Integrating Gaussians like a boss!

Abstract

This is a deep dive into Gaussian integrals.

Index Terms

Gaussian integrals, central limit theorem

I. The Gaussian function

My strongest memory of the class is the very beginning, when he started, not with some deep principle of nature, or some experiment, but with a review of Gaussian integrals. Clearly, there was some calculating to be done.

—Joe Polchinski, on Richard Feynman's quantum mechanics class[1].

Due to the central limit theorem, Gaussian functions show up everywhere[2]. One needs to deal with its integrals in various forms. This is an extensive tutorial on Gaussian integrals so that you can evaluate them like a boss! This post is inspired by the masterpiece book "The Principles of Deep Learning Theory" from Robert at al [3], and I took the liberty of borrowing some ideas from their work, including the quoted remarks from Polchinski.

Let us first show how a Gaussian distribution looks like.

$$\mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1}$$

where μ is the mean value and σ is the variance.

Assume you are on a beach, and you don't remember the form of the Gaussian function. What would you do in such a desperate situation? Normal distribution is symmetric around its mean value. So we must have a $(x-\mu)^2$ in the exponent. Note that you can never ever put a parameter which has a dimension in the exponent. You can think of x as length. $(x-\mu)^2$ has unit Length^2. Therefore you need to divide $(x-\mu)^2$ by some parameter that has the unit Length^2, and that is $2\sigma^2$. The factor 2 is a convention. That fixes the exponent as $-\frac{(x-\mu)^2}{2\sigma^2}$, where you need the negative sign so that the density function is normalizable. How about the prefactor? Remember that normal distribution is a density, and we will soon integrate it over the parameter x. Therefore, it needs to have a unit of Length^{-1} so that once you multiply it with dx you get probability, which is a unitless number. So you need $1/\sigma$ in front of the exponential. $\sqrt{2\pi}$ is for normalizing the distribution, which we will derive later. And that is how you can remind yourself the form of the normal distribution.

II. The pesky error function

Whenever you define a probability density function, you need integrals of if with finite limits. Unfortunately for the Gaussian function, the integral cannot be evaluated in terms of elementary functions. However, there is a dedicated function created just for that, which is defined as follows:

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x d\tau \, e^{-\tau^2}. \tag{2}$$

The area under the Gaussian distribution between two arbitrary limits can be written as:

$$\begin{aligned}
\int_{x_1}^{x_2} dx \mathcal{N}_{\mu,\sigma}(x) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} dx e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{\pi}} \int_{\frac{x_1-\mu}{\sqrt{2}\sigma}}^{\frac{x_2-\mu}{\sqrt{2}\sigma}} d\tau e^{-\tau^2} = \frac{1}{\sqrt{\pi}} \int_{\frac{x_1-\mu}{\sqrt{2}\sigma}}^{0} d\tau e^{-\tau^2} + \frac{1}{\sqrt{\pi}} \int_0^{\frac{x_2-\mu}{\sqrt{2}\sigma}} d\tau e^{-\tau^2} \\
&= \frac{1}{2}\text{erf}\left(\frac{x_2-\mu}{\sqrt{2}\sigma}\right) - \frac{1}{2}\text{erf}\left(\frac{x_1-\mu}{\sqrt{2}\sigma}\right), \tag{3}
\end{aligned}$$

email: quarktetra@gmail.com
Find the interactive HTML-document here.

and that is how we can compute the integral with finite limits. A special case of the integral in Eq. (3) is realized when $x_1 \to -\infty$, which gives the cumulative distribution function:

$$\text{CDF}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} d\tau e^{-\frac{(\tau-\mu)^2}{2\sigma^2}} = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]. \tag{4}$$

