Envelope testing in spatial statistics

Jiří Dvořák

Department of Probability and Mathematical Statistics, MFF UK, Prague

Course on Spatial statistics

Situation

- Observed data described by $F_{data}(r), \ r \in [r_{min}, r_{max}]$,
- example: random set, $F(r) =$ area of the set dilated by r,

Situation

- Observed data described by $F_{data}(r), \; r \in [r_{min}, r_{max}]$,
- example: random set, $F(r) =$ area of the set dilated by r,
- we have a hypothesis that the data comes from a given model (null hypothesis, H_0),

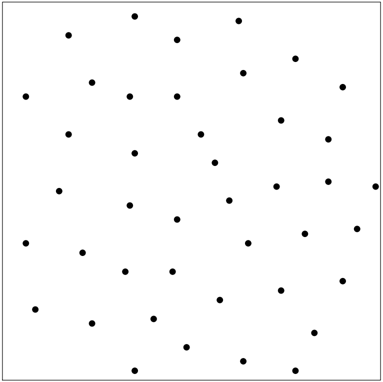
Situation

- Observed data described by $F_{data}(r),\ r \in [r_{min}, r_{max}]$,
- example: random set, $F(r) =$ area of the set dilated by r,
- we have a hypothesis that the data comes from a given model (null hypothesis, H_0),
- F_{H_0} difficult or impossible to work with,

Situation

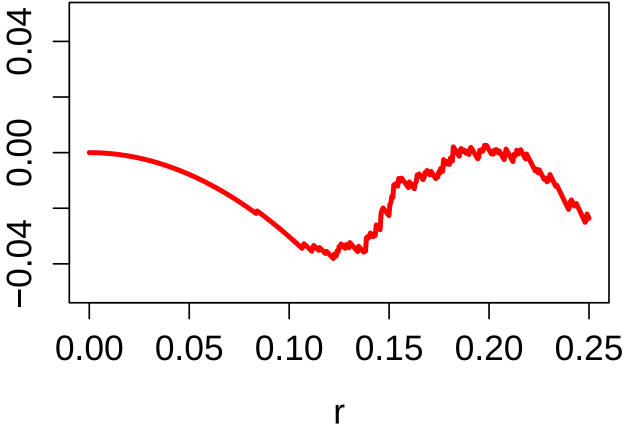
- Observed data described by $F_{data}(r), \; r \in [r_{min}, r_{max}]$,
- example: random set, $F(r) =$ area of the set dilated by r,
- we have a hypothesis that the data comes from a given model (null hypothesis, H_0),
- F_{H_0} difficult or impossible to work with,
- we can simulate from the model.

