

## Not to be turned in: Computing with Gaussian probability distributions

This is not an assigned problem set, and nothing from this needs to be turned in. It is just a set of problems you might want to work through to understand computations involving multi-dimensional Gaussian distributions. It comes from an old problem set, so it reads like a problem set, but this is just for you. It goes through the material taught in class (and so some of it should be well known to you and just need skimming), but goes beyond it to show how to do the integrals to compute arbitrary moments of a Gaussian as well as to normalize the distribution. If you aim to be a theorist, you should know this material.

The things to actually do are in red, the rest is explanation or guidance.

---

Basic to many aspects of theory is understanding and being able to compute with Gaussian distributions, so as part of this homework we'll take some time to do basic computations on Gaussians. Even where the homework doesn't require you to do anything, you should take the time to do whatever you need to do to prove to yourself or satisfy yourself that what's stated is true.

**Notation:** Where I use vectors and matrices, I will use boldface small letters, *e.g.*  $\mathbf{x}$  to represent vectors, and boldface capital letters, *e.g.*  $\mathbf{C}$ , to represent matrices.  $\mathbf{x}$  is a column vector; its transpose  $\mathbf{x}^T$  is a row vector. Plain letters represent numbers, *e.g.*  $x$  is a scalar; the  $i^{\text{th}}$  element of the vector  $\mathbf{x}$  is  $x_i$ ; the  $ij^{\text{th}}$  element of the matrix  $\mathbf{C}$  is  $C_{ij}$ . Thus  $\mathbf{xy}^T$  is a matrix with  $ij^{\text{th}}$  element equal to  $x_i y_j$ ; while  $\mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i$  (for vectors of real numbers).

## One-dimensional Gaussian distributions and integrals:

You know the Gaussian distribution for a random variable  $z$  with mean  $m$  and standard deviation  $\sigma$ :

$$P(z) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2} \frac{(z-m)^2}{\sigma^2}} \quad (1)$$

Compute the properties of the one-dimensional Gaussian distribution:

- a. Show that the probability distribution is correctly normalized:  $\int_{-\infty}^{\infty} P(z) dz = 1$ . Letting  $p(z) = e^{-\frac{1}{2} \frac{(z-m)^2}{\sigma^2}}$ , you want to show that  $\int_{-\infty}^{\infty} p(z) dz = (2\pi\sigma^2)^{1/2}$ . To do this (1) Change variables  $z \rightarrow x = z - m$  and note that the value of the integral is unchanged; (2) Take the square of the integral:  $\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy p(x)p(y)$  (where  $x$  and  $y$  both have zero mean); (3) Consider  $x$  and  $y$  as the coordinates of a 2-d space, and change variables to polar coordinates  $r, \theta$ :  $r = \sqrt{x^2 + y^2}$ ,  $\theta = \tan^{-1}(y/x)$ , where  $r$  goes from 0 to  $\infty$  and  $\theta$  goes from 0 to  $2\pi$ . Recall that  $dx dy = r dr d\theta$ . (4) Do the  $\theta$  integral; (5) Transform variables  $r \rightarrow p = r^2$ . Note that  $r dr = \frac{1}{2} dp$ . (6) Do the  $p$  integral.

- b. Compute the first few moments of  $z$ . The  $n^{\text{th}}$  moment of a random variable  $z$  is the average of  $z^n$ , which we'll denote  $\langle z^n \rangle$  and which is equal to  $\int_{-\infty}^{\infty} z^n P(z) dz$ . Define  $x = z - m$ .

The first moment  $\langle z \rangle$ , is the mean. To compute the mean, note that  $\langle x \rangle = \langle z \rangle - m$  (note that  $m$  is a constant so that  $m = \langle m \rangle$ ), and show by symmetry that  $\langle x \rangle = 0$  (for example, show that  $\langle -x \rangle = \langle x \rangle$ ).

