18.01A Topic 16: Normal distributions.
Read: SN: P section 4

Read the supplementary notes

Standard Normal Random Variables

This is the famous bell curve, it has many uses throughout math and science. The most basic involves many measurements of the same quantity.

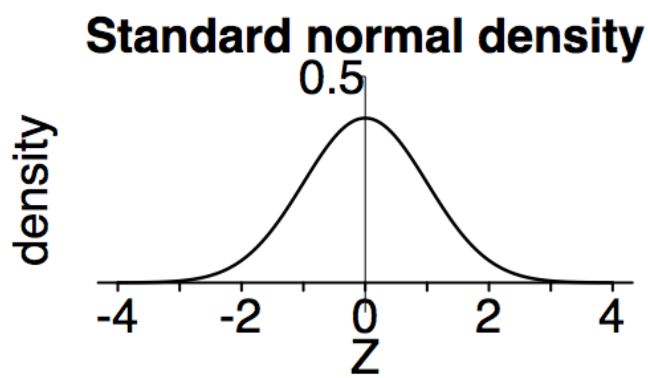
i) Standard notatation: Z

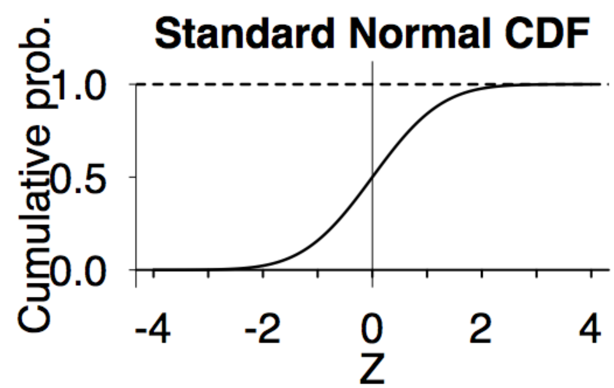
ii) Values: $(-\infty, \infty)$.

iii) Probability density function: $\phi(z) = \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}$.

iv) Cumulative distribution function: $\Phi(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} \mathrm{e}^{-t^2/2}\, dt$.

v) $P(a \le Z \le b) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{a}^{b} \mathrm{e}^{-t^2/2}\, dt = \Phi(b) - \Phi(a)$.





Notes

1. Z, $\phi(z)$ and $\Phi(z)$ are all standard notation.

2. Need the factor of $\dfrac{1}{\sqrt{2\pi}}$ so total probability is 1. (We will show this in 18.02.)

3. $\Phi(z)$ has no closed form, we must compute it numerically. See the table in the supplementary notes.

4. $\Phi(-z) = \displaystyle\int_{-\infty}^{-z} \phi(t)\, dt =$ (by symmetry) $\displaystyle\int_{z}^{\infty} \phi(t)\, dt = 1 - \Phi(z)$.

5. $P(-a < Z < a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1$.

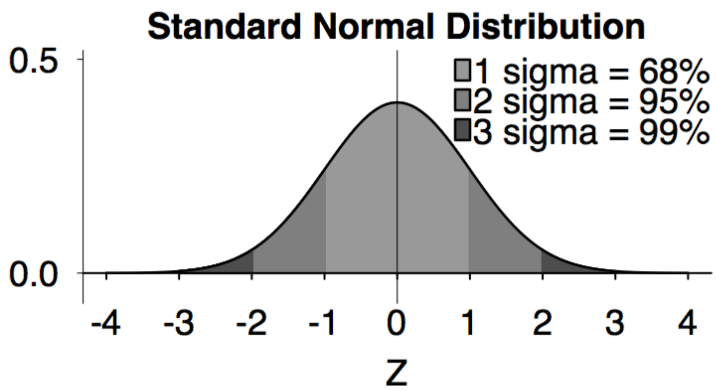
Theorem: $E(Z) = 0, \quad \sigma(Z) = 1$.

Proof: $z\phi(z)$ is an odd function $\Rightarrow E(Z) = \displaystyle\int_{-\infty}^{\infty} z\phi(z)\, dz = 0$.

The proof for σ is in the pset.

