Covariance matrices

Introduction and challenges

Pierre Burger

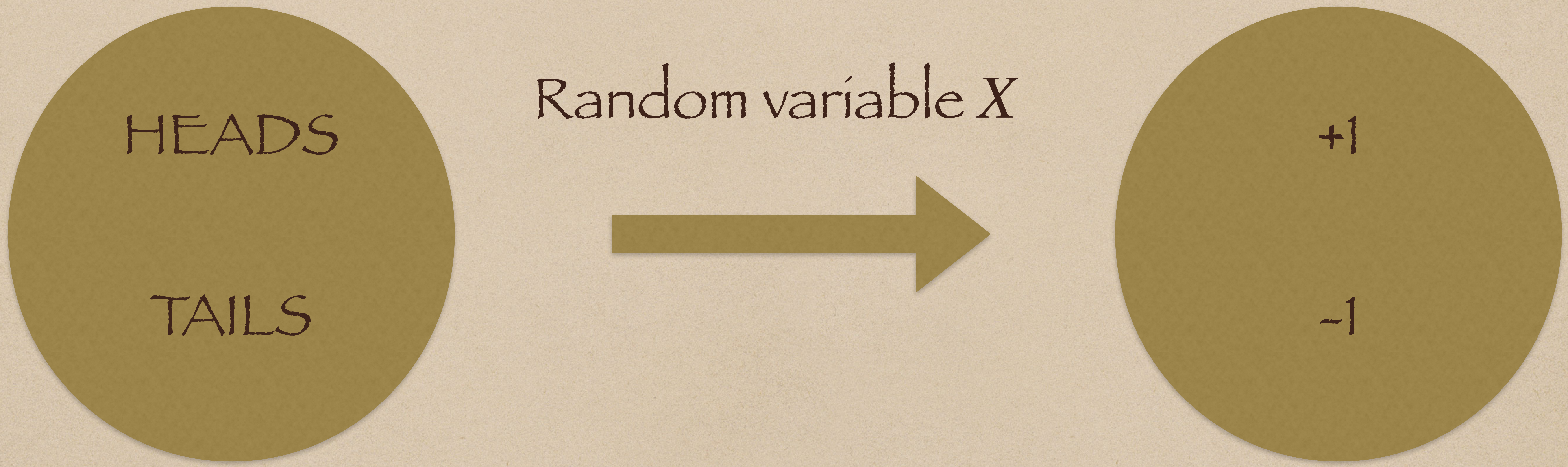
Random Variable

- A random variable is a quantity or object which depends on random events.
- Mathematically: A random variable X is a measurable function $X : \Omega \rightarrow E$ that is a mapping from observable space Ω to the measurable space E.

Wikipedia

Random Variable

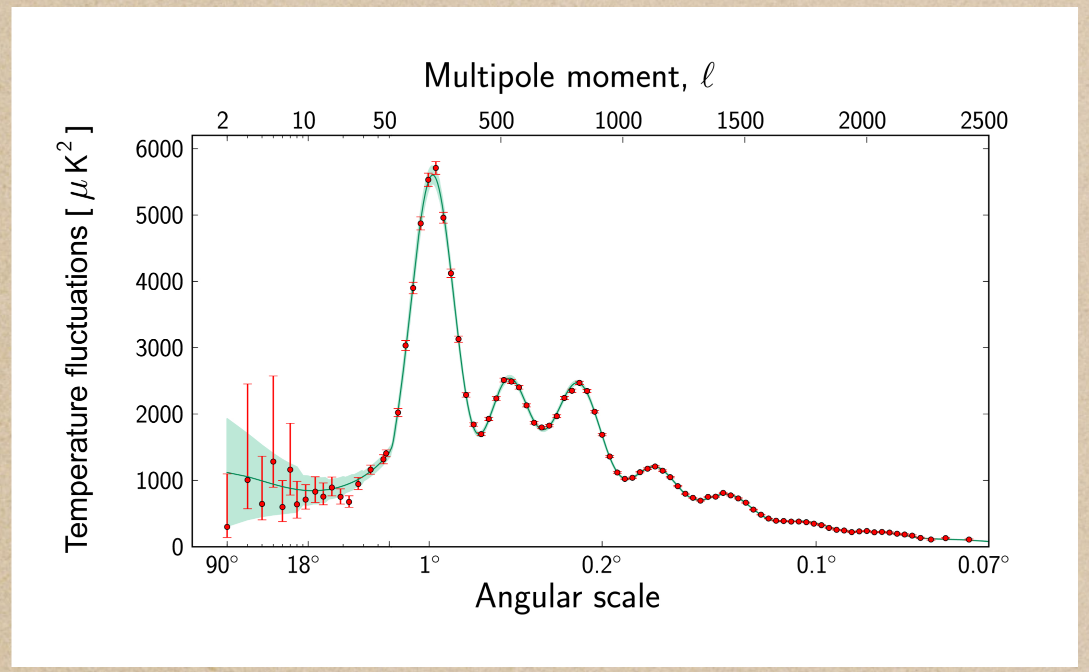
Observable space Ω Measurable space E



Random Variable

Observable space Ω

Measurable space E



ESA and the Planck Collaboration

Random variable Ω_m

0.32

0.31

0.41

0.28

0.19

0.23

NASA/ESA and The Hubble Heritage Team (STScI/AURA)

Probability density function

- The probability that a random variable takes a value between a and b is given by its probability density function, $f_X(x)$:

$$\Pr[a \le X \le b] = \int_a^b f_X(x)\,\mathrm{d}x$$

- The probability density function is normalised

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$$

- The expectation value is given by:

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\,f_X(x)\,\mathrm{d}x$$

- And the variance:

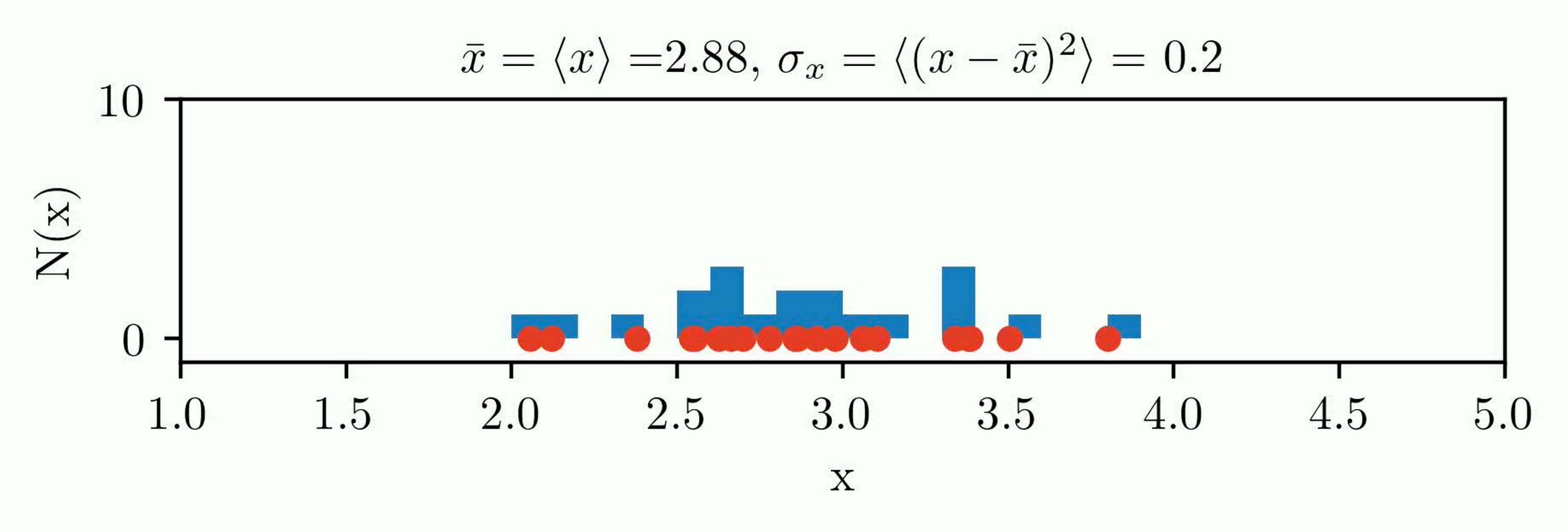
$$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = E\left[X^2\right] - E[X]^2$$

Wikipedia

Random Variable

The 1D case:

	Truth	Estimator
	$\mathrm{E}[X] = 3$	$\mathrm{E}[X] \approx \bar{x} = \frac{1}{n}\sum_i^n x_i$
	$\mathrm{Var}(X) = 0.25$	$\mathrm{Var}(X) = \sigma_x^2 \approx \frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$



Random Variable

The 2D case: Two random variables X and Y

Joint pdf: $\Pr[a \leq X \leq b, c \leq Y \leq d] = \int_a^b \int_c^d f_{XY}(x, y)\, \mathrm{d}x\, \mathrm{d}y$

Marginalised pdf: $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\ \mathrm{d}y$

If X is independent from Y it follows that $f_{XY}(x, y) = f_X(x) f_Y(y)$

Random Variable

The 2D case:

Compared to the 1D case we can define the cross-variance between X and Y:

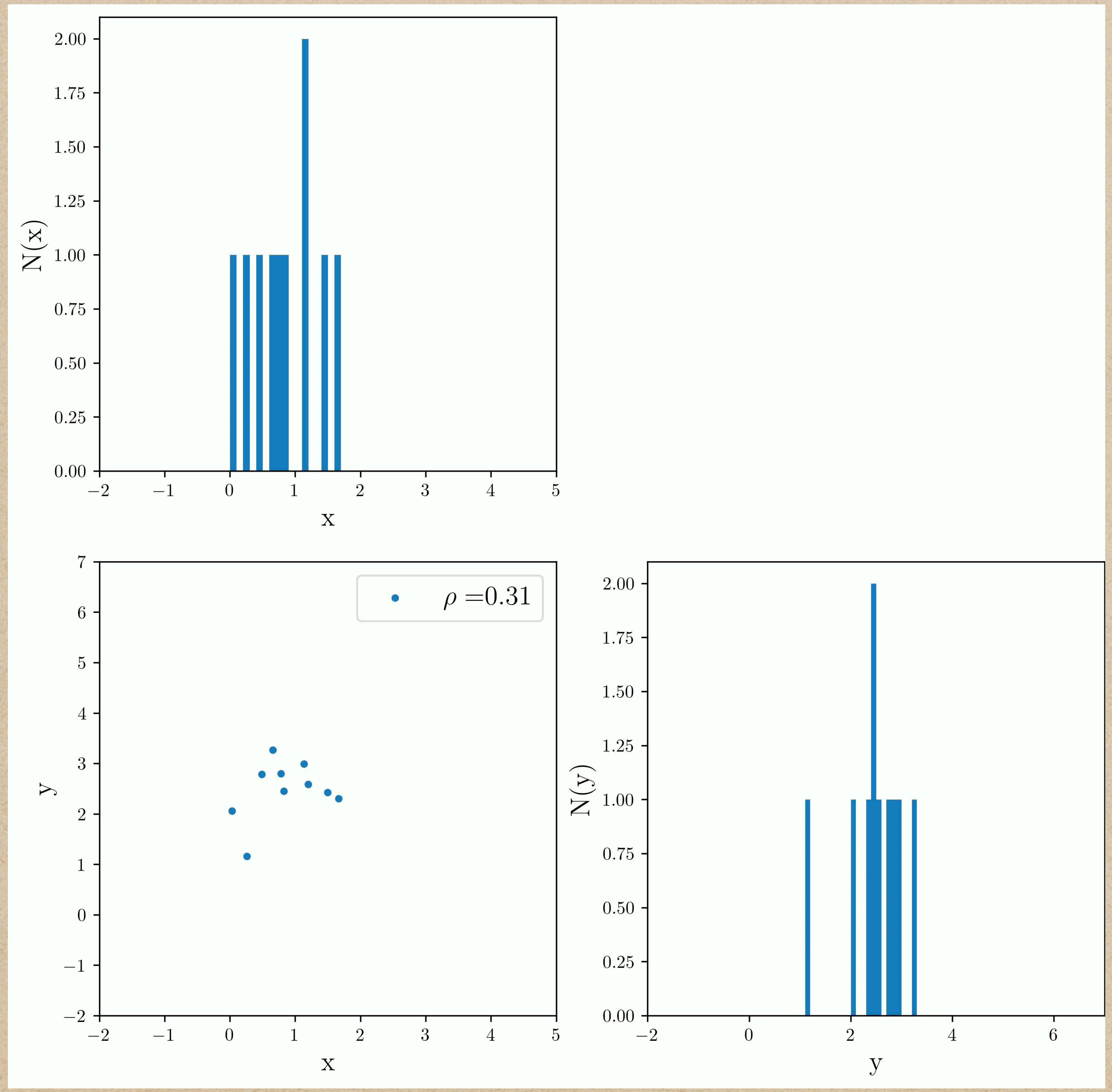
$$\mathrm{Var}(X, Y) = E\left[(X - E[X])(Y - E[Y])\right] \approx \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

A summary of all variances is the so called <u>covariance matrix</u>

$$\mathrm{Cov}(X, Y) = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Var}(Y, X) \\ \mathrm{Var}(X, Y) & \mathrm{Var}(Y) \end{pmatrix}$$

