Hypothesis testing of quantum Monte Carlo calculations

Markus Wallerberger*

in collaboration with:

Alexander Gaenko, Emanuel Gull (University of Michigan)

***mwallerb@umich.edu**

How can we know if a stochastic algorithm works in a way that can be automatized?

Testing of deterministic algorithms

- **Contract-oriented programming**: invariants testing
- **Unit tests**: small, user-facing, correctness tests
- **Fuzzing**: large-scale reliability tests
- **Benchmarks**: checks for "physical" cases[1]

[1] *e.g., J.P.F. LeBlanc et al., Phys. Rev. X 5, 041041 (2015); K. Lejaeghere et al, Science (2016)*

Testing of stochastic algorithms

- **Contract-oriented programming**: invariants testing
- ~~**Unit tests**: small, user-facing, correctness tests~~
- ~~**Fuzzing**: large-scale reliability tests~~
- **Benchmarks**: checks for "physical" cases[1]

Testing of stochastic algorithms

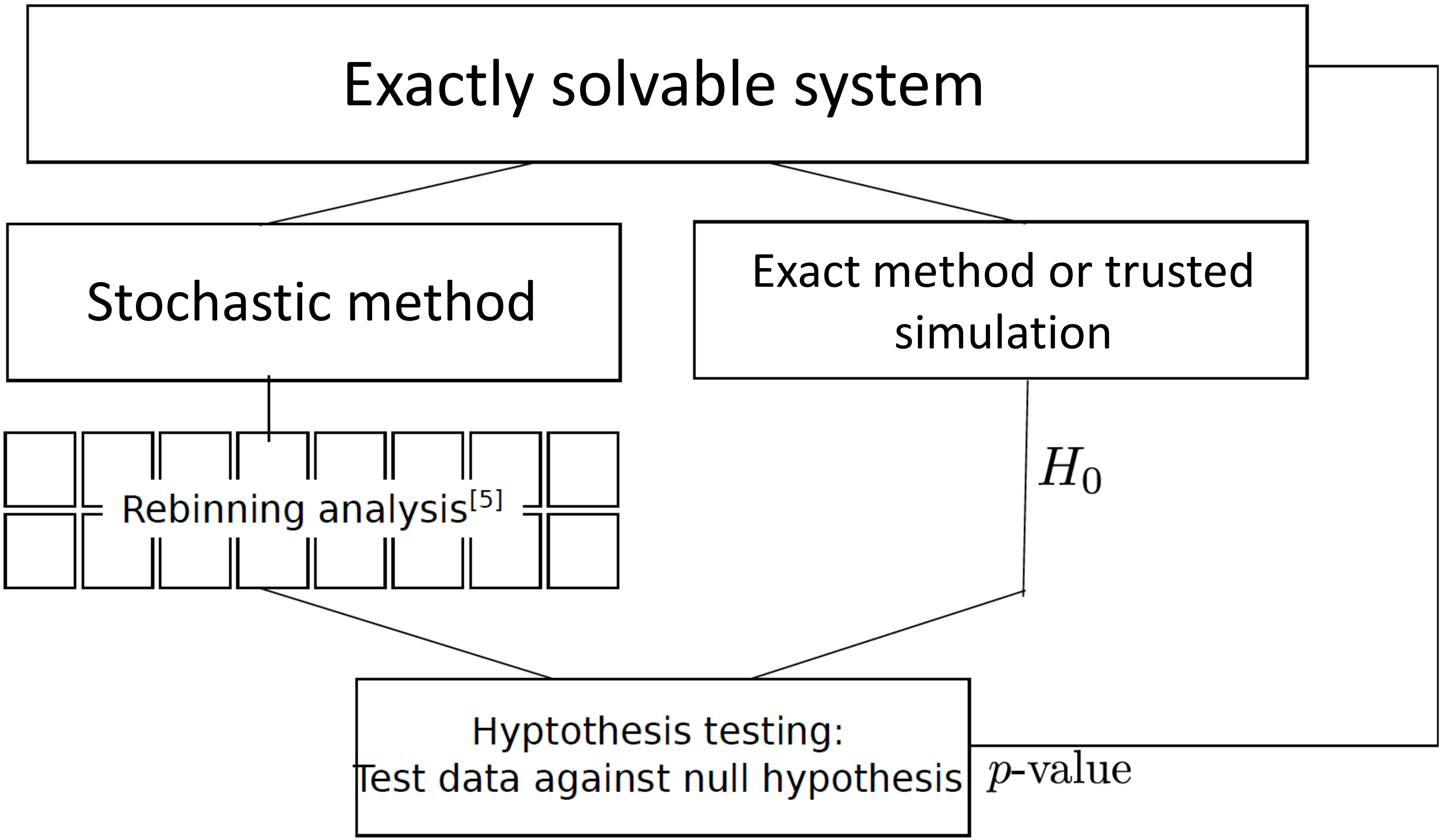
- **Contract-oriented programming**: invariants testing
- ~~**Unit tests**: small, user-facing, correctness tests~~
- ~~**Fuzzing**: large-scale reliability tests~~
- **Benchmarks**: checks for "physical" cases[1]
- **Example**: Solvers for the Anderson model
  - 1960s bath level discretization[1] ~ 200 LOC
  - 1980s imaginary time discretization[1] ~ 2,000 LOC
  - 2000s no systematic approximation[2] ~ 20,000 LOC

[1] *review: A Georges et al., Rev. Mod. Phys. 68, 13 (1996)* [2] *review: E. Gull et al., Rev. Mod. Phys. 83, 349 (2011)*

Statistical Hypothesis testing

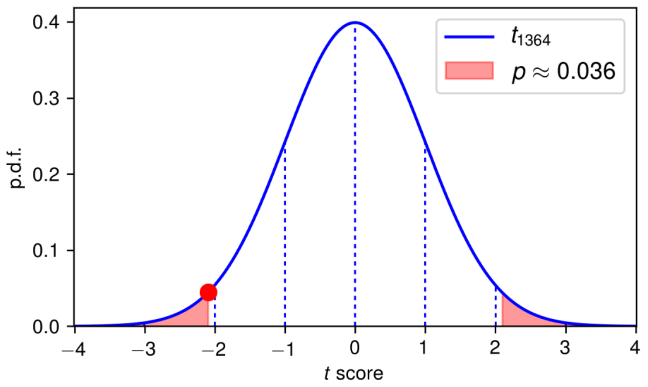
- Ubiquitous in life sciences etc.
  - (aside: formal validity from a frequentist p.o.v.?)
- Verification of stochastic algorithms: urban simulations[1] and image recognition[2]
- H_0 ... simulation follows trusted result
- H_0 rejected = failed test

[1] *H. Ševčíková et al., Proc. Int. Symp. Softw. Test. Anal., p. 215 (2006)*
[2] *K. Subr and J. Arvo, Proc. 15th Pacif. Conf. Comput. Graph. Appl., p. 106 (2007)*


Exactly solvable system
Stochastic method
Exact method or trusted simulation
Rebinning analysis[5]
H_0
Hyptothesis testing:
Test data against null hypothesis
p-value

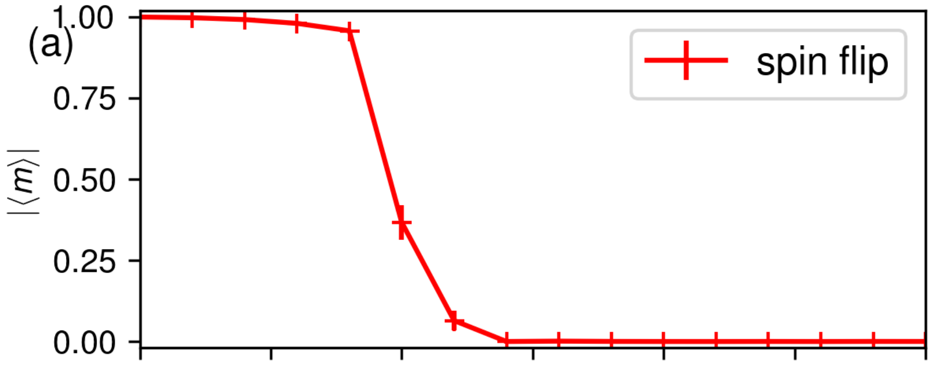
Simple scalar test

- null hypothesis $H_0 : \mathrm{E}[\hat{X}] = y$
- alternative $H_1 : \mathrm{E}[\hat{X}] \neq y$
- score $$z = \frac{\langle X \rangle - y}{\sigma_X / \sqrt{N}} \sim t_{N-1},$$
- p-value $$p = 2P^{-1}(-|z|)$$
- Student's t test

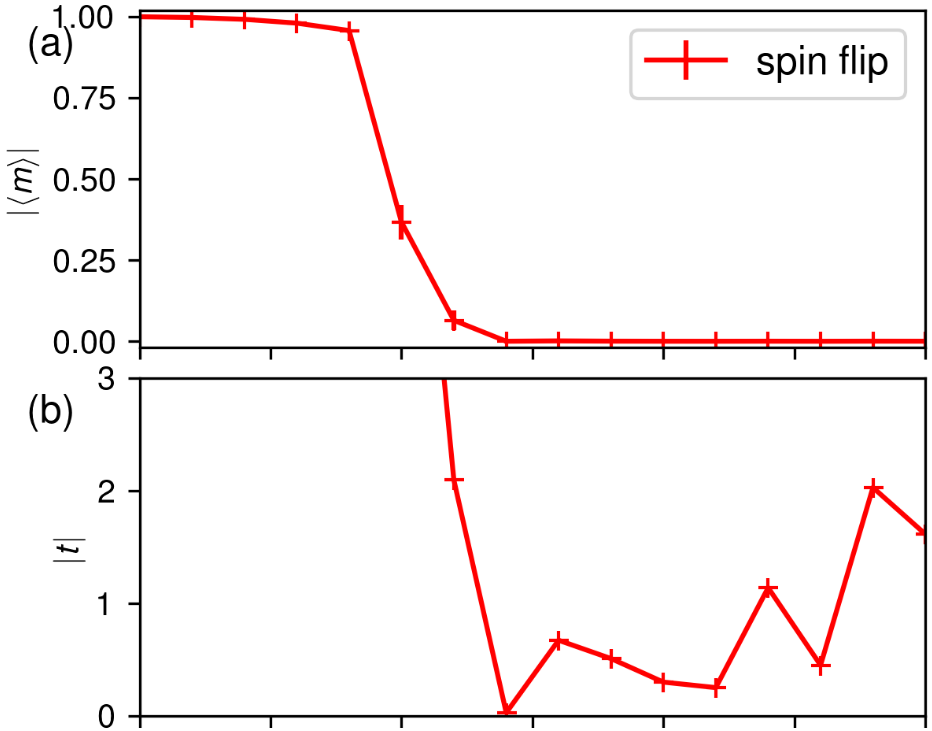


Step 1: Testing

- $$H = -\sum_{\langle ij \rangle} \sigma_i \sigma_j,$$
- finite square lattice L x L, periodic boundary cond.
- h=0: we know from SU(2) that: m = 0
- Test against Monte Carlo estimator <m>
- Single spin flips