Ex.: CELLS dataset

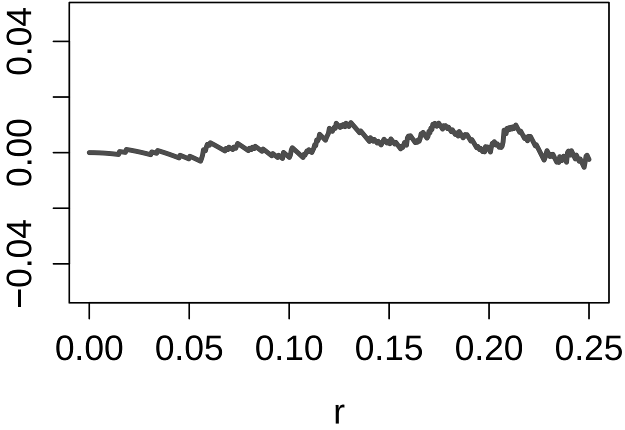


$$F(r) = K(r) - \pi r^2, \ r \in [0; 0.25]$$

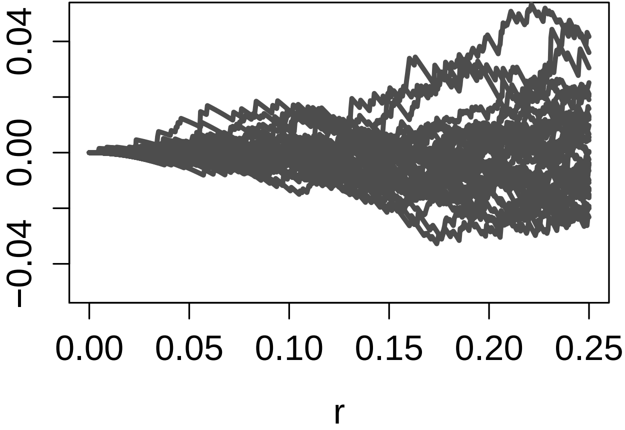
Ex.: data curve



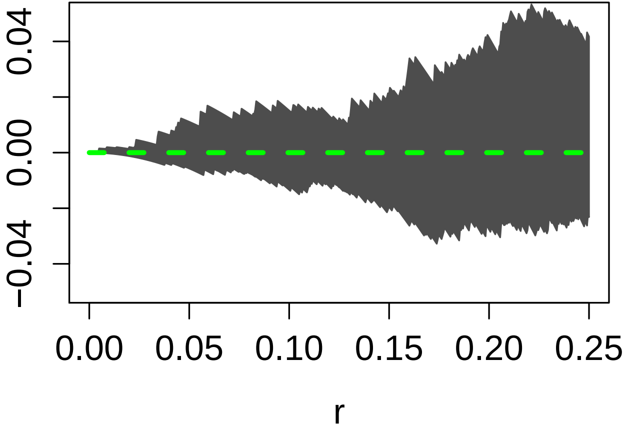
Ex.: simulated curve (Poisson process)



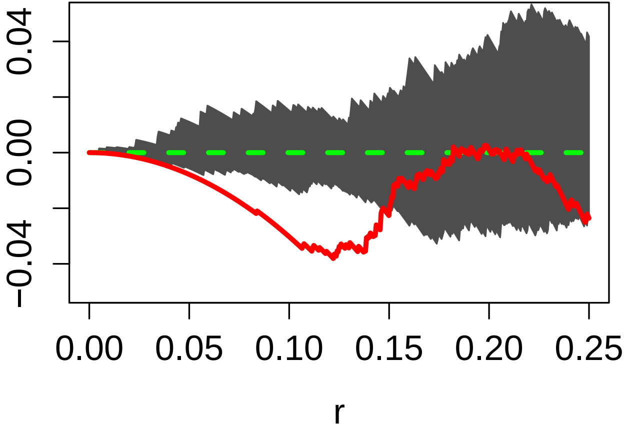
Ex.: 39 simulated curves



Ex.: envelope and F_{H_0}



Ex.: envelope test



Envelope test (Ripley, 1977)

- Reject H_0 if the data curve leaves the envelope,

Envelope test (Ripley, 1977)

- Reject H_0 if the data curve leaves the envelope,
- 39 simulated curves + 1 data curve $\Rightarrow$ probability of type I error is 5 % if we look at a single value r_0,
- if we look at more values, prob. of type I error gets higher,

Envelope test (Ripley, 1977)

- Reject H_0 if the data curve leaves the envelope,
- 39 simulated curves + 1 data curve $\Rightarrow$ probability of type I error is 5 % if we look at a single value r_0,
- if we look at more values, prob. of type I error gets higher,
- problem of multiple comparison (multiple testing),
- **suggestion:** use envelopes as visual tool, avoid “testing”,

Envelope test (Ripley, 1977)

- Reject H_0 if the data curve leaves the envelope,
- 39 simulated curves + 1 data curve $\Rightarrow$ probability of type I error is 5 % if we look at a single value r_0,
- if we look at more values, prob. of type I error gets higher,
- problem of multiple comparison (multiple testing),
- **suggestion:** use envelopes as visual tool, avoid “testing”,
- **result:** ecologists, biologists, etc., use envelopes to formally test their hypotheses.

