

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

Marcin Korzeń

Obliczenia numeryczne na zmiennych losowych

Szczecin 2013

Recenzenci

PROF. DR HAB. INŻ. OLGIERD HRYNIEWICZ

DR HAB. RYSZARD KOZERA

Opracowanie redakcyjne

KATARZYNA MITAN

WYDANO ZA ZGODĄ

REKTORA ZACHODNIOPOMORSKIEGO UNIWERSYTETU TECHNOLOGICZNEGO
W SZCZECINIE

ISBN 978-83-7663-126-4

Wydawnictwo Uczelniane Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie
70-311 Szczecin, al. Piastów 50, tel. 91 449 47 60, e-mail: wydawnictwo@zut.edu.pl

Druk PPH „ZAPOL” Dmochowski, Sobczyk Sp.j., 71-062 Szczecin, al. Piastów 42, tel. 91 434 10 21
e-mail: zarzad@zapol.com.pl

Spis treści

Lista symboli	7
Przedmowa	9
1. Wprowadzenie	11
1.1. Obliczenia probabilistyczne oraz wnioskowanie	11
1.2. Cel badań	15
1.3. Arytmetyka zmiennych losowych i propagacja niepewności	16
1.4. Przegląd stosowanych rozwiązań	17
1.4.1. Arytmetyka interwałowa	17
1.4.2. Propagacja błędów	17
1.4.3. Histogramy i dyskretyzacja	18
1.4.4. Arytmetyka rozmyta	18
1.4.5. Arytmetyka zmiennych losowych	19
1.5. Istniejące biblioteki i oprogramowanie do obliczeń na zmiennych niepewnych	19
1.6. Wnioskowanie	21
2. Operacje na zmiennych losowych niezależnych	23
2.1. Wstęp	23
2.2. Operacje arytmetyczne na zmiennych losowych	24
2.3. Funkcje zmiennych losowych	25
2.4. Operacje arytmetyczne z użyciem funkcji kawałkami gładkich	26
2.4.1. Funkcje kawałkami gładkie	26
2.4.2. Dodawanie i odejmowanie	27
2.4.3. Mnożenie	29
2.4.4. Dzielenie	30
2.4.5. Funkcje zmiennych losowych	31
2.5. Minimum i maksimum, rozkłady statystyk pozycyjnych	31
2.6. Zmienne dyskretne i mieszane	33
2.7. Wybór sposobu reprezentacji	35
2.7.1. Transformacje i momenty	35
2.7.2. Numeryczna reprezentacja funkcji gęstości i dystrybuanty	36
2.8. Klasy rozkładów zamknięte ze względu na operacje arytmetyczne i potęgowanie	37

3. Obliczenia na zmiennych losowych zależnych	43
3.1. Wstęp	43
3.2. Wnioskowanie probabilistyczne	44
3.2.1. Sformułowanie zagadnienia	44
3.2.2. Stosowane rozwiązania	44
3.3. Wielowymiarowe modele zależności	45
3.3.1. Wielowymiarowy rozkład normalny	45
3.3.2. Kopuły	46
3.3.3. Kopuły archimedesowe	49
3.3.4. Próba uporządkowana, łączny rozkład dwóch statystyk pozycyjnych	51
3.4. Dwuwymiarowe miary zależności pomiędzy zmiennymi	52
3.5. Obliczenia na zmiennych zależnych	54
3.5.1. Twierdzenie o zamianie zmiennych	54
3.5.2. Wyrażenia algebraiczne zawierające dwie zmienne losowe	56
3.5.3. Skrajne zależności pomiędzy zmiennymi, kopuły M i W	59
3.5.4. Ograniczenia na dystrybuantę wyniku	61
4. Metody numeryczne stosowane w obliczeniach probabilistycznych	63
4.1. Wstęp	63
4.2. Interpolacja i aproksymacja	64
4.2.1. Interpolacja barycentryczna	64
4.2.2. Wymierna interpolacja barycentryczna	66
4.3. Obliczenia z użyciem wielomianów Czebyszewa	67
4.3.1. Związki z szeregami Fouriera	67
4.3.2. Przyrostowy algorytm interpolacji	71
4.3.3. Całkowanie szeregów Czebyszewa	74
4.3.4. Różniczkowanie szeregów Czebyszewa	74
4.3.5. Miejsca zerowe, ekstrema	75
4.4. Całkowanie numeryczne	77
4.4.1. Całki na przedziale skończonym	77
4.4.2. Przedziały nieskończone, funkcje z osobliwościami	79
4.5. Dokładność aproksymacji	81
4.5.1. Rezultaty teoretyczne	81
4.5.2. Dokładność obliczeń zmiennopozycyjnych	82
4.5.3. Ograniczenia formuły barycentrycznej	84
4.6. Aproksymacja funkcji gęstości, funkcje kawałkami gładkie	85
4.6.1. Funkcje na przedziale skończonym	85
4.6.2. Funkcje na przedziałach nieskończonych, funkcje z osobliwościami	86
4.6.3. Aproksymacja w klasie funkcji $\mathcal{F}$	89
4.7. Pakiet chebfun	89
4.8. Obliczenia numeryczne w pakiecie PaCAL	91

