

**Grzegorz Kończak**

Uniwersytet Ekonomiczny w Katowicach

# **SPRAWDZANIE JEDNORODNOŚCI JAKOŚCI MATERIAŁÓW NIEKSZTAŁTNYCH Z WYKORZYSTANIEM ROZKŁADÓW WARTOŚCI EKSTREMALNYCH**

## **Wprowadzenie**

Klasyczne metody statystyczne bardzo często odwołują się do założenia normalnego rozkładu analizowanych charakterystyk oraz niezależności pomiarów. Tak szczególne znaczenie rozkładu normalnego jest związane m.in. z centralnym twierdzeniem granicznym. Przy stosowaniu metod statystycznych w kontroli jakości zwykle zakłada się dodatkowo, że przedstawiany do kontroli materiał jest jednorodny.

Teoria statystycznej kontroli jakości przedstawia wiele różnych metod sprawdzania jakości produktów. W zdecydowanej większości dotyczą one obiektów policzalnych, występujących w sztukach, opakowaniach itp. W artykule podjęto analizę zagadnienia badania jakości towarów materiałów niekształtnych.

Jako materiał niekształtny będzie określany materiał, z którego nie sposób bezpośrednio wyodrębnić poszczególne elementy, opakowania itp. Przykładem materiału niekształtnego może być np. zwał węgla kamiennego w punkcie sprzedaży lub ziarno w silosach. W takim rozumieniu materiał niekształtny może mieć postać regularną np. sześciennego bloku. W artykule przedstawiono propozycję metody pozwalającej na weryfikację hipotezy głoszącej, że sprawdzany materiał jest jednorodny ze względu na badaną charakterystykę.

## **1. Rozkłady wartości ekstremalnych**

Rozkładem granicznym dla wielu spotykanych w praktyce rozkładów jest rozkład normalny. Tak jest np. z rozkładem średniej arytmetycznej z próby prostej, jeśli tylko wartość oczekiwana i wariancja badanej zmiennej są skończone. Wraz ze wzrostem liczebności próbki  $n$  rozkład średniej z próby prostej zmierza do rozkładu

normalnego. Często w badaniach kontroli jakości nie interesuje nas jednak przeciętny poziom jakości, lecz występowanie materiałów o szczególnie niskiej (lub wysokiej) wartości sprawdzanej charakterystyki.

### 1.1. Dokładny rozkład wartości maksymalnych

Niech  $(X_1, X_2, \dots, X_n)$  będzie  $n$ -elementową próbą losową pobraną niezależnie z populacji o dystrybuancie  $F(x)$ . Oznaczając przez  $X_M$  zmienną losową określoną następująco:

$$X_M = \max(X_1, X_2, \dots, X_n) \quad (1)$$

dystrybuantę  $G(x)$  tej zmiennej losowej można wyznaczyć następująco:

$$\begin{aligned} G(x) &= P(X_M < x) = P(X_1 < x \wedge X_2 < x \wedge \dots \wedge X_n < x) = \\ &= P(X_1 < x) \cdot P(X_2 < x) \cdot \dots \cdot P(X_n < x) = F^n(x) \end{aligned}$$

W praktyce zwykle nie ma możliwości odwołania się do rozkładu dokładnego. Tak będzie np. w sytuacji, gdy nie jest znany rozkład analizowanej zmiennej lub jego złożona postać nie daje możliwości wyznaczenia rozkładu dokładnego na podstawie wzoru (1). W takich przypadkach niezbędne jest odwołanie się do rozkładu granicznego. Postać graniczna rozkładu wartości ekstremalnych (maksima lub minima) z  $k$  próbek o liczebności  $n$  może przyjmować wyłącznie jedną z trzech form: rozkładu Gumbela, rozkładu Frecheta lub rozkładu Weibulla (por. Castillo et al., 2005).

### 1.2. Rozkład Gumbela

Rozkład Gumbela występuje bardzo często w praktyce, a w szczególności podczas obserwacji np. wartości maksymalnych lub minimalnych różnych pomiarów. Funkcja gęstości zmiennej losowej  $X$  o rozkładzie Gumbela dla wartości minimalnych z parametrami  $\lambda$  i  $\delta$  jest zadana wzorem:

$$f(x) = \frac{1}{\delta} e^{-\frac{(x-\lambda)}{\delta}} e^{-\frac{(x-\lambda)}{\delta}} \quad \text{dla } x \in R \quad (2)$$

