Table of Contents

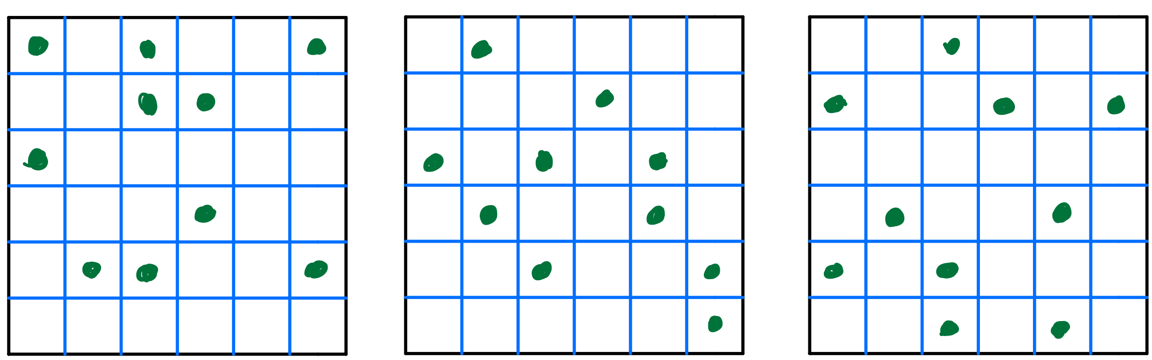
Lecture 1: Probability Theory

Motivation: Our goal in this class is to connect the properties and behavior of molecules to bulk phenomena. Much more can and will be said about this task. For now, we are going to lay a bit of mathematical framework. We need the math to understand the physics.

The mathematical subjects we need are probability theory, statistics, and stochastic processes. There are too many molecules for us to count and track each one. Also, molecules undergo chaotic motion, in the sense that thier position and velocities are hard to predict from their initial state. As such, we are going to treat the collective state of the molecules as a random variable.

This is an innovative leap made by the founders of stat mech.

Example: A lattice gas



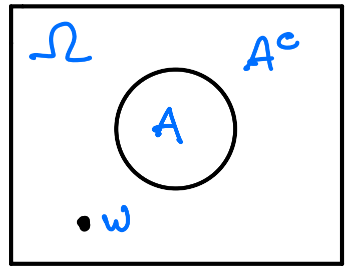
State 1 State 2 State 3

Instead of tracking the trajectories, look at "random" states.

I. Probability Basics

A. Set Notation (a brief review)

Venn diagram



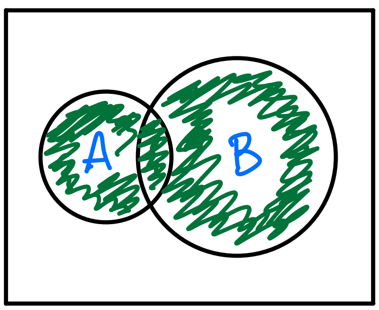
Let Ω be a set of points

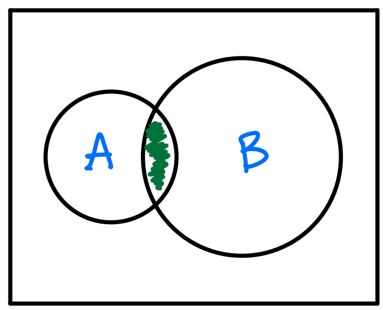
A is a subset of Ω, $A \subset \Omega$

ω is a point in Ω, $\omega \in \Omega$

$A^c = \{\omega \in \Omega : \omega \notin A\}$, the compliment

Suppose $A \subset \Omega$ and $B \subset \Omega$.

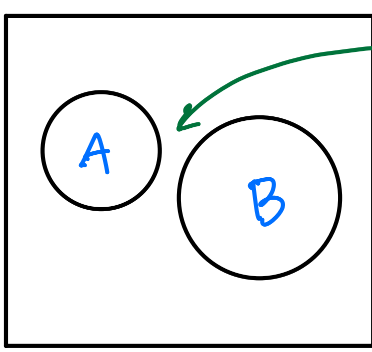




$A \cup B$ "or"
union

$A \cap B$ "and"
intersection

If $A \cap B$ is $\emptyset$, the null set, then we say that A and B are disjoint or mutually exclusive



no area of overlap

B. Probability Axioms

There are several different ways to define probability. By probability I mean a number $P(A)$ assigned to event A that corresponds to how frequently A occurs.

1. classical probability: (e.g. we know beforehand)
$$P(A) = \frac{|A|}{|\Omega|}$$
← "cardinality" = number

2. statistical probability: (e.g. we measure)
$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$
where n are repeated trials in Ω and $n_A \in A$.

3. subjective probability: "degree of belief"
   - Can apply to non-random processes (e.g. a presidential election). This is the domain of philosophers and social scientists. Can be useful, but not our focus here.

4. Axiomatic probability: mathematically rigorous "measure" originally formulated by Kolmogorov

(i.) $P(\emptyset) = 0$, "impossible event"

(ii.) $P(\Omega) = 1$, "sure event"

(iii.) $P(A) \geq 0$, non-negative probabilities

(iv.) If $A_n \cap A_m = \emptyset$ for $n \neq m$ then "countable additivity"
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$
The probabilities of mutually exclusive events add.

these axioms give probabilities certain properties that must always be true. These include:

$$P(A^c) = 1 - P(A)$$

$$P(A) \le P(B) \quad \text{if } A \subset B$$

P(A or B)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

P(A and B)

<u>Example</u>: Deck of cards

$P(\text{ace of spades}) = 1/52$

$P(\text{ten}) = 4/52$

$$P(\text{ace of spades or ten}) = P(\text{ace of spades} \cup \text{ten})$$
$$= P(\text{ace of spades}) + P(\text{ten}) - \underbrace{P(\text{ace of spades} \cap \text{ten})}_{P(\emptyset) = 0}$$
$$= 1/52 + 4/52 = \frac{5}{52}$$

A final comment on the meaning of "probability" and "randomness": Sometimes the ideas of probability and randomness are misunderstood to mean that no order or laws apply. In fact, the lack of "determinism" does not imply this at all. The theories of probability, validated by many experiments, give very precise and orderly laws that predict the range of behavior of single events and the frequency of many events. Even if the event is not deterministic, the probability is.

C. Conditional Probability

Sometimes we want to know the probability of something, given that another event happened. This is called a conditional probability. It is defined as:

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

"A given B"

P (A and B) "joint probability"

Note that:

$$P(A|B) \neq P(B|A)$$

The probability that (for example) I am wet given that it rained is not the same as the probability that it rained, given I am wet. (I might have taken a shower!)

Also, note that $P(B) \neq 0$, or the conditional probability is undefined.

Finally, in the case that the two events are independent, the following holds:

$$P(A|B) = P(A)$$

it doesn't matter what B is on A.

Using the definition, independence implies:

$$P(A \cap B) = P(A)\,P(B)$$

Joint probability / "and"

Kaznessis notation: $P(A \cap B) = P(A,B)$

Contrast this with mutual exclusivity:

$$P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B) \quad \text{"or"}$$

Kaznessis notation: $P(A \cup B) = P(A+B)$

It often occurs that I don't know the joint probability, $P(A \cap B)$. But a little algebra helps us:

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

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

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

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

Bayes' Theorem

"posterior" "informativeness" $= \frac{\text{Likelihood}}{\text{marginal probability}}$ "prior"

Also, it is often the case that I might not know $P(B)$, but I know other conditional probabilities. In this case, I can use the law of total probability:

$$P(B) = P(B|A)\,P(A) + P(B|A^c)\,P(A^c)$$

"not A"

Combining with Bayes' theorem gives

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|A^c)\,P(A^c)}$$

<u>Example:</u> Suppose I go in for a covid test. Only 1% of people have covid. The test is 95% accurate if one has covid, but has a 10% false positive rate. The test comes back positive. What is the probability that I actually have covid?

