1 Introduction

Why do we care about statistic and probability in physics?

- Experimental uncertainty: how confident are you about your measurements? Fitting data to models. Is your experimental data consistent with your model?
- Physics is statistical: statistical mechanics (thermodynamics) and quantum mechanics (statistical ensembles).

Goal: a broad exposure to statistical terms and concepts that pop up in research and a "from first principles" treatment of these concepts.

2 Events, Sample Spaces, RVs

- EVENT: a possible result of a measurement or experiment. Probability describes the statistics of an event happening.
- SAMPLE SPACE: the set of all events. For example, for a coin toss, the event space E is given by $E = \{H, T\}$.
- RANDOM VARIABLE: a variable that represents the numerical result of a random process. *A RV takes on different values by chance.* Each value the RV can take represents an event in the sample space. For example: if X is a RV representing the result of a dice roll, then after rolling a 5, X has value $X = 5$. X has a 1/6 of taking any value between 1–6.
- Function of random variables: suppose X, Y are random variables. We can define another random variable as a function of X, Y,

  $$Z = f(X, Y) \tag{2.1}$$

  NOTE: the probability distribution of functions of random variables are not trivial. For example, let X, Y be independent RVs for a coin toss (0 tails, 1 heads) and $Z = X + Y$.

  $$P(Z = 0) = P(X = 0 \cap Y = 0) = 1/4 \tag{2.2}$$
  $$P(Z = 1) = P(X = 0 \cap Y = 1) + P(X = 1 \cap Y = 0) = 1/2 \tag{2.3}$$
  $$P(Z = 2) = P(X = 1 \cap Y = 1) = 1/4 \tag{2.4}$$

  In general,

  $$P(Z = n) = \sum_{x,y:f(X=x,Y=y)=n} P(X = x \cap Y = y) \tag{2.5}$$

3 Probability

- For equally likely outcomes, the probability of event A is,

  $$P(A) = \frac{\#A}{\#\text{total}} \tag{3.1}$$

- Complements: $P(A^C) = 1 - P(A)$ (draw diagram)
- Union ($\cup$) and intersection ($\cap$)
- Inclusion-exclusion rule: (for 2 sets)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{3.2}$$

  This can be generalized: the cardinality of the union of n sets:

  1. Include the cardinalities of the sets.
  2. Exclude the cardinalities of the pairwise intersections.
  3. Include the cardinalities of the triple-wise intersections.
  4. Exclude the cardinalities of the quadruple-wise intersections.
  5. ... continue, until the n-tuple-wise intersection

- Conditional probability: $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$
- Bayes rule (from above):

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)P(A|B)}{P(A)} \tag{3.3}$$

- INDEPENDENCE: if events A, B are independent,

$$P(A \cap B) = P(A)P(B) \tag{3.4}$$

4 Expectation and uncertainty

- The EXPECTATION value of a RV is the average value of the RV. For RV X,

$$E(X) = \sum_{x'} x' P(X = x') \tag{4.1}$$

- Properties:

  1. $E(X + Y) = E(X) + E(Y)$ where X, Y do not have to be independent (there's a cool proof just by rearranging terms in the sum pg 6)
  2. $E(a \cdot C) = a \cdot P(X)$ where a is a constant.
  3. In general, $E[f(X, Y)] \neq f(E(Y), E(Y))$. For example, $E(X \cdot Y) \neq E(X)E(Y)$.

