

# АВТОРСКА СПРАВКА

на трудовете на доц. дмн Евгения Стоименова

представени за конкурса за професор по математика  
научна специалност 01.01.10 "Теория на вероятностите и математическа  
статистика"

**1. Рангови критерии за проверка на непараметрични статистически хипотези** [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. 9 от статиите са самостоятелни, а 4 в съавторство. Новите резултати са публикувани в [1, 2, 3, 4].

Статиите от тази група са посветени на непараметрични критерии за сравняване на две или повече разпределения. В задачата за две разпределения нулевата хипотеза е че двете разпределения съвпадат срещу общата алтернатива, че разпределението на втората извадка е изместено, например, наляво по отношение на разпределението на първата извадка. Без предположения за конкретен вид на разпределенията, статистиките на критериите за проверка на тези хипотези могат да се основават само на ранговете на обединената извадка. Нека  $X_1, \dots, X_m$  и  $Y_1, \dots, Y_n$  са две независими извадки от двете разпределения, съответно. Образоваме вариационния ред на обединената извадка и определяме ранга на всяко от наблюденията. Векторът от ранговете на всички  $m + n$  наблюдения е пермутация в  $S_{m+n}$ . Ако двете извадки са от едно и също разпределение, то ранговете на първата извадка ще са близки до ранговете на втората извадка, докато при алтернативата – ранговете на първата извадка ще са предимно измежду най-малките по големина рангове. Тогава критерият може да се дефинира чрез разстоянието между вектора от наредените рангове на първата и втората извадка и вектора  $(1, 2, \dots, m + n)$ . При подходящ избор на разстояние в  $S_{m+n}$  се получават редица известни рангови критерии.

Критерият за две извадки, изследван в [1] е предложен в по-ранната статия [10]. Статистиката на критерия има вида  $M_r = \max\{R_m - m, m + 1 - S_1\}$ , където  $R_m$  е максималния ранг от първата извадка, а  $S_1$  е минималния ранг от втората извадка в обединената извадка. Статистиката е породена от разстояние върху класове от пермутации. В статията [1] е получено точното разпределение за неравни извадки при Леманова алтернатива. Лемановата алтернатива е частен случай на алтернатива на отместване, която за някои практически важни разпределения е аналогична на параметрична хипотеза с параметър на отместване (тя включва експоненциално разпределение, разпределение на Вейбул и по-общо - разпределение с пропорционална хазартна функция). Разпределението при алтернативата е използвано за определяне на мощността на критерия. Мощността на критерия е сравнена с мощността на някои други популярни рангови критерии.

В статиите [2, 3] са предложен и изследван клас от рангови статистики, които са обобщение на статистиките на предхождане (Precedence tests). Рангови статисти-

ки, които се базират на броя на наблюденията от една извадка, които предхождат (превъзхождат) някоя фиксирана стойност или случайна величина наричаме рангови статистики на предхождане (превъзхождане). За нива на предхождане и превъзхождане се използват порядкови статистики по следния начин. Нека  $X_1, \dots, X$  и  $Y_1, \dots, Y_n$  са две независими извадки от разпределения  $F$  и  $G$ , съответно. Дефинираме статистиките на превъзхождане  $A_r$  и  $B_s$ , съответно като: броя на елементите на  $Y$ -извадката по-големи от  $(m - s)$ -тата поредна стойност на  $X$ -извадката и броя на елементите на  $X$ -извадката непревъзхождащи  $(1 + r)$ -тата поредна стойност на  $Y$ -извадката.

В статията [4] са дефинирани статистики на предхождане и превъзхождане по отношение на една и две извадки. Направен е паралел с подобни статистики, дефинирани чрез наблюденията на една извадка по отношение на нива, определени със случайни величини. В много приложни задачи нивата и статистиките са независими. Разглежданите от нас статистики  $A_r$  и  $B_s$  са зависими, и за да могат да се прилагат в статистически модели, се изисква явен вид на тяхното съвместно разпределение. Идеите са представени в пленарен доклад на международната конференция CDAM'10 в Минск. В [3] е изследвано съвместното разпределение на статистики на предхождане и превъзхождане при две извадки. Изведено е разпределението на статистиката с нива определени от фиксирани порядкови статистики при еднакви разпределения на двете извадки. Направени са симулационни изследвания на съвместното разпределение когато разпределението на втората извадка е изместено спрямо първото чрез параметър на отместване. Статията е по покана на редакторите на *Austrian Journal of Statistics*.