The question is asking for $P(\text{covid} \mid +)$

Bayes theorem says:

$$P(\text{covid} \mid +) = \frac{P(+ \mid \text{covid})}{P(+)} P(\text{covid})$$

$P(\text{covid}) = 0.01 \qquad P(\text{not covid}) = 0.99$

$P(+ \mid \text{covid}) = 0.95 \qquad P(+ \mid \text{not covid}) = 0.1$

$P(+) = P(+ \mid \text{covid})\,P(\text{covid}) + P(+ \mid \text{not covid})\,P(\text{not covid})$

$= 0.95 \cdot 0.01 + 0.1 \cdot 0.99 = 0.1085$

$P(\text{covid} \mid +) = \frac{0.95}{0.1085} \cdot \underbrace{0.01}_{\text{prior}} = 0.0875 \approx 8.8\%$

prior was low. + test makes it 8x more likely.

What if false positive rate was 0.5%?

$P(+ \mid \text{not covid}) = 0.005 \Rightarrow P(+) = 0.01445$

$P(\text{covid} \mid +) = \frac{0.95}{0.01445} \cdot 0.01 = 0.657 \approx 66\%$

much better test.

D. Combinatorics

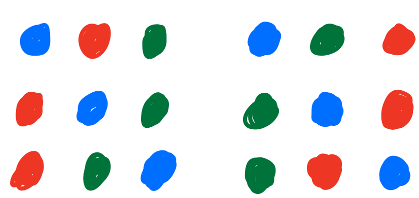
Combinatorics is the mathematical discipline concerned with counting. We can combine some of the more elementary insights in combinatorics with the classical model of probability: $P(A) = |A|/|\Omega|$ to obtain probabilities. This is often useful in games, but is also insightful for molecules.

The fundamental principle of combinatorics is that if event A has cardinality n_A (number of ways to do event A) and event B has cardinality n_B, then the cardinality of event A+B is $n_A \times n_B$.

Case 1: arranging (ordering) objects

(1a) order n different objects $= n!$ (distinuishable)

Example:



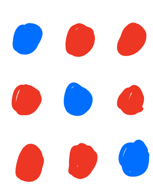
$3! = 6$

(1b) order n objects of which p are identical $= n!/p!$
(indistinguishable)

Example:

$n = p = 3$

$3!/3! = 1$



$n = 3$

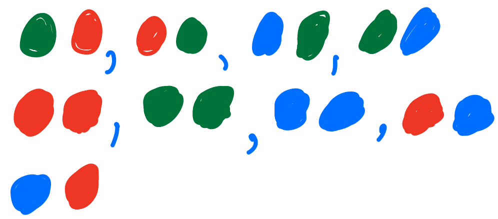
$p = 2$

$\frac{3!}{2!} = 3$

General formula : $\dfrac{n!}{p!\,q!\,r!\ldots}$ for objects $p, q, r, \ldots$

(1c) order k objects from n with replacement $= n^k$

Example :  $n=3$, $k=2$

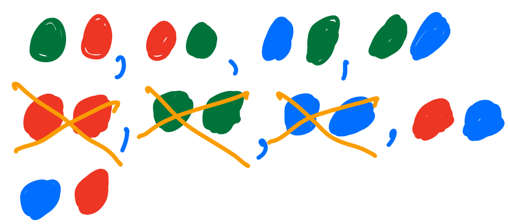


$3^2 = 9$

(1d) order k objects from n w/out replacement $= \dfrac{n!}{(n-k)!}$

"Permutations", $_nP_k$

Example :  $n=3$, $k=2$

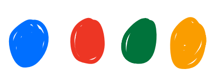


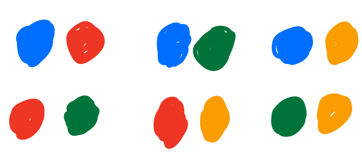
$$\frac{3!}{(3-2)!} = \frac{3\cdot2\cdot1}{1} = 6$$

<u>case 2</u>: Selecting objects (no ordering)

(2a) select k objects from n w/out replacement $= \dfrac{n!}{k!\,(n-k)!} = \dbinom{n}{k}$ binomial coefficient

"combinations", $_nC_k$

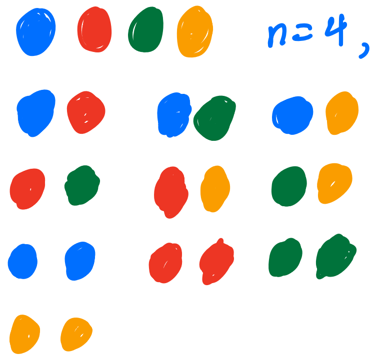
Example :  $n=4$, $k=2$



$$\frac{4!}{2!\,(4-2)!} = \frac{4\cdot3\cdot2\cdot1}{2\cdot1\cdot2\cdot1} = 6$$

(2b) select k objects from n with replacement $= \binom{n-1+k}{k}$

Example: $n=4$, $k=2$



change from 2a

$$\frac{(4-1+2)!}{2!\,(4-1)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot3\cdot2\cdot1} = 10$$

summary: 3 different types of properties

- ordering vs. not ordering
- distinguishable vs. indistinguishable
- replacement vs. no replacement

Stirling's Approximation

An important question for molecules: How do we calculate $n!$ when n is a big number?

$$\lim_{n\to\infty} n! = \sqrt{2\pi n}\, n^n e^{-n}$$

more often we write this as a log:

$$\ln n! = \ln\left[\sqrt{2\pi n}\, n^n e^{-n}\right] = \frac{1}{2}\ln(2\pi n) + n\ln n - n\ln e$$

$$= \frac{1}{2}\ln 2\pi + (n+\tfrac{1}{2})\ln n - n$$

$$\approx n\ln n - n$$

(when $n\to\infty$)

$$\boxed{\ln n! \approx n\ln n - n}$$

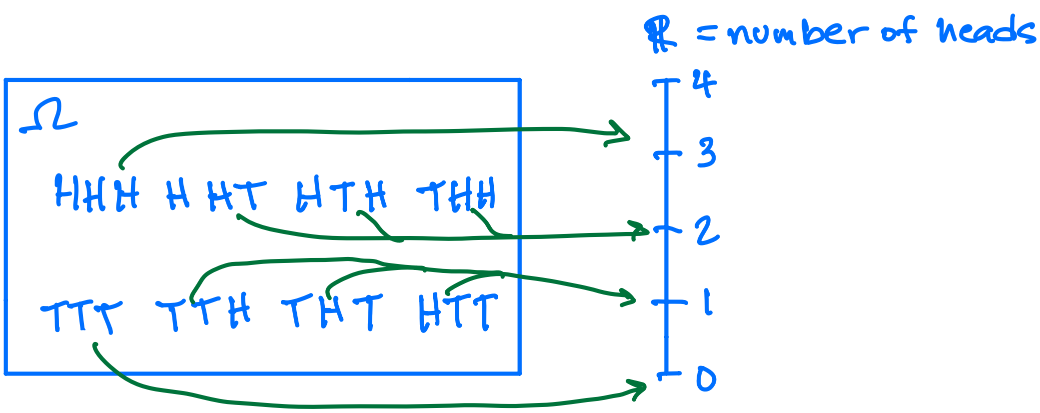
II. Probability Distributions

A. Random Variables

It is often inconvenient to directly deal with probability spaces and event sets. Random variables help us solve this problem.

Random variable are a mapping (i.e. a function) for assigning certain points in an event space Ω numerical values. This facilitates defining probabilities.

Example: Consider an event space Ω for a sequence of three coin tosses.



$$X(\omega) := \begin{cases} 0, & \omega = TTT \\ 1, & \omega \in \{TTH, THT, HTT\} \\ 2, & \omega \in \{HHT, HTH, THH\} \\ 3, & \omega = HHH \end{cases}$$

$P(X=0) = 1/8$

$P(X=1) = 3/8$

$P(X=2) = 3/8$

$P(X=3) = 1/8$

$$P(X=a) = P(\omega \in \Omega : X(\omega) = a)$$

Random variables generally come in two varieties: discrete and continuous.

discrete: the values of $X(\omega)$ are distinct real numbers, usually integers

continuous: the values of $X(\omega)$ are distributed across the real number line

B. Probability distribution functions

Now that we have random variables, we can define functions of those R.V.s that correspond to probabilities.

