Efficient tests for the similarity of dose response curves

Kathrin Möllenhoff, Ruhr-Universität Bochum
Holger Dette, Ruhr-Universität Bochum

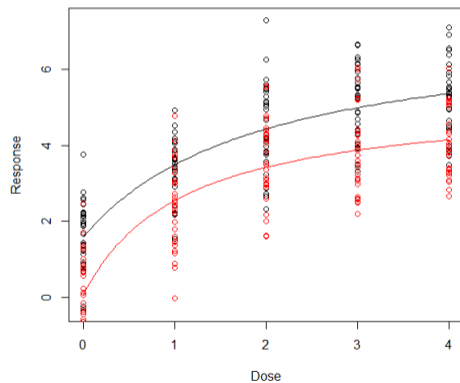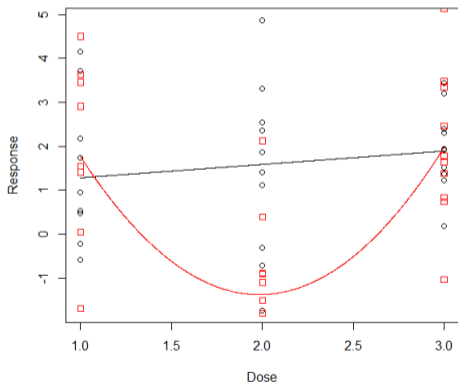Utrecht, 2015

Outline

Comparing curves - The setting

Comparing curves - The setting II

- Two dose response curves (from two samples)

$$Y_{ijk} = m_i(x_{ij}, \vartheta_i) + \varepsilon_{ijk} \; ; \quad i = 1, 2;$$
$$j = 1, \dots, \ell_i \; ; \quad k = 1, \dots, n_{ij},$$

- ε_{ijk} independent $\sim \mathcal{N}(0, \sigma_i^2)$ $\quad (i = 1, 2)$
- $x_{ij} \in \mathcal{X}$ (dose levels)
- $\vartheta_i \in \mathbb{R}^{d_i}$ (parameter in model m_i)

Problem: Are the dose response curves m_1 and m_2 similar?

Comparing curves - The setting II

- Two dose response curves (from two samples)

$$Y_{ijk} = m_i(x_{ij}, \vartheta_i) + \varepsilon_{ijk} \; ; \quad i = 1, 2;$$
$$j = 1, \ldots, \ell_i \; ; \quad k = 1, \ldots, n_{ij},$$

- ε_{ijk} independent $\sim \mathcal{N}(0, \sigma_i^2)$ $(i = 1, 2)$
- $x_{ij} \in \mathcal{X}$ (dose levels)
- $\vartheta_i \in \mathbb{R}^{d_i}$ (parameter in model m_i)

Problem: Are the dose response curves m_1 and m_2 similar?

Measures of similarity

- d: a *metric* measuring the distance between m_1 and m_2.
- Hypothesis of similarity:

$$H_0 : d(m_1, m_2) > \Delta \quad \text{versus} \quad H_1 : d(m_1, m_2) \leq \Delta$$

(here Δ is a pre-specified constant).

- Examples
    - maximum difference

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m_2(x, \vartheta_2)|$$

    - L_2-distance

$$d_2(m_1, m_2) = \Big(\int_{\mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2))^2 dx \Big)^{1/2}$$

Measures of similarity

- d: a *metric* measuring the distance between m_1 and m_2.
- Hypothesis of similarity:

$$H_0 : d(m_1, m_2) > \Delta \quad \text{versus} \quad H_1 : d(m_1, m_2) \le \Delta$$

(here Δ is a pre-specified constant).

- Examples
  - maximum difference

  $$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m_2(x, \vartheta_2)|$$

  - L_2-distance

  $$d_2(m_1, m_2) = \Big(\int_{\mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2))^2 dx \Big)^{1/2}$$

Example

- EMAX and Log-linear model ($\mathcal{X} = [0,1]$)

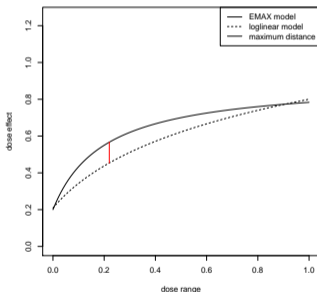$$m_1(x, \vartheta_1) = 0.2 + \frac{0.7x}{0.2 + x}, \quad m_2(x, \vartheta_2) = 0.74 + 0.33 \log(x + 0.2)$$