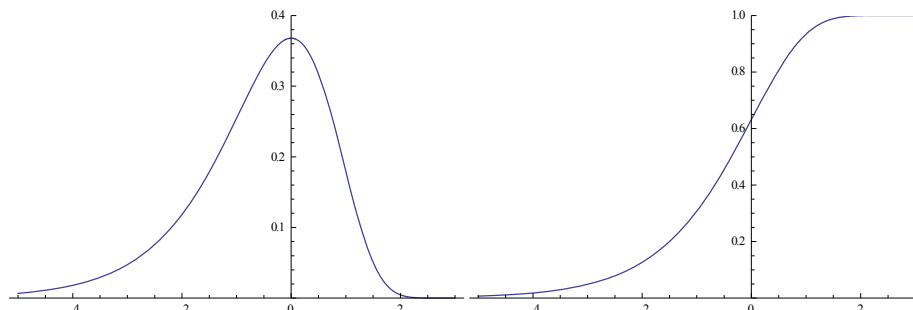
Parametry  $\lambda$  i  $\delta$  to odpowiednio dominanta oraz parametr rozproszenia. Dystrybuanta tej zmiennej losowej wyraża się następującym wzorem:

$$F(x) = 1 - e^{-e^{-\frac{(x-\lambda)}{\delta}}} \quad \text{dla } x \in R \quad (3)$$

Wartość oczekiwana  $\mu$  i wariancja  $\sigma^2$  zmiennej losowej  $X$  są następujące:

$$\mu = \lambda - 0,57772\delta \quad \text{oraz} \quad \sigma^2 = \frac{\pi^2\delta^2}{6} \quad (4)$$

Funkcję gęstości i dystrybuantę zmiennej losowej o rozkładzie Gumbela dla wartości minimalnych z parametrami  $\lambda = 0$  i  $\delta = 1$  przedstawia rys. 1.



Rys. 1. Gęstość prawdopodobieństwa i dystrybuanta zmiennej losowej o rozkładzie Gumbela wartości minimalnych z parametrami  $\lambda = 0$  i  $\delta = 1$

Funkcja gęstości zmiennej losowej  $X$  o rozkładzie Gumbela dla wartości maksymalnych z parametrami  $\lambda$  i  $\delta$  jest zadana wzorem:

$$f(x) = \frac{1}{\delta} e^{-\frac{(x-\lambda)}{\delta}} e^{-\frac{(x-\lambda)}{\delta}} \quad \text{dla } x \in R \quad (5)$$

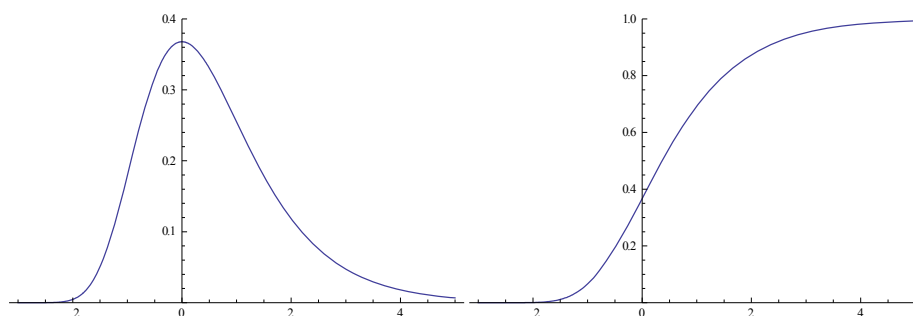
Dystrybuanta tej zmiennej losowej wyraża się następującym wzorem:

$$F(x) = e^{-e^{-\frac{(x-\lambda)}{\delta}}} \quad \text{dla } x \in R \quad (6)$$

Wartość oczekiwana i wariancja zmiennej losowej  $X$  są następujące:

$$\mu = \lambda + 0,57772\delta \quad \text{oraz} \quad \sigma^2 = \frac{\pi^2 \delta^2}{6} \quad (7)$$

Funkcję gęstości i dystrybuantę zmiennej losowej o rozkładzie Gumbela dla wartości maksymalnych z parametrami  $\lambda = 0$  i  $\delta = 1$  przedstawia rys. 2.