В статията [2] се разглежда нов клас от рангови критерии за проверка на хомогенност на две извадки от две разпределения срещу едностранна алтернатива, че втората случайна величина е стохастично по-голяма от първата. Разгледан е случая на равни извадки. Малки стойности на статистиката  $M_r = \max\{n - A_r, n - B_r\}$  водят до отхвърляне на нулевата хипотеза в полза на алтернативата за стохастична наредба. Изведено е разпределението на  $M_r$ , получени са в явен вид съвместното разпределение на  $A_r$  и  $B_r$  при Леманова алтернатива. Изведено е разпределението на  $M_r$  при нулевата хипотеза и при Леманова алтернатива. Изследвана е мощността на критериите със статистики  $M_r$  при последователни стойности на  $r$  и са направени сравнения на мощността на критериите с някои от критериите на предхождане. Последните резултати в това направление е изследването на клас от рангови критерии, които са обобщение на критерий, изследван от Тюки, Хайек и Шидак. Статистиката на критерия има вида  $A_r + B_s$ , където  $A_r$  и  $B_s$  са подходящо подбрани статистики на предхождане и превъзхождане при две извадки.

Статиите [5, 6, 7, 8, 9, 10, 11, 12] съдържат предхождащи резултати по тематиката за рангови критерии. Резултатите от тези статии са свързани с рангови статистики, породени от метрики в пермутационната група и са основа за първите две глави на дисертацията за дмн. Съответните критерии за проверка на статистически хипотези са удобни в редица задачи включително за сравняване на две и повече извадки

от разпределения с неизвестен параметър на положението и др. Конструирането на непараметрични рангови критерии и обобщаването им за други задачи за проверка на хипотези е една важна статистическа задача. В статията [10] е построен клас от рангови статистики, породени от Чебишевата метрика в пермутационната група. Чебишевата метрика е равна на максимума на абсолютната стойност на разликата между компонентите на две пермутации. Разпределението на тази метрика в  $S_n$  е изучавано в статията [11] при предположение за равномерно разпределение на пермутациите. Метричният подход за конструиране на рангови статистики позволява да се обобщават по естествен начин статистики, дефинирани за някоя непараметрична задача за проверка на хипотези до статистики за други непараметрични хипотези.

Всяка метрика във фактор-пространството  $S_{n+m}/S_n \times S_m$  поражда по естествен начин статистически критерии за сравняване на разпределенията на две извадки. В задачата за сравняване на  $k$  разпределения,  $k > 2$ , имаме  $k$  извадки с обеми  $n_1, \dots, n_k$ , съответно, а критериите могат да се основават на метрики в  $S_N/S_{n_1} \times S_{n_k}$ , ( $N = \sum n_i$ ). При избор на подходящи разстояния се получат редица от известните критерии, в това число статистиката на Mann-Whitney, статистиката на Wilcoxon, медианният критерий на Mood, критерия на Колмогоров-Смирнов и др. Критерият за две извадки с наредена алтернатива, породен от Чебишевата метрика, е обобщен за сравняване на две разпределения срещу двустранна алтернатива, за сравняване на повече от две разпределения срещу наредена и ненаредена алтернатива и за проверка на тренд.

В статията [9] са изследвани асимптотичните свойства на ранговата статистика за две извадки от [10] при нулевата хипотеза. Резултатите са приложени за доказване на аналогични свойства на  $E$ -статистиката на Шидак. В [8] е получено разпределението на статистиката по две извадки при Лемановата алтернатива, а в [6] е изследвана мощността на такива критериите срещу Лемановата алтернатива и са направени сравнения с критерии от класа на критериите на предхождане. В [5] е изследвано съвместното разпределение на статистики на предхождане и превъзхождане при две извадки при нива на предхождане и превъзхождане, определени от екстремалните наблюдения от двете извадки. В [7] е изследвана мощността на няколко критерия, основаващи се на превъзхождащи наблюдения срещу алтернатива на отместване. Мощността на критериите е оценена чрез Монте Карло метод. Резултатите са представени в пленарен доклад на международната конференция SIS'09 в Брюксел. Метрика с фиксиран вектор е използвана в [12] за дефиниране на статистика на непараметричен критерий за проверка на тренд при частично нареждане. Доказано е асимптотичното разпределение на статистиката на критерия.

