STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 10: Special Continuous Distributions: Uniform, Normal (Gaussian)

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

February 26, 2024

Outline

1. Previously...
   - Continuous Random Variables
   - Expected Value of Continuous Random Variables
   - Variance of Continuous Random Variables
2. Uniform Distribution
3. Normal (Gaussian) Distribution
4. Standard Normal Distribution

Previously...

Continuous Random Variables

- Cumulative Distribution Function (CDF) of Y:

$$F(y) = P(Y \leq y), \quad -\infty < y < \infty.$$

- Probability Density Function (PDF) of Y:

$$f(y) = \frac{dF(y)}{dy} = F'(y).$$

- PDF to CDF:

$$F(y) = \int_{-\infty}^{y} f(t)dt.$$

- Probability that Y falls in the interval $[a, b]$ is
  - Using the PDF:

$$P(a \leq Y \leq b) = \int_{a}^{b} f(y)dy$$

  - Using the CDF:

$$P(a \leq Y \leq b) = F(b) - F(a)$$

Expected Value of Continuous Random Variables

- Expected value of Y: is a measure of central tendency

$$\mu = E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

- Expected value of functions of Y: Let $g(Y)$ be a real-valued function of Y.

$$E\{g(Y)\} = \int_{-\infty}^{\infty} g(y) f(y) dy$$

- Expected value of a constant: Let c be a constant. Then $E(c) = c$.
- Expected value of a scaled Y: Let $g(Y)$ be a function of Y, and c be a constant.

$$E\{cg(Y)\} = cE\{g(Y)\}$$

- Expected value of a sum of random variables: Let $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y.

$$E\{g_1(Y) + g_2(Y) + \ldots + g_k(Y)\} = E\{g_1(Y)\} + E\{g_2(Y)\} + \ldots + E\{g_k(Y)\}$$

Variance of Continuous Random Variables

- Variance of Y: is a measure of the dispersion or scatter of the values of the random variable about the mean μ.

$$\sigma^2 = V(Y) = E\{(Y - \mu)^2\}$$

- More useful formula to compute the variance:

$$\sigma^2 = V(Y) = E(Y^2) - \mu^2$$

Note: The formula above can also be written as $\sigma^2 = V(Y) = E(Y^2) - \{E(Y)\}^2$.

- Standard deviation of Y:

$$\sigma = \sqrt{V(Y)}$$

Uniform Distribution

Uniform Distribution

Definition 3.7: Uniform Distribution

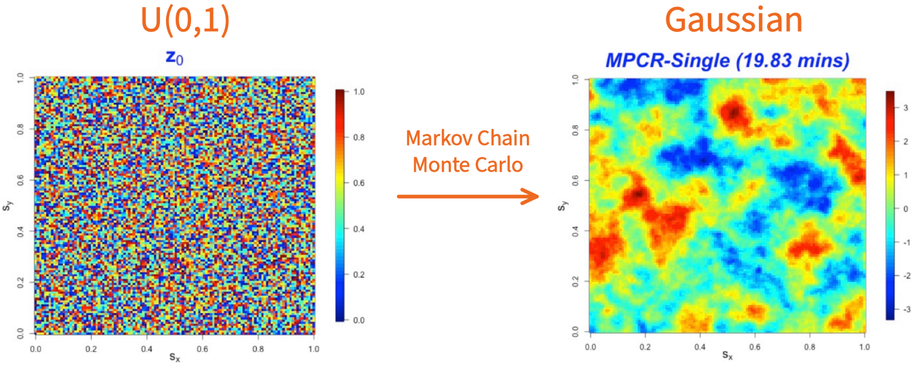
If $\theta_1 < \theta_2$, a random variable Y is said to have a continuous *uniform probability distribution* on the interval (θ_1, θ_2) if and only if the density function of Y is

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2, \\ 0, & \text{elsewhere.} \end{cases}$$

- Notation: $Y \sim \mathrm{U}(\theta_1, \theta_2)$, read as: "$Y$ is a uniform random variable on the interval (θ_1, θ_2)."
- CDF: $F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt = \begin{cases} 0, & y < \theta_1 \\ \frac{y - \theta_1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2 \\ 1, & y > \theta_2. \end{cases}$

Usage:

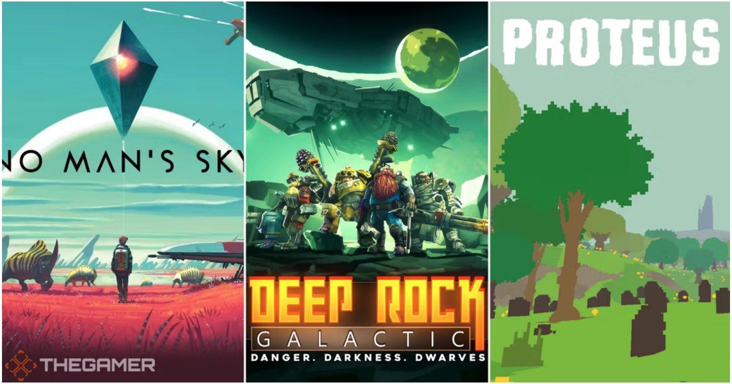
Random number generator: As starting point for generating values from more complicated distributions...



Uniform Distribution

Usage:

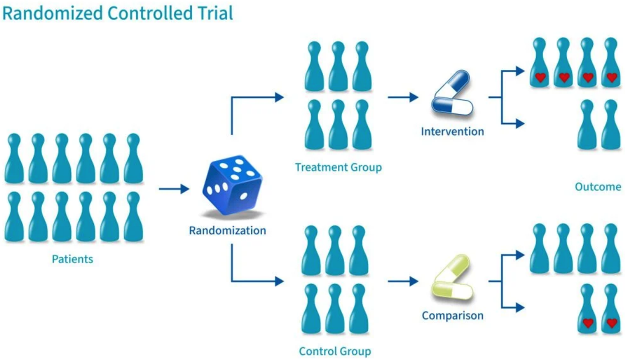
Procedural Generation: Introducing randomness into computer gaming algorithms to produce unpredictable or unique contents...



Source: https://www.thegamer.com/best-procedurally-generated-games-minecraft

Usage:

Randomized Control Trial: randomly assigning participants to either an experimental group or a control group to measure the effectiveness of an intervention or treatment....

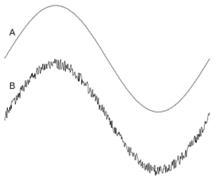


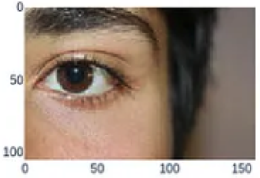
Source: https://www.simplypsychology.org/randomized-controlled-trial.html

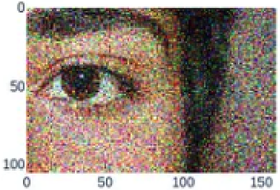
Uniform Distribution

Usage:

Data Privacy: Adding or multiplying a random number to confidential quantitative attributes







Source: https://medium.com/@ms_somanna/guide-to-adding-noise-to-your-data-using-python-and-numpy-c8be815df524

Uniform Distribution

Theorem 4.6

If $\theta_1 < \theta_2$ and Y is a random variable uniformly distributed on the interval (θ_1, θ_2), then

$$\mu = E(Y) = \frac{\theta_1 + \theta_2}{2} \quad \text{and} \quad \sigma^2 = V(Y) = \frac{(\theta_2 - \theta_1)^2}{12}.$$

Proof:

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y f(y) dy && \text{def'n of expected value} \\
&= \int_{\theta_1}^{\theta_2} y \left(\frac{1}{\theta_2 - \theta_1}\right) dy && f(y) = \begin{cases} \frac{1}{\theta_2-\theta_1}, & \theta_1 \le y \le \theta_2, \\ 0, & \text{elsewhere} \end{cases} \\
&= \left(\frac{1}{\theta_2 - \theta_1}\right) \frac{y^2}{2}\Bigg|_{\theta_1}^{\theta_2} \\
&= \frac{\theta_2^2 - \theta_1^2}{2(\theta_2 - \theta_1)} \\
&= \frac{(\theta_2 - \theta_1)(\theta_2 + \theta_1)}{2(\theta_2 - \theta_1)} && \text{diff. of two squares} \\
&= \frac{\theta_2 + \theta_1}{2}.
\end{aligned}$$

(cont'd next slide...)

