_1(\lambda), \quad \hat{\lambda}_n = 1/\overline{X}_n$$

Asymptotyczna normalność estymatora $\hat{\lambda}_n$ oznacza punktową zbieżność dystrybuant F_n zmiennych $Y_n = \sqrt{n}\left(\hat{\lambda}_n - \lambda\right)$ do dystrybuanty F rozkładu $N(0, \lambda^2)$, t.j.

$$\forall t \in \mathbb{R} \qquad F_n(t) \xrightarrow{n\to\infty} F(t).$$

Asymptotyczna normalność — ilustracja przykładu



Estymatory największej wiarygodności

$$\mathcal{F} = \{f_\theta(x) : \theta \in \Theta\}, \qquad \Theta \subset \mathbb{R}^k,$$

$X_1, \ldots, X_n$ próba losowa (niezależne zmienne losowe o tym samym rozkładzie) z rozkładu odpowiadającemu pewnej wartości θ.

Definicja (Funkcja wiarygodności)

Funkcja wiarygodności $\mathcal{L}_n$ to funkcja określona formułą

$$\mathcal{L}_n(\theta) = \mathcal{L}_n(x_1, \ldots, x_n; \theta) = f_\theta(x_1) \times \cdots \times f_\theta(x_n).$$

Logarytmiczna funkcja wiarygodności ℓ_n to

$$\ell_n(\theta) = \ell_n(x_1, \ldots, x_n; \theta) = \ln(\mathcal{L}_n(\theta)).$$

$$\mathcal{F} = \{f_\theta(x) \colon \theta \in \Theta\}, \qquad \Theta \subset \mathbb{R}^k,$$

$X_1, \ldots, X_n$ próba losowa z rozkładu odpowiadającemu pewnej (nieznanej) wartości θ.

Definicja

Estymator największej wiarygodności $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ parametru θ to zmienna losowa maksymalizująca funkcję wiarygodności $\mathcal{L}_n$, czyli

$$\hat{\theta}_n = \arg\max_\theta \mathcal{L}_n(X_1, \ldots, X_n; \theta).$$

Równoważnie

$$\hat{\theta}_n = \arg\max_\theta \ell_n(X_1, \ldots, X_n; \theta).$$

Estymator n. w. dla rozkładu Bern(p)

$$\mathcal{F} = \{f_p(x) : p \in [0,1]\}, \qquad f_p(1) = p,\ f_p(0) = 1 - p.$$

Dla zwartości zapisu $\mathcal{L}_n(p)$ zamiast $\mathcal{L}_n(X_1, \ldots, X_n; p)$.

$$\begin{aligned}\mathcal{L}_n(p) = f_p(X_1) \times \cdots \times f_p(X_n) &= p^{\text{liczba } 1}(1-p)^{\text{liczba } 0} \\ &= p^{X_1 + \cdots + X_n}(1-p)^{n-(X_1+\cdots+X_n)}.\end{aligned}$$

$$\ell_n(p) = (X_1 + \cdots + X_n)\ln p + (n - (X_1 + \cdots + X_n))\ln(1-p).$$

$$\frac{\mathrm{d}\ell_n(p)}{\mathrm{d}p} = \frac{X_1 + \cdots + X_n}{p} - \frac{n - (X_1 + \cdots + X_n)}{1-p}.$$

$$\hat{p}_n = \frac{X_1 + \cdots + X_n}{n} = \overline{X}_n = m_1.$$

Estymator n. w. dla rozkładu $N(\mu, \sigma^2)$

$$\mathcal{F} = \left\{ f_{\mu,\sigma^2}(x) \colon (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty) \right\},$$

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Logarytmiczna funkcja wiarygodności

$$\ell_n(\mu, \sigma^2) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2.$$

$$\frac{\partial \ell_n(\mu, \sigma^2)}{\partial \mu} = 0 \Longrightarrow \hat{\mu} = \overline{X}_n.$$

$$\frac{\partial \ell_n(\mu, \sigma^2)}{\partial \sigma^2} = 0 \Longrightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

Estymator n. w. dla rozkładu Unif$([0,\theta])$

$$\mathcal{F} = \{f_\theta(x) \colon \theta > 0\}, \quad f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{dla } x \in [0,\theta] \\ 0 & \text{w p.p.} \end{cases}$$

Funkcja wiarygodności

$$\mathcal{L}_n(\theta) = \frac{1}{\theta^n} \mathbb{1}_{[0,\theta]}\Big(\max(X_1, \dots, X_n)\Big).$$

$$\hat{\theta}_n = \max(X_1, \dots, X_n) = X_{(n)}.$$

Metoda momentów daje w tym wypadku

$$\hat{\theta}_n = 2\overline{X}_n.$$

Estymator n. w. dla rozkładu Γ

$$\mathcal{F} = \left\{ f_{\alpha,\beta}(x) \colon (\alpha, \beta) \in (0,\infty) \times (0,\infty) \right\},$$

$$f_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta},$$

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} \mathrm{d}u, \qquad \alpha > 0.$$

Logarytmiczna funkcja wiarygodności:

$$\ell_n(\alpha,\beta) = (\alpha - 1) \sum_{k=1}^n \ln X_k - \sum_{k=1}^n X_k/\beta - n\alpha \ln \beta - n \ln \Gamma(\alpha).$$

$$\frac{\partial \ell_n}{\partial \beta} = 0 \Longrightarrow \hat{\beta} = \frac{\overline{X}_n}{\hat{\alpha}}.$$

$$\frac{\partial \ell_n}{\partial \alpha} = 0 \Longrightarrow \overline{(\ln X)}_n - \ln \overline{X}_n = \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} - \ln \hat{\alpha}.$$

Powyższe równanie rozwiązuje się numerycznie ...

Definicja

Obciążenie estymatora $\hat{\theta}_n$ parametru θ to wielkość

$$\mathbb{E}\hat{\theta} - \theta.$$

Estymator $\hat{\theta}_n$ parametru θ nazywany jest estymatorem nieobciążonym, jeśli jego obciążenie jest zerowe, czyli

$$\mathbb{E}\hat{\theta} = \theta.$$

Przykład: $\mathcal{F} = \{\text{Unif}([0,\theta]) \colon \theta > 0\}$
Estymator (parametru θ) m. m. jest nieobciążony:

$$\mathbb{E}\hat{\theta}_n = \mathbb{E}\left(2\overline{X}_n\right) = \frac{2}{n}\mathbb{E}(X_1 + \cdots + X_n) = \frac{2}{n}n\frac{\theta}{2} = \theta.$$

Estymator (parametru θ) n. w. jest obciążony:

$$\mathbb{E}\hat{\theta}_n = \mathbb{E}\left(\max(X_1, \ldots, X_n)\right) < \theta.$$

Własności estymatorów największej wiarygodności

Twierdzenie

Estymator n. w. dla parametru θ (przy pewnych założeniach co do regularności modelu) jest

- *zgodny, tzn.*

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}} \theta,$$

- *asymptotycznie normalny, tzn. (przypadek skalarnego parametru)*

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{d} N(0, \sigma^2_{MLE}),$$

- *ekwiwariantny, tzn. jeśli $\hat{\theta}_n$ jest estymatorem n. w. dla θ, to $g(\hat{\theta}_n)$ jest estymatorem n. w. dla $g(\theta)$.*
- *asymptotycznie optymalny (asymptotycznie efektywny)*

Funkcja informacji Fishera

$$\mathcal{F} = \{f_\theta(x) \colon \theta \in \Theta\}, \qquad \Theta \subset \mathbb{R}^k,$$

X_θ zmienna losowa o rozkładzie zadanym fun. gęst. $f_\theta(x)$.

