

Rachunek prawdopodobo- bieństwa i statystyka

Wykład 12

Rozkład χ^2

Niech zmienna losowa x ma rozkład normalny $\mathcal{N}(x; \mu, \sigma)$. Znajdziemy rozkład zmiennej:

$$u = \left(\frac{x - \mu}{\sigma} \right)^2$$

Standaryzując zmienną losową $x \rightarrow \frac{x - \mu}{\sigma}$ otrzymujemy standaryzowany rozkład Gaussa:

$$\mathcal{N}(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Rozkład zmiennej $u = x^2$ ma więc postać:

$$\mathcal{X}_1(u) = \left(\mathcal{N}(-\sqrt{u}) + \mathcal{N}(\sqrt{u}) \right) \left| \frac{dx}{du} \right| = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{u}{2}\right)$$

Rozważmy dwie niezależne zmienne losowe x_1 i x_2 z tego samego rozkładu normalnego, i znajdziemy rozkład zmiennej losowej:

$$u = \left(\frac{x_1 - \mu}{\sigma} \right)^2 + \left(\frac{x_2 - \mu}{\sigma} \right)^2$$

Szukany rozkład ma postać:

$$\mathcal{X}_2(u) = \int_0^u \mathcal{X}_1(u-t) \mathcal{X}_1(t) dt = \frac{1}{2\pi} \exp\left(-\frac{u}{2}\right) \int_0^u \frac{dt}{\sqrt{t(u-t)}} = \frac{1}{2} \exp\left(-\frac{u}{2}\right)$$

Rozkład χ^2

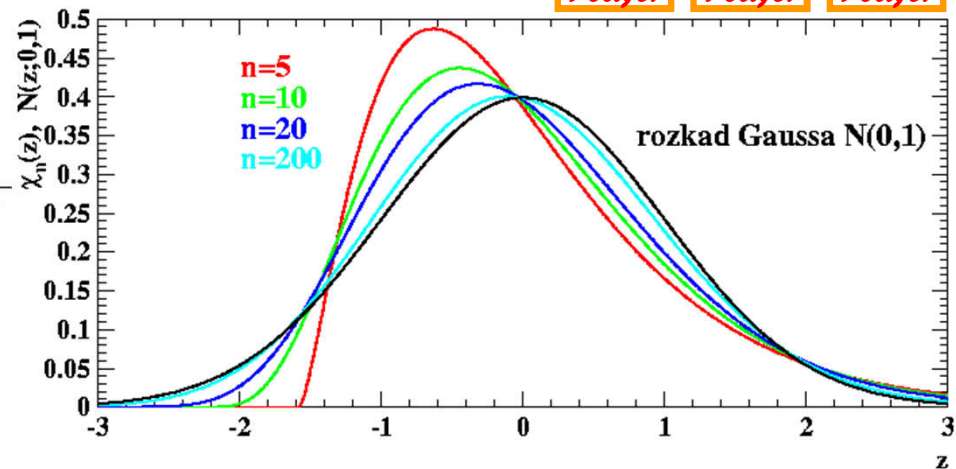
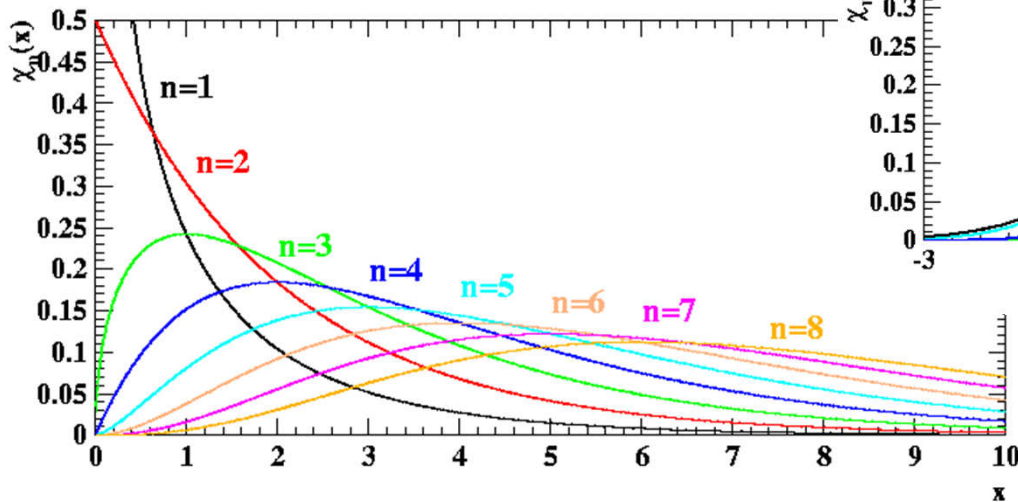
Postępując tak dalej znajdujemy rozkład sumy kwadratów n niezależnych, standaryzowanych zmiennych losowych każda z rozkładu Gaussa, zwany **rozkładem χ^2** :

