Komentarz do wykładów z 23. oraz 30. kwietnia

Różne przybliżenia wariancji

Zakładamy, że dane są niezależne obserwacje pochodzące z tego samego rozkładu $\mathrm{N}(\mu, \sigma^2)$. Wiadomo,[1] że zmienna $\frac{nS_\mu^2}{\sigma^2}$ ma rozkład $\chi^2(n)$ a także[2] $\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$.
Wspomnijmy również, że $\chi^2(k) \equiv \mathrm{Gamma}(1/2, k/2)$. Stąd, dla rozkładu $\chi^2(k)$: $M_{\chi^2(k)}(t) = (1-2t)^{-k/2}$.
To umożliwia sformułowanie – w miarę prostego do dowodu – twierdzenia

Twierdzenie 1. *Jeżeli zmienna losowa $X \sim \chi^2(k)$, to $\mathrm{E}(X) = k$ oraz $\mathrm{V}(X) = 2k$.*

Rozpatrzmy trzy estymatory wariancji σ^2, mianowicie $S_{n-1}^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$, $S_n^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$ oraz $S_{n+1}^2 = \frac{1}{n+1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$. Z początkowej uwagi wynika, że zmienne $\frac{(n-1)S_{n-1}^2}{\sigma^2}$, $\frac{nS_n^2}{\sigma^2}$, $\frac{(n+1)S_{n+1}^2}{\sigma^2}$ mają rozkład $\chi^2(n-1)$ każda.

Wartość oczekiwana $\mathrm{E}\left(\frac{nS_n^2}{\sigma^2}\right) = n-1$. Stąd, $\mathrm{E}(S_n^2) = \frac{n-1}{n}\sigma^2$. Wartość oczekiwana przybliżenia S_n^2 jest zatem różna od wartości przybliżanego parametru σ^2. O takim estymatorze mówimy, że jest **estymatorem obciążonym**. Jeżeli $n \to \infty$, to $\mathrm{E}(S_n^2) \to \sigma^2$. Mówimy wówczas, że S_n^2 jest estymatorem **asymptotycznie nieobciążonym** parametru σ^2.

Ponieważ $S_{n-1}^2 = \frac{n}{n-1}S_n^2$, więc $\mathrm{E}(S_{n-1}^2) = \sigma^2$. Mówimy, że S_{n-1}^2 jest **nieobciążonym estymatorem** parametru σ^2. Podobnie $S_{n+1}^2 = \frac{n}{n+1}S^2$ czyli $\mathrm{E}(S_{n+1}^2) = \frac{n-1}{n+1}\sigma^2$ (S_{n+1}^2 jest estymatorem obciążonym dla σ^2). Najlepszym estymaterem (uwzględniając wartość oczekiwaną) jest zatem S_{n-1}^2, najgorszym S_{n+1}^2.

Porównajmy teraz wariancje rozpatrywanych estymatorów. Wiemy, że $\frac{n\,S_n^2}{\sigma^2} \sim \chi^2(n-1)$. Stąd wynika, że $\mathrm{V}(S_n^2) = \frac{2(n-1)}{n^2}\sigma^4$, $\mathrm{V}(S_{n-1}^2) = \frac{2(n-1)}{(n-1)^2}\sigma^4$ oraz $\mathrm{V}(S_{n+1}^2) = \frac{2(n-1)}{(n+1)^2}\sigma^4$. Im mniejsza wariancja, tym bardziej zmienna losowa jest "stabilna". Najlepszym estymatorem (kierując się wariancją) jest S_{n+1}^2, najgorszym S_{n-1}^2.

Estymatory największej wiarygodności (MLE estimators)

Zakładamy że dane są niezależne obserwacje zmiennej losowej X o tym samym rozkładzie. Funkcja gęstości zmiennej X ma postać $f(x;\theta)$, gdzie θ to parametr/parametry rozkładu. Można też uważać, że dane są niezależne zmienne $X_1, \ldots, X_n$ o tym samym rozkładzie. Prawdopodobieństwo zdarzenia można zatem zapisać jako funkcję wiarygodności L względem zmiennej θ

$$L(x;\theta) = P\left(X_1 = x_1, \ldots, X_n = x_n\right) = \prod_{k=1}^{n} f(x_k;\theta). \tag{1}$$

Zakładamy, że obserwowane wartości są typowe (najbardziej prawdopodobne). Chcemy zatem aby funkcja wiarygodności osiągała maksimum w pewnym punkcie $\hat{\theta}$. Oważ wyznaczona wartość $\hat{\theta}$ nazywana jest *estymatorem największej wiarygodności parametru* θ.

