Własności EX, D^2X i DX przy przekształceniach liniowych

Zwiększenie wartości zmiennej losowej o wartość stałą: $Y=X+a$

$EY=EX+a$

$D^2Y=D^2X$

Przemnożenie wartości zmiennej losowej przez wartość stałą: $Y=a*X$

$EY=a*EX$

$D^2Y=a^2*D^2X$

Przekształcenie **$Y=a*X+b$**

$EY=a*EX+b$

$D^2Y=a^2*D^2X$

Standaryzacja zmiennej losowej

$Z={}^{X-EX}/_{DX}$

$EZ=0$

$D^2Z=1$

Podstawowe rozkłady zmiennej losowej skokowej

1. Rozkład dwupunktowy (zero-jedynkowy)

Funkcja rozkładu prawdopodobieństwa jest następująca:

$$P(X=1)=p \qquad\qquad P(X=0)=1-p$$

Wartość oczekiwana i wariancja w tym rozkładzie wynoszą:

$$EX = p, \qquad\qquad D^2X = p(1-p) = pq$$

2. Rozkład dwumianowy (Bernoulliego)

Rozkład ten otrzymujemy, gdy mamy serię n powtórzeń rozkładu zero-jedynkowego. Zmienna losowa realizuje *n+1* wartości z funkcją rozkładu prawdopodobieństwa postaci:

$$P(X=k)=\binom{n}{k}p^k q^{n-k}$$, gdzie $k = 0, 1, 2, \ldots, n$

Parametry tej zmiennej są następujące:

$$EX = np, \quad D^2X = npq, \quad DX = \sqrt{npq}$$

Przykłady:

Liczba skiełkowanych nasion, liczba prosiąt w miocie, liczba owoców uszkodzonych w czasie transportu.
Przykład liczbowy: Żarówka oświetlająca firmowy parking przepali się w nocy (i będzie wymieniona rano) z prawdopodobieństwem 1,5%.

Podać spodziewaną liczbę żarówek potrzebnych na 2 lata (730 dób) do oświetlenia parkingu. Ile wynosi odchylenie standardowe?
$EX=n*p=730*1,5\%\approx 11$
$DX=(n*p*q)^{0,5}=(730*1,5\%*98,5\%)^{0,5}\approx 3,3$

Jakie jest prawdopodobieństwo, że zużyjemy 10 żarówek? Jakie, że wystarczy 12 żarówek?
$P_{n=730}(X=10) \approx 12\%$
$P_{n=730}(X\leq 12) = F_X(12) \approx 70\%$
$P_{n=730}(X<12)=P_{n=730}(X\leq 12)–P_{n=730}(X=12)\approx 59\%$

Ile żarówek należy nabyć, aby wystarczyły na 2 lata z prawdopodobieństwem 90%? a 95%?
$P_{n=730}(X\leq x_0)=90\% \Rightarrow x_0=15$
$P_{n=730}(X\leq x_0)=95\% \Rightarrow x_0=17$

3.Rozkład Poissona (rozkład rzadkich zdarzeń)

Zmienna losowa X przyjmująca wartości k = 0, 1, 2, …n,… (liczba zdarzeń) ma rozkład Poissona jeśli jej funkcja rozkładu prawdopodobieństwa ma postać:

$$P\ (x=k)=\frac{l^k}{k!}e^{-l}$$

gdzie l to prawdopodobieństwo pojedynczego zdarzenia w badanej jednostce czasu. Parametry tej zmiennej wyrażają się wzorami:

$$EX = l, \quad D^2X = l, \quad DX = \sqrt{l}$$

Jest to rozkład graniczny dla rozkładu Bernoulliego, gdy n–>∞ oraz n*p–> l (stałe).

Zastosowanie:

- w przypadkach zjawisk „rzadkich w czasie i przestrzeni”,
- w przybliżeniach rozkładu dwumianowego, jeśli:
  $n \to \infty$ (n>100), $p < 0.1$, $n \cdot p = \lambda$.

