5.2 Continuous random variables

It is often convenient to think of a random variable as having a whole (continuous) interval for its set of possible values.

The devices used to describe continuous probability distributions differ from those that describe discrete probability distributions.

Examples of continuous random variables:

Z = the amount of torque required to loosen the next bolt (not rounded)

T = the time you'll wait for the next bus

C = the outdoor temperature at 3:17pm tomorrow

L = the length of the next manufactured part

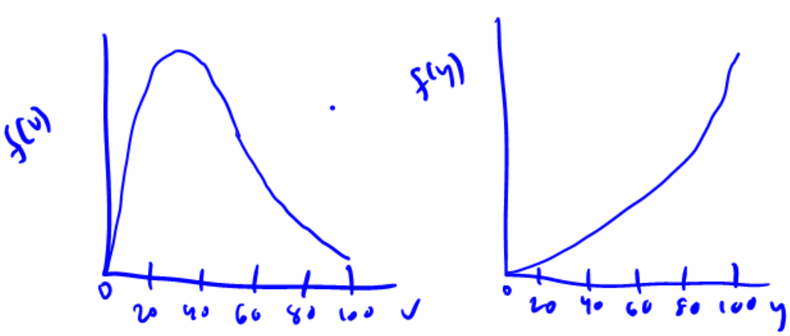
V = % yield of the next run of a process

Y = % yield of the next run of some better process

how do we mathematically distinguish V and Y?

- each has the same range $0\% \le V, Y \le 100\%$
- there are uncountably many possible values in this range.

Distribution!



The process Y will yield more product per run on average than process V.

5.2.1 Probability density functions and cumulative distribution functions

A *probability density function (pdf)* is the continuous analogue of a discrete random variable's probability mass function (pmf).

Definition 5.12. A *probability density function (pdf)* for a continuous random variable X is a 1 nonnegative function $f(x)$ with

$$\int_{-\infty}^{\infty} f(x) = 1$$

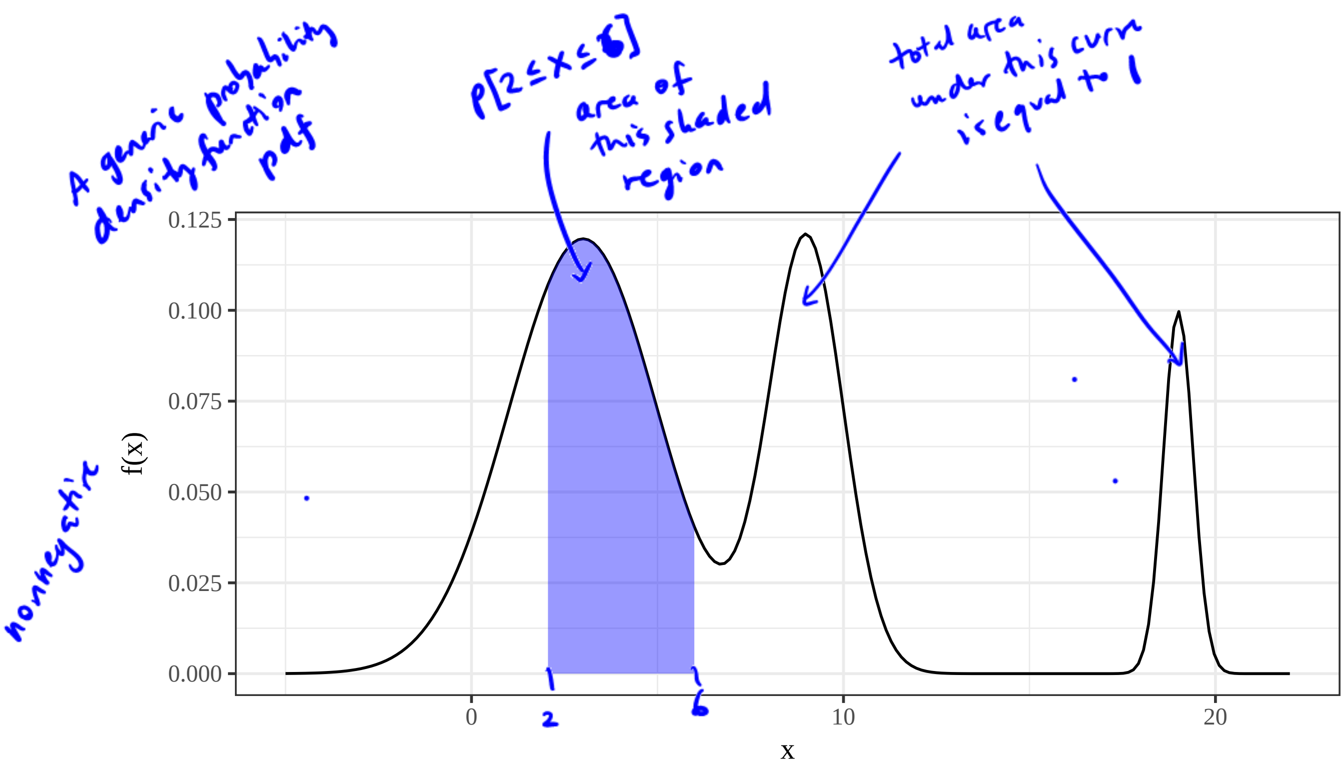
2

and such that for all $a \leq b$,

$$P[a \leq X \leq b] = \int_{a}^{b} f(x)dx.$$

3

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P[a \leq X \leq b] = \int_a^b f(x)dx$, $a \leq b$