erf(x) is readily available in Python, R and Javascript. In Figure 1 we plot Gaussian probability density function (PDF) and cumulative distribution function (CDF).
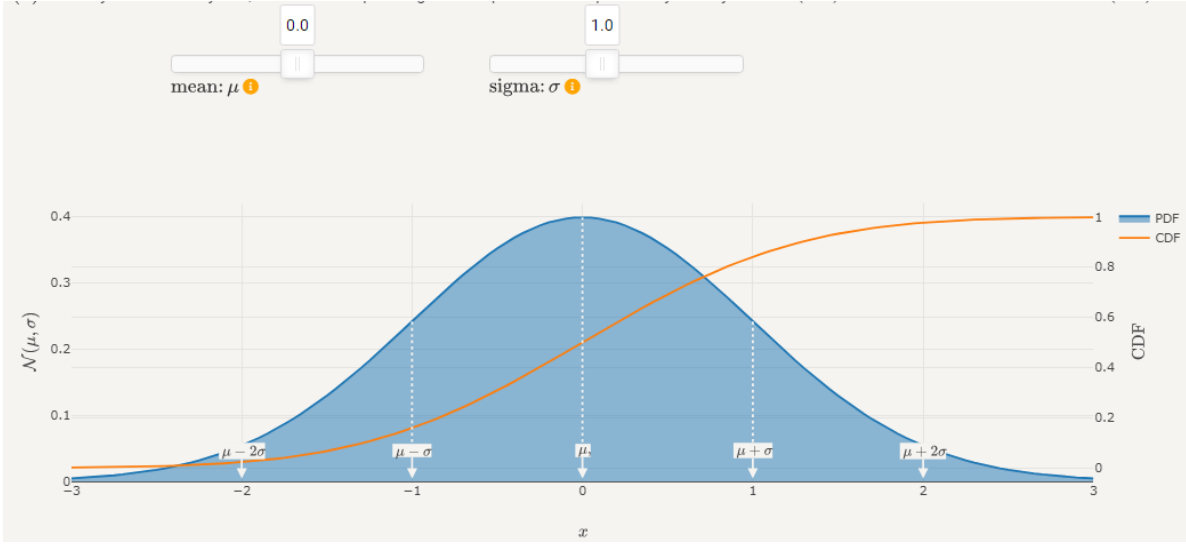
mean: μ
sigma: σ

Figure 1: Interactive plot showing the normal distribution PDF and CDF.

III. THE CENTRAL LIMIT THEOREM

Gaussian function appears everywhere, including statistics, physics, and neural networks [3]. Why is that? It is because of the central limit theorem, which can be loosely defined as follows: If you add up many random variables, their properly normalized sum converges toward a normal distribution irrespective of the type of distribution of the original variables. That is a very strong statement, and justifies the use of normal distribution if you think the end quantity is the sum of many underlying contributors. In the special case of independent and identically distribute random variables X_i with $i = 1, 2, \cdots n$, the central limit theorem can be written as follows:

$$\frac{1}{\sqrt{n}\sigma} \sum_{i=0}^{n} (X_i - \mu) \sim \mathcal{N}(0,1). \tag{5}$$

IV. THE NORMALIZATION FACTOR

The first integral we will look at is the most basic one. In order to be a proper distribution function, it has to be normalized. Therefore we need to compute the factor that appears in front of the exponential in Eq. (1). We need to evaluate the following integral:

$$I \equiv \int_{-\infty}^{\infty} dx e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{6}$$

There are many different ways computing the integral in Eq. (6), and I will only discuss the ones I like the most.

A. Squaring the integral to polarize it

Here is our integral to evaluate:

$$I = \int_{-\infty}^{\infty} dx e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \sqrt{2}\sigma \int_{-\infty}^{\infty} dx e^{-x^2}. \tag{7}$$

We will do a neat trick and square the integral:

$$I^2 = 2\sigma^2 \int_{-\infty}^{\infty} dx e^{-x^2} \int_{-\infty}^{\infty} dy e^{-y^2} = 2\sigma^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy e^{-(x^2+y^2)}. \tag{8}$$