Uniform Distribution

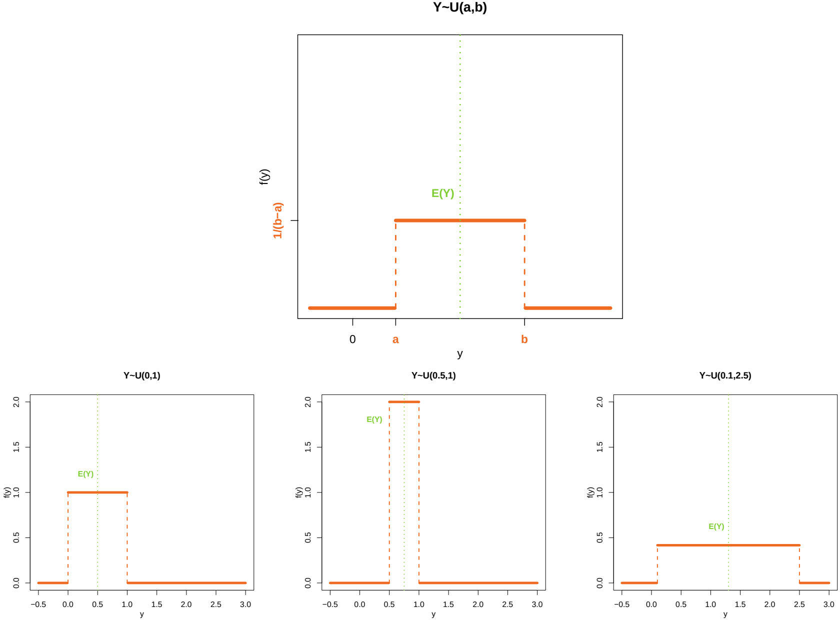
Proof:

$$
\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y)dy && \text{def'n of expected value} \\
&= \int_{\theta_1}^{\theta_2} y^2 \left(\frac{1}{\theta_2-\theta_1}\right) dy && f(y) = \begin{cases} \frac{1}{\theta_2-\theta_1}, & \theta_1 \le y \le \theta_2, \\ 0, & \text{elsewhere} \end{cases} \\
&= \left(\frac{1}{\theta_2-\theta_1}\right) \frac{y^3}{3}\Big|_{\theta_1}^{\theta_2} \\
&= \frac{\theta_2^3-\theta_1^3}{3(\theta_2-\theta_1)} \\
&= \frac{(\theta_2-\theta_1)(\theta_2^2+\theta_2\theta_1+\theta_1^2)}{3(\theta_2-\theta_1)} && \text{diff. of two cubes} \\
&= \frac{\theta_2^2+\theta_2\theta_1+\theta_1^2}{3}.
\end{aligned}
$$

$$
\begin{aligned}
V(Y) &= E(Y^2) - \{E(Y)\}^2 && \text{def'n of variance} \\
&= \frac{\theta_2^2+\theta_2\theta_1+\theta_1^2}{3} - \left(\frac{\theta_2+\theta_1}{2}\right)^2 \\
&= \frac{\theta_2^2+\theta_2\theta_1+\theta_1^2}{3} - \frac{\theta_2^2+2\theta_2\theta_1+\theta_1^2}{4} \\
&= \frac{4(\theta_2^2+\theta_2\theta_1+\theta_1^2)-3(\theta_2^2+2\theta_2\theta_1+\theta_1^2)}{12} \\
&= \frac{4\theta_2^2+4\theta_2\theta_1+4\theta_1^2-3\theta_2^2-6\theta_2\theta_1-3\theta_1^2)}{12} \\
&= \frac{\theta_2^2-2\theta_2\theta_1+\theta_1^2}{12} \\
&= \frac{(\theta_2-\theta_1)^2}{12}.
\end{aligned}
$$

□

Uniform Distribution



Uniform Distribution

Example 1:

A continuous random variable X is uniformly distributed over the interval $[b, 4b]$ where b is a constant.

a. What is $E(X)$?

Solution:

$E(X) = \frac{b+4b}{2} = \frac{5b}{2}$. Theorem 4.6: If $X \sim U(\theta_1, \theta_2)$, then $E(X) = \frac{\theta_1+\theta_2}{2}$.

Uniform Distribution

Example 1:

A continuous random variable X is uniformly distributed over the interval $[b, 4b]$ where b is a constant.

- (b) Show that $V(X) = \frac{3b^2}{4}$?

Solution:

$V(X) = \frac{(4b-b)^2}{12} = \frac{(3b)^2}{12} = \frac{9b^2}{12} = \frac{3b^2}{4}$. Theorem 4.6: If $X \sim U(\theta_1, \theta_2)$, then $V(X) = \frac{(\theta_2-\theta_1)^2}{12}$.

