1.4.3 Uniform Distribution

A **uniform** continuous random variable lies in a fixed *closed interval* $[a, b]$ and is *equally likely* to lie in any subinterval of length d no matter where the subinterval lies in $[a, b]$.

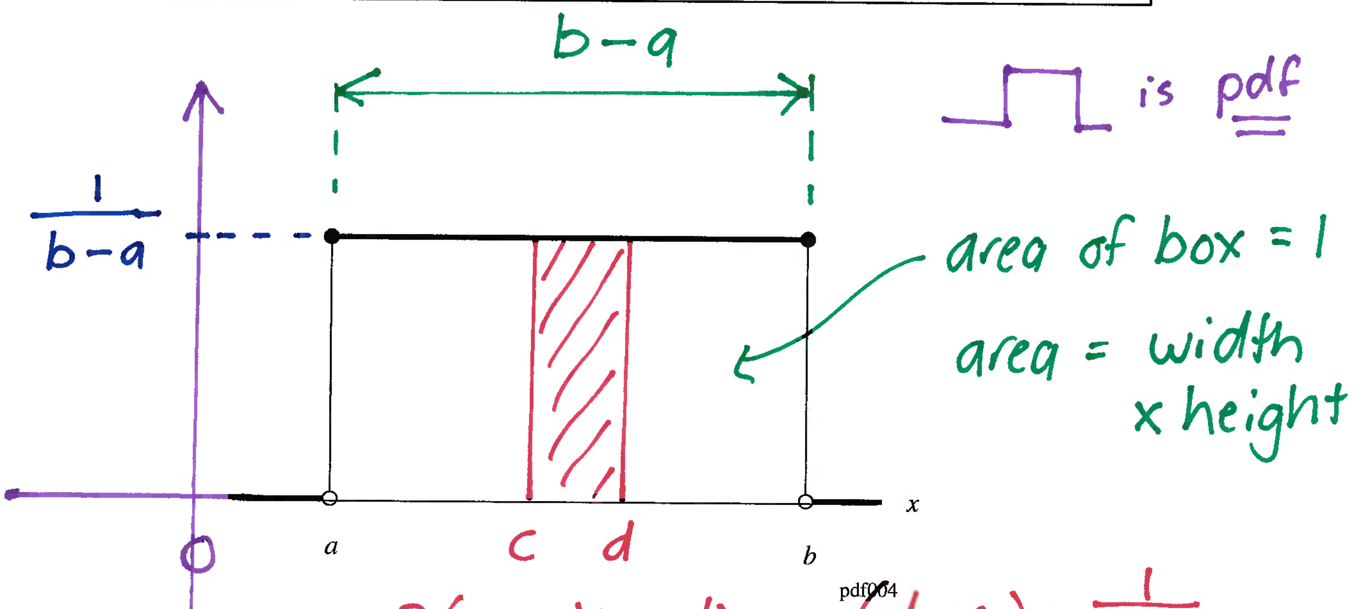
■ A continuous random variable X has a

uniform distribution on $[a, b]$

if the pdf of X is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

We write $X \sim U(a, b)$.



$$P(c \leq X \leq d) = \underbrace{(d-c)}_{\text{width}} \times \underbrace{\frac{1}{b-a}}_{\text{height}}$$

- If $X \sim U(a,b)$ then

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

cdf

$$E(X) = \frac{a+b}{2}$$

→ halfway between a and b

$$\text{var}(X) = \frac{(b-a)^2}{12}$$

Question —

What does R know about $U(a,b)$?

```
dunif(x,min,max) # pdf
punif(q,min,max) # cdf
qunif(p,min,max) # quantile
runif(n,min,max) # random variate
```

$= P(X \leq q)$

given p, what is x so that $P(X \leq x) = p$

$$x \xrightarrow{\text{punif}} p \xrightarrow{\text{qunif}} x$$

observation world — probability world "quantiles" — observation world "follows" or "obeys"

If $X \sim U(a,b)$ then we need to know a and b to calculate probabilities.

Example 1.29 The continuous random variable X is uniformly distributed on the interval $[a, b]$. Given $E(X) = 4$ and $\text{var}(X) = 3$, find

(a) a and b (b) $P(X > 5)$

Solution.