(i) Probability mass function (pmf)

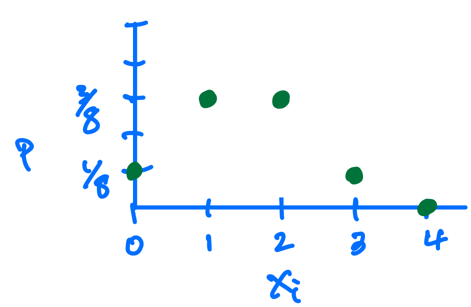
If X is discrete, the function we define is a probability mass function (pmf):

$$p(x_i) := P(X = x_i)$$

A pmf must satisfy the properties:

$$0 \leq p(x_i) \leq 1 \quad , \quad \sum_i p(x_i) = 1$$

Example: PMF for the coin toss example above



(ii) Probability density functions (pdf)

For a continuous RV, we instead have a probability density function (pdf):

$$P(a \leq X \leq b) = \int_a^b f(t)\, dt$$

A pdf must satisfy these properties:

$$f(x) \geq 0 \quad , \quad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

"Paradox" for pdfs: with continuous variables, the probability at $x=a$ is always zero:

$$P(X=a) = \int_a^a f(x)\, dx = 0$$

This is a little disconcerting. In reality, we are always interested in some range:

$$P(a \leq x \leq a+\varepsilon) = \int_a^{a+\varepsilon} f(x)\, dx \neq 0$$

(iii) Cumulative distribution function (cdf)

Another type of function is also useful for continuous RVs, called the cumulative distribution function (cdf) or sometimes just the "distribution" function:

$$P(X \leq x) := F(x)$$

The cdf obeys these properties:

$$F(-\infty)=0, \quad F(+\infty)=1$$

If $x_1 < x_2$ then $F(x_1) \le F(x_2)$ F is monotonic

$$P(a < x \le b) = F(b) - F(a)$$

Aside: there is no function called the "probability distribution function." This is ambiguous. I prefer to use this term for all of them. Instead, use pdf (density), cdf, or pmf where appropriate.

(iv) Relationships between the pmf, pdf, and cdf

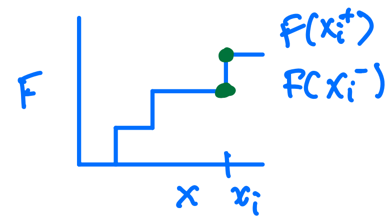
It is helpful to understand the relationships between these different functions.

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \qquad f(x) = \frac{dF}{dx} \qquad \text{continuous}$$

$$F(x) = \sum_{i}^{x_i \le x} p(x_i) \qquad p(x_i) = F(x_i^+) - F(x_i^-) \qquad \text{discrete}$$

$$f(x) = \sum_i p(x_i)\, \delta(x - x_i)$$

Dirac delta function



C. Examples of pmfs, pdfs, and cdfs

There are many examples of important distributions that arise in stat thermo and in the physical sciences more generally. See the accompanying Jupyter notebook for several plots and some discussion.

Discrete Examples (pmfs)

- Uniform
- Poisson
- Bernoulli
- Geometric
- Binomial

Continuous Examples (pdfs & cdfs)

- uniform
- exponential
- Gaussian / normal

D. The Expectation Operator and Moments

It is often inconvenient to use the entire pmf or pdf of a random variable. As such, we want a few numbers to tell us about the pmf/pdf. Additionally, these numbers, moments of the distributions, often correspond to physical observations as well.

(i) Expectation Operator

The expectation operator is motivated by the conventional idea of an average. It is defined as:

$$E[X] = \sum_i x_i \, p(x_i) \qquad \text{discrete}$$

$$E[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx \qquad \text{continuous}$$

$m = E[X]$ is the expectation or mean of X

There are a few useful properties to note about the expectation operator:

$$E[aX+bY] = aE[X] + bE[Y]$$ Linearity

$$E[g(X)] = \sum_i g(x_i)\,p(x_i)$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx$$

Law of the unconscious statistician (LOTUS)

↑ don't need to find the pmf/pdf for $Y = g(X)$

(ii) Moments

The mean is not the only piece of useful information that can be determined using the expectation operator. The n^{th} moment of the pmf/pdf is defined as:

$$E[X^n] = \sum_i x_i^n\,p(x_i) \quad \text{discrete}$$
$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)\,dx \quad \text{continuous}$$

notation: let $m_n = E[X^n]$

$m = m_1$

Also useful are the n^{th} centered moments:

$$E[(X-m)^n] = \sum_i (x_i - m)^n\,p(x_i) \quad \text{discrete}$$
$$E[(X-m)^n] = \int_{-\infty}^{\infty} (x-m)^n f(x)\,dx \quad \text{continuous}$$

The most common centered moments have names:

$n=2$: variance, $\sigma^2 = E[(X-m)^2] = E[X^2] - m^2$

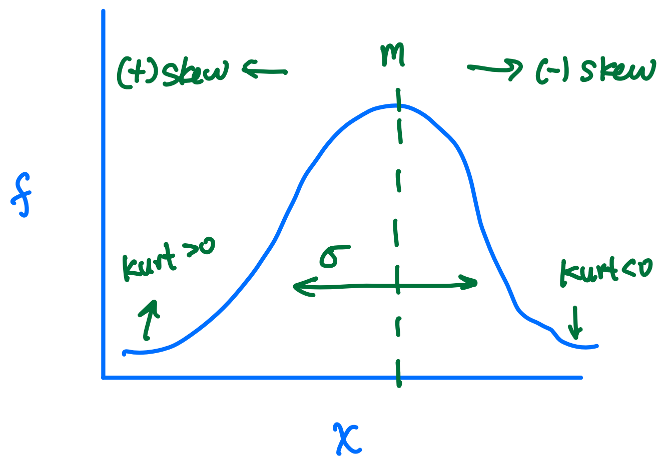
$n=3$: skewness, $E[(X-m)^3]/\sigma^3$

$n=4$: kurtosis, $E[(X-m)^4]/\sigma^4$

skewness & kurtosis are scaled, too

Recall also that $\sqrt{\sigma^2} = \sigma$ is the standard deviation.

Example: Intuition about what moments mean



m: center of mass

σ: width

skew: symmetry about m

kurt: "fat" or "thin" tails
kurt = 0 for normal

It is sometimes convenient to define a normalized R.V.

$$Y = \frac{X - m}{\sigma}$$

This has the nice properties that

$$E[Y] = 0, \quad E[Y^2] = 1, \quad \text{skew} = E[Y^3]$$
$$\text{kurt} = E[Y^4]$$

E. Characteristic Functions

Sometimes pmfs, pdfs, and cdfs are not very convenient for computing needed quantities. These include the calculation of moments and the calculation of pdfs of functions of RVs.

Characteristic functions (and related) are very useful tools for doing the above. They also provide some insight into moments and some fundamental theorems that make them worth learning. They are also very common in stat thermo.

We define the characteristic function of a random variable to be the Fourier transform of the pdf

$$\phi(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\, dx \,. \qquad i = \sqrt{-1}$$

Alternatively, $\phi(\omega)$ can be written using the expectation operator

$$\phi(\omega) = E[e^{i\omega x}]$$

The characteristic function has several important properties:

(1) It's inverse is $f(x)$.

There are different FT conventions. Note this one.

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega x} \phi(\omega)\, d\omega$$

The mapping between $f(x)$ and $\phi(\omega)$ is one-to-one, meaning $\phi(\omega)$ contains all the same information as $f(x)$.