Deviation tests (Diggle, 1979; Ripley, 1979)

Summarize $F(r)$ into a single number u:

$$u_{data} = \int_{r_{min}}^{r_{max}} \left(F_{data}(r) - F_{H_0}(r)\right)^2 \, \mathrm{d}r,$$

$$u_{data} = \max_{r \in [r_{min}, r_{max}]} |F_{data}(r) - F_{H_0}(r)|.$$

- F_{H_0} known analytically or estimated as the mean of the simulated curves,

Deviation tests (Diggle, 1979; Ripley, 1979)

Summarize $F(r)$ into a single number u:

$$u_{data} = \int_{r_{min}}^{r_{max}} \left(F_{data}(r) - F_{H_0}(r)\right)^2 \,\mathrm{d}r,$$

$$u_{data} = \max_{r \in [r_{min}, r_{max}]} |F_{data}(r) - F_{H_0}(r)|.$$

- F_{H_0} known analytically or estimated as the mean of the simulated curves,
- compare u_{data} to the values u_i from n simulations,
- does u_{data} behave extremally w.r.t. u_i?

Deviation tests (Diggle, 1979; Ripley, 1979)

Summarize $F(r)$ into a single number u:

$$u_{data} = \int_{r_{min}}^{r_{max}} \left(F_{data}(r) - F_{H_0}(r)\right)^2 \, \mathrm{d}r,$$

$$u_{data} = \max_{r \in [r_{min}, r_{max}]} |F_{data}(r) - F_{H_0}(r)|.$$

- F_{H_0} known analytically or estimated as the mean of the simulated curves,
- compare u_{data} to the values u_i from n simulations,
- does u_{data} behave extremally w.r.t. u_i?
- p-value estimated as $\frac{1}{n}\{\#i : u_i > u_{data}\}$,
- other deviation measures also possible.

Properties of deviation tests

Advantages:

- formal test with given prob. of type I error,
- gives p-value.

Properties of deviation tests

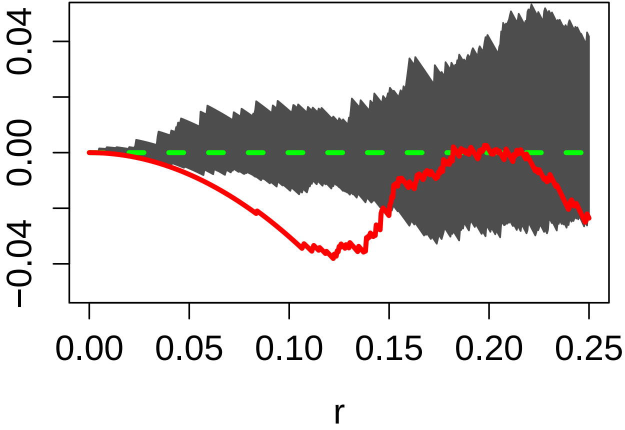
Advantages:

- formal test with given prob. of type I error,
- gives p-value.

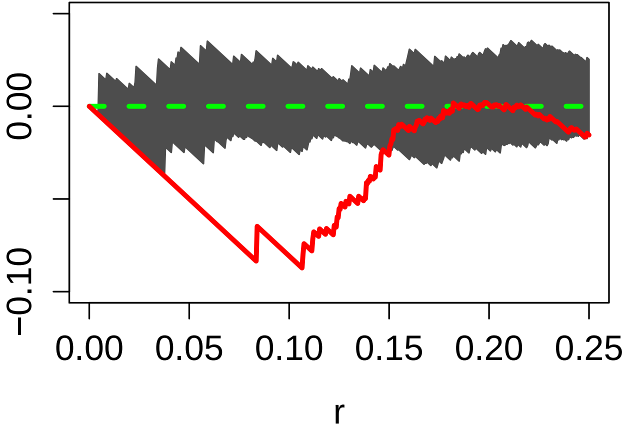
Disadvantages:

- If H_0 rejected, no clear indication of the reason,
- no information which values of r caused rejection (important for ecologists in order to form new hypotheses),
- problems with non-constant variance (width of the envelope) and/or assymetry.

Problem: non-constant variance



Variance might be stabilized



State-of-the-art (up to 2006)

- Envelopes from 39 or 99 simulations,
- deviation tests for formal testing.