(a) $$E(X) = \frac{a+b}{2} \qquad \text{var}(X) = \frac{(b-a)^2}{12}$$

$$\frac{a+b}{2} = 4 \text{ giving } a + b = 8 \text{ so } b = 8 - a$$

$$\begin{aligned}
\frac{(b-a)^2}{12} &= 3 \\
(b-a)^2 &= 36 \\
((8-a)-a)^2 &= 36 \\
(8-2a)^2 &= 36 \\
8-2a &= -6 \quad \text{or} \quad 8-2a = 6 \\
2a &= 14 \quad \text{or} \quad 2a = 2 \\
a &= 7 \quad \text{or} \quad a = 1
\end{aligned}$$

When $a = 7$ we have $b = 1$, but require $a < b$. Therefore must have $\boxed{a = 1 \text{ and therefore } b = 7.}$

(b) $P(X > 5) = 1 - P(X \le 5) = 1 - \frac{5-1}{7-1} = 1 - \frac{4}{6} = \frac{1}{3}$

```
> 1-punif(5,1,7)
[1] 0.3333333
```

Use punif to look up probabilities

□

1.4.4 Normal Distribution

"normal" as very commonly found in nature

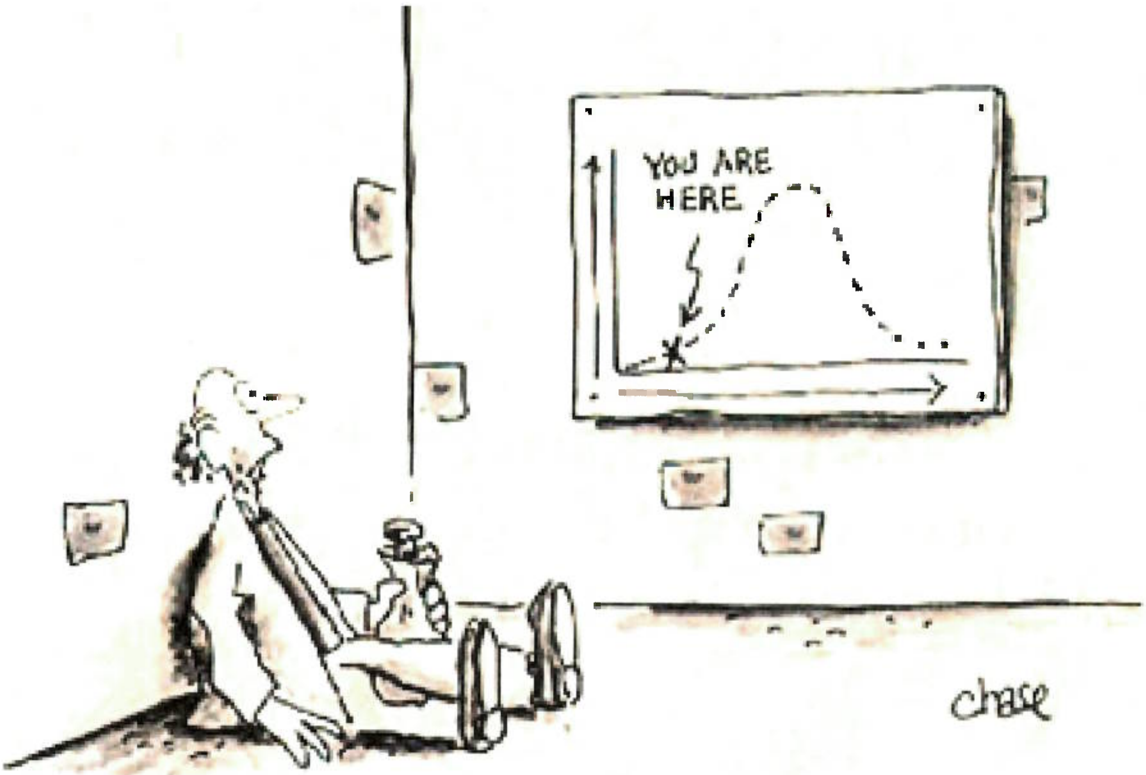


Image from http://www.sjsu.edu/faculty/gerstman/hs167/URHere.jpg

History.
The normal distribution was discovered in 1733 by Abraham De Moivre (1667–1754) in his investigation of approximating coin tossing probabilities. He named the PDF of his discovery the **exponential bell-shaped curve**.