Przykłady:

Liczba (ilość) wad na metrze kwadratowym wyprodukowanego materiału, liczba bakterii w jednostce objętości (mleka), liczba awarii różnych urządzeń /wypadków przy pracy w jednostce czasu.

Przykład liczbowy: średnia liczba wymienianych żarówek, oświetlających miejsce pracy w przeciągu 2 lat, wynosi 11 (oznacza to, że na ok. 66 dób wystarcza średnio ok. 1 żarówka, i będzie wymieniona natychmiast). Te 11 żarówek wystarczy średnio na 2*365 dni (n=720, p=11/730=1,5%), 2*365*24=17520 godzin (n=17520 p=11/17520) albo 1 051 200 minut pracy (n=1051200, $p \approx 10^{-5}$).
Dla dwóch lat $l=11$, ale po przeliczeniu dla roku $l=11/2=5,5$ a dla 66 dób $l \approx 1$.

Podać spodziewaną liczbę żarówek potrzebnych na 2 lata. Ile wynosi odchylenie standardowe?
$EX=l=11$ $DX=l^{0,5}=11^{0,5}\approx 3,3$

Jakie jest prawdopodobieństwo, że zużyjemy 10 żarówek? Jakie, że wystarczy 12 żarówek?
$P_{\lambda=11}(X=10)\approx 12\%$ $P_{\lambda=11}(X\leq 12)\approx 69\%$

Podstawowe rozkłady zmiennej losowej ciągłej

1. Rozkład wykładniczy

Zmienna losowa X ma rozkład wykładniczy o parametrze $\lambda>0$, gdy prawdopodobieństwo zmiany stanu (zajścia zdarzenia) w ustalonym przedziale czasu jest stałe (rozkład „bez pamięci” – nieważne, ile czasu już obserwujemy, prawdopodobieństwo zmiany zajścia zdarzenia w najbliższym okresie jest ciągle takie samo). Gęstość funkcji rozkładu i funkcja dystrybuanty dane są wzorami:

$f(x)=\lambda e^{-\lambda x}$ $F(x)=1-e^{-\lambda x}$

a wartość oczekiwana i wariancja

$EX=\lambda^{-1}$ $D^2X= \lambda^{-2}$

przykłady:

Rozkład promieniotwórczy pierwiastków

Przykład liczbowy: Prawdopodobieństwo przepalenia żarówki w ciągu doby wynosi 1,5% (przejdzie ze stanu „sprawna” do „przepalona”).

Jaka jest średnia liczba dób działania żarówki?
$EX=\lambda^{-1}=1/0{,}015\approx 67$

Jakie jest prawdopodobieństwo, że właśnie wkręcona żarówka będzie świecić krócej niż 50 dni? Dłużej niż 100 dni?
$P_{\lambda=0,015}(X\leq 50)\approx 53\%$ $P_{\lambda=0,015}(X\geq 100)\approx 22\%$

2. Rozkład normalny – najważniejszy rozkład zmiennej losowej ciągłej.

Zmienna losowa X ma rozkład normalny o wartości średniej *m* i wariancji σ^2, co zapisujemy **X~ N(m, $\boldsymbol{\sigma^2}$)**, jeżeli jej funkcja gęstości rozkładu f(x) wyraża się wzorem:

$$f(x) = \frac{1}{s\sqrt{2p}} e^{-\frac{(x-m)^2}{2s^2}}$$

, x, m $\in$ R, a $\sigma > 0$

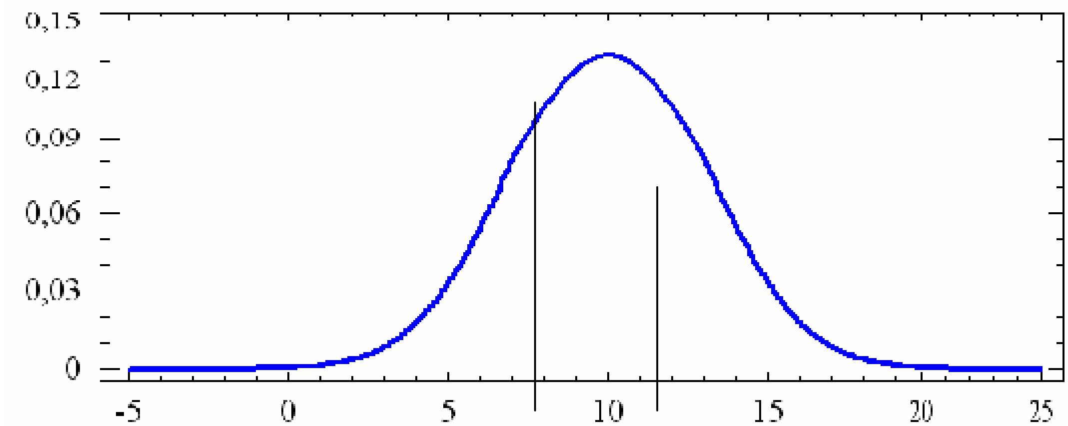
gdzie m = EX, a σ = DX.

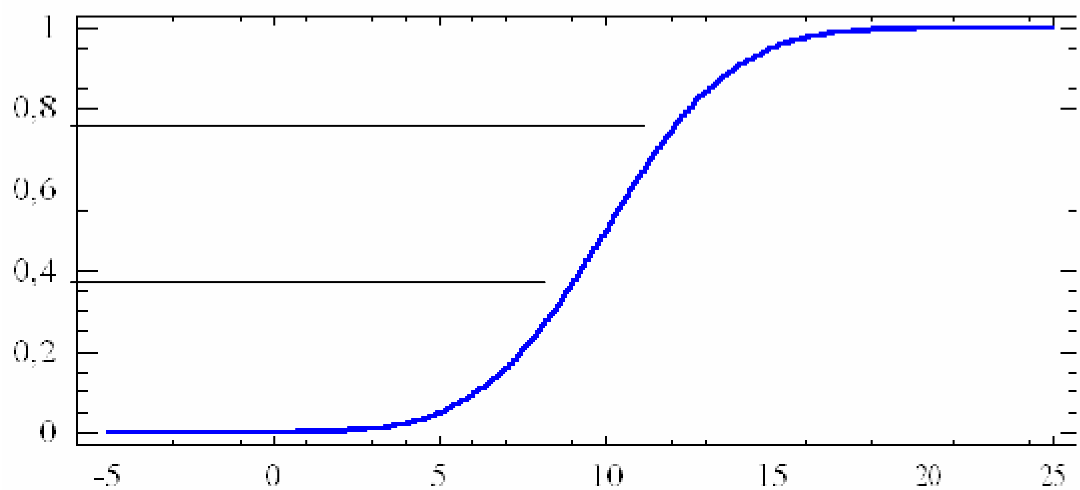
Zastosowanie – wszędzie tam, gdzie badana wielkość (zmienna losowa) jest sumą wielu zmiennych losowych, z których żadna nie dominuje.

Przykłady: zawartość białka w mięsie, ciężar ciała, błędy pomiarowe.