Definicja (Funkcja informacji Fishera)

Funkcją informacji Fishera (informacją Fishera) dla rodziny $\mathcal{F}$ ze skalarnym ($k = 1$) parametrem θ nazywamy odwzorowanie

$$\Theta \ni \theta \mapsto I(\theta) = \mathbb{E}\left(\ell'(\theta)\right)^2 = \mathbb{E}\left(\frac{\partial \ln f_\theta(X_\theta)}{\partial \theta}\right)^2 = \mathbb{E}\left(\frac{f'_\theta(X_\theta)}{f_\theta(X_\theta)}\right)^2.$$

W przypadku wektorowego parametru θ ($k > 1$) $I(\theta)$ jest macierzą o i, j-tym elemencie określonym formułą

$$\mathbb{E}\left(\frac{\partial \ell(\theta)}{\partial \theta_i}\frac{\partial \ell(\theta)}{\partial \theta_j}\right).$$

$$I(\theta) = \mathbb{E}\left(\ell'(\theta)\right)^2.$$

Fakt

Przy pewnych założeniach co do regularności modelu

$$\mathbb{E}\left(\ell''(\theta)\right) = \mathbb{E}\left(\frac{\partial^2 \ln f_\theta(X_\theta)}{\partial \theta^2}\right) = -I(\theta).$$

Twierdzenie (Asymptotyczna normalność est. n. w.)

Jeśli $\hat{\theta}_n$ *jest estymatorem n. w., to*

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{I(\theta)}\right).$$

Równoważnie:

$$\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} N(0,1).$$

Co więcej

$$\sqrt{nI(\hat{\theta})}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} N(0,1).$$

Rozkład Bern(p)

Funkcja gęstości prawdopodobieństwa

$$f_p(x) = p\mathbb{1}_{\{1\}}(x) + (1-p)\mathbb{1}_{\{0\}}(x) = p^x(1-p)^{1-x}.$$

$$\ell'(p) = (x\ln p + (1-x)\ln(1-p))' = \frac{x}{p} - \frac{1-x}{1-p}.$$

$$\ell''(p) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}.$$

$$I(p) = -\mathbb{E}\ell''(p) = \frac{\mathbb{E}X}{p^2} + \frac{1-\mathbb{E}X}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.$$

Rozkład Exp(λ)

Funkcja gęstości prawdopodobieństwa

$$f_\lambda(x) = \lambda e^{-\lambda x} \quad \text{dla } x \geq 0.$$

$$\ell(\lambda) = \ln \lambda - \lambda x, \qquad \ell''(\lambda) = -\frac{1}{\lambda^2}.$$

$$I(\lambda) = -\mathbb{E}\left(-\frac{1}{\lambda^2}\right) = \frac{1}{\lambda^2}.$$

Rozkład Pois(λ)

Funkcja gęstości prawdopodobieństwa

$$f_\lambda(x) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda} \qquad x = 0,\ 1,\ \ldots$$

$$\ell(\lambda) = x\ln\lambda - \lambda - \ln x!, \qquad \ell''(\lambda) = -\frac{x}{\lambda^2}.$$

$$I(\lambda) = -\mathbb{E}\left(-\frac{X}{\lambda^2}\right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$

Nierówność Crámera-Rao

Niech $X_1, \ldots, X_n$ próba losowa (zmienne niezależne o tym samym rozkładzie odpowiadającym nieznanej wartości parametru θ), $S = S(X_1, \ldots, X_n)$ statystyka oraz

$$m(\theta) = \mathbb{E} S(X_1, \ldots, X_n).$$

Twierdzenie (Crámera-Rao)

$$\mathbb{V}(S) = \mathbb{E}(S - m(\theta))^2 \geq \frac{(m'(\theta))^2}{nI(\theta)},$$

przy czym równość wtw, gdy

$$S(X_1, \ldots, X_n) = m(\theta) + t(\theta)\ell_n'(X_1, \ldots, X_n; \theta),$$

dla pewnej funkcji $t(\theta)$.

Definicja

Niech $S = S(X_1, \ldots, X_n)$ będzie statystyką oraz $m(\theta) = \mathbb{E}S$. S nazywamy efektywnym estymatorem wielkości $m(\theta)$, jeśli

$$\mathbb{E}(S - m(\theta))^2 = \frac{(m'(\theta))^2}{nI(\theta)},$$

tzn. w nierówności Crámera-Rao zachodzi równość.

Estymator efektywny wielkości $m(\theta)$ nie musi istnieć!

Niech $\hat{\theta}$ będzie nieobciążonym ($\mathbb{E}\hat{\theta} = \theta$) estymatorem największej wiarygodności. Z nierówności Crámera-Rao wiadomo, że

$$\mathbb{V}(\hat{\theta}) \geq \frac{1}{nI(\theta)}, \text{ czyli } n\mathbb{V}(\hat{\theta}) \geq \frac{1}{I(\theta)}.$$

Wiemy również, że

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{I(\theta)}\right).$$

Stąd

$$\mathbb{V}(\sqrt{n}(\hat{\theta} - \theta)) = n\mathbb{V}(\hat{\theta}) \to \frac{1}{I(\theta)}.$$

Ta własność nazywa się asymptotyczną efektywnością estymatorów największej wiarygodności.

Przedziały ufności

Przedziały ufności

$X_1, \ldots, X_n$ próba losowa odp. pewnej (nieznanej) wartości parametru $\theta \in \Theta$.

Definicja

Losowy przedział

$$\Big[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)\Big]$$

nazywamy przedziałem ufności na poziomie ufności γ dla parametru θ, jeśli dla każdego $\theta \in \Theta$ zachodzi nierówność

$$\mathbb{P}\left(\theta \in \Big[A(X_1, \ldots, X_n), B(X_1, \ldots, X_n)\Big]\right) \geq \gamma.$$

Poziom ufności wyraża się często przez $1 - \alpha$ zamiast γ.