Figure: *EMAX and log-linear model.*

**A conservative test for similarity of curves
(Gsteiger, Bretz, Liu (2011))**

Testing for similarity (Gsteiger et al. (2011))

- Use the maximum distance

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m_2(x, \vartheta_2)|$$

- Construct a uniform confidence band for the difference

$$m_1(\cdot, \vartheta_1) - m_2(\cdot, \vartheta_2)$$

  by a bootstrap approach

- Reject H_0 (i.e. decide for similarity) if the confidence band is contained in the rectangle $\mathcal{X} \times [-\Delta, \Delta]$

Example

- Log-linear and EMAX model ($\mathcal{X} = [0,1]$)

$$m_1(x, \vartheta_1) = 0.2 + \frac{0.7x}{0.2 + x}, \quad m_2(x, \vartheta_2) = 0.74 + 0.33 \log(x + 0.2)$$
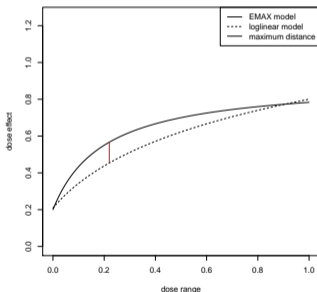
Figure: *EMAX and log-linear model.*

Testing for similarity (Gsteiger et al. (2011))

- Simulated confidence bands ($n_1 = n_2 = 100$; 100 runs)

Figure: *Confidence bands for the difference of the EMAX and loglinear model.*

Example
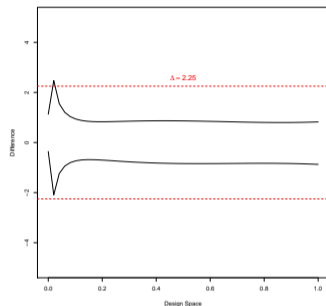
- $\Rightarrow$ $H_0 : d_\infty(m_1, m_2) < 2.25$ can not be rejected!
- Simulated power for the hypothesis

$$H_0 : \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m_2(x, \vartheta_2)| > 3$$

  is 0.15.
- **The test does not have any power** !

Efficient tests for similarity
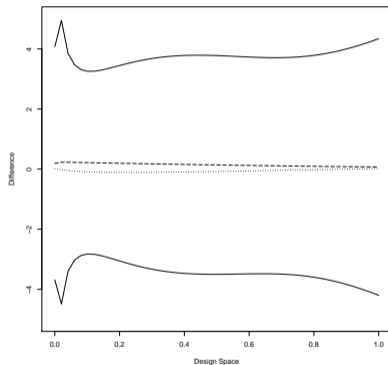
Tests based on distances

Basic idea: Estimate the distance between m_1 and m_2 directly and decide for similarity for small values of the resulting estimate

- (parametric) estimates of m_1 and m_2:

$$\hat{m}_1 = m_1(\cdot, \hat{\vartheta}_1), \ \hat{m}_2 = m_2(\cdot, \hat{\vartheta}_2)$$

- estimate of the distance between m_1 and m_2:

$$\hat{\boldsymbol{d}} = d(\hat{m}_1, \hat{m}_2)$$

Tests based on distances

$\hat{\boldsymbol{d}}$: estimate the distance between m_1 and m_2

- Decide for similarity, i.e. reject the hypothesis

$$H_0 : d(m_1, m_2) > \Delta$$

whenever

$$\boxed{\hat{\boldsymbol{d}} = d(\hat{m}_1, \hat{m}_2) \le q}$$

- **Problem:** how do we find the critical value q?

Estimating the distance between two curves

- L_2-distance
$$\hat{\mathbf{d}}_2 = d_2(\hat{m}_1, \hat{m}_2) = \Big(\int_{\mathcal{X}} (m_1(x, \hat{\vartheta}_1) - m_2(x, \hat{\vartheta}_2))^2 dx \Big)^{1/2}$$

- Empirical process theory $(n_1, n_2 \to \infty)$
$$d_2(\hat{m}_1, \hat{m}_2) - d_2(m_1, m_2) \overset{a}{\sim} \mathcal{N}\Big(0, \frac{\tau^2}{n_1 + n_2}\Big)$$

where

$$\tau^2 = \int_{\mathcal{X} \times \mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2)) k_{\lambda_1, \lambda_2}(x, y)$$
$$\times (m_1(y, \vartheta_1) - m_2(y, \vartheta_2)) dx\, dy$$