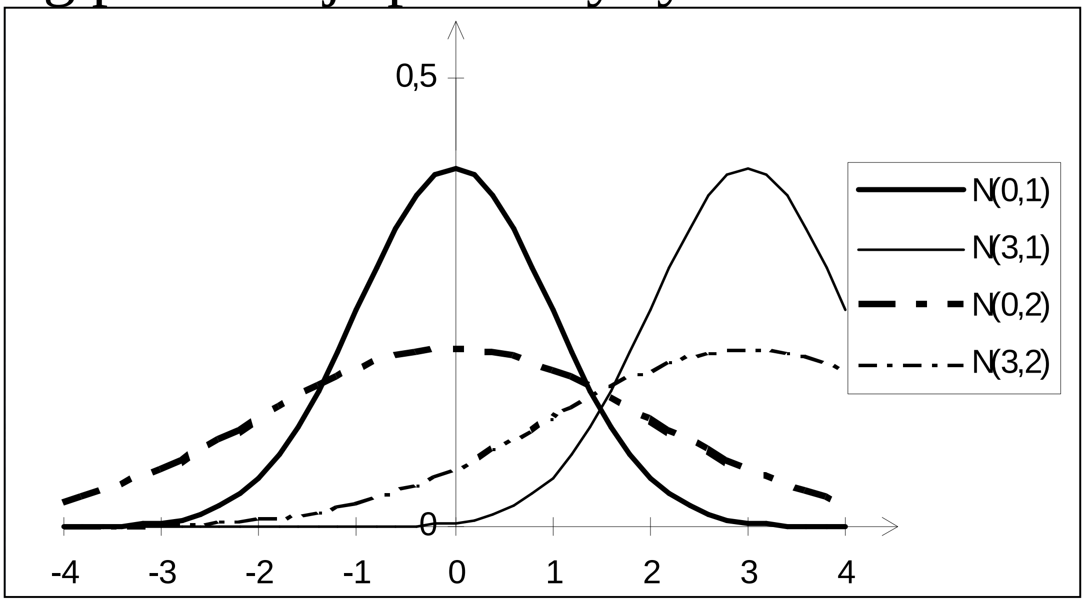
Niech X~ N(10, 3^2):

Wykresy przedstawiają funkcję gęstości rozkładu prawdopodobieństwa dla zmiennej losowej X oraz dystrybuantę tej zmiennej losowej





Wpływ parametrów rozkładu na wykres f.g.p. ilustruje poniższy rysunek:



Dystrybuanta rozkładu normalnego ma postać:

$$F(x_0) = \frac{1}{s\sqrt{2p}} \int_{-\infty}^{x_0} e^{-\frac{(x-m)^2}{2s^2}} dx$$

Dystrybuanta charakteryzuje się następującymi własnościami:

$$P(X \le b) = F(b)$$

$$P(X > a) = 1 - P(X \le a) = 1 - F(a)$$

$$P(a < X \le b) = F(b) - F(a)$$

Przy czym dla rozkładów ciągłych (więc i normalnego) $P(X = a) = 0$, a dla rozkładu normalnego standaryzowanego tj. N(0,1) $F(-z) = 1 - F(z)$

Standaryzacja zmiennej. Jeśli zmienna losowa X ma rozkład $N(m, s^2)$, to **zmienna standaryzowana** $Z = \frac{X - m}{s}$ ma rozkład N(0,1). Na tej podstawie można wyznaczyć prawdopodobieństwa przedziałów:

$$P(a < X \leq b) = F(b) - F(a) = F\left(\frac{b-m}{s}\right) - F\left(\frac{a-m}{s}\right)$$

Wartości $F\left(\frac{b-m}{s}\right)$ i $F\left(\frac{a-m}{s}\right)$ odczytuje się z **tablic** dystrybuanty standardowego rozkładu normalnego.

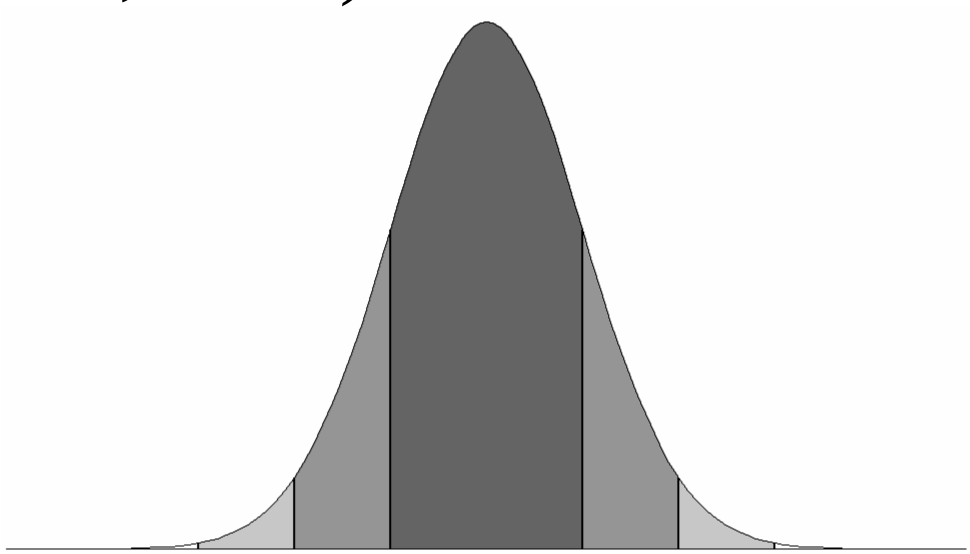
Funkcja gęstości rozkładu normalnego podlega **regule „trzech sigm"**, którą stosuje się np. w kontroli jakości czy usuwaniu obserwacji odstających (niewiarygodnych).