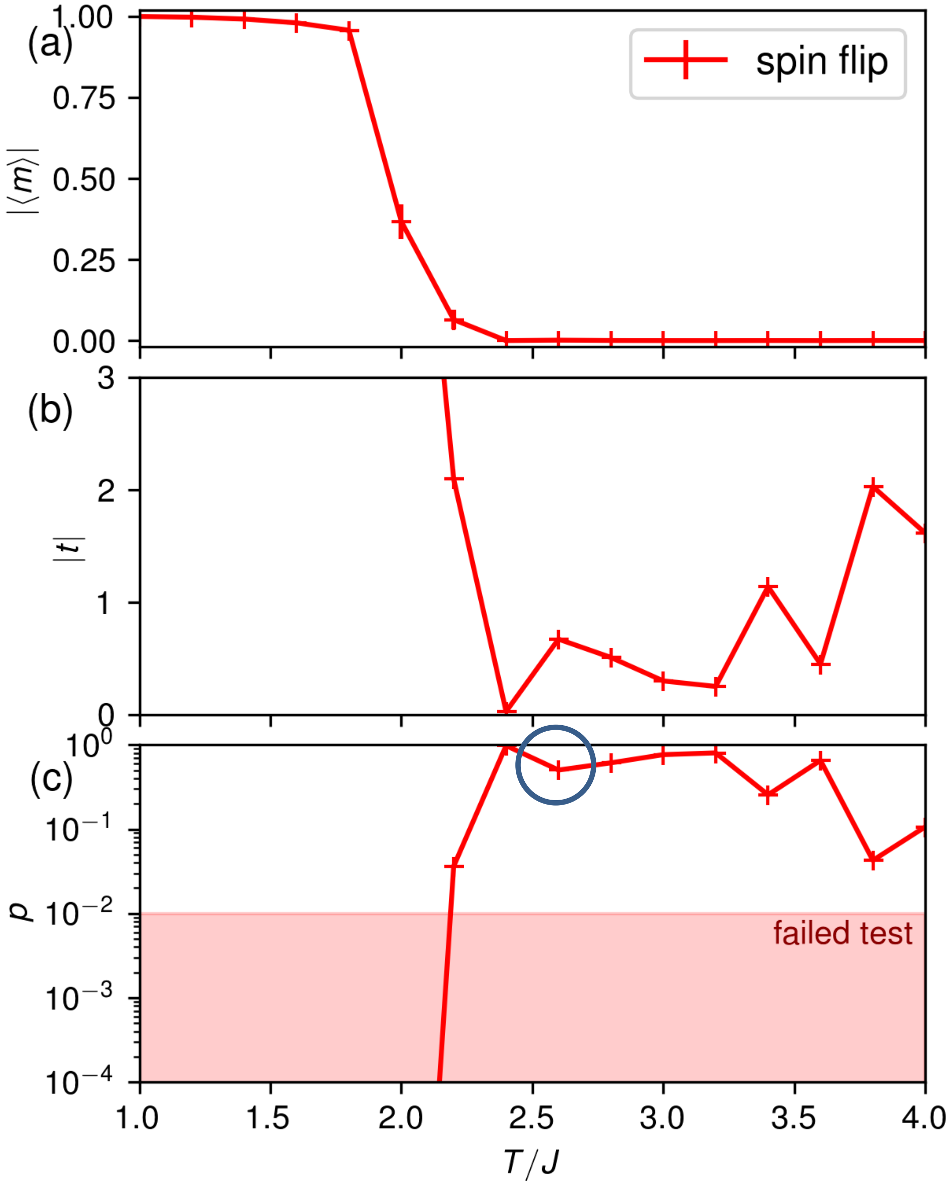
$$H_0 : \langle m \rangle = 0$$
$$H_1 : \langle m \rangle \neq 0$$



$$H_0 : \langle m \rangle = 0$$

$$H_1 : \langle m \rangle \neq 0$$

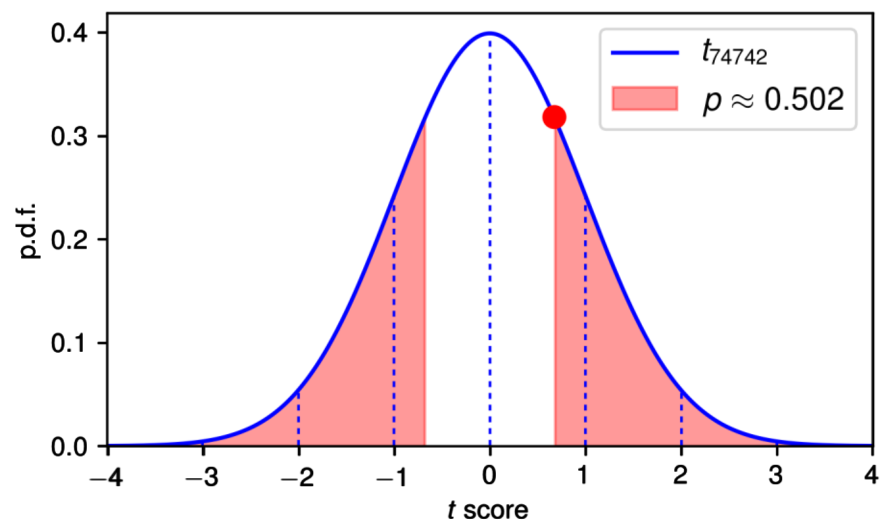
$$\frac{\langle m \rangle - 0}{\sigma_M / \sqrt{N}} \sim t_{N-1}$$