Random Variable

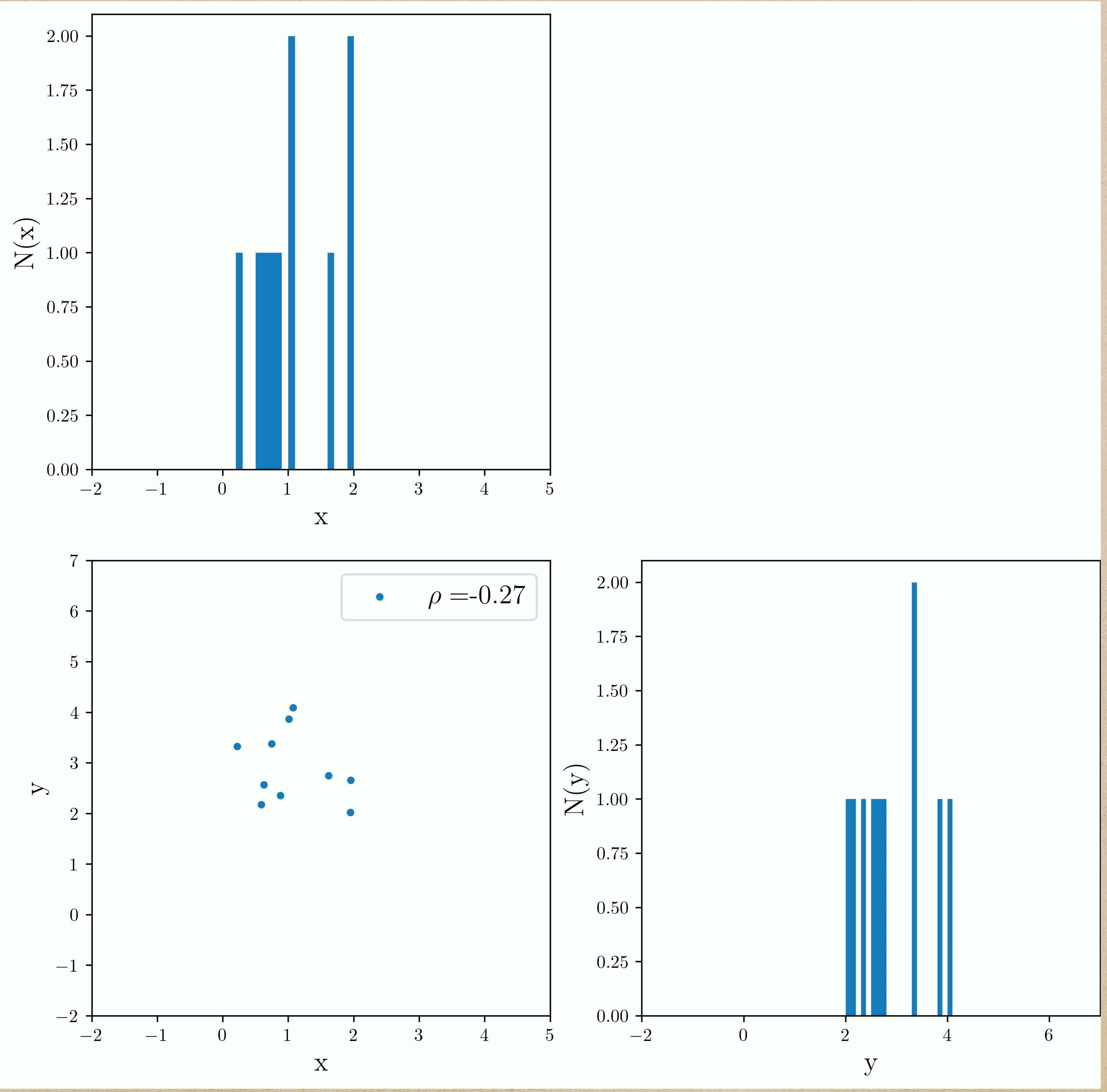
The 2D case:

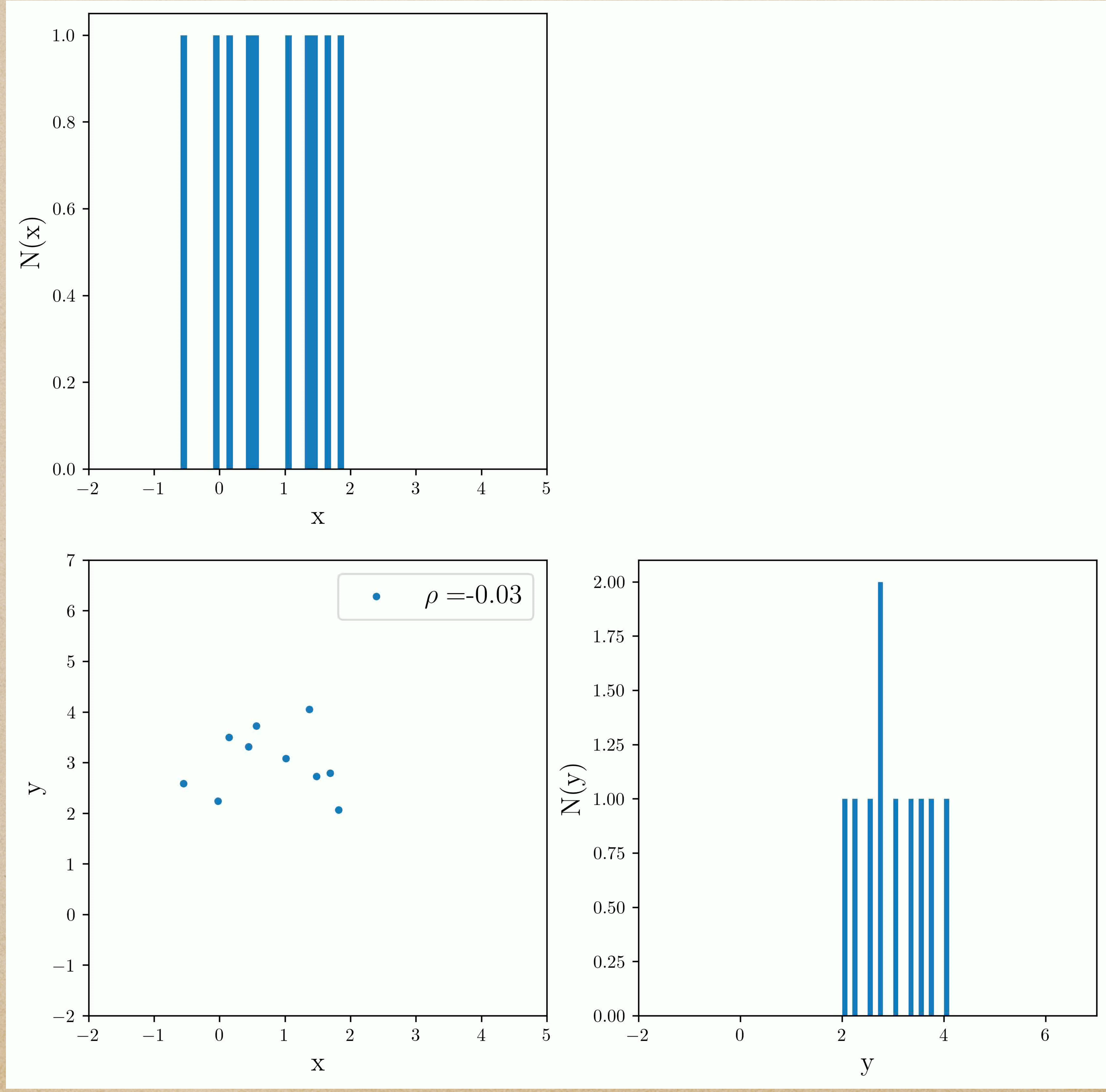


$$\rho(X, Y) = \frac{\mathrm{Var}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

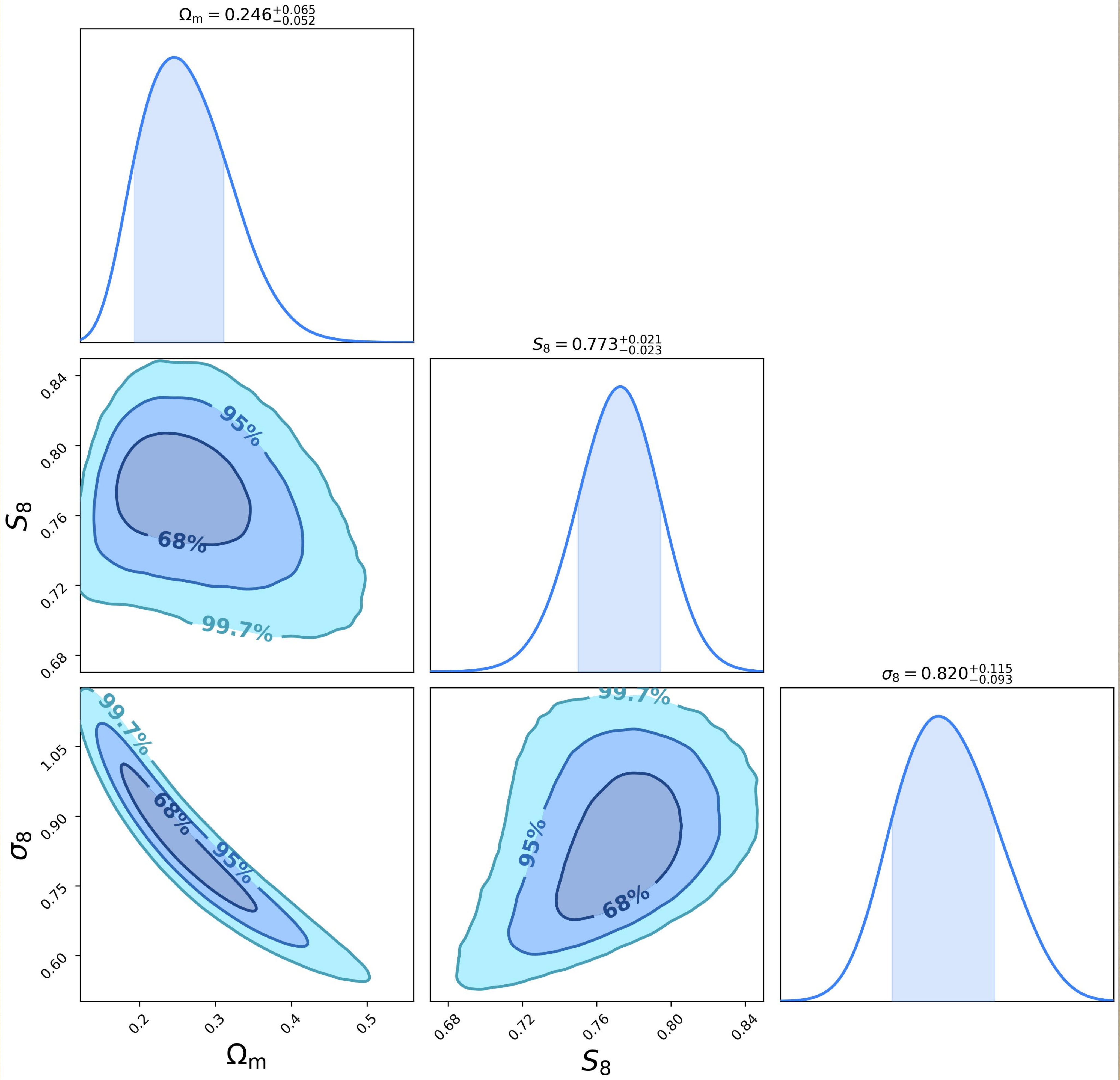
$$\mathrm{Corr}(X, Y) = \begin{pmatrix} 1 & \rho(X, Y) \\ \rho(X, Y) & 1 \end{pmatrix}$$