It doesn't look like we made progress, but we notice that $dxdy$ is the differential area, and switch to polar coordinates ,as in Fig. 2, to see if that helps.
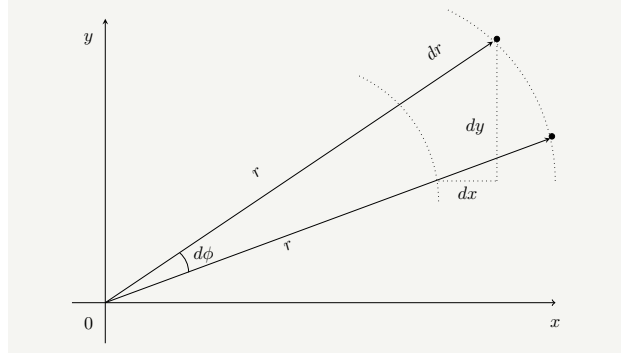

Figure 2: The integral over dA can be evaluated in Cartesian coordiates with $dA = dxdy$ or in polar coordinates $dA = rdrd\phi$.

In the polar coordinates we have $x = r\cos\phi$, $y = r\sin\phi$, and $dA = dxdy = rdrd\phi$.

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy e^{-(x^2+y^2)} = \int_0^{\infty} e^{-r^2} r dr \int_0^{2\pi} d\phi = \pi \tag{9}$$

Therefore the integral in Eq. (7) has the value

$$I = \int_{-\infty}^{\infty} dx e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \sqrt{2\pi}\sigma, \tag{10}$$

which is the inverse of the coefficient in the Gaussian distribution in Eq. (1) ensuring that it is properly normalized to 1.

B. Complex contour integral, like a boss!

It is possible to compute the Gaussian integral using complex calculus. As e^{-x^2} has no zeros itself, we consider a function $\frac{e^{-x^2}}{g(x)}$, and upgrade the real parameter x to a complex parameter z, and integrate $\frac{e^{-z^2}}{g(z)}$ in on a carefully built contour in Fig. 3. The trick is to combine two pieces of complex line integrals to give what we want at the end.

Figure 3: Animated complex contour integration.

V. THE MEAN, SIGMA, AND ALL THAT

$\mathcal{N}_{\mu,\sigma}(x)$ has a parameter μ which is the value around which the distribution is symmetric, see Fig. 1. We can show that the expected value of a random number $X \sim \mathcal{N}_{\mu,\sigma}$ is μ by evaluating the following integral:

$$
\begin{aligned}
\mathbb{E}\left[x\right] &\equiv \int_{-\infty}^{\infty} dx\, x \mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, x e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, (x-\mu+\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \underbrace{\int_{-\infty}^{\infty} dx\, (x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{0} + \mu \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \mu,
\end{aligned}
\tag{11}
$$

where the first integral vanishes since the integrand is odd.

In order to compute the variance, we have to figure out how to compute the expected value of $(x-\mu)^2$. We can be a bit more generous and compute the expected value of $(x-\mu)^{2n}$ where n is a positive integer. Note that odd powers of $(x-\mu)$ will have 0 as their expected value as they are antisymmetric.

$$
\begin{aligned}
\mathbb{E}\left[(x-\mu)^{2n}\right] &\equiv \int_{-\infty}^{\infty} dx\, (x-\mu)^{2n} \mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, (x-\mu)^{2n} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, (x-\mu)^{2n} e^{-\alpha(x-\mu)^2} \Big|_{\alpha=\frac{1}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}(-1)^n \frac{d^n}{d\alpha^n} \int_{-\infty}^{\infty} dx\, e^{-\alpha(x-\mu)^2} \Big|_{\alpha=\frac{1}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma}(-1)^n \frac{d^n}{d\alpha^n} \sqrt{\pi}\alpha^{-\frac{1}{2}} \Big|_{\alpha=\frac{1}{2\sigma^2}} = \frac{1}{\sqrt{2}\sigma}\frac{1}{2} \times \frac{3}{2} \times \cdots \times \frac{2n-1}{2}\alpha^{-\frac{2n+1}{2}} \Big|_{\alpha=\frac{1}{2\sigma^2}} \\
&= 1 \times 3 \times 5 \times \cdots \times \frac{2n-1}{2}\sigma^{2n} = \sigma^{2n}(2n-1)!!,
\end{aligned}
\tag{12}
$$

where the double factorial term is defined as $(2n-1)!! \equiv (2n-1)(2n-3)\cdots 1 = \frac{(2n)!}{2^n n!}$.

