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: 14 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 ($[\mathrm{kg\,m^{-1}}]$), 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:

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

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 Multivariate Gaussian

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

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

We denote a random variable sampled from $f(x;\mu,\sigma)$ as $X \sim \mathcal{N}(\mu, \sigma)$.
Suppose that we have multiple random variables $\{X_i\}_{i=1,\dots,n}$, each normally distributed ($X_i \sim \mathcal{N}(\mu_i, \sigma_i)$), with covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ defined as:

$$\Sigma_{ij} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

Their joint pdf is given by a **multivariate normal distribution** :

$$f(x_1, \dots, x_n; \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

0.3 Moments and Gaussians

We want now to compute the moment generating function for a multivariate gaussian, that is the value of the integral:

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \int_{\mathbb{R}^n} e^{\boldsymbol{t}\cdot\boldsymbol{x}} f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)\, \mathrm{d}^n x \tag{1}$$

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

Recall that the **gaussian integral**, i.e. the 0-th moment of a normal univariate distribution is:

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

Proof. The integral as is can't be computed in terms of elementary functions. However, its square can be calculated:

$$\left(\int_{-\infty}^{\infty} \mathrm{d}x \exp\left(-\frac{a}{2}x^2\right)\right)^2 = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y \exp\left(-\frac{2}{2}(x^2+y^2)\right)$$

Transforming to polar coordinates:

$$= \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \mathrm{d}r \exp\left(-\frac{a}{2}r^2\right) r = -\frac{2\pi}{a} \exp\left(-\frac{a}{2}r^2\right)\Big|_0^{\infty} = \frac{2\pi}{a}$$

and we arrive at the desired result by simply taking the square root.

Consider now the integral of the multivariate case, with $\boldsymbol{\mu} = \mathbf{0}$ (meaning we applied a translation from the general case):

$$Z(\Sigma) = \int_{\mathbb{R}^n} \mathrm{d}^n\boldsymbol{x} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right)$$

Notice that the inverse of the covariance matrix $\Sigma^{-1} \equiv A$ is a symmetric positive-definite matrix, thus can be used to define a quadratic form:

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

The integral can be computed by applying a change of variables, *rotating* $\boldsymbol{x}$ such that A becomes diagonal:

$$\boldsymbol{y} = O\boldsymbol{x}; \qquad O \in \mathbb{R}^{n\times n}; \quad O^T = O^{-1},\ \det(O) = 1$$

where O is an orthogonal matrix, with a set of orthogonal eigenvectors of A as columns, such that:

$$OAO^{-1} = \operatorname{diag}(a_1, \ldots, a_n)$$

with a_i being the eigenvalues of A.
Note that, as $\det(O) = 1$, the volume element in the integral does not change. So, by substituting:

$$\boldsymbol{x} = O^{-1}\boldsymbol{y}; \qquad \boldsymbol{x}^T = \boldsymbol{y}^T(O^{-1})^T = \boldsymbol{y}^T O$$

in the integral, we get:

$$
\begin{aligned}
Z(A) &= \int_{\mathbb{R}^n} \mathrm{d}^n \boldsymbol{y} \exp\left(-\frac{1}{2}\boldsymbol{y}^T O A O^T \boldsymbol{y}\right) = \int_{\mathbb{R}^n} \mathrm{d}^n \boldsymbol{y} \exp\left(-\frac{1}{2}\sum_{i=1}^n a_i y_i^2\right) = \\
&= \prod_{i=1}^n \int_{\mathbb{R}} \mathrm{d} y_i \exp\left(-\frac{1}{2} a_i y_i^2\right) = (2\pi)^{n/2} \prod_{i=1}^n a_i^{-1/2} \underset{(a)}{=} (2\pi)^{n/2} (\det(A))^{-1/2}
\end{aligned} \tag{2}
$$

where in (a) we noted that the determinant of a matrix is the product of its eigenvalues.

