4 Gaussian distributions

Narayana Santhanam

The most important distribution we will consider this term is the Gaussian model. You have seen this before in EE 342, but perhaps have not been exposed to the full picture about why this is important, other than the Central Limit Theorem. We will go a little more in depth in this course. We will also pay some attention to the concept of expectations in this section, and in the process understand a little more about random variables.

4.1 Univariate Gaussians, expectation and the Moment Generating function

A univariate standard Gaussian (also called *normal*) random variable Z takes values between $-\infty$ and ∞, satisfies

$$\mathbb{E}Z = 0$$

and

$$\operatorname{var}(Z) \stackrel{\text{def}}{=} \mathbb{E}[Z^2] - (\mathbb{E}Z)^2 = 1,$$

and putting the two together, we also get $\mathbb{E}[Z^2] = 1$.

Remark It is common to write $\mathbb{E}[Z^2]$ (expectation of Z^2) as simply $\mathbb{E}Z^2$, *i.e.,* the expectation operator simply operates on whatever follows it (here Z^2). If we wanted to write "square of the expectation of Z", we would write $(\mathbb{E}Z)^2$. For an arbitrary random variable W: (expectation of W^2, $\mathbb{E}W^2$) and (square of the expectation of W, $(\mathbb{E}W)^2$) are obviously not the same, in fact in general

$$\mathbb{E}W^2 \geq (\mathbb{E}W)^2$$

with equality iff W is a constant random variable (*i.e.,* a random variable that takes on only one value with probability 1). □

The pdf of the standard Gaussian (normal) is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \qquad z \in \mathbb{R},$$

where we use $\mathbb{R}$ to denote the set of all real numbers. We often denote this pdf by $\mathcal{N}(0,1)$, where $\mathcal{N}$ stands for normal, the first parameter in the paranthesis (here, 0) denotes the expectation and the second parameter (here, 1) the variance.

The indefinite integral of $e^{-z^2/2}$ does not exist in closed form, but the definite integrals

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi},$$

(do you know how to prove the above?) and from the fact that $\mathbb{E}Z^2 = 1$, we must have

$$\int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = \sqrt{2\pi}.$$

By the end of the section you will know more definite integrals related to the Gaussian.

If Z is a standard Gaussian, then $\sigma Z + \mu$ is a Gaussian random variable too, with mean μ and variance σ^2. Its pdf (which you can obtain by the standard argument of transformation of random variables) is

$$f_W(w) = \frac{1}{\sqrt{2\pi}} \exp(-(w-\mu)^2/2), \qquad w \in \mathbb{R}.$$

We of course now deduce from the fact that the pdf must integrate to 1 that

$$\int_{-\infty}^{\infty} \exp(-(w-\mu)^2/2) = \sqrt{2\pi},$$

from the fact that $\mathbb{E}W = \mu$ that

$$\int_{-\infty}^{\infty} w \exp(-(w-\mu)^2/2) = \mu\sqrt{2\pi},$$

and from the fact that $\mathbb{E}W^2 = \mathrm{var}(W) + (\mathbb{E}W)^2 = \sigma^2 + \mu^2$ that

$$\int_{-\infty}^{\infty} w^2 \exp(-(w-\mu)^2/2) = (\sigma^2 + \mu^2)\sqrt{2\pi}.$$

To gain more insights into random variables in general and the Gaussian random variable in particular, let us study the moment generating function of Z,

$$\mathbb{E}\exp(tZ)$$

where $\mathbb{E}$ denotes expectation. We went over the expectation of a function of a random variable in class, and concluded

$$\mathbb{E}\exp(tZ) = \int_{-\infty}^{\infty} e^{tz} f_Z(z) dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz - z^2/2} dz = e^{t^2/2}$$

where the last equality follows by completing the square in the exponent,

$$tz - z^2/2 = -(z-t)^2/2 + t^2/2,$$

followed by using~(**??**).