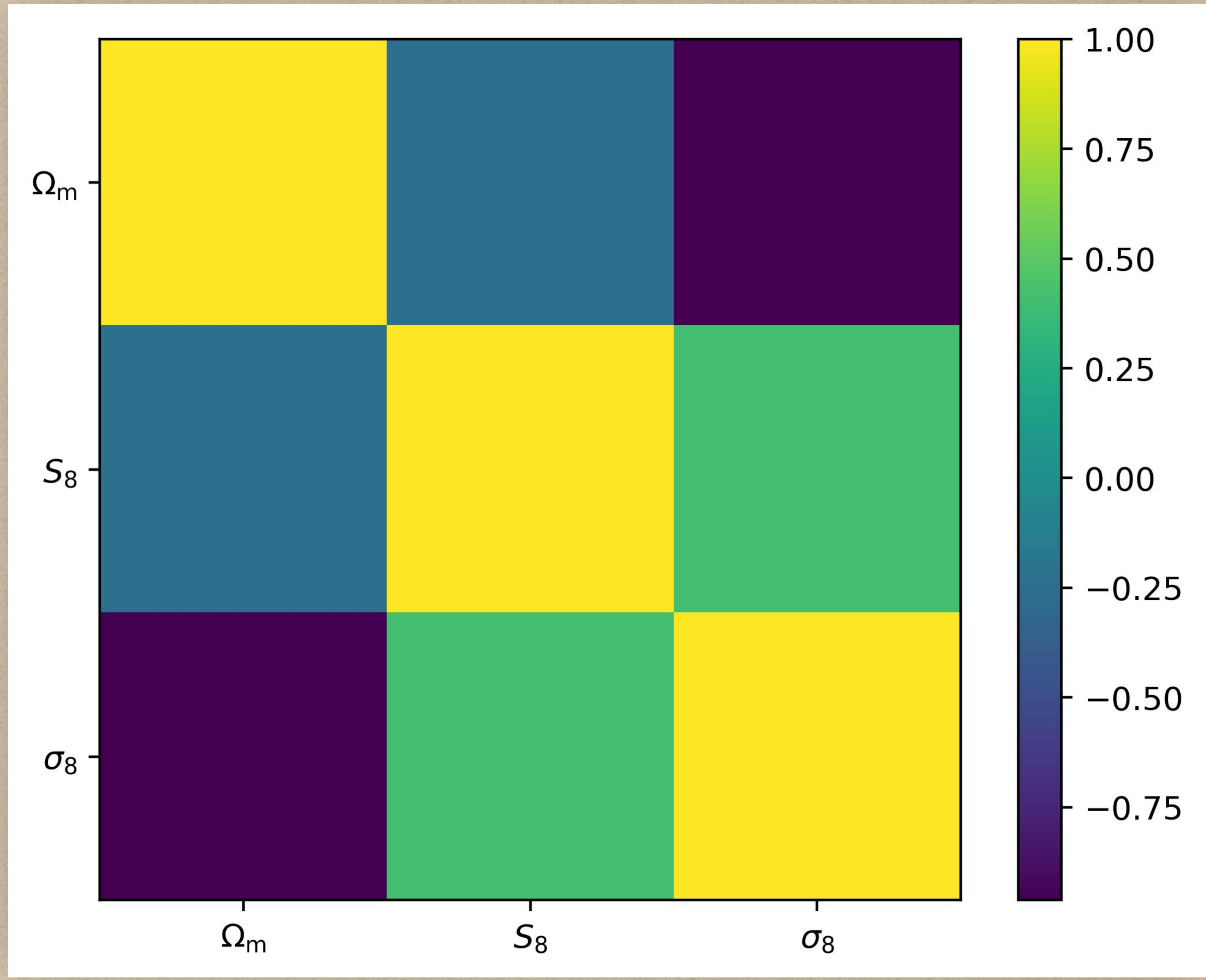
Random Variable





Random Variable: Example

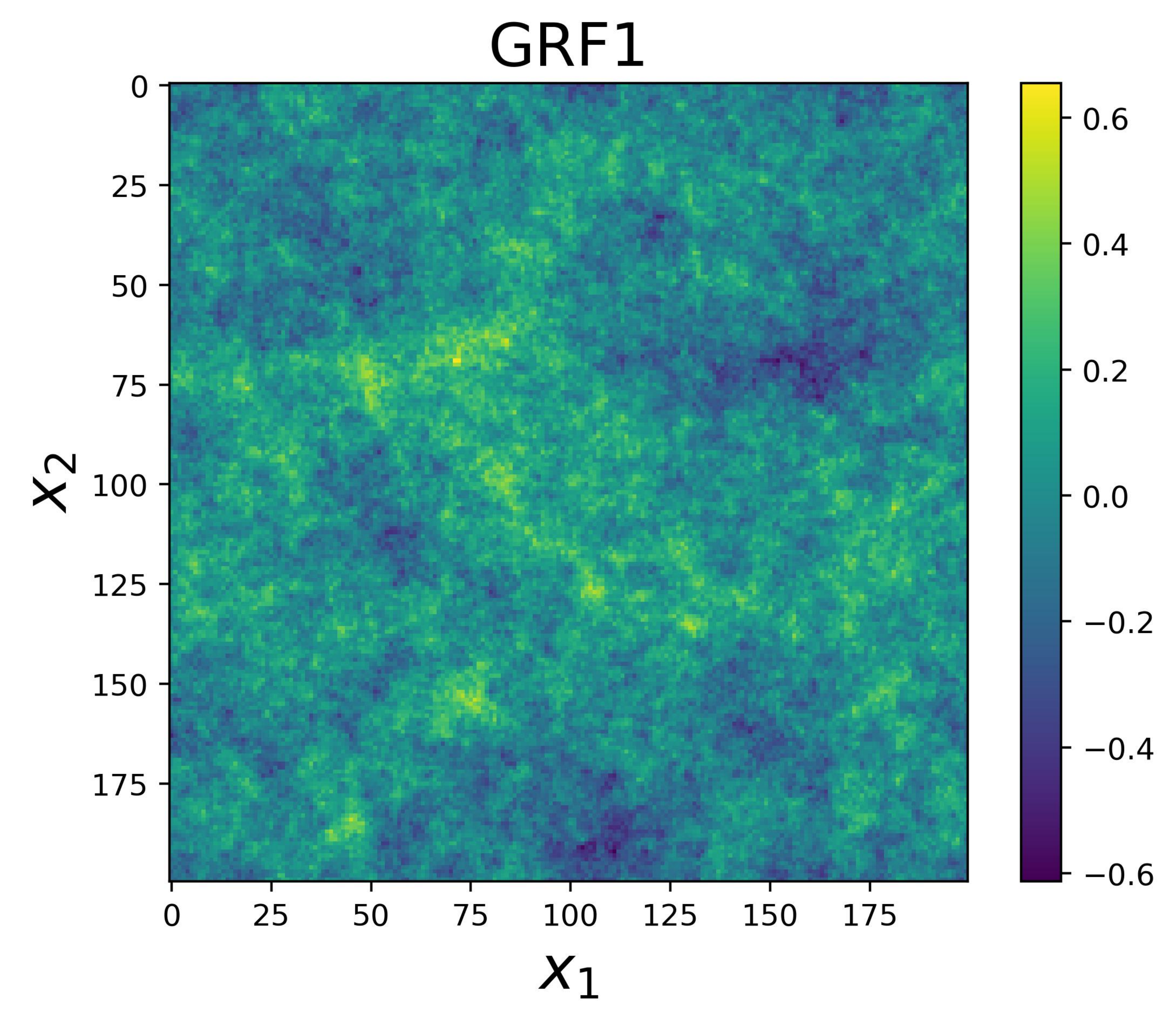


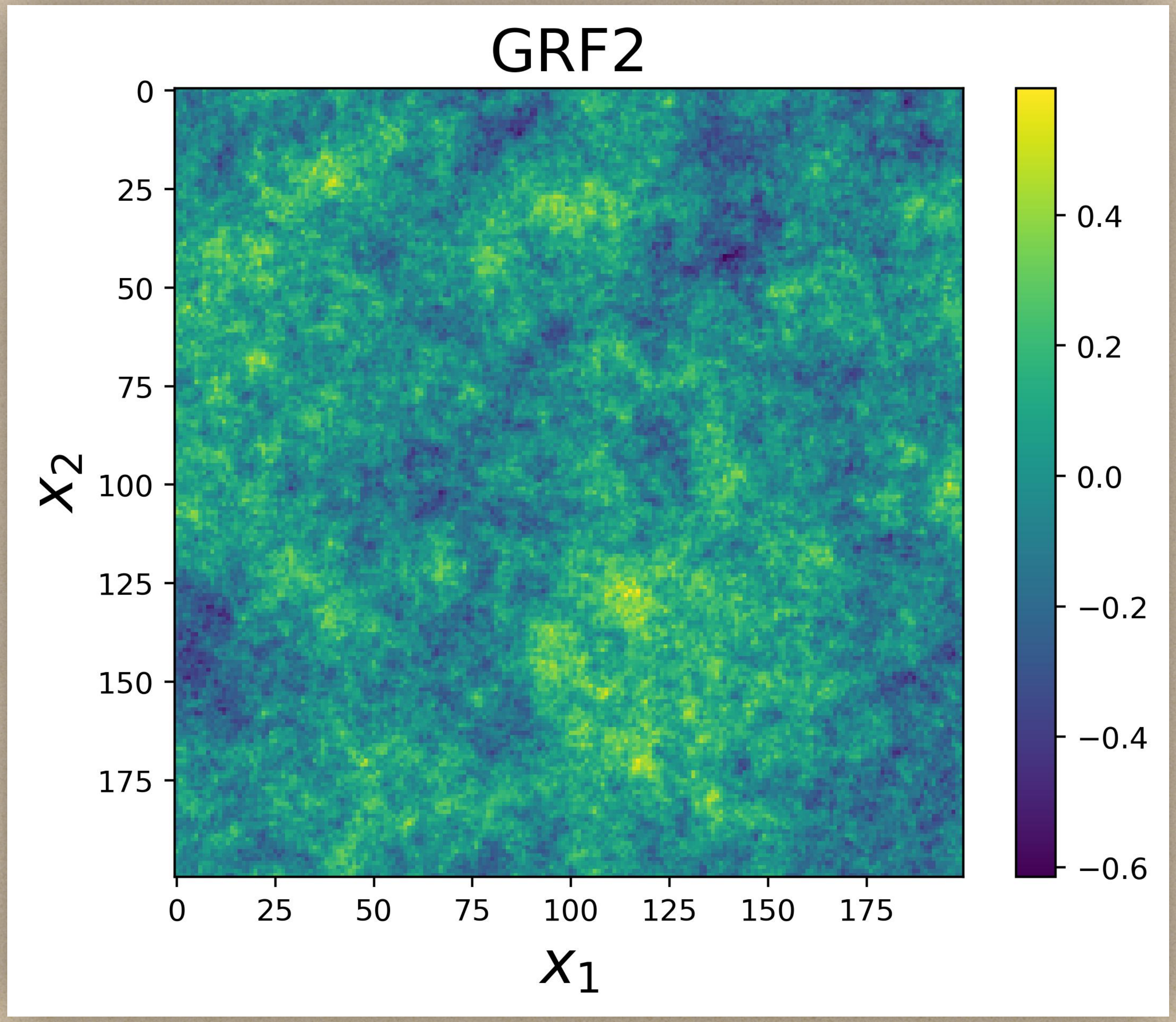


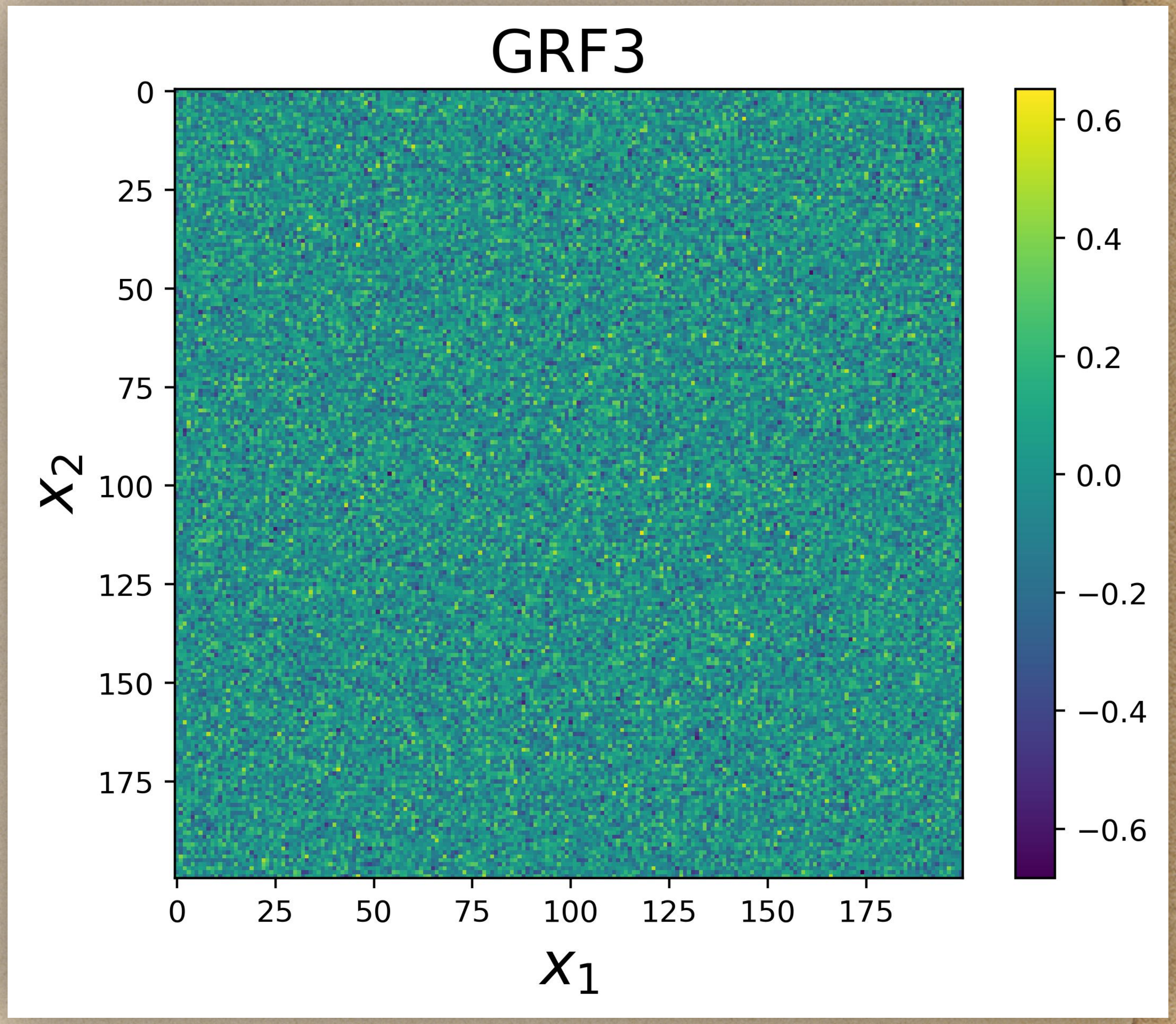
Random Fields

- A random field is a statistical ensemble of all fields with the same statistical properties.
- The individual realisations $\delta(\boldsymbol{x}, t)$ are different due to different initial conditions.
- Mathematically the random field is characterised by the joint probability distribution $p(\delta_1 \ldots \delta_n)\, \mathrm{d}\delta_1 \ldots \mathrm{d}\delta_n$, where δ_i is $\delta(\boldsymbol{x}, t)$ at grid point i.