(2) The derivatives of $\phi(\omega)$ are related to the moments.

$$\phi^{(n)}(\omega=0) = i^n E[x^n] \qquad \phi^{(n)}(\omega) = \frac{d^n\phi}{d\omega^n}$$

evaluate at $\omega = 0$ not centralized

Quick proof for 1st moment:

$$\frac{d\phi}{d\omega} = \frac{d}{d\omega} E[e^{i\omega x}] = E[ix e^{i\omega x}]$$

$$\left.\frac{d\phi}{d\omega}\right|_{\omega=0} = E[ix] = i\,E[x]$$

(3) The sum of independent RVs are products.

$$Z = X + Y$$ where X and Y are independent RVs "factorization property"

$$\phi_Z(\omega) = \phi_X(\omega)\,\phi_Y(\omega)$$ * see proof below

In real-space, this results in a convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\, dy$$ ← property of Fourier transforms of products

* Quick Proof:

$$\phi_Z(\omega) = E\left[e^{i\omega z}\right] = E\left[e^{i\omega(x+y)}\right]$$

$$= E\left[e^{i\omega x} e^{i\omega y}\right] = E\left[e^{i\omega x}\right] E\left[e^{i\omega y}\right]$$

independence

$$= \phi_X(\omega)\,\phi_Y(\omega)$$

(i) Other Transforms

The characteristic function is not the only transform that you will see. Let me briefly discuss a few others.

The moment generating function (mgf) is given by

$$M(s) = E\left[e^{sx}\right] = \int_{-\infty}^{\infty} e^{sx} f(x)\,dx$$ ← two-sided Laplace transform

- The characteristic function is a "Wick rotation" of the mgf:

$$\phi(\omega) = M(i\omega)$$

- Derivatives of the mgf are the moments.

$$M^{(n)}(s=0) = E[X^n] = m_n \qquad \leftarrow \text{no pesky } i$$

- The Taylor series of the mgf contains the moments.

$$M(s) = 1 + E[X]s + E[X^2]\frac{s^2}{2!} + E[X^3]\frac{s^3}{3!} + \dots$$
$$= 1 + m_1 s + m_2\frac{s^2}{2!} + m_3\frac{s^3}{3!} + \dots$$

If the mgf and the pdf are one-to-one, and the mgf is specified by the moments, then specifying all of the moments is equivalent to specifying the pdf. This is called the "moment problem". There are some important mathematical conditions that must be satisfied though for this to be true. (Carleman's condition.)
I don't understand

The cumulant generating function (cgf) is given by

$$K(s) = \ln E[e^{sx}]$$

- Derivatives of the cgf are called cumulants

$$\left.\frac{d^n K}{ds^n}\right|_{s=0} = K^{(n)}(0) = k_n$$

$$K(s) = k_1 s + k_2\frac{s^2}{2!} + k_3\frac{s^3}{3!} + \dots$$

cumulants have "decreasing informational content" at higher order.

k_1 is the mean, k_2 is the variance, and k_3 is the 3rd central moment. H.O.T are not in general central moments, but they are polynomial functions of them. cumulants have a special significance in statistical thermodynamics related to "connectedness" in n-body correlations. We'll hopefully talk about this soon.

The probability generating function (pgf) is the version for discrete RVs. It is defined as:

$$G(z) = E\left[z^x\right] = \sum_{n=0}^{\infty} z^n \, p(x_n)$$

- The pgf can be done using a "z-transform", a discrete analogue of the Laplace transform

$$\mathcal{Z}\left[f_n\right] = \sum_{n=0}^{\infty} z^{-n} f_n$$

Definition of the z-transform

- Derivatives of G give moments

$$\left.\frac{d^n G}{dz^n}\right|_{z=1} = G^{(n)}(1)$$

"n^{th} factorial moment"

$$= E\left[x(x-1)(x-2)\dots(x-n+1)\right].$$

- One can get probabilities from derivatives, too

$$\frac{1}{n!}\left.\frac{d^n G}{dz^n}\right|_{z=0} = \frac{G^{(n)}(0)}{n!} = p(x_n = n) = p(n)$$

<u>Example</u>: Characteristic function and mgf of a normal distribution

Normal distribution is given by

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

Use the definition of the characteristic function

$$\phi(\omega) = E\left[e^{i\omega x}\right] = \int_{-\infty}^{\infty} e^{i\omega x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx$$

change of variables: $y = \frac{x-m}{\sigma} \rightarrow x = y\sigma + m$, $dy = \frac{dx}{\sigma}$

$$\phi(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[i\omega(y\sigma+m)\right] \exp\left(-\frac{y^2}{2}\right) dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{i\omega m} \int_{-\infty}^{\infty} \exp(i\omega\sigma y) \exp\left(-y^2/2\right) dy$$

using a F.T. table (Mathematica)

$$\int_{-\infty}^{\infty} e^{i\omega\sigma y} \exp(-y^2/2)\, dy = \sqrt{2\pi} \exp\left(-\frac{\omega^2\sigma^2}{2}\right)$$

putting it together

$$\phi(\omega) = \frac{1}{\cancel{\sqrt{2\pi}}} e^{i\omega m} \cdot \cancel{\sqrt{2\pi}}\, e^{-\omega^2\sigma^2/2}$$

$$\boxed{\phi(\omega) = \exp\left(i\omega m - \omega^2\sigma^2/2\right)}$$

what is the mgf?

need to "pull out" an $i\omega \to s$ or $i\omega = s$
$\omega = -is$

$$M(s) = \exp\left[sm - (-is)^2\sigma^2/2\right]$$

$$= \exp\left(sm - (-s^2)\sigma^2/2\right) = \exp\left(sm + s^2\sigma^2/2\right)$$

$$\boxed{M(s) = \exp\left(sm + s^2\sigma^2/2\right)}$$

What are the moments?

$$\frac{dM}{ds} = \exp\left(sm + s^2\sigma^2/2\right)\cdot\left(m + s\sigma^2\right) = M\left(m + s\sigma^2\right)$$

$$\frac{d^2M}{ds^2} = M\left(m + s\sigma^2\right)^2 + M\sigma^2$$

$$\frac{d^3M}{ds^3} = M\left(m + s\sigma^2\right)^3 + 2M\left(m + s\sigma^2\right)\sigma^2 + M\left(m + s\sigma^2\right)\sigma^2$$

$$\frac{d^4M}{ds^4} = M\left(m + s\sigma^2\right)^4 + 3M\left(m + s\sigma^2\right)^2\sigma^2 + 2M\sigma^2\left(m + s\sigma^2\right)^2 + 2M\sigma^4 + M\left(m + s\sigma^2\right)^2\sigma^2 + M\sigma^4$$

$$m_1 = \left.\frac{dM}{ds}\right|_{s=0} = \cancel{M(0)}^{1}\cdot m = \boxed{m}$$

$$m_2 = \left.\frac{d^2M}{ds^2}\right|_{s=0} = \cancel{M(0)}^{1}\cdot m^2 + \cancel{M(0)}^{1}\sigma^2 = \boxed{m^2 + \sigma^2}$$

$$m_3 = \left.\frac{d^3M}{ds^3}\right|_{s=0} = M(0)\,m^3 + 2M(0)\,m\sigma^2 + M(0)\,m\cdot\sigma^2$$

$$= m^3 + 2m\sigma^2 + m\sigma^2 = \boxed{m^3 + 3m\sigma^2}$$

$$m_4 = \left.\frac{d^4M}{ds^4}\right|_{s=0} = m^4 + 3m^2\sigma^2 + 2m^2\sigma^2 + 2\sigma^4 + m^2\sigma^2 + \sigma^4$$

$$\boxed{m_4 = m^4 + 6m^2\sigma^2 + 3\sigma^4}$$

Note that these are polynomials of m and σ only

What is the cgf?

$$K(s) = \ln M(s) = \boxed{sm + s^2\sigma^2/2}$$

What are the cumulants?

$$\frac{dK}{ds} = m + s\sigma^2 \qquad k_1 = \left.\frac{dK}{ds}\right|_{s=0} = \boxed{m}$$

$$\frac{d^2K}{ds^2} = \sigma^2 \qquad k_2 = \left.\frac{d^2K}{ds^2}\right|_{s=0} = \boxed{\sigma^2}$$