Rys. 2. Gęstość prawdopodobieństwa i dystrybuanta zmiennej losowej o rozkładzie Gumbela wartości maksymalnych z parametrami  $\lambda = 0$  i  $\delta = 1$

### 1.3. Rozkład Weibulla

Zmienna losowa  $X$  ma rozkład Weibulla z parametrami  $\lambda$ ,  $\delta$  i  $\beta$ , jeśli jej gęstość prawdopodobieństwa wyraża się następującym wzorem:

$$f(x) = \frac{\beta}{\delta} \left( \frac{x-\lambda}{\delta} \right)^{1-\beta} e^{-\left(\frac{x-\lambda}{\delta}\right)^\beta} \quad \text{dla } x > \lambda \quad (8)$$

Dystrybuantę tej zmiennej losowej przedstawia wzór:

$$F(x) = 1 - e^{-\left(\frac{x-\lambda}{\delta}\right)^\beta} \quad \text{dla } x \geq \lambda \quad (9)$$

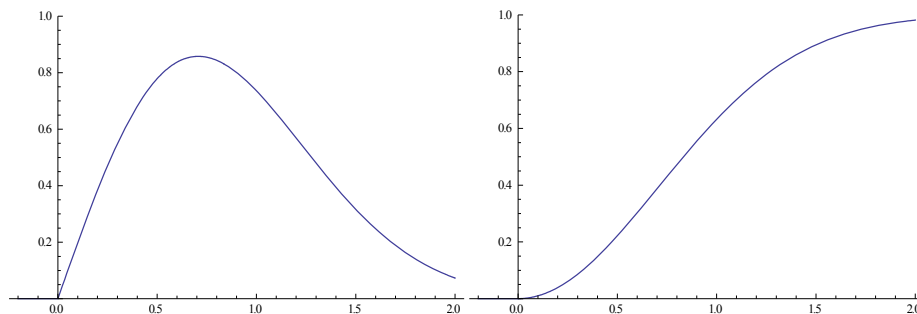
Wartość oczekiwana i wariancja zmiennej losowej  $X$  są następujące:

$$\mu = \lambda + \delta \Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{oraz} \quad \sigma^2 = \delta^2 \left[ \Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{2}{\beta}\right) \right] \quad (10)$$

gdzie  $\Gamma(x)$  jest funkcją Eulera, czyli

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Funkcję gęstości i dystrybuantę zmiennej losowej o rozkładzie Weibulla zaprezentowano na rys. 3.



Rys. 3. Gęstość prawdopodobieństwa i dystrybuanta zmiennej losowej o rozkładzie Weibulla z parametrami  $\lambda = 0$ ,  $\delta = 1$  oraz  $\beta = 2$

### 1.4. Rozkład Frcheta

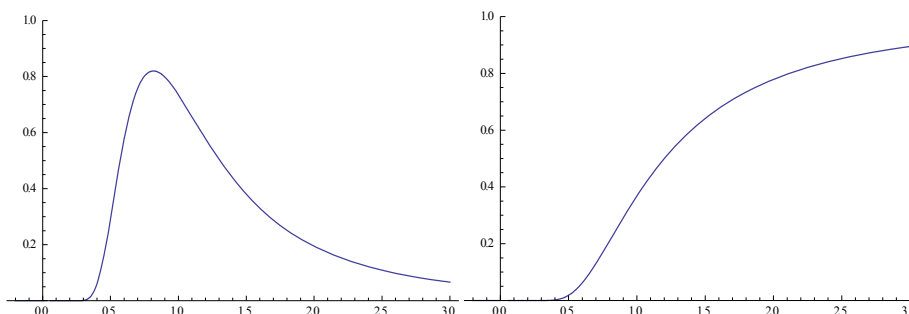
Zmienna losowa  $X$  ma rozkład Frcheta z parametrami  $\lambda$ ,  $\delta$  i  $\beta$ , jeśli jej gęstość prawdopodobieństwa jest zadana wzorem:

$$f(x) = \frac{\beta\delta}{(x-\lambda)^2} \left( \frac{\delta}{x-\lambda} \right)^{\beta-1} e^{-\left(\frac{\delta}{x-\lambda}\right)^\beta} \quad \text{dla } x > \lambda \quad (11)$$

Dystrybuanta tej zmiennej losowej wyraża się następującym wzorem:

$$F(x) = e^{-\left(\frac{\delta}{x-\lambda}\right)^\beta} \quad \text{dla } x > \lambda \quad (12)$$

Funkcję gęstości i dystrybuantę zmiennej losowej o rozkładzie Frecheta przedstawia rys. 4.



Rys. 4. Gęstość prawdopodobieństwa i dystrybuanta zmiennej losowej o rozkładzie Frecheta z parametrami  $\lambda = 0$ ,  $\delta = 1$  oraz  $\beta = 2$

W zależności od typu rozkładu badanej zmiennej  $X$  graniczny rozkład maksimum lub minimum będzie zawsze rozkładem Gumbela, Weibulla lub Frecheta. Odpowiednie zależności dla wartości maksymalnych i minimalnych przedstawia tab. 1.

