

Rozkłady Gamma i ich zastosowania

Henryk Gacki

Letnia Szkoła Instytutu Matematyki UŚI Wisła 2010

Równoważne definicje funkcji $\Gamma : (0, \infty) \rightarrow \mathbf{R}^+$ Dla $p > 0$

Równoważne definicje funkcji $\Gamma : (0, \infty) \rightarrow \mathbf{R}^+$ Dla $p > 0$

$$\blacksquare \Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$$

Równoważne definicje funkcji $\Gamma : (0, \infty) \rightarrow \mathbf{R}^+$ Dla $p > 0$

- $\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$
- $\Gamma(p) = \frac{1}{p} \prod_{n=1}^{\infty} \frac{(1 + \frac{1}{n})^p}{1 + \frac{p}{n}}$

Równoważne definicje funkcji $\Gamma : (0, \infty) \rightarrow \mathbf{R}^+$

Dla $p > 0$

$$\blacksquare \Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$$

$$\blacksquare \Gamma(p) = \frac{1}{p} \prod_{n=1}^{\infty} \frac{(1 + \frac{1}{n})^p}{1 + \frac{p}{n}}$$

$$\blacksquare \Gamma(p) = \lim_{n \rightarrow \infty} \frac{n! n^p}{p(p+1)(p+2)\dots(p+n)} -$$

Wzór Eulera – Gaussa

Równoważne definicje funkcji $\Gamma : (0, \infty) \rightarrow \mathbf{R}^+$

Dla $p > 0$

- $\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$
- $\Gamma(p) = \frac{1}{p} \prod_{n=1}^{\infty} \frac{(1 + \frac{1}{n})^p}{1 + \frac{p}{n}}$
- $\Gamma(p) = \lim_{n \rightarrow \infty} \frac{n! n^p}{p(p+1)(p+2)\dots(p+n)}$ – **Wzór Eulera – Gaussa**
- Funkcja Γ jest jedynym rozwiązaniem układu trzech równań funkcyjnych:

Równoważne definicje funkcji $\Gamma : (0, \infty) \rightarrow \mathbf{R}^+$

Dla $p > 0$

- $\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$
- $\Gamma(p) = \frac{1}{p} \prod_{n=1}^{\infty} \frac{(1 + \frac{1}{n})^p}{1 + \frac{p}{n}}$
- $\Gamma(p) = \lim_{n \rightarrow \infty} \frac{n! n^p}{p(p+1)(p+2)\dots(p+n)}$ – **Wzór Eulera – Gaussa**
- Funkcja Γ jest jedynym rozwiązaniem układu trzech równań funkcyjnych:

$$\begin{aligned}
 \text{(I)} \quad \Phi(p+1) &= p\Phi(p), & \text{(II)} \quad \Phi(p)\Phi(p + \frac{1}{2}) &= \frac{\sqrt{\pi}}{2^{2p-1}}\Phi(2p), \\
 \text{(III)} \quad \Phi(p)\Phi(1-p) &= \frac{\pi}{\sin p\pi} & \text{dla } p &\in (0, 1).
 \end{aligned}$$

Funkcja $\Gamma(p)$, $p \in (0, \infty)$ znana jest pod nazwą *całki Eulera drugiego rodzaju* - A. M. Legendre.

Własności funkcji Γ

Funkcja $\Gamma(p)$, $p \in (0, \infty)$ znana jest pod nazwą *całki Eulera drugiego rodzaju* - A. M. Legendre.

Własności funkcji Γ

- Funkcja Γ jest klasy $C^\infty((0, \infty))$ oraz

$$\Gamma^{(r)}(p) = \int_0^\infty x^{p-1} (\ln x)^r e^{-x} dx,$$

Funkcja $\Gamma(p)$, $p \in (0, \infty)$ znana jest pod nazwą *całki Eulera drugiego rodzaju* - A. M. Legendre.