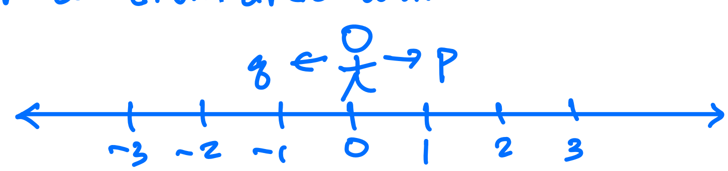
$$\frac{d^nK}{ds^n} = 0 \quad \text{for } n \geq 3 \qquad \boxed{k_3 = k_4 = k_n = 0 \ , n \geq 3}$$

F. Limiting Behavior

As we mentioned above in our discussion on Stirling's approximation, in stat thermo, we care about many, many random variables. There are two important theorems in probability that describe behavior in the limit of many independent RVs

Example: Random walk

consider a "drunkard's walk"



Each step is a "Bernoulli trial" p to the right (n_R) and q to the left (n_L).

$$p(m) = \frac{N!}{n_R!\, n_L!}\, p^{n_R} q^{n_L}$$

After $N = n_R + n_L$ steps the man is at what position $m = n_R - n_L$?

$$p(m) = \frac{N!}{\left[\frac{1}{2}(N+m)\right]!\left[\frac{1}{2}(N-m)\right]!}\, p^{1/2(N+m)} q^{1/2(N-m)}$$

What happens as $N \to \infty$?

$$\lim_{N\to\infty} p(m) = \left(\frac{2}{\pi N}\right)^{1/2} \exp\left(-\frac{m^2}{2N}\right)$$

Approaches a normal distribution with mean $= 0$ and variance $= N$

variability decreases (std dev $\sim \sqrt{N}$)

(i) Law of Large numbers

Suppose $X_1, X_2, \ldots, X_N$ are independent and identically distributed (iid) RVs of an unknown distribution with mean m and variance σ^2. The RV.

$$z = \frac{1}{N}(X_1 + X_2 + \ldots + X_N)$$

has a mean m and variance σ^2/N in the limit $N \to \infty$.

what does this mean? The mean (average) converges to what we would expect when N is very large. Also, the variance of the mean gets smaller and smaller (with rate $N^{-1/2}$). This is the "deterministic limit."

Proof:

$$Z = \frac{1}{N}(X_1 + X_2 + \cdots + X_N) = \frac{X_1}{N} + \frac{X_2}{N} + \cdots \frac{X_N}{N}$$

$M_1(s) = E[e^{sX_1/N}]$ ← moment generating functions of X_i/N

$M_2(s) = E[e^{sX_2/N}]$

$\vdots$

$M_N(s) = E[e^{sX_N/N}]$

By the factorization property:

$$M_Z(s) = M_1(s)\, M_2(s) \cdots M_N(s) = \prod_{i=1}^{N} M_i(s)$$

Expand $M_i(s)$ using the Taylor series

$$E[e^{sX_i}] = 1 + ms + \frac{\sigma^2 s^2}{2} + \theta(s^3)$$

$$E[e^{s/N \cdot X_i}] = M_i(s) = 1 + \frac{ms}{N} + \frac{\sigma^2 s^2}{2N^2} + \theta(s^3/N^3)$$

Now put both expressions together

$$M_Z(s) = \left[1 + \frac{ms}{N} + \frac{\sigma^2 s^2}{2N^2} + \theta(s^3/N^3)\right]^N$$

Recall that

$$\lim_{N\to\infty} \left[1 + \frac{z}{N}\right]^N = \exp(z)$$

Therefore

$$\lim_{N\to\infty} M_z(s) = \lim_{N\to\infty}\left[1+\frac{1}{N}\underbrace{\left(ms+\frac{\sigma^2 s^2}{2N}+O(s^3/N^2)\right)}_{z}\right]^N$$

$$\lim_{N\to\infty} M_z(s) = \exp\left(ms+\frac{\sigma^2 s^2}{2N}+\dots\right)$$

The cgf is

$$Z(s) = ms + \frac{\sigma^2 s^2}{2N}$$

$$\text{mean} = \left.\frac{dZ}{ds}\right|_{s=0} = m + \left.\frac{\sigma^2 s}{N}\right|_{s=0} = m$$

$$\text{var} = \left.\frac{d^2 Z}{ds^2}\right|_{s=0} = \frac{\sigma^2}{N}$$

$$\boxed{\text{mean} = m, \quad \text{var} = \sigma^2/N}$$

<u>(ii) Central limit theorem</u> (Lindeberg–Lévy theorem)

Suppose that $X_1, X_2, \dots, X_N$ are IID RVs with mean m and variance σ^2. If Y is defined as

$$Y = \sum_{i=1}^{N}\left(\frac{X_i - m}{\sigma\sqrt{N}}\right)$$

then the pdf of Y converges to a standard normal distribution in the limit that $N\to\infty$.

mean $=0$
var $=1$
$N(0,1)$

What does this mean? Sums of normalized independent variables <u>converge to the normal distribution</u> for large N!

Proof:

$$Y = \frac{x_1 - m}{\sigma\sqrt{N}} + \frac{x_2 - m}{\sigma\sqrt{N}} + \dots + \frac{x_N - m}{\sigma\sqrt{N}}$$

Let $z_i = \frac{x_i - m}{\sigma}$

$$Y = \frac{z_1}{\sqrt{N}} + \frac{z_2}{\sqrt{N}} + \dots + \frac{z_N}{\sqrt{N}}$$

We want to find the mgf of Y:

$$M_Y(s) = E[\exp(Ys)]$$

$$= E\left[\exp\left(\frac{z_1 s}{\sqrt{N}} + \frac{z_2 s}{\sqrt{N}} + \dots + \frac{z_N s}{\sqrt{N}}\right)\right]$$

$$= E\left[\exp\left(\frac{z_1 s}{\sqrt{N}}\right)\exp\left(\frac{z_2 s}{\sqrt{N}}\right)\cdots\exp\left(\frac{z_N s}{\sqrt{N}}\right)\right]$$

The z_i's are independent (factorization)

$$= E\left[\exp\left(\frac{z_1 s}{\sqrt{N}}\right)\right]E\left[\exp\left(\frac{z_2 s}{\sqrt{N}}\right)\right]\cdots E\left[\exp\left(\frac{z_N s}{\sqrt{N}}\right)\right]$$

$$= \prod_{i=1}^{N} \boxed{E\left[\exp\left(\frac{z_i s}{\sqrt{N}}\right)\right]}$$

Expand $\square$ as a taylor series

$$E[e^{sx}] = 1 + ms + \frac{s^2\sigma^2}{2} + O(s^3)$$

scale $s \to s/\sqrt{N}$

$m=0$, $\sigma^2=1$ for z_i

$$E\left[\exp\left(\frac{z_i s}{\sqrt{N}}\right)\right] = 1 + 0\cdot\frac{s}{\sqrt{N}} + \frac{s^2\cdot 1}{2N} + O\left(s^3/N^{3/2}\right)$$

$$= 1 + \frac{s^2}{2N} + O\left(s^3/N^{3/2}\right)$$

Now plug back into $M_Y(s)$:

$$M_Y(s) = \prod_{i=1}^{N}\left[1 + \frac{s^2}{2N} + O\left(s^3/N^{3/2}\right)\right] = \left[1 + \frac{s^2}{N} + O\left(\frac{s^3}{N^{3/2}}\right)\right]^N$$

Now, recall that

$$\lim_{N\to\infty}\left[1+\frac{x}{N}\right]^N = \exp(x)$$

Therefore

$$\lim_{N\to\infty}\left[1+\frac{s^2}{2N}\right]^N = \exp(s^2/2)$$

$$\boxed{\lim_{N\to\infty} M_Y(s) = \exp(s^2/2)}$$

The mgf for a normal distribution is

$$M(s) = \exp(ms + \sigma^2 s^2/2)$$

By comparison, the pdf is normal with mean = 0 and variance = 1.

