STATISTICS:
What have I learned from my data set? And could I have been more clever in producing/analyzing it?

GK-Series by Ln(a) Sellentin

By the end of this lecture series, we'll have...

- Seen the difference between Frequentist and Bayesian Statistics
- Assessed how many parameters a data set supports
- Contrasted model selection with “improving parameter constraints”
- Seen how to approximate likelihoods/posteriors, and worried about Gaussianity

Reminder:
When we were young...

Anfängerpraktikum...

Task: Measure the current x-times.... what is the “true” current?

The typical procedure was:

- Measure what you want to know, e.g. the current I.
- Measure it often.
- Calculate a mean and a standard deviation.
- Measure U, often, calculate mean and std.dev.
- Use U=RI to get R, using Gaussian error propagation.

This was low-level (frequentist) statistics.

Low-level because...

- What you measured was already the variable you wanted to know.
- Your “estimator” is your measurement point.
- Usually: have to estimate a **parameter**
- need **a set of multiple** data points $x_i(p,q,r)$

  to estimate $\hat{p}, \hat{q}, \hat{r}$
- → need multiple sets of data points to know how the estimates scatter
- → don't **confuse** the number and variance of data points with the number and variance of estimates of your parameters

In the practical courses, this issue did not appear (due to the given tasks).

Once, we had to conduct a fit:

- Minimize χ^2
- Which is a consequence of maximizing

$$exp(-\frac{1}{2}\chi^2)$$

(the likelihood)

→ What you are used to from the studies is Frequentist statistics.

The common aim of Frequentists and Bayesians

- Both assume there exists a “true” physical quantitiy, which we want to know.
- Both have to infere it from **imperfect** data
- Imperfections: data set is finite, noisy, potentially biased
- Both of these statistical approaches try to split imperfections from the true underlying physics

For perfect data, the two approaches would lead to identical results.

An (overly) simplistic view:

- Frequentism is good if uncertainties in the measurements dominate (which can be brought down by keeping on measuring, 1/sqrt(N)...)