[1] Notatka 6, wzór (3).
[2] Notatka 6, tw. 4.

UWAGA: Bardzo często szukamy maksimum funkcji $\ln L(x;\theta)$, wyłącznie z powodów obliczeniowych; logarytm (naturalny) jako funkcja rosnąca daje tę samą odpowiedź.

Przykład:

P1: Rozważmy n niezależnych obserwacji z rozkładu $\mathrm{Exp}(\lambda)$. Prawdopodobieństwo zdarzenia $P(X_1 = x_1, \ldots X_n = x_n)$ jest równe – z racji niezależności – $L(\lambda) = \lambda^n \exp\left(-\sum_{k=1}^{n} x_i\right)$. Chcemy znaleźć wartość λ taką iż funkcja (λ) osiąga maksimum. [a]

Maksima funkcji $L(\lambda)$ oraz $\ln L(\lambda)$ znajdują się w tym samym punkcie $\hat{\lambda}$. Obliczając – i przyrównując do zera, – pochodną $\frac{\partial \ln L}{\partial \lambda}$ otrzymujemy równanie

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} + \left(\sum_{k=1}^{n} x_i\right) = \frac{n}{\lambda} - n \cdot \bar{x} = 0,$$

skąd wynika, że $\hat{\lambda} = 1/\bar{x}$. Druga pochodna funkcji $\ln L(\lambda)$ równa $\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{n}{\lambda^2}$ jest ≤ 0 w każdym punkcie λ, czyli również dla wyznaczonej wcześniej wartości $\hat{\lambda}$, co dowodzi iż znaleźliśmy maksimum funkcji wiarygodności $\equiv$ estymator MLE $\hat{\lambda}$ parametru λ.

[a] Intuicja: to, co obserwujemy, jest najbardziej prawdopodobne.

Przykład:

P2: Rozważamy n niezależnych obserwacji z rozkładu $\mathrm{B}(n,p)$. Funkcja wiarygodności ma teraz postać

$$\begin{aligned} L(p) &= P(X_1 = x_1, \ldots, X_k = x_k) = \prod_{i=1}^{k} P(X_i = x_i) = \\ &= p^{\sum x_i} \, (1-p)^{nk - \sum x_i} \prod_{i=1}^{k} \binom{n}{x_i} = p^{k\bar{x}} \, (1-p)^{nk - k\bar{x}} \prod_{i=1}^{k} \binom{n}{x_i}. \end{aligned} \tag{2}$$

Logarytm funkcji wiarygodności ma zatem postać

$$\ln L(p) = \sum_{i=1}^{k} \ln\binom{n}{x_i} + k\bar{x} \ln p + k(n - \bar{x}) \ln(1-p),$$

a jego pochodna to

$$\frac{\partial \ln L}{\partial p} = \frac{k\bar{x}}{p} - \frac{k(n-\bar{x})}{1-p} = 0.$$

Rozwiązując powyższe równanie otrzymujemy dla estymatora MLE wyrażenie $\hat{p} = \frac{\bar{x}}{n}$. Druga pochodna funkcji wiarygodności to

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{k\bar{x}}{p^2} - \frac{k(n-\bar{x})}{(1-p)^2} < 0,$$

dla $0 < \hat{p} < n$ zatem znaleźliśmy maksimum funkcji $L(p)$.

Jeżeli $\hat{p} = 0$, to $x_1 = \ldots = x_k = 0$. Funkcja wiarygodności (2) ma postać $L(p) = (1-p)^{nk}$ i osiąga maksimum dla $p = 0$. Podobnie, jeżeli $\hat{p} = n$, to $x_1 = \ldots = x_k = n$. Funkcja wiarygodności (2) ma w tym wypadku postać $L(p) = p^{nk}$ i osiąga maksimum dla $p = 1$.

Analiza wariancji – ANalysis Of VAriance – ANOVA

Załóżmy, że dane są niezależne obserwacje zmiennej losowej $X \sim N(\mu, \sigma^2)$. Wyniki obserwacji grupujemy względem pewnej cechy jakościowej, wyróżniamy I grup. Dla każdej grupy mamy J obserwacji. Symbolem x_{ij} oznaczamy j-tą obserwację w i-tej grupie ($i = 1, \ldots, I$; $j = 1, \ldots, J$).