We are now ready to consider the more general case of (1), by simply adding a linear term in the exponential of $Z(A)$:

$$
Z(A, \boldsymbol{b}) \equiv \int_{-\infty}^{\infty} \mathrm{d}^n \boldsymbol{x} \exp\left(-\frac{1}{2}\mathbb{A}(\boldsymbol{x}) + \boldsymbol{b}\cdot\boldsymbol{x}\right) \qquad \boldsymbol{b}\cdot\boldsymbol{x} = \sum_{i=1}^n b_i x_i \tag{3}
$$

To compute this integral, a trick is to translate the maximum of the exponential to the origin. So we start by differentiating:

$$
\frac{\partial}{\partial x_i}\left(\frac{1}{2}\mathbb{A}(\boldsymbol{x}) - \boldsymbol{b}\cdot\boldsymbol{x}\right) \overset{!}{=} 0 \quad \forall i \tag{4}
$$

Note that:

$$
\begin{aligned}
\frac{\partial}{\partial x_i}\mathbb{A}(\boldsymbol{x}) &= \frac{\partial}{\partial x_i}\sum_{ab} x_a A_{ab} x_b = \sum_{ab} \delta_{ai} A_{ab} x_b + \sum_{ab} x_a A_{ab} \delta_{bi} = \\
&= \sum_b A_{ib} x_b + \sum_a x_a A_{ai}
\end{aligned}
$$

By renaming the first summation variable to a, we get:

$$
= \sum_a (A_{ia} + A_{ai}) x_a \underset{(b)}{=} 2\sum_a A_{ia} x_a = 2A\boldsymbol{x}
$$

where in (b) we used the fact that A is symmetrical ($A_{ij} = A_{ji}$).
Substituting in (4):

$$
\frac{1}{\cancel{2}}\cancel{2}\sum_j A_{ij} x_j = b_i \quad \forall i \underset{(c)}{\Leftrightarrow} A^T \boldsymbol{x} = \boldsymbol{b} \underset{(d)}{\Leftrightarrow} \boldsymbol{x}^* = A^{-1}\boldsymbol{b}
$$

In (c) we noted that b_i is the scalar product between the i-th column of A and $\boldsymbol{x}$, leading to the transpose in the matrix notation. Of course, as $A = A^T$, in (d) we simply dropped the transpose.

We can now apply the coordinate change:

$$
\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{y}
$$

Substituting in the exponential argument:

$$
\begin{aligned}
-\frac{\mathbb{A}(\boldsymbol{x})}{2}+\boldsymbol{b}\cdot\boldsymbol{x} &= -\frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x}+\boldsymbol{x}^T\boldsymbol{b} = -\frac{1}{2}(\boldsymbol{x}^*+\boldsymbol{y})^T A(\boldsymbol{x}^*+\boldsymbol{y})+(\boldsymbol{x}^*+\boldsymbol{y})^T\boldsymbol{b} = \\
&= -\frac{1}{2}\left[\boldsymbol{x}^{*T}A\boldsymbol{x}^*+\boldsymbol{y}^T A\boldsymbol{y}+\cancel{\boldsymbol{x}^{*T}A\boldsymbol{y}}+\cancel{\boldsymbol{y}^T A\boldsymbol{x}^*}\right]+\boldsymbol{x}^{*T}A\boldsymbol{x}^*+\cancel{\boldsymbol{y}^T A\boldsymbol{x}^*}
\end{aligned}
\tag{5}
$$

Note, in fact, that $\boldsymbol{y}^T A\boldsymbol{x}^* = (\boldsymbol{x}^{*T}A^T\boldsymbol{y})^T = (\boldsymbol{x}^{*T}A\boldsymbol{y})^T$ because A is symmetric, and then $(\boldsymbol{x}^{*T}A\boldsymbol{y})^T = \boldsymbol{x}^{*T}A\boldsymbol{y}$ because they are scalars.
Then:

$$
\boldsymbol{x}^{*T}A\boldsymbol{x}^* = (A^{-1}\boldsymbol{b})^T AA^{-1}\boldsymbol{b} = \boldsymbol{b}^T(A^{-1})^T\boldsymbol{b} = \boldsymbol{b}^T A^{-1}\boldsymbol{b} = \boldsymbol{b}\cdot\boldsymbol{x}^*
$$