III. Multivariate Random Variables

A. Bivariate Random Variables

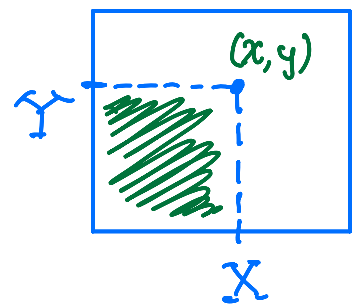
In statistical thermodynamics, we usually need to deal with functions of more than one variable. In fact, we will deal with functions of <u>many</u> variables. How do we deal with this in terms of random variables?

(i) Joint and Marginal Distributions

To describe probabilities with more than one R.V., we need a <u>joint distribution function.</u> In the bivariate case (i.e. x & y) the joint distribution describes how the probability depends on x and y and all interactions between x and y. In other words, we will want a distribution to describe $P(\mathcal{X} \cap \mathcal{Y})$ for all $\mathcal{X} \in \mathbb{R}$ and $\mathcal{Y} \in \mathbb{R}$. This will necessarily include all $P(\mathcal{X})$ and $P(\mathcal{Y})$.

The joint cumulative distribution function (cdf) is defined as

$$P(X \le x, Y \le y) = F_{xy}(x,y)$$



The joint distribution contains all the probability information for the "2D plane" of both x and y.

The marginal cumulative distribution function contains only the information for one of the variables. It can be obtained from the joint cdf:

$$F_x = F_{xy}(x, \infty)$$

$$F_y = F_{xy}(\infty, y)$$

Recall that $F(Y \le y) \to 1$ as $y \to \infty$

Similarly, we can define joint and marginal probability density functions

joint pdf

$$P((x,y) \in A) = \iint_A f_{xy}(x,y)\, dx\, dy$$

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x,y)\, dy$$

Recall that $\int_{-\infty}^{\infty} f_y(y)\, dy = 1$

We say we "integrate out" y.

$$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x,y)\, dx .$$

The definition of the joint cdf and pdf implies that

$$\frac{\partial^2 F_{xy}}{\partial x \partial y} = \frac{\partial^2 F_{xy}}{\partial y \partial x} = f_{xy}$$ joint cdf → joint pdf

and

$$\frac{\partial}{\partial x} F_{xy}(x,\infty) = \frac{\partial}{\partial x} F_x = f_x$$ marginal cdf → marginal pdf

$$\frac{\partial}{\partial y} F_{xy}(\infty,y) = \frac{\partial}{\partial y} F_y = f_y$$

Finally, for discrete RVs, we have joint and marginal pmfs

$$P(X=x_i, Y=y_j) = p_{xy}(x_i,y_j)$$ joint pmf

$$p_x(x_i) = \sum_{j=1}^{\infty} p_{xy}(x_i,y_j)$$ marginal pmf

sum over all y.

$$p_y(y_j) = \sum_{i=1}^{\infty} p_{xy}(x_i,y_j)$$

(ii) Example: Bivariate Gaussian (Normal) Distribution

$$f_{xy}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}\left(\hat{x}^2 - 2\rho\hat{x}\hat{y} + \hat{y}^2\right)\right]$$

$$\hat{x} = \frac{x-m_x}{\sigma_x} \qquad \hat{y} = \frac{y-m_y}{\sigma_y}$$

parameters: $m_x, m_y, \sigma_x, \sigma_y, \rho$: $|\rho|<1$

See the accompanying plots of the 2D normal distribution.

(iii) Moments: Correlation and covariance

The expectation operator for 2D is now a double integral

$$E[x] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \, f_{xy}(x,y)\, dx\, dy$$

$$E[y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \, f_{xy}(x,y)\, dx\, dy$$

We can also generalize LOTUS

$$E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{xy}(x,y)\, dx\, dy$$

The 2^{nd} moment is more interesting. Now, there are three different ones:

$$\underbrace{E[x^2],\ E[y^2]}_{2^{nd}\text{ moment}},\ \underbrace{E[xy]}_{\text{correlation of } x \,\&\, y}$$

Centered moments are also useful:

$$\left.\begin{aligned} \mathrm{var}(x) &= E[(x-m_x)^2] \\ \mathrm{var}(y) &= E[(y-m_y)^2] \end{aligned}\right\} \text{variance}$$

$$\mathrm{cov}(x,y) = E[(x-m_x)(y-m_y)] \quad \text{covariance}$$

This is also a correlation coefficient:

$$\rho_{xy} = E\left[\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right)\right] = \frac{\mathrm{cov}(x,y)}{[\mathrm{var}(x)\,\mathrm{var}(y)]^{1/2}}$$

Uncorrelated, therefore, means $\rho_{xy} = 0$ or $cov(x,y) = 0$

Aside: Mutually exclusive, independence, and correlation

Mutually exclusive implies $P(X \cap Y) = 0$. So, the joint cdf/pdf will give $P=0$ for those values x and y are mutually exclusive.

Independence implies $f_{xy}(x,y) = f_x(x) f_y(y)$. In this case the variables will always be uncorrelated. see below

Proof :
$$cov(x,y) = E[(x-m_x)(y-m_y)]$$
$$= E[x-m_x]\, E[y-m_y]$$
$$= (E[x]-m_x)(E[y]-m_y) = 0$$

Uncorrelated does not imply independent. The covariance could be zero for other reasons.

(iv) Characteristic Function

The joint pdf has a characteristic function, just like the 1D version. Here, we need a bivariate Fourier transform.

$$\phi_{xy}(\omega,v) = E\left[e^{i\omega x + ivy} \right]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\, e^{i\omega x + ivy}\, dx\,dy$$

If x and y are independent, then

$$\phi_{xy}(\omega,v) = \phi_x(\omega)\, \phi_y(v)$$

We're going to talk more about independence, etc. right now.

(v) Conditional Probability

The conditional probability density is defined as

$$f_{y|x}(y|x) = \frac{f_{xy}(x,y)}{f_x(x)}$$

← joint density

← marginal density

conditional density of y given x

$f_X(x)$ must also be non-zero

The definition above implies that

$$f_{xy}(x,y) = f_{y|x}(y|x)\, f_x(x)$$

can get joint density from conditional. Used something similar for Bayes theorem.

and a "law of total probability"

$$f_y(y) = \int_{-\infty}^{\infty} f(y|x)\, f_x(x)\, dx.$$

← can "integrate out" x

Independence means that

$$f_{y|x}(y|x) = f_y(y)$$

← f_y doesn't depend on x "drop the conditioning"

Using the above expression for the joint density gives

$$f_{xy}(x,y) = f_{y|x}(y|x)\, f_x(x) = f_y(y)\, f_x(x)$$

One can also define a conditional cdf

$$F_{y|x} = \int_{-\infty}^{y} f_{y|x}(t|x)\, dt,$$

which is also equal to

$$F_{y|x} = \frac{\partial F_{xy}}{\partial y}.$$

← how does the joint distribution change with y at fixed x

Lastly, we want to be able to compute expectations. The definition of the conditional expectation is

$$E[y|x] = \int_{-\infty}^{\infty} y\, f_{y|x}(y|x)\,dy \qquad \text{conditional mean}$$

$$E[g(y)|x] = \int_{-\infty}^{\infty} g(y)\, f_{y|x}(y|x)\,dy \qquad \text{conditional LOTUS}$$

Note: I did not put pmf formulas here. Replace $f \to p$ and integrals with sums. Also, one can swap x and y to get conditional probabilities of x at a given y.

B. Random Vectors

(i) A very quick review of matrix algebra / calculus

A vector is defined as components x_i with unit directions $\underline{e}_i$,

$$\underline{x} = \sum_i x_i \underline{e}_i = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_1, x_2, \dots, x_n]^T$$

transpose

"column vector"

A tensor is defined similarly, except its components are indexed in two dimensions A_{ij} and the directions are a unit dyad $e_i e_j$

$$\underline{\underline{A}} = \sum_i \sum_j A_{ij} \underline{e}_i \underline{e}_j = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \\ A_{n1} & & & A_{nn} \end{bmatrix}$$

A (2nd rank) tensor is a matrix with directions.