`http://www-groups.dcs.st-and.ac.uk/~history/BiogIndex.html`

History.
In 1809, Carl Friedrich Gauss (1777–1855) firmly established the importance of the normal distribution by using it to predict the location of astronomical bodies. As a result, the normal distribution then became known as the **Gaussian distribution**, a terminology that is still used, especially in engineering.

Later, in the last half of the 19th century, researchers discovered that many variables have distributions that follow or are well-approximated by a Gaussian distribution. Roughly speaking, researchers found that it is quite usual, or "normal", for a variable to have a Gaussian distribution. Consequently, following the lead of noted British statistician Karl Pearson (1857–1936), the Gaussian distribution began to be referred to as the **normal distribution**.

correlation coefficient

■ Remember : area under $f(x)$ must be 1

$\longrightarrow$ pdf is a special kind of function

Greek

mu sigma

■ A continuous random variable X has a **normal distribution** with parameters μ and σ^2 (must have $\sigma > 0$) if the pdf of X is

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

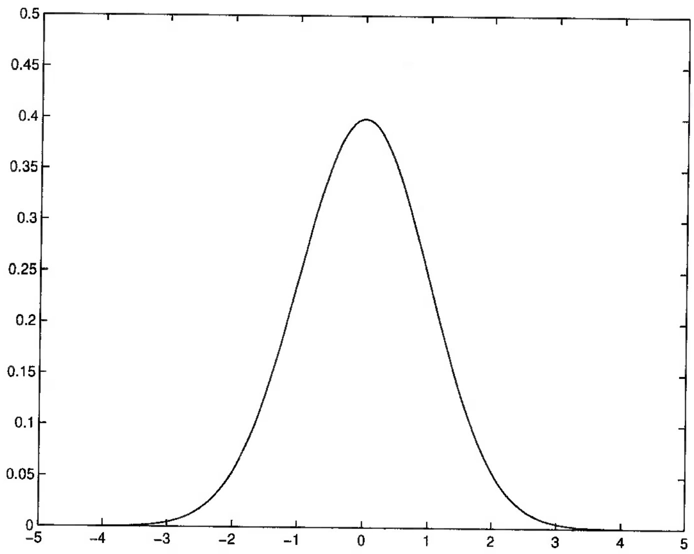
We write $X \sim N(\mu, \sigma^2)$.

mean variance

Notes —

- If $X \sim N(\mu, \sigma^2)$ then $E(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$
- The **standard normal distribution** has $\mu = 0$ and $\sigma^2 = 1$ and so we would write $Z \sim N(0,1)$ with pdf

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}$$



standard normal pdf

$\mu = 0, \sigma^2 = 1$

Standard normal distribution:

■ If $Z \sim N(0,1)$ then

$$E(Z) = 0$$
$$\text{var}(Z) = 1$$

The standard normal density (pdf) is often denoted by ϕ (Greek little "phi")

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

pdf

and the standard normal distribution function (cdf) is often denoted by Φ (Greek capital "phi")

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

which is *impossible* to write down in a closed form.

→ Use tables or software. (R)

Notes — (standardising)

- If $X \sim N(\mu, \sigma^2)$ then the random variable

  recall "z-scores" $$Z = \frac{X - \mu}{\sigma}$$

  is an $N(0,1)$ random variable

- If $Z \sim N(0,1)$ and $X = aZ + b$ then

  $$X \sim N(b, a^2)$$

X world

↕ standardise

Z world (no units)

In the "old days" ...

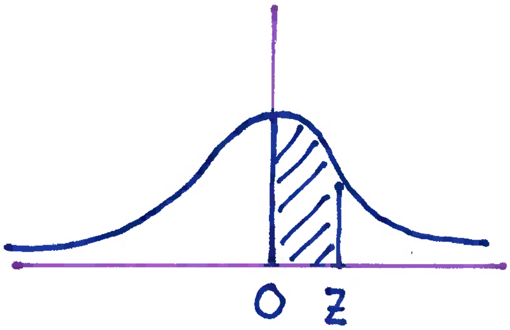


Table 1. Standard Normal Distribution

$P(0 \leq Z \leq z)$ where $Z \sim N(0,1)$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Table 2. Inverse Normal Probabilities

