# 1 Why Is It Gaussian?

Let $X$ be a normally distributed random variable with mean $\mu$ and variance $\sigma^2$. Let $Y = aX + b$, where $a$ and $b$ are non-zero real numbers. Show explicitly that $Y$ is normally distributed with mean $a\mu + b$ and variance $a^2\sigma^2$. The PDF for the Gaussian Distribution is $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

**Solution:**

Problem and solution taken from *A First Course in Probability* by Sheldon Ross, 8th edition.

Let $a > 0$.

We start with the cumulative distribution function (CDF) of $Y$, $F_Y$.

$$\begin{aligned}
F_Y(x) &= \mathbb{P}[Y \leq x] && \text{By definition of CDF} \\
&= \mathbb{P}[aX + b \leq x] && \text{Plug in } Y = aX + b \\
&= \mathbb{P}\left[X \leq \frac{x-b}{a}\right] && \text{Because } a > 0 \\
&= F_X\left(\frac{x-b}{a}\right) && \text{By definition of CDF. } F_X \text{ denotes the CDF of } X.
\end{aligned} \tag{1}$$

Let $f_Y$ denote the probability density function (PDF) of $Y$.

$$\begin{aligned}
f_Y(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_Y(x) && \text{The PDF is the derivative of the CDF.} \\
&= \frac{\mathrm{d}}{\mathrm{d}x} F_X\left(\frac{x-b}{a}\right) && \text{Plug in the result from (\textbf{??})} \\
&= \frac{1}{a} \cdot f_X\left(\frac{x-b}{a}\right) && \text{PDF is the derivative of CDF. Apply chain rule, } \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x-b}{a}\right) = \frac{1}{a}. \\
&= \frac{1}{a} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-((x-b)/a-\mu)^2/(2\sigma^2)} && X \sim \mathcal{N}(\mu, \sigma^2). \\
&= \frac{1}{a\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-(x-b-a\mu)^2/(2\sigma^2 a^2)} && \frac{x-b}{a} - \mu = \frac{1}{a}(x - b - a\mu)
\end{aligned} \tag{2}$$

We have shown that $f_Y$ equals the probability density function of a normal random variable with mean $b + a\mu$ and variance $\sigma^2 a^2$. So, $Y$ is normally distributed with mean $b + a\mu$ and variance $\sigma^2 a^2$. The proof is done for $a > 0$. The proof for $a < 0$ is similar.

## 2 Sum of Independent Gaussians

In this question, we will introduce an important property of the Gaussian distribution: the sum of independent Gaussians is also a Gaussian.

Let $X$ and $Y$ be independent standard Gaussian random variables. Recall that the density of the standard Gaussian is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

(a) What is the joint density of $X$ and $Y$?

(b) Observe that the joint density of $X$ and $Y$, $f_{X,Y}(x,y)$, only depends on the quantity $x^2+y^2$, which is the distance from the origin. In other words, the Gaussian is *rotationally symmetric*. Next, we will try to find the density of $X+Y$. To do this, draw a picture of the Cartesian plane and draw the region $x+y \le c$, where $c$ is a real number of your choice.

(c) Now, rotate your picture clockwise by $\pi/4$ so that the line $X+Y=c$ is now vertical. Redraw your figure. Let $X'$ and $Y'$ denote the random variables which correspond to the $\pi/4$ clockwise rotation of $(X,Y)$ and express the new shaded region in terms of $X'$ and $Y'$.

(d) By rotational symmetry of the Gaussian, $(X',Y')$ has the same distribution as $(X,Y)$. Argue that $X+Y$ has the same distribution as $\sqrt{2}Z$, where $Z$ is a standard Gaussian. This proves the following important fact: *the sum of independent Gaussians is also a Gaussian*. Notice that $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$ and $Z \sim \mathcal{N}(0,2)$. *In general, if $X$ and $Y$ are independent Gaussians, then $X+Y$ is a Gaussian with mean $\mu_X+\mu_Y$ and variance $\sigma_X^2+\sigma_Y^2$.*

(e) Recall the CLT:

> If $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 < \infty$, then:
>
> $$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1) \qquad \text{as } n\to\infty.$$

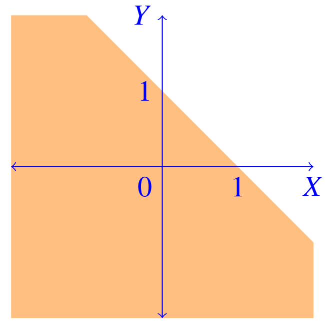
Prove that the CLT holds for the special case when the $X_i$ are i.i.d. $\mathcal{N}(0,1)$.