- Bayesianism is good when uncertainties in the data-interpretation dominate

LECTURE 1

- Some cross-calibrations between particle physics and cosmology

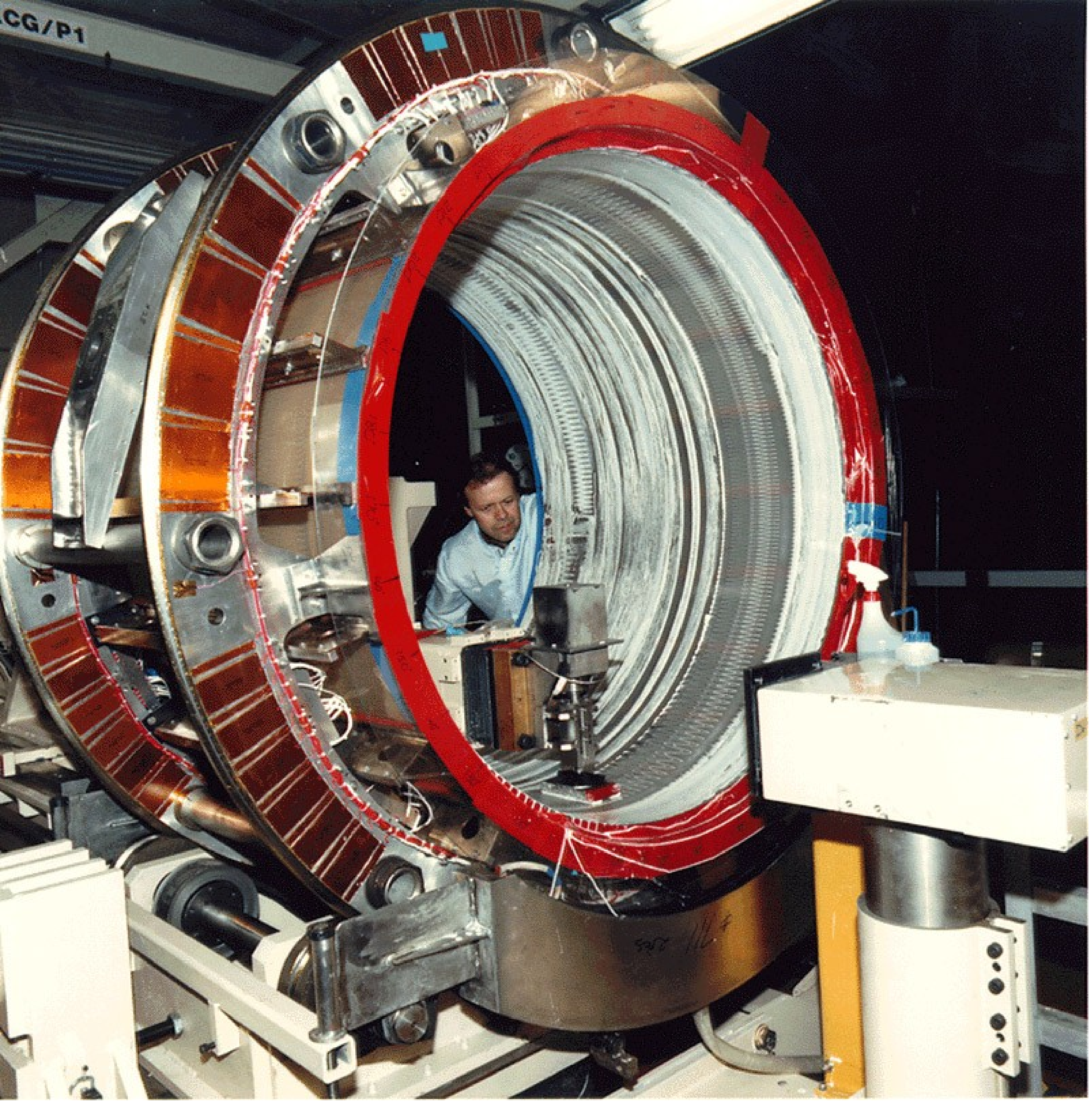
- Especially: Why Bayesian statistics

Why Baysian statistics?

A typical lifecircle in astronomy...

Let's build an X-ray satellite

Building: Mirror shells and coating with iridium



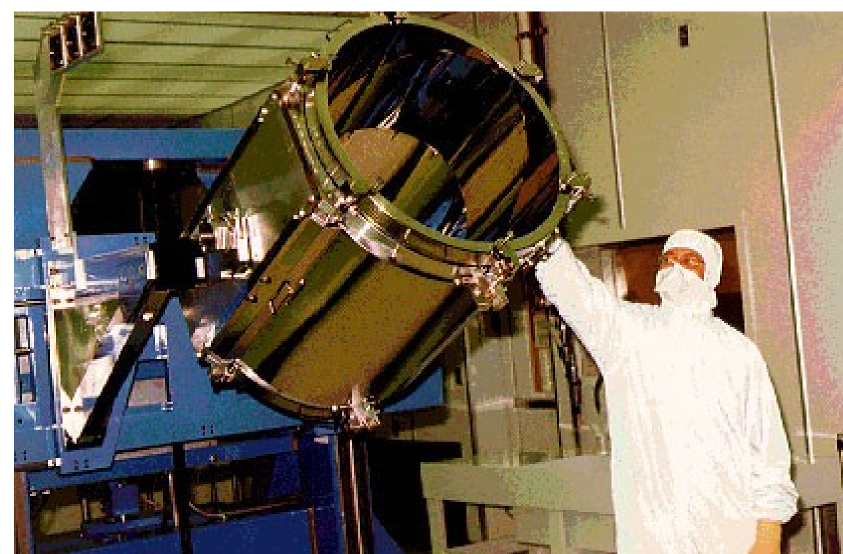


Image Credit: NASA/Chandra/Raytheon Company

Assembling

Image Credit: NASA/Chandra/Eastman-Kodak

Testing: traditional Frequentist's playground

Image Credit: Nasa/Chandra

Testing under mimicked space conditions

Image Credit: NASA/Chandra/NGST

Image Credit:NASA

Image Credit: NASA

On Orbit Calibration

Tests

Tests

Tests!