Grupa	Obserwacje				Średnie
1	x_{11}	x_{12}	$\ldots$	x_{1J}	$x_{1\bullet}$
2	x_{11}	x_{12}	$\ldots$	x_{1J}	$x_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
I	x_{I1}	x_{I2}	$\ldots$	x_{IJ}	$x_{I\bullet}$

Ostatnia kolumna powyższej tabeli zawiera średnie grup (wierszy). tzn. $x_{k\bullet} = \frac{1}{J}\sum_{j=1}^{J} x_{kj}$. Symbolem $\bar{x}$ oznaczamy średnią wszystkich obserwacji: $\bar{x} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}$. Przykłady grupowania danych: 3 grupy opon (zimowe, letnie, uniwersalne) i notujemy stopień zużycia po określonym przebiegu; skuteczność pewnego leku w grupach: początkowe stadium choroby, choroba w pełnym objawie, stan ciężki; porównujemy podobne leki od 3 producentów itp.

Zakładamy – zgodnie z założeniem początkowym, – że każda ze zmiennych losowych $X_{i\bullet}$ ma rozkład $N\left(\mu, \sigma^2/n\right)$. W istocie chcemy zaprzeczyć temu założeniu, to znaczy wywnioskować z danych iż jedna z grup (niektóre, wiele) jest różna od pozostałych. Kolejne obliczenia powinny wskazać, która grupa odróżnia się "na plus", ale to na razie odkładamy na później. To co nas interesuje, to odpowiedź w postaci: opony A są lepsze od pozostałych; pewne lekarstwo najbardziej nadaje się do któregoś stadium choroby; producent C ma najlepszy produkt.

Interesującym faktem jest iż możemy powiedzieć coś o średnich grup ($x_{i\bullet}$) na podstawie wariancji wszystkich obserwacji oraz wariancji wewnątrz grup (wierszy). Rozpocznijmy od wariancji wszystkich obserwacji. Wielokrotnie stosowaliśmy wzór

$$
\begin{array}{ccccc|c}
\sum_{k=1}^{n}(X_k-\mu)^2 & = & \sum_{k=1}^{n}\left(X_k-\bar{X}\right)^2 & + & n\cdot\left(\bar{X}-\mu\right)^2 & \text{wzór} \\
\frac{nS_\mu^2}{\sigma^2} & = & \frac{nS^2}{\sigma^2} & + & \left(\frac{\bar{X}-\mu}{\sigma\sqrt{n}}\right)^2 & \text{skróty} \\
\chi^2(n) & = & \chi^2(n-1) & + & \chi^2(1) & \text{rozkłady}
\end{array}
\tag{3}
$$

Podstawowy fakt jest taki: dla obserwacji x_{ij} zmienna losowa $\frac{nS^2}{\sigma^2} = \sum_{ij}\left(X_{ij}-\bar{X}\right)^2$ ma rozkład $\chi^2(IJ-1)$, ponieważ mamy I grup po J obserwacji oraz "znika" jeden stopień swobody, jak wynika ze wzoru (3).

Zróżnicowanie obserwacji przedstawiamy jako $\mathrm{SS}_{\mathrm{Tot}} = \sum_{i,j}(x_{ij}-\bar{x})^2$. Z dokładnością do stałej jest to wariancja wszystkich obserwacji.[3] Jest zatem: $\frac{\mathrm{SS}_{\mathrm{Tot}}}{\sigma^2} \sim \chi^2(IJ-1)$.

Zróżnicowanie grup (**zmienność międzygrupową**) można wyrazić poprzez średnie grup $x_{i\bullet}$. Ponieważ zmienne X_{ij} są niezależne, więc niezależne są też zmienne $X_{i\bullet}$. Mamy zatem I niezależnych zmiennych losowych $X_{1\bullet}, \ldots, X_{I\bullet}$. Średnia $\bar{X}$ wszystkich obserwacji jest równocześnie średnią wziętą ze średnich poszczególnych grup.[4] Traktując grupę jako "uogólnioną obserwację" stwierdzamy iż

[3] SS ≡ sum of squares.

[4] Wszystkie grupy mają tę samą liczbę obserwacji

wyrażenie SSA = $J \cdot \sum_{i=1}^{I} (x_{i\bullet} - \bar{x})^2$ - z dokładnością do stałej – ma rozkład $\chi^2(I-1)$. Do rozpatrzenia pozostaje drugi składnik zmienności, wielkość SSE = $\sum_{i,j} (x_{ij} - x_{i\bullet})^2$, nazywana **zmiennością wewnątrzgrupową**.