Przedziały ufności dla estymatora można wyznaczać:

- W sposób dokładny wykorzystując znajomość (jeśli się ją ma) analitycznej postaci rozkładu estymatora,
- W sposób przybliżony wykorzystując graniczny (normalny) rozkład estymatora,
- W sposób przybliżony przez symulacje (tzw. bootstrapping).

Wyznaczanie przedziału uf. z wykorzystaniem rozkładu granicznego

Ilość zgonów od kopnięcia konia dla 10 korpusów pruskiej kawalerii była zebrana z 20-letniego okresu (mamy 200 „korpuso-lat”).

Wyznaczanie przedziału uf. z wykorzystaniem rozkładu granicznego

ilość zgonów/rok	0	1	2	3	4
liczba „korpuso-lat”	109	65	22	3	1

$$\hat{\lambda} = (109 \times 0 + 65 \times 1 + 22 \times 2 + 3 \times 3 + 1 \times 4)/200 = 0.61$$

$$\sqrt{n}(\hat{\lambda} - \lambda) \approx N(0, 1/I(\hat{\lambda})) = N(0, \hat{\lambda}) = \sqrt{\hat{\lambda}} N(0,1)$$

Przedział ufności dla λ na poziomie ufności $\gamma = 1 - \alpha$ można zatem przybliżyć jako

$$\left[\hat{\lambda} - z_{(1+\gamma)/2}\sqrt{\hat{\lambda}/n},\ \hat{\lambda} + z_{(1+\gamma)/2}\sqrt{\hat{\lambda}/n}\right]$$
$$= \left[\hat{\lambda} - z_{1-\frac{\alpha}{2}}\sqrt{\hat{\lambda}/n},\ \hat{\lambda} + z_{1-\frac{\alpha}{2}}\sqrt{\hat{\lambda}/n}\right] \overset{\gamma=0.95}{\approx} [0.50,\ 0.72].$$

Bootstrap parametryczny

- Mamy do dyspozycji próbę, na podstawie której estymujemy parametr rozkładu $\hat{\theta}$. Chcemy zbadać jakie są własności (np. wariancja, kwantyle, rozkład) tego estymatora lub innej statystyki.
- Nie dysponując dodatkowymi danymi próbujemy wykorzystać to co mamy (bootstrapping).
- W przypadku modelu parametrycznego generujemy wiele losowych prób (liczności takiej jak oryginalna próba), dla każdej z nich wyznaczamy wartość badanej statystyki i badamy rozkład/wariancję/kwantyle/itp. tych wartości. Żeby wygenerować próby musimy znać parametr θ rozkładu!
  Zamiast θ wykorzystujemy (istota bootstrapu parametrycznego) estymatę $\hat{\theta}$ uzyskaną na podstawie oryginalnych danych.

Wyznaczanie przedziału uf. metodą bootstrapu parametrycznego

ilość zgonów/rok	0	1	2	3	4
liczba „korpuso-lat”	109	65	22	3	1

$$\hat{\lambda} = 0.61$$

Generujemy $N \gg 1$ zestawów po $n = 200$ liczb z rozkładu Pois($\hat{\lambda}$). Dla każdego zestawu wyliczamy wartość estymatora. Wyznaczamy przedział ufności $[A, B]$ (na poziomie ufności γ) tak, aby obejmował $\gamma \times N$ spośród wyznaczonych N wartości estymatora. Dla $\gamma = 0.95,\ N = 1000,\ \hat{\lambda} = 0.61$ w wyniku symulacji otrzymujemy

$$[A, B] \approx [0.45, 0.77].$$

Bootstrap nieparametryczny

- Mamy do dyspozycji próbę, na podstawie której wyznaczamy pewną statystkę T. Chcemy zbadać jakie są własności tej statystyki.
- W przypadku braku modelu parametrycznego generujemy wiele losowych prób (liczności takiej jak oryginalna próba) „z tego samego rozkładu co wyjściowa próba”. Dla każdej wygenerowanej próby wyznaczamy wartość statystyki T, a następnie badamy rozkład/wariancję/kwantyle/itp. tych wartości.
- Żeby wygenerować próby musimy znać rozkład!
  Za (nieznany!) rozkład przyjmujemy rozkład
  o dystrybuancie równej dystrybuancie (empirycznej)
  uzyskanej z wyjściowej próby!

Rozkład Normalny (Gaussa) — przypomnienie

X ma rozkład normalny (Gaussa) z parametrami μ, σ, $X \sim N(\mu, \sigma^2)$, jeśli

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Funkcję gęstości oraz dystrybuantę zmiennej o rozkładzie $N(0,1)$ oznacza się, odpowiednio, literami ϕ oraz Φ. Kwantyl rzędu α dla rozkładu $N(0,1)$ oznacza się zazwyczaj przez z_α.

Estymacja μ, znane σ^2

$$X_k = \mu + \epsilon_k, \qquad k = 1, \dots, n,$$

$\epsilon_1, \dots, \epsilon_n$ niezależne zmienne losowe $\sim N(0, \sigma^2)$.

X_k mogą modelować pomiary badanej wielkości μ, gdzie błąd pomiaru (np. związany z przyrządem) ma rozkład $N(0, \sigma^2)$.

Mamy do czynienia z rodziną (parametryczną):

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} : \mu \in \mathbb{R} \right\}, \quad \sigma^2 \text{ ustalone.}$$

$\overline{X}_n$ jest efektywnym estymatorem μ. Jego wariancja, to

$$\mathbb{V}(\overline{X}_n) = \frac{1}{nI(\mu)} = \frac{\sigma^2}{n}.$$

Przedziały ufności dla μ gdy σ^2 znane

$$\hat{\mu} = \overline{X}_n$$

Estymator $\hat{\mu} = \overline{X}_n$ ma rozkład

$$N\left(\mu, \frac{\sigma^2}{n}\right).$$

Przedziałem ufności dla μ na poziomie ufności $\gamma = 1 - \alpha$ jest zatem

$$\left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{(1+\gamma)/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{(1+\gamma)/2}\right]$$
$$= \left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right].$$

Przypadkiem gdy σ^2 nie jest znane zajmiemy się za chwilę…

Estymacja σ^2, znane μ

$$X_k = \mu + \epsilon_k, \qquad k = 1, \ldots, n, \quad \epsilon_1, \ldots, \epsilon_n \text{ niezależne } \sim N(0, \sigma^2).$$

X_k mogą modelować pomiary znanej wielkości (wzorca) μ , gdzie błąd pomiaru (np. związany z przyrządem) ma rozkład $N(0, \sigma^2)$.