There are numerous algebraic and calculus operations performed on vectors and tensors. We only need to review a few. For these examples, lets consider

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \underline{\underline{A}} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \underline{\underline{B}} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Transpose:

$$(A_{ij})^T = A_{ji} \qquad \underline{\underline{A}}^T = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}$$

← swap rows and columns

sums:

$$\underline{x} + \underline{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

element wise addition
same for matrix/tensor

inner prodcut:

$$\underline{\underline{A}} \cdot \underline{\underline{B}} = \sum_{k=1}^{n} A_{ik} B_{kj} = \begin{bmatrix} A_{11} B_{11} + A_{12} B_{21} & A_{11} B_{12} + A_{12} B_{22} \\ A_{21} B_{11} + A_{22} B_{21} & A_{21} B_{12} + A_{22} B_{22} \end{bmatrix}$$

"matrix multiplication"

$$\underline{x} \cdot \underline{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2$$

"dot product"

Trace:

$$Tr(A) = \sum_{i=1}^{n} A_{ii} = A_{11} + A_{22} \qquad \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

sum of the diagonal elements

Norm:

$$\| \underline{x} \| = (\underline{x} \cdot \underline{x})^{1/2} = (x_1^2 + x_2^2)^{1/2}$$

"distance formula"

(ii) Distributions

Just as we saw with the case with two variables, we can define mutivariate joint cdfs, pdfs, and pmfs.

Joint cdf:

$$F(x_1, x_2, \dots, x_n) = F(\underline{x}) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$$

Joint pdf:

$$P(\underline{x} \in V) = \underset{V}{\iint \cdots \int} f(x_1, x_2, \dots x_n)\, dx_1\, dx_2 \dots dx_n$$

← some "hypervolume" V

$$= \int_V f(\underline{x})\, d\underline{x}$$

note that as before: $f(\underline{x}) = \dfrac{\partial^n F}{\partial x_1 \partial x_2 \dots \partial x_n}$

Joint pmf:

$$P(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = p(x_1, x_2, \dots x_n) = p(\underline{x})$$

Now, if we want to eliminate one (or more) variables from the distribution, the process is called marginalization.

cdf marginalization:

$$F(x_1, x_2, \dots, x_{n-1}) = F(x_1, x_2, \dots, x_{n-1}, \infty)$$

pdf marginalization:

$$f(x_1, x_2, \dots, x_{n-1}) = \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_{n-1}, x_n)\, dx_n$$

pmf marginalization:

$$p(x_1, x_2, \dots, x_{n-1}) = \sum_{i=1}^{\infty} p(x_1, x_2, \dots, x_{n-1}, x_n^i)$$

Finally, we can also have multivariate conditional distributions

conditional pdf:

$$f(x_1, x_2, \dots, x_n) = f(x_n \mid x_1, x_2, \dots, x_{n-1})\, f(x_1, x_2, \dots, x_{n-1})$$

we can further decompose the conditional pdf into successive conditional probabilities. This is called the chain rule.

not related to calculus

$$f(x_1, x_2, \dots, x_n) = f(x_1) \prod_{i=2}^{n} f(x_i \mid x_1, x_2, \dots, x_{i-1})$$

Example with $n=3$:

$$f(x_1, x_2, x_3) = f(x_3 \mid x_1, x_2)\, f(x_1, x_2)$$

definition of conditional

$$= f(x_3 \mid x_1, x_2)\, \underbrace{f(x_2 \mid x_1)\, f(x_1)}$$

now do these two

$$= f(x_1)\, f(x_2 \mid x_1)\, f(x_3 \mid x_1, x_2)$$

re-arrange like formula.

Independence means:

$$f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n)$$

This is a pretty big assumption in multivariate cases. Sometimes less drastic assumptions such as pairwise independence are used.

Example: Full vs. Pairwise Independence for $n=3$

$$\text{Full: } f(x_1, x_2, x_3) = f(x_1) f(x_2) f(x_3)$$

$$\begin{aligned}\text{Pairwise: } f(x_1, x_2, x_3) &= f(x_3 \mid x_1, x_2) f(x_1, x_2) \\ &= f(x_3 \mid x_1, x_2) \underbrace{f(x_2 \mid x_1)}_{f(x_2 \mid x_1) = f(x_2)} f(x_1) \\ &= f(x_1) f(x_2) \underbrace{f(x_3 \mid x_1, x_2)}_{\text{This one stays}}\end{aligned}$$

(iii) Expectation and moments

The expectation operator is a straightforward generalization of the bivariate expectation.

$$E[g(\underline{x})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\underline{x}) f(\underline{x}) \, d\underline{x}$$

$dx_1 \, dx_2 \ldots dx_n$

mean:

$$E[\underline{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \underline{m}$$

mean vector

2^{nd} moments give a tensor/matrix called the correlation matrix.

$$E[\underline{x}\,\underline{x}] = \begin{bmatrix} E[x_1 x_1] & E[x_1 x_2] & \cdots & E[x_1 x_n] \\ E[x_2 x_1] & E[x_2 x_2] & \cdots & E[x_2 x_n] \\ \vdots & & \ddots & \vdots \\ E[x_n x_1] & E[x_n x_2] & \cdots & E[x_n x_n] \end{bmatrix} = \underline{\underline{R}}$$

Centered second moments give a covariance matrix.

$$E[(\underline{x} - \underline{m})(\underline{x} - \underline{m})] = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) & \cdots & cov(x_1, x_n) \\ cov(x_2, x_1) & cov(x_2, x_2) & \cdots & cov(x_2, x_n) \\ \vdots & & \ddots & \vdots \\ cov(x_n, x_1) & cov(x_n, x_2) & \cdots & cov(x_n, x_n) \end{bmatrix}$$

$$= \underline{\underline{C}}$$

- Both $\underline{\underline{R}}$ and $\underline{\underline{C}}$ are symmetric matrices.
- Diagonal elements of $\underline{\underline{C}}$ are the variance
- If $c_{ij} = 0$ for $i \neq j$, then x_i and x_j are uncorrelated.
  - if $\underline{\underline{C}}$ is diagonal, all x_i and x_j are uncorrelated.

One can also compute a cross-covariance matrix from two different random vectors

$$C_{xy} = E[(\underline{x} - \underline{m}_x)(\underline{y} - \underline{m}_y)]$$

Finally, one can calculate the characteristic function using an n-dimensional Fourier transform.

$$\phi(\underline{\omega}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i\underline{\omega}\cdot\underline{x}} f(\underline{x})\, d\underline{x} \qquad \underline{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}$$

$$f(\underline{x}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-i\underline{\omega}\cdot\underline{x}} \, \phi(\underline{\omega}) \, d\underline{\omega}$$

The mean and components of the correlation matrix can be found using derivatives of $\phi(\underline{\omega})$.

$$\frac{\partial}{\partial \omega_i} \phi(\underline{\omega}=0) = i \, E[\omega_i]$$

$$\frac{\partial^2}{\partial \omega_i \partial \omega_j} \phi(\underline{\omega}=0) = -E[\omega_i \omega_j]$$

(iv) n-Component Normal Distribution

The multivariate normal pdf is given by

$$f(\underline{x}) = \frac{1}{[(2\pi)^n \det \underline{\underline{C}}]^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{m})\cdot\underline{\underline{C}}^{-1}\cdot(\underline{x}-\underline{m})\right]$$

inverse of covariance matrix

The characteristic function is given by:

$$\phi(\underline{\omega}) = \exp\left(i\underline{x}\cdot\underline{m} - \frac{1}{2}\,\underline{x}\cdot\underline{\underline{C}}\cdot\underline{x}\right)$$

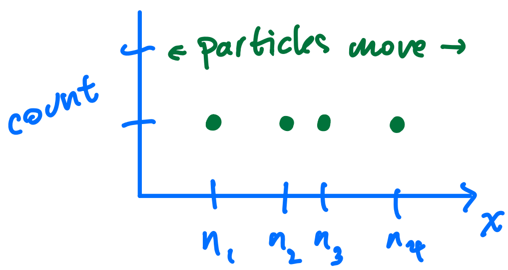
A few useful facts:

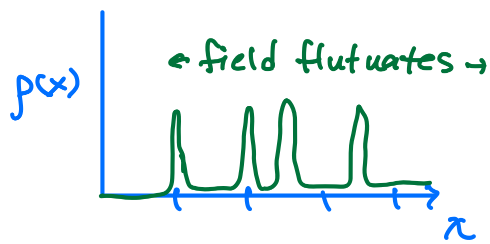
- For a Gaussian random vector, uncorrelated = independent (because $\underline{\underline{C}}$ is diagonal)
- Level sets of the density are ellipsoids centered at m.

