Summary of Probability

Mathematical Physics I

Rules of Probability

The probability of an event is called *P(A)*, which is a positive number less than or equal to 1. The total probability for all possible exclusive outcomes should be 1.

The *joint probability* of two different events is denoted *P(A,B).*

Conditional probability (probability of A given B)

$$P(A|B) \equiv \frac{P(A,B)}{P(B)}$$

Independence

$$P(A|B) = P(A) \qquad P(A,B) = P(A)P(B)$$

Law of total probability

$$P(A) = \sum_i P(A,B_i) = \sum_i P(A|B_i)P(B_i)$$

Bayes' Theorem

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

Expectation value of a function

$$E(g) = \sum_{i=1}^{N} g(A_i)P(A_i)$$

Probability density function (p.d.f.)

Here we have assumed that the outcomes are discrete, but they also hold for continuous measurements. In this case, the probability is described by a *p.d.f.*, such that

$$P(x \le X \le x + dx) \; = \; f(x)dx.$$

Sums over all possible discrete outcomes become integrals over possible results,

$$\int_{-\infty}^{\infty} f(x)dx = 1 \qquad f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad E(g) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Cumulative probability function

$$F(x) \equiv P(X \le x) = \int_{-\infty}^{x} f(y)dy$$

Common probability distributions

A few distribution functions arise quite often. Here are a few, along with the corresponding mean ($\mu \equiv E(x)$) and variance ($\sigma^2 \equiv E(x^2) - E(x)^2$.) These help describe the average and the width (or dispersion) of the distributions.

Uniform distribution

One common distribution is uniform; this occurs for example in coin flips or dice rolls.

For discrete distributions, the probability of any event is $p_i = 1/N$, where N is the number of possible outcomes. Examples include drawing a random card or rolling a die.

For continuous distributions, the p.d.f. is $f(x) = 1/\lambda$, where $\lambda = x_{max} - x_{min}$ is the range possible values and must be finite. The mean of this distribution is $\mu = (x_{min} + x_{max})/2$, and the variance is $\sigma^2 = \lambda^2/12$.

Binomial distribution

Suppose you have a number of independent trials with two possible outcomes (A and B), such as flipping coins, with probabilities given as p and $1 - p$. The distribution of the number of times A is seen given n trials is given by the binomial distribution:

$$\boxed{p_r = \frac{n!\ p^r(1-p)^{n-r}}{(n-r)!\ r!}}$$

The mean of this distribution is $= n\,p$, and its variance is $\sigma^2 = n\,p\,(1-p)$.

Poisson distribution

This is a discrete distribution described by

$$p_n = \frac{\lambda^n e^{-\lambda}}{n!} \qquad (n = 0, 1, 2, 3, \cdots)$$

This often occurs in counting statistics, and is the limit of other important distributions, like the binomial distribution in the large *n* limit.

$$\equiv E(n) = \textstyle\sum_0^\infty n p_n = \lambda \qquad\qquad \sigma^2 \equiv E(n^2 - E(n)^2) = \textstyle\sum_0^\infty (n^2 - \lambda^2)\ p_n = \lambda$$

Gaussian distribution

The Gaussian distribution is also known as the normal distribution, $N(\mu, \sigma^2)$, and is one of the most important distributions in statistics:

$$f(x) = \frac{1}{(2\pi)^{1/2}\sigma}\; e^{-(x-\mu)^2/2\sigma^2}$$

$\equiv mean \qquad \sigma^2 \equiv variance \qquad \sigma \equiv root\ mean\ squared\ (r.m.s.)\ value$

Moments (or integrals) of the normal distribution:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \qquad E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu \qquad E(x^2 - \mu^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx = \sigma^2$$

The *Central Limit theorem* says that if you add many like variables together, the distribution of their sums will approach a Gaussian, even if the original distributions are not Gaussian!

The probability of falling within a gap near the peak of the Gaussian distribution is:

Range	***Probability of falling in range***
$\mu - \sigma \leq x \leq \mu + \sigma$	*68.27%*
$- 2\sigma \leq x \leq \mu + 2\sigma$	*95.45%*
$- 3\sigma \leq x \leq \mu + 3\sigma$	*99.73%*

Such calculations can be made using the cumulative integral of the Gaussian function, which is closely related to the ***Error function***. This is usually compiled in tables like in Appendix H of Jordan & Smith.