Mamy do czynienia z rodziną (parametryczną):

$$\mathcal{F} = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} : \sigma^2 > 0 \right\}, \quad \mu \text{ znane.}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

jest efektywnym estymatorem dla σ^2 z wariancją

$$\mathbb{V}(\hat{\sigma}^2) = 1/(nI(\sigma^2)) = 2\sigma^4/n.$$

Rozkład χ^2

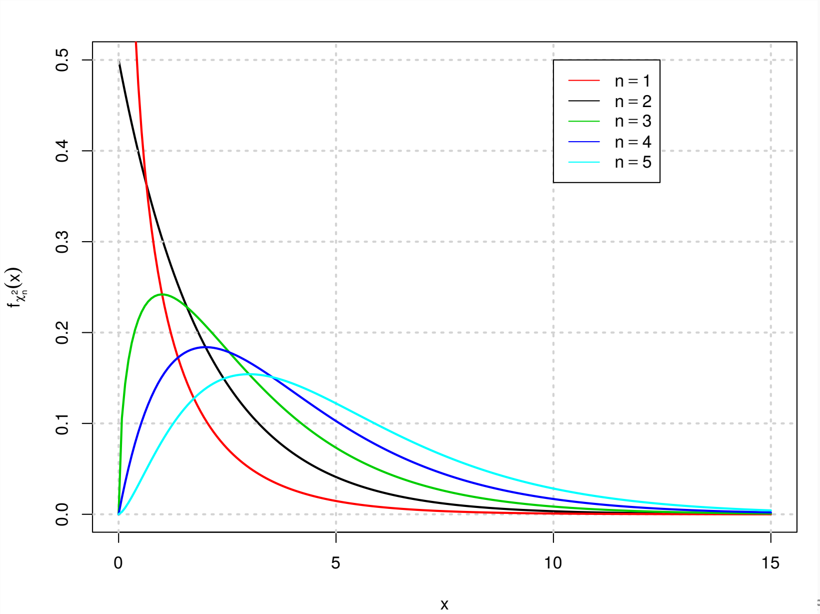
Definicja

Niech $X_1, \ldots, X_n$ niezależne zmienne losowe o rozkładzie $N(0,1)$. Rozkładam χ^2 o n stopniach swobody (rozkładem χ_n^2) nazywamy rozkład zmiennej losowej

$$X_1^2 + \cdots + X_n^2.$$

$$\chi_n^2 = \text{Gamma}\left(\frac{n}{2}, 2\right).$$

Rozkład χ^2 — wykresy funkcji gęstości



Przedziały ufności dla σ^2 (μ znane)

$X_k \sim N(\mu, \sigma^2)$ niezależne, μ znane.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

$$\frac{n\hat{\sigma}^2}{\sigma^2} \text{ ma rozkład } \chi^2_n.$$

Przedziałem ufności dla σ^2 na poziomie ufności γ jest zatem

$$\left[\frac{n\hat{\sigma^2}}{F^{-1}_{\chi^2_n}(1-b)}, \frac{n\hat{\sigma^2}}{F^{-1}_{\chi^2_n}(a)}\right], \quad a, b \geq 0,\ a + b = 1 - \gamma.$$

Można wziąć np. $a = b = (1-\gamma)/2$.

Estymacja σ^2, nieznane μ i σ^2

$$X_k = \mu + \epsilon_k, \qquad k = 1, \dots, n, \quad \epsilon_1, \dots, \epsilon_n \text{ niezależne } \sim N(0, \sigma^2).$$

X_k mogą modelować pomiary nieznanej wielkości (wzorca) μ, gdzie błąd pomiaru (np. związany z przyrządem) ma rozkład $N(0, \sigma^2)$ z nieznanych odchyleniem standardowym σ.

Mamy do czynienia z rodziną (parametryczną):

$$\mathcal{F} = \left\{ \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} / \sqrt{2\pi\sigma^2} : \mu \in \mathbb{R},\ \sigma^2 > 0 \right\}.$$

Estymator $\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2$ nie jest już estymatorem nieobciążonym. Nieobciążonym i asymptotycznie efektywnym estymatorem parametru σ^2 jest za to

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2.$$

Przedziały ufności dla μ oraz σ^2

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2, \qquad X_k \sim N(\mu, \sigma^2) \text{ niezależne.}$$

$\frac{(n-1)S^2}{\sigma^2} \sim$ Gamma $\left(\frac{n-1}{2}, 2\right) = \chi^2_{n-1}$.

Przedziałem ufności dla σ^2 na poziomie ufności γ jest zatem

$$\left[\frac{(n-1)S^2}{F^{-1}_{\chi^2_{n-1}}(1-b)}, \frac{(n-1)S^2}{F^{-1}_{\chi^2_{n-1}}(a)}\right], \quad a, b \geq 0,\ a + b = 1 - \gamma.$$

Można wziąć np. $a = b = (1-\gamma)/2$.

Przedziały ufności dla μ, gdy σ^2 nieznane

$$X_k \sim N(\mu, \sigma^2) \text{ niezależne.}$$

$$\hat{\mu} = \overline{X}_n \sim N(\mu, \frac{\sigma^2}{n}) \text{ czyli } \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\sigma^2}} \sim N(0,1).$$

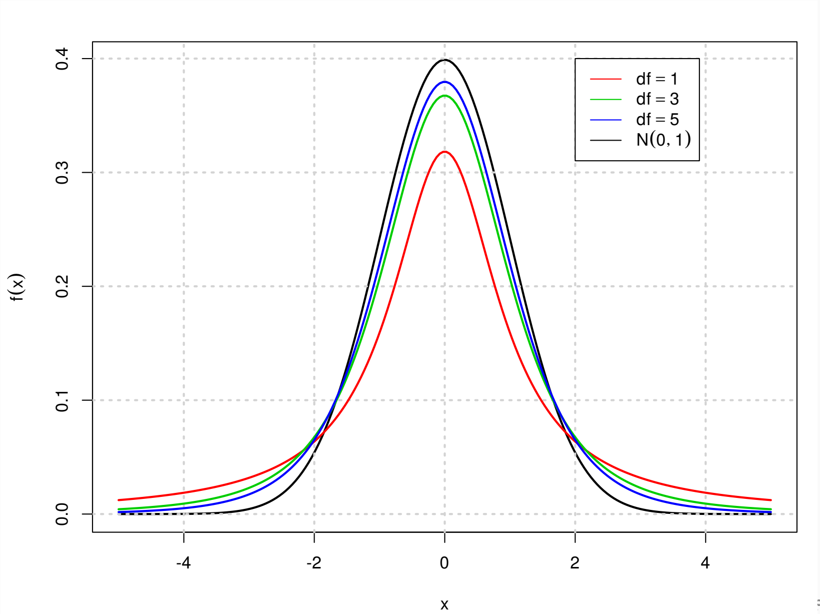
Nie znamy σ^2! Możemy za to rozważyć zmienną

$$t_{n-1} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{S^2}}.$$

Rozkład tej zmiennej nazywany jest rozkładem Studenta (t-Studenta) o n-1 stopniach swobody. Przedziałem ufności dla μ na poziomie ufności γ jest zatem

$$\left[\overline{X}_n - \frac{S}{\sqrt{n}} F^{-1}_{t_{n-1}}\left(\frac{1+\gamma}{2}\right), \overline{X}_n + \frac{S}{\sqrt{n}} F^{-1}_{t_{n-1}}\left(\frac{1+\gamma}{2}\right)\right].$$