Uniform Distribution

Example 1:

A continuous random variable X is uniformly distributed over the interval $[b, 4b]$ where b is a constant.

c. Find $V(3-2X)$.

Solution:

$V(3-2X) = (-2)^2 V(X) = (4)\frac{3b^2}{4} = 3b^2.$ Properties of variance and $V(X) = \frac{3b^2}{4}$ from b).

Uniform Distribution

Example 2:

The continuous random variable X is uniformly distributed over the interval $[-4, 6]$.

(a) Find $P(X \leq 2.4)$.

Solution:

$$
\begin{aligned}
P(X \leq 2.4) &= \int_{-\infty}^{2.4} f(x)dx \quad \text{probability = area under the curve} \\
&= \int_{-4}^{2.4} \frac{1}{6-(-4)} dx \quad \text{PDF of uniform: } f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq x \leq \theta_2, \\ 0, & \text{elsewhere} \end{cases} \\
&= \frac{1}{10} x \Big|_{-4}^{2.4} = \frac{1}{10}(2.4 - (-4)) = 0.64.
\end{aligned}
$$

Example 2:

The continuous random variable X is uniformly distributed over the interval $[-4, 6]$.

- (b) The continuous random variable Y is uniformly distributed over the interval $[a, 4a]$. Find the value of a such that $P\left(X \leq \frac{8}{3}\right) = P\left(Y \leq \frac{8}{3}\right)$.

Solution:

$$P\left(X \leq \frac{8}{3}\right) = \int_{-\infty}^{\frac{8}{3}} f(x)dx$$ probability = area under the curve

$$= \int_{-4}^{\frac{8}{3}} \frac{1}{6-(-4)} dx$$ PDF of uniform: $f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq x \leq \theta_2, \\ 0, & \text{elsewhere} \end{cases}$

$$= \frac{1}{10} x \Big|_{-4}^{\frac{8}{3}} = \frac{1}{10}\left(\frac{8}{3} - (-4)\right) = \frac{1}{10}\left(\frac{20}{3}\right) = \frac{2}{3}.$$

$$P\left(Y \leq \frac{8}{3}\right) = \int_{-\infty}^{\frac{8}{3}} f(y)dy$$ probability = area under the curve

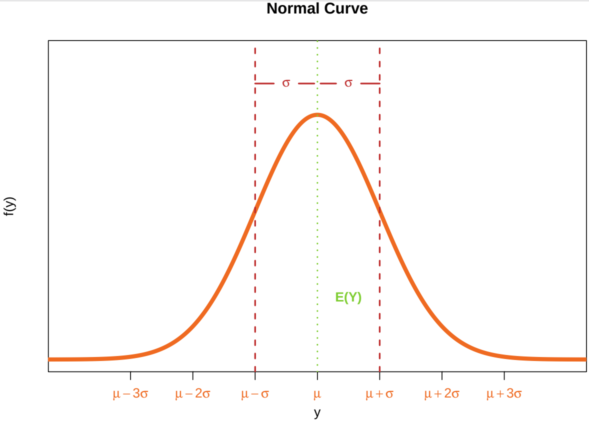
$$= \int_{a}^{\frac{8}{3}} \frac{1}{4a-a} dy$$ PDF of uniform: $f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2, \\ 0, & \text{elsewhere} \end{cases}$

$$= \frac{1}{3a} y \Big|_{a}^{\frac{8}{3}} = \frac{1}{3a}\left(\frac{8}{3} - a\right) = \frac{8}{9a} - \frac{1}{3}.$$

$$\Rightarrow \frac{2}{3} = \frac{8}{9a} - \frac{1}{3} \Rightarrow a = \frac{8}{9}.$$

Normal (Gaussian) Distribution

Gaussian Distribution



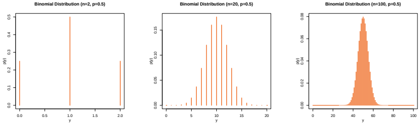
- The normal curve is perhaps the most important probability graph in all of statistics.
  - bell-shaped curve
  - high in the middle (mean, μ)
  - gradually tails off in each direction
- More values at the center of the distribution and few in the tails

Abraham de Moivre

Source: Cambridge University Library

- He was solving a gambling problem:
  - $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$.
  - $p(60) + p(61) + p(62) + \ldots$.