$p = P(0 \leq Z \leq z)$ where $Z \sim N(0,1)$

p	z	p	z	p	z	p	z	p	z
0.00	0.0000	0.10	0.2533	0.20	0.5244	0.30	0.8416	0.40	1.2816
0.01	0.0251	0.11	0.2793	0.21	0.5534	0.31	0.8779	0.41	1.3408
0.02	0.0502	0.12	0.3055	0.22	0.5828	0.32	0.9154	0.42	1.4051
0.03	0.0753	0.13	0.3319	0.23	0.6128	0.33	0.9542	0.43	1.4758
0.04	0.1004	0.14	0.3585	0.24	0.6433	0.34	0.9945	0.44	1.5548
0.05	0.1257	0.15	0.3853	0.25	0.6745	0.35	1.0364	0.45	1.6449
0.06	0.1510	0.16	0.4125	0.26	0.7063	0.36	1.0803	0.46	1.7507
0.07	0.1764	0.17	0.4399	0.27	0.7388	0.37	1.1264	0.47	1.8808
0.08	0.2019	0.18	0.4677	0.28	0.7722	0.38	1.1750	0.48	2.0537
0.09	0.2275	0.19	0.4959	0.29	0.8064	0.39	1.2265	0.49	2.3263
0.470	1.8808	0.480	2.0537	0.490	2.3263	0.4990	3.0902	0.49990	3.7190
0.471	1.8957	0.481	2.0749	0.491	2.3656	0.4991	3.1214	0.49991	3.7455
0.472	1.9110	0.482	2.0969	0.492	2.4089	0.4992	3.1559	0.49992	3.7750
0.473	1.9268	0.483	2.1201	0.493	2.4573	0.4993	3.1947	0.49993	3.8082
0.474	1.9431	0.484	2.1444	0.494	2.5121	0.4994	3.2389	0.49994	3.8461
0.475	1.9600	0.485	2.1701	0.495	2.5758	0.4995	3.2905	0.49995	3.8906
0.476	1.9774	0.486	2.1973	0.496	2.6521	0.4996	3.3528	0.49996	3.9444
0.477	1.9954	0.487	2.2262	0.497	2.7478	0.4997	3.4316	0.49997	4.0128
0.478	2.0141	0.488	2.2571	0.498	2.8782	0.4998	3.5401	0.49998	4.1075
0.479	2.0335	0.489	2.2904	0.499	3.0902	0.4999	3.7190	0.49999	4.2649

In the "old days" (continued)...

Example 1.30 Suppose $Z \sim N(0,1)$.

Numbers in boxes represent numbers looked up in the standard normal table (see page 1080).

(a)
$$\begin{aligned} P(Z \leq 1.5) &= P(Z \leq 0) + P(0 \leq Z \leq 1.5) \\ &= 0.5 + \boxed{0.4332} \\ &= 0.9332 \end{aligned}$$

(all values 4dp)

(b)
$$\begin{aligned} P(Z \geq 1.5) &= 0.5 - P(0 \leq Z \leq 1.5) \\ &= 0.5 - \boxed{0.4332} \\ &= 0.0668 \end{aligned}$$

Alternatively, note that $P(Z \geq 1.5) = 1 - P(Z \leq 1.5)$

(c)
$$\begin{aligned} P(Z < -2) &= P(Z > 2) \\ &= 0.5 - P(0 \leq Z \leq 2) \\ &= 0.5 - \boxed{0.4772} \\ &= 0.0228 \end{aligned}$$

(d)
$$\begin{aligned} P(-2 \leq Z \leq 1) &= P(-2 \leq Z \leq 0) + P(0 \leq Z \leq 1) \\ &= P(0 \leq Z \leq 2) + P(0 \leq Z \leq 1) \\ &= \boxed{0.4772} + \boxed{0.3414} \\ &= 0.8185 \end{aligned}$$

(e) Find the value a such that $P(0 \leq Z \leq a) = 0.35$.

Directly from the inverse standard normal table (see page 1081).

$$a = 1.0364$$

dnorm gives pdf not probabilities

Question —

What does R know about $N(\mu, \sigma^2)$?

```
dnorm(x,mean,sd) # pdf (be careful with sd not var)
pnorm(q,mean,sd) # cdf
qnorm(p,mean,sd) # quantile
rnorm(n,mean,sd) # random variate
```

So in the previous example . . .