$$\mathcal{X}_n(u) = \frac{1}{(\sqrt{2})^n \Gamma\left(\frac{n}{2}\right)} u^{\frac{n}{2}-1} \exp\left(-\frac{u}{2}\right), \quad n > 0, u \geq 0$$

gdzie indeks n oznacza liczbę stopni swobody.

Wartość oczekiwana i wariancja:

$$\mathcal{E}[u] = n \quad \mathcal{V}[u] = 2n$$



Math Player Math Player Math Player

Zachowanie dla dużych n :

$$z = \frac{u - n}{\sqrt{2n}} \Rightarrow u = n + z\sqrt{2n}$$

Rozkład χ^2

Jak zmieni się rozkład zmiennej losowej u w przypadku gdy:

$$u = \left(\frac{x_1 - \bar{x}}{\sigma} \right)^2 + \left(\frac{x_2 - \bar{x}}{\sigma} \right)^2 + \dots + \left(\frac{x_n - \bar{x}}{\sigma} \right)^2$$

Wychodząc z łącznego rozkładu p-twa zmiennych x_1, x_2, \dots, x_n :

$$f(x_1, x_2, \dots, x_n; \mu, \sigma) = \prod_{i=1}^n \mathcal{N}(x_i; \mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

można pokazać, że łączny rozkład p-twa zmiennych u oraz \bar{x} ma postać:

$$g(u, \bar{x}; \mu, \sigma) = \frac{1}{(\sqrt{2})^{n-1} \Gamma\left(\frac{n-1}{2}\right)} u^{\frac{n-1}{2}-1} \exp\left(-\frac{u}{2}\right) \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} \exp\left\{-\frac{(\bar{x} - \mu)^2}{2 \frac{\sigma^2}{n}}\right\} = \chi_{n-1}(u) \mathcal{N}\left(\bar{x}; \mu, \frac{\sigma}{\sqrt{n}}\right)$$

Wnioski:

1. Zmienne u i \bar{x} są statystycznie niezależne.
2. Rozkład zmiennej u to rozkład χ^2 o $n-1$ stopniach swobody (użycie średniej zmniejsza liczbę niezależnych składników w sumie).

Rozkład χ^2

Korzystając z rozkładu zmiennej losowej u znajdujemy rozkład zmiennej:

$$v \equiv s_x^2 = \frac{\sigma^2}{n-1} u = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

który ma postać:

$$f(v; \sigma, n) = \frac{\frac{n-1}{\sigma^2}}{(\sqrt{2})^{n-1} \Gamma\left(\frac{n-1}{2}\right)} \left(\frac{(n-1)v}{\sigma^2}\right)^{\frac{n-1}{2}-1} \exp\left(-\frac{(n-1)v}{2\sigma^2}\right)$$