For higher moments, it's simpler to use the centered moments, the moments of  $x = z - m$ . The second centered moment is the variance,  $\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 P(x) dx$ ; the third centered moment is called the *skew*, and the fourth the *kurtosis*. To compute these moments, we use the fact that  $y = \frac{x}{\sigma}$  is a zero-mean Gaussian variable with unit variance. Thus, if we can compute the moments of  $y$ , e.g.,  $\langle y^n \rangle$ , then we can compute the moments of  $z$ , e.g. by expanding the polynomial in the expression  $\langle (\frac{z-m}{\sigma})^n \rangle = \langle y^n \rangle$ , and noting that  $m$  and  $\sigma$  are constants that can be taken out of the averages. For example, multiplying both sides of the equation by  $\sigma^n$  tells us immediately that the centered moments are  $\langle x^n \rangle = \sigma^n \langle y^n \rangle$ ; and expanding for  $n = 2$  shows that  $\langle z^2 \rangle = m^2 + \sigma^2 \langle y^2 \rangle$  (verify all this for yourself).

To compute the moments of  $y$ , a useful trick<sup>1</sup> is to compute  $\langle e^{ky} \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dy e^{-\frac{1}{2}y^2 + ky}$ . Then we can compute any moment as  $\langle y^n \rangle = \frac{d^n}{dk^n} \langle e^{ky} \rangle |_{k=0}$ , since each application of  $\frac{d}{dk}$  pulls down one factor of  $y$  into the integral defining  $\langle e^{ky} \rangle$ , and doing this  $n$  times and then setting  $k = 0$  gives the integral that defines  $\langle y^n \rangle$ .

To compute  $\langle e^{ky} \rangle$ , you complete the square in the exponential by subtracting and adding  $\frac{1}{2}k^2$ ; you'll end up with  $e^{(y-k)^2}$ , which you can integrate by using the change of variable  $(y - k) \rightarrow y$ ; times  $e^{\frac{1}{2}k^2}$ , which comes out of the integral, giving the answer  $e^{\frac{1}{2}k^2}$ . Carry out these steps if they're not familiar to you.

Use all of this apparatus to show that  $\langle z \rangle = m$ ;  $\langle x^2 \rangle = \sigma^2$ ;  $\langle x^3 \rangle = 0$ ;  $\langle x^4 \rangle = 3\sigma^4$ .

Note: since  $y = x/\sigma$ , we have also established that  $\langle e^{kx/\sigma} \rangle = e^{\frac{1}{2}k^2}$ . Letting  $p = k/\sigma$ , this in turn gives  $\langle e^{px} \rangle = e^{\frac{1}{2}p^2\sigma^2}$ . We could have computed this directly by completing the square in the exponential of  $P(x)$  rather than  $P(y)$ . By the same reasoning as above,  $\langle x^n \rangle = \frac{d^n}{dp^n} \langle e^{px} \rangle |_{p=0}$ . This gives the same answers as  $\langle x^n \rangle = \sigma^n \langle y^n \rangle$ .

---

<sup>1</sup>This works for a Gaussian, but for other distributions  $\int dy e^{ky} P(y)$  may not converge. A more general trick is to compute  $\langle e^{iky} \rangle$  (where  $i = \sqrt{-1}$ ), which is the Fourier transform of  $P(y)$  and is called the *characteristic function* of  $y$ ; and then the moments can be computed as  $\langle y^n \rangle = \frac{1}{i^n} \frac{d^n}{dk^n} \langle e^{iky} \rangle |_{k=0}$ .

## Multi-dimensional Gaussian distributions:

The general form for a Gaussian distribution of an  $N$  – dimensional random variable  $\mathbf{s}$  with mean  $\mathbf{m}$  and covariance matrix  $\mathbf{C}$  (meaning that  $C_{ij} = \langle (s_i - m_i)(s_j - m_j) \rangle$ ; in particular, note that  $\mathbf{C}$  is a symmetric matrix), is as follows:

$$P(\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^N \text{Det } \mathbf{C}}} e^{-\frac{1}{2}(\mathbf{s}-\mathbf{m})^T \mathbf{C}^{-1}(\mathbf{s}-\mathbf{m})} \quad (2)$$

( $\mathbf{C}^{-1}$  means the matrix inverse of  $\mathbf{C}$ ). The determinant of a matrix  $\mathbf{C}$  is the product of its eigenvalues. (The matrix  $\mathbf{C}$  can be any symmetric matrix that is positive definite, meaning that all of its eigenvalues are greater than 0; as we will see later, the eigenvalues correspond to variances, which must be positive.<sup>2</sup> Given this assumption, we'll show that  $\mathbf{C}$  is in fact the covariance matrix of  $\mathbf{s}$ .) Verify that for  $N = 1$  this reduces to the formula for a one-dimensional Gaussian distribution (for  $N = 1$ , the determinant is just the value of the matrix's single element).