Twierdzenie 2.

$$SS_{\text{Tot}} = \text{SSA} + \text{SSE}. \tag{4}$$

KOMENTARZE:

- Teza twierdzenia to: wariancja całkowita dzieli się na sumę wariancji pomiędzy grupami i wariancji wewnątrz grup.
- Jeżeli większość wariancji znajduje się wewnątrz grup, to skłonni jesteśmy uznać, że średnie grup są takie same (albo zbliżone do siebie).
- Na odwrót: jeżeli wariancja między grupami przeważa nad wariancją wewnątrz grup to można sądzić, że średnie grup różnią się.
- W podsumowaniu: na podstawie wariancji (a raczej jej podziału na dwa składniki) można wyciągnąć wnioski o średnich w obrębie grup.

Dowód.

$$\begin{aligned} SS_{\text{Tot}} &= \sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} (x_{ij} - x_{i.\bullet} + x_{i\bullet} - \bar{x})^2 = \\ &= J \cdot \sum_{i} (x_{i\bullet} - \bar{x})^2 + \sum_{i,j} (x_{ij} - x_{i.\bullet})^2 + 2 \cdot \sum_{i,j} (x_{ij} - x_{i.\bullet}) \cdot (x_{i\bullet} - \bar{x}). \end{aligned}$$

Trzeci składnik w ostatniej równości można przekształcić do postaci

$$\begin{aligned} &\sum_{i,j} (x_{ij} - x_{i.\bullet}) \cdot (x_{i\bullet} - \bar{x}) = \sum_{i} (x_{i\bullet} - \bar{x}) \sum_{j} \cdot (x_{ij} - x_{i.\bullet}) = \\ &= (x_{i\bullet} - \bar{x}) \cdot (n \cdot x_{i\bullet} - n \cdot x_{i\bullet}) = 0. \end{aligned}$$

Stąd wynika już, że

$$SS_{\text{Tot}} = \sum_{i,j} (x_{ij} - \bar{x})^2 = J \cdot (x_{i\bullet j} - \bar{x})^2 \sum_{i,j} + (x_{ij} - x_{i.\bullet})^2 = \text{SSA} \ + \ \text{SSE}. \tag{5}$$

□

2-czynnikowa ANOVA

Załóżmy, że dane są niezależne obserwacje zmiennej losowej $X \sim N(\mu, \sigma^2)$. Wyniki obserwacji grupujemy względem cechy jakościowej (czynnika) A oraz cechy jakościowej B, wyróżniamy odpowiednio I oraz J grup. Dla każdej kombinacji grup mamy jedną obserwację.[5] Symbolem x_{ij} oznaczamy j-tą obserwację dla której cecha A przyjęła i-tą wartość, natomiast cecha B wartość j-tą ($i = 1, \ldots, I$; $j = 1, \ldots, J$).

Grupa	1	2	...	J	Średnie
1	x_{11}	x_{12}	...	x_{1J}	$x_{1\bullet}$
2	x_{11}	x_{12}	...	x_{1J}	$x_{2\bullet}$
⋮	⋮	⋮		⋮	⋮
I	x_{I1}	x_{I2}	...	x_{IJ}	$x_{I\bullet}$
Średnie	$x_{\bullet 1}$	$x_{\bullet 2}$	...	$x_{\bullet J}$	

[5] Mówimy wówczas o 2-czynnikowej analizie ANOVA bez powtórzeń.

Symbole $x_{i\bullet}, x_{\bullet j}$ oznaczają – odpowiednio – średnią wartość i-tej grupy czynnika A oraz średnią wartość j-tej grupy czynnika B. Symbol $\bar{x}$ oznacza średnią wszystkich obserwacji. Niech ponadto

$$\begin{aligned} \mathrm{SSTot} &= \sum_{ij} (x_{ij} - \bar{x})^2, & \mathrm{SSA} &= J \cdot \sum_i (x_{i\bullet} - \bar{x})^2 \\ \mathrm{SSB} &= I \cdot \sum_j (x_{\bullet j} - \bar{x})^2, & \mathrm{SSE} &= \sum_{ij} (x_{ij} - x_{i\bullet} - x_{\bullet j} + \bar{x})^2. \end{aligned} \tag{6}$$

Twierdzenie 3.

$$\mathrm{SSTot} = \mathrm{SSA} + \mathrm{SSB} + \mathrm{SSE}. \tag{7}$$

Dowód. Zauważmy, że $\mathrm{SSTot} = \sum_{ij} (x_{ij} - \bar{x})^2 = \sum_{ij} \left(\underbrace{x_{ij} - x_{i\bullet} - x_{\bullet j} + \bar{x}}_{(a)} + \overbrace{x_{i\bullet} - \bar{x}}^{(b)} + \underbrace{x_{\bullet j} - \bar{x}}_{(c)} \right)^2$.