Własności funkcji Γ

- Funkcja Γ jest klasy $C^\infty((0, \infty))$ oraz

$$\Gamma^{(r)}(p) = \int_0^\infty x^{p-1} (\ln x)^r e^{-x} dx,$$

- Z zależności $\Gamma(p+1) = p\Gamma(p)$ dostajemy
$$\Gamma(n+1) = n! \quad \text{dla } n \in \mathbb{N},$$

Funkcja $\Gamma(p)$, $p \in (0, \infty)$ znana jest pod nazwą *całki Eulera drugiego rodzaju* - A. M. Legendre.

Własności funkcji Γ

- Funkcja Γ jest klasy $C^\infty((0, \infty))$ oraz

$$\Gamma^{(r)}(p) = \int_0^\infty x^{p-1} (\ln x)^r e^{-x} dx,$$

- Z zależności $\Gamma(p+1) = p\Gamma(p)$ dostajemy

$$\Gamma(n+1) = n! \quad \text{dla } n \in \mathbb{N},$$
- Z definicji Γ :

$$\Gamma\left(\frac{1}{2}\right) = \pi^{\frac{1}{2}},$$

Funkcja $\Gamma(p)$, $p \in (0, \infty)$ znana jest pod nazwą *całki Eulera drugiego rodzaju* - A. M. Legendre.

Własności funkcji Γ

- Funkcja Γ jest klasy $C^\infty((0, \infty))$ oraz

$$\Gamma^{(r)}(p) = \int_0^\infty x^{p-1} (\ln x)^r e^{-x} dx,$$

- Z zależności $\Gamma(p+1) = p\Gamma(p)$ dostajemy

$$\Gamma(n+1) = n! \quad \text{dla } n \in \mathbb{N},$$
- Z definicji Γ : $\Gamma\left(\frac{1}{2}\right) = \pi^{\frac{1}{2}},$
- Jeśli $\gamma \sim \mathcal{N}(0, 1)$ to

$$E|\gamma|^p = \frac{2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right),$$

Gęstość gamma

Gęstości gamma, skoncentrowane na $[0, \infty)$ określone są wzorem

$$f_{p,\lambda}(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0. \end{cases}$$

dla dowolnych $p, \lambda > 0$.

Gęstość gamma

Gęstości gamma, skoncentrowane na $[0, \infty)$ określone są wzorem

$$f_{p,\lambda}(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0. \end{cases}$$

dla dowolnych $p, \lambda > 0$.

Szczególne przypadki rozkładów gamma

Gęstość gamma

Gęstości gamma, skoncentrowane na $[0, \infty)$ określone są wzorem

$$f_{p,\lambda}(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0. \end{cases}$$

dla dowolnych $p, \lambda > 0$.

Szczególne przypadki rozkładów gamma

- $f_{1,\lambda}$ to gęstość wykładnicza,

Gęstość gamma

Gęstości gamma, skoncentrowane na $[0, \infty)$ określone są wzorem

$$f_{p,\lambda}(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0. \end{cases}$$

dla dowolnych $p, \lambda > 0$.

Szczególne przypadki rozkładów gamma

- $f_{1,\lambda}$ to gęstość wykładnicza,
- $f_{\frac{1}{2},\frac{1}{2}}$ to gęstość kwadratu rozkładu normalnego,

Gęstość gamma

Gęstości gamma, skoncentrowane na $[0, \infty)$ określone są wzorem

$$f_{p,\lambda}(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0. \end{cases}$$

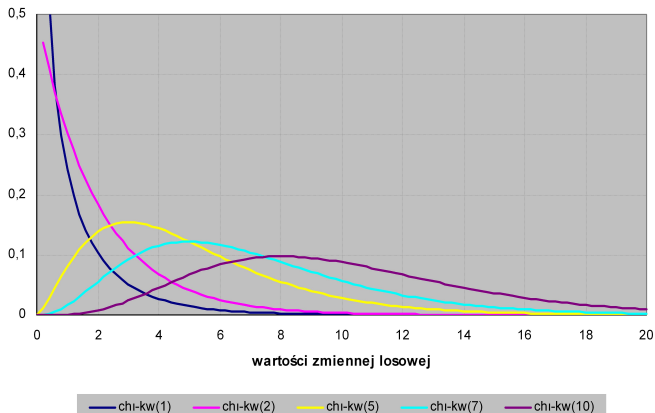
dla dowolnych $p, \lambda > 0$.