There is another technique which is known as Feynman's method. It is similar to what we just did: we differentiated the integrand with respect to the parameter we called α. However, it was kind of an obvious thing to do. Feynman's method takes this one step further. You insert a parameter by hand and poke that parameter before you set it to the appropriate value(typically 0 or 1) at the end. It is an incredibly powerful technique and it is work repeating the calculations to illustrate how it works. Let's consider the following integral:

$$
\begin{aligned}
\mathbb{E}\left[e^{J(x-\mu)}\right] &= \int_{-\infty}^{\infty} dx\, e^{J(x-\mu)} \mathcal{N}_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, e^{-\frac{(x-\mu)^2}{2\sigma^2}+J(x-\mu)} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, e^{-\frac{(x-\mu)^2-2\sigma^2 J(x-\mu)}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, e^{-\frac{[x-\mu-\sigma^2 J(x-\mu)]^2}{2\sigma^2}+\frac{J^2\sigma^2}{2}} = e^{\frac{J^2\sigma^2}{2}}.
\end{aligned}
\tag{13}
$$

Note that we can now drop in the term $(x-\mu)^n$ by differentiating the integral with respect to J.

$$
\begin{aligned}
\mathbb{E}\left[(x-\mu)^n\right] &= \left[\left(\frac{d}{dJ}\right)^{2n} \mathbb{E}\left[e^{J(x-\mu)}\right]\right]_{J=0} = \left[\left(\frac{d}{dJ}\right)^{2n} \int_{-\infty}^{\infty} dx\, e^{J(x-\mu)} \mathcal{N}_{\mu,\sigma}(x)\right]_{J=0} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx\, (x-\mu)^{2n} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \left[\left(\frac{d}{dJ}\right)^{2n} e^{\frac{J^2\sigma^2}{2}}\right]_{J=0} \\
&= \left[\left(\frac{d}{dJ}\right)^{2n} \left\{\sum_{i=0}^{\infty} \frac{1}{i!}\left(\frac{\sigma^2}{2}\right)^i J^{2i}\right\}\right]_{J=0} = \sigma^{2n}\frac{(2n)!}{2^n n!} \\
&= \sigma^{2n}(2n-1)!!,
\end{aligned}
\tag{14}
$$

where the only contribution from the summation comes from $i=n$ since J is set to zero after differentiation.

VI. THE MATRIX REPRESENTATION

There are problems involving Gaussians with multiple variables, and ff the variables are not connected, one can simply separate the integrals to evaluate them individually. For example a Gaussian with

two variables will be of the form $exp\left\{-(x-\mu_x)^2/(2\sigma_x^2)-(y-\mu_y)^2/(2\sigma_y^2)\right\}$ which can be separated as $exp\left\{-(x-\mu_x)^2/(2\sigma_x^2)\right\}\times exp\left\{-(y-\mu_y)^2/(2\sigma_y^2)\right\}$, and the integrals simply decouple. However, if there are cross terms of the form $x\,y$, we need to come up with a new set of coordinates to eliminate the cross term. In the jargon of linear algebra, this is referred to as diagonalization. In order to keep the notation simple, let us now upgrade the parameter x to a vector $\mathbf{x} = (x_1, x_2, \cdots, x_N)^T$, and consider the following integral:

$$I = \int_{-\infty}^{\infty} d^N\mathbf{x}\, e^{-\mathbf{x}^T M \mathbf{x}}, \tag{15}$$

where M is an N-by-N symmetric positive definite matrix, which is not necessarily a diagonal one. However we can rotate our vector $\mathbf{x}$ with a rotation matrix R, i.e., $\mathbf{y} = M\mathbf{x}$ or equivalently $\mathbf{x} = M^T\mathbf{y}$, such that the rotation operation diagonalizes the matrix M. It works like below:

$$\mathbf{x}^T M \mathbf{x} = \mathbf{y}^T R M R^T \mathbf{y} \equiv \mathbf{y}^T \Lambda \mathbf{y}, \tag{16}$$