Wartość oczekiwana zmiennej v : $\mathcal{E}[v] = \mathcal{E}[s_x^2] = \frac{\sigma^2}{n-1} \mathcal{E}[u] = \frac{\sigma^2}{n-1} (n-1) = \sigma^2$

Wariancja zmiennej v : $\mathcal{V}[v] = \mathcal{V}[s_x^2] = \left(\frac{\sigma^2}{n-1}\right)^2 \mathcal{V}[u] = \left(\frac{\sigma^2}{n-1}\right)^2 2(n-1) = \frac{2}{n-1} \sigma^4$

Wartość oczekiwana i wariancja estymatora s_x :

$$\mathcal{E}[s_x] \cong \sqrt{\frac{n}{n-1}} \sigma \quad \mathcal{V}[s_x] \cong \frac{1}{2n} \sigma^2 \quad \hat{\mathcal{V}}[s_x] = \frac{1}{2n} s_x^2$$

Podobna analiza prowadzi do wniosku, że statystyką o $n-1$ stopniach

swobody jest też: $u = \frac{n(n-1)}{\sigma^2} s_x^2$ gdzie $s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$

Rozkład Studenta

Aby uniknąć problemu nieznanego rozkładu normalnego w zmiennej typu χ^2 , rozważmy statystykę postaci (Studenta):

$$\frac{\bar{x} - \mu}{s_{\bar{x}}} = \left\| s_{\bar{x}}^2 = \frac{\sigma^2}{n(n-1)} u \right\| = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n(n-1)} u}} = \frac{\frac{\bar{x} - \mu}{\sigma}}{\sqrt{\frac{u}{n-1}}}$$

Uogólnijmy rozważaną zmienną losową na zmienną: $t = \frac{\frac{x - \mu}{\sigma}}{\sqrt{\frac{u}{n}}} \rightarrow t = \frac{z}{\sqrt{\frac{u}{n}}}$

gdzie x jest zmienną z rozkładu normalnego o parametrach μ i σ , a z jest standaryzowaną zmienną gaussowską. Natomiast zmienna u podlega rozkładowi χ^2 o n stopniach swobody.

Łączna gęstość p-twa zmiennych z oraz u ma postać:

$$h_n(z, u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \frac{1}{(\sqrt{2})^n \Gamma\left(\frac{n}{2}\right)} u^{\frac{n}{2}-1} \exp\left(-\frac{u}{2}\right)$$

Rozkład Studenta

W standardowy sposób znajdujemy rozkład zmiennej losowej t:

$$\begin{cases} \mathbf{t} = \frac{\mathbf{z}}{\sqrt{\mathbf{u}/n}} \\ \mathbf{v} = \mathbf{u} \end{cases} \Rightarrow \begin{cases} \mathbf{z} = \mathbf{t} \sqrt{\frac{\mathbf{v}}{n}} \\ \mathbf{u} = \mathbf{v} \end{cases} \Rightarrow J = \begin{vmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{t}} & \frac{\partial \mathbf{z}}{\partial \mathbf{v}} \\ \frac{\partial \mathbf{u}}{\partial \mathbf{t}} & \frac{\partial \mathbf{u}}{\partial \mathbf{v}} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{\mathbf{v}}{n}} & \frac{\mathbf{t}}{2\sqrt{\mathbf{v}n}} \\ \mathbf{0} & \mathbf{1} \end{vmatrix} = \sqrt{\frac{\mathbf{v}}{n}}$$

Łączny rozkład p-twa zmiennych t oraz v:

$$\begin{aligned} g_n(t, v) &= h_n z((t, v), u(t, v)) |J(t, v)| = \\ &= \frac{1}{\sqrt{2\pi} (\sqrt{2})^n \Gamma\left(\frac{n}{2}\right)} v^{\frac{n}{2}-1} \sqrt{\frac{v}{n}} \exp\left(-\frac{v}{2} - \frac{t^2 v}{2n}\right) = \frac{1}{\sqrt{\pi n} (\sqrt{2})^{n+1} \Gamma\left(\frac{n}{2}\right)} v^{\frac{n+1}{2}-1} \exp\left(-\frac{v}{2} \left(1 + \frac{t^2}{n}\right)\right) \end{aligned}$$

