Nonparametric and Resampling Methods

Lukas Meier

18.01.2016

Overview

- Nonparametric tests
- Randomization tests
- Asymptotic approximations of estimators
- Jackknife and Bootstrap

Nonparametric Tests

Introduction

- Up to now we basically always used a **parametric family**, like the normal distribution $\mathcal{N}(\mu, \sigma^2)$ for modeling random data.
- Based on observed data $x_1, \ldots, x_n$ we were able to
  - calculate **parameter estimates** $\widehat{\mu}, \widehat{\sigma}^2$.
  - derive **confidence intervals** for the parameters.
  - perform **tests**, e.g. $H_0 : \mu = \mu_0, H_A : \mu \neq \mu_0$.
- All calculations were based on the assumption that the observed data comes from the corresponding parametric family.
- This is a **strong** assumption that has to be checked, e.g. by using QQ-plots.

- Why did we need that parametric family/model?
- Remember: a **probability model** describes our ideas about the possible **outcomes** of a random variable and the corresponding **probabilities**.
- The model is needed for determining the **statistical uncertainty** of an estimate or for deriving the rejection region of a test.
- Remember: Derivation of rejection region for Z-Test (based on normal assumption!).
- Wouldn't it be nice to have a method that would need (basically) **no parametric assumptions**?

Nonparametric Statistics

- The field of **nonparametric statistics** is about methods that do **not** make parametric assumptions about the data generating process.
- It includes (among others):
  - Nonparametric **regression** models
  - Nonparametric **tests** / **distribution free tests**
  - Nonparametric **density estimation**
  - ...

Example: Nonparametric Regression

- Regression model so far

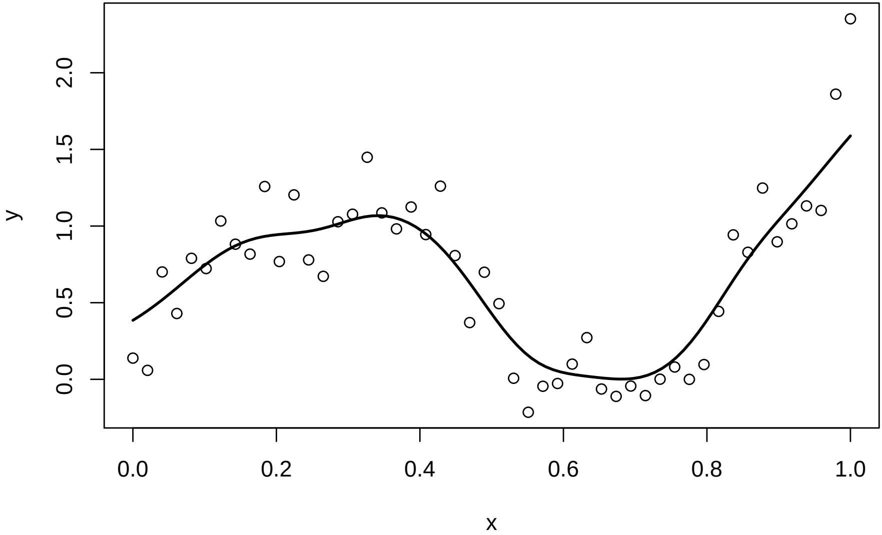
$$Y_i = \underbrace{\beta_0 + \beta_1 x_i^{(1)} + \cdots + \beta_p x_i^{(p)}}_{\text{structural part}} + \underbrace{E_i}_{\text{random part}}$$

- The structural part can be written as a function $h(x_i; \beta)$.

- Nonparametric regression relaxes the assumptions about the function h. Can be any "smooth" function, "no formula required".

- See picture on next slide.

- This will be treated in the block course "Nonparametric Regression".



y
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
x

Nonparametric Tests

- Nonparametric tests are tests that do **not** assume that the random variables follow a parametric family (normal, exponential, ...).
- More precise term: **distribution free tests**.
- Distribution free also means: The test statistic has the same distribution for all distributional assumptions for the observations.
- We will have a look at
  - Rank-based tests
  - Rank correlation
  - Goodness-of-fit tests

Rank-Based Tests: Overview

- Remember: rank = position of observation in **ordered** sample.
- Therefore, the ranks are **always** the numbers 1 to n (with the exception of ties)
- Overview
  - Signed-rank test of **Wilcoxon** (as seen in introductory course) for a simple sample or for paired samples.
  - Rank-sum test of **Wilcoxon, Mann and Whitney (Mann Whitney U-Test)** for 2 independent samples.
  - Extensions for ANOVA
    - one-way: **Kruskal-Wallis test**
    - block designs: **Friedman test**
  - Rank correlations

