There will be 8 lectures by Prof. Baiesi - the first two about some mathematical tools, and then about scattered arguments, to show how to apply the theoretical physics framework to various topics.

(Lesson 3 of 10/10/19) Compiled: 12 ottobre 2019

0.1 Moments and Generating Functions

Consider a continuous function $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto f(x)$. The n-th **moment** of f about a point $c \in \mathbb{R}$ is defined as the integral:

$$\mu_n = \int_{-\infty}^{\infty} (x-c)^n f(x) \, \mathrm{d}x$$

Moments provide a way to quantify, in a certain sense, the *shape* of f. For example, if $f(x)$ is a linear density ($[\mathrm{kg}\,\mathrm{m}^{-1}]$), then the 0-th moment is the total mass, the first one (with $c = 0$) is the center of mass, and the second is the *moment of inertia*.

Moments are especially useful if $f(x)$ is a probability density function (pdf), i.e. a non-negative normalized function. In this case the first moment about 0 is the **mean**:

$$\mu_1 \equiv \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x = \mathrm{E}[X] \equiv \mu; \qquad X \sim f$$

where X is a random variable sampled from f. Note that, if not specified, a moment is intended to be centered around $c = 0$ (it is a *raw moment* or *crude moment*).

The *central second moment*, that is μ_2 with $c = \mu$ is the **variance**:

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x) \, \mathrm{d}x \equiv \mathrm{E}[(X-\mu)^2] = \mathrm{Var}[X]$$

A **moment-generating function** of a real-valued random variable is a certain function $f\colon \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{x} \mapsto f(\boldsymbol{x})$ that can be used to *compute* the moment of the distribution where X comes from.
More precisely, for a random variable X, the moment-generating function M_X is defined as:

$$M_X(t) \equiv \mathrm{E}[e^{tX}], \quad t \in \mathbb{R}$$

In fact, recall that:

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \dots$$

Hence, as the *expected value* is a linear operator:

$$\begin{aligned} M_X(t) = \mathrm{E}[e^{tX}] &= 1 + t\,\mathrm{E}[X] + \frac{t^2\,\mathrm{E}[X^2]}{2!} + \dots = \\ &= 1 + t\mu_1 + \frac{t^2 \mu_2}{2!} + \dots \end{aligned}$$

Note that the distribution's moments are the coefficients of the power series that defines $M_X(t)$.

In fact, the more general definition of a **generating function** is that of a power-series with "hand-picked" coefficients a_n, such that by simply knowing the function one can compute a_n in an iterative way.

To recover a certain μ_n we start by differentiating M_X n times with respect to t, such that the first $n-1$ terms vanish:

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n} M_X(t) = \underbrace{\frac{n(n-1)\dots 1}{n!}}_{=1} \mu_n + \frac{(n+1)n \dots 2}{(n+1)!} t \mu_{n+1} + \dots$$

Then, by setting $t = 0$, all μ_r with $r > n$ vanish, leaving only the desired μ_n:

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n} M_X(t)\Big|_{t=0} = \mu_n$$

Finally, we note that a moment-generating function can be constructed even for a multi-dimensional vector $\boldsymbol{X} = (X_1, \dots, X_n)^T$ of random variables, by simply taking a scalar product in the exponential:

$$M_{\boldsymbol{X}}(\boldsymbol{t}) \equiv \mathrm{E}\left(e^{\boldsymbol{t}^T \boldsymbol{X}}\right) \qquad \boldsymbol{t} \in \mathbb{R}^n$$

0.2 Gaussian Integrals and Moments

Consider now a normal pdf in $d = 1$:

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

If we want to know the moment generating function, we have to compute the value of the integral:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x; \mu, \sigma)\, \mathrm{d}x$$

Let's start from the easiest case, and work our way out to the most general one. Recall that:

$$\int_{-\infty}^{\infty} \exp\left(-\frac{a}{2}x^2\right) \mathrm{dx} = \sqrt{\frac{2}{a}\pi}$$

We will show now how to generalize this result to many dimensions.
Let $\vec{x}$ be a vector of n dimension: $\vec{x} = (x_i)_{i=1,\dots,n}$. Suppose that x_i are random variables, that can be correlated - that is $p(x_i|x_j) \neq p(x_i)p(x_j)$. All the information about that is contained in the *covariance matrix*, $A_{ii} = \mathrm{var}(x_i)$ and $A_{ij} = A_{ji} =$

$\text{cov}(x_i, x_j)$.
The covariance matrix induces a metric:

$$\mathbb{A}(\vec{x}) = \sum_{i,j=1}^{n} x_i A_{ij} x_j$$

We will prove that:

$$Z(\mathbb{A}) = \int d^n x \exp\left(-\frac{1}{2}\mathbb{A}(\vec{x})\right) = (2\pi)^{n/2} \prod_{i=1}^{n} a_i^{-1/2} = (2\pi)^{n/2} (\det A)^{-1/2}$$

where a_i are the eigenvalues of A.