Example 5.17 (Compass needle)**.** Consider a de-magnetized compass needle mounted at its center so that it can spin freely. It is spun clockwise and when it comes to rest the angle, θ, from the vertical, is measured. Let

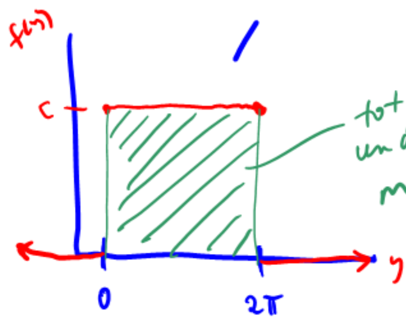
$$Y = \text{the angle measured after each spin in radians}$$

What values can Y take? $[0, 2\pi]$

What form makes sense for $f(y)$?

$$f(y) \doteq \begin{cases} c & y \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases}$$

Y has a positive probability between 0 and 2π and it is equally likely to land on any angle (can spin freely)



total area under the curve $f(y)$ must equal 1

✓

we say that Y is distributed Unif$(0, 2\pi)$

If this form is adopted, that what must the pdf be?

$$1 = \int_{-\infty}^{\infty} f(y)\,dy = \underbrace{\int_{-\infty}^{0} 0\,dy}_{0} + \int_0^{2\pi} c\,dy + \underbrace{\int_{2\pi}^{\infty} 0\,dy}_{0}$$

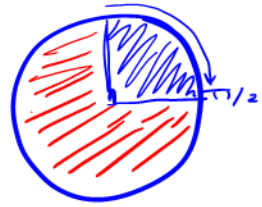
$$= cy\Big|_0^{2\pi} = 2\pi c$$

$$\Rightarrow c = 1/2\pi$$

$$\text{Thus} \quad f(y) = \begin{cases} \frac{1}{2\pi} & 0 \le y \le 2\pi \\ 0 & \text{otherwise} \end{cases} \quad ✓$$

Using this pdf, calculate the following probabilities:

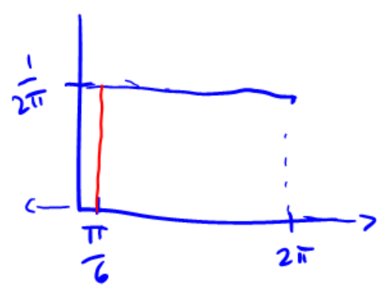
1. $P[Y < \frac{\pi}{2}] = P[-\infty < y < \frac{\pi}{2}]$
$$= \int_{-\infty}^{\pi/2} f(y)\,dy$$
$$= \int_{-\infty}^{0} 0\,dy + \int_{0}^{\pi/2} \frac{1}{2\pi}\,dy$$
$$= \frac{1}{2\pi} \cdot \frac{\pi}{2} = \frac{1}{4}$$



2. $P[\frac{\pi}{2} < Y < 2\pi] = \int_{\pi/2}^{2\pi} f(y)\,dy$
$$= \int_{\pi/2}^{2\pi} \frac{1}{2\pi}\,dy$$
$$= \frac{1}{2\pi} \cdot 2\pi - \frac{1}{2\pi} \cdot \frac{\pi}{2} = 1 - \frac{1}{4} = 3/4$$

3. $P[\frac{\pi}{6} < Y < \frac{\pi}{4}] = \int_{\pi/6}^{\pi/4} f(y)\,dy$
$$= \int_{\pi/6}^{\pi/4} \frac{1}{2\pi}\,dy$$
$$= \frac{\pi}{4} \cdot \frac{1}{2\pi} - \frac{\pi}{6} \cdot \frac{1}{2\pi} = \frac{1}{24} \approx .04167$$

4. $P[Y = \frac{\pi}{6}] = P[\frac{\pi}{6} \le y \le \frac{\pi}{6}]$
$$= \int_{\frac{\pi}{6}}^{\pi/6} f(y)\,dy = \int_{\pi/6}^{\pi/6} \frac{1}{2\pi}\,dy$$
$$= \frac{1}{2\pi}\left(\frac{\pi}{6} - \frac{\pi}{6}\right) = 0$$



In fact, for any $\underline{\underline{\text{continuous}}}$ random variable X, and any real number a,

$$P[X = a] = 0.$$

Definition 5.13. The *cumulative distribution function (cdf)* of a continuous random variable X is a function F such that

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(t)dt$$

$F(x)$ is obtained from $f(x)$ by integration, and applying the fundamental theorem of calculus yields

$$\frac{d}{dx}F(x) = f(x).$$

That is, $f(x)$ is obtained from $F(x)$ by differentiation.

As with discrete random variables, F has the following properties:

1. $F(x) \geq 0$ for all real x.

2. F is monotonically increasing

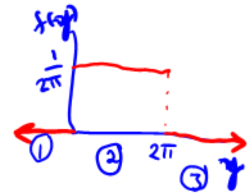
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

different

4. F is continuous.

Example 5.18 (Compass needle, cont'd)**.** Recall the compass needle example, with

$$f(y) = \begin{cases} \frac{1}{2\pi} & 0 \le y \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$



Find the cdf.