Szczególne przypadki rozkładów gamma

- $f_{1,\lambda}$ to gęstość wykładnicza,
- $f_{\frac{1}{2},\frac{1}{2}}$ to gęstość kwadratu rozkładu normalnego,
- $f_{\frac{n}{2},\frac{1}{2}}$ to gęstość χ^2 z n -stopniami swobody.

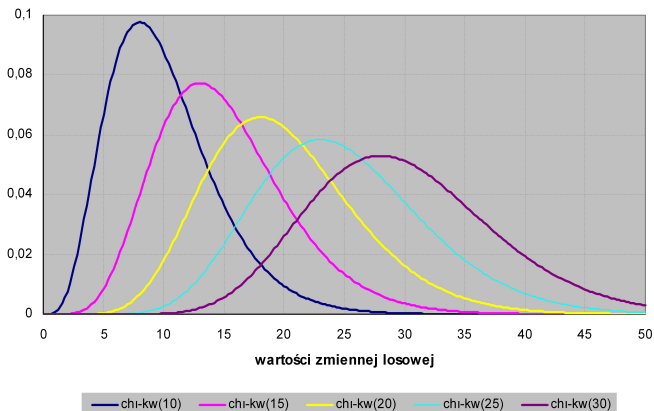
Wykresy funkcji gęstości rozkładu chi-kwadrat z 1, 2, 5, 7 i 10 stopniami swobody

Wykresy funkcji gęstości rozkładu chi-kwadrat z 1, 2, 5, 7 i 10 stopniami swobody



Wykresy funkcji gęstości rozkładu chi-kwadrat z 10, 15, 20, 25 i 30 stopniami swobody

Wykresy funkcji gęstości rozkładu chi-kwadrat z 10, 15, 20, 25 i 30 stopniami swobody



Uwaga

W statystyce $\chi^2 = \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2$ jest wariancją z próbki wybranej z populacji o rozkładzie normalnym z wartością oczekiwaną 0.

Uwaga

W statystyce $\chi^2 = \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2$ jest wariancją z próbki wybranej z populacji o rozkładzie normalnym z wartością oczekiwaną 0.

Pewne rozkłady pokrewne występujące w statystyce

Uwaga

W statystyce $\chi^2 = \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2$ jest wariancją z próbki wybranej z populacji o rozkładzie normalnym z wartością oczekiwaną 0.

Pewne rozkłady pokrewne występujące w statystyce

- Jeżeli $\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{Y}_1, \dots, \mathbf{Y}_m$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym $\mathcal{N}(0, 1)$ oraz

$$\mathbf{X} := \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2, \quad \mathbf{Y} := \mathbf{Y}_1^2 + \dots + \mathbf{Y}_m^2 \quad \text{to}$$

Uwaga

W statystyce $\chi^2 = \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2$ jest wariancją z próbki wybranej z populacji o rozkładzie normalnym z wartością oczekiwaną 0.

Pewne rozkłady pokrewne występujące w statystyce

- Jeżeli $\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{Y}_1, \dots, \mathbf{Y}_m$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym $\mathcal{N}(0, 1)$ oraz

$$\mathbf{X} := \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2, \quad \mathbf{Y} := \mathbf{Y}_1^2 + \dots + \mathbf{Y}_m^2 \quad \text{to}$$

\mathbf{X} i \mathbf{Y} mają rozkłady o gęstościach odpowiednio $f_{\frac{n}{2}, \frac{1}{2}}$, $f_{\frac{m}{2}, \frac{1}{2}}$ oraz zmienna losowa $\frac{\mathbf{X}}{\mathbf{Y}}$ ma rozkład o gęstości

Uwaga

W statystyce $\chi^2 = \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2$ jest wariancją z próbki wybranej z populacji o rozkładzie normalnym z wartością oczekiwaną 0.

