Von der Le Cam Theorie zur Datenanalyse: Resamplingmethoden für studentisierte Statistiken

Arnold Janssen
Heinrich-Heine Universität Düsseldorf
Mathematisches Institut

Wien, 22. November 2013

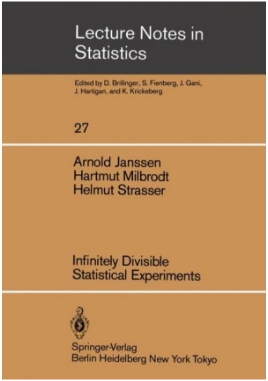
Teilprojekte: Förderung durch die DFG

Outline

Vorwort. Der Vortrag gibt einen Überblick über Ergebnisse für Permutationstests und Bootstrapstatistiken. Aus didaktischen Gründen werden die Ergebnisse durch einschlägige Beispiele illustriert und die Resultate werden exemplarisch vorgestellt. Dazu sind einige Vereinfachungen der Modelle vorgenommen worden. Die Referenzen enthalten die Ergebnisse jedoch in voller Allgemeinheit.

Einführung und Datenbeispiele

- Entscheidende Begegnungen:
- Zusammenarbeit mit Helmut Strasser



- Beitrag zur Le Cam Theorie (Likelihood Prozesse)

- Beispiel 2.5: Fraktale Brownsche Bewegung als asymptotischer log-Likelihood Prozess

(2.5) Examples.

(1) A Gaussian shift experiment on a Hilbert space $(H, \langle\cdot,\cdot\rangle)$ is a Gaussian experiment with kernel $\langle\cdot,\cdot\rangle$. All Gaussian shift experiments with $H = \mathbb{R}^1$ are given by the kernels $K: (s,t) \longmapsto a\,s\,t$ $(a > 0)$.

(2) Let $T = \mathbb{R}^1$ and $0 < \rho \leq 2$. Then

$$K: (s,t) \longmapsto a\,(|s|^\rho + |t|^\rho - |s-t|^\rho) \quad (a > 0)$$

is a positive definite and symmetric kernel. If $\rho = 2$, then this reduces to the situation of (1). If $\rho < 2$, the corresponding

Gaussian experiments arise e.g. as weak limits of experiments $(\mathbb{R}^n, \mathbb{B}^n, \{P_{n^{-1/\alpha}t} : t \in \mathbb{R}^1\})$, $n \in \mathbb{N}$, where

$$\frac{dP_\theta}{d\lambda^1}(x) := C(\rho) \cdot \exp(-|x-\theta|^\rho), \quad x \in \mathbb{R}^1, \quad \theta \in \mathbb{R}^1,$$

and $\alpha := \frac{\rho-1}{2}$. According to LeCam, 1969 (p. 109) this situation is "... plus compliquée mais plus interessante". It has been studied by several authors, e.g. Pflug, 1982, and Strasser, 1984 a. In the latter paper it has been shown that, under some additional invariance conditions, all kernels on $\mathbb{R}^1$ are of the form given above.

Stochastic Processes and their Applications 13 (1982) 45–57
North-Holland Publishing Company

A STATISTICALLY IMPORTANT GAUSSIAN PROCESS

Georg PFLUG
Institute of Statistics, University of Vienna, A-1010 Wien, Austria

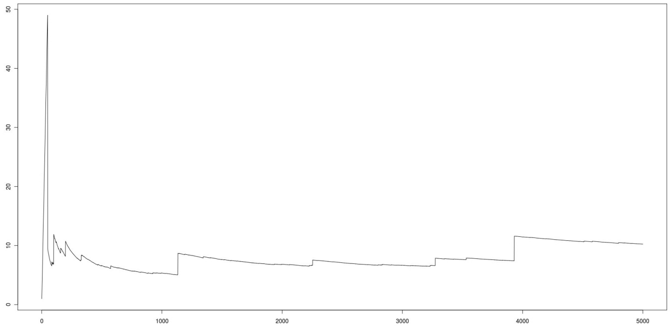
Acknowledgement

The author wishes to thank Prof. H. Strasser for the encouragement to this work and very useful hints and comments.

- Le Cam Theorie: Klare Strukturierung statistischer Probleme
- Von der Theorie bis zur Datenanalyse
- Mathematische Statistik: Architekt von Verfahren
- Daten: Rohstoffe für statistische Datenanalysen
- 3 Datenbeispiele:
  - (i) **1-Stichprobenbeispiel**: $x_1, \ldots, x_n$ Realisierungen von i.i.d. Zufallsvariablen $X_1, \ldots, X_n$.