Tabela 1

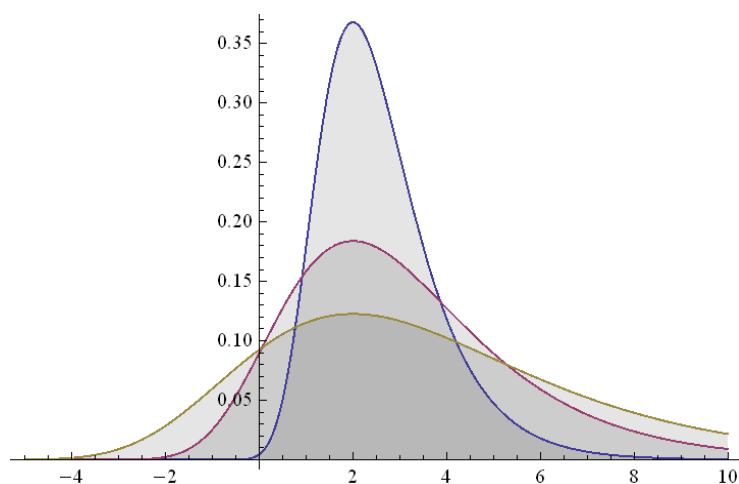
Typ rozkładu granicznego maksimum i minimum w zależności od rozkładu badanej zmiennej

Rozkład badanej zmiennej	Rozkład graniczny	
	Maksimum	Minimum
Normalny	Gumbela	Gumbela
Wykładniczy	Gumbela	Weibulla
Log-normalny	Gumbela	Gumbela
Gamma	Gumbela	Weibulla
Jednostajny	Weibulla	Weibulla
Cauchy'ego	Frecheta	Frecheta
Pareto	Frecheta	Weibulla

Źródło: Na podstawie (Castillo et al., 2005).

W dalszych rozważaniach będzie rozważany wyłącznie rozkład Gumbela dla wartości maksymalnych jako opisujący rozkład zmiennej losowej  $X_M = \max_{i=1,2,\dots,n} (X_1, X_2, \dots, X_n)$  gdzie  $(X_1, X_2, \dots, X_n)$  jest próbą prostą pochodzącą z rozkładu, dla którego granicznym rozkładem wartości maksymalnych

jest rozkład Gumbela. Gęstości rozkładu Gumbela dla wartości maksymalnych dla różnych wartości parametrów przedstawia rys. 5.



Rys. 5. Rozkład Gumbela – gęstości w zależności od wartości parametrów rozkładu

## 2. Sprawdzenie jednorodności jakości materiałów niekształtnych

Niech do oceny jakości przedstawiany będzie pewien materiał niekształtny. Obserwowane są wartości zmiennej losowej  $X$  charakteryzującej jakość badanego materiału. W poniższych rozważaniach przyjęto, że badana zmienna ma rozkład, dla którego granicznym rozkładem wartości maksymalnych jest rozkład Gumbela (np. rozkład normalny, logarytmiczno-normalny, wykładniczy, gamma). Przyjęto, że występuje swobodny dostęp do dowolnej części badanego materiału niekształtnego. Celem jest ocena jednorodności materiału ze względu na wyróżnioną badaną charakterystykę. Wiele procedur statystycznej kontroli jakości wymaga, aby badany materiał charakteryzował się jednorodnością. Problemem jednorodności danych dla mieszanek zajmował się m.in. R.L. Schaeffer (1971). Często w analizach jednorodności danych wykorzystywane są metody próbkowania, jak bootstrap (por. np. Boss i Brownie, 1989 oraz Kończak, 2006) czy testy permutacyjne (Good, 1994). W poniższych rozważaniach jednorodności danych wykorzystane zostaną własności rozkładów ekstremalnych.

Jeżeli materiał ze względu na badaną charakterystykę jest jednorodny, to wyznaczone wartości maksymalne z  $k$  próbek o jednakowych liczebnościach  $n$ , przy powyższych założeniach, będą obserwacjami z rozkładu Gumbela. Przebieg procesu sprawdzenia jednorodności badanego materiału jest następujący:

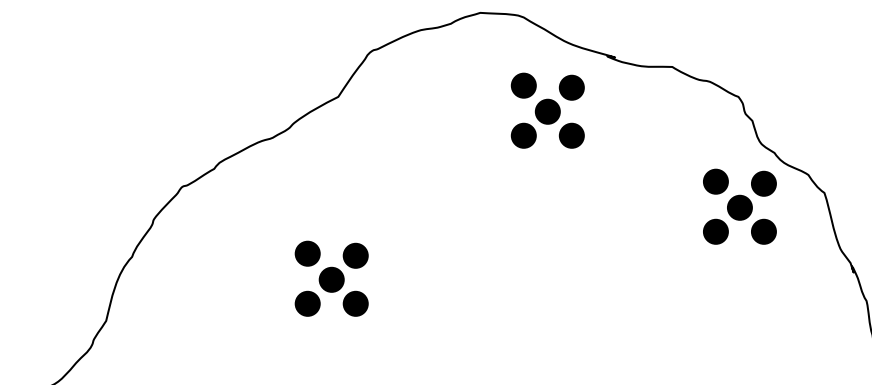
1. Wyznaczane jest losowo  $k$  punktów\* (obszarów) w materiale niekształtnym.
2. Z każdego otoczenia wyróżnionych punktów pobierana jest próbka o liczebności  $n$  (pobieranie próbek o liczebności  $n = 5$  z  $k = 3$  obszarów schematycznie przedstawiono na rys. 6).
3. Dla każdej z próbek wyznaczana jest wartość badanej charakterystyki. Otrzymane dla  $i$ -tego ( $i = 1, 2, \dots, k$ ) obszaru wyniki są oznaczane przez  $(x_{i1}, x_{i2}, \dots, x_{in})$ .
4. Dla każdej z próbek wyznaczana jest maksymalna zaobserwowana wartość  $\hat{x}_i = \max\{x_{i1}, x_{i2}, \dots, x_{in}\}$  dla  $i = 1, 2, \dots, k$  badanej charakterystyki. Tym sposobem otrzymywany jest ciąg wartości maksymalnych  $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k)$ .
5. Obliczana jest średnia oraz odchylenie standardowe z otrzymanych wartości maksymalnych

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k \hat{x}_i \quad \text{oraz} \quad S(x) = \sqrt{\frac{1}{k-1} \sum_{i=1}^k (\hat{x}_i - \bar{x})^2}$$

6. Estymowane są parametry rozkładu Gumbela  $\lambda$  i  $\delta$ . Oszacowanie jest przeprowadzane na podstawie wzoru (7), czyli

$$\delta \approx \frac{\sqrt{6}S}{\pi} \quad \text{oraz} \quad \lambda \approx \bar{x} - 0,57772\delta \approx \bar{x} - 0,57772 \frac{\sqrt{6}S}{\pi}$$

7. Testem Kołmogorowa-Smirnowa (Kanji, 2006) weryfikowana jest hipoteza o zgodności danych z próby (maksima) z rozkładem Gumbela dla wartości maksymalnych.



Rys. 6. Schemat pobierania próbek o liczebności  $n = 5$  z  $k = 3$  obszarów materiału niekształtnego

\* W artykule poprzez losowy wybór punktu P rozumiany jest przypadkowy wybór, realizowany poprzez losowanie z rozkładu jednostajnego, współrzędnych  $(x_1, x_2, x_3)$  w przestrzeni trójwymiarowej.

### 3. Analiza symulacyjna

Funkcjonowanie przedstawionej procedury zostało poddane analizie symulacyjnej. W symulacjach uwzględniono dwa rodzaje rozkładu badanej zmiennej: normalny i logarytmiczno-normalny. W obu przypadkach rozkładem granicznym wartości ekstremalnych jest rozkład Gumbela wartości maksymalnych (por. tab. 1). W analizach symulacyjnych rozważano próbki pobierane z losowych obszarów sześciennego bloku o boku 1 z wymienionych rozkładów charakteryzujących się jednorodnością (oznaczenia –  $N_0$  i  $LN_0$ ) oraz niejednorodnością (oznaczenia –  $N_1$ ,  $LN_1$ ,  $N_2$  i  $LN_2$ ) o specyfikacji jak poniżej:

$N_0$  – rozkład normalny  $N(10; 1)$

$LN_0$  – rozkład log-normalny  $LN(0; 1)$

$N_1$  – rozkład normalny  $N(10 + x_1 + x_2 + x_3, 1)$

$LN_1$  – rozkład log-normalny  $LN(x_1 + x_2 + x_3, 1)$

$N_2$  – rozkład normalny  $N(10 + 2x_1 + 2x_2 + 2x_3, 1)$

$LN_2$  – rozkład log-normalny  $LN(2x_1 + 2x_2 + 2x_3, 1)$

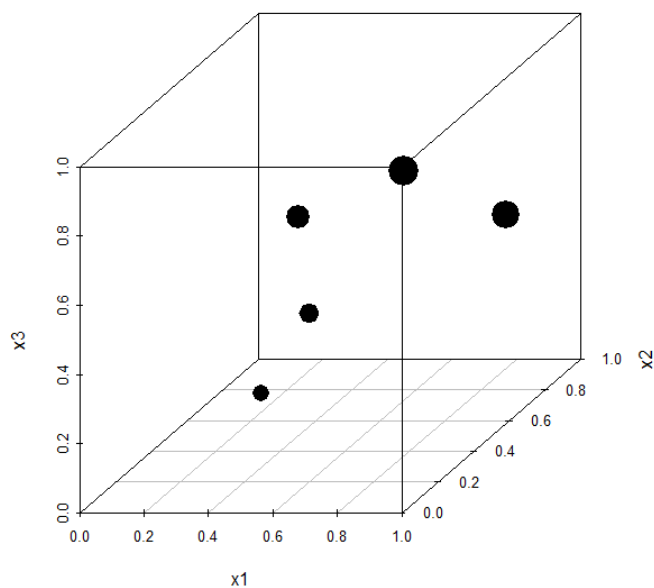
gdzie  $x_1, x_2, x_3 \in (0,1)$  są współrzędnymi punktu badanego bloku sześciennego.

Blok sześcienny został przedstawiony na rys. 7. Schematycznie zaprezentowano również niejednorodności badanej charakterystyki. Większe wartości zmiennej symbolicznie oznaczono na rys. 7 większymi kropkami.

Wszystkie analizy symulacyjne wykonano w programie R (WWW1). We wszystkich przypadkach pobierano losowo  $k = 5, 10$  oraz  $30$  próbek o liczebnościach  $n = 5$ .

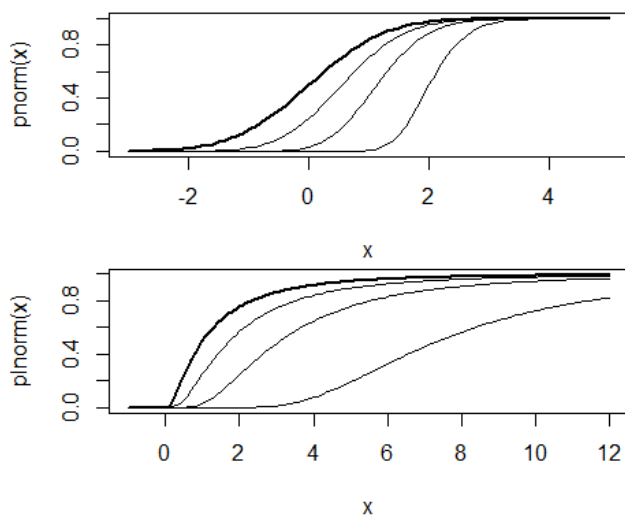
We wszystkich rozważanych przypadkach w analizach symulacyjnych sprawdzano zgodność rozkładu próbkowego z dwoma postaciami rozkładów: granicznym rozkładem, czyli rozkładem Gumbela dla wartości maksymalnych, oraz rozkładem dokładnym. Dystrybuanty dokładnych rozkładów wartości maksymalnych dla próbek o liczebnościach  $n = 1, 2, 5$  i  $30$  dla danych pochodzących z rozkładu normalnego oraz logarytmiczno-normalnego wyznaczono na podstawie wzoru (1) i przedstawiono na rys. 8.





Rys. 7. Schemat materiału, z którego symulacyjnie pobierano próbki z symbolicznie zaznaczoną niejednorodnością badanej charakterystyki (dot.  $N_1$ ,  $LN_1$ ,  $N_2$  i  $LN_2$ )

Wyniki analizy symulacyjnej oceny prawdopodobieństw odrzucenia hipotezy głoszącej, że rozkład próbkowy jest zgodny z rozkładem granicznym Gumbela przedstawia tab. 2, a że jest zgodny z rozkładem dokładnym – tab. 3. Wyniki te zostały również zobrazowane na rys. 9.



Rys. 8. Dystrybuanty rozkładów dokładnych wartości maksymalnych dla próbek o liczebnościach  $n = 1, 2, 5$  i  $30$  pochodzących z rozkładów normalnego (u góry) i logarytmiczno-normalnego (u dołu).

W przypadkach gdy badany materiał jest jednorodny ( $N_0$  oraz  $LN_0$ ), odwołanie się do dokładnego rozkładu wartości maksymalnych pozwala na uzyskanie właściwego rozmiaru testu (prawdopodobieństwa odrzucenia hipotezy o zgodności rozkładów są zbliżone do  $\alpha = 0,05$ ). Właściwego rozmiaru testu nie zapewnia porównanie wyników z próby z rozkładem granicznym. W przypadku występowania niejednorodności materiału zarówno poprzez wykorzystanie rozkładu dokładnego, jak i granicznego postaci, czyli rozkładu Gumbela otrzymujemy wyniki podobne. W większości przypadków większą mocą charakteryzuje się test, gdzie wykorzystywany jest rozkład dokładny wartości maksymalnych.