Reguła „trzech sigm" - jeżeli zmienna losowa ma rozkład normalny to:

- $P(|X-m| < \sigma) = 0.683$
- $P(|X-m| < 2\sigma) = 0.955$
- $P(|X-m| < 3\sigma) = 0.997$

Oznacza to, że dla realizacji zmiennej losowej o dowolnym rozkładzie normalnym

- około 70% obserwacji mieści się w granicach jednego odchylenia standardowego wokół średniej, w przedziale $(m - \sigma; m + \sigma)$,
- 95% obserwacji mieści się w granicach dwóch odchyleń standardowych, w przedziale $(m - 2\sigma; m + 2\sigma)$,
- 99,7% w granicach trzech odchyleń standardowych, w przedziale $(m - 3\sigma; m + 3\sigma)$.



Reguła trzech sigm jest wykorzystywana w badaniach statystycznych do eliminacji obserwacji niewiarygodnych. Obserwacje niewiarygodne to obserwacje, których wartość różni się od średniej o więcej niż trzy odchylenia standardowe. Przyjmuje się, iż zmienne, które odbiegają tak znacznie od średniej mogą być skutkiem błędu pomiaru.

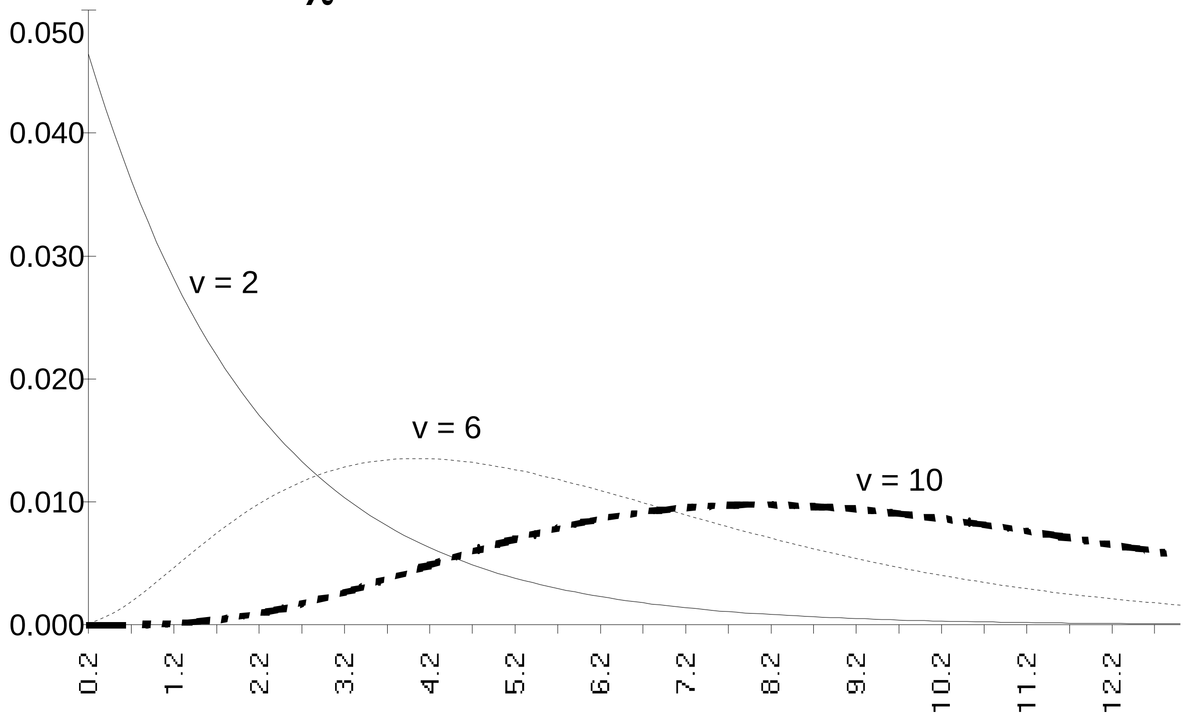
3. Rozkłady związane z rozkładem normalnym