    $(x_1, \ldots, x_n) =$
    $(16, 2, 1, 2, 1, 32, 1, 2, 1, 4, 2, 4, 1, 1, 1, 4, 1, 4, 1, 1, 1, 128,$
    $8, 2, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 2, 1, 512, 1, 1, 8, 2, 8, 256, \ldots$
    $\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$
    $1, 64, 4, 4, 1, 1, 1, 4, 1, 2, 1, 2, 4, 2, 2, 8, 1, 2, 1, 2, 4, 2, 1, 2).$

- Problem: $x_1, \ldots, x_n$ gegeben. Was kann über die Verteilung von $\sum_{i=1}^n X_i$ (oder $(\sum_{i=1}^n X_i - a)/b$) gesagt werden?
- Konfidenzintervalle, kritische Werte von Tests
- Sequentielle Mittel $k \mapsto \frac{1}{k}\sum_{i=1}^k x_i$ für $50 \leq k \leq n = 5000$



- Weitere Techniken der Datenanalyse...
- STOPP! Keine Vorabsuche nach Unregelmäßigkeiten in den Daten

$$\begin{aligned}&\text{Datenanalyse} \longleftrightarrow \text{Hypothesengenerierung}\\ &= \text{Reduktion auf eine propädeutische Vorstudie}\end{aligned}$$

- Lösungsansatz: Vorschlag aus der Praxis
  Bootstrapmethode (tool box)
- Ziel: Mathematische Analyse

- (ii) Medizinische Datensätze mit zensierten Daten:

(a) Kidney Daten
Infektionszeiten für 2 Gruppen von Dialyse Patienten

(b) 2 Gruppen von Zungenkrebspatienten mit unterschiedlichem DNA-Status (siehe Anhang)

Different catheterization procedures: percutant and surgical placements

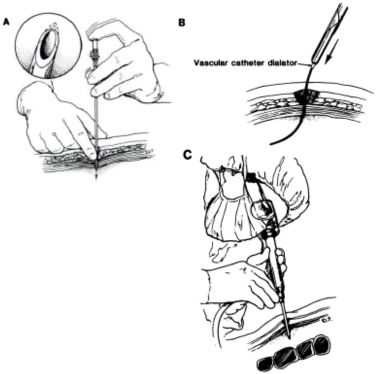


Figure 1. Technical considerations (see Text for details). (A) The tip of a 6-inch, 14-gauge angiocatheter is dulled (insert), and the catheter is used to enter the peritoneal cavity. (B) The peritoneoscopic guide, precannulated with a 7 French vascular catheter dilator, is advanced over a flexible guidewire. (C) Peritoneoscopic visualization of the peritoneal cavity.

Source: Klein/Moeschberger

TABLE 1.2

Times to infection (in months) of kidney dialysis patients with different catheterization procedures

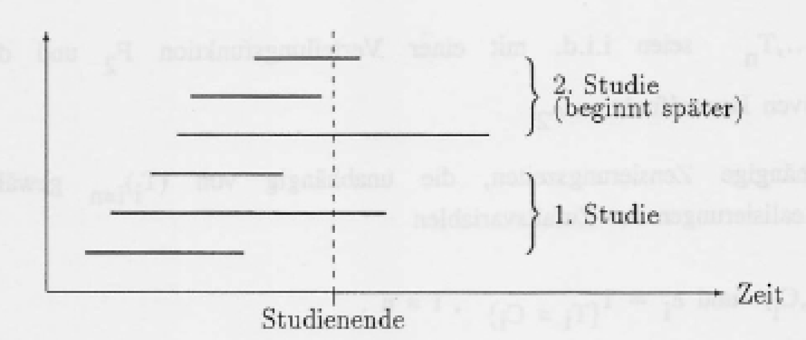
Surgically Placed Catheter
Infection Times: 1.5, 3.5, 4.5, 4.5, 5.5, 8.5, 8.5, 9.5, 10.5, 11.5, 15.5, 16.5, 18.5, 23.5, 26.5   *Censored Observations:* 2.5, 2.5, 3.5, 3.5, 3.5, 4.5, 5.5, 6.5, 6.5, 7 5, 7.5, 7.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 12.5, 13.5, 14.5, 14.5, 21.5, 21.5, 22.5, 22.5, 25.5, 27.5

Percutaneous Placed Catheter
Infection Times: 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 2.5, 2.5, 3.5, 6.5, 15.5   *Censored Observations:* 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.5, 1.5, 1.5, 1.5, 2.5, 2.5, 2.5, 2.5, 2.5, 3.5, 3.5, 3.5, 3.5, 3.5, 4.5, 4.5, 4.5, 5.5, 5.5, 5.5, 5.5, 5.5, 6.5, 7.5, 7.5, 7.5, 8.5, 8.5, 8.5, 9.5, 9.5, 10.5, 10.5, 10.5, 11.5, 11.5, 12.5, 12.5, 12.5, 12.5, 14.5, 14.5, 16.5, 16.5, 18.5, 19.5, 19.5, 19.5, 20.5, 22.5, 24.5, 25.5, 26.5, 26.5, 28.5