C. Random Fields

We will see in statistical thermo that we have many random variables, and sometimes, especially if they are indistinguishable, it makes more sense to think of a random field, rather than discrete molecules.

Example:





(i) vector – function analogy

An infinite-dimensional vector is a function.

$$\lim_{N\to\infty} [f_1, f_2, \ldots f_N]^T = f(x)$$

I don't want to go into too many formal details. Think about programming for a minute if you need an example.

Similarly, an infinite-dimensional matrix is a linear operator (e.g. a differential operator).

$$\lim_{N\to\infty} \underline{\underline{L}} \cdot \underline{f} = \mathcal{L} f(x) = g(x)$$

example:

$$\lim_{N\to\infty} \frac{1}{\Delta x} \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & -1 \\ -1 & 1 & 0 & 0 & \cdots & \cdots \\ 0 & -1 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & -1 & 1 & \cdots & \\ & & & & \ddots & \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} = \frac{df}{dx}$$

Functions are vectors in a Hilbert vector space, and derivatives are linear operators on that space.

A linear operator maps a function to another one.

Just like we do vector algebra for finite dimensional vectors, we can do it for functions too.

inner product:

$$\underline{x} \cdot \underline{y} = \sum_i x_i y_i \qquad (f,g) = \int_{-\infty}^{\infty} f(x)\, g(x)\, dx$$

functional:

A vector function maps a vector to a number:

$\underline{x} \mapsto f(\underline{x})$

A functional maps a function to a number:

$f(x) \mapsto F[f]$

Example: a definite integral

$$F[f] = \int_0^1 f(x)\, dx$$

if $f(x) = x$, $F[f] = \int_0^1 x\, dx = \frac{x^2}{2}\Big|_0^1 = 1/2$

if $f(x) = x^2$, $F[f] = \int_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = 1/3$

Just like we do multivariable calculus, we can also do calculus with functionals.

functional derivative:

$$dF = \sum_{i=1}^{N} \frac{\partial F}{\partial f_i} df_i \;\; \xRightarrow{\lim_{N \to \infty}} \;\; dF = \int \frac{\delta F}{\delta f} \overbrace{\phi(x)\, dx}^{\delta f}$$

partial derivative

functional derivative.

How does the number $F[f]$ change as we vary $f(x)$?

functional integral:

$$\lim_{N\to\infty} \iint \cdots \int F(f_1, f_2, \ldots f_N)\, df_1\, df_2 \ldots df_N = \int Df\; F[f]$$

Functional Integral.

Add up all the numbers $F[f]$ over all the different functions $f(x)$ that could contribute.

(ii) Distributions

Analogous to the definition for random vectors, we can define a probability density functional, $P[f]$,

$$\int Df\; P[f] = 1 \;.$$

$P[f]$ is the joint pdf.

space of all functions

pdf

Usually, like we've seen many times, $P[f]$ has an exponential

$$P[f] = \frac{1}{Z}\, e^{-S[f]}$$

S is called the "action" functional

$$Z = \int Df\; e^{-S[f]}$$

Z is the normalization constant. (keeps $\int Df\, P[f] = 1$)

It is often difficult to compute Z, but it usually cancels out, so we don't always need to do it.

It is hard to generalize a cdf or a pmf in this context. Like we saw with random vectors, we can marginalize this probability density functional by integrating out some degrees of freedom.

let $f(x) = g(x) + h(x)$ split into two functions, e.g. long vs. short wavelength modes.

$$P_m[h] = \int Dg\; P[f] \leftarrow \text{joint}$$

marginal "integrate out" modes in $g(x)$

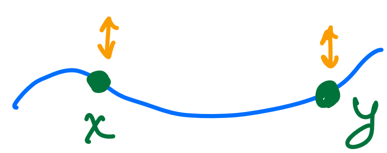
This has application in coarse-graining and in a process called "renormalization."

(ii) Expectation

The expectation operator is defined in a way very analogous to random vectors

$$E[\Theta[f]] = \int Df\; \Theta[f]\; P[f]$$
$$= \frac{1}{Z} \int Df\; \Theta[f]\; e^{-S[f]}$$

- $E[f]$ is the mean value of the field. Often $\bar{f}$ or $\langle f \rangle$
- $E[f^2(x)]$ is the fluctuation of the field.
- $E[f(x) f(y)]$ is the correlation between different points in space.



- $E[f(x) g(x)]$ is the correlation between two different fields.

(iii) Example: The Gaussian probability density functional

$$P[f] = \frac{1}{Z} \exp\left[-\frac{1}{2} \int dx \int dy\; \hat{f}(x)\, G^{-1}(x,y)\, \hat{f}(y) \right]$$

Action: $S[f] = \frac{1}{2} \int dx \int dy\; \hat{f}(x)\, G^{-1}(x,y)\, \hat{f}(y)$

two inner products

Normalization:

$$Z = \int Df \, \exp\left[-\tfrac{1}{2}\int dx \int dy \, \hat{f}(x)\, G^{-1}(x,y)\, \hat{f}(y)\right]$$

$$\hat{f}(x) = f(x) - E[f(x)] \qquad \hat{f}(y) = f(y) - E[f(y)]$$

$G(x,y) = E[\hat{f}(x)\hat{f}(y)]$, covariance function

$G^{-1}(x,y)$: inverse covariance or "propagator."
This is usually a differential operator

Example: $G(x,y) = \delta(x-y)$ "white noise field"
Each point is independent

In statistical mechanics $G(x,y)$ is related to the structure factor, which can be measured by scattering experiments.

(iv.) Characteristic Functional

The characteristic functional is one of the most useful objects when working with random fields. This is because we cannot usually compute functional integrals, but we can calculate functional derivatives to get moments.

The characteristic functional is defined as

$$\phi[J] = E\left[e^{i\int dx\, J(x) f(x)}\right]$$ ← inner product.

$$= \int Df \, e^{i\int dx\, J(x) f(x)} \, P[f]$$

$J(x)$ is the new function, analogous to the Fourier variable in finite dimensions.

Example : Characteristic Functional of a Gaussian field

$$\phi[J] = E\left[e^{i\int dx\, J(x) f(x)}\right] \qquad \text{As "simple" as it gets!}$$

$$= \exp\left[-\frac{1}{2}\int dx \int dy\; J(x)\, G(x,y)\, J(y)\right]$$

Functional derivatives of $\phi[J]$ give moments of the pdf.

$$E\left[f(x_1) f(x_2) \cdots f(x_n)\right] = (-i)^n \left.\frac{\delta^n \phi[J]}{\delta J(x_1)\, \delta J(x_2) \cdots \delta J(x_n)}\right|_{J=0}$$

Example :

$$G(x,y) = -\left.\frac{\delta^2 \phi[J]}{\delta J(x_1) \delta J(x_2)}\right|_{J=0}$$

(v.) Final comments

- One can define a conditional probability density

$$P[f \mid g] = \frac{P[f,g]}{P[g]}$$

- We have been cavalier about boundary conditions for the fields. The boundary conditions can be. ← e.g. changes bounds of integrals.
  1. $f(x) \to 0$ as $x \to \pm\infty$ (fast enough)
  2. Bounded to a finite domain: Dirichlet, Neumann or Robin condition on the boundary.
  3. Periodic boundaries
  4. Asymptotic matching (matches non-zero at $\pm\infty$)