4.9.	Przykłady numeryczne	94
4.9.1.	Funkcje ciągłe na przedziale domkniętym	94
4.9.2.	Funkcje z osobliwościami, funkcje na przedziałach ograniczonych	98
4.9.3.	Funkcje z ciężkimi ogonami	102
5.	Projekt PaCAL	105
5.1.	Cel i główne założenia projektu	105
5.2.	Instalacja i pierwsze uruchomienie	106
5.3.	Obliczenia na zmiennych niezależnych	107
5.3.1.	Szczegóły implementacyjne	107
5.3.2.	Rozkłady jednowymiarowe zdefiniowane w PaCAL	108
5.3.3.	Metody klasy Distr	113
5.3.4.	Arytmetyka zmiennych losowych niezależnych	115
5.3.5.	Zmienne dyskretne i mieszane	116
5.3.6.	Rozkłady warunkowe	117
5.3.7.	Estymacja parametrów	118
5.4.	Semantyka	118
5.5.	Praca z biblioteką PaCAL	119
5.6.	Rozkłady wielowymiarowe, obliczenia na zmiennych zależnych	120
6.	Przykłady i zastosowania	123
6.1.	Testowanie dokładności	123
6.2.	Przykłady numeryczne	126
6.2.1.	Środowisko testowe	126
6.2.2.	Porównanie z pakietami obliczeń symbolicznych	126
6.2.3.	Porównanie z pakietem chebfun	128
6.2.4.	Rozwiązania teoretyczne	130
6.2.5.	Rozkłady niecentralne	132
6.3.	Możliwości i zastosowania pakietu PaCAL	135
6.3.1.	Wnioskowanie statystyczne	135
6.3.2.	Pomiary wielkości fizycznych	136
6.3.3.	Agregacja ocen eksperckich	138
6.4.	Obliczenia na zmiennych zależnych	143
6.4.1.	Rozstęp z próby, rozstęp międzykwantylowy	143
6.4.2.	Skrajne możliwe wartości sumy zmiennych losowych	144
7.	Podsumowanie	147
	Bibliografia	149
	Indeks	157
	Summary	161
	Zusammenfassung	163

Numerical computations on random variables

Summary

The central theme of the monograph is the automation of numerical computations on random variables. Calculations in the model of independent variables and simpler models of dependent random variables were discussed thoroughly, including distributions defined with using of copulas or specific models, like the joint distribution of order statistics. The theoretical properties of computations on independent random variables were presented from the point of view of numerical implementation. The range of applicability of the methods presented is wide and it covers most of the distributions, occurring in practical problems, including: continuous variables, discrete variables and also mixed case.