State-of-the-art (up to 2006)

- Envelopes from 39 or 99 simulations,
- deviation tests for formal testing.

Loosmore & Ford (2006):

- illustrate by example that prob. of type I error in the global envelope test may be unacceptably high (0.74 or even more),

State-of-the-art (up to 2006)

- Envelopes from 39 or 99 simulations,
- deviation tests for formal testing.

Loosmore & Ford (2006):

- illustrate by example that prob. of type I error in the global envelope test may be unacceptably high (0.74 or even more),
- to do so they explain how prob. of type I error can be estimated *a posteriori*, i.e. after the simulations,
- idea: how many curves contribute to the shape of the envelope?

State-of-the-art (up to 2006)

- Envelopes from 39 or 99 simulations,
- deviation tests for formal testing.

Loosmore & Ford (2006):

- illustrate by example that prob. of type I error in the global envelope test may be unacceptably high (0.74 or even more),
- to do so they explain how prob. of type I error can be estimated *a posteriori*, i.e. after the simulations,
- idea: how many curves contribute to the shape of the envelope?
- recommendation: use deviation tests, not envelopes.

State-of-the-art (up to 2006)

- Envelopes from 39 or 99 simulations,
- deviation tests for formal testing.

Loosmore & Ford (2006):

- illustrate by example that prob. of type I error in the global envelope test may be unacceptably high (0.74 or even more),
- to do so they explain how prob. of type I error can be estimated *a posteriori*, i.e. after the simulations,
- idea: how many curves contribute to the shape of the envelope?
- recommendation: use deviation tests, not envelopes.

(see Section 2.3 of paper Grabarnik et al., 2011)

State-of-the-art (2011)

Grabarnik et al. (2011):

- Take up the approach of Loosmore & Ford (2006),
- idea: try increasing the number of simulations, hoping that the prob. of type I error will decrease,
- it might be necessary to use, say, 1999 simulations to reduce the probability under 0.1.

State-of-the-art (2011)

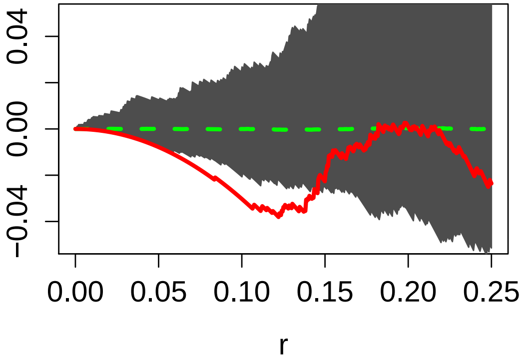
Grabarnik et al. (2011):

- Take up the approach of Loosmore & Ford (2006),
- idea: try increasing the number of simulations, hoping that the prob. of type I error will decrease,
- it might be necessary to use, say, 1999 simulations to reduce the probability under 0.1.

Disadvantages:

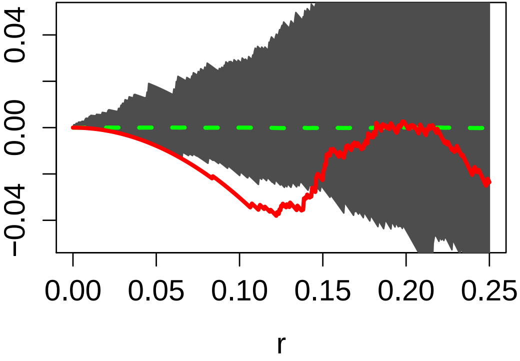
- Iterative procedure,
- no p-value provided (we do not know how strong is the reason for rejection).