Signed-Rank Test of Wilcoxon

- First, back to basics: **sign-test** (see blackboard)
- Observations x_i are interpreted as i.i.d. realizations of random variables X_i, $i = 1, \ldots, n$.
- Assumption: $X_i \sim \mathcal{F}$, where $\mathcal{F}(\mu)$ is a **continuous** and **symmetric** distribution with median (=mean) μ.
- No parametric assumption such as normal, exponential, log-normal, etc. is needed here.
- $H_0 : \mu = \mu_0$, $H_A : \mu \neq \mu_0$ (or one-sided alternative)

Signed-Rank Test of Wilcoxon: Details

- Determine differences to "reference value" μ_0: $D_i = X_i - \mu_0$.
- Calculate absolute values: $|D_1|, \ldots, |D_n|$.
- Replace by the corresponding ranks $R_i = \text{rank}(|D_i|)$
- Test statistic = sum of all R_i's that have a corresponding *positive* D_i.

- The distribution of the test statistic under H_0 does not depend on the distribution $\mathcal{F}$.
- Why? The sign of D_i does **not** depend on $\mathcal{F}$ (under H_0).
- The ranks are always the numbers 1 to n.
- Due to the symmetry assumption, "the left and the right-hand side should show the same mixture of ranks" (under H_0).
- I.e., under H_0 we can go through the numbers 1 to n and "toss a coin" whether it goes to "the left or the right-hand side of μ_0". This determines the distribution under H_0!
- Check rejection region in table or use computer.

Signed-Rank Test of Wilcoxon: Toy-Example

- $H_0 : \mu = 9, H_A : \mu \neq 9.$

- Data and calculations:

i	1	2	3	4	5	6	7	8
x_i	7.0	13.2	8.1	8.2	6.0	9.5	9.4	8.7
d_i	−2.0	**4.2**	−0.9	−0.8	−3.0	**0.5**	**0.4**	−0.3
$\lvert d_i \rvert$	2.0	4.2	0.9	0.8	3.0	0.5	0.4	0.3
r_i	6	**8**	5	4	7	**3**	**2**	1

- Value of test statistic:

$$8 + 3 + 2 = 13.$$

- Rejection region is (see e.g. table in book of Stahel)

$$\{0, 1, 2, 3\} \cup \{33, 34, 35, 36\}$$

Signed-Rank Test of Wilcoxon: R

- In R: `wilcox.test(x, mu = 9)`
- Output:

```
Wilcoxon signed rank test
data:  x
V = 13, p-value = 0.5469
alternative hypothesis: true location is not equal to 9
```

Signed-Rank Test of Wilcoxon: Technical Details

- If there are **ties** (in the $|d_i|$'s), we assign each of the elements the average of the applicable ranks.
- If there are **zeros** (in the $|d_i|$'s), we **ignore** them and adjust the sample-size accordingly.

Mann-Whitney U-Test

- Test to compare two **independent** samples (like two-sample t-test)
- Data: $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ (two groups).
- $H_0 : X_i, Y_j$ i.i.d. $\sim \mathcal{F}$ where $\mathcal{F}$ is **any** continuous distribution, the **same** for all observations.
- $H_A : X_i \sim \mathcal{F}_X, Y_j \sim \mathcal{F}_Y$, where

$$F_X(x) = F_Y(x - \delta).$$

  I.e. the distribution functions are simply **shifted** versions of each other ("location shift").

Mann-Whitney U-Test: Details

- **Combine** data from both groups.
- Replace every observation with its rank (with respect to the **combined** data).
- Calculate sum of ranks of each group $\rightsquigarrow S^{(1)}, S^{(2)}$.
- Define $T^{(k)} = S^{(k)} - n_k(n_k + 1)/2$, $k = 1, 2$.
- Use
$$U = \min\left\{T^{(1)}, T^{(2)}\right\}$$
as test statistic (R uses $T^{(1)}$).
- Compare with table.