Rozkład brzegowy zmiennej t:

$$\begin{aligned} \mathcal{S}_n(t) &= \int_0^{\infty} g_n(t, v) dv = \left\| q = v \left(1 + \frac{t^2}{n}\right) \right\| = \\ &= \frac{1}{\sqrt{\pi n} (\sqrt{2})^{n+1} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \int_0^{\infty} q^{\frac{n+1}{2}-1} \exp\left(-\frac{q}{2}\right) dq = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \end{aligned}$$

Rozkład Studenta

Rozkład Studenta o n stopniach swobody:

$$S_n(t) = \int_0^{\infty} g_n(t, \nu) d\nu = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \quad n = 1, 2, \dots \quad -\infty < t < \infty$$

Dla $n=1$ rozkład Studenta przechodzi w rozkład Cauchy'ego:

$$S(t) = \frac{1}{\pi} \frac{1}{1+t^2}$$

Uwaga: Dla rozkładu Cauchy'ego nie istnieją wartość oczekiwana ani wariancja.

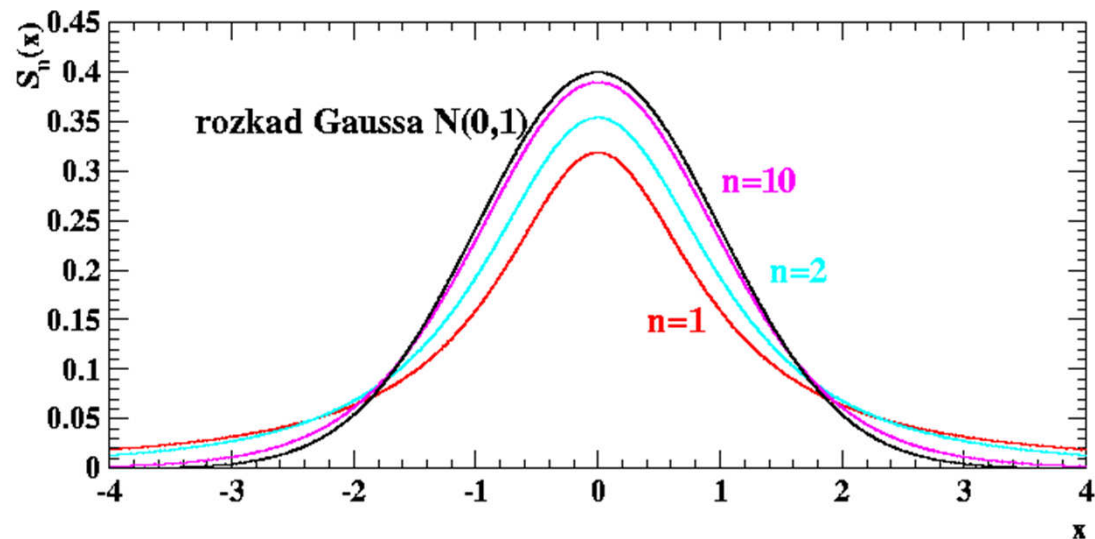
Math
Player

Wartość oczekiwana rozkładu Studenta (dla $n > 1$):

$$\mathcal{E}[t] = \int_{-\infty}^{\infty} t S_n(t) dt = 0$$

Wariancja (dla $n > 2$):

$$\mathcal{V}[t] = \int_{-\infty}^{\infty} t^2 S_n(t) dt = \frac{n}{n-2}$$



Rozkład Studenta

Przykład: Sprawdzamy rzetelność producenta cukru. Wazymy 5 toreb cukru (których waga netto powinna wynosić 1 kg) otrzymując $\bar{x} = 960$ g oraz $s_{\bar{x}} = 10$ g.

Czy producent jest uczciwy w ramach trzech odchyłeń standardowych?

Rozkład normalny: $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

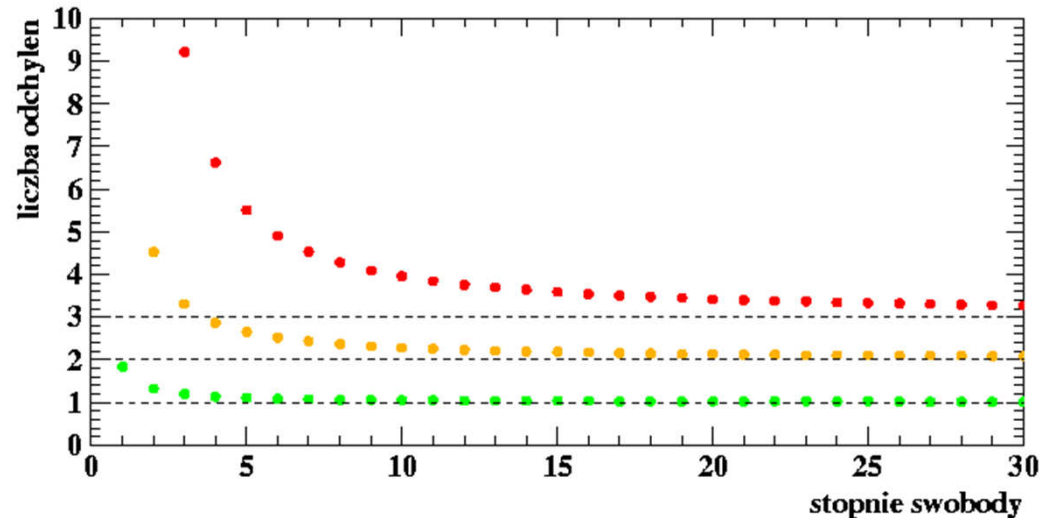
Obliczamy wartość statystyki Studenta: $t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{960 - 1000}{10} = -4$

Obliczmy p-two:

$$\begin{aligned} P(960 \leq \bar{x} \leq 1040) &= \\ &= P(-4 \leq t \leq 4) \cong 0.9839 \end{aligned}$$

Dla rozkładu Studenta:

$$\begin{aligned} P(-6.6 \leq t \leq 6.6) &= \\ P(934 < \bar{x} < 1066) &\cong 0.9973 \end{aligned}$$



Rozkład \mathcal{F} Snedecora-Fishera

Rozważmy dwie zmienne losowe u_1 oraz u_2 każda o rozkładzie χ^2 , odpowiednio o n i m stopniach swobody. Konstruujemy statystykę Fishera:

$$\mathbf{F} = \left(\frac{\mathbf{u}_1}{n} \right) / \left(\frac{\mathbf{u}_2}{m} \right) = \frac{m}{n} \frac{\mathbf{u}_1}{\mathbf{u}_2}$$

Łączna gęstość p-twa zmiennych u_1 oraz u_2 ma postać:

$$h_{n,m}(u_1, u_2) = \frac{1}{(\sqrt{2})^n \Gamma\left(\frac{n}{2}\right)} (u_1)^{\frac{n}{2}-1} \exp\left(-\frac{u_1}{2}\right) \frac{1}{(\sqrt{2})^m \Gamma\left(\frac{m}{2}\right)} (u_2)^{\frac{m}{2}-1} \exp\left(-\frac{u_2}{2}\right)$$

Interesuje nas rozkład zmiennej losowej \mathbf{F} :

$$\begin{cases} \mathbf{F} = \frac{m}{n} \frac{\mathbf{u}_1}{\mathbf{u}_2} \\ \mathbf{v} = \mathbf{u}_2 \end{cases} \Rightarrow \begin{cases} \mathbf{u}_1 = \frac{n}{m} \mathbf{F} \mathbf{v} \\ \mathbf{u}_2 = \mathbf{v} \end{cases} \Rightarrow \mathbf{J} = \begin{vmatrix} \frac{\partial \mathbf{u}_1}{\partial \mathbf{F}} & \frac{\partial \mathbf{u}_1}{\partial \mathbf{v}} \\ \frac{\partial \mathbf{u}_2}{\partial \mathbf{F}} & \frac{\partial \mathbf{u}_2}{\partial \mathbf{v}} \end{vmatrix} = \begin{vmatrix} \frac{n}{m} \mathbf{v} & \frac{n}{m} \mathbf{F} \\ \mathbf{0} & \mathbf{1} \end{vmatrix} = \frac{n}{m} \mathbf{v}$$

$$g_{n,m}(\mathbf{F}, \mathbf{v}) = \frac{1}{(\sqrt{2})^{n+m} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{m}{2}\right)} \frac{n}{m} \left(\frac{n}{m} \mathbf{F}\right)^{\frac{n}{2}-1} \mathbf{v}^{\frac{1}{2}(n+m)-1} \exp\left(-\frac{\mathbf{v}}{2} \left(\frac{n}{m} \mathbf{F} + 1\right)\right)$$

Rozkład \mathcal{F} Snedecora-Fishera

Rozkład brzegowy zmiennej F nazywamy rozkładem \mathcal{F} - Fishera o (n,m) stopniach swobody:

$$\mathcal{F}_{n,m}(F) = \frac{\Gamma\left(\frac{1}{2}(n+m)\right) n^{\frac{n}{2}} m^{\frac{m}{2}} F^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)(m+nF)^{\frac{1}{2}(n+m)}}, \quad n, m > 0, \quad F \geq 0$$

Wartość oczekiwana (dla $m > 2$): $\mathcal{E}[F] = \frac{m}{m-2}$

Wariancja (dla $m > 4$):

$$\mathcal{V}[F] = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$$

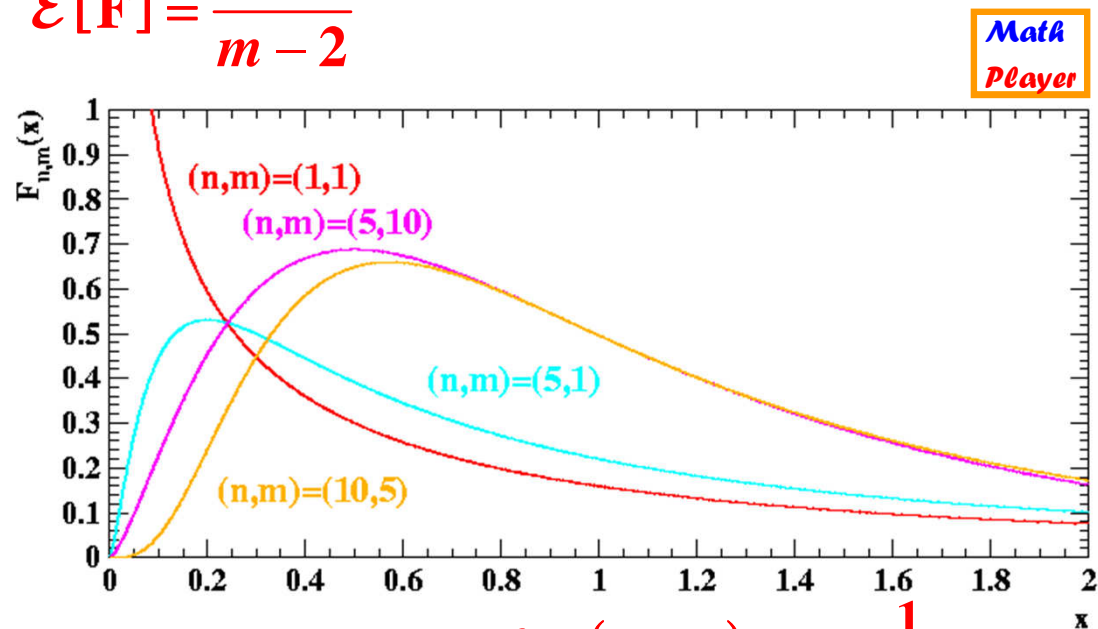
Zachowania graniczne:

$n=1$: rozkład $S_m(t)$

$m \rightarrow \infty$: rozkład $\chi_n(nF)$

$n \rightarrow \infty$ i $m \rightarrow \infty$: r. Gaussa

Uwaga: Kwantyle rozkładu \mathcal{F} - Fishera spełniają relację $f_{1-\alpha}(n_2, n_1) = \frac{1}{f_\alpha(n_1, n_2)}$



Rozkład \mathcal{F} Snedecora-Fishera

Przykład: Porównanie odchyłeń standardowych. Konstruujemy dwie statystyki u_x oraz u_y podlegające rozkładowi χ^2 o $n-1$ i $m-1$ st. swobody:

$$u_x = \frac{n-1}{\sigma_x^2} s_x^2 = \frac{1}{\sigma_x^2} \sum_{i=1}^n (x_i - \bar{x})^2 \quad u_y = \frac{m-1}{\sigma_y^2} s_y^2 = \frac{1}{\sigma_y^2} \sum_{i=1}^m (y_i - \bar{y})^2$$

Z wielkości tych możemy zbudować zmienną F Fishera:

$$F = \left(\frac{u_x}{n-1} \right) / \left(\frac{u_y}{m-1} \right) = \left(\frac{\frac{n-1}{\sigma_x^2} s_x^2}{n-1} \right) / \left(\frac{\frac{m-1}{\sigma_y^2} s_y^2}{m-1} \right) = \frac{s_x^2 / \sigma_x^2}{s_y^2 / \sigma_y^2}$$

Zmienna ta podlega rozkładowi \mathcal{F} -Fishera o $(n-1, m-1)$ stopniach swobody i dostarcza narzędzia do testowania hipotez o równości wariacji w dwu próbach prostych wylosowanych z rozkładu normalnego.

Rozkład \mathcal{F} Snedecora-Fishera

Przykład: Zakładając, że s_x^2 i s_y^2 są estymatorami wariancji obliczonymi z próbek $n_1=10$ i $n_2=8$ elementowych pobranych z rozkładów takich, że $\sigma_x^2 = 3\sigma_y^2$, znajdź wartości a i b dla których zachodzi

$$\mathbf{P\left(a \leq \frac{s_x^2}{s_y^2} \leq b\right) = 0.90}$$

Wiemy, że zmienna losowa:

$$\frac{s_x^2 / \sigma_x^2}{s_y^2 / \sigma_y^2} = \frac{s_x^2 / 3\sigma_y^2}{s_y^2 / \sigma_y^2} = \frac{s_x^2}{3s_y^2}$$

ma rozkład \mathcal{F} -Fishera o $n_1-1=9$ i $n_2-1=7$ stopniach swobody, a więc:

$$\mathbf{0.90 = P\left(f_{0.05}(9,7) \leq \frac{s_x^2}{3s_y^2} \leq f_{0.95}(9,7)\right) = P\left(\frac{1}{f_{0.95}(7,9)} \leq \frac{s_x^2}{3s_y^2} \leq f_{0.95}(9,7)\right)}$$

Z tablic odczytujemy: $f_{0.95}(7,9) = 3.29$ oraz $f_{0.95}(9,7) = 3.68$

Ostatecznie więc otrzymujemy:

$$\mathbf{0.90 = P\left(\frac{1}{3.29} \leq \frac{s_x^2}{3s_y^2} \leq 3.68\right) = P\left(0.912 \leq \frac{s_x^2}{s_y^2} \leq 11.04\right)}$$