PHYS114 Reference

Krish A. Patel

February 20, 2026

1 Data Files, Software, and Figures

1.1 Data Files

- **Binary**:

  - Advantages: compact file size.
  - Disadvantages: need encoders (otherwise difficult to read).
  - Microsoft 365 files (.xlsx, .docx, etc.), which are proprietary but have been open-sourced for interoperability with, for example, Apple and Google's suites.
  - **.fits**: Well-known, well-defined, and used by the astronomy community.
  - **.cdf**: Common Data Format, standardized encoding.
  - **.sav**: IDL's proprietary format.

- **ASCII**:

  - Advantages: built-in encoding table and, thus, easy to read.
  - Disadvantages: large file size.
  - .txt, .csv, etc.

1.2 Software

- A wide array of scientific software exists such as MATLAB and IDL.

- However, these tend to require expensive licenses.

- In contrast, Python is free and open-source.

- However, Python comes with no warranty, meaning that if something goes wrong, the user is responsible.

1.3 Figures

A professional-quality figure

- is grayscale, relying on various marker sizes and shapes (e.g. circle, diamond, etc.) as well as line styles (solid, dashed, etc.) to differentiate different series.

- contains a title, axes labels with units, and a legend as necessary.

- uses space on the plot effectively.

- has proper axis scaling.

- is high resolution (usually 600 DPI or more).

2 Introduction to Research

Some information is specific to the New Jersey Institute of Technology (NJIT).

- A typical external federal grant is $130,000 for 3 years.

- Professors must bring in grants to be tenured.

- Professors also have a "2 + 2" teaching requirement: 2 courses in the fall, and 2 courses in the spring.

- If a professor supports a research assistant (RA), covering their **stipend**, **tuition**, and **fringe**, they go down to 1 + 2.

- Supporting another RA reduces this to 1 + 1.

- In contrast, teaching assistants (TAs) are paid for by the university, not the professor.

- A researcher or RA typically cannot have another job if they are already being funded by a grant for maximal dedication.

- Usually, grant money goes to the university, not the professor directly.

- PhDs in the United States usually take 6 years because new knowledge must be created to attain it.

- A typical physics graduate student coursework:
  - Fall: Electricity and Magnetism I, Classical Mechanics, Mathematical Methods
  - Spring: Electricity and Magnetism II, Quantum Mechanics, Statistical Mechanics / Thermodynamics

- Qualifying exams in the summer (June for NJIT) determine whether a student remains in the graduate program.

- Qualifying exams are difficult (typically 50% pass rate for NJIT).

- There is one last redemption exam in January for those who fail.

- In the 1980s, program officers would evaluate grant proposals, but from 1990s onward, panels of qualified individuals perform this task.

2.1 Federal Funding Agencies

- Top agencies that fund grants:

    - National Science Foundation (NSF)
    - Department of Energy (DOE)
    - National Aeronautics and Space Administration (NASA)
    - National Institutes of Health (NIH)
    - Department of Defense (DoD)

- All except NSF require **closure**, which means meaningful results must be found, and the research question must be answered.

- Some agencies require an American citizenship for all researchers.

2.2 Grant Format

- A small part of the grant ($\sim 10\%$) is:

    - Project summary (1 page)
    - Project description (15+ pages)
    - Export control, etc.

- But the majority of the grant ($\sim 90\%$) is paperwork

2.3 Fringe and Overhead

- **Fringe** is the cost of employing someone, which includes associated costs such as insurance, retirement, etc.

- **Overhead** is the money used to support the research enterprise of an institution, which includes the cost of utilities, supplies, subscriptions, etc.

- The fringe at NJIT went from 43% to 86% in 5 years.

- The cost of research has skyrocketed due to rising insurance premiums and unionization.

3 Forward Model and Photometer

3.1 Forward Model

A mathematical or quantitative description of the problem you're solving. One can use synthetic data and example numbers to test.
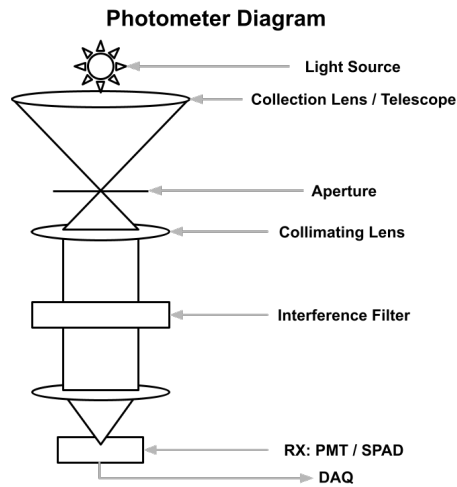
3.2 Photometer

Figure 1: A diagram of a photometer.

- Measures the number of photons from a source.

- IF stands for interference filter.

- RX stands for receiver.

- TX stands for transmitter.

- The receiver can either be a **photomultiplier tube (PMT)** or a **single-photon avalanche diode (SPAD)**.

- **Charge-coupled devices (CCDs)** can be found in smartphones, but scientific-grade CCDs are much more sensitive and expensive.

- **Photon count formula**:

$$N = N_{\text{signal}} + N_{\text{noise}}$$

where

- N is the total number of photons.
- N_{signal} is the number of photons from the desired source.
- N_{noise} is the number of photons from other (undesirable) sources.

- **Noise photon count formula**:

$$N_{\text{noise}} = N_C \Delta t$$

where

- N_C [counts/s] is the noise count rate.
- $\Delta t\,[s]$ is the **integration time**, the time interval at which the detector collects photons.

- **Signal photon count formula**:

$$N_{\text{signal}} = B A \Delta t\, \Omega T_A \nu_{RX}$$

$$\Omega = \frac{A}{r^2} = 2\pi(1 - \cos\alpha)$$

where

- $B\,[\text{photons}/(\text{s} \cdot \text{m}^2 \cdot \text{str})]$ is a constant.
- $A\,[\text{m}^2]$ for N_{signal} is the area of the telescope.
- $\Delta t\,[s]$ is the integration time.
- $\Omega\,[\text{str}]$ is the solid angle.
- T_A is the atmospheric transmission.
- ν_{RX} is the receiver efficiency.
- A for N_{noise} is the area of the detector.
- r is the focal length of the telescope.
- α is the half-angle of the cone of light collected by the telescope, also known as the field of view.

Note: $A\Omega$ is a conversed quality called the **etendue**, and it will be explored later.

4 Distributions and Uncertainty

- For **nonstationary** PDFs, the moments are in terms of time.

- For **wide-sense stationary (WSS)** PDFs, only the **first two moments** are stationary, which represents most real-life phenomena.

- For **stationary** PDFs, all the moments are constant and *not* in terms of time.

- **Moment generating function (continuous)**:

$$n^{\text{th}} \text{ moment} = \int_{-\infty}^{\infty} x^n \, \text{PDF}(x) \, \partial x$$

- **Moment generating function (discrete)**:

$$n^{\text{th}} \text{ moment} = \sum_{i=1}^{\infty} x_i^n \, \text{PDF}(x_i)$$

- This class only focuses on the Gaussian, Binomial, and Poisson distributions. The rest are extra information.

- Remember that probabilities multiply. The probability of reproducing an entire dataset is

$$\prod_{i=1}^{N} \text{PDF}(x_i) = p_1 p_2 \ldots p_N$$

- General process of data collection and processing:

  1. Determine PDF.
  2. Create estimators for moments from data.
  3. Determine if estimators are biased or unbiased using **expectation values**.
  4. Determine the chi-squared goodness of fit.

- From strongest to weakest: independent, uncorrelated, and no covariance.

4.1 Gaussian Distribution

Also known as the normal distribution. Represents humans and many natural phenomena.

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- μ and σ are parameters.
- 1st moment (mean): μ
- 2nd moment (variance): σ^2
- Estimator for 1st moment:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Estimator for 2$^{\text{nd}}$ moment:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

- 3rd moment (skewness)

- 4th moment (kurtosis)

- Uncertainty for 1$^{\text{st}}$ moment:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

This signals diminishing returns for increasing N.

- However, the above is assuming WSS. In a case without WSS:

$$\bar{x} = \frac{\sum_{i=1}^{N} \left(\frac{x_i}{\sigma_i^2} \right)}{\sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} \right)}$$

$$\sigma_{\bar{x}}^2 = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

4.2   Binomial Distribution

$$\text{PDF}(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

- Trials must be **independent**.

- For p large and n large, similar to the **Gaussian distribution**.

- For p low and n large, similar to the **Poisson distribution**.

- n and p are parameters.