- UNCERTAINTY: Variance is expectation value of the squared deviation from the mean μ.

$$Var(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2 \tag{4.2}$$

  Properties of variance:

1. $Var(aX + b) = a^2 \cdot Var(X)$

In physics, we call the standard deviation the uncertainty,

$$SD(x) = \sqrt{Var(X)} \tag{4.3}$$

Covariance: for random variables X, Y

$$Cov(X, Y) = E[(x - E(x))(y - E(y))] \tag{4.4}$$

For X, Y independent, $Cov(X, Y) = 0$

5 Distributions

- BERNOULLI trial: tossing a weighted coin with probability p of getting heads,

$$\text{Bernoulli}(p) = \begin{cases} 1, & p \\ 0, & 1 - p = q \end{cases} \tag{5.1}$$

- BINOMIAL distribution: the probability of k successes in n Bernoulli(p) trials.

$$\text{Binomial}(p) = \binom{n}{k} p^k q^{n-k}; \quad \text{for } k = 0, 1, \ldots, n \tag{5.2}$$

n *choose* k: the number of ways to choose k successful trials among n total trials.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{5.3}$$

Mean: np, Variance: npq

- Intuitive picture: the binomial theorem.

$$(p + q)^n = \sum_{k=0}^{\infty} \binom{n}{k} p^k q^{n-k} \tag{5.4}$$

- The binomial distribution can be generalized:

	2 outcomes per trial	**multiple outcomes per trial**
Independent trials (w/replacement)	binomial	multinomial
Dependent trails (wo/replacement)	hypergeometric	multivariate hypergeometric

- NORMAL distribution. Q: how tall is the Campanille? If I were to give you a meter stick, ladder, and a well-written liability release form, you could measure for yourself! If you were to measure the height 20 times, we would expect a slightly different value each time. If we were to plot the measured values, we would expect a distribution that looks like a normal distribution.

  How can we derive the normal distribution. Idea: (1) errors are symmetric about the true value, so the true value must be the mean, (2) the larger the error, i.e. farther from the mean, the less probable we are to measure that value. Can we express this as a differential equation for the probability density?

  ~~Data is normally distributed if the rate that the probability density falls is proportional to the distance from the mean.~~ (Almost)

  $$\frac{\mathrm{d}f}{\mathrm{d}x} = -k(x-\mu) \quad \implies \quad f(x) = -\frac{k}{2}(x-\mu)^2 \tag{5.5}$$

  The probability density here is a parabola, and will go negative. Instead,

  Data is normally distributed if the *rate* that the probability density falls is *proportional* to the distance from the mean *and* the probability density at that location.

  $$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = -k(x-\mu)f(x) \tag{5.6}$$

  $$\int \frac{1}{f(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x}\mathrm{d}x = \int -k(x-\mu)\mathrm{d}x \tag{5.7}$$

  $$\ln(f) = -\frac{k}{2}(x-\mu)^2 + \ln(A) \tag{5.8}$$

  $$f(x) = Ae^{-\frac{1}{2}(x-\mu)^2} \tag{5.9}$$

  By normalization, we will find,

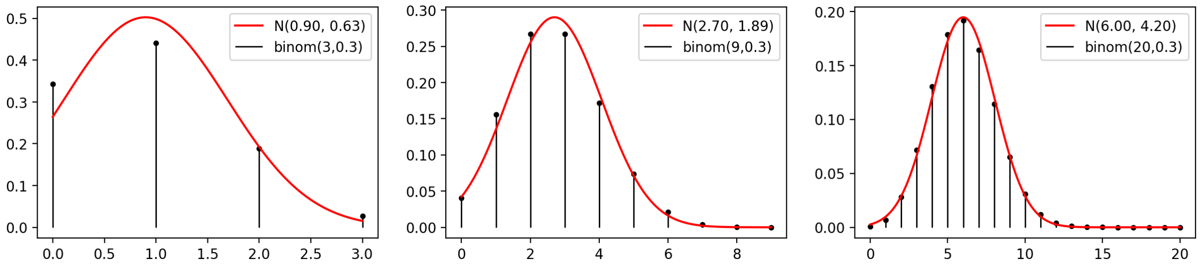
  $$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(x-\mu)^2} \tag{5.10}$$

  Notice that $f(x)$ is the probability density function. This means the probability is given by,

  $$P(x \in [x_0, x_1]) = \int_{x_0}^{x_1} f(x)\mathrm{d}x \tag{5.11}$$

  We can say the RV X is normally distributed with the notation $X \sim N(\mu, \sigma^2)$

- Normal approximation to the binomial: for n sufficiently large, the normal distribution approximates the binomial with mean np and variance npq.