## 2. Непараметрично оценяване на плътността на разпределение [13, 14, 15].

Резултатите в тези статии са свързани с непараметрично оценяване на вероятностната плътност когато данните са интервално ограничени (цензурирани). Те са част от съдържанието на трета глава на дисертацията за дмн. Трите работи са самостоя-

телни. Интервално ограничените данни са резултат от наблюдаване на непрекъснатата случайна величина посредством функция, която съпоставя на всяка наблюдавана стойност интервал, който я съдържа. Такива данни възникват естествено в практиката. В статията [13] направено обобщение на оценката на Надарая-Уотсън за оценяване на вероятностната плътност по цензурирани наблюдения. Чрез нея се коригира максимално правдоподобната оценка (МПО) в интервалите, в които оценката на плътността е занижена поради цензурирането. Оценката е състоятелна при условие за асимптотики за редки данни (sparse asymptotics), докато МПО не е. Резултатите са представени в пленарен доклад на международната конференция CDAМ в Минск. В статията [14] е предложено в модифицираната оценка на Надарая-Уотсън да се използват ядра с променливи дължини. Построен е алгоритъм за интерактивно графично изследване на различните оценки. Повечето непараметрични техники (все още) не са включени в стандартните статистически програмни системи. Ето защо активните изследвания в областта на непараметричното оценяване на функции е свързано и с практическа реализация на алгоритмите. Статията е по покана на редакторите на *Austrian Journal of Statistics*. В статията [15] е направена връзка между интервално цензурираните данни и размитите числа (fuzzy numbers). Реализиран е алгоритъм за интерактивно изучаване на хистограмни оценки по цензурирани данни.

### **3. Приложно статистическо моделиране – [16, 17, 18, 19, 20, 21, 22].**

Резултатите от тази група статии са свързани със статистически модели, приложими някои области на механиката и инженерните науки. Статиите имат научно-приложен характер и са в съавторство с колеги от съответните области. Статиите [16, 17, 18] не са включвани в предходящи конкурси и дисертации.

В статията [17] е изследвано влиянието на епоксидното покритие на стоманобетонни пръти от стоманобетонни структури върху издръжливостта на връзките между бетона и арматурата. През последните години, са известни редица публикации, посветени на нарастващата нужда от оценка на съществуващите стоманобетонни структури които съдържат главно обикновени пръти. В анализа, представен в тази статия, предполагаме, че покриването с епоксидна смола на обичайни пръти със закалено покритие създава нови комплексни свойства на връзката между бетона и стоманата. Установихме, че тези връзки могат да бъдат опростени и моделирани по аналогия с деформирани стоманени пръти. За да се оцени влиянието на грапава повърхностна текстурата при опън, разгледахме редица физически параметри характеризиращи връзката: относителна грапавост на площта, средният ъгъл на неравностите и средната дебелина на покритието, както и стандартното отклонение на дебелината на покритие. Оценяването е извършено чрез анализ на разтегателен тестови данни от експериментално изследване на проби от стоманобетонни пръти.

В статията [16] е изследване влиянието на геометричните параметри на капков емитер, състоящ се от лабиринт от канали с триъгълен профил при освобождаването

на водата. При моделирането на такъв капков емитер е изследвано влиянието на параметри като разстояние между зъбчетата, височина на назъбеност, дълбочината на лабиринта и др. Движението на водния поток е числено симулирано в съответствие с кубичен централен дизайн с помощта на стандартни програми. Чрез статистически модел е получено, че увеличението на разстоянието между зъбчетата влияе позитивно на освобождаването на водата. По-малко позитивно влияние има съвместното влияние на разстоянието между зъбчетата и дълбочината на лабиринта.