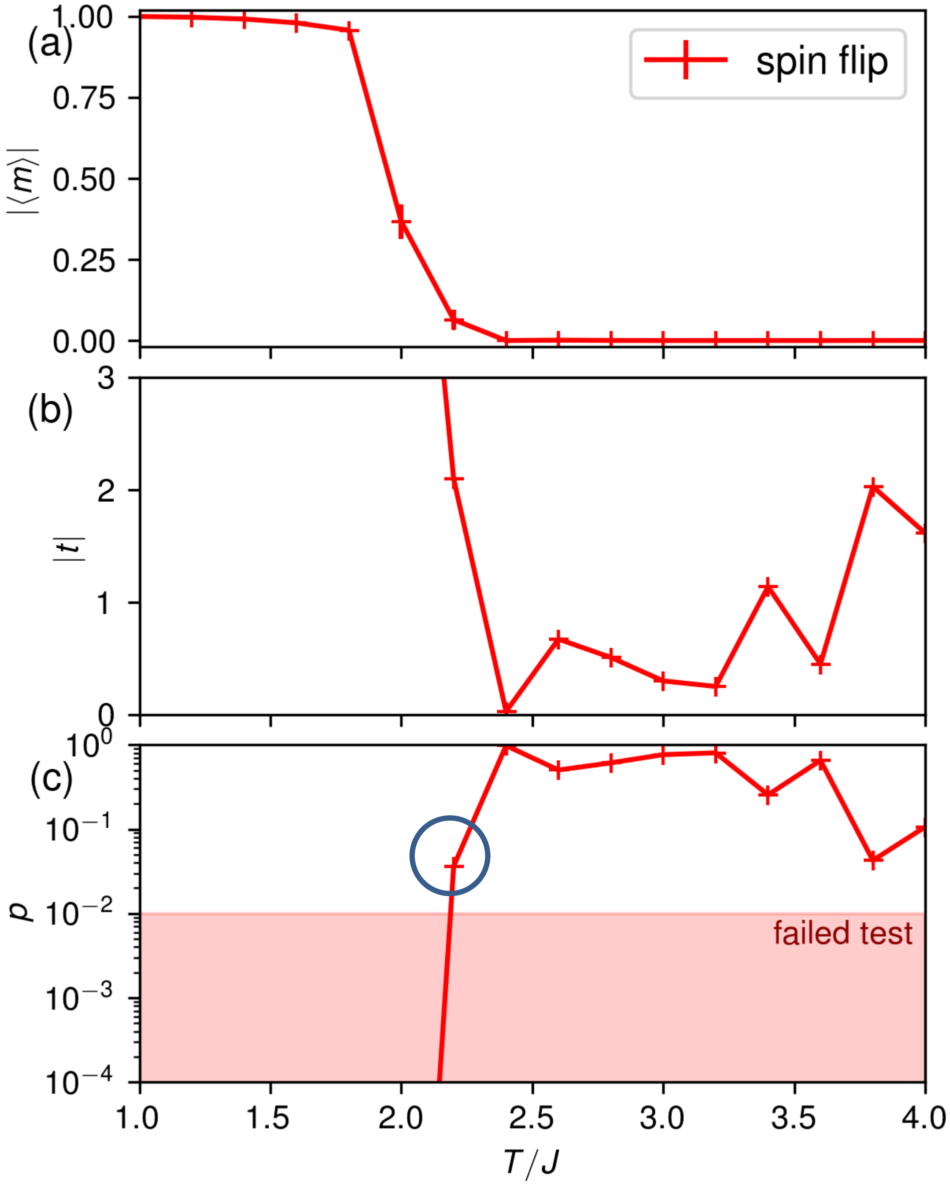


$$H_0 : \langle m \rangle = 0$$

$$H_1 : \langle m \rangle \neq 0$$

$$\frac{\langle m \rangle - 0}{\sigma_M / \sqrt{N}} \sim t_{N-1}$$

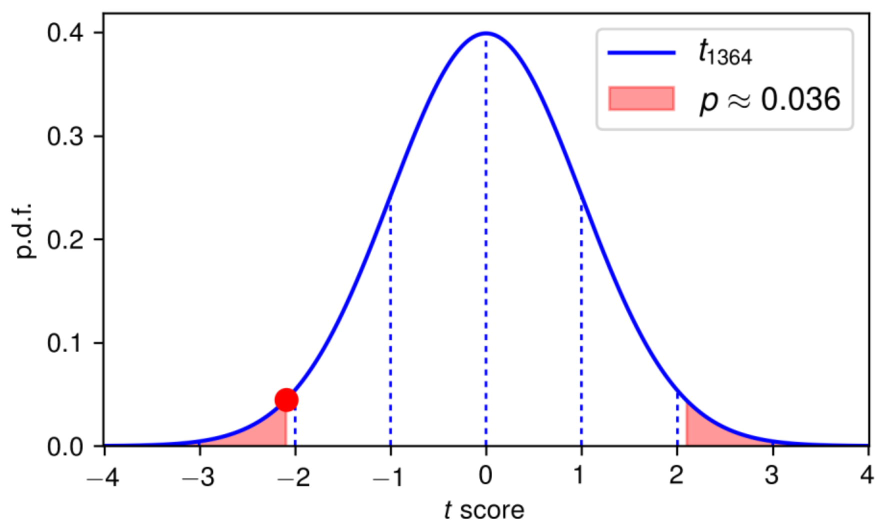




$$H_0 : \langle m \rangle = 0$$

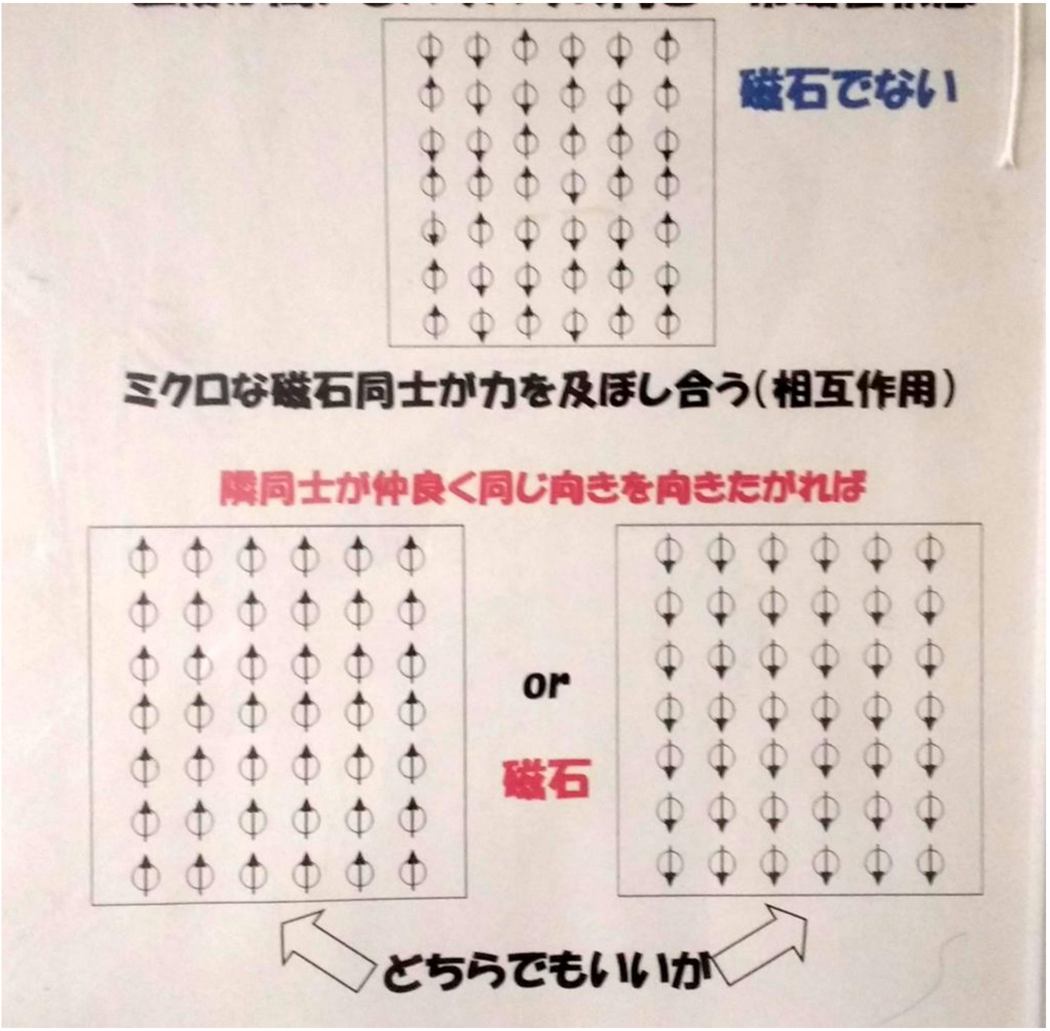
$$H_1 : \langle m \rangle \neq 0$$

$$\frac{\langle m \rangle - 0}{\sigma_M / \sqrt{N}} \sim t_{N-1}$$

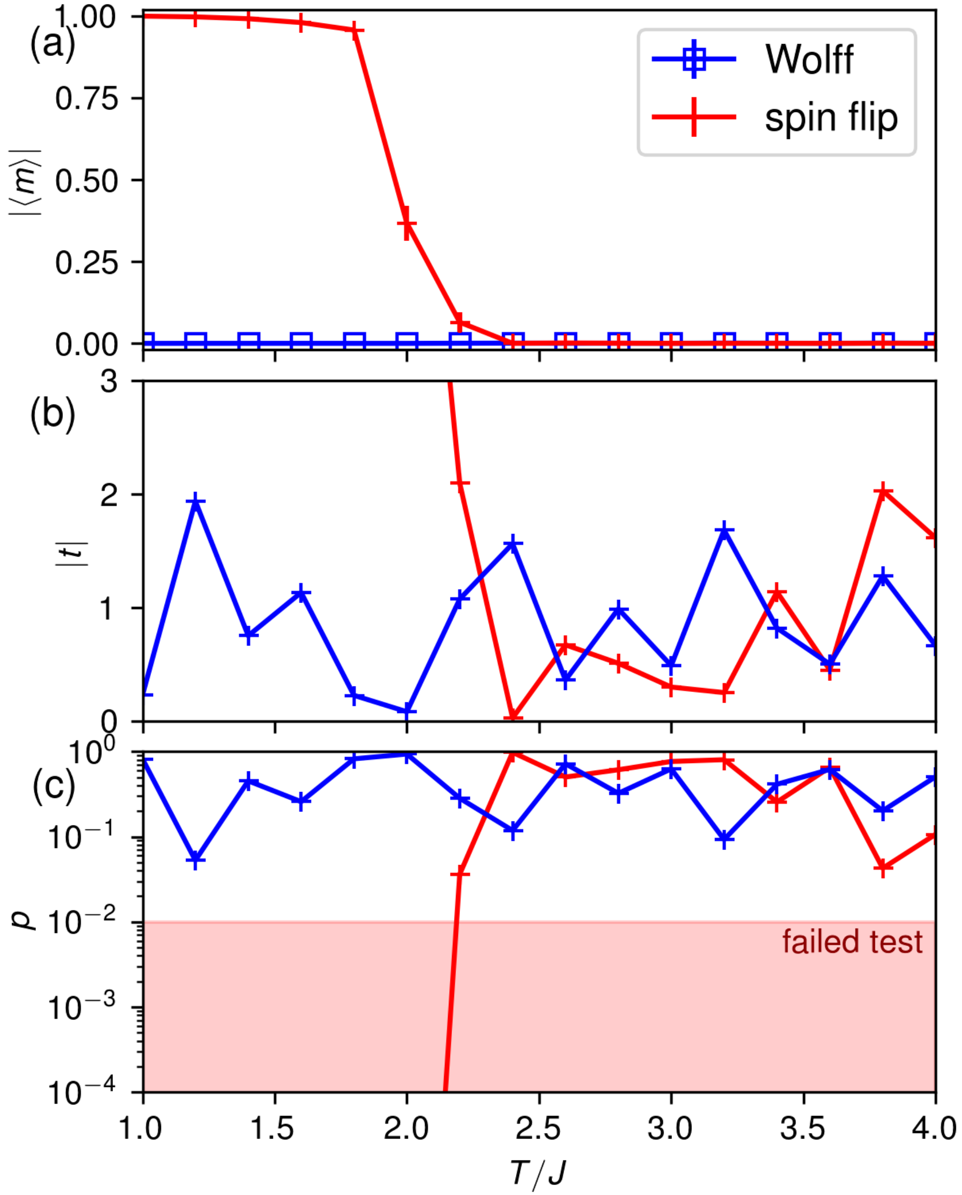


Step 2: Debugging

- Single spin flips are bad!