Refined envelope test, n = 999



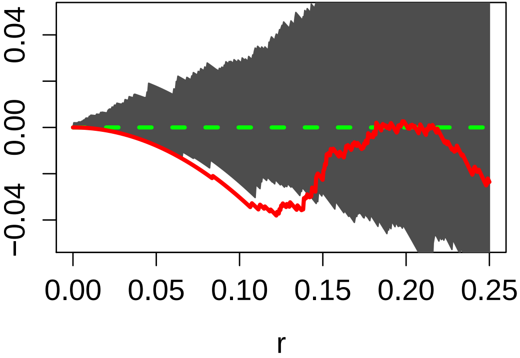
Probability of type I error: 0.1231

Refined envelope test, n = 1999



Probability of type I error: 0.0670

Refined envelope test, n = 2999



Probability of type I error: 0.0460

State-of-the-art (2017)

Myllymäki et al. (2017), exact envelope test:

- increase the number of simulations even more (say 3999),

State-of-the-art (2017)

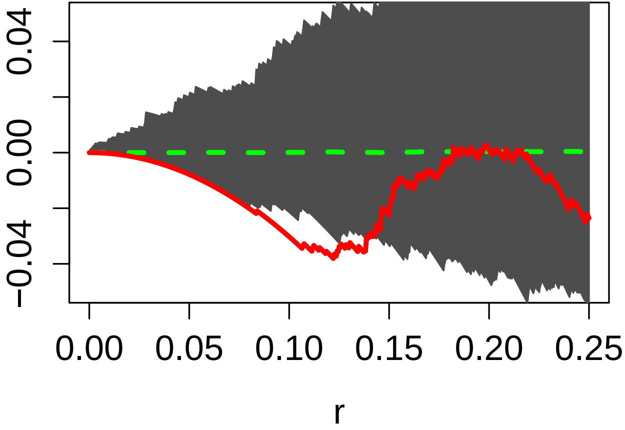
Myllymäki et al. (2017), exact envelope test:

- increase the number of simulations even more (say 3999),
- idea: define kth upper and lower envelope:

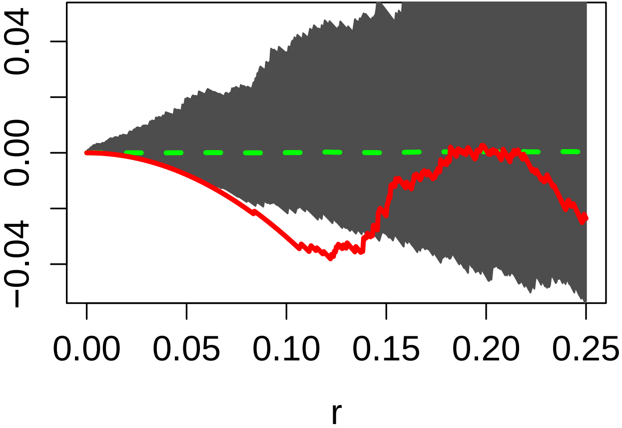
$$F^k_{upp}(r) = \max\nolimits^k_{i \in \{data, 1, \dots, n\}} F_i(r),$$
$$F^k_{low}(r) = \min\nolimits^k_{i \in \{data, 1, \dots, n\}} F_i(r),$$

- $\max^k$ – kth largest value, $\min^k$ – kth lowest value,
- first envelope is the broadest (most extreme), second, third etc. envelopes are more and more narrow.

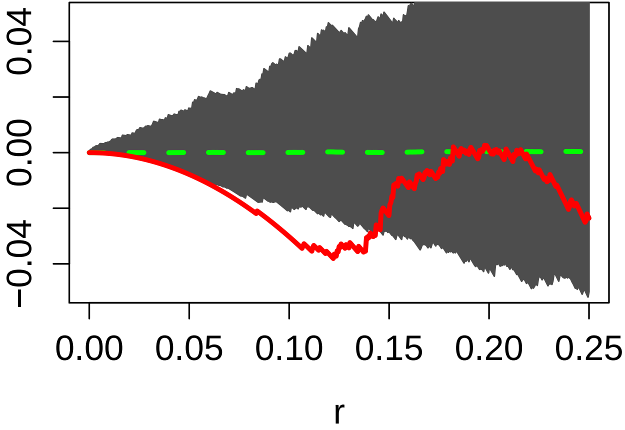
Exact envelope test, n = 4999, 1st envelope