For $y < 0$

$$F(y) = P[Y \le y] = \int_{-\infty}^{y} f(t)\,dt = \int_{-\infty}^{y} 0\,dt = 0$$

For $0 \le y \le 2\pi$

$$F(y) = P[Y \le y] = \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{y} f(t)\,dt$$

$$= \cancel{\int_{-\infty}^{0} 0\,dt}^{\,0} + \int_0^y \frac{1}{2\pi}\,dt = \frac{y}{2\pi}$$

For $y > 2\pi$

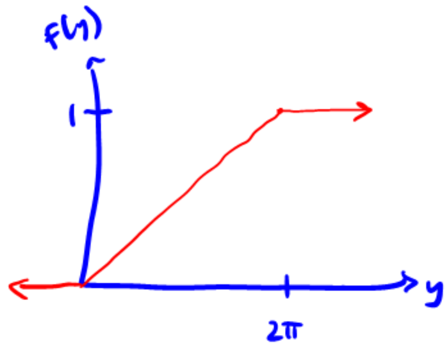
$$F(y) = P(Y \le y) = \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{2\pi} f(t)\,dt + \int_{2\pi}^{y} f(t)\,dt$$

$$\int_{-\infty}^{y} f(t)\,dt$$

$$= \cancel{\int_{-\infty}^{0} 0\,dt}^{\,0} + \int_{0}^{2\pi} \frac{1}{2\pi}\,dt + \cancel{\int_{2\pi}^{y} 0\,dt}^{\,0}$$

$$= 2\pi \cdot \frac{1}{2\pi} = 1$$

$$\Rightarrow F(y) = \begin{cases} 0 & y < 0 \\ y/2\pi & 0 \le y \le 2\pi \leftarrow \\ 1 & y > 2\pi \end{cases}$$



Calculate the following using the cdf:

$F(1.5)$

$0 \le 1.5 \le 2\pi \Rightarrow \quad F(1.5) = \frac{1.5}{2\pi} = \frac{3}{4\pi} \approx 0.2387$

$P[Y \le \frac{4\pi}{5}] = F(\frac{4\pi}{5}) = \frac{4\pi/5}{2\pi} = \frac{2}{5} = 0.4$

$1 = P(X > x) + P(X \le x)$

$P(X \text{ is on the number line}) = 1$
$= P(X > x \text{ or } X \le x) = P(X > x) + P(X \le x)$

x

For any r.v. X, discrete or continuous, $P(X > x) = 1 - P(X \le x)$

$P[Y > 5] = 1 - P[Y \le 5] = 1 - F(5) = 1 - \frac{5}{2\pi} \approx 0.2042$

$P[\frac{\pi}{3} < Y \le \frac{\pi}{2}] = \int_{\pi/3}^{\pi/2} f(y)\,dy = \int_{-\infty}^{\pi/2} f(y)\,dy - \int_{-\infty}^{\pi/3} f(y)\,dy = F(\frac{\pi}{2}) - F(\frac{\pi}{3})$

$= \frac{\pi}{2} \cdot \frac{1}{2\pi} - \frac{\pi}{3} \cdot \frac{1}{2\pi}$

$= \frac{1}{12} \approx 0.08333$

$\frac{\pi}{3}$ $\frac{\pi}{2}$

5.2.2 Quantiles

Recall: The p quantile of a distribution of data is a number such that a fraction p of the distribution lies the left and a fraction 1−p of the dist'n lies to the right.

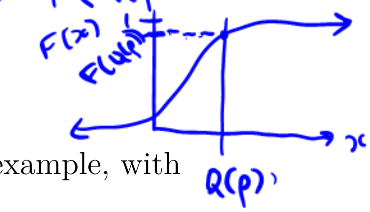
Definition 5.14. The *p-quantile of a random variable*, X, is the number $Q(p)$ such that $P[X \le Q(p)] = p$.

In terms of the cumulative distribution function (for a continuous random variable),

$$\rightarrow p = \underline{P[X \le Q(p)]} = F(Q(p))$$

$$\text{i.e.} \quad F^{-1}(p) = Q(p)$$

$p = F(Q(p))$



Example 5.19 (Compass needle, cont'd)**.** Recall the compass needle example, with

$$f(x) = \begin{cases} \frac{1}{2\pi} & 0 \le y \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$

$Q(.95)$:

$$\begin{aligned} 0.95 &= P(Y \le Q(.95)) \\ &= \int_{-\infty}^{Q(.95)} f(y)\,dy \\ &= \int_{-\infty}^{0} 0\,dy + \int_0^{Q(.95)} \frac{1}{2\pi}\,dy \\ &= \frac{1}{2\pi} Q(.95) \end{aligned}$$

$$\Rightarrow Q(.95) = 0.95 \cdot 2\pi \approx 5.9690$$

On average, 95% of the needle spins will be below 5.9690 radians.

You can also calculate quantiles directly from the cdf.

$$F(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\pi} y & 0 \le y \le 2\pi \\ 1 & \text{otherwise} \end{cases}$$

$Q(.25)$:

$$0.25 = P[Y \le Q(.25)] = F(Q(.25)) = \frac{Q(.25)}{2\pi}$$

$$\Rightarrow Q(.25) = .25 \cdot 2\pi = \frac{\pi}{2} \approx 1.5708 \text{ radians}$$

"median"

$Q(.5)$ $$0.5 = P[Y \le Q(.5)] = F(Q(.5)) = \frac{Q(.5)}{2\pi}$$

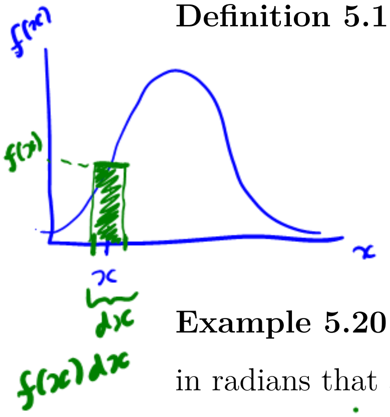
$$\Rightarrow Q(.5) = .5 \cdot 2\pi = \pi \approx 3.1416 \text{ radians.}$$

5.2.3 Means and variances for continuous distributions

It is possible to summarize continuous probability distributions using

1. plot of probability density function $f(x)$ [kind of idealized probability histogram]
2. mean (measure of location)
3. variance (measure of spread)

Definition 5.15. The *mean* or *expected value* of a continuous random variable X is



Sometimes denoted μ $\longrightarrow$

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x) dx.$$

reasoning: the probability in a small interval of length dx around x is approximately $f(x)dx$.