Proof. We can simply apply a coordinate transform through an orthogonal matrix O, with $|\det O| = 1$:

$$x'_i = \sum_j O_{ij} x_j$$

So that A becomes a diagonal matrix with eigenvalues a_i on the diagonal. Then $Z(\mathbb{A})$ becomes the multi-dimensional integral of independent gaussians, and we can integrate wrt to each variable separately, arriving at the final result.

Let's now consider a more general case:

$$Z(A, \vec{b}) = \int_{-\infty}^{\infty} d^n x \exp\left(-\frac{1}{2}\mathbb{A}(\vec{x}) + \vec{b}\cdot\vec{x}\right)$$

where $\vec{b}\cdot\vec{x} = \sum_i x_i b_i$.
We start by finding the minimum of the exponent, by differentiation:

$$\frac{\partial}{\partial x_i}\left(\frac{1}{2}\mathbb{A}(\vec{x}) - \vec{b}\cdot\vec{x}\right) = 0 \quad \forall i$$

The solution for the minimum is:

$$\sum_j A_{ij} x_j = b_i \Rightarrow x_i^* = \sum_j (A^{-1})_{ij} b_i$$

Then we apply a coordinate change:

$$\vec{x} \to \vec{y}; \vec{x} = \vec{x}^* + \vec{y}$$

so that:

$$-\frac{\mathbb{A}(\vec{x})}{2} + \vec{b}\cdot\vec{x} = -\frac{\mathbb{A}(\vec{y})}{2} + w_2(\vec{b})$$

with:

$$w_2(\vec{b}) = \frac{1}{2}\sum_{i,j=1}^{n} b_i (A^{-1})_{ij} b_i = \frac{1}{2}\vec{b}\cdot\vec{x}^*$$

And so:

$$Z(A, \vec{b}) = e^{w_x(\vec{b})} \int d^n y \exp\left(-\frac{1}{2}\mathbb{A}(\vec{y})\right) = e^{w_2(\vec{b})}(2\pi)^{n/2}(\det A)^{-1/2}$$

Exercise Compute $Z(A)$ and $Z(A, \vec{b})$ with:

$$A = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}; \quad b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(Solve the exercise, and put in a notebook to bring to the exam!)

0.2.1 Gaussian expectation values

We want to compute the average value of the *product* of multiple random variables, following a gaussian distribution of covariance matrix A, centered on 0 ($\vec{b} = 0$).

$$\langle x_{k_1} x_{k_2} \dots x_{k_l} \rangle \equiv \frac{1}{Z(A)} \int d^n x x_{k_1} x_{k_2} \dots x_{k_l} \exp\left(-\frac{1}{2}\mathbb{A}(\vec{x})\right)$$

where the factor $Z(A)^{-1}$ is just a normalization constant.
We then introduce a term $\vec{b}$ and compute the derivatives:

$$\frac{\partial}{\partial b_k} Z(A, \vec{b}) = \int d^n x x_k \exp\left(+\frac{\mathbb{A}}{2} + \vec{b} \cdot \vec{x}\right)$$

In physics, we say that b_k is "coupled" to x_k, and that $Z(A, \vec{b})$ is used as "generating function" for $\vec{x}$.
Then we have:

$$\begin{aligned} \langle x_{k_1} x_{k_2} \dots x_{k_l} \rangle &= (2\pi)^{-n/2}(\det A)^{-1/2} \left[\frac{\partial}{\partial b_{k_1}} \frac{\partial}{\partial b_{k_2}} \dots \frac{\partial}{\partial b_{k_l}} Z(A, \vec{b})\right]_{\vec{b}=0} = \\ &= \frac{\partial}{\partial b_{k_1}} \frac{\partial}{\partial b_{k_2}} \dots \frac{\partial}{\partial b_{k_l}} e^{w_2(\vec{b})}\bigg|_{\vec{b}=0} \end{aligned}$$

0.2.2 Wick's Theorem

From the previous formula we know that:

$$\frac{\partial}{\partial b_i} \text{ pulls down a } b_i$$

So, if we set $b_i = 0$, the result will also be 0, as $\langle x_i \rangle = 0$.
However, with more derivatives, the result can be non-zero, due to correlations between different x_j:

$$\langle x_i x_j \rangle \text{ may be} \neq 0$$

But always, for an odd number of derivatives, due to the symmetry of the gaussian, we have:

$$\langle x_i x_j x_k \rangle = 0$$

So the expectation value of the product of different random variables, sampled from the same gaussian distribution centered on 0, is only non-zero for an even number of variables.

(Reference: "Path integrals in Physics" vol. 1, Chaichian & Demichev)

The contribution from every pair (k_p, k_q) from the list of indices $K = \{k_1, k_2, \ldots, k_l\}$ is associated with the covariance $(A^{-1})_{k_p,k_q}$. Then we arrive at Wick's theorem:

$$\langle x_{k_1} x_{k_2} \ldots x_{k_l} \rangle = \sum_{P \in \sigma(K)} A^{-1}_{k_{P_1} k_{P_2}} A^{-1}_{k_{P_3} k_{P_4}} \ldots A^{-1}_{k_{P_{l-1}} k_{P_l}} = \sum_{P \in \sigma(K)} \langle x_{k_{P_1}} x_{k_{P_2}} \ldots \langle x_{k_{P_{l-1}}} x_{k_{P_l}} \rangle \rangle$$

where P is a permutation of indices K.
This is true only for gaussians.