**Solution:**

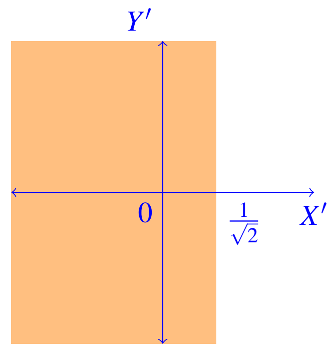
(a) By independence, we have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right).$$

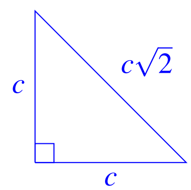
(b) We draw the line for $c=1$.



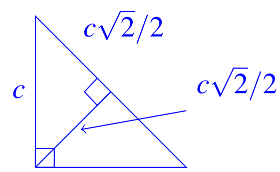

(c) Here is the new figure after the rotation (for $c = 1$).




For general $c \in \mathbb{R}$, the new region is $\{X' \le c/\sqrt{2}\}$. To see why, draw the triangle: We want




to find the distance between the origin and the long side of the triangle, and we can do so by adding a diagonal:




(d) We observe that $\mathbb{P}(X+Y \le c) = \mathbb{P}(X' \le c/\sqrt{2}) = \mathbb{P}(\sqrt{2}X' \le c)$, where $X'$ is a standard Gaussian by rotational symmetry, so this proves the claim.

(e) Here, $\mu = 0$ and $\sigma = 1$. So, by the previous part,

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \sim \frac{1}{\sqrt{n}} \mathcal{N}(0,n) \sim \mathcal{N}(0,1).$$

## 3 Hypothesis testing

We would like to test the hypothesis claiming that a coin is fair, i.e. $P(H) = P(T) = 0.5$. To do this, we flip the coin $n = 100$ times. Let $Y$ be the number of heads in $n = 100$ flips of the coin. We decide to reject the hypothesis if we observe that the number of heads is less than $50 - c$ or larger than $50 + c$. However, we would like to avoid rejecting the hypothesis if it is true; we want to keep the probability of doing so less than 0.05. Please determine $c$. (*Hints: use the central limit theorem to estimate the probability of rejecting the hypothesis given it is actually true.*)

You might need to use the table in the Appendix. Source: `http://cosstatistics.pbworks.com/w/page/27425647/Lesson`

**Solution:**

Let $X_i$ be the random variable denoting the result of the $i$-th flip:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th flip is head,} \\ 0 & \text{if the } i\text{-th flip is tail.} \end{cases}$$

Then we have $Y = \sum_{i=1}^{n} X_i$. If the hypothesis is true, then $\mu = \mathbb{E}[X_i] = \frac{1}{2}$ and $\sigma^2 = \text{var}(X_i) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. By central limit theorem, we know that

$$P\left(\frac{Y - n\mu}{\sqrt{n\sigma^2}} \leq z\right) \approx \Phi(z)$$
$$P\left(\frac{Y - 100 \cdot \frac{1}{2}}{\sqrt{100 \cdot \frac{1}{4}}} \leq z\right) \approx \Phi(z)$$
$$P\left(\frac{Y - 50}{5} \leq z\right) \approx \Phi(z)$$

where

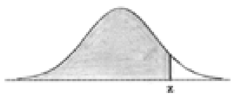
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x.$$

We will reject the hypothesis when $|Y - 50| > c$. We also want $P(|Y - 50| > c) < 0.05$, or equivalently $P(|Y - 50| \leq c) > 0.95$. We have

$$P(|Y - 50| \leq c) = P\left(\frac{|Y - 50|}{5} \leq \frac{c}{5}\right) \approx 2\Phi(\frac{c}{5}) - 1.$$

The reason this is $\approx 2\Phi(\frac{c}{5}) - 1$ is because the probability we are looking for is the probability that $Y$ is within $\frac{c}{5}$ standard deviations of the mean. By an area argument, we can see that this is $\Phi(\frac{c}{5}) - (1 - \Phi(\frac{c}{5})) = 2\Phi(\frac{c}{5}) - 1$. Let $2\Phi(\frac{c}{5}) - 1 = 0.95$, so $\Phi(\frac{c}{5}) = 0.975$ or $\frac{c}{5} = 1.96$. That is $c = 9.8$ flips. So we see that if we observe more that $50 + 10 = 60$ or less than $50 - 10 = 40$ heads, we can reject the hypothesis.

## 4 Appendix




# Probability Content from -oo to Z

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |

Table 1: Table of the Normal Distribution