So $\mathrm{E}X \approx \sum x f(x) dx$

$\mathrm{E}X = \lim_{dx \to 0} \sum x f(x) dx = \int x f(x) dx$

Example 5.20 (Compass needle, cont'd)**.** Calculate $\mathrm{E}Y$ where Y is the angle from vertical in radians that a spun needle lands on.

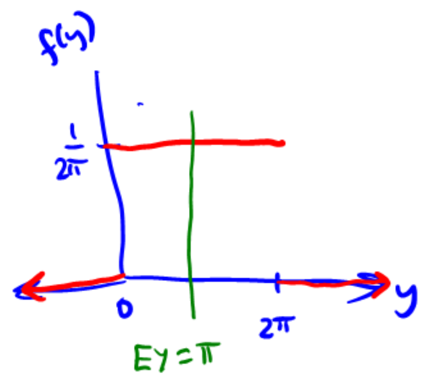
$$f(y) = \begin{cases} \frac{1}{2\pi} & 0 \le y \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$

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

$$= \int_{-\infty}^{0} y f(y) dy + \int_{0}^{2\pi} y f(y) dy + \int_{2\pi}^{\infty} y f(y) dy$$

$$= \cancel{\int_{-\infty}^{0} y \cdot 0 \, dy} + \int_{0}^{2\pi} y \cdot \frac{1}{2\pi} dy + \cancel{\int_{2\pi}^{\infty} y \cdot 0 \, dy}$$

$$= \left[\frac{y^2}{4\pi} \right]_0^{2\pi} = \frac{(2\pi)^2}{4\pi} = \pi$$



EY is the "center of mass" of the distribution

Example 5.21. Calculate EX where X follows the following distribution

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{3}e^{-x/3} & x \geq 0 \end{cases}$$

$$EX = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{\infty} x f(x)\,dx$$

$$= \int_{-\infty}^{0} x \cdot \cancel{0}\,dx + \int_{0}^{\infty} x \tfrac{1}{3} e^{-x/3}\,dx$$

$$\begin{array}{c|c} + \; x & \frac{1}{3}e^{-x/3} \\ \hline - \; 1 & -e^{-x/3} \\ + \; 0 & 3e^{-x/3} \end{array}$$

$$= -xe^{-x/3} - 3e^{-x/3} \Big|_0^{\infty}$$

x goes to ∞ slower than e^{-x} goes to 0

$$= \lim_{x\to\infty} -xe^{-x/3} + \cancel{0 e^{-0/3}}^{\,0} - \lim_{x\to\infty} 3e^{-x/3} + 3\underbrace{e^{-0/3}}_{=3}$$

$$= 0 + 0 - 0 + 3$$

$$= 3$$

Definition 5.16. The *variance* of a continuous random variable X is

$$\text{Var}X = \overbrace{\int_{-\infty}^{\infty} (x - \text{E}X)^2 f(x)dx}^{=E((X-EX)^2)} = \underbrace{\overbrace{\int_{-\infty}^{\infty} x^2 f(x)dx}^{=E(X^2)} - (\text{E}X)^2}_{\text{computationally easier}}.$$

The *standard deviation* of X is $\sqrt{\text{Var}X}$.

Example 5.22 (Library books). Let X denote the amount of time for which a book on 2-hour hold reserve at a college library is checked out by a randomly selected student and suppose its density function is

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Calculate EX and Var$X$.

$$EX = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{2} x f(x)\,dx + \int_{2}^{\infty} x f(x)\,dx$$

$$= \cancel{\int_{-\infty}^{0} x \cdot 0\,dx} + \int_{0}^{2} x \cdot 0.5x\,dx + \cancel{\int_{2}^{\infty} x \cdot 0\,dx}$$

$$= \int_{0}^{2} 0.5x^2\,dx$$

$$= \left[0.5\frac{x^3}{3}\right]_0^2 = \frac{8}{6} \approx 1.333$$

$$(EX)^2 \ne E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \int_0^2 x^2\, 0.5x\,dx$$

$$= 0.5\frac{x^4}{4}\Big|_0^2 = \frac{16}{8} = 2$$

$$\Rightarrow \text{Var}\, X = E(X^2) - (EX)^2 = 2 - \left(\frac{8}{6}\right)^2 = \frac{2}{9}$$

Example 5.23 (Ecology)**.** An ecologist wishes to mark off a circular sampling region having radius 10m. However, the radius of the resulting region is actually a random variable R with pdf

$$f(r) = \begin{cases} \frac{3}{2}(10-r)^2 & 9 \le r \le 11 \\ 0 & \text{otherwise} \end{cases}$$

Calculate ER and SD(R).

