§16. Gaussian Random Variables

Definition 1

A random variable X is **Gaussian (normal)** if it has the pdf

$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty, \qquad \sigma > 0$$

and $X \sim N(\mu, \sigma)$

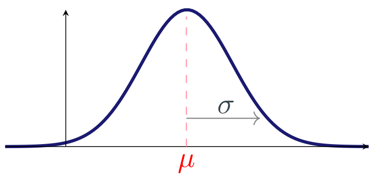


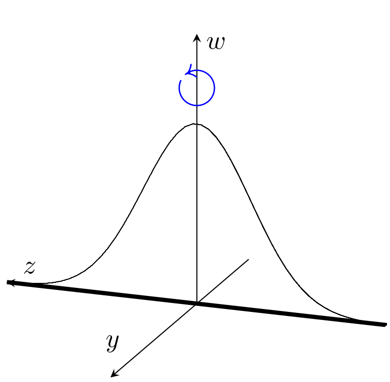
Figure 1: Normal Distribution. Observe the very thin tails

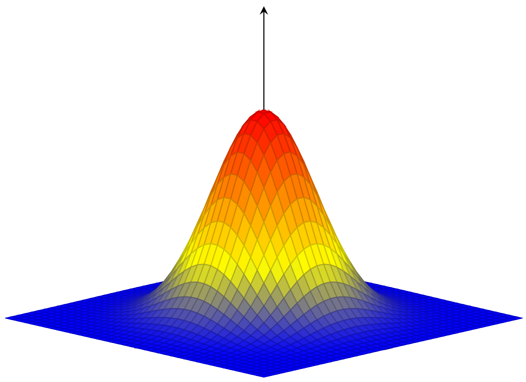
Let's check that this is a valid pdf.

Let $z = \dfrac{x-\mu}{\sigma}$. Then

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz = I$$

Let's rotate $\dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = w$ about the w–axis.





The resulting surface is $w = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2+y^2}{2}}$

The volume under this surface is:

$$\text{Vol.} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{z^2+y^2}{2}}\,dz\,dy = \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx \right) \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy \right) = I^2 \sqrt{2\pi}$$

Using cylindrical shells

$$\text{Vol.} = \frac{1}{\sqrt{2\pi}} \int_0^\infty 2\pi r e^{-\frac{r^2}{2}}\, dr = \sqrt{2\pi} \int_0^\infty r e^{-\frac{r^2}{2}}\, dr = \sqrt{2\pi}$$

$\sqrt{2\pi} = I^2\sqrt{2\pi} \quad \Rightarrow \quad I^2 = 1 \quad \Rightarrow \quad I = 1$; this is a valid density.

Next, let's compute the mean and variance

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \qquad \text{Let } z = \frac{x-\mu}{\sigma} \quad , \quad x = \mu + z\sigma \\ &= \mu \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\, dz + \sigma \cdot \frac{1}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}}\, dz}_{0} \\ &= \mu \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \qquad , \quad z = \frac{x-\mu}{\sigma} \\ &= \int_{-\infty}^{\infty} \frac{(\sigma z + \mu)^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz \\ &= \underbrace{\sigma^2 \int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz}_{integration\, by\, parts = 1} + 2\mu\sigma \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}}\, dz}_{0} + \mu^2 \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz}_{1} \\ &= \sigma^2 + \mu^2 \end{aligned}$$

$\text{Var}(X) = E(X^2) - E(X)^2 = \sigma^2$

Theorem 1

If $X \sim N(\mu, \sigma)$ then $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

Proof.

