Mathematical Toolkit **Autumn 2023**

Lecture 16: November 16, 2023

Lecturer: Madhur Tulsiani

1 Gaussian Random Variables

A Gaussian random variable X is defined through the density function

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where μ is its mean and σ^2 is its variance, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$. To see the definition gives a valid probability distribution, we need to show $\int_{-\infty}^{\infty} \gamma(x)dx = 1$. It suffices to show for the case that $\mu = 0$ and $\sigma^2 = 1$. First we show the integral is bounded.

Claim 1.1 *$I = \int_{-\infty}^{\infty} e^{-x^2/2}dx$ is bounded.*

Proof: We see that

$$I = \int_{-\infty}^{\infty} e^{-x^2/2}dx = 2\int_{0}^{\infty} e^{-x^2/2}dx \le 2\int_0^2 1dx + 2\int_2^{\infty} e^{-x}dx = 4 + 2e^{-2},$$

where we use the fact that I is even and after $x = 2$, $e^{-x^2/2}$ is upper bounded by e^{-x}. ∎

Next we show that the normalization factor is $\sqrt{2\pi}$.

Claim 1.2 *$I^2 = 2\pi$.*

Proof:

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}dx \int_{-\infty}^{\infty} e^{-y^2/2}dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy \\
&= \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r dr d\theta \quad (\text{let } x = r\cos\theta \text{ and } y = r\sin\theta) \\
&= 2\pi \int_0^{\infty} e^{-s}ds \quad (\text{let } s = r^2/2) \\
&= 2\pi\,.
\end{aligned}$$

∎

This completes the proof that the definition gives a valid probability distribution. Before proceeding to applications of Gaussian random variables, we prove the following fact which we will use repeatedly.

Proposition 1.3 *Let $Z = c_1X_1 + c_2X_2$, where $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ are independent. Then $Z \sim \mathcal{N}(0, c_1^2 + c_2^2)$.*

Proof: By a simple change of variable, we can check that the density function for c_1X_1 is $\frac{1}{\sqrt{2\pi}|c_1|}e^{-\frac{x^2}{2c_1^2}}$, which shows that $c_1X_1 \sim \mathcal{N}(0, c_1^2)$, and similarly $c_2X_2 \sim \mathcal{N}(0, c_2^2)$.
Next, we can check that if X and Y are independent random variables with densities f and g, then for $Z = X + Y$, we have

$$\mathbb{P}\left[Z \le t\right] \;=\; \int_{\infty}^{t} \left(\int_{-\infty}^{\infty} f(x) \cdot g(z-x)dx \right) dz\,,$$

which gives the density of Z as $h(z) = \int_{-\infty}^{\infty} f(x) \cdot g(z-x)dx$. Taking $X = c_1X$ and $Y = c_2X_2$, we get the density of $Z = c_1X_1 + c_2X_2$ is

$$h(z) \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,|c_1|} \cdot e^{-\frac{x^2}{2c_1^2}} \cdot \frac{1}{\sqrt{2\pi}\,|c_2|} \cdot e^{-\frac{(z-x)^2}{2c_2^2}}\, dx\,.$$

We leave it as an exercise to show that the above integral gives

$$h(z) \;=\; \frac{1}{\sqrt{2\pi(c_1^2 + c_2^2)}} \cdot e^{-\frac{z^2}{2(c_1^2+c_2^2)}}\,,$$

which implies $c_1X_1 + c_2X_2 \sim \mathcal{N}(0, c_1^2 + c_2^2)$. ■

One can obtain the following corollary using an inductive application of the above proposition.

Corollary 1.4 *Let $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$ be independent standard Gaussian random variables. Then, for any vector of coefficients $c = (c_1, \ldots, c_n)$, we have*

$$Z \;=\; c_1X_1 + \cdots + c_nX_n \;\sim\; \mathcal{N}(0, \|c\|^2)\,,$$

where $\|c\|^2 = c_1^2 + \cdots + c_n^2$.

Remark 1.5 *Note that we need independence for general statements of the form "linear combination of Gaussians is a Gaussian", and that the statement can fail when the Gaussians are not independent. For example, consider the random variables*

$$X_1 \sim \mathcal{N}(0,1) \qquad \textit{and} \qquad X_2 = \begin{cases} X_1 & \textit{if } |X_1| \le 1 \\ -X_1 & \textit{if } |X_1| > 1 \end{cases}$$

One can check that if we look at the variables X_1 and X_2, they are Gaussian random variables with mean 0 and variance 1. However, we have that

$$X_1 + X_2 \;=\; \begin{cases} 2X_1 & \textit{if } |X_1| \leq 1 \\ 0 & \textit{otherwise} \end{cases} .$$

Thus, the linear combination is zero with positive probability, and is not a Gaussian distribution.

When a collection of Gaussian random variables $X_1, \ldots, X_n$ satisfies that their linear combinations are also Gaussian, they are called "jointly Gaussian random variables. Thus, we proved that inpdendent Gaussian random variables are also jointly Gaussian.

We prove a useful lemma for later use.

Lemma 1.6 *For $X \sim \mathcal{N}(0,1)$ and $\lambda \in (0, 1/2)$,*

$$\mathbb{E}\left[e^{\lambda \cdot X^2}\right] = \frac{1}{\sqrt{1-2\lambda}} .$$

Proof:

$$\begin{aligned} \mathbb{E}\left[e^{\lambda \cdot X^2}\right] &= \int_{-\infty}^{\infty} e^{\lambda \cdot x^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-2\lambda)x^2/2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \frac{dy}{\sqrt{1-2\lambda}} \quad (\text{let } y = \sqrt{1-2\lambda} x) \\ &= \frac{1}{\sqrt{1-2\lambda}} \end{aligned}$$

■

Recall that a Gaussian random variable X is defined through the density function

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} ,$$

where μ is its mean and σ^2 is its variance, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$.