$$ER = \int_{-\infty}^{\infty} r\, f(r)\, dr$$

$$= \int_9^{11} r \cdot \frac{3}{2}(10-r)^2\, dr$$

$$= \frac{3}{2}\int_9^{11} (100r - 20r^2 + r^3)\, dr$$

$$= \frac{3}{2}\left[100\frac{r^2}{2} - 20\frac{r^3}{3} + \frac{r^4}{4}\right]_9^{11}$$

$$= \frac{3}{2}\left[100\frac{11^2}{2} - 20\frac{11^3}{3} + \frac{11^4}{4} - 100\frac{9^2}{2} + 20\frac{9^3}{3} - \frac{9^4}{4}\right]$$

$$= 10$$

$$E(R^2) = \int_{-\infty}^{\infty} r^2 f(r)\, dr$$

$$= \int_9^{11} r^2\, \frac{3}{2}(10-r)^2\, dr$$

$$= \frac{3}{2}\int_9^{11} 100r^2 - 20r^3 + r^4\, dr$$

$$= \frac{3}{2}\left[100\frac{r^3}{3} - 20\frac{r^4}{4} + \frac{r^5}{5}\right]_9^{11} = 100.6$$

$$\text{Var}R = E(R^2) - (ER)^2 = 100.6 - 10^2 = 0.6$$

$$SD(R) = \sqrt{\text{Var}R} = \sqrt{6.6} \approx 0.7746$$

Why does $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$?

For any function g of a continuous random variable X,

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$ where $f(x)$ is the pdf of X.

So,

$$EX^2 = \int_{-\infty}^{\infty} x^2 f(x) dx$$ where $g(X) = X^2$

Example 5.24 (Ecology, cont'd). Calculate the expected *area* of the circular sampling region.

R = the radius of the circular sampling region

A = the area of the circular sampling region

$= \pi R^2$

$$EA = E\pi R^2 = \int_{-\infty}^{\infty} \pi r^2 f(r)\, dr$$ where $g(R) = \pi R^2$, the sampling area

$$= \pi \int_{-\infty}^{\infty} r^2 f(r) dr$$

$$= \pi \cdot 100.6$$

For a linear function, $g(X) = aX + b$, where a and b are constants,

$$E(aX + b) = \int_{-\infty}^{\infty} (ax+b) f(x)\, dx$$

$$= \int_{-\infty}^{\infty} axf(x)dx + \int_{-\infty}^{\infty} bf(x)\, dx$$

$$= a \underbrace{\int_{-\infty}^{\infty} xf(x)dx}_{EX} + b \underbrace{\int_{-\infty}^{\infty} f(x)\, dx}_{1}$$

$$= aEX + b$$

$$\mathrm{Var}(aX + b) = E\left[(aX+b)^2\right] - \left[E(aX+b)\right]^2$$

$$= E\left[a^2X^2 + 2abX + b^2\right] - \left[aEX + b\right]^2$$

$$= \int_{-\infty}^{\infty} (a^2x^2 + 2abx + b^2) f(x)\, dx - \left(a^2[EX]^2 + 2ab\,EX + b^2\right)$$

$$= a^2E(X^2) + 2a\cancel{b}EX + \cancel{b^2} - \left(a^2(EX)^2 + 2\cancel{ab}EX + \cancel{b^2}\right)$$

$$= a^2E(X^2) - a^2(EX)^2$$

$$= a^2\left[E(X^2) - (EX)^2\right]$$

$$= a^2\, \mathrm{Var}X$$

Example 5.25 (Ecology, cont'd)**.** Calculate the expected value and variance of the *diameter* of the circular sampling region. D = diameter of circular sampling region

$= 2 \cdot R$

$$g(R) = \underbrace{2}_{a} \cdot R + \underbrace{0}_{b}$$

$$ED = Eg(R) = E[2 \cdot R + 0] = 2ER + 0 = 20$$

$$\mathrm{Var}\, D = \mathrm{Var}(g(R)) = \mathrm{Var}(2 \cdot R + 0) = 2^2\, \mathrm{Var}\, R = 4 \cdot .6 = 2.4$$

Definition 5.17. *Standardization* is the process of transforming a random variable, X, into the signed number of standard deviations by which it is is above its mean value.

$$Z = \frac{X - \mathrm{E}X}{\mathrm{SD}(X)}$$

subtracting the mean and dividing by the s.d.

Z has mean 0

$$\begin{aligned} EZ &= E\left[\frac{X - EX}{SD(X)}\right] \\ &= E\left[\frac{1}{SD(X)}X - \frac{EX}{SD(X)}\right] \\ &= \frac{1}{SD(X)}EX - \frac{EX}{SD(X)} \\ &= 0 \end{aligned}$$

Z has variance (and standard deviation) 1

$$\begin{aligned} \mathrm{Var}\,Z &= \mathrm{Var}\left(\frac{X - EX}{SD(X)}\right) \\ &= \mathrm{Var}\left(\underbrace{\frac{1}{SD(X)}}_{a}X \underbrace{- \frac{EX}{SD(X)}}_{b}\right) \\ &= \left(\frac{1}{SD(X)}\right)^2 \mathrm{Var}X \\ &= \frac{1}{\mathrm{Var}X}\mathrm{Var}X = 1 \end{aligned}$$

5.2.4 A special continuous distribution

Just as there are a number of useful discrete distributions commonly applied to engineering problems, there are a number of standard continuous probability distributions.