To see why the moment generating function is useful, observe the Taylor series expansion

$$e^{tZ} = 1 + tZ + \ldots + \frac{t^n Z^n}{n!} + \ldots = \sum_{t=0}^{\infty},$$

from which we get that

$$\mathbb{E}e^{tZ} = 1 + t\mathbb{E}Z + \ldots + \frac{t^n \mathbb{E}Z^n}{n!} + \ldots,$$

The quantity $\mathbb{E}Z^n$ is the n'th moment of Z (so the expectation $\mathbb{E}Z$ is also called the first moment. The moment generating function effectively determines the pdf (we will not qualify this formally, but this would be true of most distributions we will encounter), so knowing all moments of Z tells you everything you need to know about the pdf of Z. In that sense, one could consider the expectation $\mathbb{E}Z$ and variance of Z to be first and second order detail about the random variable.

For the Gaussian case, observe that

$$e^{t^2/2} = \sum_{t=0}^{\infty} \frac{(t^2/)^n}{n!} = \sum_{t=0}^{\infty} \frac{t^{2n}}{2^n n!} = \sum_{k=0}^{\infty} \frac{t^{2k}(2k)!}{2^k k!(2k)!},$$

implying that if $Z \sim \mathcal{N}(0,1)$, then for all n

$$\mathbb{E}Z^n = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \frac{(2k)!}{2^k k!} & \text{if } n = 2k. \end{cases}$$

Verify that if $W \sim \mathcal{N}(\mu, \sigma^2)$, then the moment generating function of W is $\exp(t\mu + \frac{1}{2}\sigma^2 t^2)$. We will need this in the multivariate case.

4.2 Multivariate Gaussians (pdf)

Suppose $X_1, \ldots, X_k$ are continuous random variables, and each of X_i is Gaussian. This setup by itself is not that useful, it needs an additional component. We define $X_1, \ldots, X_k$ to be a multivariate Gaussian (also jointly Gaussian, or sometimes, just Gaussian) if all linear combinations $a_1X_1 + \cdots + a_kX_k$, $a_i \in \mathbb{R}$ happen to be univariate Gaussian random variables.

We will show below a counterexample of two Gaussian random variables X_1 and X_2 that are *not* jointly/multivariate Gaussian. We will then show that independent Gaussians are multivariate Gaussian. Finally, we will derive the pdf of multivariate Gaussians.

Counterexample Before we unravel this definition, let us look at the following counterexample. $X_1 \sim N(0,1)$ is a standard Gaussian, and W is a Bernoulli 1/2 random variable independent of X_1. Let $X_2 = (2W-1)X_1$, namely if $W = 1$, $X_2 = X_1$ and if $W = 0$, $X_2 = -X_1$. Then $X_2 \sim N(0,1)$ (prove it). But X_1 and X_2 are not jointly Gaussian, because $X_1 + X_2$ assigns the number 0 probabilty 1/2. Now, Gaussian pdfs can be degenerate (when variance =0, the pdf assigns probability 1 to the mean). But no pdf, leave alone a Gaussian one, can ever assign probability 1/2 to any element of its support.

Independent normals If $X_1, \ldots, X_k$ are independent normal variables, $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then they are jointly Gaussian. Then for all $a_1, \ldots, a_k$, $a_i \in \mathbb{R}$, we have the moment generating function of $a_1X_1 + a_2X_2 + \ldots + a_kX_k$ to be

$$\begin{aligned} M(t) &= \mathbb{E} \exp\left(t(a_1X_1 + a_2X_2 + \ldots + a_kX_k)\right) \\ &= \prod_{i=1}^{k} \mathbb{E} \exp\left(ta_iX_i\right) \\ &= \prod_{i=1}^{k} \exp\left(ta_i\mu_i + \frac{1}{2}t^2 a_i^2 \sigma_i^2\right) \\ &= \exp\left(t\left(\sum_i a_i\mu_i\right) + \frac{1}{2}t^2\left(\sum_j a_j^2\sigma_j^2\right)\right) \end{aligned} \tag{1}$$

where the second equality above is because X_i are independent. The moment generating function $M(t)$ calculated above is simply the moment generating function of a Gaussian random variable with mean $\sum_i a_i\mu_i$ and variance $\sum_j a_j^2\sigma_j^2$, and thus we conclude $a_1X_1 + a_2X_2 + \ldots + a_kX_k$ is Gaussian. Therefore, independent normals are also jointly Gaussian.