Bardzo ważną rolę w statystyce odgrywają trzy rozkłady zmiennych losowych bazujące na zmiennych o standardowych rozkładach normalnych (N(0,1)). Są to rozkłady:

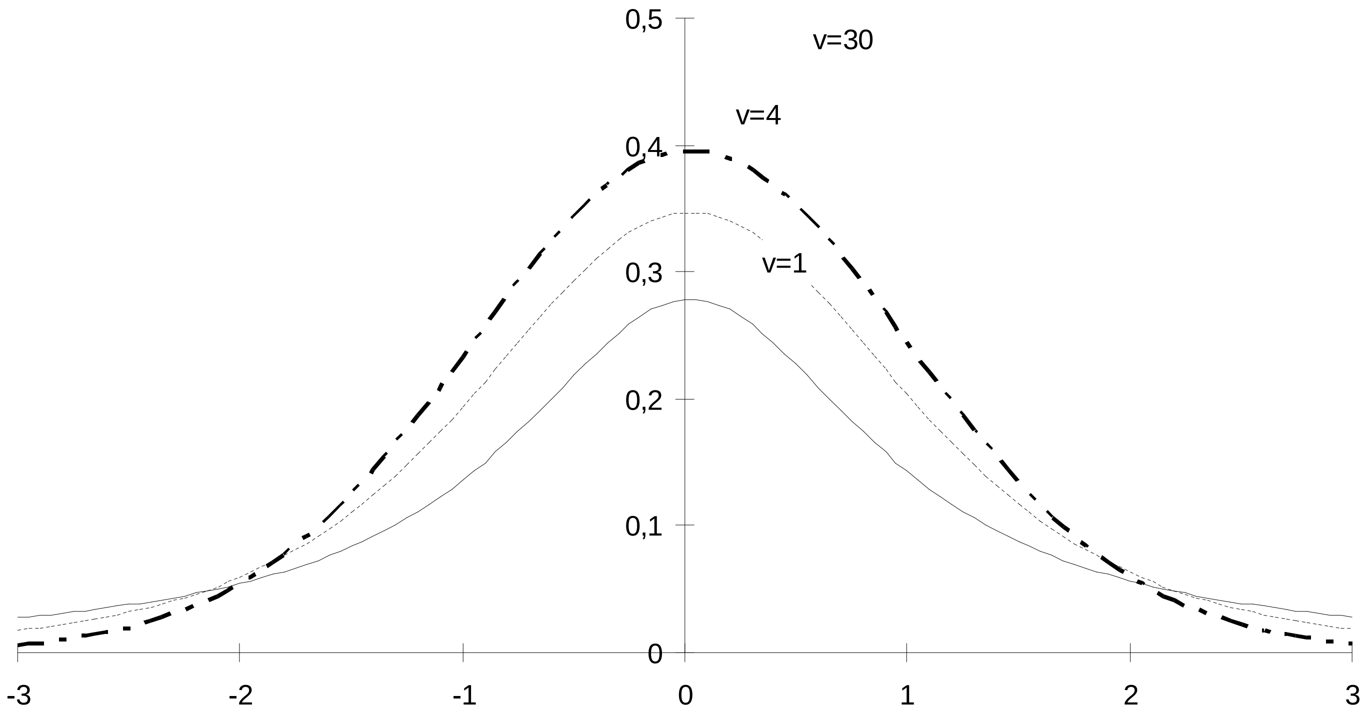
- χ^2 – Persona (chi-kwadrat),
- t – Studenta,
- F – Fishera – Snedecora.