Random Fields: Example







Correlation functions

- The n-point correlation function is defined by:

$$\xi^{(n)}(\boldsymbol{x}_1, \dots \boldsymbol{x}_n, t) = \langle \delta(\boldsymbol{x}_1, t) \dots \delta(\boldsymbol{x}_n, t) \rangle$$

$$= \int \mathrm{d}\delta(\boldsymbol{x}_1, t) \dots \mathrm{d}\delta(\boldsymbol{x}_n, t)\, \delta(\boldsymbol{x}_1, t) \dots \delta(\boldsymbol{x}_n, t)\, p[\delta(\boldsymbol{x}_1, t) \dots \delta(\boldsymbol{x}_n, t)]$$

- This requires the average over all possible realisations.
- Using the Ergodicity Principle (Peebles 1980) the ensemble average are equivalent to spatial averages:

$$\xi^{(n)}(\boldsymbol{x}_1, \dots \boldsymbol{x}_n, t) = \frac{1}{V^n} \int_V \mathrm{d}\boldsymbol{x}_1 \dots \int_V \mathrm{d}\boldsymbol{x}_n\, \delta(\boldsymbol{x}_1, t) \dots \delta(\boldsymbol{x}_n, t)$$

Correlation functions

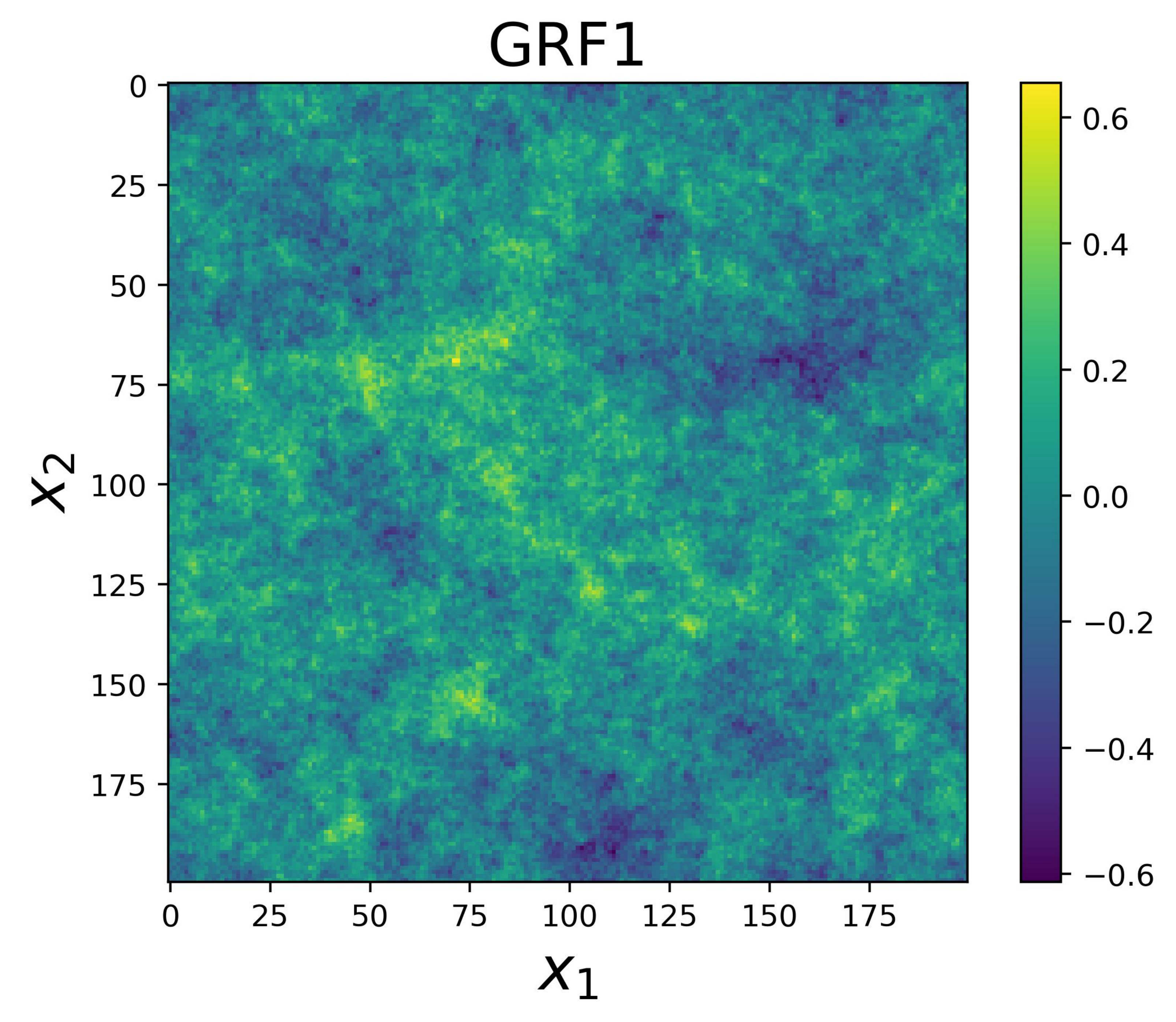
- The most famous is the 2-point correlation function

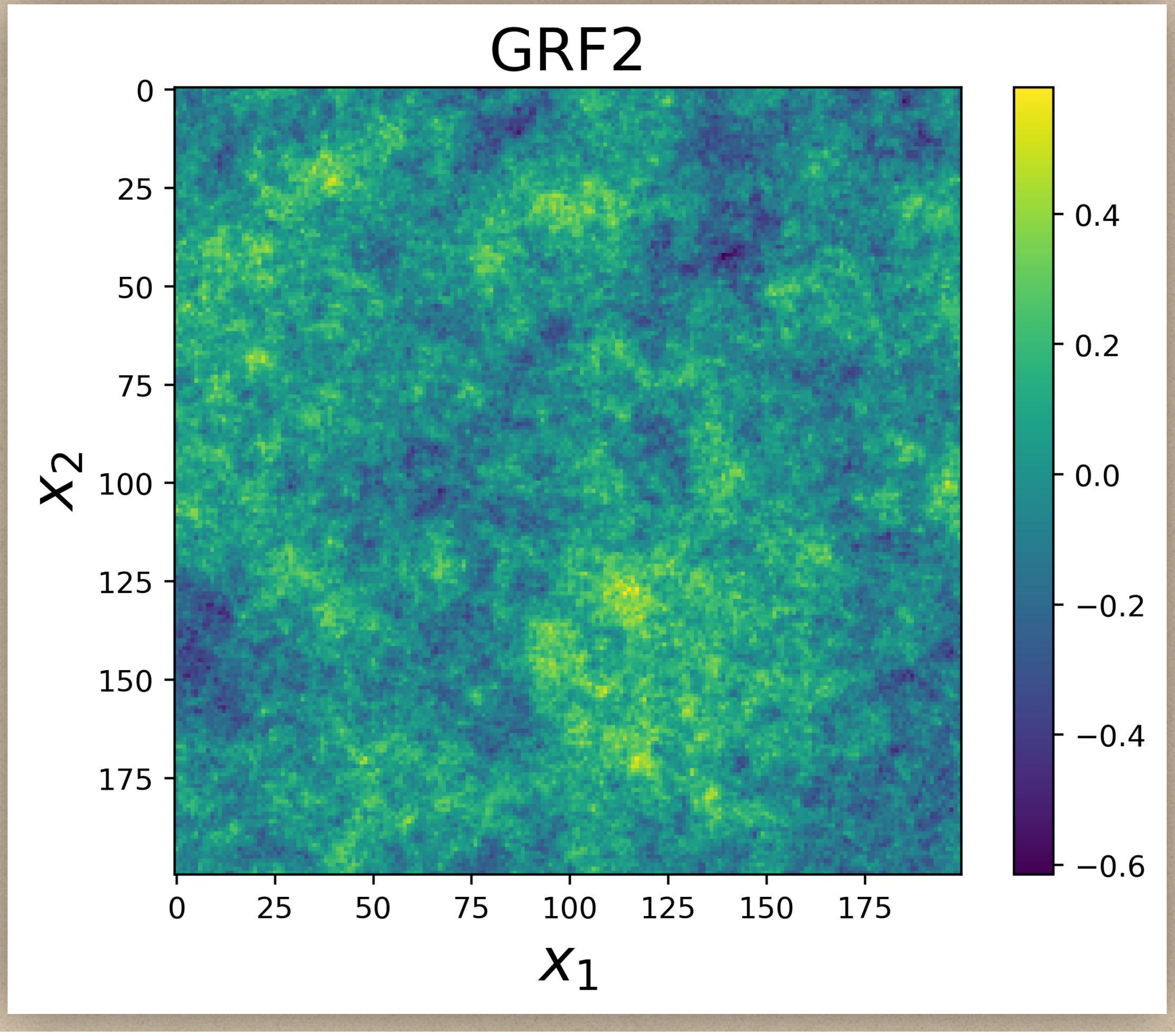
$$\xi(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = \frac{1}{V^2} \int_V \mathrm{d}\boldsymbol{x}_1 \int_V \mathrm{d}\boldsymbol{x}_2 \, \delta(\boldsymbol{x}_1, t) \, \delta(\boldsymbol{x}_2, t)$$

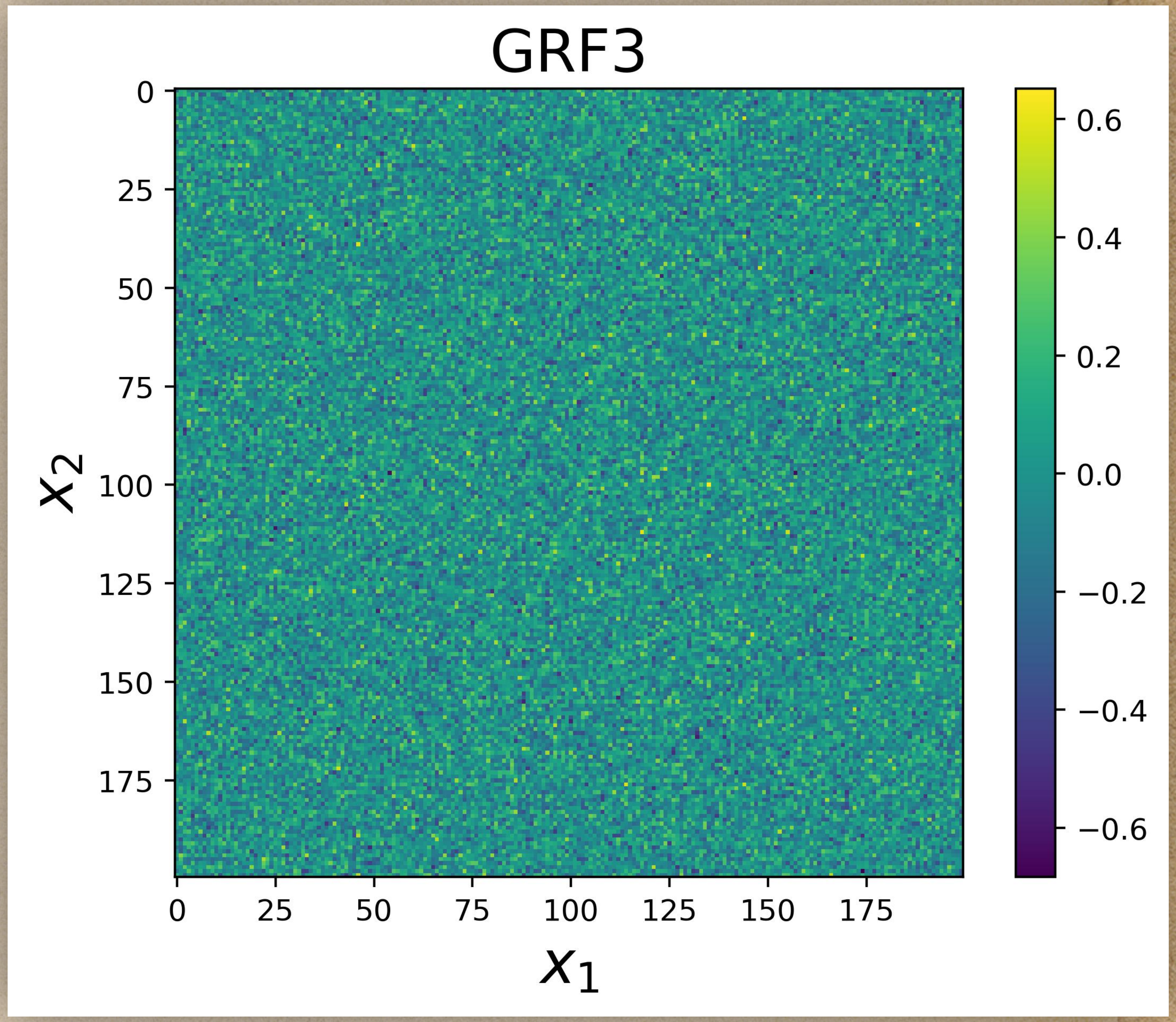
- For an isotropic and homogeneous field, the correlation functions simplifies to