where

$$\Lambda = R M R^T \tag{17}$$

is a diagonal matrix $\frac{1}{\lambda_i}\delta_{ij}$. Since rotation matrices do not change the volume, the integral measure stays the same. Therefore the integrals decouple in the $\mathbf{y}$ space:

$$
\begin{aligned}
I &= \int_{-\infty}^{\infty} d^N\mathbf{x}\, e^{-\frac{1}{2}\mathbf{x}^T M \mathbf{x}} = \int_{-\infty}^{\infty} d^N\mathbf{y}\, e^{-\frac{1}{2}\mathbf{y}^T \Lambda \mathbf{y}} = \int_{-\infty}^{\infty} d^N\mathbf{y}\, e^{-\sum_{i,j}\frac{1}{2\lambda_i}\mathbf{y}_i\delta_{ij}\mathbf{y}_i} = \int_{-\infty}^{\infty} d^N\mathbf{y}\, e^{-\sum_i \frac{y_i^2}{2\lambda_i}} \\
&= \prod_i \int_{-\infty}^{\infty} dy_i\, e^{\frac{y_i^2}{2\lambda_i}} = \prod_i \sqrt{2\pi\lambda_i} = \sqrt{\prod_i (2\pi\lambda_i)} = \sqrt{|2\pi\Lambda^{-1}|} = \sqrt{|2\pi M^{-1}|},
\end{aligned}
\tag{18}
$$

where we used the fact that determinant of a matrix, $|M|$, is the multiplication of its eigenvalues. Putting the reciprocal of this factor infront gives the normalized Gaussian distribution with multiple variables:

$$\mathcal{N}_{\mu,M}(\mathbf{x}) = \frac{1}{\sqrt{|2\pi M^{-1}|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T M (\mathbf{x}-\mu)}, \tag{19}$$

where we introduced the mean value vector $\mu = (\mu_1, \mu_2, \cdots, \mu_N)$.

VII. The Wick contraction

Computation of the moments is very similar to the case of single variable Gaussian. We introduce $\mathbf{J}$ vector and couple it to $\mathbf{x} - \mu$. The exponent becomes:

$$-\frac{1}{2}(\mathbf{x}-\mu)^T M(\mathbf{x}-\mu) + \mathbf{J}^T(\mathbf{x}-\mu) = -\frac{1}{2}\left(\mathbf{x}-\mu-(M^{-1})^T\mathbf{J}\right)^T M\left(\mathbf{x}-\mu-M^{-1}\mathbf{J}\right) + \frac{1}{2}\mathbf{J}^T M^{-1}\mathbf{J}. \tag{20}$$

Similar to what we did in Eq. (13), we consider the following integral:

$$
\begin{aligned}
\mathbb{E}\left[e^{\mathbf{J}^T(\mathbf{x}-\mu)}\right] &= \int_{-\infty}^{\infty} d^N\mathbf{x}\, e^{\mathbf{J}(\mathbf{x}-\mu)}\mathcal{N}_{\mu,M}(\mathbf{x}) = \frac{1}{\sqrt{|2\pi M^{-1}|}} \int_{-\infty}^{\infty} d^N\mathbf{x}\, e^{\mathbf{J}(\mathbf{x}-\mu)}\mathcal{N}_{\mu,M}(\mathbf{x}) \\
&= \frac{1}{\sqrt{|2\pi M^{-1}|}} \int_{-\infty}^{\infty} d^N\mathbf{x}\, e^{-\frac{1}{2}\left(\mathbf{x}-\mu-(M^{-1})^T\mathbf{J}\right)^T M\left(\mathbf{x}-\mu-M^{-1}\mathbf{J}\right)+\frac{1}{2}\mathbf{J}^T M^{-1}\mathbf{J}} \\
&= e^{\frac{1}{2}\mathbf{J}^T M^{-1}\mathbf{J}} = e^{\frac{1}{2}\sum_{k,l} J_k M_{kl}^{-1} J_l}.
\end{aligned}
\tag{21}
$$