$F(z) = P(Z \le z) = P\left(\frac{X-\mu}{\sigma} \le Z\right) = P(X \le \mu + z\sigma) = F(\mu + z\sigma)$

Remember that $F(x) = \int_{-\infty}^{z} p(t)\, dt \quad \Rightarrow \quad F'(z) = p(z)$

$$\begin{aligned} P(z) = F'(z) = F'(\mu + z\sigma) \cdot \sigma &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(\mu + z\sigma - \mu)^2}{2\sigma^2}} \cdot \sigma \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \end{aligned}$$

Thus, $Z = \dfrac{X-\mu}{\sigma}$ is standard normal. □

The **error function** is defined as

$$\operatorname{erf}(X) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

The **cumulative** of the standard normal is

$$\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2}\, dt = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)$$

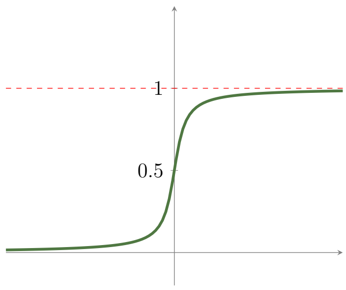


Figure 2: Graph of $\operatorname{erf}(X)$

Example 1

a. $P(Z < -1.82) = 0.0344$

b. $P(Z < 1.09) = 0.8621$

c. $P(Z > 0.64) = 1 - P(Z \leq 0.64)$
$= 1 - 0.7389 = 0.2611$

d. $P(0.12 < Z \leq 2.08) = P(Z \leq 2.08) - P(Z < 0.12) = 0.9812 - 0.578 = 0.4334$

Example 2

The student grades in Calculus 2 are normally distributed with $\mu = 68$ and $\sigma = 13$. What is the probability that the Cal 2 grade of a random student is

a. $X > 90$

b. $X < 40$

c. $70 < X < 75$

Solution

a. $Z = \dfrac{90 - 68}{13} = 1.69$

$$\begin{aligned} P(X > 90) &= P(Z > 1.69) \\ &= 1 - 0.9545 \\ &= 0.0455 \end{aligned}$$

b. $Z = \dfrac{40 - 68}{13} = -2.15$

$$\begin{aligned} P(X < 40) &= P(Z < -2.15) \\ &= 0.0158 \end{aligned}$$

c. $Z = \dfrac{70.68}{13} = 0.15 \qquad Z = \dfrac{75-68}{13} = 0.54$

$$\begin{aligned} P(70 < X < 75) &= P(0.15 < Z < 0.54) \\ &= 0.7054 - 0.5596 \\ &= 0.1458 \end{aligned}$$

Example: Cont'd

$X = 75 \quad P \stackrel{?}{=} 0$. We are approximating a discrete distribution with continuous. To be discussed in detail in the next lecture.

Example: Cont'd

What are the top 10% and bottom 10% percentiles of the Cal 2 grades distribution.

$$P_{10}: \quad Z = -1.28 \qquad x = \mu + z\sigma = 68 + (-1.28)13 = 51.36$$

$$P_{90}: \quad Z = 1.28 \qquad x = \mu + z\sigma = 68 + (1.28)13 = 84.64$$

Example 3

The average weight of Canadian men is $\mu = 178.2\,cm$ with $\sigma = 7.4\,cm$. Assume Gaussian distribution of heights.

a. What is the probability that a randomly selected Canadian man is more than $2\,m$ tall?

b. J.T. is $188.0\,cm$ tall. At what percentile of the Canadian male population is his height?

c. What is the top 20% percentile of the height distribution of Canadian males?

Solution

a. $Z = \dfrac{200 - 178.2}{7.4} = 2.95$

$$\begin{aligned} P(X > 200) &= P(Z > 2.95) \\ &= 1 - 0.9984 \\ &= 0.6016 \end{aligned}$$

b. $Z = \dfrac{188 - 178.2}{7.4} = 1.32$

$P(Z < 1.32) = 0.9066 \sim P_{91}$; The ninety-first percentile

c. $P_{80} \to Z = 0.84$

$x = \mu + z\sigma = 178.2 + 0.84(7.4) = 184.42\,cm.$

Note: That for any Gaussian distribution we have:

$P(\mu - \sigma \le X \le \mu + \sigma) = P(-1 \le z \le 1) = 0.6827$ Chebychev says nothing

$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(-2 \le z \le 2) = 0.9545$ At least$1 - \dfrac{1}{2^2} = 0.75$

$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = P(-3 \le z \le 3) = 0.9973$ At least $1 - \dfrac{1}{3^2} = 0.8889$