We again will compute the properties of this distribution. To do so, you need to understand the following (please review any of this material that is not clear to you; my online math notes are one source):

- a. Because  $\mathbf{C}$  is a symmetric matrix, it has a complete basis of real orthonormal eigenvectors  $\mathbf{e}_\mu$  with corresponding real eigenvalues  $\lambda_\mu$ ,  $\mu = 1, \dots, N$  (I'll use Greek letters, *e.g.*  $\mu, \nu$ , to label the eigenvector basis, and Roman letters, *e.g.*  $i, j$ , for indices in the original basis). This means that  $\mathbf{C}\mathbf{e}_\mu = \lambda_\mu\mathbf{e}_\mu$  for all  $\mu$  and  $\mathbf{e}_\mu \cdot \mathbf{e}_\nu = \delta_{\mu\nu}$  for all  $\mu$  and  $\nu$  ( $\delta_{\mu\nu}$  is the Kronecker delta function, equal to 1 if  $\mu = \nu$  and 0 otherwise).
- b. Let  $\mathbf{U}$  be the matrix whose columns are the eigenvectors. Then  $\mathbf{U}^T\mathbf{U} = \mathbf{1}$  where  $\mathbf{1}$  is the identity matrix (to see this, note that the  $\mu\nu^{\text{th}}$  element of  $\mathbf{U}^T\mathbf{U} = \mathbf{1}$  is the dot product of the  $\mu^{\text{th}}$  row of  $\mathbf{U}^T$  and the  $\nu^{\text{th}}$  column of  $\mathbf{U}$ , or in other words it is  $\mathbf{e}_\mu \cdot \mathbf{e}_\nu$  and thus is  $\delta_{\mu\nu}$ ). That is,  $\mathbf{U}^T$  is the inverse of  $\mathbf{U}$ , so also  $\mathbf{U}\mathbf{U}^T = \mathbf{1}$ . This is the definition of an *orthogonal* matrix, which represents a rigid rotation and/or reflection without any stretching – all angles between vectors and lengths of vectors are preserved under transformation by  $\mathbf{U}$ . (To show this, it suffices to show that  $\mathbf{U}$  preserves all dot products, since  $\text{length}^2$  is the dot product of a vector with itself, while the cosine of the angle between two vectors is their dot product divided by the product of their lengths. This is easy to show: for any vectors  $\mathbf{x}$  and  $\mathbf{y}$ ,  $\mathbf{U}\mathbf{x} \cdot \mathbf{U}\mathbf{y} = (\mathbf{U}\mathbf{x})^T\mathbf{U}\mathbf{y} = \mathbf{x}^T\mathbf{U}^T\mathbf{U}\mathbf{y} = \mathbf{x}^T\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ .)

---

<sup>2</sup>Any covariance matrix is positive definite if the distribution includes the whole space: it is of the form  $\langle \mathbf{x}\mathbf{x}^T \rangle$ , so for any eigenvector  $\mathbf{e}_i$  with eigenvalue  $\lambda_i$ ,  $\mathbf{e}_i^T \langle \mathbf{x}\mathbf{x}^T \rangle \mathbf{e}_i = \mathbf{e}_i^T (\lambda_i \mathbf{e}_i) = \lambda_i$ , but also  $\mathbf{e}_i^T \langle \mathbf{x}\mathbf{x}^T \rangle \mathbf{e}_i = \langle (\mathbf{e}_i \cdot \mathbf{x})^2 \rangle > 0$ .