Let us first shift the variables by their mean values to simplify the notation a bit: $\mathbf{z} = \mathbf{x} - \mu$. In the $\mathbf{z}$ coordinates, we can compute the moments easily as follows:

$$
\begin{aligned}
\mathbb{E}\left[z_{i_1} z_{i_2} \cdots z_{i_{2n}}\right] &= \left[\frac{d}{dJ_{i_1}} \frac{d}{dJ_{i_2}} \cdots \frac{d}{dJ_{i_{2n}}} \mathbb{E}\left[e^{\mathbf{Jz}}\right]\right]_{\mathbf{J}=0} = \left[\frac{d}{dJ_{i_1}} \frac{d}{dJ_{i_2}} \cdots \frac{d}{dJ_{i_{2n}}} \int_{-\infty}^{\infty} d^N \mathbf{z}\, e^{\mathbf{Jz}} \mathcal{N}_{\mathbf{0},M}\right]_{\mathbf{J}=0} \\
&= \left[\frac{d}{dJ_{i_1}} \frac{d}{dJ_{i_2}} \cdots \frac{d}{dJ_{i_{2n}}} e^{\frac{1}{2}\mathbf{J}^T M^{-1} \mathbf{J}}\right]_{\mathbf{J}=0} = \left[\frac{d}{dJ_{i_1}} \frac{d}{dJ_{i_2}} \cdots \frac{d}{dJ_{i_{2n}}} \left\{\sum_{i=0}^{\infty} \frac{1}{i!} \left(\frac{1}{2} \sum_{k,l} J_k M_{kl}^{-1} J_l\right)^i\right\}\right]_{J=0} \\
&= \frac{1}{2^n n!} \frac{d}{dJ_{i_1}} \frac{d}{dJ_{i_2}} \cdots \frac{d}{dJ_{i_{2n}}} \left(\sum_{k,l} J_k M_{kl}^{-1} J_l\right)^n.
\end{aligned}
\tag{22}
$$

Note that the derivative operator will isolate the corresponding indices of the matrix M^{-1}. For example:

$$
\mathbb{E}\left[z_{i_1} z_{i_2}\right] = \frac{1}{2} \frac{d}{dJ_{i_1}} \frac{d}{dJ_{i_2}} \sum_{k,l} J_k M_{kl}^{-1} J_l = M_{i_1 i_2}^{-1},
\tag{23}
$$

where we got a factor of 2 from the differentiation which canceled the prefactor. In fact, this cancellation will hold for any n. Therefore we can write the expectation value as summation of all pairs coupled with the matrix M^{-1}:

$$
\mathbb{E}\left[z_{i_1} z_{i_2} \cdots z_{i_{2n}}\right] = \sum_{\text{pairs}} M_{p_1^1 p_2^1}^{-1} M_{p_1^2 p_2^2}^{-1} \cdots M_{p_1^n p_2^n}^{-1}
\tag{24}
$$

where p_1^k and p_2^k for a given k represent two indices paired together, and the sum is over all such possible pairs. For example, for $n = 2$ we get:

$$
\mathbb{E}\left[z_{i_1} z_{i_2} z_{i_3} z_{i_4}\right] = \sum_{\text{pairs}} M_{p_1^1 p_2^1}^{-1} M_{p_1^2 p_2^2}^{-1} = M_{i_1 i_2}^{-1} M_{i_3 i_4}^{-1} + M_{i_1 i_3}^{-1} M_{i_2 i_4}^{-1} + M_{i_1 i_4}^{-1} M_{i_2 i_3}^{-1}.
\tag{25}
$$

The formula in (24) is the Wick's theorem and it appears frequently in quantum field theory.

[1] J. Polchinski, "Memories of a theoretical physicist." 2006 [Online]. Available: http://arxiv.org/abs/1708.09093
[2] "The probability integral." [Online]. Available: https://www.york.ac.uk/depts/maths/histstat/normal_history.pdf. [Accessed: 24-Jan-2022]
[3] D. A. Roberts, S. Yaida, and B. Hanin, "The Principles of Deep Learning Theory," Jun. 2021 [Online]. Available: http://arxiv.org/abs/2106.10165