Estimating the distance between two curves

Testing for similarity of two dose response curves:

- If $\hat{\tau}^2$ is an estimate of τ^2, then

$$H_0 : d_2(m_1, m_2) > \Delta$$

  is rejected, whenever

$$d_2(\hat{m}_1, \hat{m}_2) \leq \Delta + \frac{\hat{\tau}}{\sqrt{n_1 + n_2}} u_\alpha$$

- **It is very difficult to estimate τ^2, especially for small sample sizes**

Bootstrap test for similarity

Generate data under the **null hypothesis**

- Estimate ϑ_1, ϑ_2 under the restriction

$$H_0 : \Big(\int_{\mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2))^2 dx \Big)^{1/2} > \Delta$$

$$\rightsquigarrow \quad \textbf{constrained} \text{ ML-estimates } \ \hat{\hat{\vartheta}}_1, \hat{\hat{\vartheta}}_2$$

- Generate bootstrap data (parametric bootstrap)

$$Y_{ijk}^* = m_i(t_{ij}, \hat{\hat{\vartheta}}_i) + \hat{\sigma}_i \varepsilon_{ijk}^*$$

where ε_{ijk}^* are i.i.d. $\sim \mathcal{N}(0, 1)$

Bootstrap test for similarity

Generate data under the **null hypothesis**

- Estimate ϑ_1, ϑ_2 under the restriction

$$H_0 : \Big(\int_{\mathcal{X}} (m_1(x, \vartheta_1) - m_2(x, \vartheta_2))^2 dx \Big)^{1/2} > \Delta$$
$$\rightsquigarrow \text{ \textbf{constrained} ML-estimates } \ \hat{\hat{\vartheta}}_1, \hat{\hat{\vartheta}}_2$$

- Generate bootstrap data (parametric bootstrap)

$$Y_{ijk}^* = m_i(t_{ij}, \hat{\hat{\vartheta}}_i) + \hat{\sigma}_i \varepsilon_{ijk}^*$$

where ε_{ijk}^* are i.i.d. $\sim \mathcal{N}(0, 1)$

Bootstrap test for similarity

Bootstrap test:

- Calculate $\hat{\vartheta}_1^*, \hat{\vartheta}_2^*$ from the bootstrap data Y_{ijk}^*
- Calculate

$$\hat{\mathbf{d}}_2^* = \Big(\int_{\mathcal{X}} (m_1(x, \hat{\vartheta}_1^*) - m_2(x, \hat{\vartheta}_2^*))^2 dx \Big)^{\frac{1}{2}}$$

- Repeat this procedure B times
- If $\hat{\mathbf{d}}_2^{*(1)}, \ldots, \hat{\mathbf{d}}_2^{*(\mathbf{B})}$ denote the ordered bootstrap replicates the hypothesis

$$H_0 : d_2(m_1, m_2) > \Delta$$

is rejected, whenever

$$\hat{\mathbf{d}}_2 \leq \ \hat{\mathbf{d}}_2^{*(\lfloor \mathbf{B}\alpha \rfloor)}$$

Bootstrap test for similarity

- Theoretical properties:
  - ▸ the bootstrap test has asymptotic level α
  - ▸ the bootstrap test is consistent

- More precisely: for increasing sample sizes $n_1, n_2 \to \infty$

$$d_2(m_1, m_2) = \Delta \ \rightsquigarrow \quad \mathbb{P}(\text{"rejection"}) \ \to \alpha$$
$$d_2(m_1, m_2) > \Delta \ \rightsquigarrow \quad \mathbb{P}(\text{"rejection"}) \ \to 0$$
$$d_2(m_1, m_2) < \Delta \ \rightsquigarrow \quad \mathbb{P}(\text{"rejection"}) \ \to 1$$

Finite sample properties

- $m_1(x, \vartheta_1) = \delta + \frac{5x}{x+1}, \qquad m_2(x, \vartheta_2) = \frac{5x}{1+x}, \qquad \mathcal{X} = [0, 4]$
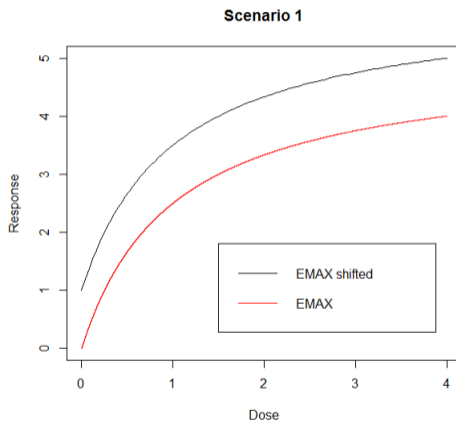