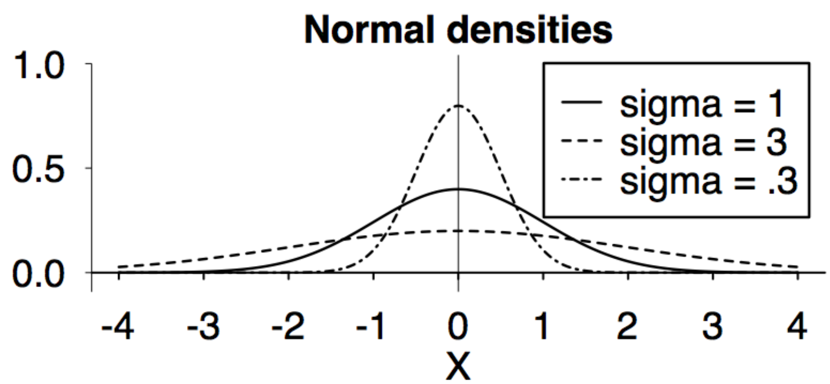
(continued)

Important numbers: $P(-1 < Z < 1) \approx .68, \quad P(-2 < Z < 2) \approx .95.$
I.e., the probability Z is within one standard deviation of the mean is .68 and the probability Z is within two standard deviations of the mean is .95.



Normal Distributions: We can scale and shift Z.
$X_{m,\sigma} = \sigma Z + m$ is called **normal** with mean m and standard deviation σ. ($Z = X_{0,1}$.)

Theorem:
1. $X_{m,\sigma}$ has probability density function $\phi_{m,\sigma}(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$.
2. $E(X_{m,\sigma}) = m, \quad \sigma(X_{m,\sigma}) = \sigma$.
3. $P(-\sigma < X_{m,\sigma} - m < \sigma) \approx .68, \quad P(-2\sigma < X_{m,\sigma} - m < 2\sigma) \approx .95$.



Proof: We transform $X_{m,\sigma}$ to standard normal –this is the same type of argument we will use repeatedly.

1. $P(a < X_{m,\sigma} < b) = P(a < \sigma Z + m < b) = P\left(\dfrac{a-m}{\sigma} < Z < \dfrac{b-m}{\sigma}\right) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{(a-m)/\sigma}^{(b-m)/\sigma} e^{-z^2}\, dx$. Make the change of variable $x = \sigma z + m$, which gives

$$P(a < X_{m,\sigma} < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx. \quad \blacksquare$$

(continued)

2. $E(X_{m,\sigma}) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx.$

The same change of variable gives

$E(X_{m,\sigma}) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} \sigma z e^{-z^2/2}\, dz + \dfrac{m}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} e^{-z^2/2}\, dz = m.$

3. Is similar to 2.

Example: Suppose the average height of a male in Boston is a normal r.v. with mean 68 inches and with a standard deviation of 4. What is the percentage of men over 6 feet tall (6 feet = 72 inches)?

answer: Let X = height of a randomly chosen person.

We want $P(X > 72)$.

Note that $72 = 68 + 4 = \mu + \sigma$.

For a normal random variable $P(X > \mu + \sigma) = .16$.

Reason 1: Look in the table $P(Z < 1) = .84 \Rightarrow P(Z > 1) = .16$.

Reason 2: 68% are within one std. dev. of the mean

$\Rightarrow$ 32% are outside (above or below) 1 std. dev. from the mean

$\Rightarrow$ (by symmetry) 16% are more than 1 std. dev. above the mean.

Final answer: 16%.

What percentage of Boston males will be able to walk through a 76 inch door frame without ducking?

answer: We want $P(X < 76)$.

76 inches is 2σ above the mean. We know 95% are within 2σ of the mean. So only 2.5% are more than 2σ above the mean $\Rightarrow P(X < 76) \approx 97.5\%$.

Example: Suppose the lifetime in hours of a brand of flashlight batteries is a normal r.v. X with mean 100 and standard deviation 12. Give an interval in which X lies 95% of the time.

answer: We know $P(m - 2\sigma < X < m + 2\sigma) = .95 \Rightarrow P(76 < X < 124) = .95$.

$\Rightarrow$ $\boxed{\text{our interval is } 76 < X < 124.}$

Table of values for $\Phi(z)$

z:	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	
$\Phi(z)$	0.5000	0.5398	0.5793	0.6179	0.6554	0.6915	0.7257	0.7580	0.7881	0.8159	
z:	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	
$\Phi(z)$	0.8413	0.8643	0.8849	0.9032	0.9192	0.9332	0.9452	0.9554	0.9641	0.9713	
z:	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
$\Phi(z)$	0.9772	0.9821	0.9861	0.9893	0.9918	0.9938	0.9953	0.9965	0.9974	0.9981	0.9987