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt$$

Learning with Bayesian inference

Two general approaches to statistics exist, known as the ***frequentist*** and ***Bayesian*** methods. The classical approach is the frequentist method, and most texts focus on this kind of analysis. However, the Bayesian approach is becoming more widely accepted.

When the data are good, these two approaches will generally result in the same conclusions. However, there are some deep philosophical differences between the approaches, mostly based on what is meant by probability and whether it is absolute or subjective. The Bayesian approach requires an explicit assumption of the believability of a model (called a ***prior***) which could be different for different people. It then uses observations to update these beliefs using Bayes' theorem.

Bayes' theorem shows us how to update our confidence in a theory or model (M) given the results of a new experiment, observation or data (D) :

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

Here, $P(M)$ is our evaluation of the probability of the model before the new data; this is known as the ***prior distribution,*** and is usually based on earlier data or theoretical ideas.

$P(M|D)$ is the new evaluation of the probability of the model after the new data is considered. It is called the ***posterior distribution.***

$P(D|M)$ is called the ***likelihood*** and it describes how likely the data would be if the model were true. This is usually more straightforward to calculate.

$P(D)$ is a normalising factor called the ***Bayesian evidence***.

Data and estimators of mean and variance

Suppose we have a sample of N independent measurements of a quantity (x_i); it is useful to have statistics which describe its distribution:

Sample mean estimator

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

Sample variance estimator

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

The mean is the average value, while the square root of the variance (known as the *standard deviation* or root mean square deviation) describes the error in the estimate of that value. We tend to write this as, $x = \bar{x} \pm s$.

We really want to know the properties of the underlying distribution, not of the sample itself. We hope that our estimators we calculate from the sample will tend to the true properties as the number of samples increases. If this is true, the estimator is *unbiased.*

Propagation of Errors

Often we must estimate the errors on quantities that aren't measured directly, but are related to things which are measured. The errors in these derived quantities can be calculated:

$$f(u,v): \quad \sigma_f^2 = \sigma_u^2 \left(\frac{\partial f}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial f}{\partial v}\right)^2$$

Sums and differences:

$$f = au \pm bv: \qquad \sigma_f^2 = a^2\sigma_u^2 + b^2\sigma_v^2$$

Products and quotients:

$$f = auv; \frac{au}{v}: \qquad \frac{\sigma_f^2}{f^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2}$$

Note that these equations ignore potential correlations between *u* and *v* measurements.

Rule of thumb: When adding or subtracting, the absolute errors add in quadrature. When multiplying or dividing, the percentage errors add in quadrature.

Goodness of fit test (χ^2 or chi-squared test)

Given a collection of data *(x_i, $y_i \pm \sigma_i$)*, we can evaluate the quality of the fit of a model for that data *y(x)* by use of the chi-squared statistic:

$$\chi^2 \equiv \sum (y_i - y(x_i))^2/{\sigma_i}^2$$

This is closely related to the likelihood of the data given the model, assuming the errors are distributed with a Gaussian distribution. The smaller chi-squared is, the larger the likelihood is.

Typically, the chi-squared statistic is expected to be around the number of ***effective degrees of freedom*** which is the number of data points *minus* the number of parameters in the model being fit. If these are much different, then it may be indicating a problem with the model.

We usually try to find the best fit parameters for a model by finding those which minimise the chi-squared value (thereby maximising the likelihood of the data.) This is sometimes called a ***least-squares fit*** to the data, and is also known as a ***regression.***

The most common application is to a linear model for the data, that is assuming that $y = a + bx$. For a linear model, it is easy to find the parameters which minimise the chi-squared.

$$\chi^2 \equiv \sum (y_i - a - bx_i)^2/{\sigma_i}^2$$

The best-fit parameters of the model are relatively simple to calculate if we assume all the errors are the same ($\sigma_i = \sigma$):

$$\Delta \equiv N \sum {x_i}^2 - \left(\sum x_i\right)^2$$

$$\hat{a} = \frac{1}{\Delta}\left(\sum {x_i}^2 \sum y_i - \sum x_i \sum x_i y_i\right)$$

$$\hat{b} = \frac{1}{\Delta}\left(N \sum x_i y_i - \sum x_i \sum y_i\right)$$

$${\sigma_a}^2 = \frac{\sigma^2}{\Delta} \sum {x_i}^2$$

$${\sigma_b}^2 = \frac{N\sigma^2}{\Delta}$$