Figure: *Two shifted EMAX models.*

Simulation of the nominal level

- $m_1(x, \vartheta_1) = \delta + \frac{5x}{x+1}$ $\qquad m_2(x, \vartheta_2) = \frac{5x}{1+x}$ $\qquad \mathcal{X} = [0, 4]$
- $H_0 : d_2(m_1, m_2) > 1$ versus $H_1 : d_2(m_1, m_2) \leq 1$

		$\sigma^2 = 0.25$		$\sigma^2 = 1$	
n	$d_2(m_1, m_2)$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
10	1	0.055	0.117	0.044	0.098
10	1.5	0.000	0.000	0.003	0.011
10	2	0.000	0.000	0.000	0.000
50	1	0.048	0.106	0.061	0.088
50	1.5	0.000	0.000	0.000	0.000
50	2	0.000	0.000	0.000	0.000

Table: Simulated level of the new bootstrap test

Simulation of power

		$\sigma^2 = 0.5$		$\sigma^2 = 1$	
n	$d_2(m_1, m_2)$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
10	0	0.713	0.855	0.387	0.555
10	0.2	0.680	0.794	0.346	0.503
10	0.5	0.438	0.574	0.238	0.364
20	0	0.976	0.985	0.739	0.874
20	0.2	0.949	0.974	0.661	0.792
20	0.5	0.707	0.818	0.440	0.578

Table: Simulated power of the new bootstrap test

Comparison of the two tests

		New test		Gsteiger et al.	
n	$d_2(m_1, m_2)$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
50	1	0.062	0.086	0.000	0.000
50	1.5	0.000	0.000	0.000	0.000
50	2	0.000	0.000	0.000	0.000
150	1	0.053	0.106	0.000	0.000
150	1.5	0.000	0.000	0.000	0.000
150	2	0.000	0.001	0.000	0.000

Table: Simulated level of the new bootstrap test and the test of Gsteiger et al. (2011)

Comparison of two tests

		New test		Gsteiger et al.	
n	$d_2(m_1, m_2)$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
50	0	0.984	0.999	0.006	0.011
50	0.2	0.967	0.991	0.000	0.001
50	0.5	0.774	0.971	0.000	0.000
150	0	1.000	1.000	0.632	0.725
150	0.2	1.000	1.000	0.586	0.720
150	0.5	1.000	1.000	0.332	0.447

Table: Simulated power of the new bootstrap test and the test of Gsteiger et al. (2011)

Conclusions

- The test for similarity of Gsteiger et al. (2011) is extremely conservative
- In the second part of this talk a substantially more powerful test has been proposed
  - estimate the distance and the variance of the test statistic directly
  - generate quantiles by parametric bootstrap

- other distances were also considered, e.g.

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m(x, \vartheta_2)|$$

these results are now available as preprint and the test was implemented in an R package

Conclusions

- The test for similarity of Gsteiger et al. (2011) is extremely conservative
- In the second part of this talk a substantially more powerful test has been proposed
  - estimate the distance and the variance of the test statistic directly
  - generate quantiles by parametric bootstrap

- other distances were also considered, e.g.

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m(x, \vartheta_2)|$$

these results are now available as preprint and the test was implemented in an R package

Conclusions

- The test for similarity of Gsteiger et al. (2011) is extremely conservative
- In the second part of this talk a substantially more powerful test has been proposed
  - estimate the distance and the variance of the test statistic directly
  - generate quantiles by parametric bootstrap

- other distances were also considered, e.g.

$$d_\infty(m_1, m_2) = \max_{x \in \mathcal{X}} |m_1(x, \vartheta_1) - m(x, \vartheta_2)|$$

these results are now available as preprint and the test was implemented in an R package

Acknowledgments

Thank you very much!

References:
Gsteiger, S. , Bretz, F. and Liu, W.(2011): Simultaneous Confidence Bands for Nonlinear Regression Models with Application to Population Pharmacokinetic Analyses, Journal of Biopharmaceutical Statistics, 21: 4, 708 725