Статиите [20], [21] и [22] са посветени на модели от областта на геотехниката. Характеристичната крива почва-вода (SWCC) описва зависимостта между водното съдържание в почвата и хидравличното налягане в почвата. Известните в литературата модели имат емпиричен произход и са подходящи за различни видове почви. Моделите се ограничават до подбор на параметрична функция, описваща характеристичната крива. Тъй като зависимостта е нелинейна, оценките на параметрите се получават чрез нелинеен регресионен анализ. В статията [22] се предлага двуфазен регресионен модел за описване на SWCC. Допълнително се предполага, че точката на свързване на двете регресионни криви е неизвестна. Предложен е алгоритъм за оценяване на този допълнителен параметър заедно с параметрите на двете регресионни уравнения. В статията [21] се разглежда фамилия от нелинейни модели, които допускат линеаризация и позволяват прилагането на линеен регресионен анализ за оценка на неизвестните параметри. Популярните модели от литературата се включват в тази фамилия. Линеаризацията позволява оценка на грешката на прогнозиране на първоначалните нелинейни модели, при условие че данните от всяка конкретна почва удовлетворяват регресионните условия. Предложени са нови модели, описващи SWCC. Моделите са сравнени с класическите модели на ван Генухтен и Фредлунд. В статията [20] е предложен модел, описващ SWCC, който включва хистерезиса при двата процеса на изсушаване и намокряне. Моделът включва две независими променливи, налягане и дълбочина на уреда за отчитане на налягането. Моделът позволява да се предвижда водното съдържание и сканиращите криви при различни дълбочини.

Статията [19] е от областта на геокосмичната хидродинамика. В нея се оценява политропния индекс при слънчев вятър. Предложена е оценка на базата на регресионен модел за логаритъма на плътността на плазмата и температурата. Количественото определение на политропния подход се състои в дефиниране на газово-динамичните уравнения за запазване на енергията на флуида, което включва флуидното налягане, плътността и политропният индекс  $\gamma$ , който трябва да се определи. Обикновено коефициентът  $\gamma$  се определя по експериментални данни. Уравнението е приложимо при предположение, че разглеждаме сгъстен флуиден елемент или достатъчно хомогенна флуидна подобласт. Затова политропният индекс  $\gamma$  се оценява за всеки хомогенен участък на относителната плътност. Методът е приложен за анализ на реални онлайн данни, получавани от спътник.

В статията [18] се изследва влиянието на генотипа, условията на съхранение и др.

фактори върху кълняемостта и растежа на житни растения. Статистическите задачи са породени от изследванията на група от учени генетици. Приложен е дисперсионен анализ за няколко значими задачи на генетичното изследване. В някои от случаите бе потвърдено влиянието на определени фактори изследвани върху няколко показателя на растежа (корен, стъбло и др.) Направени са сравнения на качествата на растения, отглеждани при сухи условия с тези на контролна група, както и ефекта на нови генетични комбинации с този на диви родственици. Установено е, че растежа на растенията зависи от генотипа, както се вижда от високата им статистически значимост на два главни ефекта и при двете изследвани групи от растения. Размерът на основния генотипен ефект върху параметрите на растежа зависи от потенциала на водата, което се потвърждава статистически от анализа на факторните взаимодействия.

#### **4. Модели на тестове за постижения – [23]**

Книгата **“Измерителни качества на тестове”** [23] е в обем от 176 стр., състои се от 5 глави с графики и приложения. Предназначена е главно за специалисти от областта на образованието, ангажирани с конструирането на тестове за постижения. Ролята на тестовите като инструмент за измерване на постижения налага определени изисквания за тяхното качество. Това неизбежно включва оценка на точността на измерване, което от своя страна се основава на статистически анализ на данни. Традиционната измерителна характеристика на тест – неговата надеждност, е дефинирана като компонента на класическия модел на тест. Представени са основните статистически оценки на надеждността – алфа на Кронбах, формула на Кудер-Ричардсън и др. Важно място е отделено на надеждността на тестове, съставени от като композиция на от няколко теста. Изяснена е ролята на надеждността за постигане на крайната цел в изследването на тестове – осигуряване на желана точност на измерване. Друга основна тема, подробно застъпена в книгата, е изследването на измерителните качества на компонентите на теста, тестовите задачи. Освен популярните мерки за надеждност на задача, се дискутират съвременни мерки, основани на корелационни коефициенти. В книгата са включени и редица приложни теми като методика за анализ на учебни тестове, препоръки за тълкуване на резултатите от психометричния анализ, определяне на големината на извадката, осигуряваща желаната точност на оценките на параметрите на теста и др. Книгата стана популярна сред създателите на тестове и специалисти от областта. Тя се използва като учебно пособие в няколко магистърски курса по психология и методика на обучението по математика, физика, история, български език и др.

септември 2011

Подпис: