Probability and statistics lectures (PHYS 6350)

J.D. Romano

Fall 2025

Abstract

The following notes are based on material from Boas (Chapter 15), Mathews and Walker (Chapter 14), Sec III of Living Reviews in Relativity article by Romano and Cornish

1 Suggested problems

From Boas, Chpt 15: 1.2, 1.5, 2.13, 3.3, 3.5, 3.18, 3.19, 4.10, 5.9, 5.11, 5.13, 5.16, 6.13, 6.14, 6.15, 6.16, 7.9, 7.12, 7.13, 8.8, 9.1, 9.2, 9.4, 9.7, 9.10, 10.1, 10.2, 10.3

2 Motivation

- Mathematics is based on *deduction*: one starts with a set of postulates (assumed to be true) and then derives the valid consequences of these postulates using the rules of logic.
- Mathematical statements are either true or false, with no uncertainty.
- Which postulates to start with is the creative aspect of mathematics.
- Example: Euclidean (flat space) vs non-Euclidean (curved space) geometry depends on whether or not we include the postulate that "two lines that are initially parallel never intersect and never diverge".
- Science, on the other hand, is based on *induction*, trying to make general statements about the state of nature from observations, which are *incomplete* and *imprecise* (due to measurement noise). Thus, our conclusions about nature are necessarily uncertain.
- Probabilistic inference (aka "plausible inference", "statistical inference") is the way that science deals with uncertainty.

3 Preliminaries

- Start by considering games of chance: flipping coins, rolling die, drawing cards from a deck, selecting colored balls from a bag, ... These are all examples of "repeatable, identical experiments", which is needed for the "classical" (or frequentist) interpretation of probability.
- According to the frequentist definition, probabilities cannot be assigned to hypotheses or parameters describing the state of nature, which have fixed but unknown values.
- NOTE: A Bayesian has a more general definition of probability, interpreting it as "degree of belief" (or confidence, plausibility) in any proposition (need not be associated with repeatable identical experiments). So a Bayesian can assign probabilities to hypotheses or parameters describing the state of nature.
- In what follows, we will restrict attention to the frequentist interpretation of probability, but note that all of the following results apply as well to the more general Bayesian interpretation.

- An example of an "experiment" would be rolling two die and recording the numbers on the the two faces of the die.
- For such experiments, the probability of an event A is defined as the long-run relative frequency of A occurring in a set of repeatable, identical experiments:

$$P(A) \equiv \lim_{N\to\infty} N(A)/N \tag{1}$$

  Here $N(A)$ is the number of times that the event A occurred and N is the total number of times that the experiment was performed.
- Sample space: set of *mutually-exclusive*, *collectively-exhaustive* outcomes of an experiment.
- Mutually-exclusive means that outcome A and B do not occur simultaneously (i.e., the joint probability of A and B is zero). Collectively-exhaustive means all possible outcomes are accounted for.
- Examples: for two successive flips of a coin, we can consider two different sample spaces:

$$S_1 \equiv \{tt, th, ht, hh\}\,, \qquad S_2 \equiv \{0h, 1h, 2h\} \tag{2}$$

  where $0h = tt$, $1h = \{th, ht\}$, $2h = hh$.
- Events in S_1 are equally likely (an example of a *uniform* sample space), while the events in S_2 have different probabilities (non-uniform sample space):

$$P(A \in S_1) = \{0.25, 0.25, 0.25, 0.25\}\,, \qquad P(A \in S_2) = \{0.25, 0.50, 0.25\} \tag{3}$$

3.1 Algebra of probability

- Probabilities satisfy the following:

  (i) Possible values:

$$0 \le P(A) \le 1 \tag{4}$$

  with 1 or 0 if A is certain to occur or certain not to occur.

  (ii) Sum rule:

$$P(A) + P(\bar{A}) = 1 \tag{5}$$

  where $\bar{A}$ is the complement of A (i.e., all outcomes different from A, with no overlap).

  (iii) Product rule:

$$P(A, B) = P(A|B)P(B) \tag{6}$$

  where $P(A, B)\,(= P(B, A))$ is the *joint probability* for both A and B to occur, and $P(A|B)$ is the *conditional probability* for A to occur assuming that B occurs.
- Note that $P(B|A) \neq P(A|B)$.
- Example: $A =$ "person is a woman"; $B =$ "person is pregnant". Then $P(B|A) \approx 0.30$, but $P(A|B) = 1$.
- If the conditional probability $P(A|B)$ is independent of B (i.e., $P(A|B) = P(A)$), then A and B are said to be *statistically independent*.
- It follows that if A and B are statistically independent then

$$P(A, B) = P(A)P(B) \tag{7}$$

- If A and B are *mutually exclusive*, then

$$P(A, B) = 0 \tag{8}$$

- Define $P(A+B)$ to be the probability that "events A or B or both A and B occur". Then

$$P(A+B) = P(A) + P(B) - P(A,B) \tag{9}$$

- Example: Consider a sample space for which events A and B both occupy $2/3$ of the total area, with $1/3$ overlap. In terms of probabilities, we have

$$P(A) = P(B) = 2/3\,, \quad P(A,B) = 1/3\,, \quad P(A|B) = P(B|A) = 1/2 \tag{10}$$

Note also that

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A)\,, \quad P(A+B) = P(A) + P(B) - P(A,B) = 1 \tag{11}$$

3.2 Bayes' theorem

- Bayes' theorem:

$$P(B|A)P(A) = P(A|B)P(B) \tag{12}$$

or, equivalently,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{13}$$

- Bayes' theorem is an immediate consequence of $P(A,B) = P(B,A)$.

- Applications of Bayes' theorem:

Exercise: Suppose on your last visit to the doctor's office you took a test for some rare disease. This type of disease occurs in only 1 out of 10,000 people, as determined by a random sample of the population. You are told that the test you took is rather effective in that it can correctly identify the presence of the disease 95% of the time, but it gives false positive 1% of the time.

Suppose the test came up positive. What is the probability that you have the disease?

Solution:

Let H = have the disease; $+$ = test positive

Information:

$$P(H) = 0.0001\,, \quad P(\bar{H}) = 0.9999\,, \quad P(+|H) = 0.95\,, \quad P(+|\bar{H}) = 0.01\,, \tag{14}$$

Calculate

$$P(H|+) = \frac{P(+|H)P(H)}{P(+)} \tag{15}$$

where

$$P(+) = P(+|H)P(H) + P(+|\bar{H})P(\bar{H}) = 0.95 \times 0.0001 + 0.01 \times 0.9999 \approx 0.01 \tag{16}$$

Thus,

$$P(H|+) \approx \frac{0.95 \times 0.0001}{0.01} \approx 0.01 \tag{17}$$

so 1% chance that you have the disease given that you tested positive

- Exercise: (Boas, Chpt 15, Sec 3, Example 3):

A prelim exam is given to students before they take a course. You are told that 95% of the students pass the course, so 5% fail the course. You are also told that 96% of the students who pass the course passed the prelim; while 25% of the students who fail the course passed the prelim.

What is the probability that a student will pass the course if they fail the prelim?

Solution:

Let A = student fails prelim and B = student passes course.

From the above information, we have:

$$P(B) = 0.95\,, \quad P(\bar{B}) = 0.05\,, \quad P(\bar{A}|B) = 0.96\,, \quad P(\bar{A}|\bar{B}) = 0.25 \tag{18}$$

Using the sum rule, we can also conclude:

$$P(A|B) = 0.04\,, \quad P(A|\bar{B}) = 0.75 \tag{19}$$

Thus,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{20}$$

where $P(A|B)$ and $P(B)$ are given above, and

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) = 0.04 \times 0.95 + 0.75 \times 0.05 = 0.0755 \tag{21}$$

leading to $P(B|A) \approx 0.50$.

4 Permutations, combinations, ...

- The total number of permutations of n objects equals

$$P(n) \equiv n(n-1)\cdots 1 = n! \tag{22}$$

since there are n possible choices for the first object, $n-1$ choices for the second object, ...

- The number of permutations of n objects taken m at a time

$$P(n,m) \equiv n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!} \tag{23}$$

- The number of *combinations* of n objects taken m at a time ($m \leq n$) (so ignoring order):

$$C(n,m) \equiv \frac{P(n,m)}{P(m)} = \frac{n!}{(n-m)!\,m!} \equiv \binom{n}{m} \tag{24}$$

- $C(n,m)$ are the standard binomial coefficients, which appear when expanding $(a+b)^n$, namely

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^{(n-m)}b^m = a^n + na^{(n-1)}b + \frac{1}{2!}n(n-1)a^{(n-2)}b^2 + \cdots + b^n \tag{25}$$

- The above ideas are useful in statistical mechanics, where we consider distributing m particles in n states. (The m particles can be thought of as "balls" that we need to distribute among n "boxes").

- Maxwell-Boltzmann statistics: Consider (classical) distinguishable particles, like gas molecules in a container. The total number of equally-likely arrangements of m such particles in n different states is simply

$$nnn\cdots n = n^m \tag{26}$$

since there are n choices for each particle.

- Fermi-Dirac statistics: Consider m identical particles, for which only one or zero such particles can occupy a state (i.e., Pauli exclusion principle for fermions). Then the total number of equally-likely arrangements of m such particles in n different states is simply

$$\binom{n}{m} = \frac{n!}{(n-m)!\,m!} = \frac{n(n-1)\cdots(n-m+1)}{m!} \tag{27}$$

because each particle can occupy only one state, and the order of the particles doesn't matter since the particles are identical.

- Bose-Einstein statistics: Consider m identical particles, for which more than one particles can occupy the same state (i.e., bosons). Then the total number of equally-likely arrangements of m such particles in n different states is

$$\binom{n+m-1}{m} = \frac{(n+m-1)!}{(n-1)!\,m!} = \frac{(n+m-1)(n+m-2)\cdots n}{m!} \tag{28}$$

  Note that in comparison with the Fermi-Dirac result, the total number of equally-likely arrangements is larger for bosons since the effective number of states has increased from n to $n+m-1$.

5 Probability distributions

- A random variable is a function on sample space, e.g., x is the number of heads obtained when flipping a coin n times or the sum of the number of two die that are rolled repeatedly.
- The different possible values x_i of a random variable x have different probabilities $P_i = P(x_i)$.
- Random variables can be either discrete or continuous.
- A continuous random variable x has a *probability distribution* or *probability density function* $p(x)$, which has dimension of 1/dimension of x. This means that

$$p(x_0)\,\mathrm{d}x = \mathrm{Prob}(x \in [x_0, x_0 + \mathrm{d}x]) \tag{29}$$

- More generally,

$$\int_a^b \mathrm{d}x\, p(x) = \mathrm{Prob}(a \le x \le b) \tag{30}$$

5.1 Mean, variance, etc.

- *Mean* of x:

$$\mu \equiv \sum_i x_i P(x_i) \qquad \text{or} \qquad \mu \equiv \int_{-\infty}^{\infty} \mathrm{d}x\, xp(x) \tag{31}$$

  also denoted by $\langle x \rangle$.

- *Variance* of x:

$$\mathrm{Var}(x) \equiv \sum_i (x_i - \mu)^2 P(x_i) \qquad \text{or} \qquad \mathrm{Var}(x) \equiv \int_{-\infty}^{\infty} \mathrm{d}x\, (x-\mu)^2 p(x) \tag{32}$$

  also denoted by σ^2. Note that $\sigma \equiv \sqrt{\sigma^2}$ is called the *standard deviation*.

- Exercise: Prove that

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \tag{33}$$

- *Cumulative density function*:

$$F(x_i) \equiv \sum_{x_j \le x_i} P(x_j) \qquad \text{or} \qquad F(x) \equiv \int_{-\infty}^{x} \mathrm{d}y\, p(y) \tag{34}$$

- *nth-moment* of a probability distribution:

$$\langle x^n \rangle = \int_{-\infty}^{\infty} \mathrm{d}x\, x^n p(x) \tag{35}$$

- *Characteristic function*:

$$\phi(k) \equiv \int_{-\infty}^{\infty} \mathrm{d}x\, e^{ikx} p(x) \tag{36}$$

which is just the Fourier transform of $p(x)$.

- Inverse relation:

$$p(x) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}k\, e^{-ikx} \phi(k) \tag{37}$$

which is the inverse Fourier transform.

- The characteristic function is a *moment-generating function*, which can be seen by expanding the rhs of the definition of $\phi(k)$ as a power series in k and then performing the integrals over x:

$$\phi(k) = \langle e^{ikx} \rangle = 1 + ik\langle x \rangle - \frac{k^2}{2!}\langle x^2 \rangle + \cdots + \frac{(ik)^n}{n!}\langle x^n \rangle + \ldots \tag{38}$$

This leads to

$$\langle x^n \rangle = (-i)^n \frac{\mathrm{d}^n \phi}{\mathrm{d}k^n}\bigg|_{k=0} \tag{39}$$

5.2 Joint probability distributions

- Let x and y be continuous random variables with joint probability distribution $p(x, y)$. Then the probability distributions for x and y separately are given by the integrals

$$p_X(x) = \int_{-\infty}^{\infty} \mathrm{d}y\, p(x, y)\,, \quad p_Y(y) = \int_{-\infty}^{\infty} \mathrm{d}x\, p(x, y) \tag{40}$$

- The *covariance* of x and y is defined by

$$\mathrm{Cov}(x, y) \equiv \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle \tag{41}$$

where

$$\langle f(x, y) \rangle \equiv \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y\, f(x, y)\, p(x, y) \tag{42}$$

for any function $f(x, y)$.

- Exercise: Show that if x and y are statistically independent (i.e., $p(x, y) = p_X(x) p_Y(y)$), then $\mathrm{Cov}(x, y) = 0$.

- But note that the converse is not true. In other words, $\mathrm{Cov}(x, y) = 0$ does *not* imply in general that x and y are statistically independent.

- An exception to this last statement is for x and y described by a *multivariate Gaussian* distribution (which we will describe in the next section). For this case, $\mathrm{Cov}(x, y) = 0$ implies that x and y are statistically independent, described by a product of 1-dimensional Gaussian distributions for x and y separately.

- For convenience, denote the two random variables by x_1, x_2. Then the characteristic function for the joint probability distribution $p(x_1, x_2)$ is defined as

$$\phi(k_1, k_2) = \int \mathrm{d}x_1 \int \mathrm{d}x_2\, e^{i(k_1 x_1 + k_2 x_2)} p(x_1, y_2) \tag{43}$$

which is a 2-d Fourier transform of $p(x_1, y_1)$.

- Expanding the exponential on the rhs leads to

$$\phi(k_1, k_2) = 1 + i\left(k_1\langle x_1\rangle + k_2\langle x_2\rangle\right) - \frac{1}{2!}\left(k_1^2\langle x_1^2\rangle + k_2^2\langle x_2^2\rangle + 2k_1k_2\langle x_1x_2\rangle\right) + \cdots \tag{44}$$

  in terms of the moments of x_1 and x_2.

- It immediately follows from the above definition of $\phi(k_1, k_2)$ that if x_1 and x_2 are statistically independent, then $\phi(k_1, k_2)$ has the form of a product

$$\phi(k_1, k_2) = \phi_1(k_1)\phi_2(k_2) \tag{45}$$

- The above definitions and results applies to an arbitrary number of random variables $x_1, x_2, \cdots, x_n$.

- Consider a new random variable via $z \equiv f(x, y)$, for some function $f(x, y)$. Then the characteristic function $\phi_Z(k)$ of $z = f(x, y)$ is given by

$$\phi_Z(k) = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y\, e^{ikf(x,y)} p(x, y) \tag{46}$$

- Exercise: Using the above expression for $\phi_Z(k)$, show that the probability distribution $p_Z(z)$ for $z = f(x, y)$ is given by

$$p_Z(z) = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y\, \delta(z - f(x, y))\, p(x, y) \tag{47}$$

- Example: Assume that x and y are statistically-independent random variables having probability distribution $p_X(x)$ and $p_Y(y)$ and characteristic functions $\phi_X(k)$ and $\phi_Y(k)$. Define a new random variable

$$z \equiv x + y \tag{48}$$

  Show that the characteristic function $\phi_Z(k)$ for $z = x + y$ is the product of the characteristic functions for x and z:

$$\phi_Z(k) = \phi_X(k)\phi_Y(k) \tag{49}$$

- We will use the above result in the next section when we prove the *central limit theorem*—i.e., that the probability distribution of the average of n independent samples of a random variables x with probability distribution $p(x)$ is Gaussian in the limit $n \to \infty$, *independent* of the form of the probability distribution $p(x)$.

6 Gaussian distribution

- The familiar symmetric "bell-shaped" curve that people often draw is for a *Gaussian probability distribution*:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{50}$$

- The above distribution is properly normalized and has mean and variance given by the parameters μ and σ^2 appearing in the definition of $p(x)$.

- Exercise: Prove those statements.

- The *mode* (correponding to the maximum $p(x_{\text{mode}})$) and *median* (corresponding to total probability equal to 50% below (or above) x_{median}) agree with the mean for a Gaussian probability distribution.

- Exercise: Prove that the characteristic function $\phi(k)$ of the Gaussian distribution is given by

$$\phi(k) = e^{ik\mu} e^{-k^2\sigma^2/2} \tag{51}$$

  Recall that we saw this earlier in the topic on Fourier transforms—namely, that the Fourier transform of a Gaussian is another Gaussian.

- Central-limit theorem:

  Consider n independent samples x_i of a random variable x described by *any* probability distribution $p(x)$. Let μ and σ^2 denote the mean and variance of the random variable x (and hence the mean and variance of each of the random samples x_i).

  Define the new random variable

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{52}$$

  which is just the arithmetic mean of the N independent samples of x. Given this definiton of $\bar{x}$, it follows that

$$\langle \bar{x} \rangle = \mu\,, \qquad \mathrm{Var}(\bar{x}) = \sigma^2/n \tag{53}$$

  The central limit theorem says then says that in the limit $n \to \infty$, the probability distribution of $\bar{x}$ is Gaussian-distributed with mean μ and variance σ^2/n as given above.

- Exercise: Prove the central limit theorem (see Mathews and Walker, Section 14.5).

  Proof: The prove begins by showing that the characteristic function $\phi_Y(k)$ of the random variable $y \equiv \bar{x} - \mu$ is equal to the nth power of $e^{-ik\mu/n}\phi_X(k/n)$, i.e.,

$$\phi_Y(k) = \left[e^{-ik\mu/n} \phi_X\left(k/n\right) \right]^n \tag{54}$$

  where $\phi_X(k)$ is the characteristic function of x (or each random sample x_i).

  It then follows that

$$\begin{aligned} e^{-ik\mu/n}\phi_X(k/n) &= \int \mathrm{d}x\, e^{ik(x-\mu)/n} p(x) \\ &= 1 + i\frac{k}{n}\langle x-\mu\rangle - \frac{1}{2}\frac{k^2}{n^2}\langle (x-\mu)^2\rangle + \cdots \\ &= 1 - \frac{1}{2}\frac{k^2}{n^2}\sigma^2 + \cdots \end{aligned} \tag{55}$$

  as a consequence of $\langle x - \mu \rangle = 0$ and $\langle (x-\mu)^2 \rangle = \sigma^2$.

  Raising this last expression to the nth power and taking the limit $n \to \infty$ yields

$$\phi_Y(k) = \left(1 - \frac{1}{2}\frac{k^2}{n^2}\sigma^2 + \cdots \right)^n \sim e^{-\frac{k^2\sigma^2}{2n}} \tag{56}$$

  where we used the following limiting expression for e^x:

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x \tag{57}$$

  Finally, since the characteristic function of a Gaussian is also a Gaussian, we can conclude that

$$p_Y(y) \sim \frac{1}{\sqrt{2\pi\sigma^2/n}} e^{-\frac{1}{2}\frac{y^2}{\sigma^2/n}} = \frac{1}{\sqrt{2\pi\sigma^2/n}} e^{-\frac{1}{2}\frac{(\bar{x}-\mu)^2}{\sigma^2/n}} \tag{58}$$

- Multi-variate Gaussian distribution:

  Consider two random variables x, y. Then the multivariate Gaussian distribution for x and y is

$$p(x) = \frac{1}{\sqrt{\det(2\pi C)}} e^{-\frac{1}{2}d^T C^{-1} d} \tag{59}$$

  where

$$d = \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}, \qquad C = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \tag{60}$$

  with C being the covariance matrix of x and y. Here, μ_x, μ_y and σ_x^2, σ_y^2 are the means and variances of x, y, respectively, and $0 \le \rho < 1$ is the *correlation coefficient* which enters in to expression for the covariance $\mathrm{Cov}(x, y)$ of x and y.