Pewne rozkłady pokrewne występujące w statystyce

- Jeżeli $\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{Y}_1, \dots, \mathbf{Y}_m$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym $\mathcal{N}(0, 1)$ oraz

$$\mathbf{X} := \mathbf{X}_1^2 + \dots + \mathbf{X}_n^2, \quad \mathbf{Y} := \mathbf{Y}_1^2 + \dots + \mathbf{Y}_m^2 \quad \text{to}$$

\mathbf{X} i \mathbf{Y} mają rozkłady o gęstościach odpowiednio $f_{\frac{n}{2}, \frac{1}{2}}$, $f_{\frac{m}{2}, \frac{1}{2}}$ oraz zmienna losowa $\frac{\mathbf{X}}{\mathbf{Y}}$ ma rozkład o gęstości

$$w(t) = \frac{\Gamma(\frac{1}{2}(m+n))}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} \cdot \frac{t^{m/2-1}}{(1+t)^{(m+n)/2}}, \quad t > 0.$$

Uwaga

Zmienna losowa $F = \frac{n}{m} \cdot \frac{X}{Y}$ o gęstość $\frac{m}{n} \cdot w\left(\frac{m}{n}x\right)$ znana jest pod nazwą statystyki F-Snedecora.

Uwaga

Zmienna losowa $F = \frac{n}{m} \cdot \frac{X}{Y}$ o gęstość $\frac{m}{n} \cdot w\left(\frac{m}{n}x\right)$ znana jest pod nazwą statystyki F-Snedecora.

- Zmienna losowa $Z = \frac{1}{2} \log F$ znana pod nazwą statystyki Z Fishera.

Uwaga

Zmienna losowa $F = \frac{n}{m} \cdot \frac{X}{Y}$ o gęstość $\frac{m}{n} \cdot w\left(\frac{m}{n}x\right)$ znana jest pod nazwą statystyki **F-Snedecora**.

- Zmienna losowa $Z = \frac{1}{2} \log F$ znana pod nazwą **statystyki Z Fishera**.
- Zmienna losowa o **rozkładzie t-Studenta** o gęstości danej wzorem

$$w_1(t) = \frac{C_n}{(1+t^2)^{(n+1)/2}} \quad t \in \mathbb{R} \quad \left(C_n = \frac{1}{\sqrt{\pi}} \cdot \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{1}{2}n\right)} \right).$$

Uwaga

Zmienna losowa $F = \frac{n}{m} \cdot \frac{X}{Y}$ o gęstość $\frac{m}{n} \cdot w\left(\frac{m}{n}x\right)$ znana jest pod nazwą statystyki **F-Snedecora**.

- Zmienna losowa $Z = \frac{1}{2} \log F$ znana pod nazwą **statystyki Z Fishera**.
- Zmienna losowa o **rozkładzie t-Studenta** o gęstości danej wzorem

$$w_1(t) = \frac{C_n}{(1+t^2)^{(n+1)/2}} \quad t \in \mathbb{R} \quad \left(C_n = \frac{1}{\sqrt{\pi}} \cdot \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{1}{2}n\right)} \right).$$

W szczególności jest to gęstość statystyki $\frac{X}{\sqrt{Y_1^2 + \dots + Y_n^2}}$, gdzie X, Y_1, \dots, Y_n są niezależnymi zmiennymi losowymi o rozkładzie normalnym $\mathcal{N}(0, 1)$.

Podstawowa własność rozkładów gamma

Podstawowa własność rozkładów gamma

Dla dowolnych $p, \lambda > 0$ zachodzi równość

$$\int_0^{\infty} \mathbf{f}_{p,\lambda}(x-y)\mathbf{f}_{q,\lambda}(y) dy = \mathbf{f}_{p+q,\lambda}(x) \quad \text{dla} \quad x \in [0, \infty).$$

Podstawowa własność rozkładów gamma

Dla dowolnych $p, \lambda > 0$ zachodzi równość

$$\int_0^{\infty} \mathbf{f}_{p,\lambda}(x-y)\mathbf{f}_{q,\lambda}(y) dy = \mathbf{f}_{p+q,\lambda}(x) \quad \text{dla} \quad x \in [0, \infty).$$

Gęstości gamma są zamknięte ze względu na operację splotu.