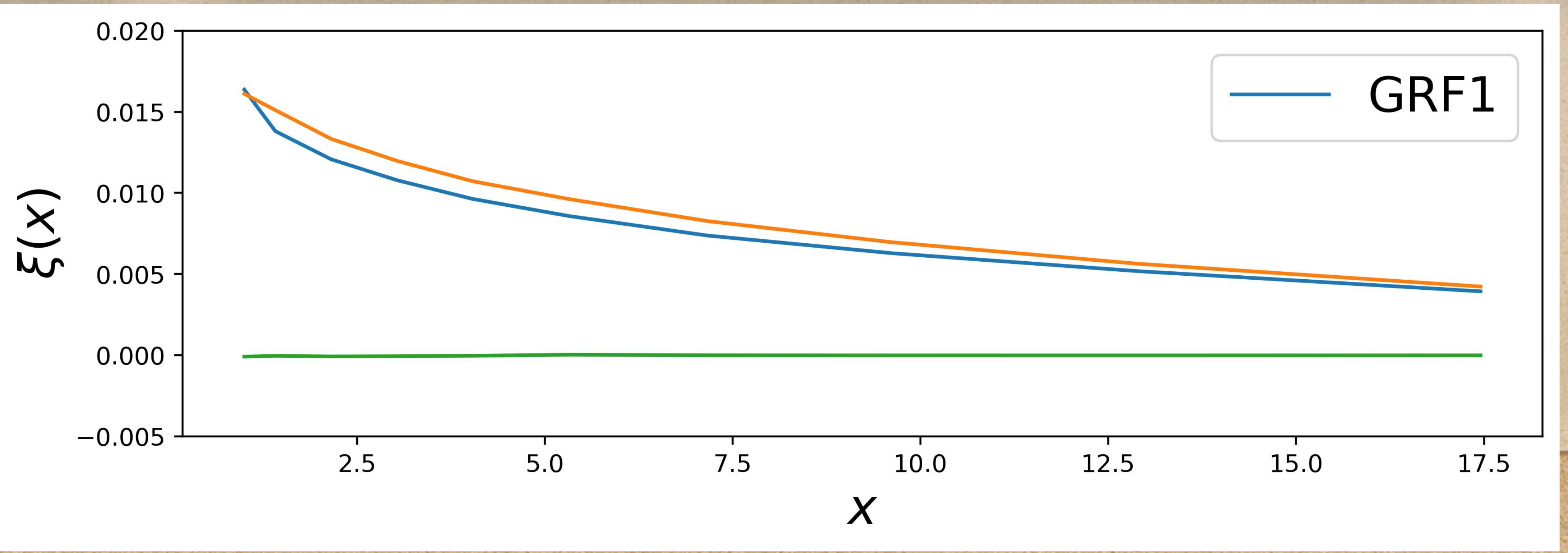
$$\xi(\boldsymbol{x}_1, \boldsymbol{x}_2, t) \to \xi(|\boldsymbol{x}_1 - \boldsymbol{x}_2|, t)$$

Correlation functions: Example









Computations of covariance matrices

- If the experiment can be repeated to measure variance from all individual results -> Frequentist approach
- But we have only one observable Universe?

There are three possible solutions

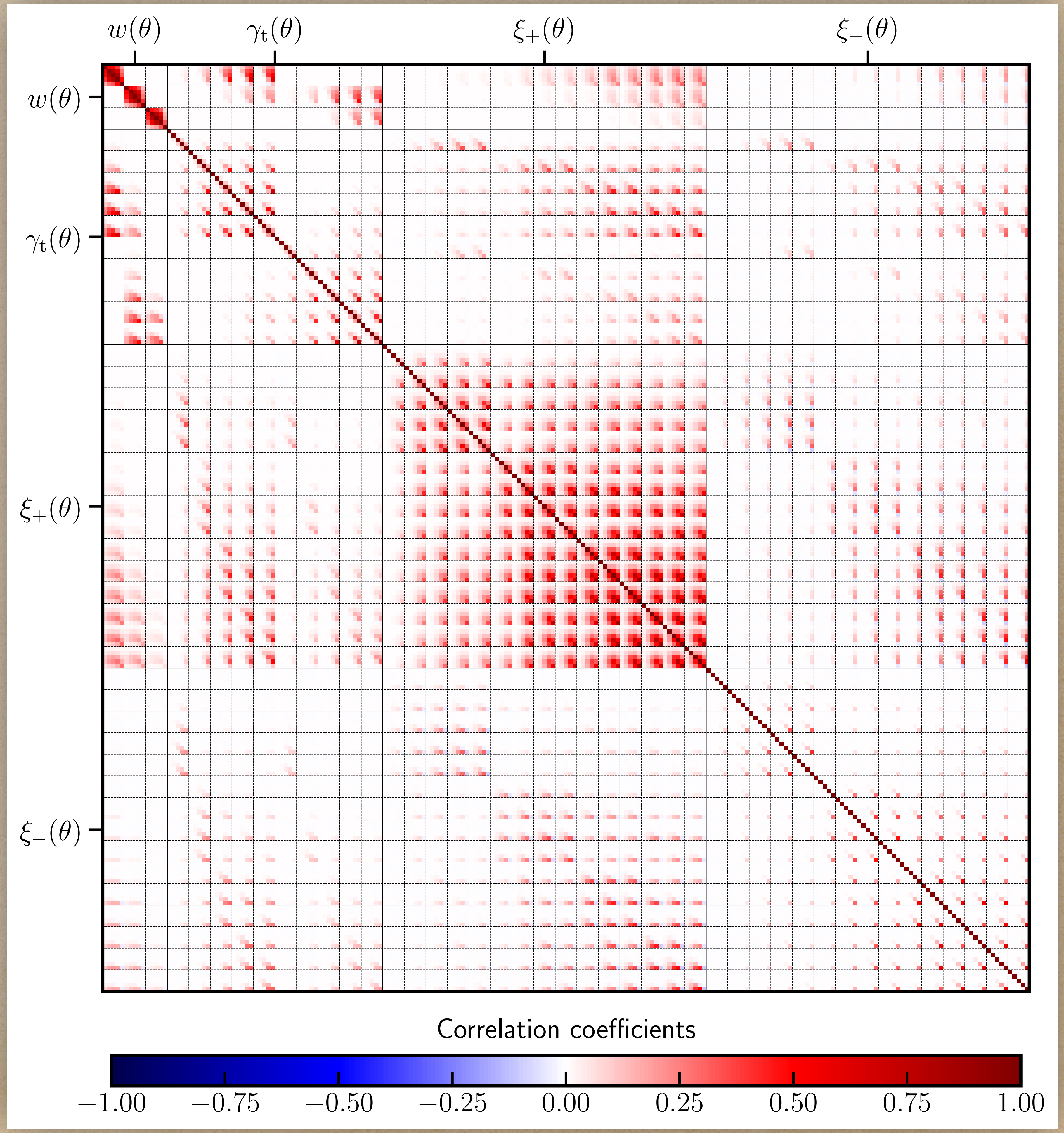
1. Analytical descriptions
2. Numerical simulations
3. The real data itself -> Jackknife/Bootstrap

Analytical descriptions

- Assume you have an estimator of your summary statistic $\Xi(R)$

- In general, covariance is then given by:

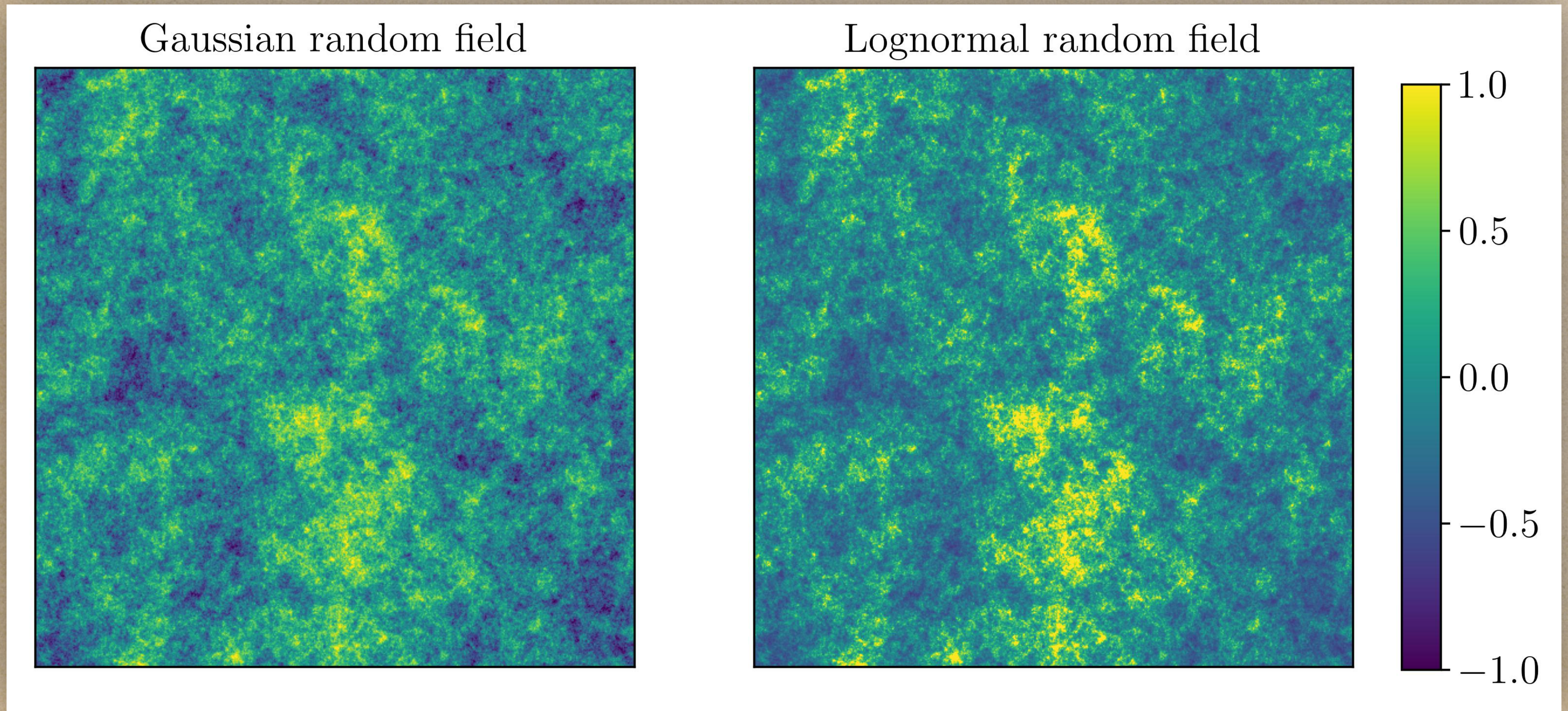
$$C_{\Xi}(R_1, R_2) = \langle \Xi(R_1)\Xi(R_2) \rangle - \langle \Xi(R_1) \rangle \langle \Xi(R_2) \rangle$$