Tests

Tests

Tests

Tests

Tests

Tests

First Light Image: Public Release

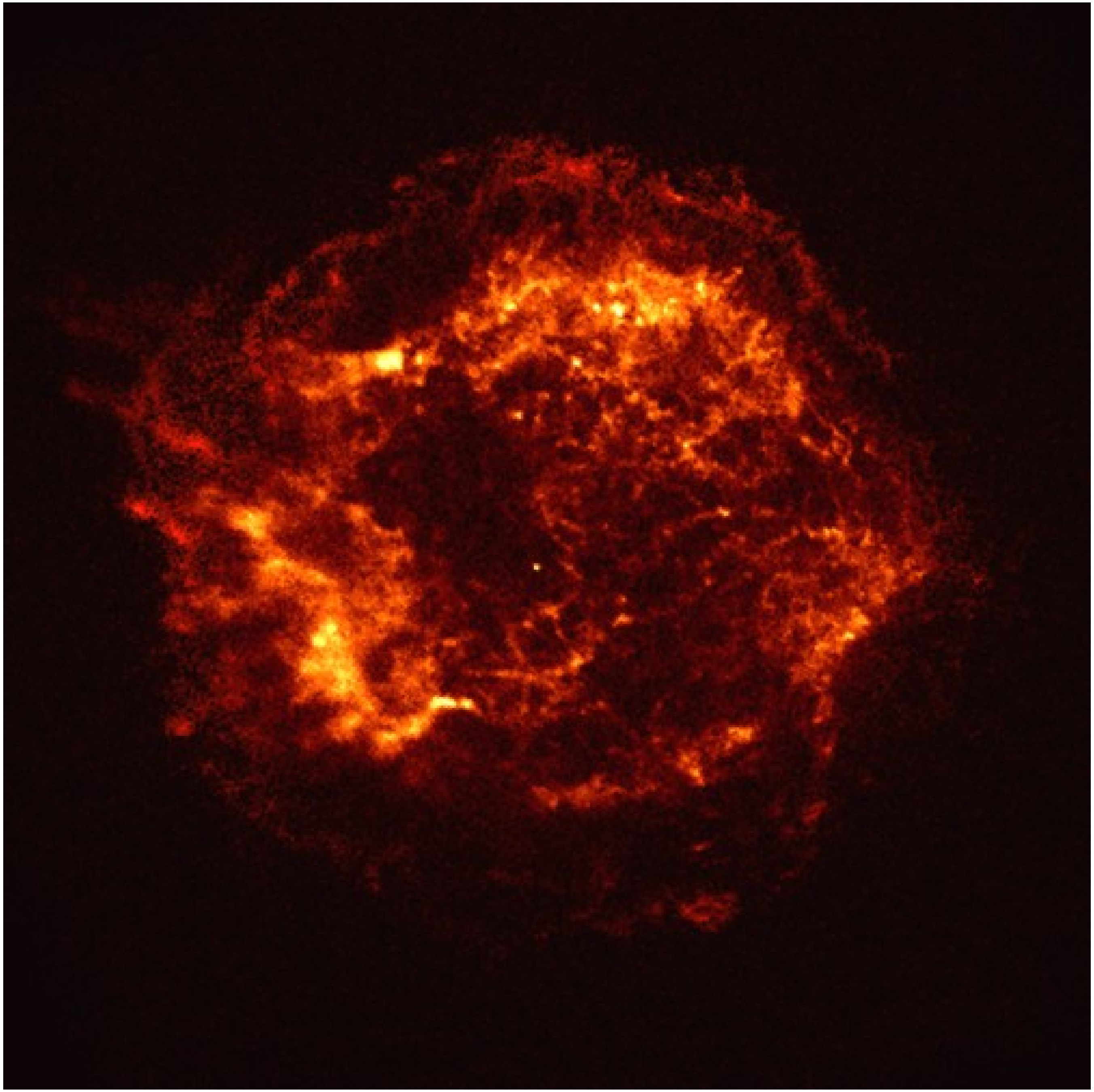