- Cluster updates!



$$H_0 : \langle m \rangle = 0$$

$$H_1 : \langle m \rangle \neq 0$$

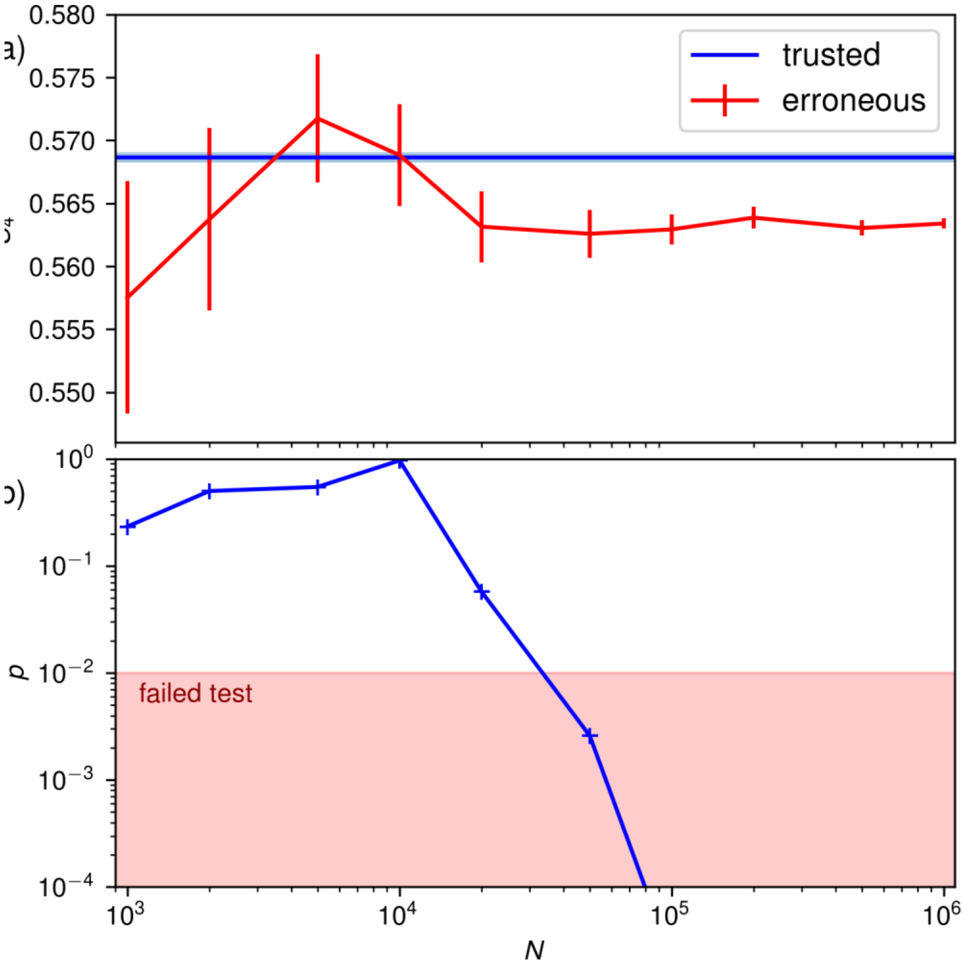
$$\frac{\langle m \rangle - 0}{\sigma_M / \sqrt{N}} \sim t_{N-1}$$

Testing against stochastic result

- null hypothesis $H_0 : \mathrm{E}[\hat{X}] = \mathrm{E}[\hat{Y}]$
- alternative $H_1 : \mathrm{E}[\hat{X}] \neq \mathrm{E}[\hat{Y}]$
- score
$$\frac{\langle X \rangle - \langle Y \rangle}{\sigma / N_\mu} \sim t_{N_X + N_Y - 2},$$
- **pooled variance**
$$\sigma^2 = \frac{(N_X - 1)\sigma_X^2 + (N_Y - 1)\sigma_Y^2}{N_X + N_Y - 2}.$$
$$N_\mu^{-1} = N_X^{-1} + N_Y^{-1}$$

Example: Ising model again

- Binder cumulant: $\hat{U}_4 = \dfrac{\langle m^4 \rangle}{1 - 3\langle m^2 \rangle^2}.$
- Hard to compute analytically
- → Compare with trusted simulation result
- Non-linear error propagation:
  - Problem for Student's t test
  - Bootstrap/Jackknife resampling as preprocessing
  - Alternative: parametric bootstrapt
- Artificial error: open boundary condition for corners



$$\hat{U}_4 = \frac{\langle m^4 \rangle}{1 - 3\langle m^2 \rangle^2}.$$

$$H_0 : \langle U_4 \rangle = \langle U_4^{\mathrm{tr}} \rangle$$

$$H_1 : \langle U_4 \rangle = \langle U_4^{\mathrm{tr}} \rangle$$

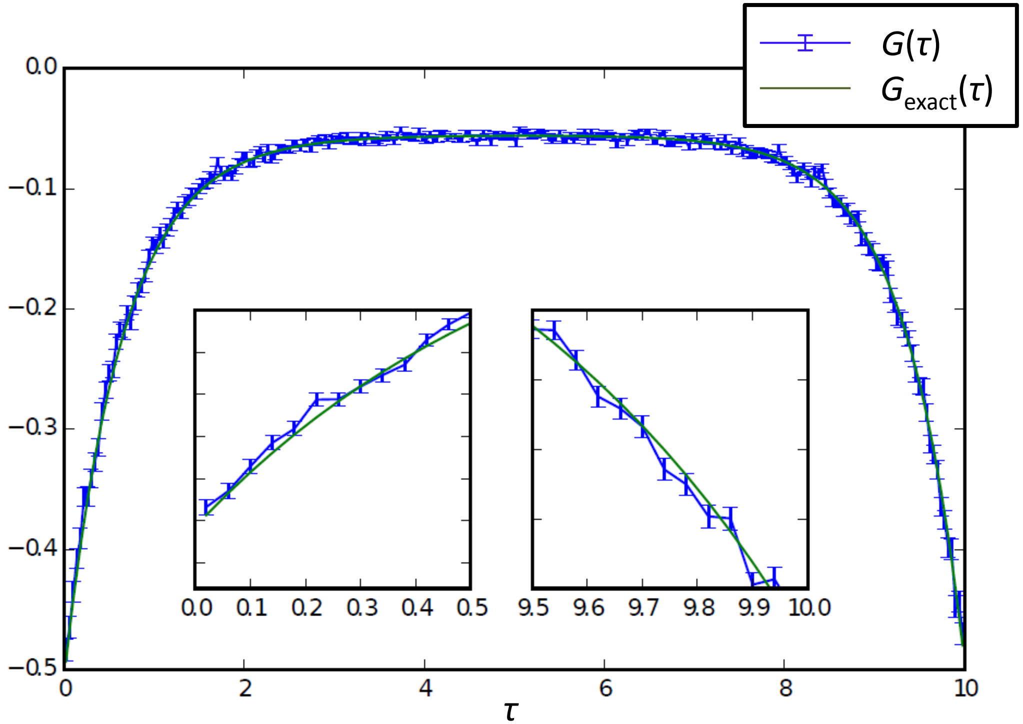
$$\frac{\langle U_4 \rangle - \langle U_4^{\mathrm{tr}} \rangle}{\sigma / N_\mu} \sim t_{N_X + N_Y - 2},$$

Testing data series

- often stricter criterion!
- null hypothesis $H_0 : \mathrm{E}[\hat{X}] = y$
- alternative $H_1 : \mathrm{E}[\hat{X}] \neq y$
- score $T^2 = N(\langle x \rangle - x_0)^\dagger \Sigma_x^{-1} (\langle x \rangle - x_0)$
- p-value $T^2 \sim \frac{n(N-1)}{N-n} F_{n,N-n}$
- Hotelling's T^2 test; generalizable for N<n[1]

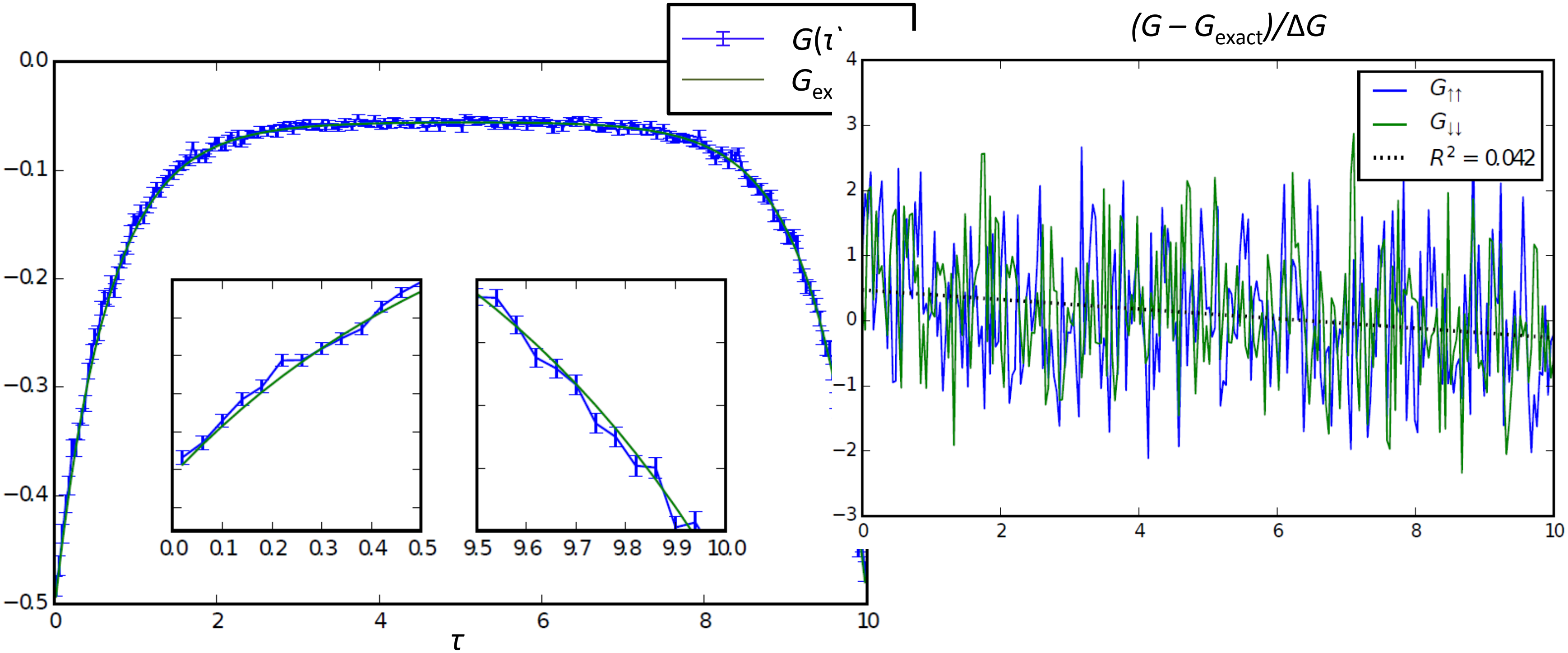
[1] *M. Marozzi: Stat. Meth. Med. Res. 25(6) 2593–2610 (2016)*

Example: chi(iw)