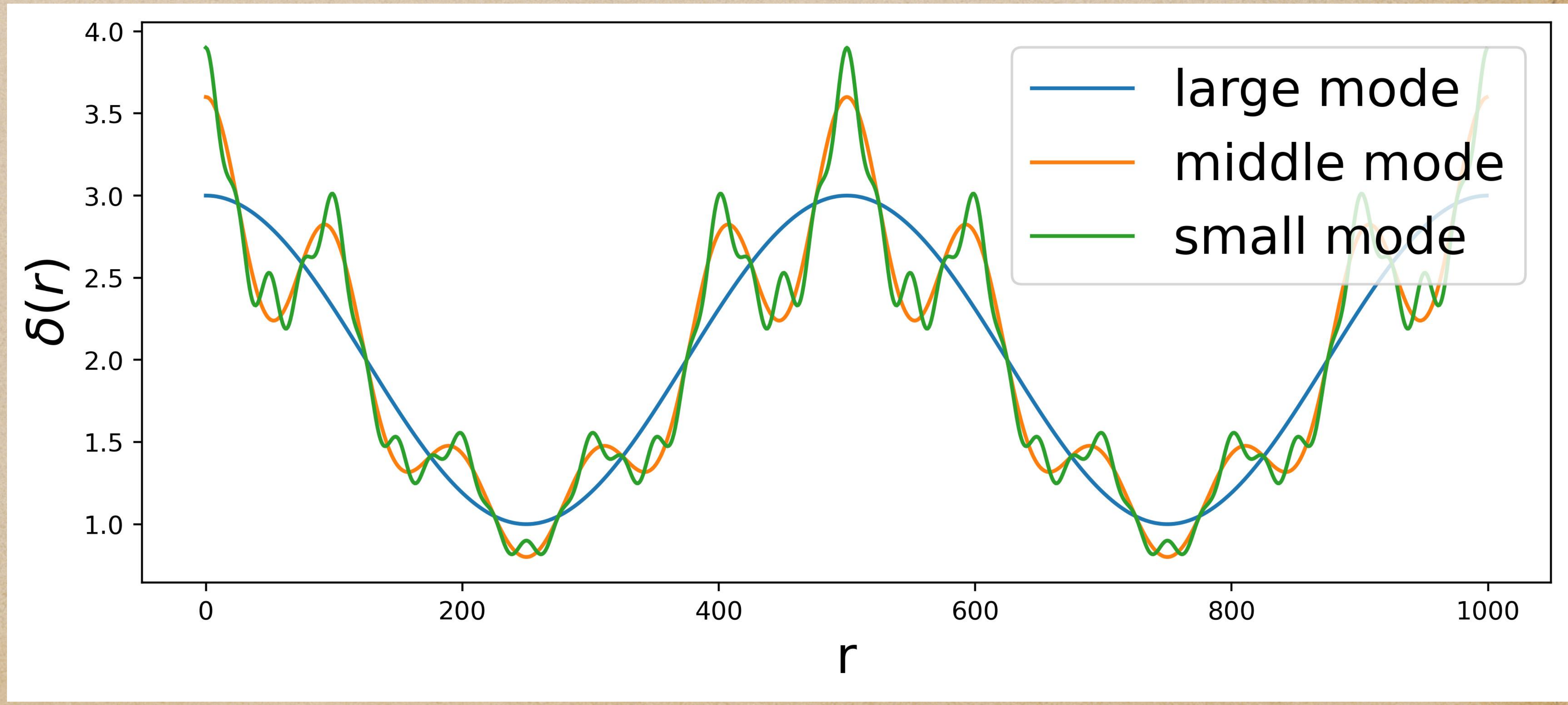


Analytical descriptions: Example

In large scale structure cosmology

$$\mathrm{C}_{\Xi}(R_1, R_2) = \mathrm{C}_{\Xi}^{\mathrm{G}}(R_1, R_2) + \mathrm{C}_{\Xi}^{\mathrm{NG}}(R_1, R_2) + \mathrm{C}_{\Xi}^{\mathrm{SSC}}(R_1, R_2)$$





Numerical simulations for covariance matrices

- Using N-body simulations it is possible to compute the sample covariance matrix

$$\mathrm{Var}(X, Y) \approx \frac{1}{n-1} \sum_{i}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Advantages compared to analytical descriptions:

- Naturally captures all non-Gaussian information
- <u>If enough realisations are used always invertible</u>
- No complicate mathematical considerations

Disadvantages:

- Usually takes longer
- Need to trust simulations
- <u>Itself is a random variable</u>

Why do we need to invert a covariance matrix?

Parameter estimation:

- Markov chain Monte Carlo:

$$P(\Theta, \boldsymbol{d}, C) \propto |C(\Theta)|^{-1/2} \exp\left[-2\chi^2(\Theta)\right] \quad \chi^2(\Theta) = [\boldsymbol{m}(\Theta) - \boldsymbol{d}]^{\mathrm{T}} C(\Theta)^{-1} [\boldsymbol{m}(\Theta) - \boldsymbol{d}]$$

- Fisher information matrix:

$$C_\Theta = \begin{pmatrix} \mathrm{Var}(\Theta_1) & \mathrm{Var}(\Theta_1, \Theta_2) \\ \mathrm{Var}(\Theta_1, \Theta_2) & \mathrm{Var}(\Theta_2) \end{pmatrix} \qquad F_{ij} = \left(\frac{\partial m(\Theta)}{\partial \Theta_i}\right)^{\mathrm{T}} C^{-1} \left(\frac{\partial m(\Theta)}{\partial \Theta_j}\right)$$

Cramer-Rao bound: $F^{-1} \leq C_\Theta$

How can we check that a matrix is invertible?

- Check if the determinate in is non-zero

$$\det(C) \neq 0 \rightarrow C^{-1} \propto 1/\det(C)$$

- Check if there are no vanishing eigenvalues

$$C\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \qquad \rightarrow D[\lambda_1 \ldots \lambda_n] = T[\boldsymbol{v}_1 \ldots \boldsymbol{v}_n]^{\mathrm{T}}\, C\, T[\boldsymbol{v}_1 \ldots \boldsymbol{v}_n]$$

$$\rightarrow \det(C) = \prod_i \lambda_i$$

How can we check that a covariance matrix is invertible?

- Check if there are no negative eigenvalues $\lambda_i \geq 0$ $\quad$ $C\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$

Why is that? Because the covariance matrix is, by definition, positive-semi definite. $\boldsymbol{x}^{\mathrm{T}} C \boldsymbol{x} \geq 0, \forall \boldsymbol{x} \in \mathbb{R}^n$

$$\boldsymbol{x}^{\mathrm{T}} C \boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}} \left[\frac{1}{n}\sum_i^n (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^{\mathrm{T}}\right] \boldsymbol{x} = \frac{1}{n}\sum_i^n \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^{\mathrm{T}} \boldsymbol{x} = \frac{1}{n}\sum_i^n \left[(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^{\mathrm{T}} \boldsymbol{x}\right]^2 \geq 0$$

So how this implies that $\lambda_i \geq 0$:

$$0 \leq \boldsymbol{v}^{\mathrm{T}} C \boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}} \lambda \boldsymbol{v} = \lambda |\boldsymbol{v}|^2 \rightarrow \lambda \geq 0$$

Invertible sample covariance matrix

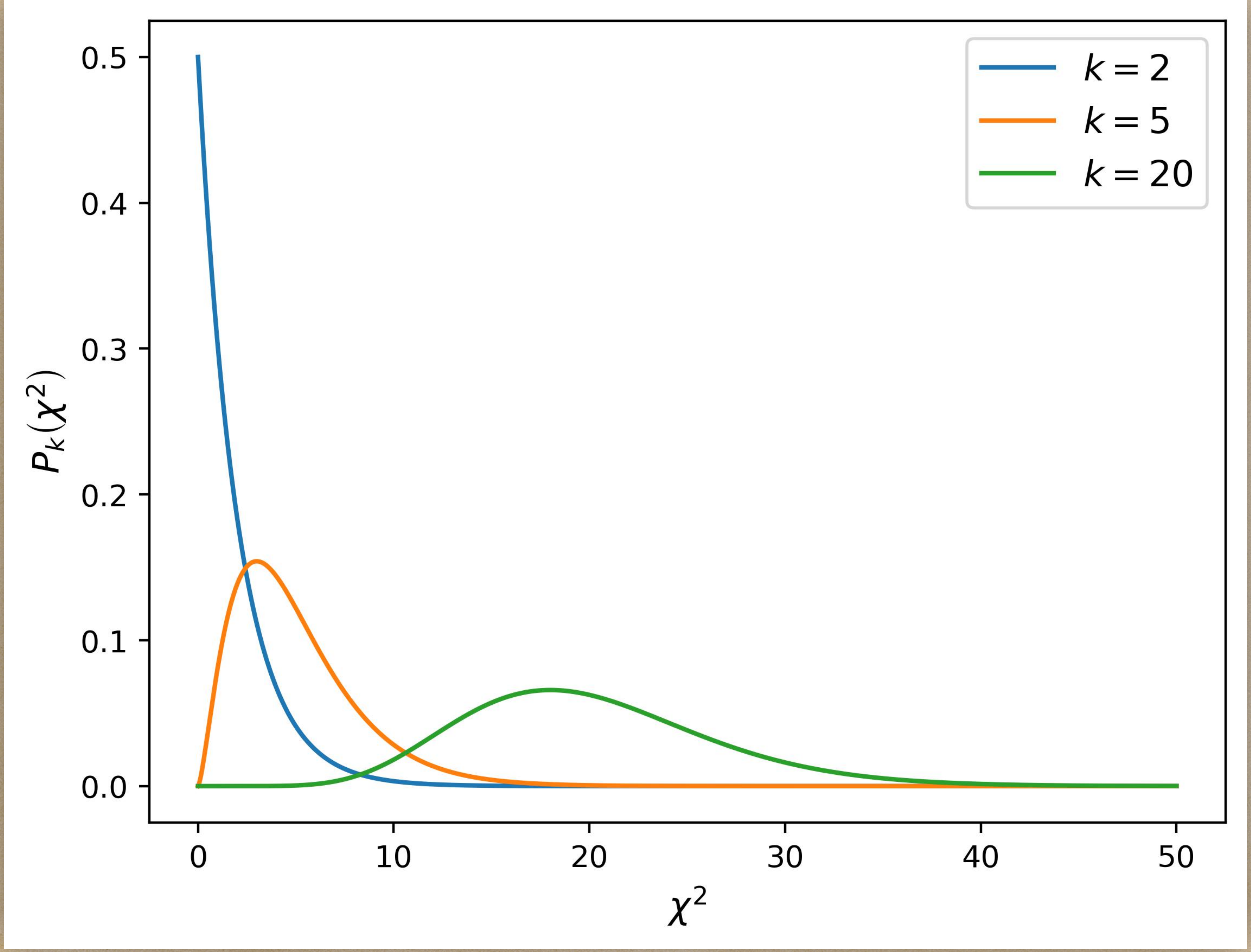
How to make sure that the sample covariance matrix has only positive eigenvalues?

m: Rank of covariance matrix

n: Number of realisations used to compute the covariance matrix

-> $n \geq m$

χ^2-distribution



$$E[\chi^2] = k$$

$$Var[\chi^2] = 2\,k$$

$$p(\chi^2 = x, k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \mathrm{e}^{-x^2/2}$$

For a goodness-of-fit

$$\chi^2/k \approx 1$$

Estimating the degrees of freedom

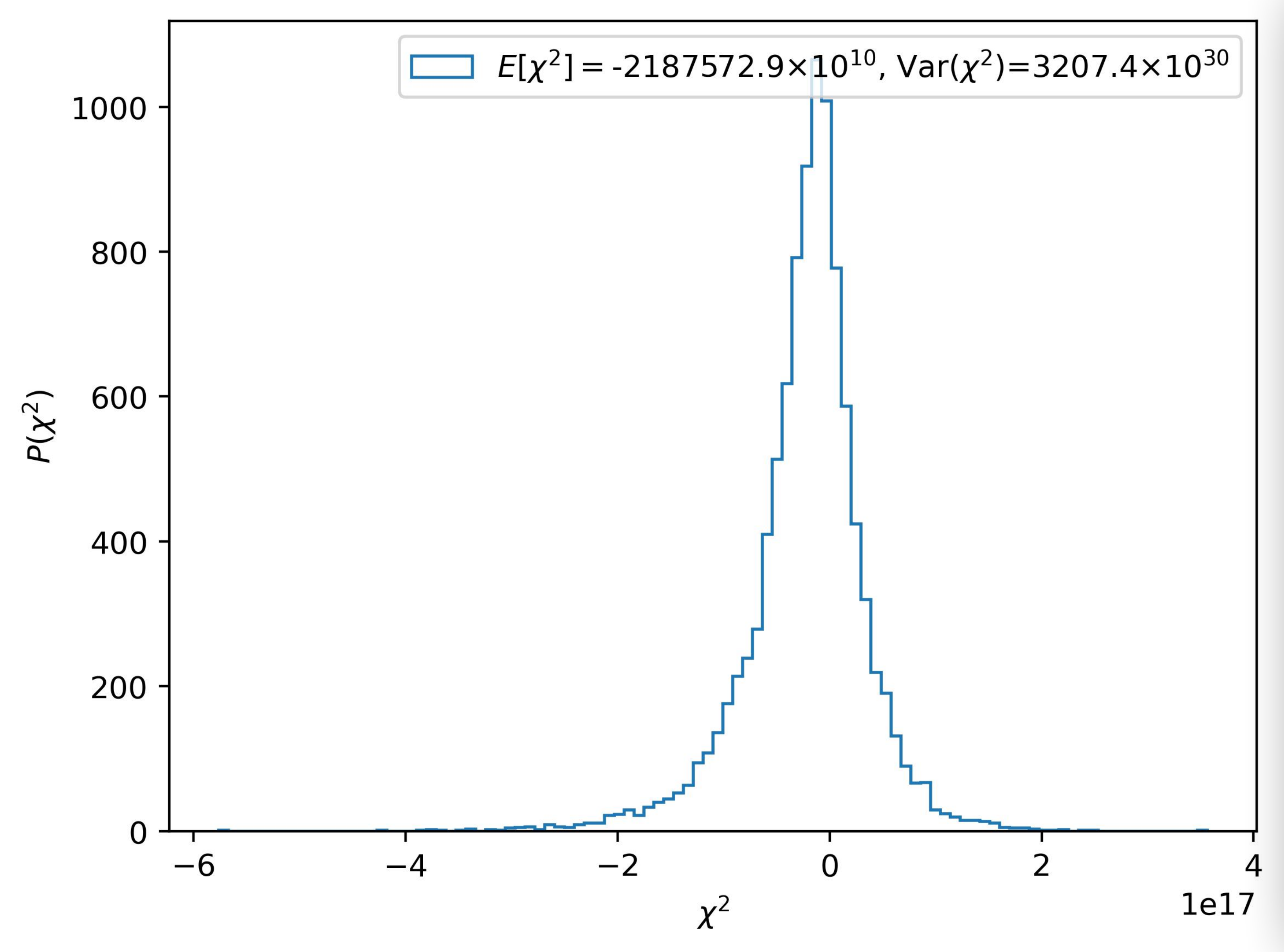
- If we do not fit the data then d.o.f.= dimension of data vector
- If you fit, a rough estimate is d.o.f.= dimension of data - number of parameters