And substituting in (5):

$$
-\frac{\mathbb{A}(\boldsymbol{x})}{2}+\boldsymbol{b}\cdot\boldsymbol{x} = -\frac{1}{2}\boldsymbol{y}^T A\boldsymbol{y}+\underbrace{\frac{1}{2}\boldsymbol{b}\cdot\boldsymbol{x}^*}_{\omega_2(\boldsymbol{b})}
$$

To simplify notation, let's define:

$$
w_2(\boldsymbol{b}) = \frac{1}{2}\sum_{i,j=1}^{n} b_i(A^{-1})_{ij}b_i = \frac{1}{2}\boldsymbol{b}\cdot\boldsymbol{x}^* \tag{6}
$$

As the change of variables involves only a translation by a constant value, the volume element in the integral does not change, leading to:

$$
Z(A,\boldsymbol{b}) = \int_{-\infty}^{\infty} \mathrm{d}^n\boldsymbol{y}\exp\left(-\frac{\mathbb{A}(\boldsymbol{y})}{2}+\omega_2(\boldsymbol{b})\right)
$$

Note that $\omega_2(\boldsymbol{b})$ is constant, thus can be extracted from the integral:

$$
= e^{\omega_2(\boldsymbol{b})}\int_{-\infty}^{\infty}\mathrm{d}^n\boldsymbol{y}\exp\left(-\frac{\mathbb{A}(\boldsymbol{y})}{2}\right) \underset{(2)}{=} e^{\omega_2(\boldsymbol{b})}(2\pi)^{n/2}(\det A)^{-1/2} \tag{7}
$$

Another way to solve the integral for $Z(A,\boldsymbol{b})$ is by using the matrix equivalent of "completing the square". We start by considering the argument of the exponential in (3):

$$
-\frac{1}{2}(\boldsymbol{x}^T A\boldsymbol{x}-2\boldsymbol{b}^T\boldsymbol{x})
$$

$\boldsymbol{x}^T A\boldsymbol{x}$ has the role of the square, and $-2\boldsymbol{b}^T\boldsymbol{x}$ that of the double product.
We can then sum and subtract a constant vector $\boldsymbol{c}$ in order to rewrite:

$$
\boldsymbol{x}^T A\boldsymbol{x}-2\boldsymbol{b}^T\boldsymbol{x}+\boldsymbol{c}-\boldsymbol{c} = \boldsymbol{y}^T A\boldsymbol{y}-\boldsymbol{c}
$$

for some $\boldsymbol{y}\in\mathbb{R}^n$.
Comparing to a generic square:

$$
(\boldsymbol{a}+\boldsymbol{b})^T A(\boldsymbol{a}+\boldsymbol{b}) = \boldsymbol{a}^T A\boldsymbol{a}+\boldsymbol{b}^T A\boldsymbol{b}+2\boldsymbol{a}^T A\boldsymbol{b}
$$

we note that $\boldsymbol{a} = \boldsymbol{x}$ and $\boldsymbol{b} = -A^{-1}\boldsymbol{b}$, leading to:

$$\boldsymbol{x}^T A\boldsymbol{x} - 2\boldsymbol{b}^T\boldsymbol{x} = (\boldsymbol{x} - A^{-1}\boldsymbol{b})^T A(\boldsymbol{x} - A^{-1}\boldsymbol{b}) - \boldsymbol{b}^T A^{-1}\boldsymbol{b}$$

Defining $A^{-1}\boldsymbol{b} \equiv \boldsymbol{x}^*$ and $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}^*$ then leads to the same calculations as before.

Exercise 1 (Multivariate Gaussian Integral)**:**

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

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

Note that $\det A = 8$, and:

$$A^{-1} = \frac{1}{8}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

So, by simply using (2) and (7):

$$\begin{aligned} Z(A, 0) &= \frac{(2\pi)^{2/2}}{\sqrt{8}} = \frac{\pi}{\sqrt{2}} \\ \frac{1}{2}\begin{pmatrix} 1 & 0 \end{pmatrix} \frac{1}{8} &\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{3}{16} \\ Z(A, \boldsymbol{b}) &= \frac{\pi}{\sqrt{2}} \exp\left(\frac{3}{16}\right) \end{aligned}$$

0.3.1 Gaussian expectation values

The result in (7) is exactly what we need to compute the moment generating function for the multivariate normal (1).