- Problem in der Praxis:
  unvollständige Daten, sogenannte zensierte Daten
- Beobachtet werden zwei Typen von Daten
  1. Vollständige Beobachtungen: Ausfallzeiten
  2. Unvollständige Beobachtungen: letzte beobachtete Lebenszeit, Ausscheiden aus der Studie aus anderen Gründen,
     z.B. Katheter fällt aus bevor es zur Infektion kommt oder keine Infektion bis zum Studienende
- **Datenstruktur:** $X = \min(T, C), \Delta = \mathbf{1}(T \leq C)$,
  T Überlebenszeit, C Zensierungsvariable (unabhängig).

Probleme:

- Die Zensierungsverteilung ist unbekannt
- Die Zensierung kann vom Gruppenstatus abhängen
- Zensierte Daten dürfen keinesfalls ignoriert werden



- (iii) Beispiel aus den Wirtschaftswissenschaften
“Dauer der Arbeitslosigkeit” (zensierte Daten).
Sind Frauen länger arbeitslos als Männer?
- Test- und Schätzprobleme

- normale Studie: 30-40 Patienten
  Diese Fallzahlen sind nur ausreichend, wenn die Studien effizient ausgewertet werden.
- Grundbegriff der medizinischen Statistik: Survival Funktion
- T Überlebenszeit (Zufallsvariable) (Eintrittszeit eines Ereignisses, z.B. Tod, Infektion, ...)
- $S(x) = P(T > x)$ **Survival Funktion** oder Überlebenskurve
  "Wahrscheinlichkeit, den Zeitpunkt x zu überleben"

$$\text{Studie} \longrightarrow \text{Daten} \longrightarrow \text{Schätzer } \hat{S} \text{ für } S$$

- 1. Beispiel: Zeit bis zur Infektion durch Katheter bei Dialysepatienten
  Dargestellt werden sogenannte Kaplan-Meier Schätzer für die Survivalfunktion, Kaplan und Meier (1958).

Survival functions for kidney dialysis patients with different catheterization procedures

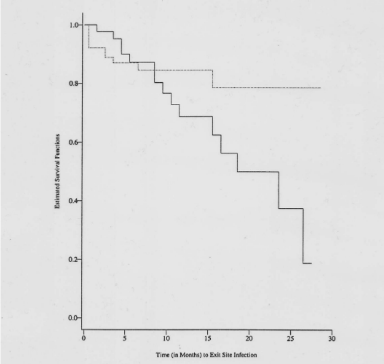


Figure 7.1 *Estimated (Infection-free) survival function for kidney dialysis patients with percutaneous (------) and surgical (——) placements of catheters.*

survival times given by randomly right censored data

Source: Klein/Moeschberger

- Kaplan and Meier (1958). Nonparametric estimation from incomplete observations. JASA $\approx$ 41.700 GS Zitate
- Mathematische Theorie: 80er Jahre mittels Martingalmethoden für Zählprozesse

Testprobleme

- Sind die Unterschiede zwischen den Gruppen signifikant?

  statistische Tests: belegen signifikante Unterschiede

  Insbesondere: handelt es sich bei den Daten
    - um Effekte, die durch zufällige Schwankungen zu erklären sind
    - oder sind die Effekte signifikant, d.h. so groß, dass diese nur durch systematische Abweichungen zu erkären sind?
- **Validität:** Wie vertrauenswürdig sind Ergebnisse statistischer Tests (p-Werte)?
- **Effizienz:** Wird die volle Information des Datensatzes verwendet?

$\Rightarrow$ hier: Das Standardverfahren (klassischer Log-Rang Test für proportionale Hazards, Cox Modelle) erkennt keine Unterschiede, p-Wert > 0.05

- Besonderheit: Kreuzende Survivalfunktion

Outline

- Modellbildung über Ausfallraten
- T Überlebenszeit mit Dichte f, $P(T > t) = \exp(-\int_0^t \lambda(u)du)$
- $\lambda(t) = f(t)/P(T \geq t)$ altersbedingte Ausfallrate, (*Hazardrate*)
- historisch: Cox-Modelle:

$\Rightarrow$ Cox (1972).
Regression models and life-tables. JRSS B $\approx$ 35.500 GS Zitate

- $\lambda_1/\lambda_2 =$ "zeitlich konstant"
- Schemper (Wien)
- Finanzmathematik: Schönbucher (2003): Credit Derivatives Pricing Models: Models, Pricing and Implementation (Wiley Finance Series)

- Beginn der Survival Analysis:
  Dissertation Aalen (1976) Berkely, Supervisor Le Cam

The Annals of Statistics
1978, Vol. 6, No. 4, 701–726

NONPARAMETRIC INFERENCE FOR A FAMILY OF COUNTING PROCESSES[1]

BY ODD AALEN

University of California, Berkeley and University of Copenhagen

Acknowledgment. The main results in this paper are from the author's Ph.D. dissertation written under the supervision of Professor Lucien Le Cam. I am also grateful to Jan M. Hoem, Søren Johansen, Niels Keiding and Mats Rudemo for helpful discussions. Freddy Christiansen has kindly permitted the use of data from his experiments.

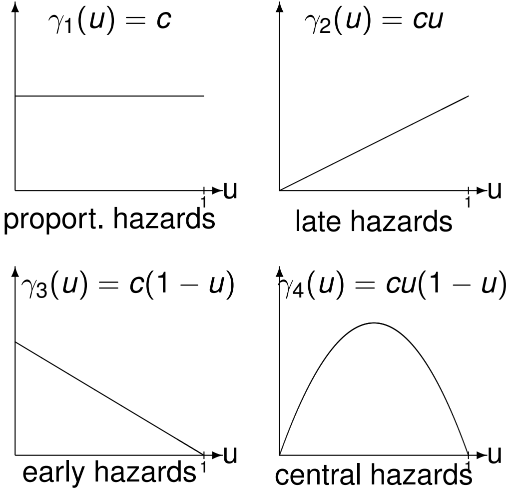
- 1006 Zitate bei Google.Scholar

- Modellbildung durch das relative Ausfallrisiko
  λ_1 altersbedingte Ausfallrate in Gruppe 1
  λ_2 altersbedingte Ausfallrate in Gruppe 2
- zeitabhängige relative Risiken für Ausfälle

$$\frac{\lambda_2(t)}{\lambda_1(t)} = 1 + \vartheta\gamma(F(t)),$$

F “baseline“ Verteilungsfunktion, Störparameter
ϑ reeller Parameter, $\gamma : [0,1] \to \mathbb{R}$ Gewichtsfunktion,
$\gamma_0 = 1$ zeitlich konstantes Risiko (Cox)

Examples of semiparametric models



Konkurrierende Testverfahren können unterschiedliche Qualität besitzen

- Beurteilung durch die Güte
  (Gütemaß: Fehlerwahrscheinlichkeit 2.Art, d.h. WS dafür, dass Unterschiede nicht erkannt werden)
- ARE (ψ): asmptotische relative Effizienz eines Testverfahrens ψ

$$\mathrm{ARE}(\psi) = \frac{N_{opt}}{N(\psi)}$$

$N(\psi)$ Anzahl der für ψ nötigen Beobachtungen, um eine vorgegebene Güte zu erreichen

N_{opt}=Minimum ($N(\varphi)$: φ alle Tests derselben Kategorie wie ψ)

"minimale Anzahl nötiger Beobachtungen"

Beispiel:
ARE(ψ)=$\frac{1}{2}$ doppelt so viele Beobachtungen sind für ψ nötig

Problem: Optimale Verfahren hängen vom Modell ab

Nichtparametrik: das genaue Modell ist unbekannt

- unendlich viele konkurrierende Modelle
- unendlich viele Tests ψ

Für diese Datensätze: der typische Log-Rang Test für proportionale Abweichungen der Hazards versagt

Grund: kreuzende Hazards liegen vor

Idee: schätze das Modell (adaptive Tests)

Mathematische Ergebnisse der Arbeitsgruppe (1988-2013)
Neuhaus (1993 Ann. Stat., 2000)

- Umsetzung des Vierpunkteprogramms von Le Cam:
  1. Konvergenz der Likelihoodprozesse (Experimente) unter lokalen Alternativen, lokal asymptotische Normalität (LAN), semiparametrisches Limesexperiment G (Gauß-Shift)
  2. Analyse des Limesexperiments G
  3. Gütevergleich von Tests für G
  4. Übertragung der Ergebnisse auf die Experimentenfolgen

- Zu jeder Gewichtsfunktion $\gamma : [0, 1] \to \mathbb{R}$ gehört ein asymptotisch optimaler Test ψ_γ (für Abweichungen des relativen Risikos in Richtung γ); Konstruktion mittels Martingalmethoden
- Berechnung der Qualität von Tests (einfachster Fall) mittels ARE

- J. (1991) Ann. Stat:
  Das Modell mit $\frac{\lambda_2(t)}{\lambda_1(t)} = 1 + \vartheta\gamma_1(F(t))$ liege vor, ein Test Ψ_{γ_2} für eine andere Richtung γ_2 werde verwendet. Dann gilt:

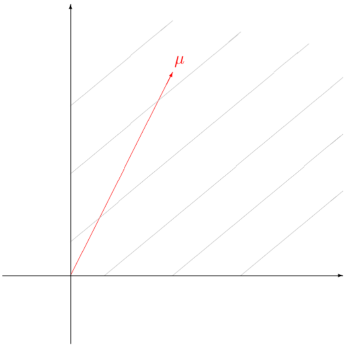
$$ARE(\Psi_{\gamma_2}) = \frac{\langle\gamma_1,\gamma_2\rangle^2}{\langle\gamma_1,\gamma_1\rangle\langle\gamma_2,\gamma_2\rangle} = \cos^2\beta,$$

  β Winkel zwischen den Vektoren γ_1 und γ_2,

$$\langle\gamma_1,\gamma_2\rangle = \int_0^1 \gamma_1(x)\gamma_2(x)\, d\eta_2(x).$$

  Nachteil: Ψ_{γ_2} wirkt hauptsächlich in Richtung γ_2

- Neues Verfahren: Bilde einen Kegel von Alternativen; verwende einen nichtparametrischen Likelihood-Ratio-Test
- Motivation $\mu = (\mu_1, \mu_2)$, $\{N(\mu, \Sigma) : \mu_i \geq 0\}$, $H_0 : \{\mu = 0\}$
- Projektionstest (schätzt die Richtung γ aus den Daten)
  gute Güte für ganze Kegel von Alternativen



- Behnen und Neuhaus (1989): Rank Tests with Estimated Scores and Their Application. B.G. Teubner, Stuttgart.

- Details: Brendel, Janssen, Mayer und Pauly (2013). Weighted logrank permutation tests for randomly right censored life science data. Erscheint in *Scand. J. Stat.*
- p-Werte für
  - Log Rang Test, $\gamma = 1$, Cox-Modell: p-Wert=0.0549
  - Projektionstest mit γ aus einem Kegel zur Erkennung der Risiken für frühe, späte, zentrale und proportionale Unterschiede der Hazards: p-Wert=0.010
- Festlegung der kritischen Werte durch Permutationstests

Outline

Typische Fragestellung für Zweistichprobenprobleme:
Ist die eine Behandlungsmethode der anderen überlegen?
(Hypothesen statistisch absichern)

Mathematische Methode: Martingale=faire Spiele

Eine Hypothese: es gibt keine Unterschiede in den Gruppen (Nullhypothese)

- durch eine zufällige (gleichverteilte) Zuweisung der Probanden in die beiden Behandlungsgruppen werden künstlich die Voraussetzungen für ein faires Spiel geschaffen!
- Unterschiede bewirken Abweichungen vom fairen Spiel
- Strasser und Weber (1999): „On the asymptotic theory of permutation statistics". Math. Methods Statist.

- Trick: Verwendung bedingter Permutationstests für <u>studentisierte</u> Statistiken unter Zensierung
  - finit gute Ergebnisse
  - asymptotisch valide
- Neuhaus (1993), Ann. Stat
- J. und Mayer (2001)
- J., Brendel, Mayer, Pauly (2013)

Einfachster Fall: Zweistichprobenproblem

- $X_i, Y_j : \Omega \longrightarrow \Omega'$
- $(\underbrace{X_1, \ldots, X_{n_1}}_{\text{Gruppe 1}}, \underbrace{Y_1, \ldots, Y_{n_2}}_{\text{Gruppe 2}}) = (Z_1, \ldots, Z_n) = \mathbf{Z}$
- $\widetilde{\mathcal{P}}_0 : (Z_1, \ldots, Z_n)$ austauschbar
- Ziel: verteilungsfreie Tests
- $T = T(\mathbf{Z})$ reelle Teststatistik vorgegeben
- $(Z_{i:n})_{i \leq n}$ Orderstatistiken

Permutationsverteilung :

$$P(T \in \cdot \,|(Z_{i:n})_{i \leq n} = (\omega_1, \ldots, \omega_n)) = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \epsilon_{T(\omega_{\pi_1}, \ldots, \omega_{\pi_n})}$$

Kritische Werte c_{perm} der Permutationsverteilung

$$\varphi_{n,perm} = \begin{cases} 1 & \left(\frac{n_1 n_2}{n}\right)^{1/2} T \begin{array}{c} > \\ \\ \leq \end{array} c_{perm} \\ 0 & \end{cases}$$

Güteuntersuchungen von Permutationstests:
Bester parametrischer Test (oracle test, benchmark test)

$$\varphi_n = \begin{cases} 1 & \left(\frac{n_1 n_2}{n}\right)^{1/2} T \begin{array}{c} > \\ \\ \leq \end{array} c_\alpha \\ 0 & \end{cases}$$

für parametrische Alternativen

Theorem

$$E_{P_0^n}\left[|\varphi_n - \varphi_{n,perm}|\right] \longrightarrow 0$$

kein Güteverlust unter lokalen Alternativen

J./ Pauls (2003) Ann. Stat.
How do permutation and bootstrap tests work?