Rozkład t-Studenta — wykresy funkcji gęstości



Estymacja w ujęciu bayesowskim

Czym właściwie jest prawdopodobieństwo

- prawdopodobieństwo obiektywne nazywane także prawdopodobieństwem w sensie częstościowym.
  Jeśli prawdopodobieństwo (częstościowe) wyrzucenia orła pewną monetą wynosi $\frac{1}{2}$, to
  $$\lim_{n\to\infty} \frac{\text{liczba orłów w pierwszych } n \text{ rzutach}}{n} = \frac{1}{2}.$$
- prawdopodobieństwo subiektywne nazywane także prawdopodobieństwem w sensie bayesowskim.
  W tym sensie prowadzący przedmiot może stwierdzić, że np. student A zaliczy przedmiot ADPS
  z prawdopodobieństwem 80%, a student B —
  z prawdopodobieństwem 99% Studenci A i B moga oceniać te prawdobodobieństwa inaczej. Co więcej prawdopodobieństwa te mogą się zmieniać wraz z czasem (napływ nowych informacji).

Prawdopodobieństwo bayesowskie

Prawdopodobieństwo bayesowskie spełnia wszystkie aksjomaty prawdopodobieństwa:

- $\mathbb{P}(A) \geq 0$ dla każdego zdarzenia A,
- $\mathbb{P}(\Omega) = 1$,
- Dla każdych rozłącznych zdarzeń $A_1, A_2, A_3, \ldots$

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Prawdopodobieństwo warunkowe

W podejściu bayesowskim kluczową rolę gra prawdopodobieństwo warunkowe:

$\mathbb{P}(A)$, to nasze przekonanie, że prawdziwe jest A.
W szczególnych przypadkach można interpretować $\mathbb{P}(A)$ jako prawdopodobieństwo a priori (przed wykonaniem eksperymentu/pomiaru).

$\mathbb{P}(A|B)$, to nasze przekonanie, że prawdziwe jest A pod warunkiem, że zachodzi B. Można je interpretować jako prawdopodobieństwo a posteriori (po wykonaniu eksperymentu/pomiaru, którego wynik opisywany jest przez B).

$$\mathbb{P}\left(A|B\right) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$

$$\mathbb{P}\left(B|A\right) = \mathbb{P}(B \cap A)/\mathbb{P}(A) = \mathbb{P}\left(A|B\right)\mathbb{P}(B)/\mathbb{P}(A).$$

Prawdopodobieństwo warunkowe a intuicja

Mamy do dyspozycji 3 monety:

1. monetę orzeł-reszka (O/R),
2. monetę orzeł-orzeł (O/O),
3. monetę reszka-reszka (R/R).

Losowo wybieramy monetę i rzucamy nią dostając orła. Jakie jest prawdopodobieństwo tego, że z drugiej strony jest też orzeł?

$$\frac{2}{3}$$

Prawdopodobieństwo warunkowe a intuicja

W pewnej zabawie mamy do wyboru 3 drzwi. Za jednymi z nich stoi samochód, który wygramy wybierając te drzwi. Za pozostałymi stoją kozy. Wybieramy jedne z drzwi (nie zaglądamy za nie), a następnie prowadzący zabawę otwiera jedne z pozostałych drzwi ukazując kozę. Następnie zadaje pytanie, czy zmieniamy nasz wybór. Zmienić wybór/ nie zmienić / wszystko jedno?

Zmienić wybór! (wygramy z prawd. $\frac{2}{3}$; gdy nie zmienimy mamy szanse $\frac{1}{3}$.).

Problemy z szacowaniem prawdopodobieństw warunkowych

Przed 100 lekarzami postawiono następujący problem dotyczący analizowania wyników mammograficznych badań przesiewowych:

- W przypadku braku pewnych dodatkowych informacji, prawdopodobieństwo, że kobieta (w odp. wieku i kondycji) ma raka piersi wynosi 1%.
- Jeśli pacjentka ma raka piersi, to prawdopodobieństwo, że radiolog na podstawie badania postawi właściwą diagnozę wynosi 80%.
- Jeśli pacjentka ma zmiany nienowotworowe, to prawdopodobieństwo, że radiolog zdiagnozuje raka wynosi 10%.

Jakie jest prawdopodobieństwo, że pacjentka z pozytywnym wynikiem mammografii ma istotnie raka piersi?

Twierdzenie (Reguła Bayesa)

Jeśli A *jest pewnym zdarzeniem oraz rozłączne zdarzenia* $B_1, \ldots, B_n$ *o niezerowym prawdopodobieństwie pokrywają całą przestrzeń probabilistyczną* Ω *(*$\bigcup_{i=1}^n B_i = \Omega$*), to*

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

wersja „ciągła":

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x) \mathrm{d}x}.$$

Parametryczny model:

$$\mathcal{F} = \{f_p(x) : p \in \Theta\}, \qquad \Theta \subset \mathbb{R}^k,$$

z parametrem $p = (p_1, \ldots, p_k)$

W podejściu bayesowskim zakładamy, że parametr p jest zmienną losową! (w związku z tym będziemy go czasem oznaczać dużą literą P) o pewnym rozkładzie zadanym np. funkcją prawdopodobieństwa:

- a priori

  $$f_{\text{prior}} = f_P : \Theta \to \mathbb{R}, \qquad p \mapsto f(p)$$

  wynikającą z naszej wiedzy i doświadczenia,
- a posteriori

  $$f_{\text{post}} = f_{P|X} : \Theta \times \mathbb{R}^n \to \mathbb{R}, \quad (p, x_1, \ldots, x_n) \mapsto f_{\text{post}}(p, x_1, \ldots, x_n)$$

  wynikającą z rozkładu a priori oraz zaobserwowanej próby

  $$X = (X_1, \ldots, X_n).$$

Konstrukcja rozkładu a posteriori

Model parametryczny $\mathcal{F} = \{f_p(x) : p \in \Theta\}$.