Podstawowa własność rozkładów gamma

Dla dowolnych $p, \lambda > 0$ zachodzi równość

$$\int_0^{\infty} \mathbf{f}_{p,\lambda}(x-y)\mathbf{f}_{q,\lambda}(y) dy = \mathbf{f}_{p+q,\lambda}(x) \quad \text{dla} \quad x \in [0, \infty).$$

Gęstości gamma są zamknięte ze względu na operację splotu.

Wniosek

Dla niezależnych zmiennych losowych $\mathbf{X}_1, \dots, \mathbf{X}_n$ o gęstościach gamma $\mathbf{f}_{p_1,\lambda}, \dots, \mathbf{f}_{p_n,\lambda}$ suma $\mathbf{X}_1 + \dots + \mathbf{X}_n$ ma gęstość $\mathbf{f}_{p_1+\dots+p_n,\lambda}$.

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

W mechanice statystycznej zakłada się, że składowe prędkości $\vec{V} = (X_1, X_2, X_3)$ cząstek w \mathbb{R}^3 , zachowują niezależność przy przejściu z jednego układu kartezjańskich do drugiego oraz $EX_i = 0$, $D^2X_i = \sigma^2$ dla $i = 1, 2, 3$, gdzie $\sigma > 0$ (**Model Maxwella**).

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

W mechanice statystycznej zakłada się, że składowe prędkości $\vec{V} = (X_1, X_2, X_3)$ cząstek w \mathbb{R}^3 , zachowują niezależność przy przejściu z jednego układu kartezjańskich do drugiego oraz $EX_i = 0$, $D^2X_i = \sigma^2$ dla $i = 1, 2, 3$., gdzie $\sigma > 0$ (**Model Maxwella**).

Twierdzenie W. P. Skitowicza -1954

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

W mechanice statystycznej zakłada się, że składowe prędkości $\vec{V} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ cząstek w \mathbb{R}^3 , zachowują niezależność przy przejściu z jednego układu kartezjańskich do drugiego oraz $E\mathbf{X}_i = 0$, $D^2\mathbf{X}_i = \sigma^2$ dla $i = 1, 2, 3$., gdzie $\sigma > 0$ (**Model Maxwella**).

Twierdzenie W. P. Skitowicza -1954

Niech $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ będą niezależnymi zmiennymi losowymi, oraz niech

$$\mathbf{Y}_1 := \sum_{k=1}^n a_k \mathbf{X}_k, \quad \mathbf{Y}_2 := \sum_{k=1}^n b_k \mathbf{X}_k, \quad \mathbf{Y}_3 := \sum_{k=1}^n c_k \mathbf{X}_k$$

gdzie $|a_k| > 0$, $|c_k| > 0$ oraz $|b_k| > 0$ dla $k = 1, \dots, n$.

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

W mechanice statystycznej zakłada się, że składowe prędkości $\vec{V} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ cząstek w \mathbb{R}^3 , zachowują niezależność przy przejściu z jednego układu kartezjańskich do drugiego oraz $E\mathbf{X}_i = 0$, $D^2\mathbf{X}_i = \sigma^2$ dla $i = 1, 2, 3$, gdzie $\sigma > 0$ (**Model Maxwella**).

Twierdzenie W. P. Skitowicza -1954

Niech $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ będą niezależnymi zmiennymi losowymi, oraz niech

$$\mathbf{Y}_1 := \sum_{k=1}^n a_k \mathbf{X}_k, \quad \mathbf{Y}_2 := \sum_{k=1}^n b_k \mathbf{X}_k, \quad \mathbf{Y}_3 := \sum_{k=1}^n c_k \mathbf{X}_k$$

gdzie $|a_k| > 0$, $|c_k| > 0$ oraz $|b_k| > 0$ dla $k = 1, \dots, n$.

Jeżeli $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ są niezależnymi zmiennymi losowymi, to $\mathbf{X}_k, k = 1, \dots, n$ mają rozkład normalny.

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

Z twierdzenie Skitowicza wnioskujemy, że w modelu Maxwella $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ są niezależne i mają rozkład normalny o tej samej wariancji σ . Przykładowo, dla $\sigma = 1$, $\mathbf{X}_i \sim \mathcal{N}(0, 1), i = 1, 2, 3$.

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

Z twierdzenie Skitowicza wnioskujemy, że w modelu Maxwella $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ są niezależne i mają rozkład normalny o tej samej wariancji σ . Przykładowo, dla $\sigma = 1$, $\mathbf{X}_i \sim \mathcal{N}(0, 1)$, $i = 1, 2, 3$.

W konsekwencji $\mathbf{X}_1^2, \mathbf{X}_2^2, \mathbf{X}_3^2$ są niezależne i mają rozkład typu gamma o gęstości $f_{\frac{1}{2}, \frac{1}{2}}$. Ze wzoru na splot gęstości Γ

$$|\vec{\mathbf{V}}|^2 = \mathbf{X}_1^2 + \mathbf{X}_2^2 + \mathbf{X}_3^2 \quad \text{ma rozkład} \quad f_{\frac{3}{2}, \frac{1}{2}}$$

Zastosowanie-rozkład Maxwella prędkości cząstek w \mathbb{R}^3

Z twierdzenie Skitowicza wnioskujemy, że w modelu Maxwella $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ są niezależne i mają rozkład normalny o tej samej wariancji σ . Przykładowo, dla $\sigma = 1$, $\mathbf{X}_i \sim \mathcal{N}(0, 1)$, $i = 1, 2, 3$.

W konsekwencji $\mathbf{X}_1^2, \mathbf{X}_2^2, \mathbf{X}_3^2$ są niezależne i mają rozkład typu gamma o gęstości $f_{\frac{1}{2}, \frac{1}{2}}$. Ze wzoru na splot gęstości Γ

$$|\vec{\mathbf{V}}|^2 = \mathbf{X}_1^2 + \mathbf{X}_2^2 + \mathbf{X}_3^2 \quad \text{ma rozkład} \quad f_{\frac{3}{2}, \frac{1}{2}}$$

Stąd rozkład $|\vec{\mathbf{V}}|$ zwany rozkładem Maxwella ma gęstość

$$f_{|\vec{\mathbf{V}}|}(x) = \begin{cases} \sqrt{\frac{2}{\pi}} * x^2 e^{-\frac{x^2}{2}} & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0. \end{cases}$$

Kluczowa własność rozkładu χ^2 - twierdzenie Pearsona

Kluczowa własność rozkładu χ^2 - twierdzenie Pearsona

Rozważmy zmienną losową \mathbf{X} o danym (znanym) rozkładzie \mathbf{P} oraz próbę o n zaobserwowanych wartościach x_1, \dots, x_n .

Kluczowa własność rozkładu χ^2 - twierdzenie Pearsona

Rozważmy zmienną losową \mathbf{X} o danym (znanym) rozkładzie \mathbf{P} oraz próbę o n zaobserwowanych wartościach x_1, \dots, x_n .

Hipoteza:

Dane tworzą n -elementową próbę zmiennej losowej \mathbf{X} .

Kluczowa własność rozkładu χ^2 - twierdzenie Pearsona

Rozważmy zmienną losową \mathbf{X} o danym (znanym) rozkładzie \mathbf{P} oraz próbę o n zaobserwowanych wartościach x_1, \dots, x_n .

Hipoteza:

Dane tworzą n -elementową próbę zmiennej losowej \mathbf{X} .

Wnioskowanie będzie oparte na własnościach **pewnej miary odchylenia** pomiędzy rozkładem hipotetycznym \mathbf{P} a rozkładem μ_n danym wzorem

$$\mu_n = \sum_{k=1}^n \frac{1}{n} \delta_{x_k}.$$

Kluczowa własność rozkładu χ^2 - twierdzenie Pearsona

Rozważmy zmienną losową \mathbf{X} o danym (znanym) rozkładzie \mathbf{P} oraz próbę o n zaobserwowanych wartościach x_1, \dots, x_n .