Nullhypothese: $\widetilde{\mathcal{P}}_0 = \{P^n \ : \ P \in \mathcal{M}\}$ permutationsinvariant
häufig liegt eine **erweiterte Nullhypothese** $\mathcal{P}_0$ vor:
$\widetilde{\mathcal{P}}_0 \subset \mathcal{P}_0$

Beispiel 1: Behrens Fischer Problem (1929)

(a) Normalverteilungsmodell

- $\mathcal{L}(X_1) = \mathcal{N}(\mu_1, \sigma_1^2), \ \mathcal{L}(Y_1) = \mathcal{N}(\mu_2, \sigma_2^2)$
- $\mathcal{P}_0 = \{\mu_1 = \mu_2, \quad \sigma_1, \sigma_2 \text{ beliebig}\}$
- (verschiedene Messgenauigkeiten in den Gruppen)
- Welch-Test (Welch (1947), Pfanzagl (1974))

(b) Verallgemeinertes Behrens Fischer Problem

- $X_1 = \mu_1 + \sigma_1 Z_1$
- $Y_1 = \mu_2 + \sigma_2 \widetilde{Z}_1$
- $\mathrm{E}[Z_1] = 0$, $\mathrm{Var}[Z_1] = 1$
- $\mathcal{P}_0 = \{\mu_1 = \mu_2, \quad \sigma_1, \sigma_2 \text{ beliebig}\}$

Permutationstest von Pitman nur für $\sigma_1 = \sigma_2$ anwendbar

Zum verallgemeinerten Behrens Fischer Problem:

- Teststatistik

$$\widetilde{T}_n \sim \frac{\bar{X} - \bar{Y}}{\left(\frac{1}{n_1} S_X^2 + \frac{1}{n_2} S_Y^2\right)^{1/2}} = c_n \frac{T_n}{V_n^{1/2}}$$

- $Z_1 \sim \mathcal{N}(0, 1)$: Es gibt **keine exakten finiten Lösungen** $\rightarrow$ Welch Test *d* Metrik auf dem Raum der W'maße (schwache Konvergenz)

Permutiere studentisierte Statistiken:

Theorem (Verallg. Behrens Fischer, J. (1997))

Es gelte $\min(n_1, n_2) \to \infty$. *Dann folgt unter* H_o

$$\mathcal{L}_{perm}(\widetilde{T}_n) \overset{P-stoch.}{\longrightarrow} \mathcal{N}(0,1)$$

im Raum der W'maße $d(\mathcal{L}_{perm}(\widetilde{T}_n), \mathcal{N}(0,1)) \overset{P-stoch.}{\longrightarrow} 0.$

$\to$ kritische Werte

- Auf $\{\mu_1 = \mu_2, \sigma_1 \neq \sigma_2\}$ (asymptotische Ergebnisse $\forall \sigma_1, \sigma_2$)
- Auf $\widetilde{\mathcal{P}}_0 = \{\mu_1 = \mu_2, \sigma_1 = \sigma_2\}$ $\forall n$ exakte α-Tests (Permutationstests)

Weiterführende Literatur

- J. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. Stat. Probab. Lett. 36, 9-21.
- J. und Pauls (2003). How do Bootstrap and permutation test work? Ann. Stat.31, 768-806.
- J. (2005). Resampling Student's t-Type Statistics. Ann.Inst.Statist.Math. 57, 507-529.
- J. und Pauls (2005). A Monte Carlo Comparision of studentized bootstrap and permutation tests for heteroscedastic two-sample problems. Computational Statistics, 20, 369-383

Outline

- Permutationsstatistiken: Spezialfall von gewichteten Bootstrap-Statistiken, J. und Pauls (2003).
- Bootstrap: Efron (1979). Bootstrap methods: another look at the jackknife. Ann. Stat. - $\approx$ 9.700 GS Zitate

  Ziehe $X_1^*, \ldots, X_{k(n)}^*$ mit Zurücklegen aus $\mathbf{X}_{k(n)} := (X_1, \ldots, X_{k(n)})$
- Approximiere $\mathcal{L}(\sum_{i=1}^{k(n)} \frac{X_i}{b_n} - a_n)$ durch $\mathcal{L}(\sum_{i=1}^{k(n)} \frac{X_i^*}{b_n} - \overline{X}_{k(n)})$
- Aber Vorsicht: **Bootstrap** kann selbst im i.i.d. Fall **inkonsistent** sein!
- ...Zurück zum ersten Datensatz $x_1, \ldots, x_n$...