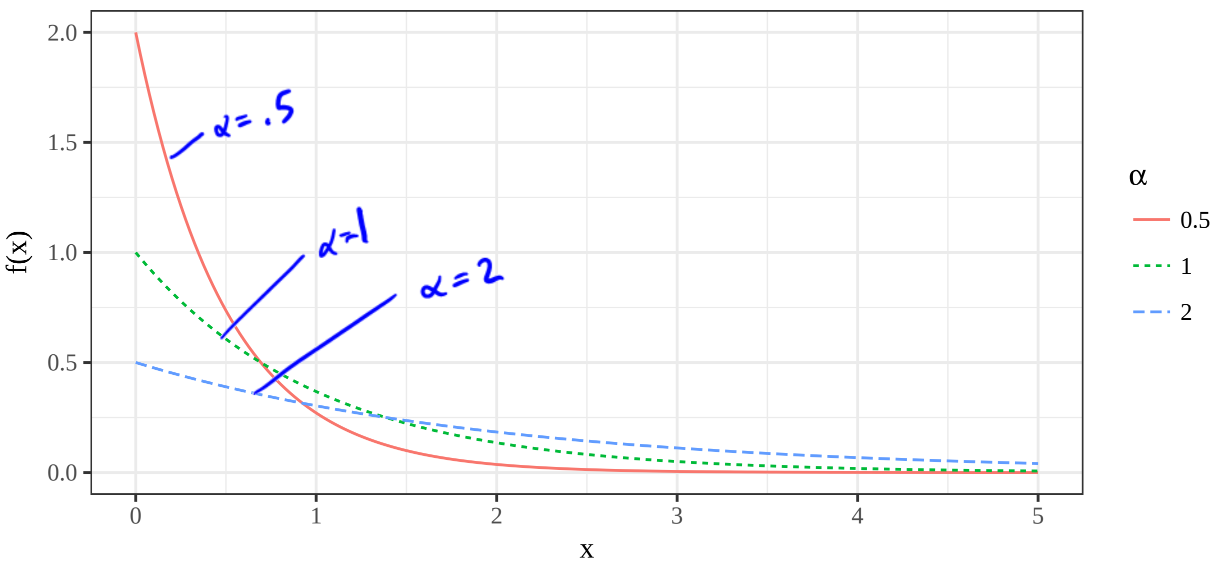
1 parameter

Definition 5.18. The *exponential*(α) *distribution* is a continuous probability distribution with probability density function

example 5.21 is an Exp(3) random variable

$$f(x) = \begin{cases} \frac{1}{\alpha} e^{-x/\alpha} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

for $\alpha > 0$.



An Exp(α) random variable measures the waiting time until a specific event that has an equal chance of happening at any point in time.

This is kind of like a continuous version of the geometric dsn.

Examples:

- Time between your arrival at the bus stop and the moment your bus comes.
- Time until next person walks inside the library
- Time until the next car accident on a stretch of highway

It is straightforward to show for $X \sim \text{Exp}(\alpha)$,

1. $\mu = \text{E}X = \int\limits_0^\infty x \frac{1}{\alpha} e^{-x/\alpha} dx = \alpha$

2. $\sigma^2 = \text{Var}X = \int\limits_0^\infty (x-\alpha)^2 \frac{1}{\alpha} e^{-x/\alpha} dx = \alpha^2$

Further, $F(x)$ has a simple formulation:

For $x < 0$:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$$

For $x \ge 0$:

$$\begin{aligned} F(x) &= P(X \le x) \\ &= \int_{-\infty}^{x} f(t)\,dt \\ &= \cancel{\int_{-\infty}^{0} 0\,dt}_{0} + \int_0^x \frac{1}{\alpha} e^{-t/\alpha}\,dt \\ &= \left[-e^{-t/\alpha}\right]_0^x \\ &= -e^{-x/\alpha} + 1 = 1 - e^{-x/\alpha} \end{aligned}$$

$$\Rightarrow F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x/\alpha} & x \ge 0 \end{cases}$$

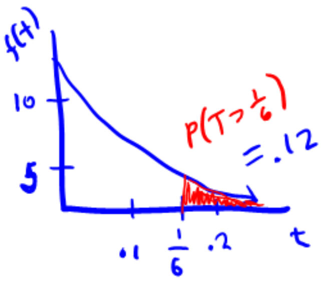
Example 5.26 (Library arrivals, cont'd)**.** Recall the example the arrival rate of students at Parks library between 12:00 and 12:10pm early in the week to be about 12.5 students per minute. That translates to a $1/12.5 = .08$ minute average waiting time between student arrivals.

$\frac{\text{students}}{\text{minute}} \Rightarrow \frac{\text{minutes}}{\text{student}}$

Consider observing the entrance to Parks library at exactly noon next Tuesday and define the random variable

$$T = \text{the waiting time (min) until the first student passes through the door.}$$

Using $T \sim \text{Exp}(.08)$, what is the probability of waiting more than 10 seconds (1/6 min) for the first arrival?