So, we can finally compute moments:

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

by simply deriving the generating function $Z(A, \boldsymbol{b})$ with respect to certain variables in $\boldsymbol{b}$. For example:

$$\langle x_k \rangle = \frac{1}{Z(A)} \frac{\partial}{\partial b_k} Z(A, \vec{b}) = \frac{1}{Z(A)} \int \mathrm{d}^n\boldsymbol{x}\, x_k \exp\left(-\frac{\mathbb{A}(\boldsymbol{x})}{2} + \boldsymbol{b}^T\boldsymbol{x}\right)$$

For the general case:

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

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

0.3.2 Wick's Theorem

From the previous formula we know that:

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

Explicitly, recall that:

$$\omega_2(\boldsymbol{b}) = \frac{1}{2} \boldsymbol{b}^T A^{-1} \boldsymbol{b}$$

and so:

$$\frac{\partial}{\partial b_i} e^{\omega_2(\boldsymbol{b})} = \frac{1}{2} e^{\omega_2 \boldsymbol{b}} \frac{\partial}{\partial b_i} \sum_{tk} b_t A^{-1}_{tk} b_k = e^{\omega_2 \boldsymbol{b}} \sum_k A^{-1}_{ik} b_k$$

If we now set $\boldsymbol{b} = \boldsymbol{0}$, the result will be 0, meaning that:

$$\langle x_i \rangle = \frac{\partial}{\partial b_i} \frac{Z(A, \boldsymbol{b})}{Z(A)} = 0$$

This result is expected, as in $Z(A, \boldsymbol{b})$ all random variables are centered in 0. However, note that if we derive one more time, with respect to some b_l:

$$\frac{\partial}{\partial b_i} \frac{\partial}{\partial b_l} e^{\omega_2(\boldsymbol{b})} = e^{\omega_2 \boldsymbol{b}} \sum_s A^{-1}_{ls} b_s \sum_k A^{-1}_{ik} b_k + e^{\omega_2 \boldsymbol{b}} A^{-1}_{il}$$

And now, if we set $\boldsymbol{b} = \boldsymbol{0}$, the result may be $\neq 0$.
Note that if we derive one more time we return to the previous situation - and the result will be also 0.

In general, every moment of odd-order is 0, 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. This result is known as the **Wick's theorem** (also known in literature as the **Isserlis theorem**).

By extending this argument, one can find a way to compute the even-order moments, leading to the following formula (which we will not prove):

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

where (k_p, k_q) are a pair of indices from $K = \{k_1, \dots, k_l\}$, and P is a permutation of K, so that (k_{P_1}, k_{P_2}) is the first pair of indices after the permutation P. The sum is over all the distinct ways of partitioning $l = 2s$ variables in pairs to obtain distinct products of s groups.
So, the total number of terms to be added will be $(2s)!/(2^s s!)$ - that is the total number of permutation of $2s$ elements, where the order within couples does not matter (2^s) and neither the order of the couples themselves $(s!)$.
Note that:

$$\frac{(2s)!}{2^s n!} = (2s-1)! = (2s-1)(2s-3)(2s-5)\dots$$

Where !! denotes the *double factorial*, not to be confused with the factorial of a factorial (which requires brackets: $(a!)!$).

Exercise 2 (Wick's theorem)**:**

> Consider a univariate normal distribution:
>
> $$f(x) = \frac{1}{Z(A)} \exp\left(-\frac{a}{2} x^2\right)$$
>
> Show that:
>
> $$\begin{aligned} \langle x^2 \rangle &= \frac{1}{a} \\ \langle x^4 \rangle &= \frac{3}{a^2} = 3(\langle x^2 \rangle)^2 \end{aligned}$$
>
> Here the A matrix is just the scalar $a = \sigma^{-2}$. As the pdf is univariate, there is only one index possible $K = \{1\}$. As $(2-1)!! = 1!! = 1$, there is only one term in the summation, thus:
>
> $$\langle x^2 \rangle = A^{-1}_{11} = \frac{1}{a}$$
>
> For the 4-th order, however, we have more combinations: $(4-1)!! = 3!! = 3 \cdot 1 = 3$. Again, there is only one possible index, so all terms will be the same:
>
> $$\langle x^4 \rangle = A^{-1}_{11} A^{-1}_{11} + A^{-1}_{11} A^{-1}_{11} + A^{-1}_{11} A^{-1}_{11} = \frac{3}{a^2} = 3(\langle x^2 \rangle)^2$$