- $x_1, \ldots, x_n$ Realisierungen des St- Petersburg-Spiels (Bernoulli, 300 Jahre alt)

  $(x_1, \ldots, x_n) =$
  $(16, 2, 1, 2, 1, 32, 1, 2, 1, 4, 2, 4, 1, 1, 1, 4, 1, 4, 1, 1, 1, 128,$
  $8, 2, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 2, 1, 512, 1, 1, 8, 2, 8, 256, \ldots$
  $\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$
  $1, 64, 4, 4, 1, 1, 1, 4, 1, 2, 1, 2, 4, 2, 2, 8, 1, 2, 1, 2, 4, 2, 1, 2).$
- St. Petersburg-Spiel:

$$P(X_1 = 2^k) = 2^{-k} \text{ for } k \in \mathbb{N}$$

- $S_{k(n)} = \sum_{i=1}^{k(n)} X_i =$ Gewinnsumme nach $k(n)$ Spielen

- Kontinuum von Grenzverteilungen von $S_{r(n)} = \sum_{i=1}^{r(n)} X_i$

Theorem (Martin-Löf, 1985; S. Csörgő und Dodunekova, 1991; S. Csörgő, 2002, 2007, 2010)

Es gilt

$$\frac{S_{r(n)}}{r(n)} - \log_2(r(n)) \xrightarrow{d} W_\gamma$$

entlang Teilfolgen $r(n) \to \infty$ *mit*

$$\exp(2\pi i\{\log_2(r(n)) - \lfloor \log_2(r(n)) \rfloor\}) \longrightarrow \exp(2\pi i\{1 + \log_2(\gamma)\})$$

für ein $1/2 < \gamma \leq 1$. *Dabei gilt* $\mathcal{L}(W_{\gamma_1}) \neq \mathcal{L}(W_{\gamma_2})$ *für* $\gamma_1 \neq \gamma_2$.

- Problem $E(X_1) = \infty$ (heavy tails)

- Ginè und Zinn (1989): Bootstrap versagt bei heavy tails.
- Vorschlag: Swanepoel (1986), S. Csörgő und Mason (1989): Reduziere den Bootstrap-Stichprobenumfang, falls X_i im Anziehungsbereich einer stabilen Verteilung liegt:

$$m(n): \quad \frac{m(n)}{k(n)} \longrightarrow 0 \quad \sum_{i=1}^{m(n)} X_i^*$$

- Untersuchung u.a. in del Barrio, Janssen und Matran (2009a, b) sowie del Barrio, Janssen und Pauly (2013).

- Allgemein: $X_1, X_2, \ldots$ i.i.d. mit

$$\sum_{i=1}^{k(n)} \frac{X_i}{b_n} - a_n \xrightarrow{d} Y(\neq \text{konstant})$$

- $\mathcal{P} = \{Z : \sum_{i=1}^{r(n)} \frac{X_i}{\beta_n} - \alpha_n \xrightarrow{d} Z\}$

Theorem (Vorsicht: Bootstrap!!)

$X_1^*, \ldots, X_{m(n)}^*$ Bootstrap-Stichprobe aus $X_1, \ldots, X_{k(n)}$. Für alle $Z \in \mathcal{P}$ gibt es $\frac{m(n)}{k(n)} \to 0$, α_n', β_n' mit

$$\sum_{i=1}^{m(n)} \frac{X_i^*}{\beta_n'} - \alpha_n' \xrightarrow{d} Z.$$

Anwendungen

St Petersburg-Spiel: $k(n) = 2^n$ Grenzwertsatz von Martin-Löf

- $\frac{m(n)}{k(n)} \to 0$ Bootstrap erreicht alle Häufungspunkte
- $\{m(n)\} \subset \{2^n\}$, so wird die Martin-Löf Grenzverteilung reproduziert

„noch schlimmer"

Wolfgang Doeblin: Es gibt i.i.d. Folge $(X_i)_{i \in \mathbb{N}}$ mit
$\mathcal{P} = \{$alle unendlich teilbaren Zuffallsvariablen$\}$

- $\frac{m(n)}{k(n)} \to 0$ Bootstrap kann alle diese reproduzieren

Positives Resultat:

Theorem

Sei $S_{k(n)}/b_n - a_n \xrightarrow{d} Y$. Für alle $m(n)$-Bootstrap Stichproben, $\frac{m(n)}{k(n)} \to 0$, aus $(X_1, \ldots, X_{k(n)})$ gilt:
Ist $\sum_{i=1}^{m(n)} \frac{X_i^*}{\beta_n} - \alpha_n \xrightarrow{d} \xi$ kovergent und lässt sich $\mathcal{L}(Y)$ aus ξ durch Transformation $Y \stackrel{d}{=} \frac{\xi}{b} - a$ reproduzieren, so ist $\mathcal{L}(Y)$ eine stabile Verteilung.