Zgodnie z regułą Bayesa

$$f_{\text{post}}(p,x) = f_{P|X}(p,x) = \frac{f_{X|P}(x,p) f_P(p)}{\text{"} f_X(x) \text{"}} = \frac{f_p(x) f_P(p)}{\int_\Theta f_p(x) f_P(p) \mathrm{d}p} = \frac{f_p(x) f_{\text{prior}}(p)}{\int_\Theta f_p(x) f_{\text{prior}}(p) \mathrm{d}p},$$

gdzie

$$f_p(x) = f_p(x_1) f_p(x_2) \dots f_p(x_n) = \mathcal{L}_n(x_1, \dots, x_n; p).$$

gęstość a posteriori $\propto$ fun. wiarygodności $\times$ gęst. a priori.

Przykład bayesowskiego podejścia do problemu

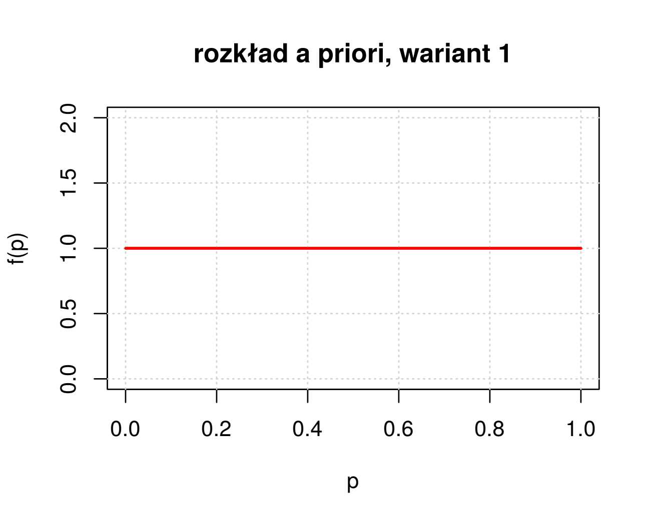
Spotykamy znajomego, który proponuje nam zakład oparty na rzutach monetą.

Prawdopodobieństwo p wypadnięcia orła to parametr modelu.

Chcemy oszacować p przed zakładem.

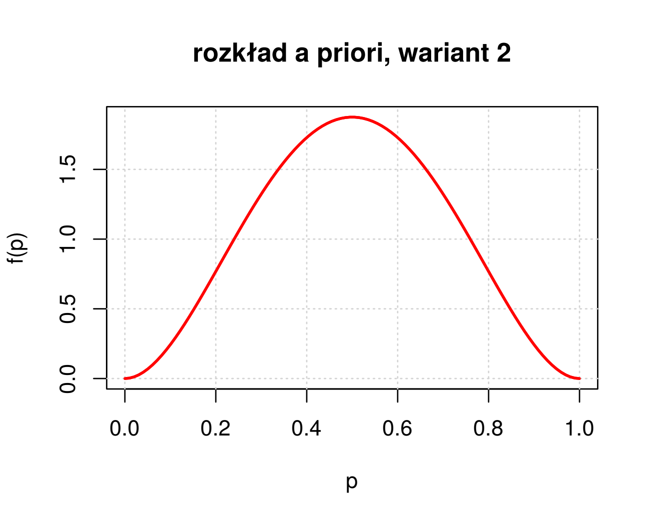
W zależności od sytuacji możemy zakładać różne rozkłady a priori.

Jeśli chcemy być „obiektywni” możemy założyć rozkład jednostajny.



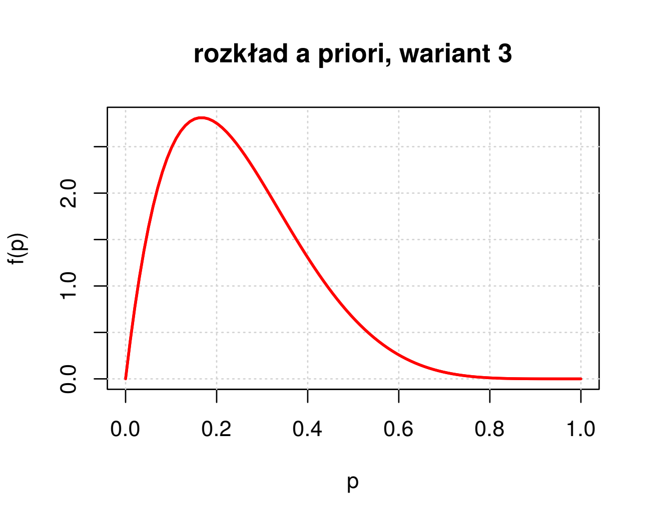
Rozkład a priori — wariant 2

Wiedząc, że monety są zazwyczaj symetryczne można zakładać rozkład skupiony wokół wartości $p = \frac{1}{2}$.



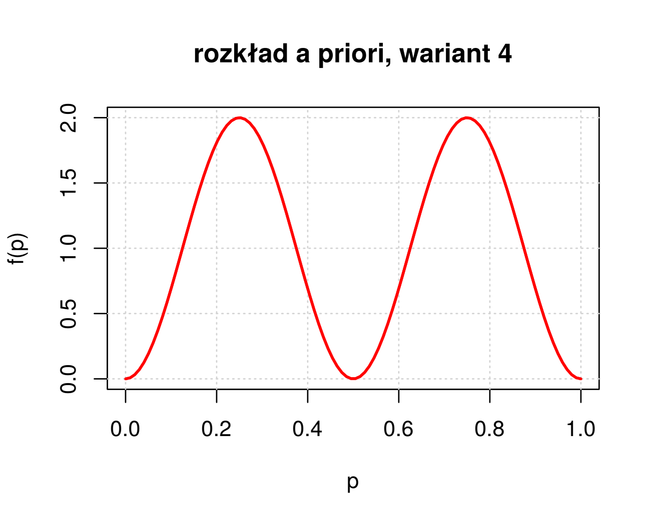
Rozkład a priori — wariant 3

Znając monetę (np. będąc świadkiem innych zakładów z wykorzystaniem tej samej monety), można zakładać rozkład:



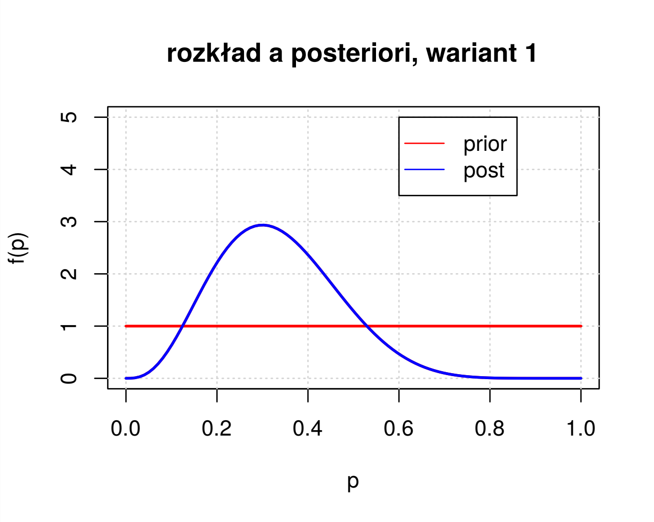
Rozkład a priori — wariant 4

Znając „przewrotność” właściciela monety można też zakładać rozkład bimodalny.