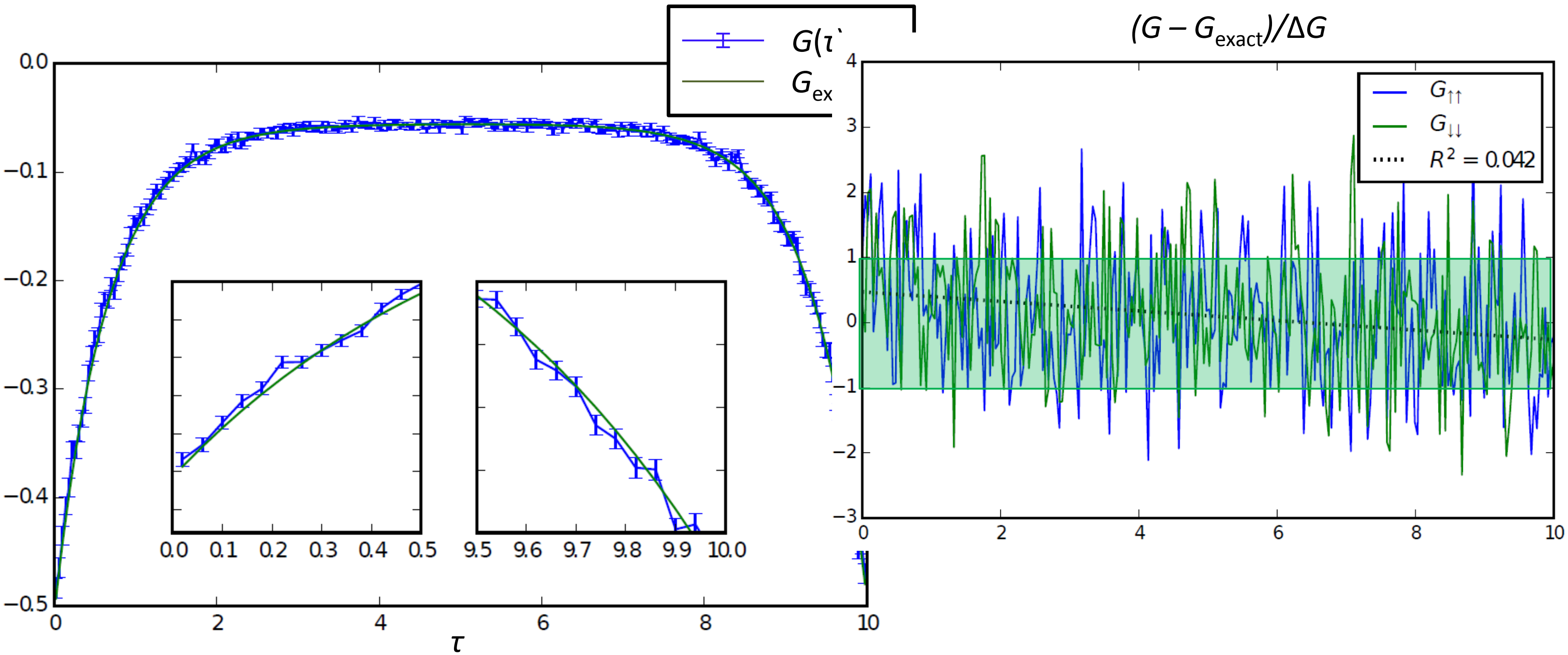
- 

- AIM: 1-orbital, 2 bath states (+/-0.5), V=1, U=1, mu=0.42

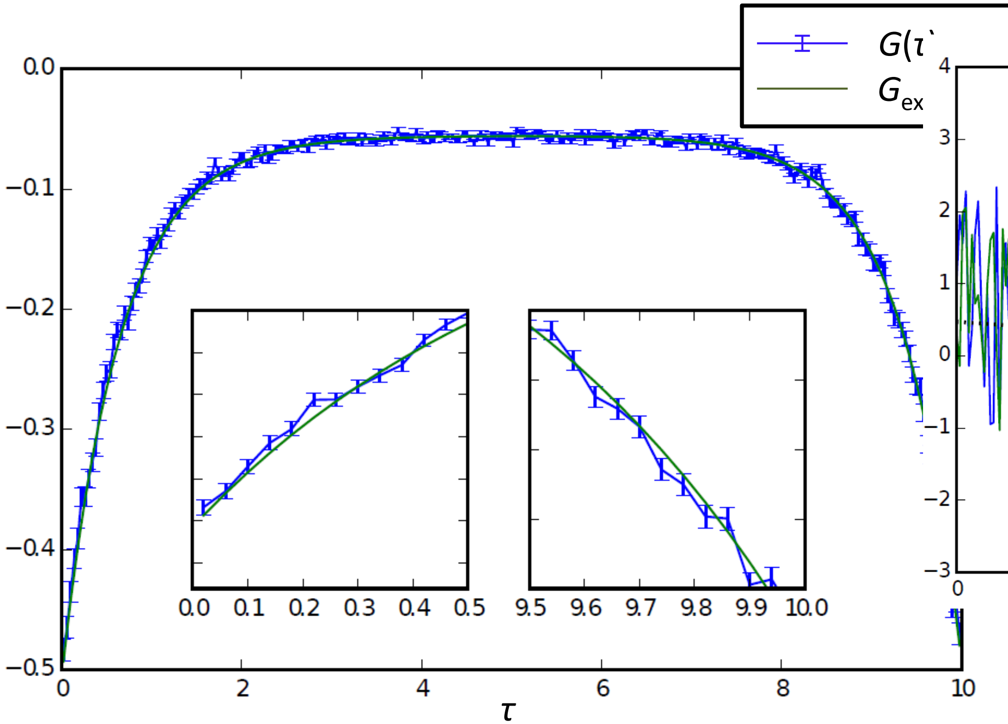
Example: chi(iw)

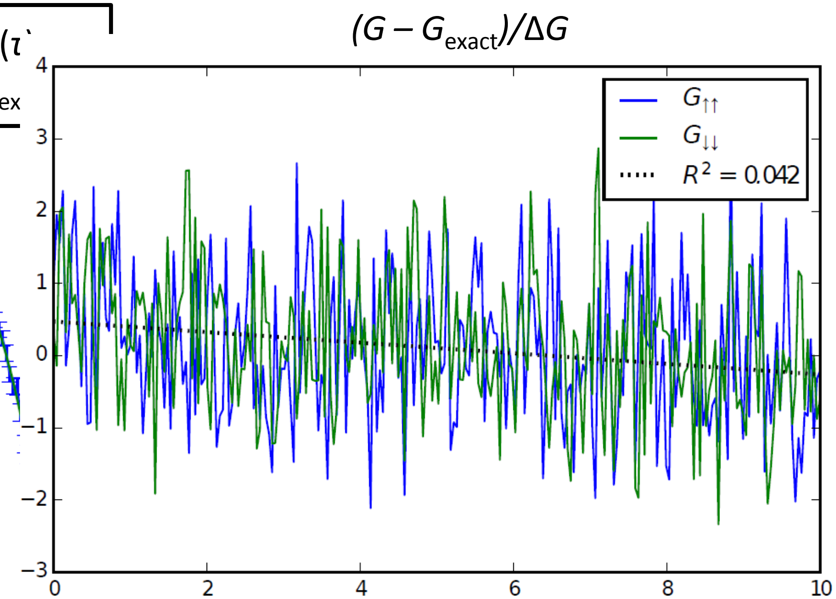
- 

Example: chi(iw)

- 

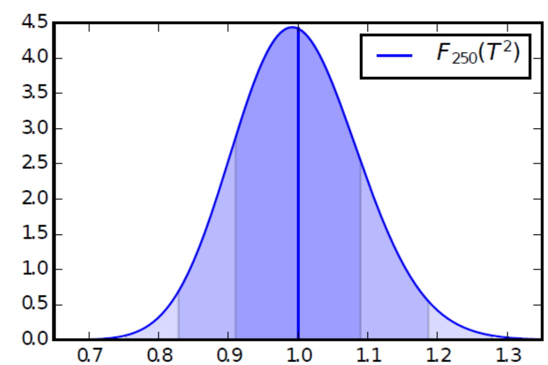
Example: chi(iw)

- 

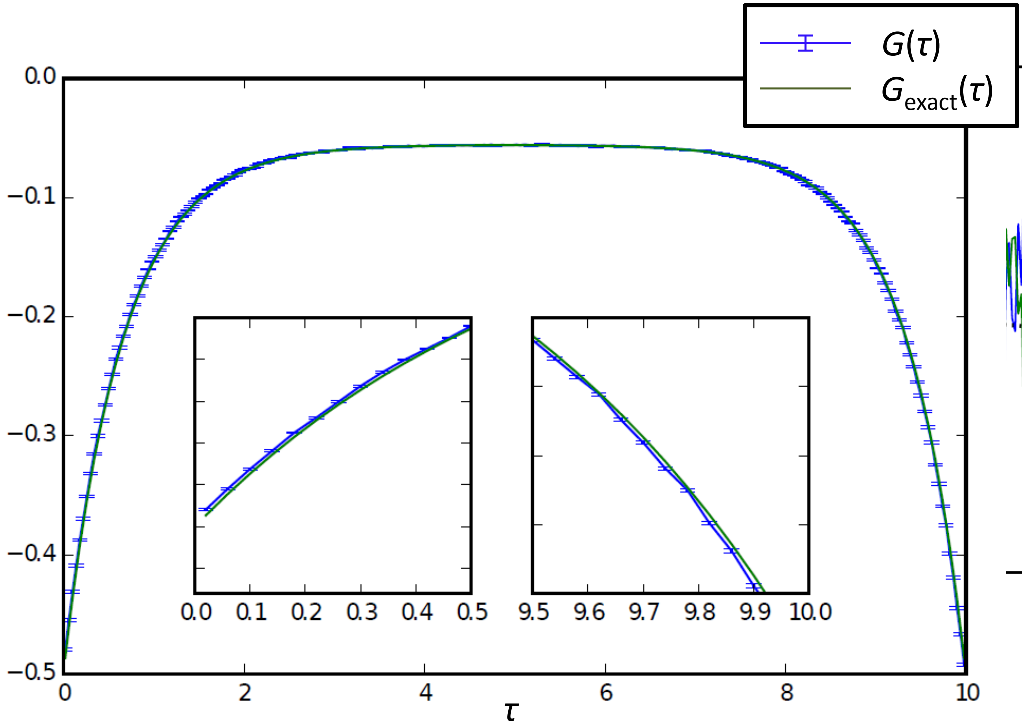


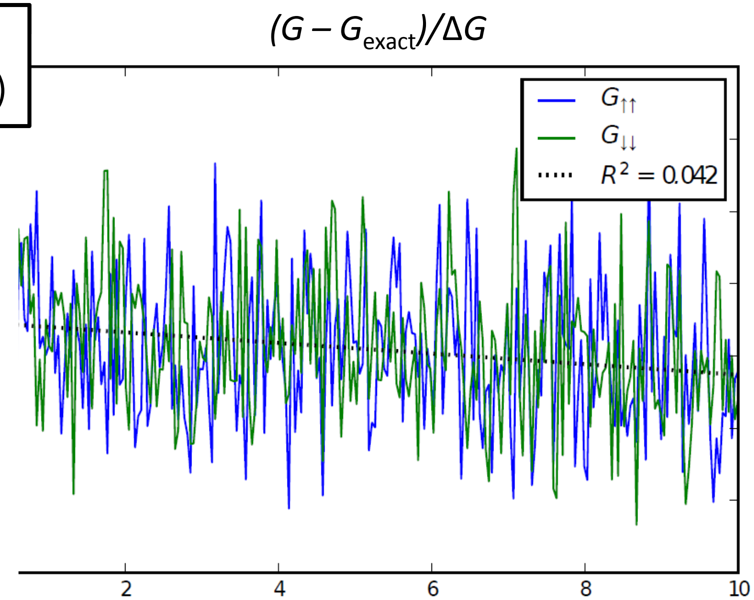
- one-sided *F* test:

$T^2 = 2.3$ (rejection at the 15σ level)



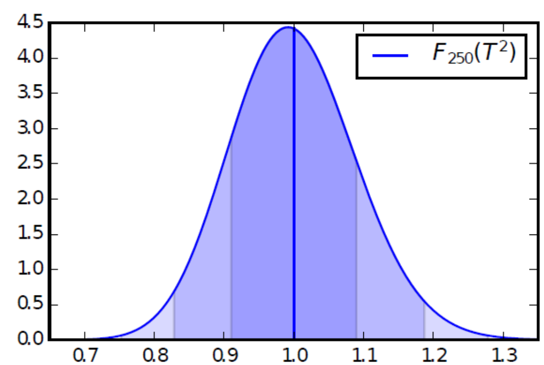
Example: chi(iw)

- 

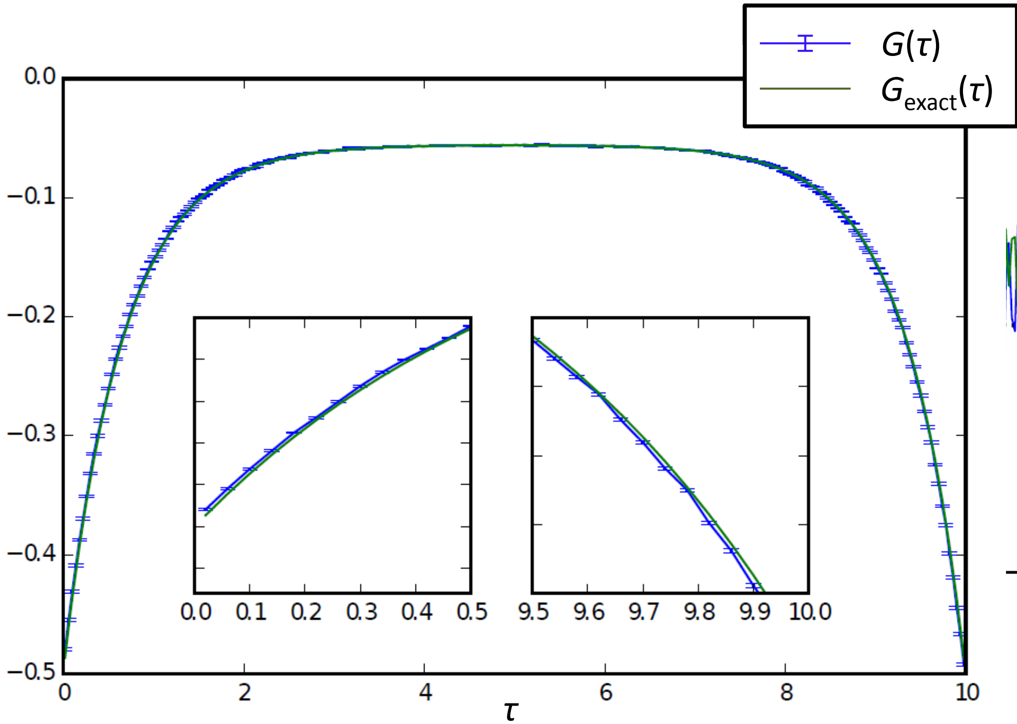


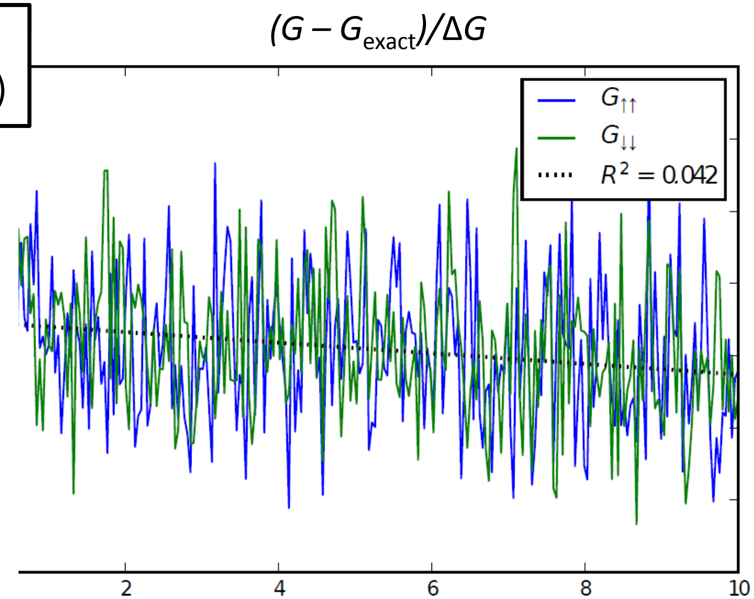
- one-sided *F* test:

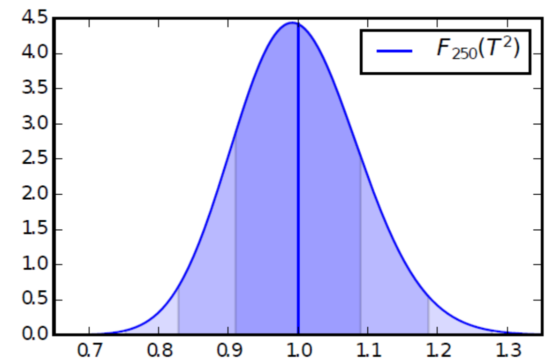
$T^2 = 2.3$ (rejection at the 15σ level)



Example: chi(iw)

- 





- **Reason**: bin shifting error.

Test for the error bars

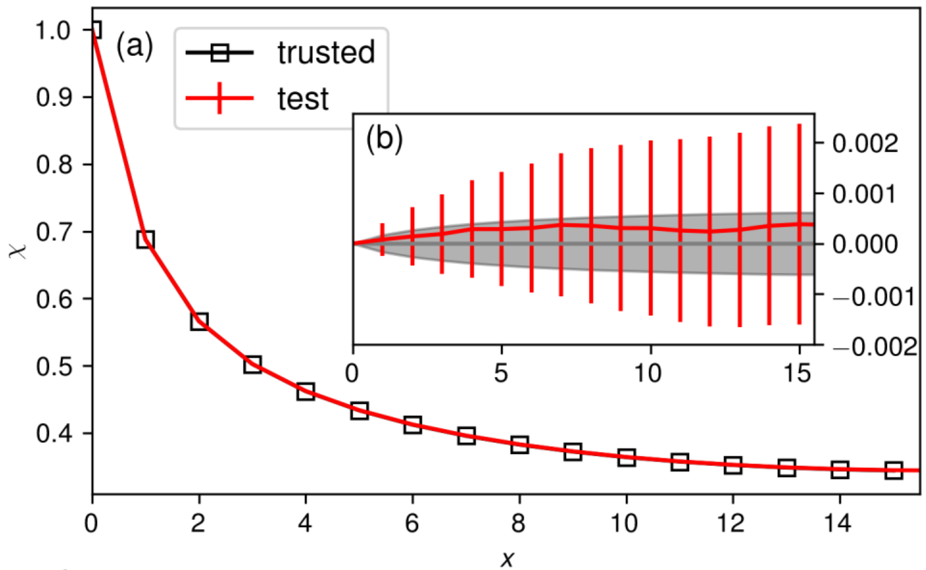
- possible for data series!
- null hypothesis $H_0 : \sigma = \sigma_0$
- lower alternate $H_1^- : \sigma < \sigma_0$ $P(F \geq t^2) < p$
- **upper alternate** $H_1^+ : \sigma > \sigma_0$ $P(F \leq t^2) < p$

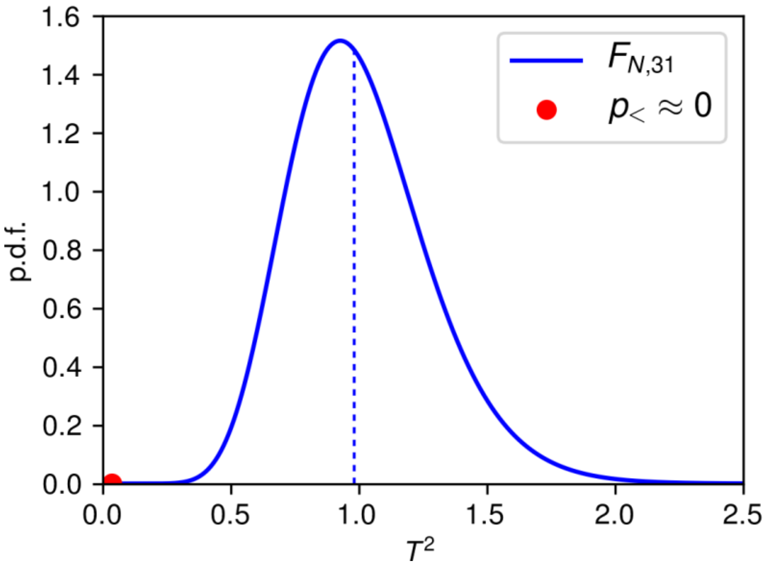
- Hotelling's T^2 test

$$T^2 = N(\langle x \rangle - x_0)^\dagger \Sigma_x^{-1}(\langle x \rangle - x_0) \qquad T^2 \sim \frac{n(N-1)}{N-n} F_{n,N-n}$$