Koautoren: Projekte Survival Analysis / Resampling

Eustasio del Barrio	Hartmut Milbrodt
Jan Beyersmann	Georg Neuhaus
Michael Brendel	Thorsten Pauls
Stefan Brenner	Markus Pauly
Holger Hebben	Jörg Rahnenführer
Andreas Knoch	Dominik Völker
Carlos Matrán	Stefan Wellek
Claus-Dieter Mayer	Wiebke Werft

und viele Diplomanden.

Aktuelle Projekte: Anwendungen von Permutationstests in der Bioinformatik

Literatur

del Barrio, E.; Janssen, A.; Matran, C. (2009). *On the low intensity bootstrap for triangular arrays of independent identically distributed random variables.* TEST 18, 283-301

del Barrio, E.; Janssen, A.; Matran, C. (2009). *Resampling schemes with low resampling intensity and their applications in testing hypotheses.* J. Statist. Plann. Inference 139, 184-202

Brendel, M.; Janssen, A.; Mayer, C.-D.; Pauly, M. (2013). *Weighted logrank permutation tests for randomly right censored life science data.* Erscheint bei Scand. J. Stat.

Janssen, A.; Pauls, T. (2005). *A Monte Carlo comparison of studentized bootstrap and permutation tests for heteroscedastic two-sample problems.* Comput. Statist. 20, 369-383

Janssen, A. (2005). *Resampling Student's t-type statistics.* Ann. Inst. Statist. Math. 57, 507-529

Janssen, A.; Werft, W. (2004). *A survey about the efficiency of two-sample survival tests for randomly censored data.* Mitt. Math. Sem. Giessen 254, 1-47

Literatur

Janssen, A.; Pauls, T. (2003). *How do bootstrap and permutation tests work?* Ann. Stat. 31, No. 3, 768-806

Janssen, A.; Mayer, C.-D. (2001). *Conditional Studentized Survival Tests for Randomly Censored Models.* Scand. J. Stat. 28, No. 2, 283-293

Janssen, A. (1997). *Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem.* Statist. Probab. Lett. 36, No.1, 9-21

Janssen, A. (1991). *Conditional rank tests for randomly censored data.* Ann. Stat. 19, No. 3, 1434-1456

Neuhaus, G. (1993). *Conditional rank test for the two-sample problem under random censorship.* Ann. Stat. 21, 1760-1779

Neuhaus, G. (2000). *A method of constructing rank tests in Survival Analysis.* J.Stat.Plan.Inf. 91, 481-497.

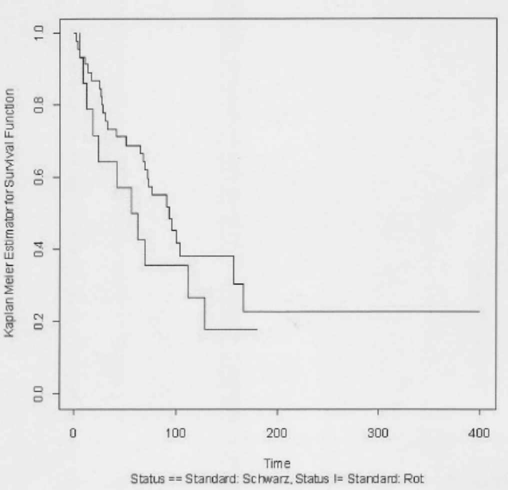
Vielen Dank für Ihre Aufmerksamkeit!

Beispiel: Überlebenszeiten von Patienten
Tumore in der Zunge

Unterschiede: DNA - Gehalt der Zellen

Gruppe 1: diploide Zellen
Gruppe 2: aneuploide Zellen (verändert)

Ist das Merkmal “aneuploide Zellen” ein signifikantes prognostisches Merkmal für die Überlebenszeit?


Kaplan Meier Estimator for Survival Function
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
Time
Status == Standard: Schwarz, Status != Standard: Rot

Tongue Data

test	p-value
(logrank test)	0.0832526
projection test	0.09