0.4 Steepest Descent Integrals

It is possible to use gaussian integrals to solve a more general set of integrals, thanks to the *Steepest Descent approximation*.
We start with an integral of the form:

$$I(\lambda) \equiv \int_S \mathrm{d}^n x \exp\left(-\frac{F(\boldsymbol{x})}{\lambda}\right) \tag{8}$$

where λ is a small parameter (the approximation is more and more accurate as $\lambda \to 0$), $F(\boldsymbol{x})$ has a global minimum in $\boldsymbol{x_0} \in (a,b)$ and $S \subseteq \mathbb{R}^n$ is a sufficiently large region.
Note that, if λ is lowered, the integral is dominated by the neighborhood of the minimum $\boldsymbol{x_0}$. In fact:

$$h(\boldsymbol{x}) \equiv \exp\left(-\frac{F(\boldsymbol{x})}{\lambda}\right); \quad \frac{h(\boldsymbol{x_0})}{h(\boldsymbol{x})} = \exp\left(-\frac{1}{\lambda}(F(\boldsymbol{x_0}) - F(\boldsymbol{x}))\right)$$

As $F(\boldsymbol{x_0}) - F(\boldsymbol{x}) < 0$, the ratio becomes exponentially higher if $\lambda \to 0$. Basically, for $\lambda \to 0$, the integrand function becomes "more and more similar to a gaussian".

To compute the integral, then, we translate the coordinates about $\boldsymbol{x_0}$:

$$\boldsymbol{x} = \boldsymbol{x_0} + \sqrt{\lambda}\boldsymbol{y} \qquad \mathrm{d}^n\boldsymbol{x} = \lambda^{n/2}\,\mathrm{d}^n\boldsymbol{y}$$

Then we perform a second order Taylor expansion about $\lambda = 0$ and $\boldsymbol{x} = \boldsymbol{x_0}$:

$$\frac{1}{\lambda}F(\boldsymbol{x}) = \frac{1}{\lambda}F(\boldsymbol{x_0}) + \cancel{\frac{1}{\lambda}\sum_i \partial_{x_i} F(\boldsymbol{x_0}) y_i \sqrt{\lambda}} + \frac{1}{\cancel{\lambda}}\frac{1}{2!}\sum_{ij}\partial^2_{x_i x_j} F(\boldsymbol{x_0}) y_i y_j \cancel{\lambda} + O(\lambda^{1/2})$$

where we cancelled the first derivative, as $\boldsymbol{x_0}$ is a stationary point for F.

Substituting back in the integral we get:

$$I(\lambda) = \lambda^{n/2} \exp\left(-\frac{F(\boldsymbol{x_0})}{\lambda}\right) \int_{S'} \mathrm{d}^n\boldsymbol{y} \exp\left[-\frac{1}{2}\sum_{ij}\partial^2_{x_i x_j} F(\boldsymbol{x_0}) y_i y_j - R(\boldsymbol{y})\right]$$

This is a gaussian integral $Z(A)$, with A being the Hessian of F evaluated at the minimum $\boldsymbol{x_0}$ (or, equivalently, at the maximum of $-F(\boldsymbol{x})$).
Now, for λ sufficiently small, we can ignore $R(\boldsymbol{y})$ and compute the integral with (2), leading to the approximation:

$$I(\lambda) \underset{\lambda \to 0}{\approx} (2\pi\lambda)^{n/2} [\det \partial^2_{x_i x_i} F(\boldsymbol{x_0})]^{-1/2} \exp\left(-\frac{F(\boldsymbol{x_0})}{\lambda}\right) \tag{9}$$

Doing this, we implicitly integrated over the entire $\mathbb{R}^n$. This is fine because, for $\lambda \to 0$, the gaussian is "peaked" in a small region around $\boldsymbol{x_0}$, and vanishes exponentially moving further away.