The main idea presented in the paper consists in performing computations in a numerically accurate way, based on numerical representations of random variables and keeping the semantics similar to symbolic calculations. For this reason, a significant part of the paper concerns the practical implementation of numerical computations and representations of random variables. In this scope, barycentric interpolation, together with the method of approximation based on Chebyshev's polynomials was described. Furthermore, the methods of approximation and integration of functions defined on infinite domains and functions with singularities were presented. A lot of attention was given to the issue of the evaluation of the numerical accuracy. The practical implementation of the described methods was also presented – **PaCAL** library written in Python programming language. The library is publicly available at website: <http://pacal.sf.net>.

Performing computation with variables defined in an imprecise manner is of a great importance and appears in both practical and theoretical applications. Presented applications includes the computation of statistic distributions, propagation of uncertainties and measurement errors in metrology as well as the aggregation of experts judgements. The numerical accuracy of the presented approach is high, generally close to the machine precision. Despite of the fact that presented solutions are general, they are often more accurate than the generally applied specific routines, as for instance, in the case of the

implementation of non-central distributions. These distributions are computed frequently with a higher accuracy than the solutions implemented in the statistical libraries of software like Python, Matlab or R.

Numerische Berechnungen mit Zufallsvariablen

Zusammenfassung

Das Leitthema der Monographie ist die Automatisierung von numerischen Berechnungen mit Zufallsvariablen. Es wurde ausführlich auf Berechnungen mit unabhängigen Variablen eingegangen und es wurden Berechnungen in einfachen Modellen von abhängigen Variablen dargestellt, die die mit Hilfe von Copulas erstellten Verteilungen umfassen, oder auch besondere Modelle, wie z.B. die Gesamtverteilung von Zufallsstatistiken. Es wurden theoretische Eigenschaften von Berechnungen mit unabhängigen Zufallsvariablen aus dem Gesichtspunkt der numerischen Implementierungen dargestellt. Der Einsatzbereich von präsentierten Methoden ist sehr breit und umfasst die Mehrheit von Verteilungen, die in praktischen Problemen auftreten, darin die Berechnungen von stetigen, diskreten und gemischten Zufallsvariablen.

Die in der Arbeit präsentierte Hauptidee besteht darin, die Berechnungen in einer numerisch genauen Weise durchzuführen und dabei auf numerischen Darstellungen von Zufallsvariablen zu basieren und unter Beibehaltung ähnlicher Semantik wie im Fall von symbolischen Berechnungen. Aus diesem Grund betrifft ein wesentlicher Teil der Arbeit praktische Implementierung von numerischen Mechanismen und Darstellungsarten von Zufallsvariablen. In diesem Bereich wurden die baryzentrische Interpolation und Approximationsmethoden beschrieben, die auf Tschebyscheff-Polynomen basieren. Darüber hinaus wurden die Methoden der Approximation und Integration von bestimmten Funktionen für unendliche Bereiche als auch von Funktionen mit Singularitäten dargestellt. Eine große Aufmerksamkeit schenkte man dem Problem der Bewertung der Genauigkeit von Berechnungen. Es wurde auch eine praktische Implementierung der beschriebenen Methoden vorgestellt – eine in der Python-Programmiersprache geschriebene **PaCAL**-Bibliothek. Diese Bibliothek ist öffentlich verfügbar unter Adresse <http://pacal.sf.net>.

Die Durchführung von Berechnungen mit Variablen, die in einer nicht präzisen Weise bestimmt wurden, hat eine große Bedeutung und erscheint in vielen sowohl praktischen als auch theoretischen Anwendungen. Der Anwendungsbereich umfasst die Bestimmung von Statistikverteilungen, Propagationen von Messunsicherheiten und Messfehlern

in der Metrologie oder auch die Aggregation von Expertenprognosen. Die Genauigkeit von erzielten Lösungen ist hoch und liegt gewöhnlich der maschinellen Präzision nahe. Obwohl die präsentierten Lösungen allgemeiner Natur sind, sind sie oft genauer als die allgemein eingesetzten Sondermethoden, wie zum Beispiel im Fall der Implementierung von Nichtzentralen-Verteilungen. Diese Verteilungen werden oft mit einer höheren Genauigkeit berechnet als Lösungen, die in statistischen Bibliotheken der Programme Python, Matlab oder R implementiert wurden.