  VERY TEDIOUS!!!
- He noticed that as n increases, the shape of the binomial distribution approaches a smooth curve.



- He found a mathematical expression for this curve.
- So instead of having to add lots of individual numbers you can just find the area under the curve...

Gaussian Distribution: History

Carl Friedrich Gauss

Portrait by Christian Albrecht Jensen, 1840

- He developed the Gaussian PDF:

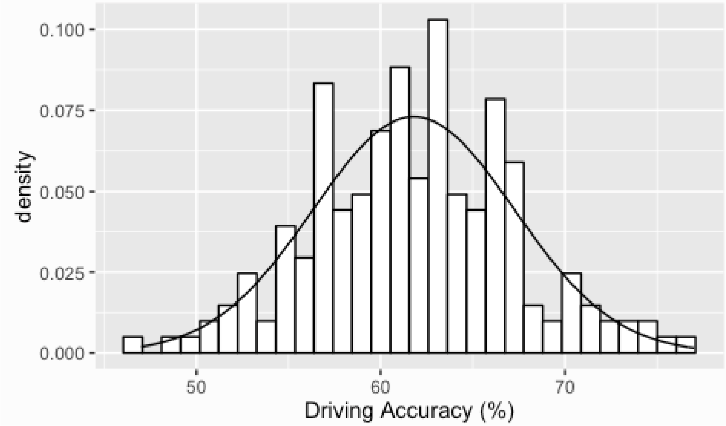
$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, -\infty \leq y \leq \infty$$

  - expected value: μ
  - variance: σ^2
- Approximating distribution to:
  - binomial
  - Poisson
  - χ^2
  - Student-t

German 10-Deutsche Mark Banknote (1993; discontinued)

Gaussian Distribution: Normality is Everywhere

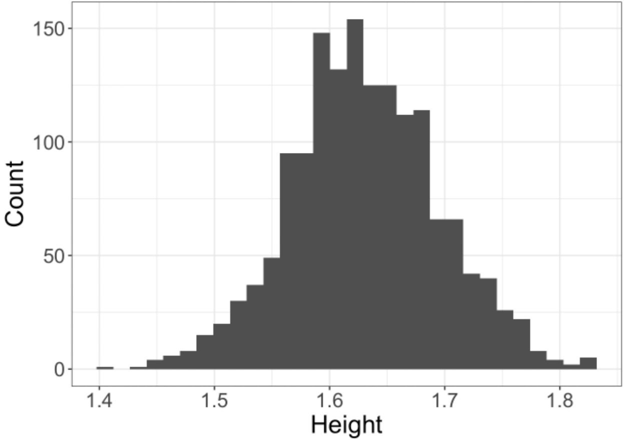
Golf



Source: https://uc-r.github.io/assumptions_normality

Gaussian Distribution: Normality is Everywhere

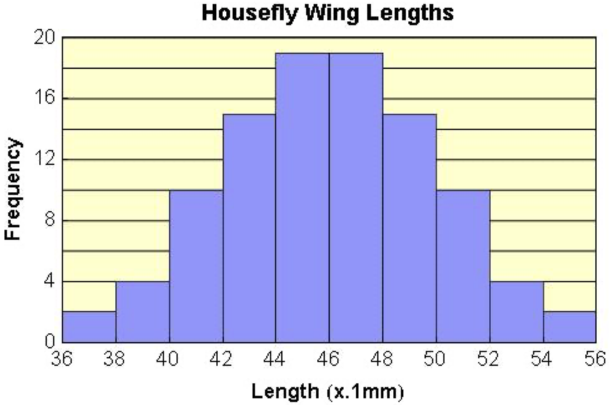
Human Height



Source:

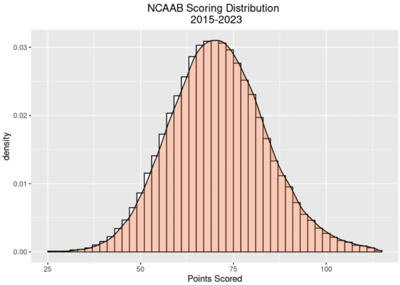
https://www.ucd.ie/ecomodel/Resources/Sheet4_data_distributions_WebVersion.html