Mann-Whitney U-Test: Toy-Example

- Data

i	1	2	3	4	5	6	7	8
x_i	6.2	9.9	7.3	6.4	10.3	11.1	10.6	10.0

j	1	2	3	4	5	6	7	8	9	10
y_j	7.4	11.7	6.7	11.0	8.1	6.5	4.3	10.2	10.7	7.1

- Ranks with respect to combined groups:

i	1	2	3	4	5	6	7	8
r_{x_i}	2	10	7	3	13	17	14	11

j	1	2	3	4	5	6	7	8	9	10
r_{y_j}	8	18	5	16	9	4	1	12	15	6

- Rank sum group X: $S^{(1)} = 2 + 10 + \cdots + 11 = 77$
  Rank sum group Y: $S^{(2)} = 8 + 18 + \cdots + 6 = 94$.

Mann-Whitney U-Test: Toy-Example

- $T^{(1)} = S^{(1)} - 8 \cdot 9/2 = 77 - 36 = 41$
  $T^{(2)} = S^{(2)} - 10 \cdot 11/2 = 94 - 55 = 39.$
- $U = \min\{41, 39\} = 39.$
- Test decision: See R-output on next slide.

Mann-Whitney *U*-Test: R

- In R: `wilcox.test(x, y, paired = FALSE)`
- Output:

```
Wilcoxon rank sum test

data:  x and y
W = 41, p-value = 0.9654
alternative hypothesis: true location shift is not
equal to 0
```

Kruskal-Wallis Test

- What if we have $g > 2$ groups? I.e., if we are in the **one-way ANOVA** situation?

- There, the typical (global) null-hypothesis is

  H_0: All groups come from the same (continuous) distribution.
  H_A: At least one group has a shifted distribution function.

- The same idea as in the previous examples is applicable.

- We "pool" all our data together and replace it with the corresponding ranks.

- The test is known as **Kruskal-Wallis Test**.

Kruskal-Wallis Test: Details

- Combine all observations of all g groups.
- Calculate the corresponding ranks (with respect to the **combined** sample).
- Sum up ranks in each group: $\rightsquigarrow S^{(1)}, \ldots, S^{(g)}$.
- Calculate test statistic that is based on $S^{(1)}, \ldots, S^{(g)}$ (without details).

Kruskal-Wallis Test: R

- In R:
  ```
  kruskal.test(Ozone ~ Month, data = airquality)
  ```
- Output:
  ```
  Kruskal-Wallis rank sum test

  data:  Ozone by Month
  Kruskal-Wallis chi-squared = 29.2666, df = 4,
  p-value = 6.901e-06
  ```
- Also have a look at boxplot...

Friedman Test

- Consider the case of an (unreplicated) **complete block design**.
- Example: Sales of 5 products (A to E) in 7 stores

	A	B	C	D	E
1	5	4	7	10	12
2	1	3	1	0	2
3	16	12	22	22	35
4	5	4	3	5	4
5	10	9	7	13	10
6	19	18	28	37	58
7	10	7	6	8	7

- Here: blockfactor = stores, treatment factor = products.

- The corresponding model is

$$Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$$

  $i = 1, \dots, a$ (treatments), $j = 1, \dots, b$ (blocks).

- $H_0 : \alpha_i = 0$ for *all* i ("treatment has **no effect**")
  H_A: at least one $\alpha_i \neq 0$.

- If the products show a consistent difference, the rankings **within** the blocks (= stores) are expected to be similar.

- E.g., in every store the same product is best-selling.

- On the other side, if there is no "product effect", the rankings within each store are **random**.

- This is the idea of the so-called **Friedman Test**.

Friedman Test: Details

- Replace Y_{ij} with the corresponding rank **within** the same block $\rightsquigarrow R_{ij}$.

- Sum up ranks for every level of the treatment factor (= products) $\rightsquigarrow S_k$.

- Define

$$U = \sum_{k=1}^{a} \left(S_k - b \cdot \underbrace{(a+1)/2}_{\text{avg. rank of treatm.}} \right)^2 .$$

- Standardize

$$T = \frac{12}{b \cdot a \cdot (a+1)} U$$

- Fact: $T \approx \chi^2_{a-1}$ under H_0.