$P(Z \leqslant 1.5)$

```
> # --(a)--
> pnorm(1.5,0,1)
[1] 0.9331928
> # --(b)--
> 1-pnorm(1.5,0,1)
[1] 0.0668072
> # --(c)--
> pnorm(-2,0,1)
[1] 0.02275013
> # --(d)--
> pnorm(1,0,1)-pnorm(-2,0,1)
[1] 0.8185946
> # --(e)--
> qnorm(0.35+0.5,0,1)
[1] 1.036433
> # --check--
> pnorm(1.036433,0,1)
[1] 0.8499999
```

□

Example 1.31 Find the following probabilities for a normal random variable X that has a mean of 10 and a standard deviation of 2.

careful: σ not σ^2

(a) $P(10 \leq X \leq 12)$

```
> pnorm(12,10,2)-pnorm(10,10,2)
[1] 0.3413447
```

(b) $P(8 \leq X \leq 10)$

```
> pnorm(10,10,2)-pnorm(8,10,2)
[1] 0.3413447
```

(c) $P(11 \leq X \leq 14)$

```
> pnorm(14,10,2)-pnorm(11,10,2)
[1] 0.2857874
```

(d) $P(9 \leq X \leq 11)$

```
> pnorm(11,10,2)-pnorm(9,10,2)
[1] 0.3829249
```

(e) $P(7 \leq X \leq 13)$

```
> pnorm(13,10,2)-pnorm(7,10,2)
[1] 0.8663856
```

□

Practical examples :

Example 1.32 A firm that manufactures and bottles apple juice has a machine that automatically fills bottles with 1.6 litres of juice. (The bottle can hold up to 1.7 litres.) Over a long period, the average amount dispensed into the bottle has been 1.6 litres. However, there is some variability in how much juice is put in each bottle; the distribution of these amounts has a standard deviation of 0.1 litres. If the litres of fill per bottle can be assumed to follow a normal distribution, find the probability that the machine will overflow any one bottle.

Let X denote the amount of liquid (in litres) dispensed into one bottle by the filling machine. Then $X \sim N(1.6, 0.1^2)$.

A bottle will overflow if the machine attempts to put more than 1.7 litres in it. We need to find $P(X > 1.7)$.

```
> 1-pnorm(1.7,1.6,0.1)
[1] 0.1586553
```

□

Example 1.33 At a temperature of $25°$C, the resistances of a type of thermistor are normally distributed with a mean of 10000 ohms and a standard deviation of 4000 ohms. The thermistors are to be checked, and those having resistances between 8000 and 15000 ohms are to shipped to a vendor. What proportion of these thermistors will actually be shipped?

Let X denote the resistance of a thermistor. Then $X \sim N(10000, 4000^2)$.

We need to find $P(8000 \leq X \leq 15000)$.

```
> pnorm(15000,10000,4000)-pnorm(8000,10000,4000)
[1] 0.5858127
```