Zauważmy, że sumowanie kwadratów wyrażeń oznaczonych jako $(a), (b), (c)$ daje składniki SSA, SSB, SSE prawej strony równania (7). Pozostaje zatem do wykazania, że sumowanie iloczynów $(a) \cdot (b), (a) \cdot (c), (b) \cdot (c)$ daje w wyniku 0.

$(b) \cdot (c) = \sum_{ij} (x_{i\bullet} - \bar{x}) \cdot (x_{\bullet j} - \bar{x}) = \sum_i (x_{i\bullet} - \bar{x}) \cdot \sum_j (x_{\bullet j} - \bar{x}) = 0$, bo $\sum_i (x_{i\bullet} - \bar{x}) = I \cdot \bar{x} - I \cdot \bar{x} = 0$.

$(a) \cdot (b) = \sum_{ij} (x_{i\bullet} - \bar{x}) \cdot (x_{ij} - x_{i\bullet} - x_{\bullet j} + \bar{x}) = \sum_i (x_{i\bullet} - \bar{x}) \cdot \sum_j (x_{ij} - x_{i\bullet} - x_{\bullet j} + \bar{x}) = (*)$

Dla ustalonego i rozpatrzmy wewnętrzne sumowanie po j daje $\sum_j (x_{ij} - x_{i\bullet}) = 0$. Stąd

$$(*) = \sum_i (x_{i\bullet} - \bar{x}) \cdot \sum_j (\bar{x} - x_{\bullet j}) = 0.$$

Dowód dla sumowania iloczynów postaci $(a) \cdot (c)$ jest praktycznie taki sam, wystarczy zamienić miejscami indeksy i, j. $\square$

Generator liczb losowych z rozkładu $\mathbf{N}(0,1)$

Przykład:

Załóżmy, że mamy do dyspozycji generator liczb losowych z rozkładu $U[0,1]$ i wylosowaliśmy dwie wartości u_1, u_2. Dwuwymiarowa zmienna losowa (U_1, U_2) ma zatem rozkład o gęstości $f_{U_1,U_2}(u_1,u_2) = 1$ dla $((u_1,u_2) \in [0,1] \times [0,1]$. Rozważmy nowe zmienne $Y_1 = -2\ln U_1,\ Y_2 = 2\pi U_2$. Oczywiście $Y_1 \in [0, \infty)$ oraz $Y_2 \in [0, 2\pi)$. Interpretując Y_1, Y_2 jako współrzędne biegunowe punktu na płaszczyźnie można powiedzieć iż losujemy kwadrat promienia i argument punktu. Wyznaczmy gęstość $f_{Y_1,Y_2}(y_1,y_2)$ zmiennej (Y_1,Y_2).

$$\begin{cases} U_1 = \exp\left(-\dfrac{Y_1}{2}\right) \\ U_2 = \dfrac{Y_2}{2\pi} \end{cases}, \text{ skąd } \operatorname{abs}(J) = \operatorname{abs}\left(\begin{vmatrix} -\dfrac{1}{2}\exp\left(-\dfrac{Y_1}{2}\right) & 0 \\ 0 & \dfrac{1}{2\pi} \end{vmatrix} \right) = \frac{1}{4 \cdot \pi} \exp\left(-\frac{Y_1}{2}\right). \tag{8}$$

W powyższym wzorze chcemy policzyć wartość bezwzględną z wyznacznika Jacobianu. Niestety, obydwie operacje (wartość bezwzględna i wyznacznik) oznaczane są często tym samym znakiem $|\ |$. Wskutek tego: $\operatorname{abs}(\det(A)) \equiv ||A||$ – co z kolei mogłoby sugerować, że mówimy o **normie** macierzy A.
Dla gęstości $f_{Y_1,Y_2}(y_1,y_2)$ mamy zatem wzór

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{4 \cdot \pi} \exp\left(-\frac{Y_1}{2}\right). \tag{9}$$