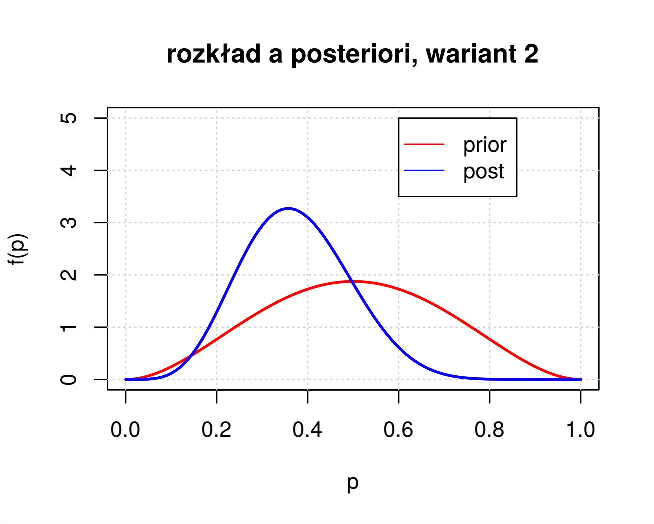


Rzucamy „próbnie” 10 razy monetą i notujemy 3 orły.

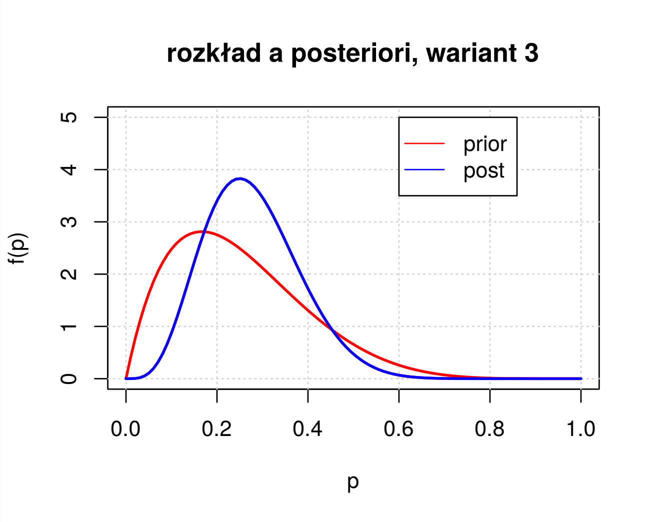
Rozkład a posteriori — wariant 1



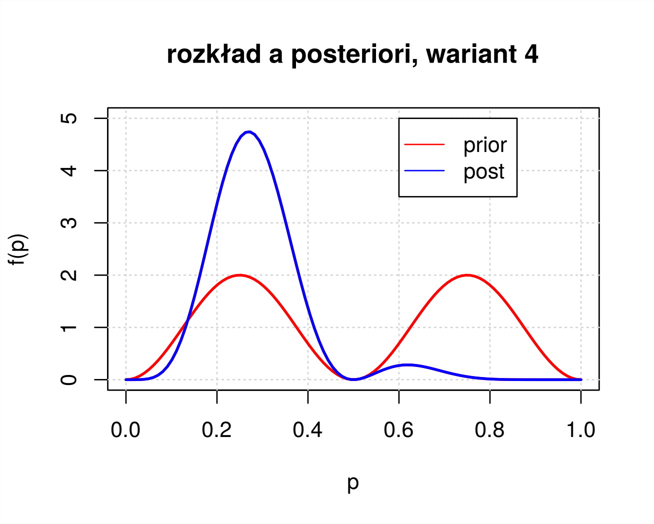
Rozkład a posteriori — wariant 2



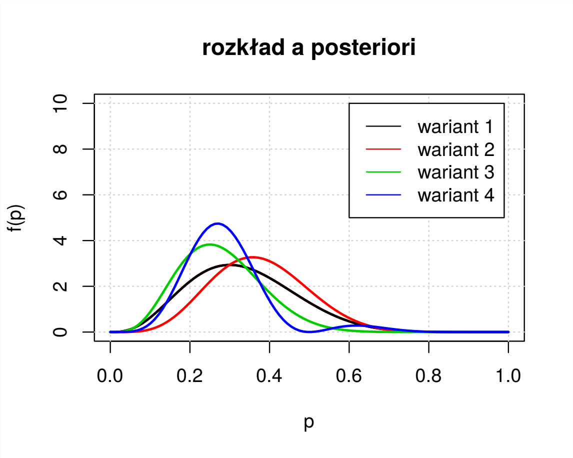
Rozkład a posteriori — wariant 3



Rozkład a posteriori — wariant 4

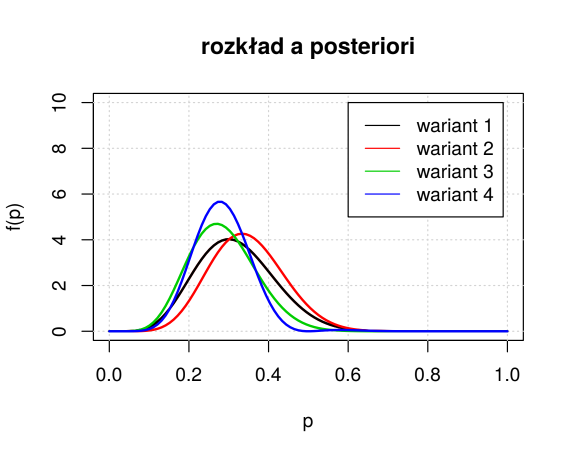


Rozkład a posteriori — podsumowanie



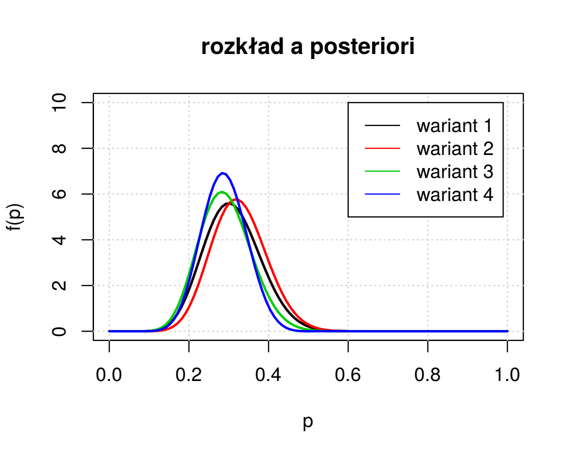
Rozkład a posteriori — podsumowanie (c.d.)