Exercise. For a one variable gaussian distribution:

$$\frac{1}{Z(A)} \exp\left(-\frac{a}{2}x^2\right)$$

prove that:

$$\begin{aligned} \langle x^2 \rangle &= \frac{1}{a} \\ \langle x^4 \rangle &= \frac{3}{a^2} = 3(\langle x^2 \rangle)^2 \end{aligned}$$

(Hint for $\langle x^4 \rangle$: $k_1 = 1 = k_2 = k_3 = k_4$).

0.3 Steepest Descent Integrals

It is possible to use gaussian integrals to solve a more general set of integrals, with the *Steepest Descent Integrals* (or *saddle point method*).

The idea is to convert a generic integral to a Gaussian one (with some approximation, depending on a controllable parameter).
We start with:

$$I(\lambda) = \int d^n x e^{-F(\vec{x})/\lambda}$$

As $\lambda \to 0$, it becomes important to study the integral behaviour around the minimum of $F(\vec{x})$, that is the maximum of $-F(\vec{x})$, which is called the "saddle point". This is because this function is analytic, and in the complex plane there are saddle points at the real coordinate of the maximum.

We do this by changing variables:

$$\vec{x} = \vec{x}_c + \sqrt{\lambda}\vec{y}$$

where $\vec{x}_c$ is the saddle point, and $\vec{y}$ is a given variable.
Then we expand around $\lambda = 0$ and $\vec{x} = \vec{x}_c$, so that:

$$\frac{1}{\lambda}F(\vec{x}) = \frac{1}{\lambda}F(\vec{x}_c) + \cancel{\frac{1}{\lambda}\sum_i \partial_{x_i} F(\vec{x}_c) y_i \sqrt{\lambda}} + \frac{1}{\lambda}\frac{1}{2!}\sum_{ij}\partial^2_{x_i x_j}F(\vec{x}_c) y_i y_j \lambda + O(\lambda^{1/2})$$

Note that, at the maximum, the first derivative is 0.
Substituting in the integral (and transforming the differential) we get:

$$I(\lambda) = \lambda^{n/2} \exp\left(-\frac{F(\vec{x}_c)}{\lambda}\right) \int d^n y \exp\left[-\frac{1}{2}\sum_{ij}\partial^2_{x_i x_j}F(\vec{x}_c) y_i y_j - R(y)\right]$$

The idea is that the integrand of $I(\lambda)$ is "more and more similar" to a gaussian as λ is lower.

Then, by taking the limit $\lambda \to 0$:

$$I(\lambda) \underset{\lambda \to 0}{\approx} (2\pi\lambda)^{n/2} [\det \partial^2_{x_i x_i} F]^{-1/2} \exp\left(-\frac{F(\vec{x}_c)}{\lambda}\right)$$

Often, in literature, the following expression can be found the case for $d = 1$:

$$I(s) = \int g(z) e^{sf(z)} dz \underset{s \to \infty}{\approx} \frac{(2\pi)^{1/2} g(z_c) e^{sf(z_c)}}{|sg''(z_c)|}$$

with $f, g \in \mathbb{R}$, and z_c is the maximum of f: $f(z_c) \geq f(z)$.
This formula is useful in physics: s can model the system's size, and $s \to \infty$ is the limit for a large system.

Example. We can then derive the Stirling approximation. Recall that:

$$s! = \int_0^\infty x^s e^{-x} dx$$

We then perform a change of variables:

$$x = zs$$

so that:

$$s! = s^{s+1} \int_0^\infty e^{s(\ln z - z)} dz$$

Then we look for the maximum of $f(z) = \ln z - z$:

$$f'(z) = \frac{d}{dz}(\ln z - z) = \frac{1}{z} - 1 \Rightarrow z_c = 1$$
$$f''(z) = -\frac{1}{z^2} \Rightarrow f''(z_c) = -1$$

Using the Steepest Descent formula we get the Stirling approximation for a factorial:

$$s! \approx \frac{\sqrt{2\pi}}{|s|^{1/2}} s^{s+1} e^{-s} \approx \sqrt{2\pi} s^{s+1/2} e^{-s}$$

Note that the integral goes from 0 to ∞. We can use the gaussian integral as the maximum is far from 0, and so we can neglect the left side.

Exercise. Compute the Steepest Descent Approximation for the following integral:

$$I(s) = \int_{-\infty}^{\infty} e^{sx - \cosh x} dx$$

Note that, for this peculiar case, the simple 1D formula does not work - and so one should proceed with the general method (full steps: find maximum, second derivative...).

Exercise.

$$I(\lambda) = \int_0^{\infty} \cos(x) \exp\left(-N\left[\left(x - \frac{\pi}{3}\right)^2 + \left(x - \frac{\pi}{3}\right)^4\right]\right) dx$$

Find also the limit:

$$I(N) \xrightarrow[N\to\infty]{} \dots$$