A nice and easy test to find the "true" d.o.f if data is Gaussian distributed:

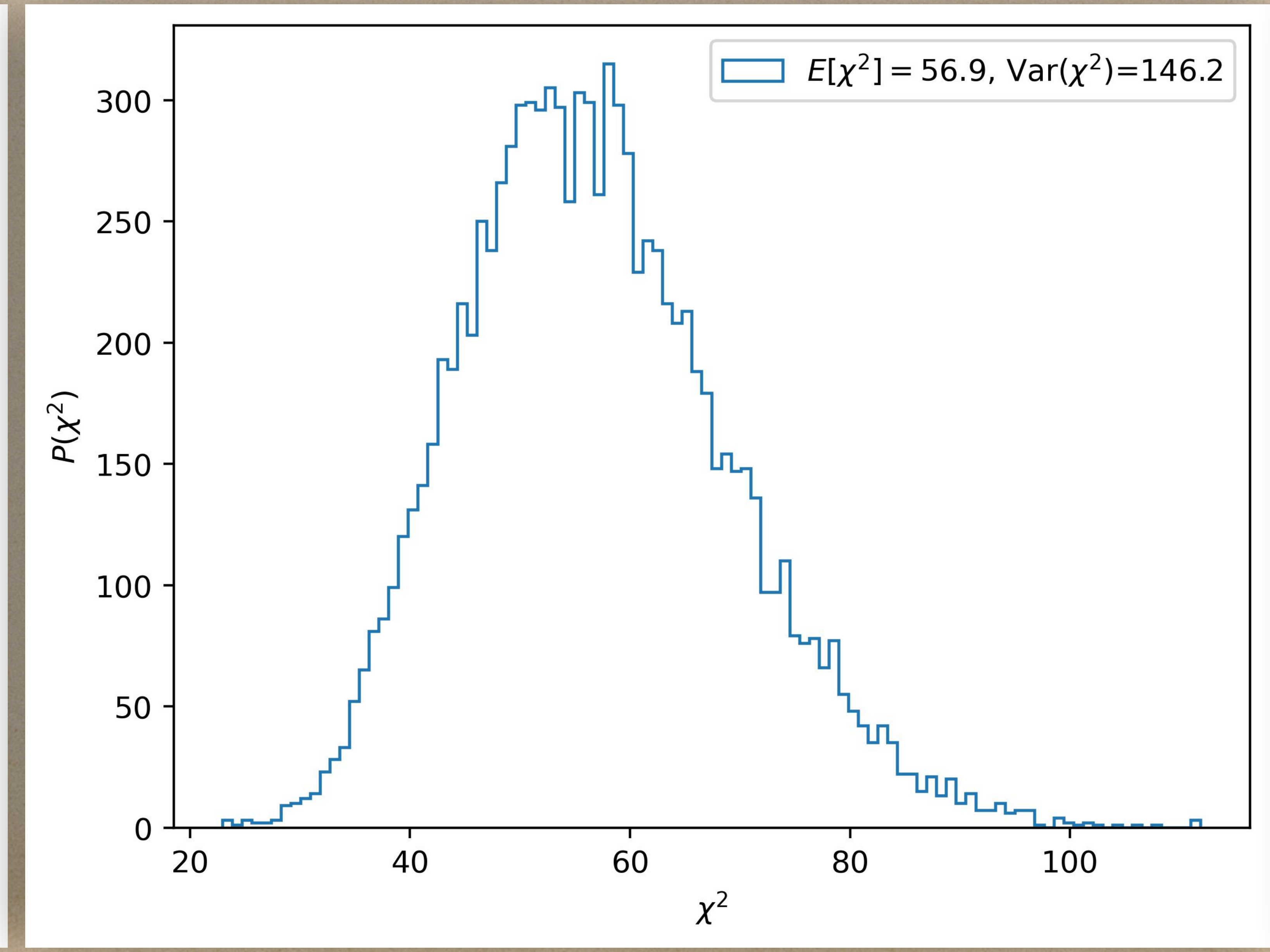
1. Draw n multivariate Gaussian random vectors using reference model and covariance matrix. -> np.random.multivariate(mean,cov,size=n)
2. Find the best χ^2 for all n random vectors
3. Measure mean χ^2 which is equal to the d.o.f

χ^2-distribution: Example 1

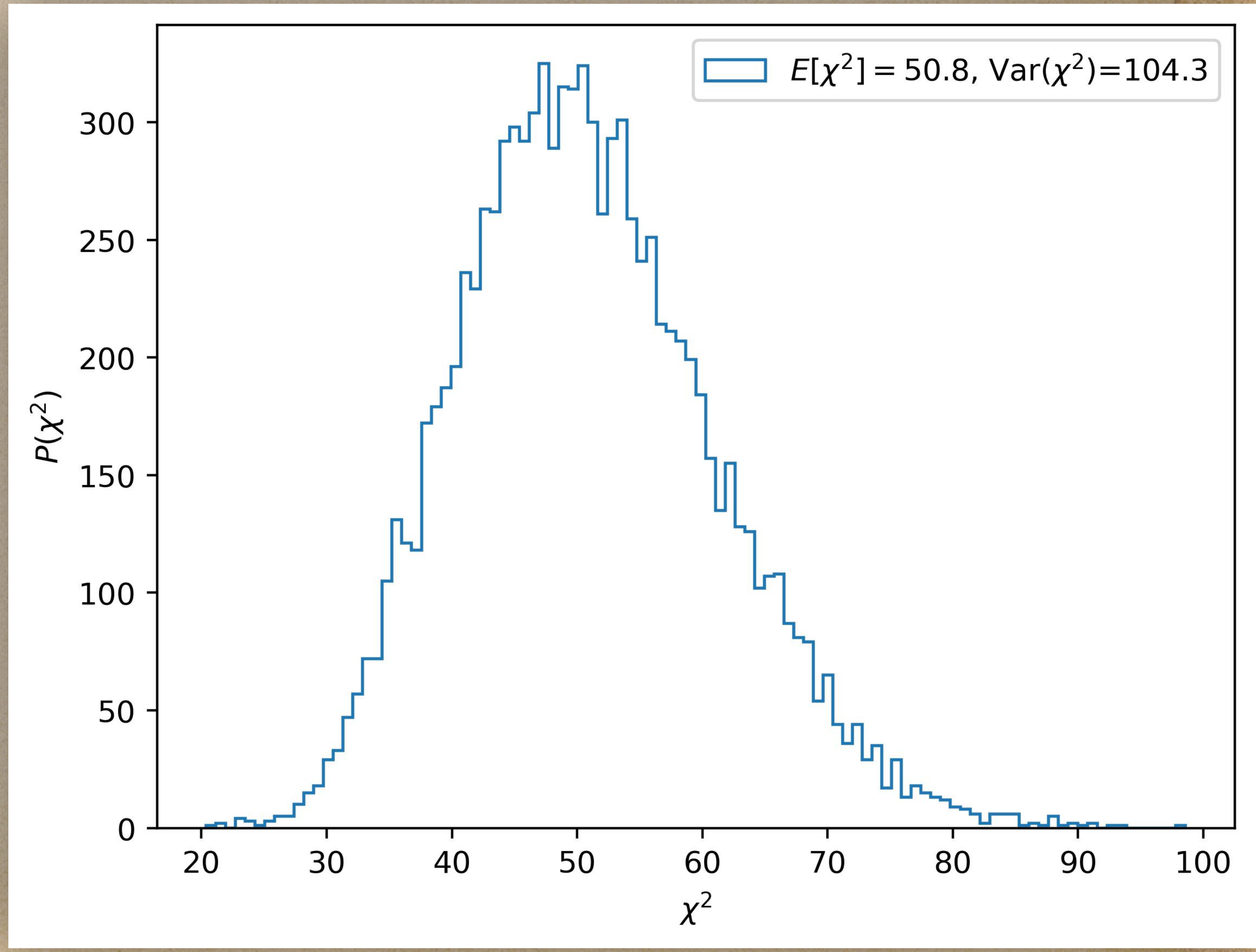
$m = 50, n = 40$



$m = 50, n = 400$



$m = 50, n = 4000$



Hartlap factor

Even with roughly 100 more realisations, the rank of the sample covariance matrix does not give the correct mean and variance.

What are we missing?

If measured from simulations, the covariance matrix itself is a random variable (Percival et al. 2020)

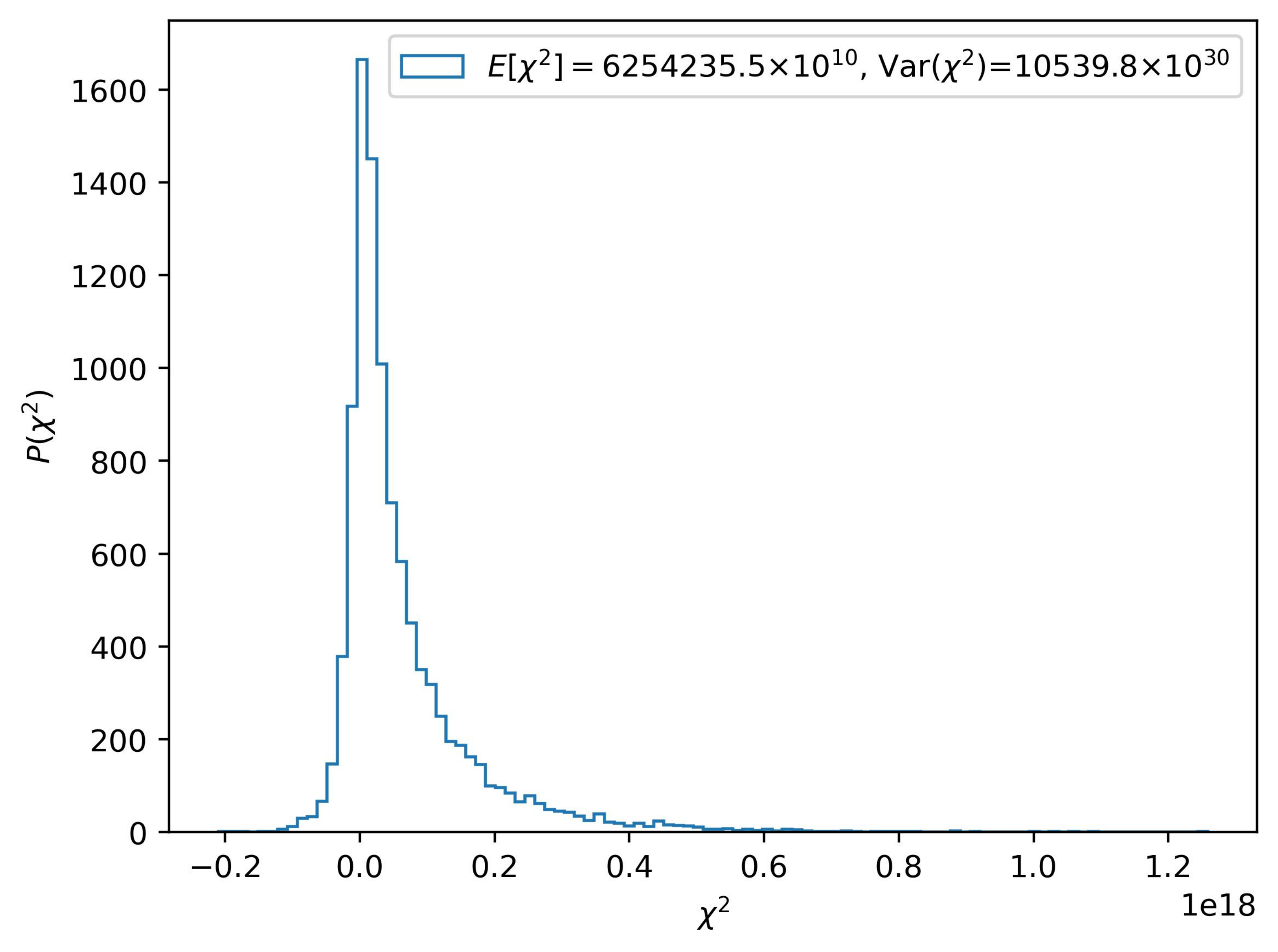
In order to debias the inverse covariance matrix we have to scale it with Hartlap factor (Kaufman 1967, Hartlap 2007)