- c. To transform to the eigenvector basis, we take any vector  $\mathbf{x} \rightarrow \mathbf{U}^T \mathbf{x}$ . If we write  $\mathbf{x}$  in the eigenvector basis as  $\mathbf{x} = \sum_{\mu} x_{\mu} \mathbf{e}_{\mu}$ , then  $x_{\mu} = \mathbf{e}_{\mu} \cdot \mathbf{x}$ . The  $\mu^{\text{th}}$  element of  $\mathbf{U}^T \mathbf{x}$  is  $\mathbf{e}_{\mu} \cdot \mathbf{x} = x_{\mu}$ , so  $\mathbf{U}^T \mathbf{x}$  gives the elements of  $\mathbf{x}$  in the eigenvector basis.
- d. As we'll see, the eigenvalues  $\lambda_{\mu}$  are the variances along the direction of  $\mathbf{e}_{\mu}$ , so let's use  $\sigma_{\mu}^2$  for the eigenvectors:  $\sigma_{\mu}^2 \equiv \lambda_{\mu}$ . The statement  $\mathbf{C} \mathbf{e}_{\mu} = \sigma_{\mu}^2 \mathbf{e}_{\mu}$  for  $\mu = 1, \dots, N$  is equivalent to  $\mathbf{C} \mathbf{U} = \mathbf{U} \Sigma$  where  $\Sigma$  is the diagonal matrix whose diagonal elements are the eigenvalues,  $\Sigma_{\mu\nu} = \delta_{\mu\nu} \sigma_{\mu}^2$ . That is,  $\mathbf{C}$  acts on the  $\mu^{\text{th}}$  column of  $\mathbf{U}$ , which is  $\mathbf{e}_{\mu}$ , to give  $\sigma_{\mu}^2$  times that column (if this is not clear: in indices,  $\mathbf{C} \mathbf{U} = \mathbf{U} \Sigma$  becomes  $\sum_j C_{ij} U_{j\mu} = \sum_j U_{ij} \Sigma_{j\mu} = \sum_j U_{ij} \delta_{j\mu} \sigma_{\mu}^2 = U_{i\mu} \sigma_{\mu}^2$ ; since  $U_{j\mu}$  is the  $j^{\text{th}}$  element of  $\mathbf{e}_{\mu}$ , this can be rewritten as  $C_{ij} (\mathbf{e}_{\mu})_j = (\mathbf{e}_{\mu})_i \sigma_{\mu}^2$ , which is the index form of the vector equation  $\mathbf{C} \mathbf{e}_{\mu} = \sigma_{\mu}^2 \mathbf{e}_{\mu}$ ). We can rewrite this as  $\mathbf{C} = \mathbf{U} \Sigma \mathbf{U}^T$  or  $\Sigma = \mathbf{U}^T \mathbf{C} \mathbf{U}$ .  $\Sigma$  is the form  $\mathbf{C}$  takes in the eigenvector basis, that is, when vectors are transformed by  $\mathbf{U}^T$ .

Similarly,  $\Sigma^{-1} = \mathbf{U}^T \mathbf{C}^{-1} \mathbf{U}$ . Since  $\Sigma$  is diagonal, we know how to compute the matrix inverse  $\Sigma^{-1}$ , we simply invert each diagonal entry: it is the diagonal matrix with the inverse eigenvalues on the diagonal. You can see this in at least two ways (1) The inverse is preserved under change of basis;  $\Sigma$  is the representation of  $\mathbf{C}$  in the eigenvector basis, so  $\Sigma^{-1}$  is the representation of  $\mathbf{C}^{-1}$  in the eigenvector basis. (2) Equivalently, a definition of the inverse  $\mathbf{C}^{-1}$  is that  $\mathbf{C}^{-1} \mathbf{e}_{\mu} = \frac{1}{\sigma_{\mu}^2} \mathbf{e}_{\mu}$  for  $\mu = 1, \dots, N$ . This leads to the same equations as for  $\mathbf{C}$  except with inverse eigenvalues replacing eigenvalues.

Then:

- a. Show that in the eigenvector basis,  $\mathbf{v} = \mathbf{U}^T \mathbf{s}$ , the probability distribution becomes a product of independent one-dimensional Gaussian distributions, one distribution for the component along each eigenvector, with variance equal to the corresponding  $\sigma_{\mu}^2$ :

$$P(\mathbf{v}) = \prod_{\mu=1}^N \frac{1}{(2\pi\sigma_{\mu}^2)^{\frac{1}{2}}} e^{-\frac{1}{2} \frac{(v_{\mu} - \langle v_{\mu} \rangle)^2}{\sigma_{\mu}^2}} \quad (3)$$

(Note: the notation  $\prod_{\mu=1}^N$  means multiply the terms together for all  $\mu$ 's from 1 to  $N$ ; much like  $\sum_{\mu=1}^n$  means sum together the terms for all the  $\mu$ 's.)