Image Credit: NASA/Chandra/SAO

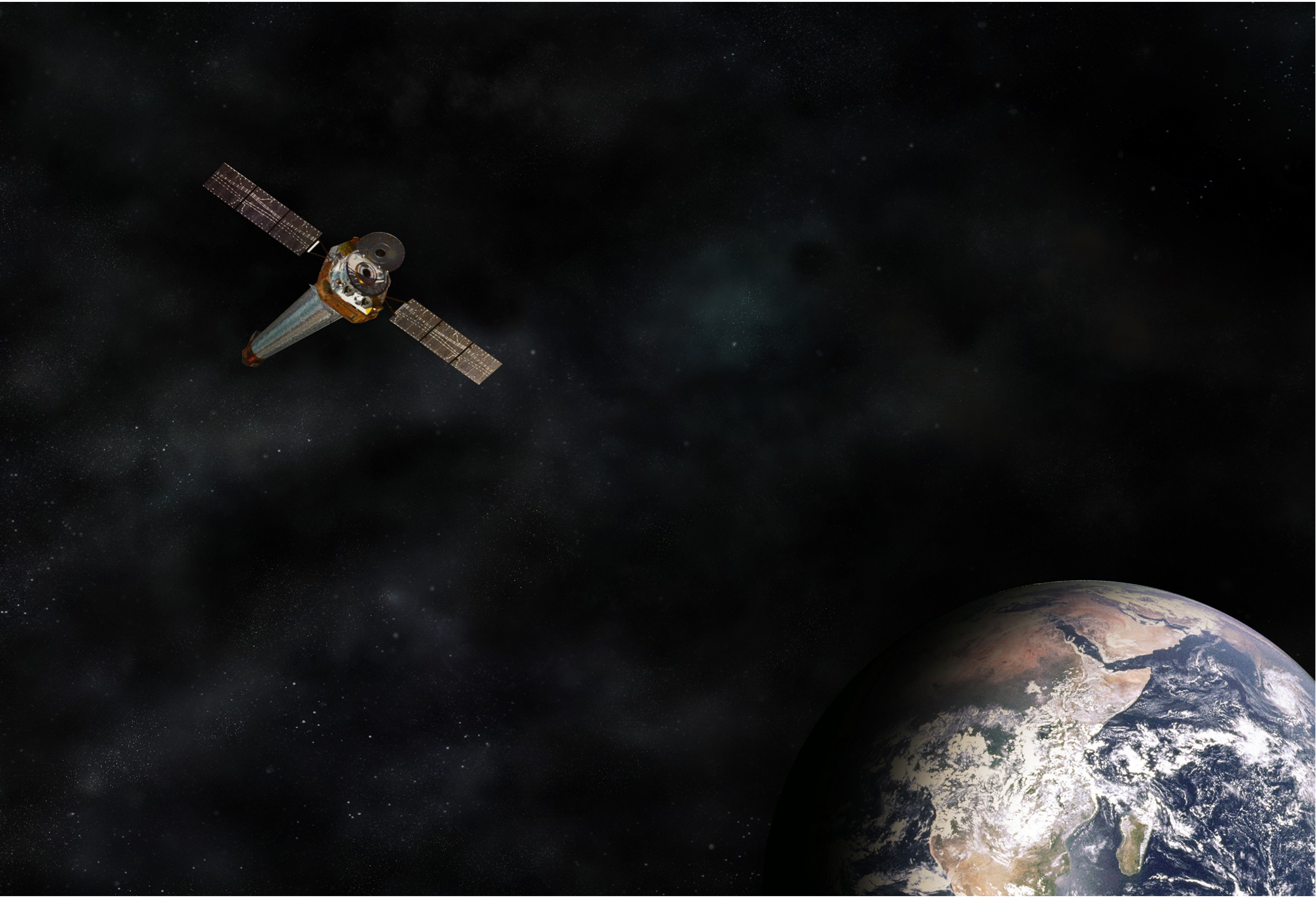


Illustration:NASA/CXC/D.Berry & M.Weiss