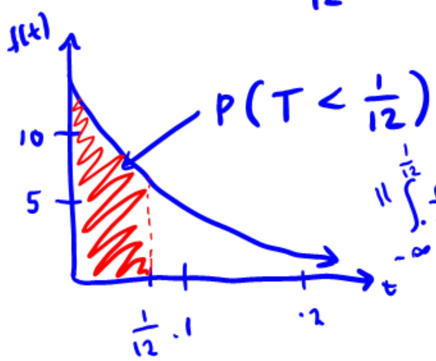


$$\begin{aligned} P(T > \tfrac{1}{6}) &= 1 - P(T \le \tfrac{1}{6}) \\ &= 1 - F(\tfrac{1}{6}) \\ &= 1 - \left(1 - e^{-\frac{1}{6}\cdot\frac{1}{.08}}\right) = .12 \end{aligned}$$

What is the probability of waiting less than 5 seconds?

cdf: $F(x) = 1 - e^{-x/\alpha}$ $x \ge 0$

5 seconds $= \frac{1}{12}$ minute



$$P(T < \tfrac{1}{12}) = F(\tfrac{1}{12}) = \left(1 - e^{-\frac{1}{12}\cdot\frac{1}{.08}}\right) \approx .6471$$

$$= \int_{-\infty}^{\frac{1}{12}} f(x)\,dx$$

Note: since T is continuous, $P(T = c) = 0$ for all c. Meaning,

$P(T < c) = P(T \le c) = F(c)$.

5.2.5 The Normal distribution

used for Normal QQ plots

We have already seen the normal distribution as a "bell shaped" distribution, but we can formalize this.

Definition 5.19. The *normal* or *Gaussian*$(\underbrace{\mu, \sigma^2}_{\text{2 parameters}})$ distribution is a continuous probability distribution with probability density

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

for $\sigma > 0$. and $\mu \in \mathbb{R}$ (literally any real number)

A normal random variable is (often) a finite average of many repeated, independent, identical trials.

- mean width of the next 50 hexamine pellets
- mean height of 30 students
- Total % yield of the next 40 runs of a chemical process.

It is not obvious, or easy to show but

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

2. $\mathrm{E}X = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \mu$

3. $\mathrm{Var}X = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2$

the parameters in the distribution are actually the mean and variance of the distribution.

The Calculus I methods of evaluating integrals via anti-differentiation will fail when it comes to normal densities. They do not have anti-derivatives that are expressible in terms of elementary functions.

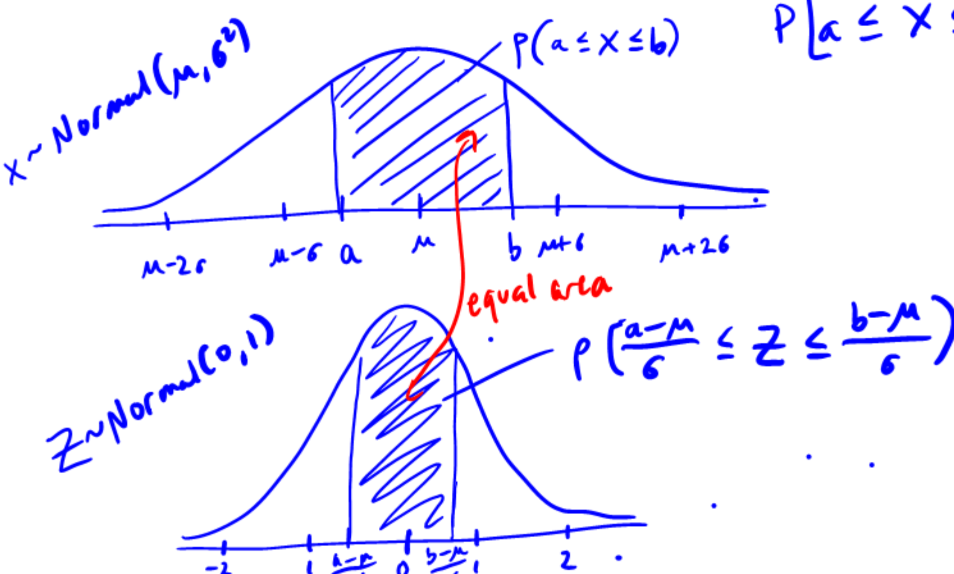
What to do? Computer or use tables of values.

The use of tables for evaluating normal probabilities depends on the following relationship. If $X \sim \text{Normal}(\mu, \sigma^2)$, "change of variable" $Z = \frac{X-\mu}{\sigma}$

$$P[a \leq X \leq b] = \underbrace{\int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz} = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]$$

where $Z \sim \text{Normal}(0, 1)$.

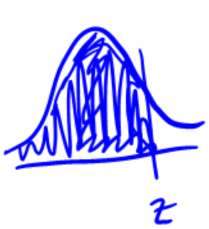
$$P[a \leq X \leq b] = P[a-\mu \leq X-\mu \leq b-\mu]$$
$$= P\left[\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right]$$
$$= P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]$$



Definition 5.20. The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *standard normal distribution*.

① Question about probability using Normal(μ,σ^2) → ② Transform to a question about $N(0,1)$ → ③ look up in tables.

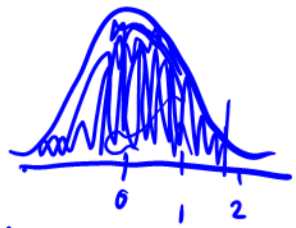
So, we can find probabilities for all normal distributions by tabulating probabilities for only the standard normal distribution. We will use a table of the **standard normal cumulative probability function**.