The Steepest Descent approximation generalizes Laplace's method for calculating integrals, which has a much simpler expression for the limited case of univariate integrals:

$$I(s) = \int g(z) e^{sf(z)} dz \underset{s\to\infty}{\approx} \frac{(2\pi)^{1/2} g(z_c) e^{sf(z_c)}}{|sf''(z_c)|^{1/2}} \tag{10}$$

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

Example 1 (Stirling approximation)**:**

We can use the Steepest Descent approximation to derive the formula for the Stirling approximation of factorials.
Recall that a factorial is merely the Γ function evaluated on $\mathbb{N}$:

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

We then perform a change of variables:

$$x = zs$$

so that:

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

This is an integral in the form:

$$\int \exp\left(-\frac{F(x)}{\lambda}\right)$$

if we let $\lambda = 1/s$ and $F(z) = z - \ln z$. So we need to find the minimum of $F(z)$:

$$\begin{aligned} F'(z) &= \frac{\mathrm{d}}{\mathrm{d}z}(z - \ln z) = 1 - \frac{1}{z} \overset{!}{=} 0 \Rightarrow z_c = 1 \\ F''(z) &= \frac{1}{z^2} \Rightarrow F''(z_c) = 1 > 0 \end{aligned}$$

We can now apply (9), leading to:

$$s! \underset{s\to\infty}{\approx} \left(\frac{2\pi}{s}\right)^{1/2} (1)^{1/2} e^{-s} = \sqrt{2\pi} s^{s+\frac{1}{2}} e^{-s}$$

Note that the same result can be obtained by using the much simpler (10), with $g(z) \equiv 1$ and $f(z) = \ln z - z$.

Exercise 3 (Steepest Descent Approximation)**:**

Compute the Steepest Descent Approximation for the following integral (for $s \to \infty$):

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

By collecting a s in the exponential argument:

$$I(s) = \int_{-\infty}^{\infty} \exp\left(s\left(x - \frac{\cosh x}{s}\right)\right)$$

we can bring back to the form of (8) with $F(x) = \cosh x/s - x$ and $\lambda = s^{-1}$. We find the minimum of $F(x)$ by differentiating:

$$F'(x) = \frac{\sinh x}{s} - 1 \stackrel{!}{=} 0 \Rightarrow x_0 = \sinh^{-1} s$$

$$F''(x) = \frac{\cosh x}{s} \Rightarrow F''(x_0) = \frac{\cosh \sinh^{-1} s}{s} = \frac{\sqrt{1+s^2}}{s} > 0$$

Finally, by applying (9) we obtain the result:

$$\begin{aligned} I(s) \underset{s\to\infty}{\approx} \sqrt{\frac{2\pi}{s}}\sqrt{\frac{s}{\sqrt{1+s^2}}} \exp\left(\frac{\sqrt{1+s^2}}{s} - \sinh^{-1} s\right) = \\ = \frac{\sqrt{2\pi}}{(1+s^2)^{1/4}} \exp\left(\frac{\sqrt{1+s^2}}{s} - \sinh^{-1} s\right) \end{aligned}$$

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

Exercise 4 (Laplace's formula)**:**

Compute:

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

in the limit $N \to \infty$.

For this exercise we can use Laplace's formula (10) with:

$$g(x) = \cos(x) \qquad f(x) = -\left[\left(x - \frac{\pi}{3}\right)^2 + \left(x - \frac{\pi}{3}\right)^4\right]$$

By looking at $f(x)$ one can see directly that it has a global maximum in $x_0 = \pi/3$. In fact:

$$f'(x) = -\left[2\left(x - \frac{\pi}{3}\right) + 4\left(x - \frac{\pi}{3}\right)^3\right] \stackrel{!}{=} 0 \Leftrightarrow x_0 = \frac{\pi}{3}$$
$$f''(x) = -\left[2 + 12\left(x - \frac{\pi}{3}\right)^2\right] \Rightarrow f''(x_0) = -2 < 0$$

And so we arrive at:

$$I(N) \underset{N\to\infty}{\approx} \frac{(2\pi)^{1/2}\cos(\pi/3)e^{N\cdot 0}}{|N(-2)|^{1/2}} = \frac{1}{2}\sqrt{\frac{\pi}{N}}$$