□

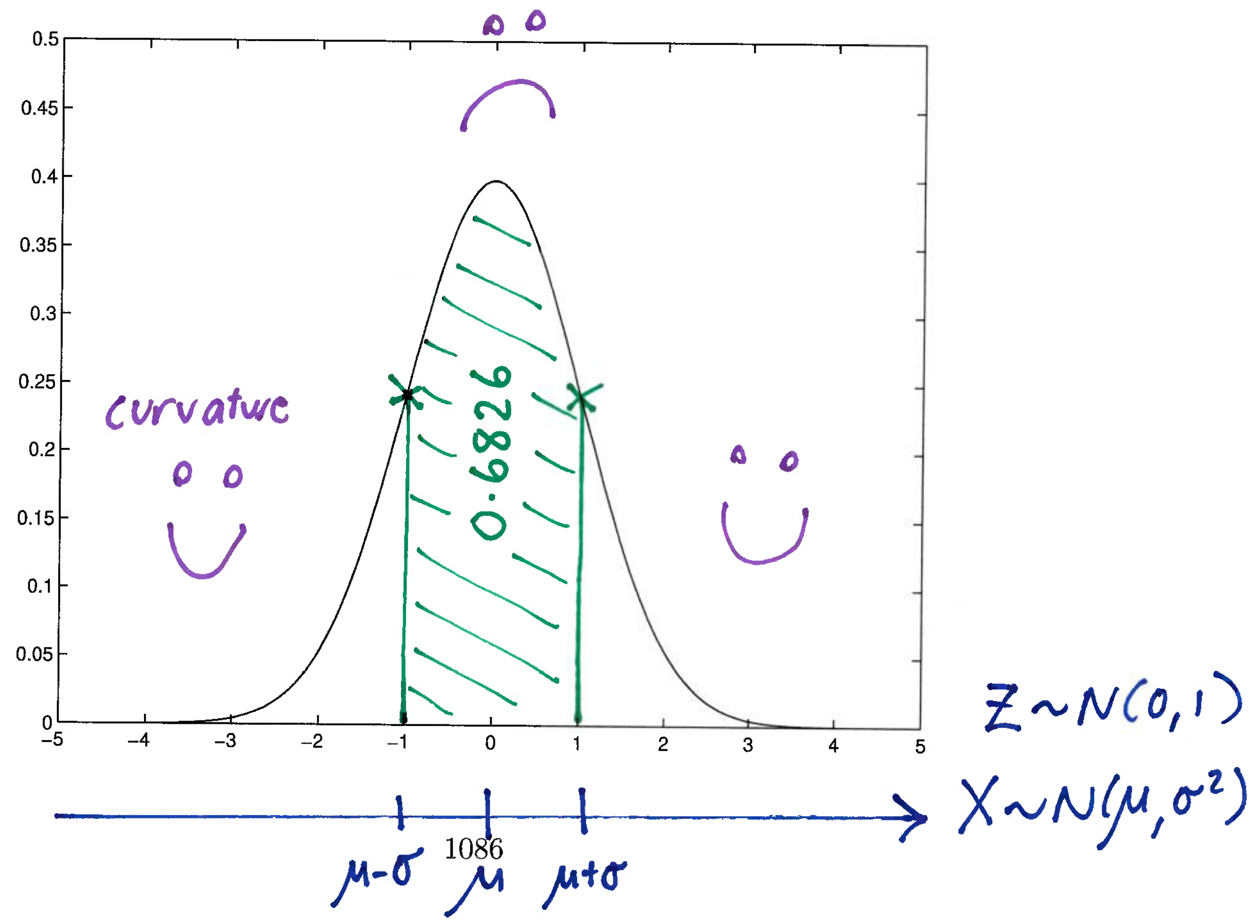
Properties of Normal Distribution —

- Points of inflection (change of curvature) are one standard deviation from the mean.
- $P(-1 < Z < 1) \;=\; 2(0.3413) \;=\; 0.6826$

  Just over $\frac{2}{3}$ of normal observations lie within one standard deviation of the mean (between points of inflection on the curve).
- $P(-2 < Z < 2) \;=\; 2(0.4772) \;=\; 0.9544$

  Just over 95% of normal observations lie within 2 standard deviations of the mean.
- $P(-3 < Z < 3) = 2(0.4987) = 0.9974$

  Or "almost all" normal observations lie within 3 standard deviations of the mean. 99.74%



Summary

- For a continuous random variable X
  - the $\boxed{\text{CDF is } \; F(u) = P(X \leq u)}$
  - which is given by the *area under* the PDF $f(x)$ to the left of $x = u$ and above the x-axis.

cdf gives probability

- Expected value (mean) and variance

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \qquad E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x$$

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

Uniform	Parameters:	a and b where $a < b$
$X \sim U(a, b)$	PDF:	$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$
	Mean:	$\frac{a+b}{2}$
	Variance:	$\frac{(b-a)^2}{12}$

Normal	Parameters:	μ and σ^2 where $\sigma > 0$
$X \sim N(\mu, \sigma^2)$	PDF:	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$
	Mean:	μ
	Variance	σ^2

X world ⟷ Z world

- Standardising: If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.
- `dunif`, `punif`, `qunif`, `runif`
- `dnorm`, `pnorm`, `qnorm`, `rnorm`

Week 5 lab worksheet

A useful way of thinking ...

Normal distribution:

$$Z = \frac{X - \mu}{\sigma}$$

X-world

$X \sim N(\mu, \sigma^2)$

Z-world

$Z \sim N(0, 1)$

$$X = \mu + \sigma Z$$

$p = \text{pnorm}(x)$

$x = \text{qnorm}(p)$

$p = \text{pnorm}(z)$

$z = \text{qnorm}(p)$

p-world

$0 \leq \text{probability} \leq 1$

"quantiles"