- POISSON distribution: A Poisson process is a simple stochastic model for arrivals. What do we mean by arrivals?

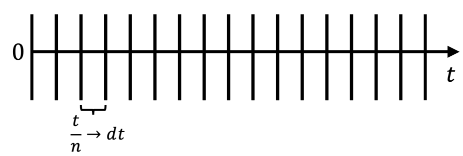
  Suppose you build a muon detector. For the size of the detector, you expect (i.e. on average) to observe λ arrivals per unit time. If you wait t seconds: what is the probability distribution of the number of muons you will have observed? This is modeled by the Poisson distribution.

  Assumptions:

  1. The average rate of an event (i.e. an arrival) is λ
  2. The events are independent
  3. Two events can not occur at the same time

  What is the probability that we observe x events in time t?

  We break up the time t into n intervals. For n large enough (i.e. when we take $n \to \infty$), the time interval is so short that, at most, one event can occur. The probability of this event occurring is $\lambda(t/n)$. Therefore, we can model each small time intervals as a Bernoulli trial with $p = \lambda t/n$



  Then, the probability of x arrivals after waiting some time t is given by the binomial distribution (x successes in n Bernoulli($\lambda t/n$) trials as $n \to \infty$),

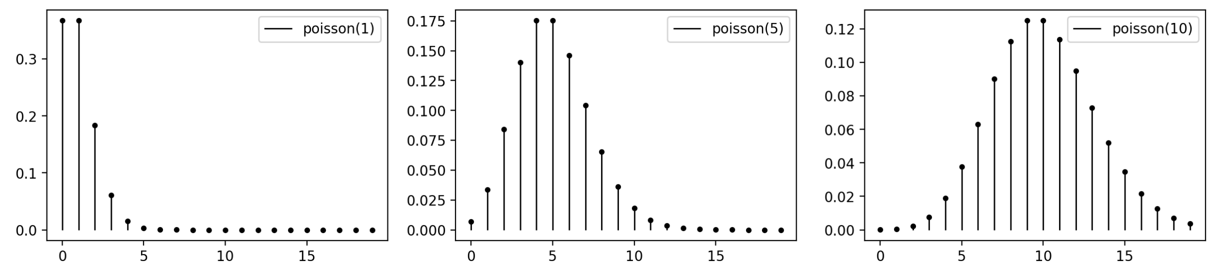
$$P(x \text{ arrivals}) = \lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x}$$

$$= \frac{n!}{x!(n-x)!} \frac{(\lambda t)^x}{n^x} \left(1 - \frac{\lambda t}{n}\right)^n \cancelto{1}{\left(1 - \frac{\lambda t}{n}\right)^{-x}} = \frac{n!}{(n-x)!\, n^x} \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

$$= \left[\frac{n(n-1)(n-2)\ldots(n-x+1)}{n^x}\right] \frac{(\lambda t)^x}{x!} e^{-\lambda t} = \cancelto{1}{\left[\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \ldots \cdot \frac{n-x+1}{n}\right]} \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

$$P(x \text{ arrivals}) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \tag{5.12}$$

The mean and variance of Poisson(λt) is λt



- The Gamma distribution is closely related to the Poisson process. You can read more in a statistics textbook. The GAMMA FUNCTION is derived from the Gamma distribution, and will be relevant when we discus the χ^2 distribution.
- Gamma function:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \mathrm{d}x \tag{5.13}$$

For $z \in \mathbb{N}$, $\Gamma(z) = (z-1)!$. The Gamma function can be considered an analytic continuation of the factorial to a larger domain.

6 Summary

- Today we discussed a basic summary of the formalism of probability, RVs, and some common distributions.
- There is one more important distribution: the χ^2 distribution, which we will discuss next week
- Next time we will also discuss curve fitting and understanding if a model is a good fit.

A Sample vs. Population Variance

See "Proof of Bessel's Correction" in local directory.