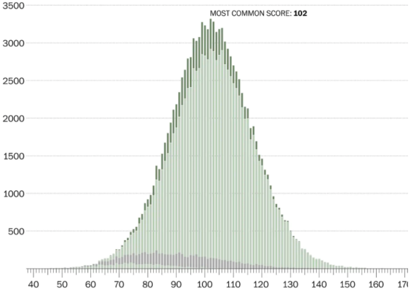
Housefly Wingspan



Source: https://seattlecentral.edu/qelp/sets/057/057.html

Sports





Sources:
https://rpubs.com/Thom9567/1012507
https://priorprobability.com/2014/12/06/nba-data-set/

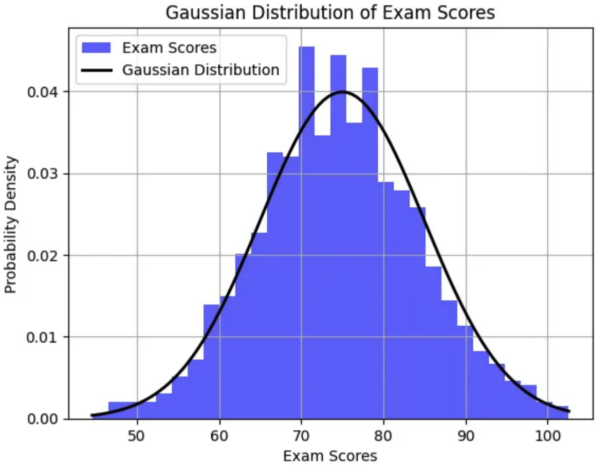
Gaussian Distribution: Normality is Everywhere

Gym

Source: https://www.reddit.com/r/mildlyinteresting/comments/9omj54/gaussian_distribution_of_usage_marks_at_my_local

Gaussian Distribution: Normality is Everywhere

Exam Scores

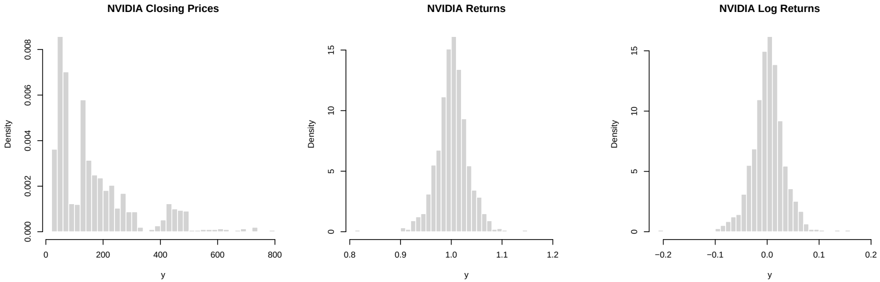


Source: https://medium.com/@akashsri306/the-gaussian-distribution-machine-learnings-secret-weapon-4f37f590718d

Gaussian Distribution: Normality is Everywhere

Even when they are NOT normal...

we make them normal!



Gaussian Distribution

Definition 4.8: Gaussian Distribution

A random variable Y is said to have a *Gaussian probability distribution* if and only if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of Y is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty \leq y \leq \infty$$