Hipoteza:

Dane tworzą n -elementową próbę zmiennej losowej \mathbf{X} .

Wnioskowanie będzie oparte na własnościach **pewnej miary odchyłeń** pomiędzy rozkładem hipotetycznym \mathbf{P} a rozkładem μ_n danym wzorem

$$\mu_n = \sum_{k=1}^n \frac{1}{n} \delta_{x_k}.$$

Przestrzeń wartości S badanej zmiennej dzielimy na r rozłącznych zbiorów S_1, \dots, S_r i określamy prawdopodobieństwa $p_i > 0$ dla $i = 1, \dots, r$

$$p_i = \mathbf{P}(S_i) \quad \text{oraz} \quad \sum_{i=1}^r p_i = 1.$$

Oznaczamy odpowiednio liczbami ν_1, \dots, ν_r liczebności danej próby w zbiorach S_1, \dots, S_r . Miara odchyień dla rozkładów μ_n oraz \mathbf{P} jest postaci

$$\chi_n^2 = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i},$$

Oznaczamy odpowiednio liczbami ν_1, \dots, ν_r liczebności danej próby w zbiorach S_1, \dots, S_r . Miara odchyień dla rozkładów μ_n oraz \mathbf{P} jest postaci

$$\chi_n^2 = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i},$$

Twierdzenie Pearsona

Oznaczamy odpowiednio liczbami ν_1, \dots, ν_r liczebności danej próby w zbiorach S_1, \dots, S_r . Miara odchyień dla rozkładów μ_n oraz \mathbf{P} jest postaci

$$\chi_n^2 = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i},$$

Twierdzenie Pearsona

Dystrybuanta χ_n^2 jest zbieżna dla $n \rightarrow \infty$ do dystrybuanty rozkładu $\chi_{\frac{1}{2}(r-1)}^2$ określonego za pomocą funkcji gęstości $f_{\frac{1}{2}(r-1), \frac{1}{2}}$.

Oznaczamy odpowiednio liczbami ν_1, \dots, ν_r liczebności danej próby w zbiorach S_1, \dots, S_r . Miara odchyień dla rozkładów μ_n oraz \mathbf{P} jest postaci

$$\chi_n^2 = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i},$$

Twierdzenie Pearsona

Dystrybuanta χ_n^2 jest zbieżna dla $n \rightarrow \infty$ do dystrybuanty rozkładu $\chi_{\frac{1}{2}(r-1)}^2$ określonego za pomocą funkcji gęstości $f_{\frac{1}{2}(r-1), \frac{1}{2}}$.

Ponadto graniczny rozkład nie zależy od rozkładu hipotetycznego \mathbf{P} .

Symetryczności monety

Symetryczności monety

W serii złożonej z n niezależnych obserwacji zdarzenie E zaszło ν - razy. Czy obserwacje zgodne są z hipotezą, że E ma w każdym doświadczeniu dane prawdopodobieństwo $p = 1 - q$?

Symetryczności monety

W serii złożonej z n niezależnych obserwacji zdarzenie E zaszło ν - razy. Czy obserwacje zgodne są z hipotezą, że E ma w każdym doświadczeniu dane prawdopodobieństwo $p = 1 - q$?

Inaczej czy nasze dane tworzą n -elementową próbę zmiennej losowej X o rozkładzie $q\delta_0 + p\delta_1$.

Symetryczności monety

W serii złożonej z n niezależnych obserwacji zdarzenie E zaszło ν - razy. Czy obserwacje zgodne są z hipotezą, że E ma w każdym doświadczeniu dane prawdopodobieństwo $p = 1 - q$?

Inaczej czy nasze dane tworzą n -elementową próbę zmiennej losowej \mathbf{X} o rozkładzie $q\delta_0 + p\delta_1$.