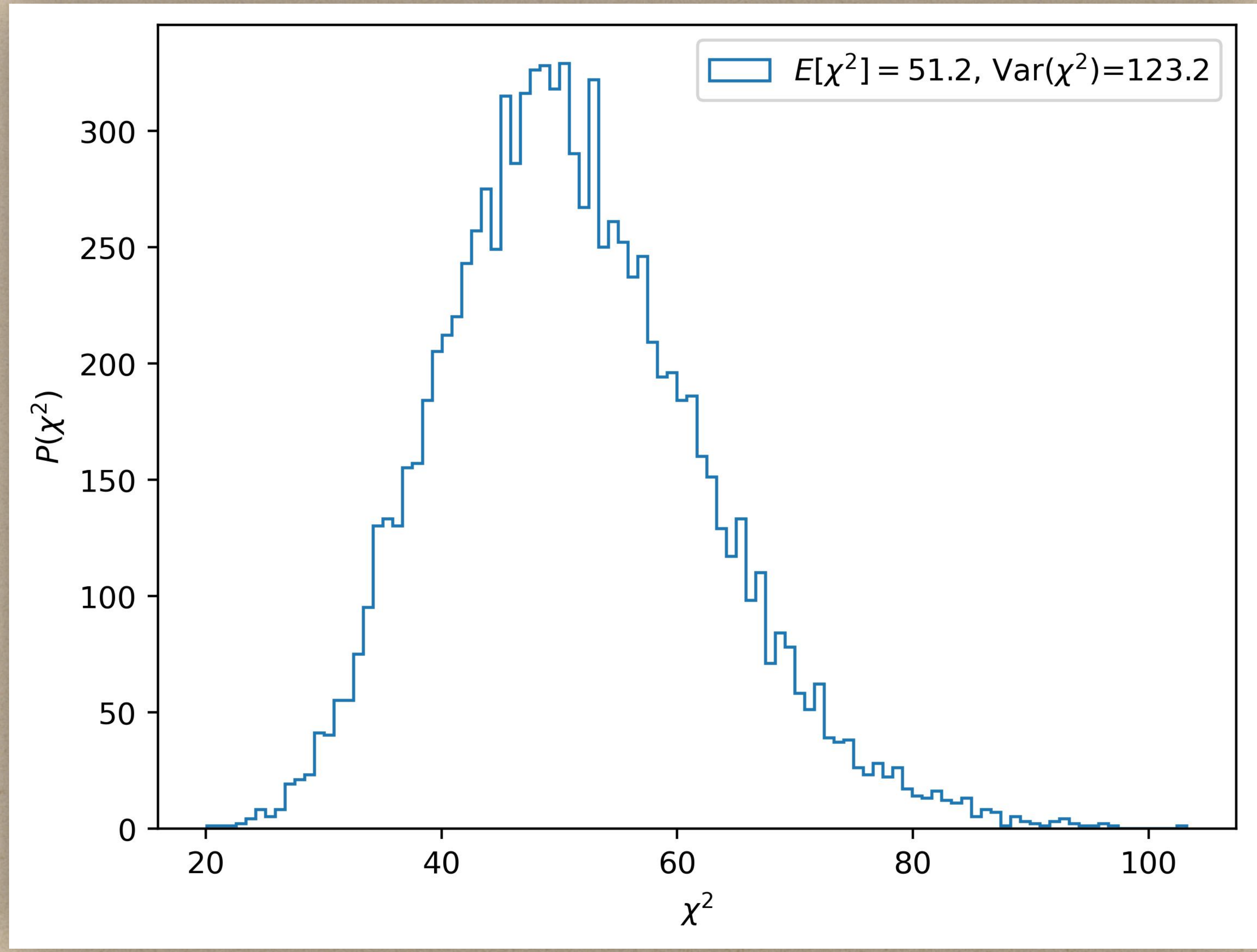
$$h = \frac{n-1}{n-m-2} \qquad \rightarrow C^{-1} = h\ \tilde{C}^{-1}$$

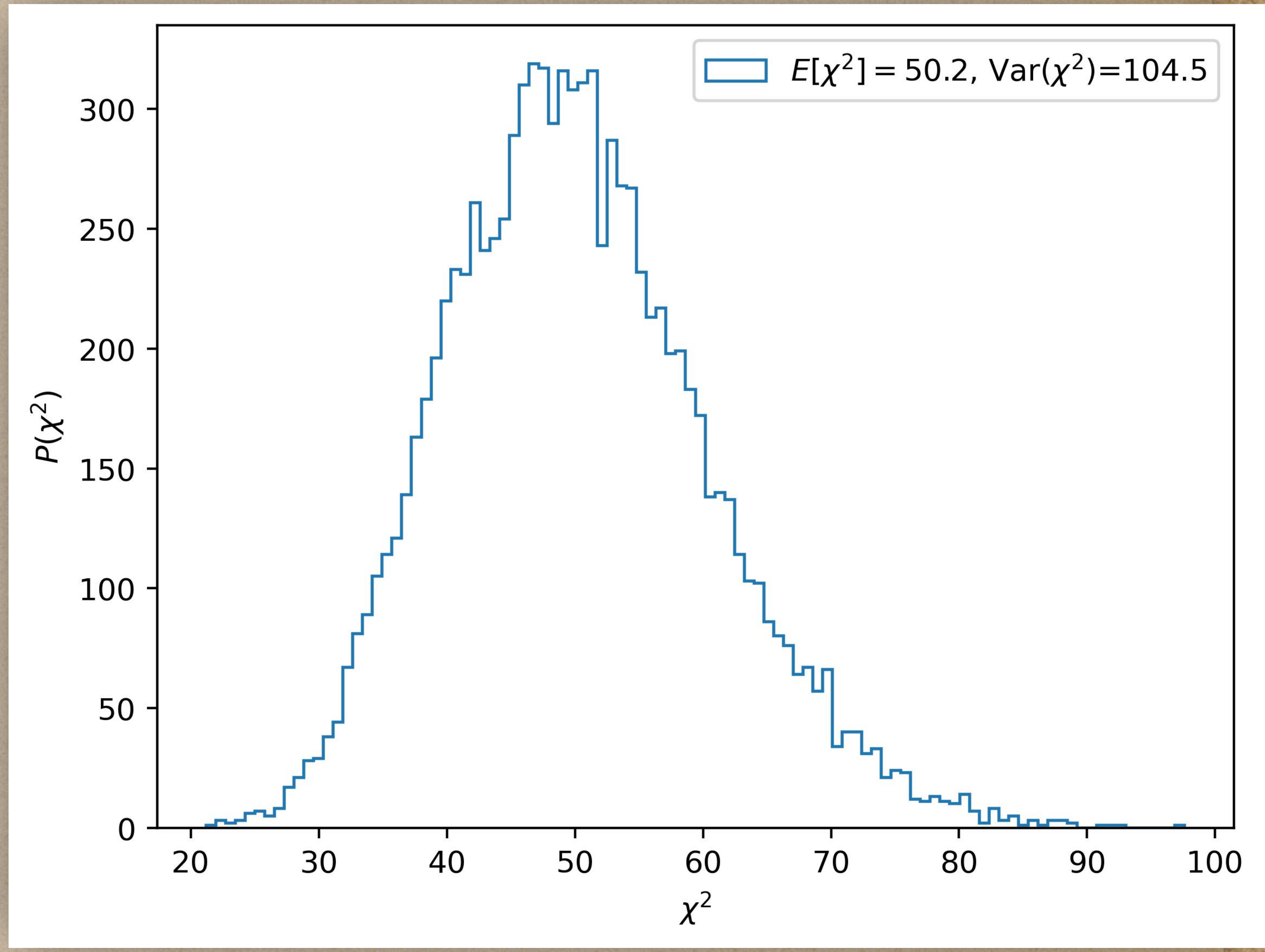
χ^2-distribution: Example 2

With Hartlap corruption

$m = 50,\ n = 40$ | $m = 50,\ n = 400$ | $m = 50,\ n = 4000$







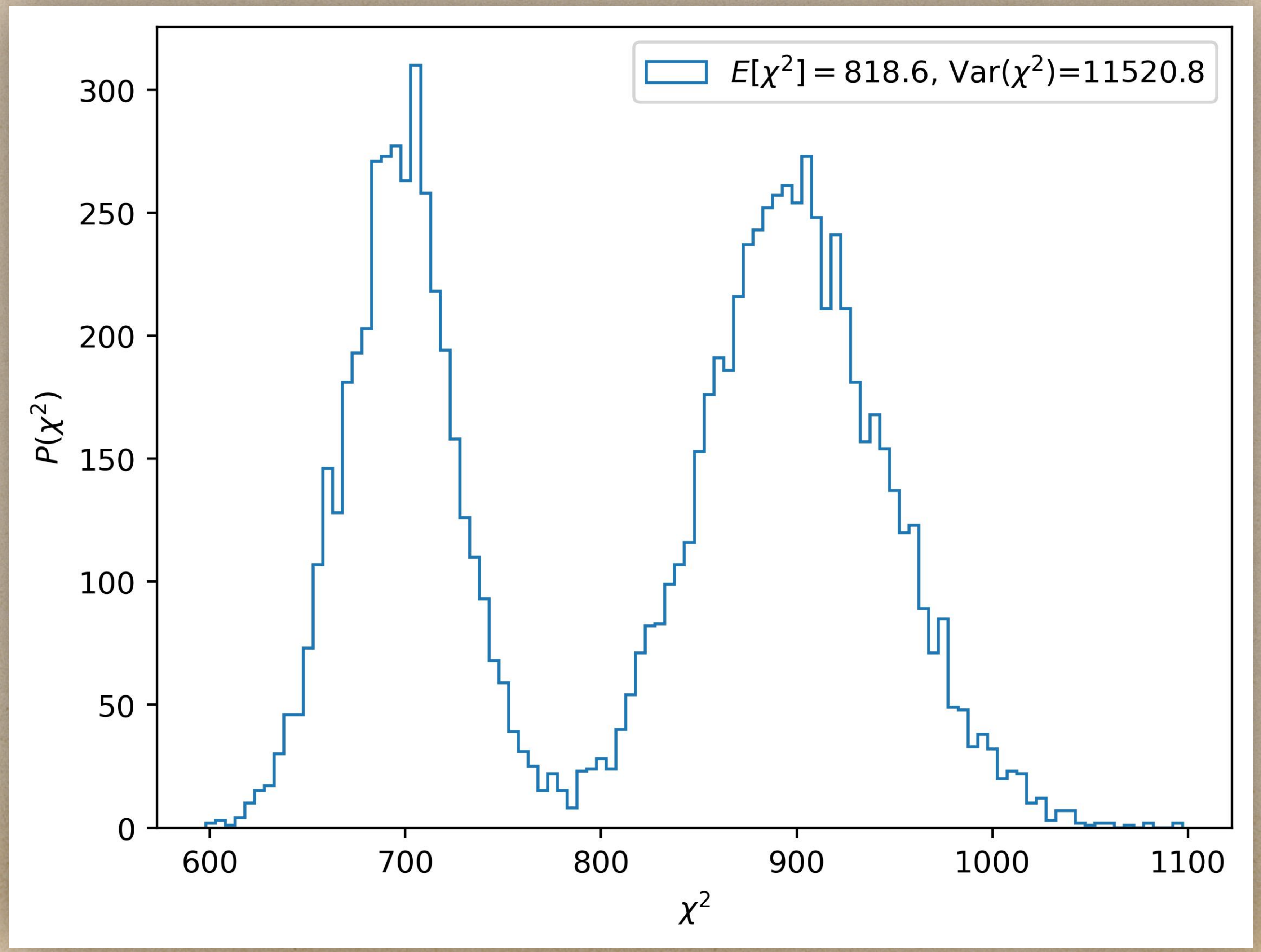
Note: Even with Hartlap correction, you need plenty of realisations that the variance converges.

$$n \gg m$$

χ^2-distribution: Example 3

Do not measure the χ^2-distribution on those data that are used for the covariance matrix -> wrong mean and variance

$$m = 900,\ n = 4000$$



Test for Gaussianity

Why do we need to check that each element of the summary statistic is drawn from a Gaussian distribution?

1. The covariance matrix is not sufficient to describe uncertainty data
2. Likelihood or Fisher analysis is only valid for Gaussian-distributed data

Use for instance the Shapiro-Wilk test:

Given test statistics W it computes p-value if data is drawn from a Gaussian distribution

$$W = \frac{\left(\sum_i^n a_i \, x_{(i)}\right)^2}{\sum_i^n (x_i - \bar{x})^2}$$

Scipy.stats.shapiro

Singular value decomposition

Suppose you have a covariance matrix with negative eigenvalues. What do you do?

1. Remove the negative eigenvalues
2. Transform covariance matrix and data/model vectors

Singular value decomposition

As we know from before we can diagonalise a covariance matrix

$$D[\lambda_1 \dots \lambda_n] = T[\boldsymbol{v}_1 \dots \boldsymbol{v}_n]^{\mathrm{T}}\, C\, T[\boldsymbol{v}_1 \dots \boldsymbol{v}_n]$$

But we do it slightly different:

$$C = U[\pm \boldsymbol{v}_1 \dots \pm \boldsymbol{v}_n]\, S[\,|\lambda_1| \dots |\lambda_n|\,]\, V^H[\pm \boldsymbol{v}_1^{\mathrm{T}} \dots \pm \boldsymbol{v}_n^{\mathrm{T}}]$$

Singular value decomposition

Remove all negative eigenvalues and corresponding eigenvectors:

$$\tilde{C} = \tilde{U}\,\tilde{S}\,\tilde{V}^H$$

Using that $(AB)^{-1} = B^{-1}A^{-1}$ and that $\tilde{U}^{-1} = \tilde{U}^{\mathrm{T}}, \left(\tilde{V}^h\right)^{-1} = \left(\tilde{V}^h\right)^{\mathrm{T}}$ we find that

$$\chi^2 = [\boldsymbol{m} - \boldsymbol{d}]^{\mathrm{T}}\tilde{C}^{-1}[\boldsymbol{m} - \boldsymbol{d}] = [\boldsymbol{m} - \boldsymbol{d}]^{\mathrm{T}}\left(\tilde{V}^H\right)^{\mathrm{T}}\,\tilde{S}^{-1}\,\tilde{U}^{\mathrm{T}}[\boldsymbol{m} - \boldsymbol{d}]$$

Questions?