- 1$^{\text{st}}$ moment: np

- 2$^{\text{nd}}$ moment: $np(1-p)$

- 3$^{\text{rd}}$ moment: $\frac{(1-p)-p}{\sqrt{np(1-p)}}$

- Estimator for 1st moment:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Estimator for 2nd moment:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

4.3 Poisson Distribution

Represents scenarios with counting and waiting for events.

$$\text{PDF}(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- λ is the parameter.
- 1st moment: λ
- 2nd moment: λ
- 3rd moment: $\frac{1}{\sqrt{\lambda}}$
- Estimator for 1st moment:

$$\bar{x} = \sum_{i=1}^{N} x_i$$

- Estimator for 2nd moment:

$$s^2 = \sum_{i=1}^{N} x_i$$

- Percent error:

$$\frac{1}{\sqrt{\bar{x}}} \times 100\%$$

4.4 Laplace Distribution

$$f(x; \mu, b) = \frac{1}{2b} \exp\left[-\frac{x-\mu}{b} \right]$$

- μ and b are parameters.
- μ is a location parameter.
- $b > 0$ is called "diversity."
- 1st moment: μ
- 2nd moment: $2b^2$
- 3rd moment: $\emptyset$

4.5 Gamma Distribution

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where Γ is the gamma function.

- α and β are parameters.

- α is the shape parameter.

- $\frac{1}{\beta}$ is the scale parameter.

- 1st moment: $\frac{\alpha}{\beta}$

- 2nd moment: $\frac{\alpha}{\beta^2}$

- 3rd moment: $\frac{2}{\sqrt{\alpha}}$

4.6 Assumptions

We often assume that

1. Stationary or WSS PDF: $\sigma_i \approx \sigma$

2. Each measurement is independent and identically distributed (IID).

3. **Method of Maximum Liklihood (MML)**

   - Also known as Maximum Likelihood Method (MLM).
   - Maximum probability of reproducing the data.
   - An alternate method is **Maximum Entropy Method (MEM)** that seeks to add as much disorder as possible to the datapoints, but this class only focuses on MML.

4. An implicit assumption is the zero-mean process. We assume or subtract off the mean to make calculations much easier.

4.7 Expectation Values

$$\langle x_i \rangle = E[x_i] = \mu \quad \text{(typically)}$$

Examples:

- We expect the $\bar{x}$ estimator to yield μ: $\langle \bar{x} \rangle = \mu$

- And similarly for s^2: $\langle s^2 \rangle = \sigma^2$

A rule for expectation values is that you ignore constants:

$$\langle c \rangle = c$$

You can also ignore summations:

$$\langle \sum_{i=1}^{N} x_i \rangle = \sum_{i=1}^{N} \langle x_i \rangle$$

4.8 Uncertainty Propagation

- Uncertainty is represented by σ. For example, the uncertainty in $\bar{x}$ is denoted as $\sigma_{\bar{x}}$.

- Uncertainty can either propagate in a **worst case** or **even-steven** (informally named by our professor) scenario.

- Basics for worst case propagation: Suppose there are height measurements h_1 and h_2 with uncertainties σ_{h_1} and σ_{h_2}. Adding the two heights together also adds their uncertainties: $(h_1 \pm \sigma_{h_1}) + (h_2 \pm \sigma_{h_2}) = (h_1 + h_2) \pm (\sigma_{h_1} + \sigma_{h_2})$. This is the simplest example, but this can get much more complex, especially for operations like multiplication.

- **Even-steven formula** (Taylor series approximation):

$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v} \right)^2 + \cdots + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u} \right) \left(\frac{\partial x}{\partial v} \right) + \ldots$$

$$\sigma_{uv}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \bar{u})(v_i - \bar{v})$$

  – σ_{uv}^2 is the covariance, which is 0 when both u and v are **independent**.

4.9 Chi-Squared Goodness of Fit

Given a histogram:

$$\chi^2 \equiv \sum_{j=1}^{n} \frac{(h(x_j) - NP(x_j))^2}{\sigma_j(h)^2} = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\sigma_i^2}$$

where

- n is the number of bins.

- $h(x_j)$ is the value of the j-th bin.

- N is the total number of measurements.

- $P(x_j)$ is the probability of the j-th bin.

- $\sigma_j(h)^2$ is the variance in the j-th bin.

- χ^2 is a measure of the goodness of fit of a model to the data.

- A smaller χ^2 indicates a better fit. Ideal value is 0, but in practice it is around n because the numerator and denominator of the summation are around the same order of magnitude.

- **reduced chi-squared**:
$$\chi^2_\nu = \frac{\chi^2}{\nu}$$
$$\nu = n - n_c$$

  - n_c is the number of constraints. Usually 1.
  - ν is the number of degrees of freedom. Usually $n-1$ with no outliers.