W tym przypadku zbiór wartości zmiennej losowej \mathbf{X} to $\{0, 1\}$. Dzielimy go na dwa zbiory $S_1 = \{1\}$ oraz $S_2 = \{0\}$ ($r=2$).

$$\chi_n^2 = \frac{(\nu - np)^2}{np} + \frac{(n - \nu - nq)^2}{nq} = \frac{(\nu - np)^2}{npq}$$

Symetryczności monety

W serii złożonej z n niezależnych obserwacji zdarzenie E zaszło ν - razy. Czy obserwacje zgodne są z hipotezą, że E ma w każdym doświadczeniu dane prawdopodobieństwo $p = 1 - q$?

Inaczej czy nasze dane tworzą n -elementową próbę zmiennej losowej \mathbf{X} o rozkładzie $q\delta_0 + p\delta_1$.

W tym przypadku zbiór wartości zmiennej losowej \mathbf{X} to $\{0, 1\}$. Dzielimy go na dwa zbiory $S_1 = \{1\}$ oraz $S_2 = \{0\}$ ($r=2$).

$$\chi_n^2 = \frac{(\nu - np)^2}{np} + \frac{(n - \nu - nq)^2}{nq} = \frac{(\nu - np)^2}{npq}$$

Na podstawie twierdzenia Pearsona zmienna ta ma rozkład asymptotycznie $\chi_{\frac{1}{2}}^2$ o gęstości $f_{\frac{1}{2}, \frac{1}{2}}$ (kwadrat rozkładu normalnego).

Symetryczności monety

W serii złożonej z n niezależnych obserwacji zdarzenie E zaszło ν - razy. Czy obserwacje zgodne są z hipotezą, że E ma w każdym doświadczeniu dane prawdopodobieństwo $p = 1 - q$?

Inaczej czy nasze dane tworzą n -elementową próbę zmiennej losowej \mathbf{X} o rozkładzie $q\delta_0 + p\delta_1$.

W tym przypadku zbiór wartości zmiennej losowej \mathbf{X} to $\{0, 1\}$. Dzielimy go na dwa zbiory $S_1 = \{1\}$ oraz $S_2 = \{0\}$ ($r=2$).

$$\chi_n^2 = \frac{(\nu - np)^2}{np} + \frac{(n - \nu - nq)^2}{nq} = \frac{(\nu - np)^2}{npq}$$

Na podstawie twierdzenia Pearsona zmienna ta ma rozkład asymptotycznie $\chi_{\frac{1}{2}}^2$ o gęstości $f_{\frac{1}{2}, \frac{1}{2}}$ (kwadrat rozkładu normalnego).

Zgadza się to z faktem, że standaryzowana zmienna losowa $\frac{(\nu - np)}{\sqrt{npq}}$ ma rozkład asymptotycznie normalny.

Symetryczności monety

Symetryczności monety

W $n = 4040$ rzutach monetą Buffon otrzymał $\nu = 2048$ orłów i $n - \nu = 1992$ reszek. Czy rzucał monetą symetryczną?

$$\chi_n^2 = \frac{(\nu - np)^2}{npq} = 0,776.$$

Symetryczności monety

W $n = 4040$ rzutach monetą Buffon otrzymał $\nu = 2048$ orłów i $n - \nu = 1992$ reszek. Czy rzucał monetą symetryczną?

$$\chi_n^2 = \frac{(\nu - np)^2}{npq} = 0,776.$$

Z tablic rozkładu $\chi_{\frac{1}{2}}^2$ dostaniemy

$$Prob(\chi_{\frac{1}{2}}^2 > 3,841) = 0.05$$

Symetryczności monety

W $n = 4040$ rzutach monetą Buffon otrzymał $\nu = 2048$ orłów i $n - \nu = 1992$ reszek. Czy rzucał monetą symetryczną?

$$\chi_n^2 = \frac{(\nu - np)^2}{npq} = 0,776.$$

Z tablic rozkładu $\chi_{\frac{1}{2}}^2$ dostaniemy

$$Prob(\chi_{\frac{1}{2}}^2 > 3,841) = 0.05$$

W związku z tym obserwacje należy uważać za zgodne z hipotezą.