Cross-correlated data

- $$\chi_{x,y} = \langle \sigma_{0,0}\sigma_{x,y} \rangle = \frac{1}{L^2}\langle \sum_{x,y,k,q} \mathcal{F}^{-1}_{x,y;k,q} |\mathcal{F}_{k,q;x',y'}\sigma_{x',y'}|^2 \rangle,$$





Cross correlated data

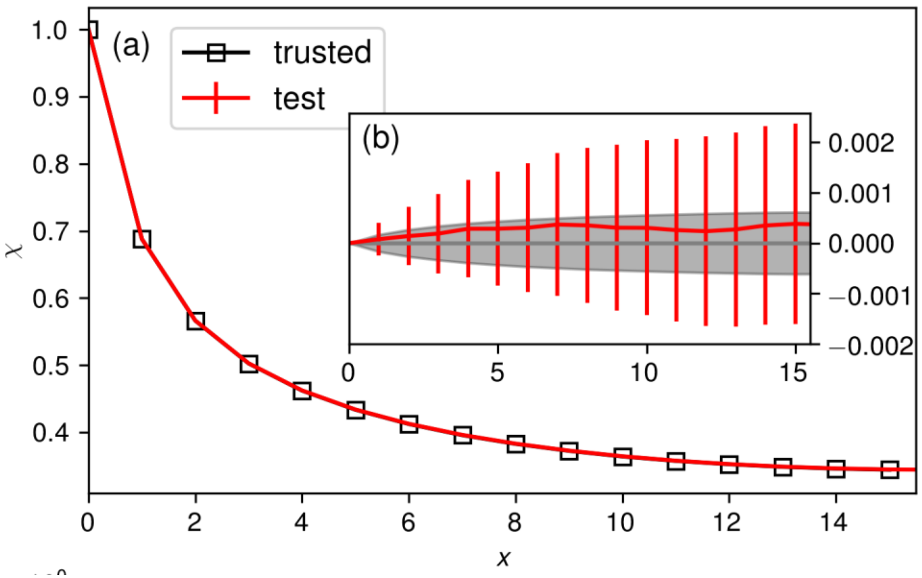
- use covariance matrix
- common complication: **duplicates**
- solution: projection to non-zero eigenvalues

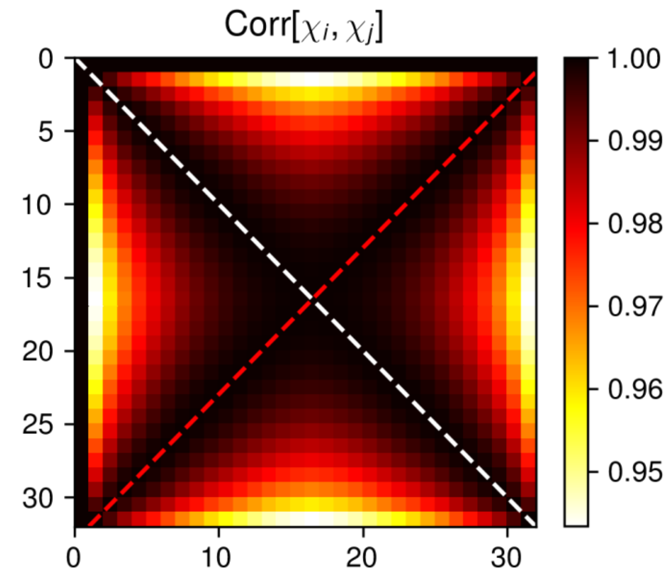
$$\Sigma = P \,\mathrm{diag}(s_1^2, \ldots, s_m^2)\; P^{\mathrm{T}},$$

- Hotelling's T^2 test

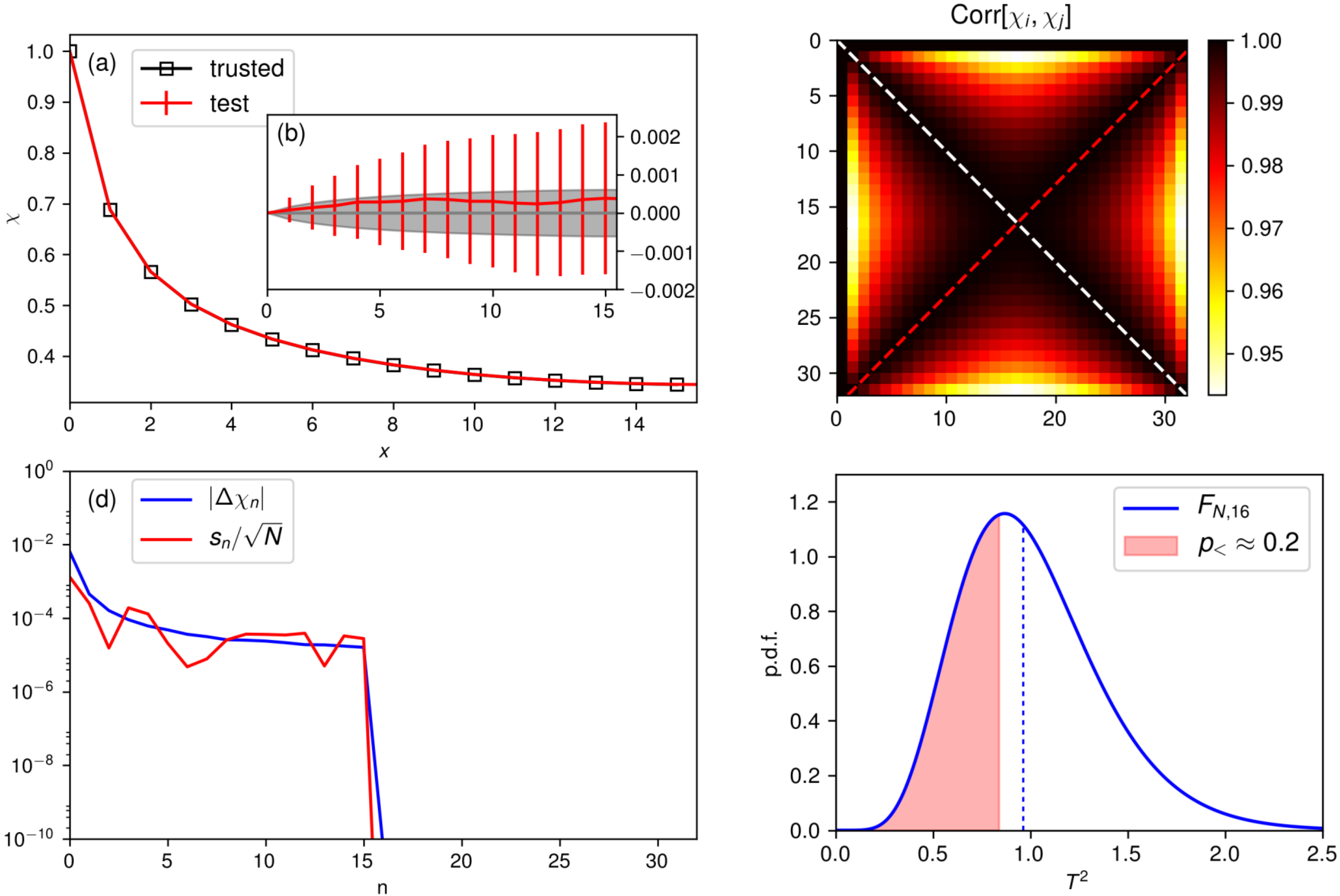
$$\sum_{i=1}^{m} \frac{|\sum_{k=1}^{n} P_{ki}(\langle X_k \rangle - y_k)|^2}{s_i^2/N} \sim \frac{m(N-1)}{N-m} F_{m,N-m} \,.$$

Cross-correlated data





Cross-correlated data



Conclusions

- Testing of stochastic codes
- Hypothesis testing powerful framework
- Should become standard tool in testing arsenal
- Outlook: stochastic fuzzing
- Outlook: part of ALPSCore testing framework[1]

[1] *A. Gaenko et al., Comput. Phys. Commun. 213, 235 (2017);* *http://www.alpscore.org*

BACKUP

Quantum Monte Carlo algorithms

- e.g., Anderson impurity model (AIM)

$$H = \underbrace{E_{ij} c_i^\dagger c_j + \tfrac{1}{2} U_{ikjl} c_i^\dagger c_j^\dagger c_l c_k}_{\text{impurity}} + \underbrace{(V_{pi} f_p^\dagger c_i + \mathrm{H.c.})}_{\text{hybridization}} + \underbrace{\epsilon_p f_p^\dagger f_p}_{\text{bath}}$$

- Solvers

– Exact diagonalization[1]	1960s	bath levels	~ 200 LOC
– Hirsch-Fye QMC[1]	1980s	imag. time	~ 2,000 LOC
– Continuous-time QMC[2]	2000s	none	~ 20,000 LOC

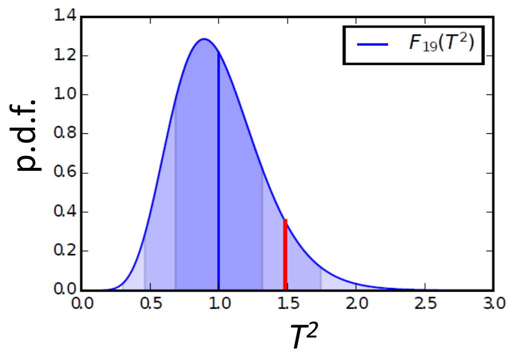
[1] *review: A Georges et al., Rev. Mod. Phys. 68, 13 (1996)*
[2] *review: E. Gull et al., Rev. Mod. Phys. 83, 349 (2011)*

Testing data series

- Usually stricter criterion
- Hotelling's T^2-test: e.g., Green's function

$$T^2 = N(\langle x \rangle - x_0)^\dagger \Sigma_x^{-1} (\langle x \rangle - x_0) \qquad T^2 \sim \frac{n(N-1)}{N-n} F_{n,N-n}$$