Friedman Test: R

- R-function: `friedman.test`
- Use formula interface (see help file) or a data matrix.
- If a matrix is used in the call, the **treatments** are given by the different **columns** and the **blocks** by the different **rows** (check details in help file).
- `friedman.test(m)`
- Output:

```
Friedman rank sum test

data:  m
Friedman chi-squared = 8.3284, df = 4, p-value = 0.08026
```

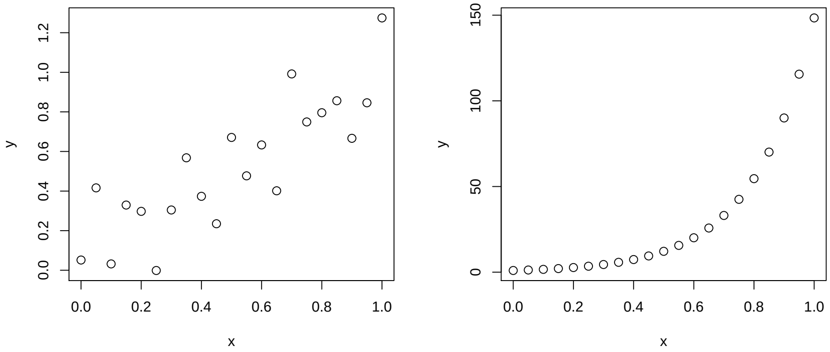
More Complex Designs

- Ideas also applicable for more complex ANOVA-designs.
- Methods not very common...
- See also approaches with randomization tests (later).
- So far for tests...what about **measures of dependence**?

Rank Correlation

- Ranks can also be used to describe the **relationship** between two variables.
- Remember: "Ordinary" (Pearson) correlation measures the strength of the **linear relationship** between two variables.
- **Rank correlation of Spearman** is defined as the Pearson correlation of the **rank-transformed data**.
- I.e., replace every observation with its rank and calculate Pearson correlation. That's all!

- In R: `cor(x, y, method = "spearman")`
- If we transform one or both variables with a monotone transformation, the Spearman correlation **will not change!**
- Hence, it measures the strength of a **monotone relationship** between two variables.
- See example on next slide.



- Pearson: Left: 0.84 / Right: 0.85
- Spearman: Left: 0.84 / Right: 1

Goodness-of-Fit Tests

- Problem in parametric statistics: Are the data compatible with the assumption about their distribution?
- Good approach: **Visual inspection** (QQ-plots) instead of formal tests.
- QQ-plot: Empirical quantiles vs. theoretical quantiles.
- Theoretical quantiles based on model distribution $F(x)$.
- Empirical quantiles based on empirical distribution function

$$\widehat{F}_n(x) = \frac{1}{n} \# \{ i \mid X_i \leq x \} .$$

Kolmogorov-Smirnov Test

- We can also try to use a statistical test.
- H_0 : Data comes from a certain pre-specified distribution.
  H_A: Data comes from (any) other distribution.
- I.e., we are a situation where we are trying to "prove" H_0.
- If we can't reject H_0 we **cannot** be confident that H_0 really holds. The test might simply have low power.
- The **Kolmogorov-Smirnov test** is based on the comparison of the two distribution functions

$$T = \max_x \left\{ \left| \widehat{F}_n(x) - F(x) \right| \right\}.$$

- The distribution of T under H_0 does not depend on the distribution of F (if F is continuous).
- Reason: We can apply any monotone transformation and the result will not change.
- The Kolmogorov-Smirnov test is interesting from a theoretical point of view, but typically has **low power**.
- Hence, if it does **not** reject, we can basically say **nothing!**
- If it rejects, we can of course believe H_A, as the type I error rate is controlled (as usual).
- Alternative: χ^2-test (typically more powerful).

Kolmogorov-Smirnov Test: R

- In R: `ks.test`
- `ks.test(x, "pnorm", mean = 0.5, sd = 1)`
- Output:

```
One-sample Kolmogorov-Smirnov test

data:  x
D = 0.3347, p-value = 0.01687
alternative hypothesis: two-sided
```

Kolmogorov-Smirnov Test: Problems

- Problem: The parameters of the theoretical distribution have to be **pre-specified** and should **not** be estimated from the data (!).
- E.g., we have to specify the values of μ and σ^2 for the normal distribution.
- From the help file of `ks.test`:

  ```
  If a single-sample test is used, the parameters specified
  in '...' must be pre-specified and not estimated from the
  data.  There is some more refined distribution theory for
  the KS test with estimated parameters (see Durbin, 1973),
  but that is not implemented in 'ks.test'.
  ```

- Not a problem with normalplots...

Summary

- Overview of discussed rank-based tests:

	2 groups	More than two groups
paired	Wilcoxon	Friedman
unpaired	Mann-Whitney U	Kruskal-Wallis

- Use rank correlation to measure strength of monotone dependence.
- Kolmogorov-Smirnov Test not very powerful.