Tabela 2

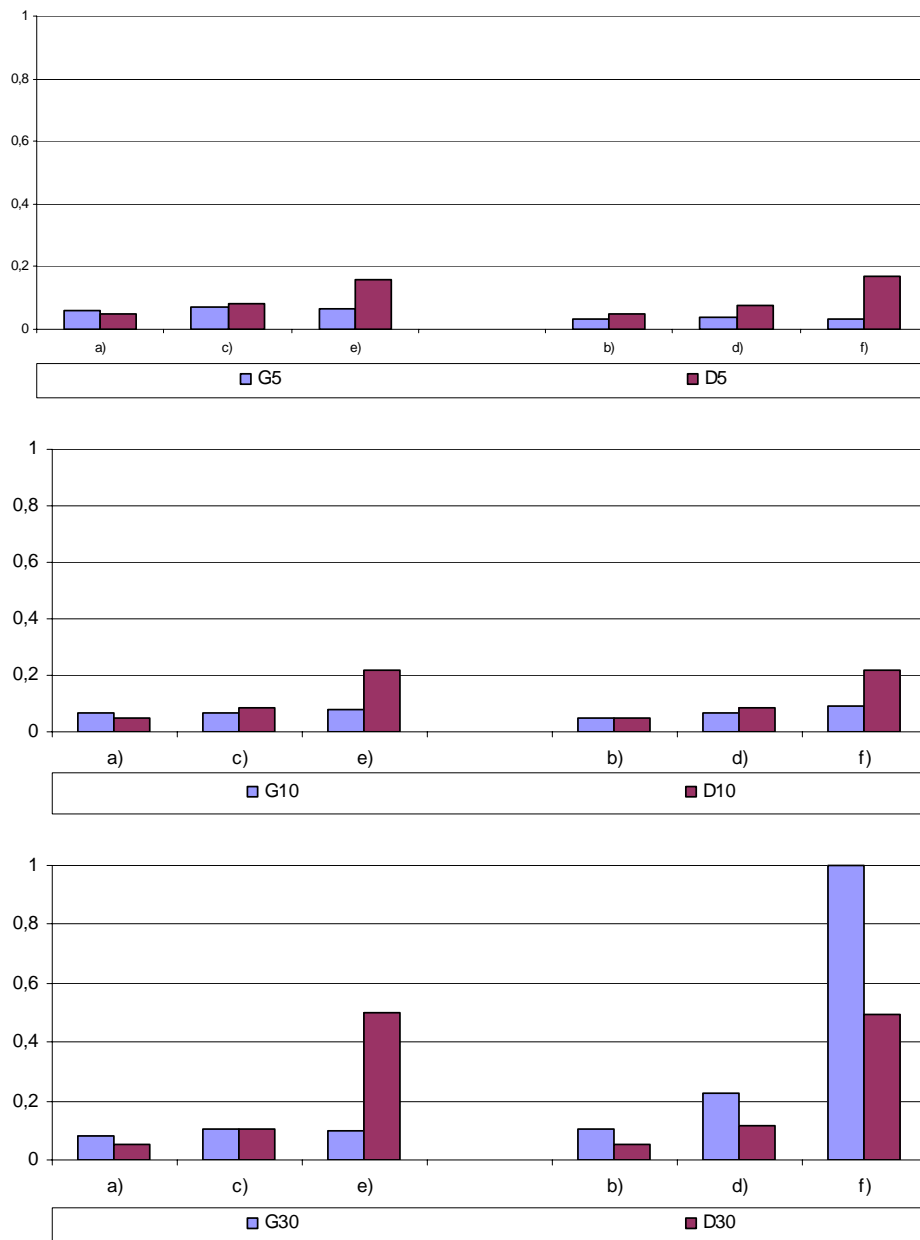
Oceny prawdopodobieństw ( $p$ ) oraz błąd oceny ( $sp$ ) odrzucenia hipotezy o zgodności badanego rozkładu z rozkładem Gumbela

Rozkład	$k = 5$		$k = 10$		$k = 30$	
	$p$	$sp$	$p$	$sp$	$P$	$sp$
$N_0$	0,058	0,008	0,068	0,008	0,079	0,007
$LN_0$	0,032	0,004	0,049	0,009	0,103	0,008
$N_1$	0,070	0,004	0,067	0,006	0,105	0,004
$LN_1$	0,037	0,007	0,065	0,009	0,226	0,010
$N_2$	0,066	0,007	0,079	0,006	0,101	0,011
$LN_2$	0,032	0,006	0,092	0,009	1,000	0,000