  - $\sigma > \sigma_0; P(F \geq t^2) < p$: systematic error
  - $\sigma < \sigma_0; P(F \leq t^2) < p$: error bars too large

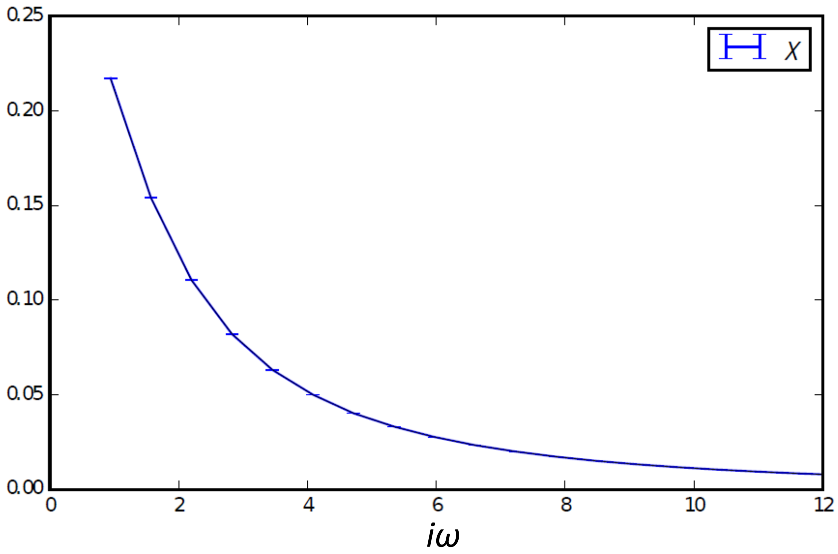


- Non-Gaussian batches/low statistics[1]

[1] *M. Marozzi: Stat. Meth. Med. Res. 25(6) 2593–2610 (2016)*

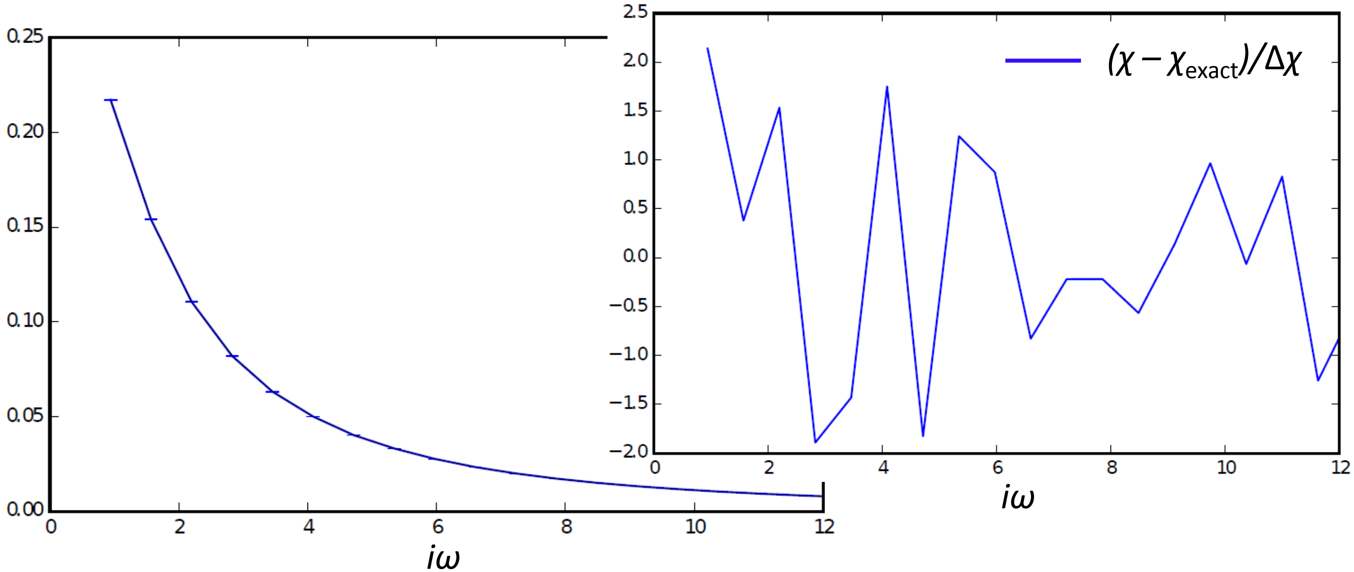
Example I:

- 1-orbital, 2 bath states (+/-0.5), V=1, U=1, mu=0.42



Example I:

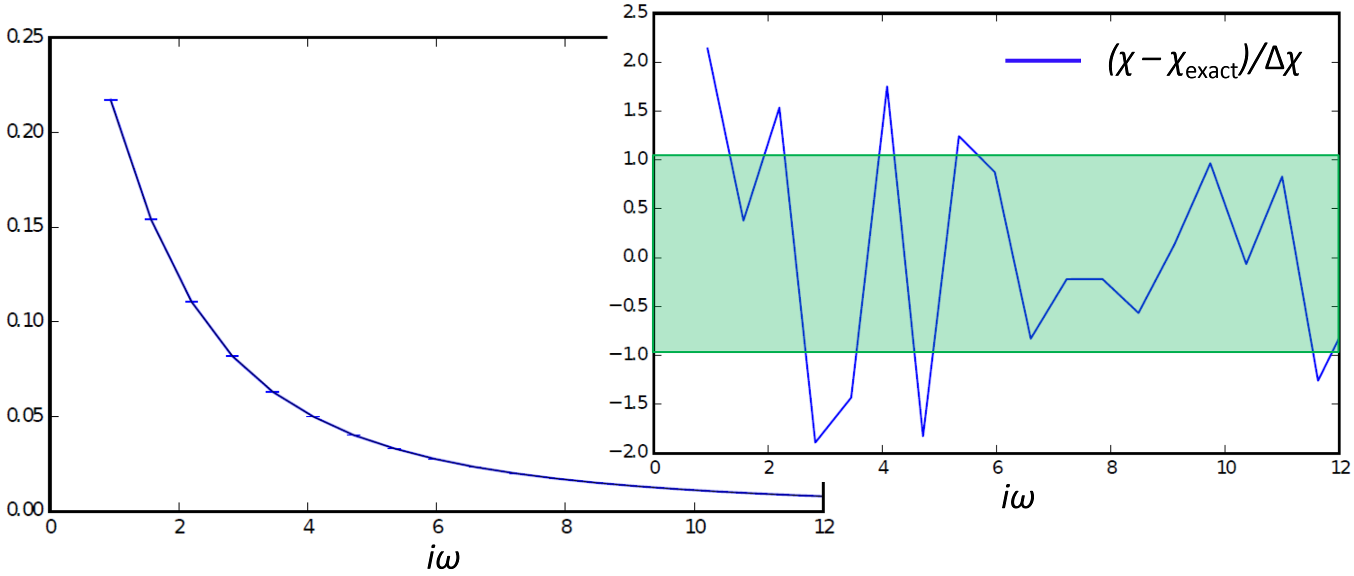
- 1-orbital, 2 bath states (+/-0.5), V=1, U=1, mu=0.42



[1] *M. Wallerberger, PhD thesis (2016)*

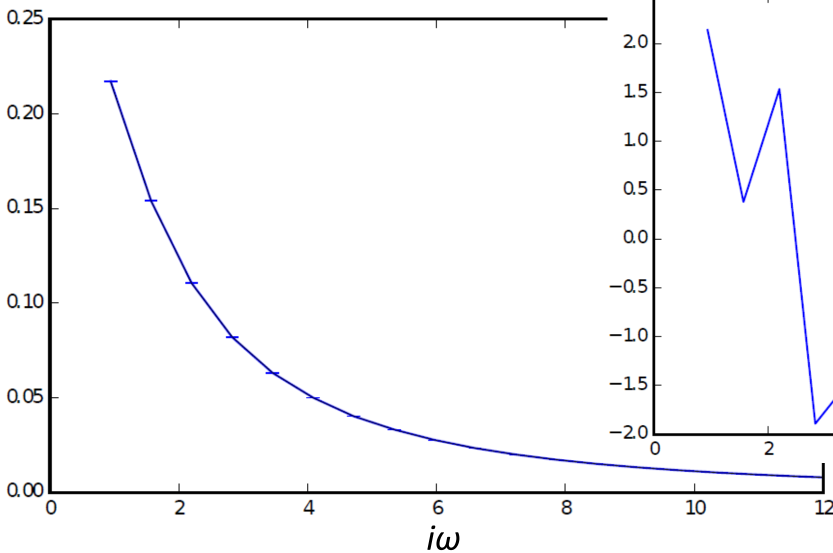
Example I:

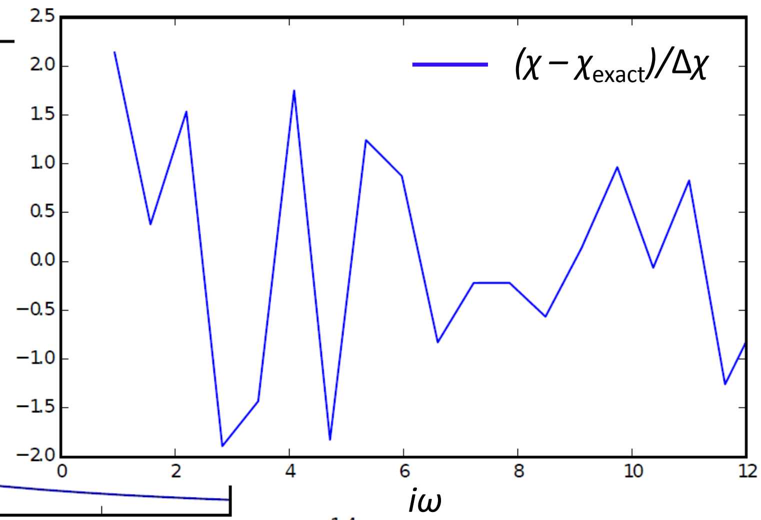
- 1-orbital, 2 bath states (+/-0.5), V=1, U=1, mu=0.42



Example I:

- 1-orbital, 2 bath states (+/-0.5), V=1, U=1, mu=0.42





- one-sided *F* test:

$$T^2 \approx 1.48 \;\; (\, p \approx 0.08 \;)$$