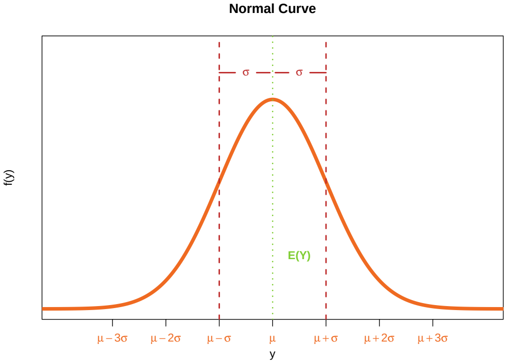
Theorem 4.7

If Y is a normally distributed random variable with parameters μ and σ, then

$$E(Y) = \mu \quad \text{and} \quad V(Y) = \sigma^2.$$

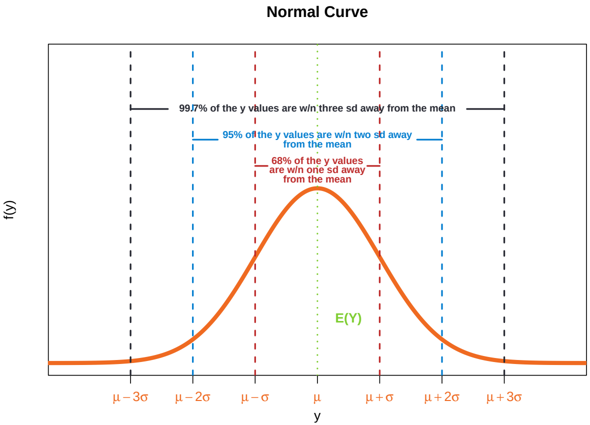
- Notation: $Y \sim \mathcal{N}(\mu, \sigma^2)$, read as: "$Y$ is normally distributed with mean μ and variance σ^2."
- The parameter μ locates the center or peak of the distribution.
- The parameter σ measures the spread of the distribution.
- symmetric at $y = \mu$

Gaussian Distribution: A Closer Look at the PDF



- The center of the curve is determined by μ.
- The width of the curve is determined by σ.
  - The larger σ is, the wider or flatter the curve will be.
  - The smaller σ is, the narrower or taller the curve will be.
- The units in the horizontal axis are given in standard deviations.
- The area under the curve to the right of the mean is 0.5.
- The area under the curve to the left of the mean is 0.5.

Gaussian Distribution: The Empirical Rule



- $P(\mu - \sigma \leq Y \leq \mu + \sigma) = 0.6827$
- $P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) = 0.9545$
- $P(\mu - 3\sigma \leq Y \leq \mu + 3\sigma) = 0.9973$

Gaussian Distribution: Areas under the PDF

- PDF: $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, -\infty \leq y \leq \infty$
- CDF: $F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)dt = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$
  - $P(a \leq Y \leq b) = \int_a^b f(y)dy = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$ (Using the PDF) OR
  - $P(a \leq Y \leq b) = F(b) - F(a)$ (Using the CDF)

Unfortunately, closed-form expressions for these integrals do not exist. These need to be solved using numerical integration techniques.

If these integrals cannot be solved, how do we compute the probabilties or areas under the normal curve?

Answer: We use the standard normal table:

- also known as the Z-table
- provides the area under the curve to the left of a z-score (values in the horizontal axis of a standard normal curve)

Standard Normal Distribution

Standard Normal Distribution

- Normal (Gaussian) PDF: $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty \leq y \leq \infty.$
- Let $Z = \frac{Y-\mu}{\sigma}$. What is the distribution of this new random variable?

$$
\begin{aligned}
F_Z(z) = P(Z \leq z) &= P\left(\frac{Y-\mu}{\sigma} \leq z\right) \\
&= P\left(Y \leq \sigma z + \mu\right) \\
&= F_Y(\sigma z + \mu). \\
f_Z(z) = \frac{d}{dz} F_Z(z) &= \frac{d}{dz} F_Y(\sigma z + \mu) \quad \text{def'n of PDF} \\
&= \sigma f_Y(\sigma z + \mu) \quad f(y) = F'(y) \text{ and chain rule} \\
&= \sigma \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma z + \mu - \mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \quad \text{(This is the standard normal (Gaussian) PDF)}
\end{aligned}
$$

- This is known as the standard normal distribution, a special normal (Gaussian) distribution where $\mu = 0$ and $\sigma = 1$.

Standard Normal Distribution

- Turns out, $Z = \frac{Y-\mu}{\sigma}$ is a standard normal random variable.
- Notation: $Z \sim \mathcal{N}(0,1)$
- PDF: $\phi(z) = f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ -\infty \leq z \leq \infty$
- CDF: $\Phi(z) = F(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$
  - $P(a \leq Z \leq b) = \int_a^b \phi(z)dz = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$ (Using the PDF)
    OR
  - $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$ (Using the CDF)

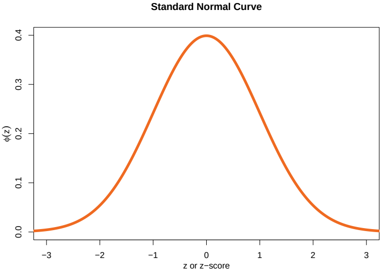
Again, unfortunately, we cannot simplify or solve these integrals.

However, FORTUNATELY, the value of the CDF $\Phi(z)$ has already been precomputed for any number z.

You can find these values in a standard normal table or Z-table.

You need to know how to use the Z-table for the exams!