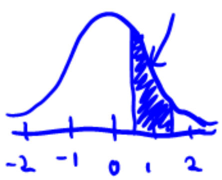
$$\Phi(z) = F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2} dt. = P[Z \leq z]$$

Example 5.27 (Standard normal probabilities)**.** $P[Z < 1.76]$

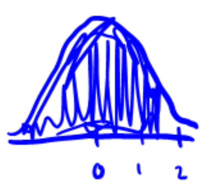
$$P[Z < 1.76] = P[Z \le 1.76]$$
$$= \Phi(1.76)$$
$$= .9608$$



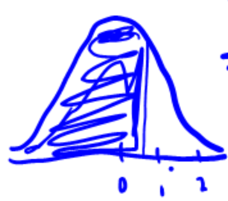
$$P[.57 < Z < 1.32] = \int_{.57}^{1.32} f(z)\,dz = \int_{-\infty}^{1.32} f(z)\,dz - \int_{-\infty}^{.57} f(z)\,dz$$
$$= P(Z \le 1.32) - P(Z \le .57)$$
$$= \Phi(1.32) - \Phi(.57)$$
$$= .9066 - .7157$$
$$= .19$$

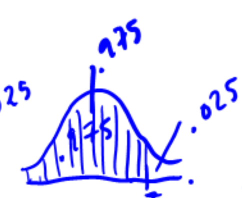


=



−





we are looking for z such that

$$P(Z \le z) = .975 = \Phi(z)$$

We can also do it in reverse, find z such that $P[-z < Z < z] = .95$.

$$\Rightarrow z = 1.96$$

Example 5.28 (Baby food)**.** J. Fisher, in his article Computer Assisted Net Weight Control (*Quality Progress*, June 1983), discusses the filling of food containers with strained plums and tapioca by weight. The mean of the values portrayed is about 137.2g, the standard deviation is about 1.6g, and data look bell-shaped. Let

$$W = \text{the next fill weight.}$$

And let $W \sim N(137.2, 1.6^2)$

Let's find the probability that the next jar contains less food by mass than it's supposed to (declared weight = 135.05g).

"standardize" W

$$P(W < 135.05) = P\left(\frac{W - 137.2}{1.6} < \frac{135.05 - 137.2}{1.6}\right)$$

$$= P(Z < -1.34)$$

$$= \Phi(-1.34)$$

$$= .0901$$

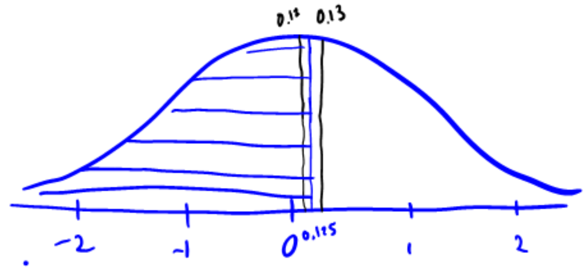
So there is about a 9% chance the next jar contains less food by mass than it is supposed to.

Table B.3
Standard Normal Cumulative Probabilities

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985

Example 5.29 (More normal probabilities)**.** Using the standard normal table, calculate the following:

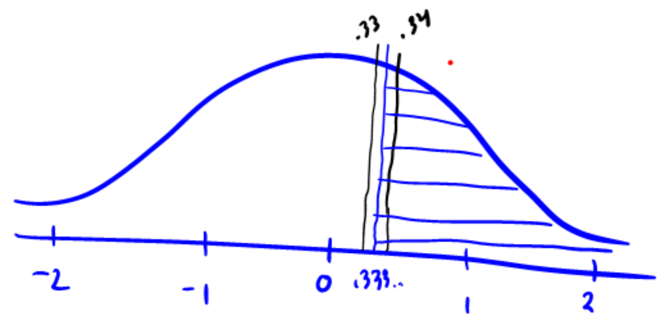


$P(X \leq 3), X \sim \text{Normal}(2, 64)$ $\quad 64 = 8^2$

standardize

$$P(X \leq 3) = P\left(\frac{X-2}{8} \leq \frac{3-2}{8}\right)$$
$$= P(Z \leq 0.125)$$
$$= \Phi(0.125)$$
$$\approx 0.5478$$

conservative = "probability is at least this value"



$P(X > 7), X \sim \text{Normal}(6, 9)$ $\quad 3^2$

$$P(X > 7) = P\left(\frac{X-6}{3} > \frac{7-6}{3}\right)$$
$$= P(Z > .3333)$$
$$\approx P(Z > .34)$$
$$= 1 - P(Z \leq .34)$$
$$= 1 - \Phi(.34)$$
$$= 1 - .6331 \doteq .3669$$

$P(Z > .3333)$ is at least $P(Z > .34)$

"probability is at least this"

$P(|X - 1| > 0.5), X \sim \text{Normal}(2, 4)$

$$= P(X-1 > 0.5 \text{ or } X-1 < -.5)$$
$$= P(X-1 > 0.5) + P(X-1 < -.5)$$
$$= P(X > 1.5) + P(X < 0.5)$$
$$= P\left(\frac{X-2}{2} > \frac{1.5-2}{2}\right) + P\left(\frac{X-2}{2} < \frac{0.5-2}{2}\right)$$

to be continued....

We can find standard normal quantiles by using the standard normal table in reverse.

Example 5.30 (Baby food, cont'd)**.** For the jar weights $X \sim \text{Normal}(137.2, 1.62^2)$, find $Q(0.1)$.

Table B.3

Standard Normal Cumulative Probabilities

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

$P(Z \leq -3.40)$ $= P(Z \leq -3.49)$

①

②

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985

Example 5.31 (Normal quantiles)**.** Find:

$Q(0.95)$ of $X \sim \text{Normal}(9, 3)$.

c such that $P(|X - 2| > c) = 0.01$, $X \sim \text{Normal}(2, 4)$

Table B.3

Standard Normal Cumulative Probabilities

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table B.3

Standard Normal Cumulative Probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9983	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

This table was generated using MINITAB.