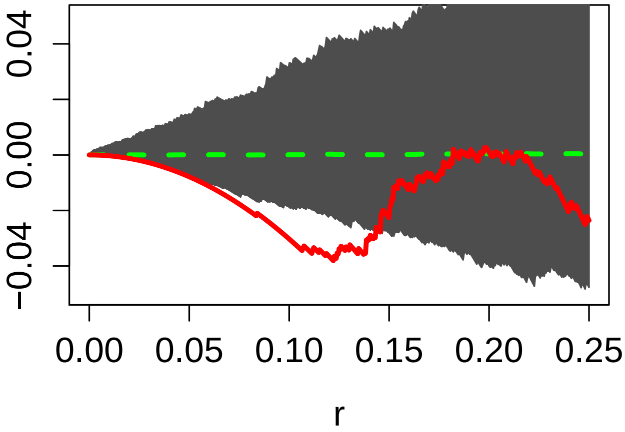
Exact envelope test, n = 4999, 2nd envelope



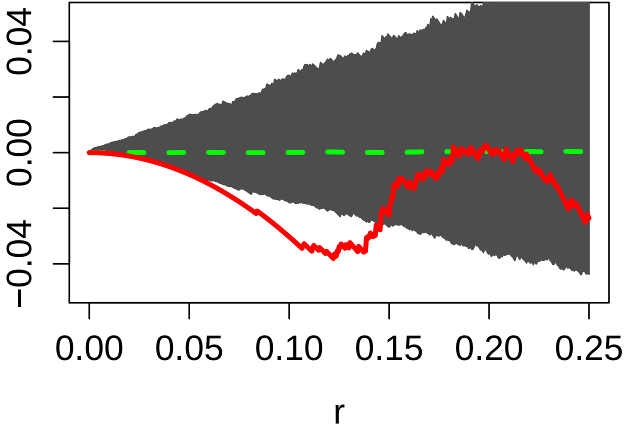
Exact envelope test, n = 4999, 3rd envelope



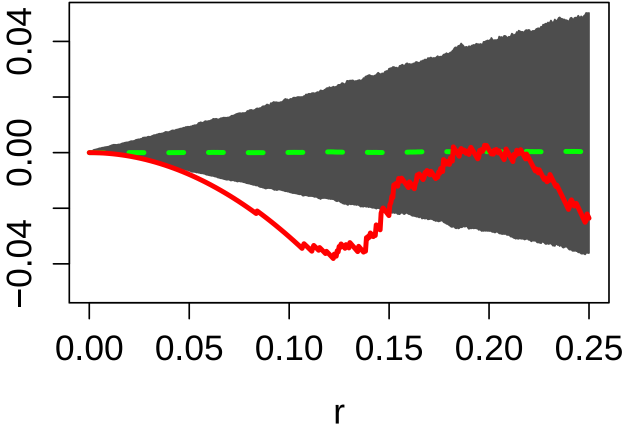
Exact envelope test, n = 4999, 5th envelope