A gdybyśmy rzucili 20 razy uzyskując 6 orłów.



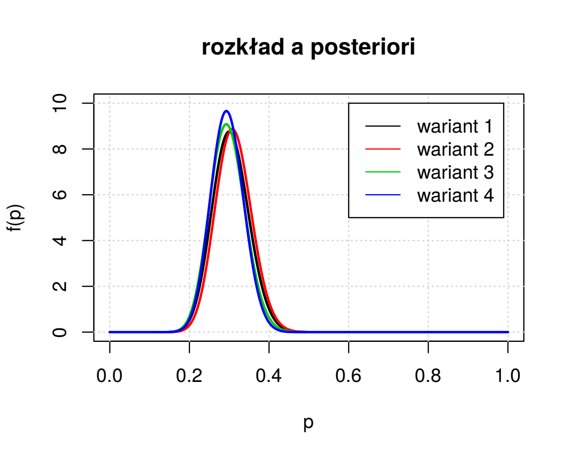
Rozkład a posteriori — podsumowanie (c.d.)

A gdybyśmy rzucili 40 razy uzyskując 12 orłów.



Rozkład a posteriori — podsumowanie (c.d.)

A gdybyśmy rzucili 100 razy uzyskując 30 orłów.



Model parametryczny $\mathcal{F} = \{f_p(x) : p \in \Theta\}$.

Mając rozkład a priori parametru p (funkcja gęstości $f_{\text{prior}}(p)$), parametryczny model oraz próbę $X = (X_1, \ldots, X_n)$ możemy wyznaczyć rozkład a posteriori parametru (funkcję gęstości $f_{\text{post}}(p, x)$):

$$f_{\text{post}}(p, x) = \frac{\mathcal{L}_n(x; p) f_{\text{prior}}(p)}{\int_\Theta \mathcal{L}_n(x; p) f_{\text{prior}}(p) \mathrm{d}p}.$$

Definicja

Bayesowskim estymatorem $\hat{p}$ parametru p nazywamy wartość oczekiwaną tego parametru wyliczoną za pomocą rozkładu a posteriori:

$$\hat{p} = \int_\Theta p f_{\text{post}}(p) \mathrm{d}p.$$

Czasami zamiast wartości oczekiwanej wg rozkładu a posteriori, definiuje się estymator bayesowski $\hat{p}$ jako dominantę (inaczej wartość modalną lub modę) rozkładu a posteriori, czyli wartość parametru p, dla której funkcja gęstości f_{post} przyjmuje wartość maksymalną.

Fakt

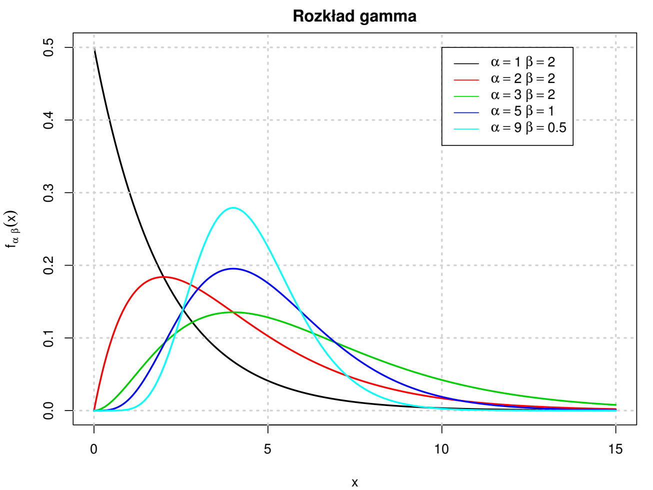
Niezależnie od wyboru (jednej z dwóch) definicji estymatora bayesowskiego, dla dużych prób (asymptotycznie) ma on te same własności co estymator największej wiarygodności.

Niezwykle użyteczne rozkłady — przypomnienie

- rozkład normalny $N(\mu, \sigma^2)$ (nośnik $\mathbb{R}$),
- rozkład gamma Gamma(α, β) (nośnik $[0, +\infty)$),
- rozkład beta Beta(α, β) (nośnik $[0, 1]$).

Rozkład gamma — przypomnienie

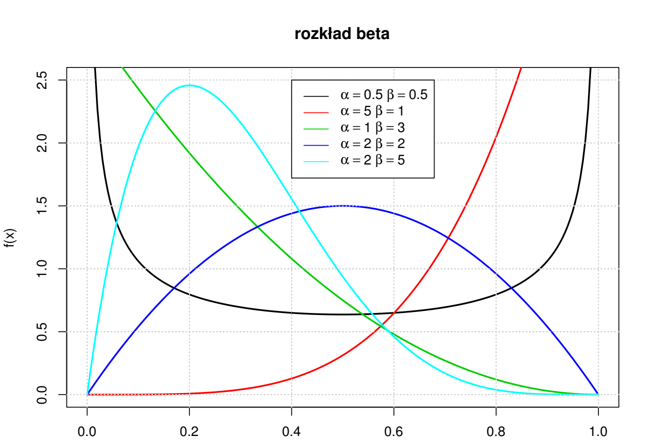
$$f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad \mathbb{E}X = \alpha\beta, \mathbb{V}X = \alpha\beta^2$$



Rozkład beta — przypomnienie

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ \alpha,\beta > 0$$

$$\mathbb{E}X = \frac{\alpha}{\alpha+\beta}, \mathbb{V}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$



Rozkład sprzężony dla modelu parametrycznego $\mathcal{F} = \{f_p(x) : p \in \Theta\}$, to taki rozkład (typ rozkładu) X, że przy rozkładzie a priori typu X dostaje się rozkład a posteriori również typu X (być może z innymi parametrami). Przykłady:

model	par.	r. sprz.	par. a priori	parametry rozkł. a posteriori
Bern	p	Beta	α, β	$\alpha + \sum_{i=1}^n x_i,\ \beta + n - \sum_{i=1}^n x_i$
Pois	λ	Γ	α, β	$\alpha + \sum_{i=1}^n x_i,\ \frac{\beta}{n\beta+1}$
N	μ	N	μ_0, σ_0^2	$\frac{\mu_0\sigma_0^{-2}+\sigma^{-2}\sum_{i=1}^n x_i}{\sigma_0^{-2}+n\sigma^{-2}},\ (\sigma_0^{-2}+n\sigma^{-2})^{-1}$
N	σ^2	Γ^{-1}	α, β	$\alpha + n/2,\ \beta + \sum_{i=1}^n (x_i-\mu)^2/2$
Exp	λ	Γ	α, β	$\alpha + n,\ \frac{\beta}{1+\beta\sum_{i=1}^n x_i}$