7 Binomial distribution

- Consider n flips of a fair coin, so that probability of heads for a single flip is $p = 1/2$ (which is also the probability of a tail). Then the probability of m heads out of n flips is given by:

$$P(m) = \binom{n}{m} \left(\frac{1}{2}\right)^m \left(\frac{1}{2}\right)^{(n-m)} = \binom{n}{m} \left(\frac{1}{2}\right)^n \tag{61}$$

- More generally, if you have n trials of some repeatable experiment, such that the probability for a "success" for each trial is p (so that the probability for "failure" is $q \equiv 1 - p$, then

$$P(m) = \binom{n}{m} p^m q^{(n-m)} = \binom{n}{m} p^m (1-p)^{(n-m)} \tag{62}$$

  is the probability of m "successes" out of n trials.

- The binomial distribution has its peak at or close to the value np, which the expected number of successes (which need not be an integer), i.e., $\langle m \rangle = np$.

- The binomial distribution also has variance $\sigma^2 \equiv \langle m^2 \rangle - \langle m \rangle^2 = np(1-p)$.

- Exercise: Prove these last two statements.

- There are two interesting limiting cases of the binomial distribution:

  (i) The limit $n \to \infty$ for fixed p (so that np also tends to infinity). In this limit

$$P(m) \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\frac{(m-np)^2}{np(1-p)}} \tag{63}$$

  which is a Gaussian probability distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$

  Note the the width σ of this distribution relative its mean μ is given by

$$\frac{\sigma}{\mu} = \frac{\sqrt{np(1-p)}}{np} = \sqrt{\frac{1-p}{np}} \tag{64}$$

  which tends to zero as $n \to \infty$.

  (ii) The limit $n \to \infty$ and $p \to 0$ in such a way that np is fixed. In this limit

$$P(m) \sim \frac{(np)^m}{m!} e^{-np} \tag{65}$$

  which is the Poission distribution with $\mu \equiv np$ (see next section).

- Exercise: Prove (i) (see Boas, Chpt 15, Problem 8.8)

  This (rather long) proof requires the use of Stirling's approximation for factorials:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n (1 + \mathcal{O}(1/n)) \tag{66}$$

  and replacing exponents like m and $n - m$ from the factorial approximations by

$$m = np + \delta = np\left(1 + \frac{\delta}{np}\right), \qquad n - m = nq - \delta = nq\left(1 - \frac{\delta}{nq}\right) \tag{67}$$

  where $q \equiv 1 - p$, and then making expansions in powers of δ/n. (See the solution for Boas, Chpt 15, Problem 8.8 for details.)

- Exercise: Prove (ii) (see Boas, Chpt 15, Problem 9.10)

  This makes use of two results:

$$\binom{n}{m} \equiv \frac{n!}{m!(n-m)!} \sim \frac{n^m}{m!}, \tag{68}$$

  and

$$p^m(1-p)^{(n-m)} = p^m(1-p)^n(1-p)^{-m} = p^m \left(1 - \frac{np}{n}\right)^n \left(1 - \frac{np}{n}\right)^{-m} \sim p^m e^{-np} \tag{69}$$

8 Poisson distribution

- As mentioned in the previous section, the Poisson distribution is given by

$$P(m) = \frac{\mu^m}{m!}\, e^{-\mu} \tag{70}$$

- Exercise: Prove that μ is the mean and variance of this distribution.
- One can interpret $P(m)$ as the probability of m events occurring in a unit time interval, where μ is the average number of events in that time interval.
- If you explicitly include the time interval Δt in the probability distribution, then

$$P(m) = \frac{(\lambda \Delta t)^m}{m!}\, e^{-(\lambda \Delta t)} \tag{71}$$

  where λ is the rate of occurrence of the events.

- Examples:

  (i) m is the number of cars that cross an intersection over a 1-minute interval.

  (ii) m is the number of decays of a radioactive substance every second.