Od współrzędnych biegunowych (Y_1, Y_2) przejdźmy teraz do współrzędnych kartezjańskich (X_1, X_2), to znaczy

$$\begin{cases} X_1 = \sqrt{Y_1}\,\cos Y_2 \\ X_2 = \sqrt{Y_1}\,\sin Y_2 \end{cases}, \text{ i stąd } J = \begin{vmatrix} \dfrac{\cos Y_2}{2\sqrt{Y_1}} & -\sqrt{Y_1}\sin Y_2 \\ \dfrac{\sin Y_2}{2\sqrt{Y_1}} & \sqrt{Y_1}\cos Y_2 \end{vmatrix} = \frac{1}{2}. \tag{10}$$

Kończymy przekształcenia dwiema uwagami:

1. Wyznaczony powyżej Jacobian należy ODWRÓCIĆ. Zwyczajowo liczymy Jacobian "starych" zmiennych względem "nowych". Tutaj: wygodniej jest wyznaczyć odwrotność Jacobianu "nowych" zmiennych względem "starych" zmiennych.
2. Korzystamy również z zależności: $Y_1 = X_1^2 + X_2^2$ [a].

Wynik końcowy ≡ gęstość $f_{X_1,X_2}(x_1, x_2)$ ma postać:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_1^2}{2}\right) \times \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_2^2}{2}\right), \tag{11}$$

co oznacza, że zmienne X_1, X_2 są niezależne i mają rozkład $\mathrm{N}(0,1)$ każda.

[a](wzór (10))

[Popularne|Ulubione] wzory i rozkłady

1. Załóżmy, że zmienne losowe X, Y są niezależne i podlegają rozkładom $X \sim \chi^2(n)$, $Y \sim \chi^2(k)$. Wówczas zmienna losowa $Z = X + Y$ podlega rozkładowi $Z \sim \chi^2(n+k)$.
2. Załóżmy, że zmienna X podlega rozkładowi $\mathrm{N}(\mu, \sigma^2)$. Niech dodatkowo $Y = \dfrac{X-\mu}{\sigma}$. Zachodzi FAKT: $X \sim \mathrm{N}(\mu,\sigma^2) \iff Y \sim \mathrm{N}(0,1)$.
3. $\mathrm{Gamma}\,(1/2, n/2) \equiv \chi^2(n)$.
4. Załóżmy, że zmienne $X_1, \dots, X_n$ są niezależne i podlegają rozkładowi $\mathrm{N}(\mu,\sigma^2)$ każda. Wówczas zmienna $Z = \sum_{k=1}^{n}\left(\dfrac{X_k-\mu}{\sigma}\right)^2$ ma rozkład $\chi^2(n)$.
5. Niezależne zmienne X, Y mają rozkłady $X \sim \chi^2(k)$, $Y \sim \chi^2(l)$ odpowiednio. Mówimy, że zmienna $F(k,l) = \dfrac{X}{Y} \cdot \dfrac{l}{k}$ ma rozkład F-Fishera z (k,l) stopniami swobody.
6. Niezależne zmienne X, Y mają rozkłady $X \sim \mathrm{N}(0,1)$, $Y \sim \chi^2(k)$ odpowiednio. Mówimy, że zmienna $t(k) = \dfrac{X}{\sqrt{Y/k}}$ ma rozkład t-Studenta z k stopniami swobody.
7. Intuicja: iloraz niezależnych i normalizowanych rozkładów χ^2 to rozkład F-Fishera zaś iloraz standardowego rozkładu normalnego i pierwiastka normalizowanego rozkładu χ^2 to rozkład t-Studenta.
8. Załóżmy, że zmienne $X_1, \dots, X_n$ są niezależne i podlegają rozkładowi $\mathrm{N}(\mu,\sigma^2)$ każda. Niech dodatkowo $S_\mu^2 = \dfrac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2$. Wówczas $\dfrac{nS_\mu^2}{\sigma^2} \sim \chi^2(n)$.
9. Załóżmy, że zmienne $X_1, \dots, X_n$ są niezależne i podlegają rozkładowi $\mathrm{N}(\mu,\sigma^2)$ każda. Niech dodatkowo $S^2 = \dfrac{1}{n}\sum_{k=1}^{n}\left(X_k-\bar{X}\right)^2$. Wówczas $\dfrac{nS^2}{\sigma^2} \sim \chi^2(n-1)$.
10. ...tysiąc i jeden wzór (jak w orientalnych baśniach).

Z poważaniem, Witold Karczewski