Exact envelope test, n = 4999, 10th envelope



Exact envelope test, n = 4999, 50th envelope

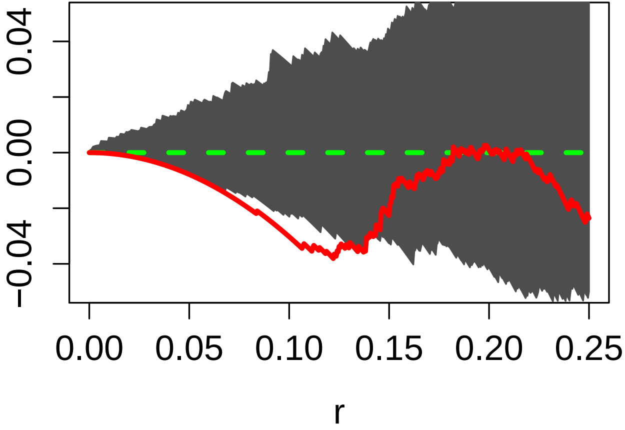


Exact envelope test

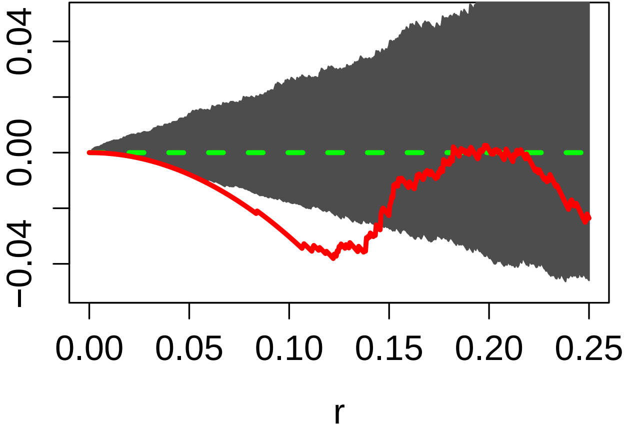
Global α-envelope (say 95%): we need to find k_α such that

$$\alpha = \mathbb{P}_{H_0}\left(F_{low}^{k_\alpha}(r) \leq F(r) \leq F_{upp}^{k_\alpha}(r)\right)$$
$$\approx \frac{1}{n}\left\{\#i : F_{low}^{k_\alpha}(r) \leq F_i(r) \leq F_{upp}^{k_\alpha}(r)\right\}.$$

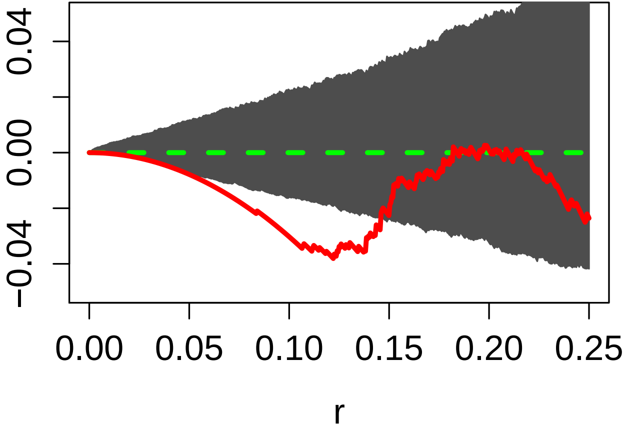
Exact envelope test, n = 4999, α=0.99



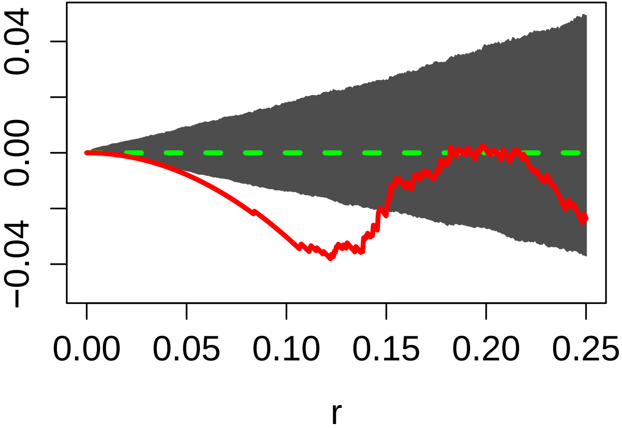
Exact envelope test, n = 4999, α=0.95



Exact envelope test, n = 4999, α=0.90



Exact envelope test, n = 4999, α=0.75



Exact envelope test

Provides p-value: let k_1 be the maximal k for which the kth envelope fully covers F_{data}. Then

$$p \approx \frac{1}{n}\left\{\#i : F_i(r) \text{ somewhere leaves } [F_{low}^{k_1}(r), F_{upp}^{k_1}(r)]\right\}.$$

Advantages:

- gives p-value,
- no iterations needed,
- intuitive interpretation of the global envelopes,
- shows reason of rejection,
- no problems with non-constant variance or asymetry.

Disadvantages:

- high number of ties implies a LOT of simulations is needed.

Extreme rank length ranking

Olympic games – ranking countries based on the number of gold/silver/bronze medals.