Tabela 3

Oceny prawdopodobieństw ( $p$ ) oraz błąd oceny ( $sp$ ) odrzucenia hipotezy o zgodności badanego rozkładu z rozkładem dokładnym wartości maksymalnych

Rozkład	$k = 5$		$k = 10$		$k = 30$	
	$p$	$sp$	$p$	$sp$	$p$	$sp$
$N_0$	0,048	0,006	0,051	0,007	0,050	0,006
$LN_0$	0,049	0,007	0,051	0,006	0,050	0,008
$N_1$	0,082	0,006	0,086	0,010	0,107	0,008
$LN_1$	0,077	0,011	0,085	0,012	0,116	0,013
$N_2$	0,157	0,018	0,218	0,013	0,501	0,009
$LN_2$	0,167	0,012	0,216	0,013	0,497	0,009



Rys. 9. Oceny prawdopodobieństw odrzucenia hipotezy o jednorodności badanego materiału dla prób o liczebnościach  $n = 5$  (u góry),  $n = 10$  (w środku) i  $n = 30$  (na dole) dla rozkładu Gumbela (G) oraz rozkładu dokładnego (D)

## Podsumowanie

W badaniach dotyczących jakości próbek przedstawianych do kontroli zwykle zakłada się jednorodność materiału ze względu na badane charakterystyki. Nie zawsze przyjęcie takiego założenia jest uzasadnione. Bez potwierdzenia jednorodności sprawdzanego materiału niezasadne jest przeprowadzanie różnych klasycznych analiz jakości. W artykule rozważano problem weryfikacji hipotezy głoszącej, że przedstawiany do kontroli materiał jest jednorodny. W tym celu weryfikowano hipotezę o zgodności rozkładu wartości maksymalnych z rozkładem dokładnym oraz z rozkładem Gumbela, który przy założonych rozkładach charakterystyk jest rozkładem granicznym wartości maksymalnych. Proponowana procedura ma na celu wykrycie istnienia obszarów o różnych wartościach badanej charakterystyki w materiałach niekształtnych. Nieco większe możliwości podjęcia prawidłowej decyzji zapewnia porównanie rozkładu próbkowego z rozkładem dokładnym, zwłaszcza dla prób o niewielkich liczebnościach. Jednak porównanie z rozkładem dokładnym wartości maksymalnych możliwe jest jedynie wówczas, gdy dysponujemy postacią funkcyjną (dystrybuantą) rozkładu badanej zmiennej. W przeciwnym przypadku jedynym wyjściem jest porównanie rozkładu próbkowego wartości maksymalnych z rozkładem granicznym (np. rozkładem Gumbela). Przeprowadzone analizy potwierdziły, że większą mocą charakteryzuje się test porównujący rozkład próbkowy z rozkładem dokładnym, ale wykorzystanie rozkładu Gumbela również przynosi dobre rezultaty.

## Bibliografia

- Boos D.D., Brownie C. (1989): *Bootstrap Methods for Testing Homogeneity of Variances*. „Technometrics”, Vol. 31, No. 1.
- Castillo E., Hadi A.S., Balakrishnan N., Sarabia J.M. (2005): *Extreme Value and Related Models with Applications in Engineering and Science*. John Wiley & Sons, New Jersey.
- Efron B., Tibshirani R. (1993): *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Good P.I. (1994): *Permutation Tests: A Practical Guide for Testing Hypotheses*. Springer-Verlag, New York.
- Kanji G.K. (2006): *100 Statistical Tests*. Sage Publications, London.
- Kończak G. (2006): *O teście równości wariancji wykorzystującym metodę bootstrap*. W: *Taksonomia 13. Klasyfikacja i analiza danych – teoria i zastosowania*. Red. K. Jajuga, M. Walesiak. Wydawnictwo Akademii Ekonomicznej, Wrocław, s. 543-552.
- Scheaffer R.L. (1971): *A Test for the Homogeneity of a Mixture*. „Technometrics”, Vol. 13, No. 2.
- (WWW1): <http://www.r-project.org>.

## THE USE OF EXTREME VALUE DISTRIBUTIONS IN CHECKING THE QUALITY SHAPELESS MATERIALS HOMOGENEITY

### Summary

For quality control it is essential that the control samples are homogeneous. In practice this is impossible, and the requirement can be reduced to the condition that the samples were taken from the same population. The study presented in the paper is an analysis of the issue of testing the quality of the shapeless material. As a shapeless material is referred to the material from which it is impossible to directly extract the individual elements, packages, etc. This paper proposes a method to verify the hypothesis that the tested material is homogeneous due to the observed characteristics.