Poniżej podane są wykresy funkcji gęstości prawdopodobieństwa trzech wymienionych rozkładów (dla wybranych charakterystyk *v* i *u*).

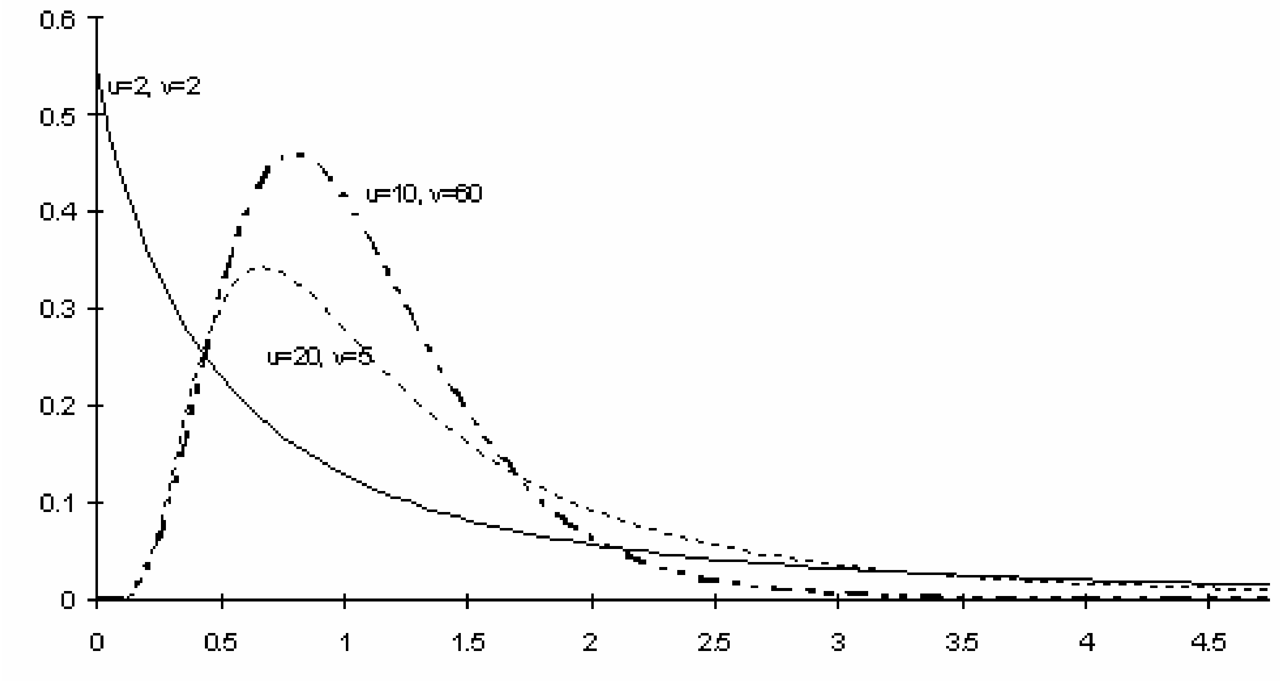
1. Rozkład χ2 – Persona



2. Rozkład t – Studenta



3. Rozkład F – Fishera – Snedecora



Funkcje gęstości omawianych trzech rozkładów oparte są na funkcji gamma $\Gamma(v)$.

Ze statystykami opartymi na tych rozkładach związane są podstawowe działy statystyki matematycznej, takie jak: przedziały ufności, weryfikacja hipotez, analiza wariancji czy analiza regresji.

Wartości krytyczne wymienionych rozkładów, potrzebne we wnioskowaniu statystycznym, są zawarte w tablicach statystycznych.