Standard Normal Distribution: A Closer Look at the PDF



- A standard normal distribution always has a mean of zero.
- The unit in the horizontal axis is in standard deviations.
- The z-score is a value which the standard normal random variable Z can take.
- The z-score value = -1 is:
  - 1 standard deviation away from the mean and
  - falls below the mean since the sign is negative.
- The z-score value $= 2$ is:
  - 2 standard deviations away from the mean and
  - found above the mean since the sign is positive.

Reading the Z-table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831

- Row headings: z-score up to the first decimal place.
- Column headings: second decimal place of the z-score.
- Cells: areas under the standard normal curve to the left of every z-score.

Reading the Z-table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831

- Row headings: z-score up to the first decimal place.
- Column headings: second decimal place of the z-score.
- Cells: areas under the standard normal curve to the left of every z-score.

What is $\Phi(-2.23)$ or $P(Z \leq -2.23)$?

Answer: $0.01287 = 1.287\%$

Reading the Z-table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786

What is $P(-1 \leq Z \leq 1)$?

Answer: $P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1)$ probability = area under the curve

$= \Phi(1) - \Phi(-1)$ Standard normal CDF

$= 0.84134 - 0.15866 = 0.68268$

Reading the Z-table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361

What is $P(-2 \leq Z \leq 2)$?

Answer: $P(-2 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq -2)$ probability = area under the curve

$= \Phi(2) - \Phi(-2)$ Standard normal CDF

$= 0.97725 - 0.02275 = 0.9545$

Reading the Z-table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976

What is $P(-3 \leq Z \leq 3)$?

Answer: $P(-3 \leq Z \leq 3) = P(Z \leq 3) - P(Z \leq -3)$ probability = area under the curve

$= \Phi(3) - \Phi(-3)$ Standard normal CDF

$= 0.99865 - 0.00135 = 0.9973$

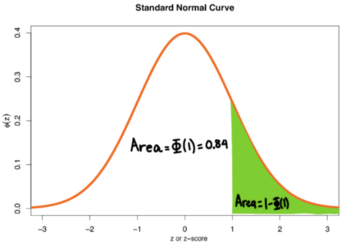
Reading the Z-table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214

What is $P(Z > 1)$?

Answer: $P(Z > 1) = 1 - P(Z \leq 1) = 1 - \Phi(1) = 1 - 0.84134 = 0.15866$



Converting Normal to Standard Normal ($Y \to Z$)

- What if I need $P(a \leq Y \leq b)$ where $Y \sim \mathcal{N}(\mu, \sigma^2)$?

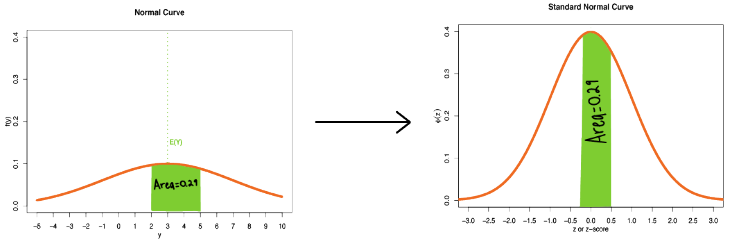
$$\begin{aligned} P(a \leq Y \leq b) &= P(Y \leq b) - P(Y \leq a) && \text{probability = area under the PDF} \\ &= P\left(\frac{Y-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) - P\left(\frac{Y-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) && \text{standardization won't change the inequality} \\ &= P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z \leq \frac{a-\mu}{\sigma}\right) && \text{def'n of standard normal r.v.} \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) && \text{standard normal CDF} \end{aligned}$$

- This tells us that to compute $P(a \leq Y \leq b)$, we need to transform Y to Z and find the z-scores of a and b (or the equivalent numbers of a and b in the standard normal graph).

Example

Let $Y \sim \mathcal{N}(3, 16)$, what is $P(2 < Y < 5)$?

$$\begin{aligned} P(2 < Y < 5) &= P\left(\frac{2-3}{\sqrt{16}} \leq \frac{Y-3}{\sqrt{16}} \leq \frac{5-3}{\sqrt{16}}\right) \quad \text{standardize } Y \\ &= P\left(-\frac{1}{4} \leq Z \leq \frac{2}{4}\right) \\ &= \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{4}\right) \\ &= 0.69146 - 0.40129 = 0.29017. \end{aligned}$$



Questions?

Homework Exercises: 4.61, 4.71, 4.73, 4.77, 4.81

Solutions will be discussed this Friday by the TA.