Of course, we know the joint pdf of independent normals $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$

from the independence property,

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \prod_j f_{X_j}(x_j) = \frac{1}{(2\pi)^{k/2}\prod_j \sigma_j} \exp\left(-\sum_i \frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right).$$

It is useful to rewrite it in a slightly different way. Let

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix},$$

and let

$$\mu = \mathbb{E}\mathbf{X} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}, \text{ and } cov(\mathbf{X}) = \mathbb{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T] \stackrel{\text{def}}{=} K,$$

where K is a diagonal matrix whose diagonal entries are $\sigma_1^2,\ldots,\sigma_k^2$. The determinant of K, $|K| = \prod_i \sigma_i$ and K^{-1} is a diagonal matrix whose entries are $\frac{1}{\sigma^2},\ldots,\frac{1}{\sigma_k^2}$. Using these observations, we can rewrite the pdf of $\mathbf{X}$, using the notation $\mathbf{x} = (x_1,\ldots,x_k)$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|K|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T K^{-1}(\mathbf{x}-\mu)\right). \quad (2)$$

4.2.1 PDF of Jointly Gaussian random variables

We will now derive the pdf of joint/multivariate Gaussian random from its definition. The route to the pdf here will be via the moment generating function and the fact that we know the pdf of independent (and therefore automatically multivariate) Gaussian random variables. In what follows, we assume

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix},$$

and that

$$\mathbb{E}\mathbf{X} = \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}, \text{ and } cov(\mathbf{X}) = \mathbb{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T] = K$$

for some positive definite matrix K. Review positive definite matrices from module in EE345.

Moment Generating function Generalizing the univariate case, the MGF of a multivariate random vector $\mathbf{X} = (X_1, \dots, X_k)$ is defined to be function $M : \mathbb{R}^k \to \mathbb{R}$, where for any $\mathbf{z} \in \mathbb{R}^k$

$$M_{\mathbf{X}}(\mathbf{z}) = \mathbb{E} \exp(\mathbf{z}^T \mathbf{X}).$$

MGF of jointly Gaussian random variables By definition of multivariate Gaussian vectors, we know that an linear combination of the components of $\mathbf{X}$ is Gaussian.

Show that for all $\mathbf{z} \in \mathbb{R}^k$, $\mathbf{z}^T\mathbf{X}$ is univariate Gaussian, with mean equal to $\mathbf{z}^T\mu$ and variance $\mathbf{z}^T K \mathbf{z}$.

Given the above result, note that $M(\mathbf{z})$, the moment generating function for $\mathbf{X}$ is simply the moment generating function of the univariate Gaussian, $\mathbf{z}^T\mathbf{X}$, evaluated at $t = 1$. Using the result for mgfs of univariate Gaussians, we therefore get

$$M_{\mathbf{X}}(\mathbf{z}) = M_{\mathbf{z}^T\mathbf{X}}(1) = \exp\left(\mathbf{z}^T\mu + \frac{1}{2}\mathbf{z}^T K \mathbf{z}\right).$$

We are almost home. Since K is symmetric, we can use its spectral decomposition to write $K = V \Lambda V^T$, where V is a unitary matrix (ie V is a square invertible matrix whose inverse is V^T) and Λ is a diagonal vector with the eigenvalues of K on its diagonal. The eigenvalues are all > 0 since K is positive definite, so Λ^{-1} exists as well.

Let $\mathbf{Y} = V^T(\mathbf{X} - \mu)$ so that $\mathbf{X} = V\mathbf{Y} + \mu$. Note that $\mathbb{E}Y = \mathbf{0}$ and $cov(Y) = \Lambda$. $\mathbf{Y}$ is a multivariate Gaussian as well (why?) and its mgf is therefore for any $\mathbf{z} = (z_1, \dots, z_k)$,

$$M_{\mathbf{Y}}(\mathbf{z}) = \exp\left(\frac{1}{2} z^T \Lambda \mathbf{z}\right) = \exp\left(\frac{1}{2}\sum_j z_i^2 \lambda_i\right),$$

which from~(1) is exactly the mgf of independent $\mathcal{N}(0, \lambda_i)$ Gaussians. Since mgfs are uniquely associated with pdfs, we have that the components of $\mathbf{Y}$ are independent $\mathcal{N}(0, \lambda_i)$ Gaussians, and the pdf of $\mathbf{Y}$ is from~(2), for any $\mathbf{y} \in \mathbb{R}^k$,

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{2\pi}^{k/2} |\Lambda|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{y}^T \Lambda^{-1} \mathbf{y}\right).$$