To show this, note that  $(\mathbf{s}^T - \langle \mathbf{s}^T \rangle) \mathbf{C}^{-1} (\mathbf{s} - \langle \mathbf{s} \rangle) = (\mathbf{v}^T - \langle \mathbf{v}^T \rangle) \Sigma^{-1} (\mathbf{v} - \langle \mathbf{v} \rangle)$ , which can be shown for example by inserting  $\mathbf{1} = \mathbf{U} \mathbf{U}^T$  on each side of  $\mathbf{C}^{-1}$  in the left term (note that since  $\mathbf{v} = \mathbf{U}^T \mathbf{s}$  for each instance of  $\mathbf{v}$  and  $\mathbf{s}$ , the same relationship also holds for their averages). You also need to remember that under transformation of variables, probability distributions transform as  $P(\mathbf{s}) d^N \mathbf{s} = P(\mathbf{v}) d^N \mathbf{v}$ . If  $\mathbf{v} = \mathbf{U}^T \mathbf{s}$ , the differentials transform in turn as  $d^N \mathbf{v} = |\text{Det}(\mathbf{U}^T)| d^N \mathbf{s}$ . Finally, all of the eigenvalues

of an orthogonal matrix must have absolute value 1, since the matrix does not change the length of any vector, so the determinant of an orthogonal matrix has absolute value 1.

Use this to show that the multi-dimensional Gaussian as written above is correctly normalized,  $\int d^N \mathbf{s} P(\mathbf{s}) = 1$ .

- b. **Set up the framework for computing the moments, and use to compute  $\langle x_i x_j \rangle$ .** A simple way is to work in the eigenvector basis using zero-mean variables. Set  $\mathbf{x} = \mathbf{s} - \langle \mathbf{s} \rangle$ ,  $\mathbf{t} = \mathbf{U}^T \mathbf{x} = \mathbf{v} - \langle \mathbf{v} \rangle$ , and show that  $\langle e^{\mathbf{k} \cdot \mathbf{t}} \rangle = \langle e^{k_1 t_1 + k_2 t_2 + \dots + k_N t_N} \rangle = \langle e^{k_1 t_1} \rangle \langle e^{k_2 t_2} \rangle \dots \langle e^{k_N t_N} \rangle = e^{\frac{1}{2} k_1^2 \sigma_1^2} e^{\frac{1}{2} k_2^2 \sigma_2^2} \dots e^{\frac{1}{2} k_N^2 \sigma_N^2} = e^{\frac{1}{2} \mathbf{k}^T \Sigma \mathbf{k}}$ . Rewrite this as  $\langle e^{\mathbf{k}^T \mathbf{U}^T \mathbf{x}} \rangle = e^{\frac{1}{2} \mathbf{k}^T \Sigma \mathbf{k}}$ . Let  $\mathbf{p} = \mathbf{U} \mathbf{k}$ , to find  $\langle e^{\mathbf{p} \cdot \mathbf{x}} \rangle = e^{\frac{1}{2} \mathbf{p}^T \mathbf{U} \Sigma \mathbf{U}^T \mathbf{p}} = e^{\frac{1}{2} \mathbf{p}^T \mathbf{C} \mathbf{p}}$ .

The moments can then be found as follows: write an arbitrary product of elements of  $\mathbf{x}$  as  $x_i x_j \dots x_k$ . Then  $\langle x_i x_j \dots x_k \rangle = \frac{d}{dp_i} \frac{d}{dp_j} \dots \frac{d}{dp_k} \langle e^{\mathbf{p} \cdot \mathbf{x}} \rangle |_{\mathbf{p}=0} = \left( \frac{d}{dp_i} \frac{d}{dp_j} \dots \frac{d}{dp_k} e^{\frac{1}{2} \mathbf{p}^T \mathbf{C} \mathbf{p}} \right) |_{\mathbf{p}=0}$ .

Use this to show that  $\mathbf{C}$  is indeed the covariance of the distribution:  $\langle x_i x_j \rangle = C_{ij}$ .

For the bold: use this to prove, or at least get a feel for, Wick's theorem. The average of a product of  $N$  elements of  $\mathbf{x}$  is called an  $N$ -point function. Wick's theorem states that, for a Gaussian distribution, any  $N$ -point function for  $N$  even is equal to a sum, over all distinct ways of grouping the  $N$  elements into pairs, of the products of the two-point functions of the pairs (and for  $N$  odd, the  $N$ -point function is zero). For example:  $\langle x_1 x_2 x_3 x_4 \rangle = \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle = C_{12} C_{34} + C_{13} C_{24} + C_{14} C_{23}$ .