This quantity is computed with the R statement

```
pnorm(4.5 / sqrt(140 / 3)) - pnorm(-16.5 / sqrt(140 / 3))
```

The use of the continuity correction results in two digits (and almost three digits) of accuracy—a significant improvement over the naive application of the central limit theorem.

Statisticians are often interested in determining the appropriate sample size n in the presence of limited information about a population distribution. The central limit theorem can be used to determine how many observations should be collected.

Example 8.15 Cassie is interested in determining the average gas mileage for a fleet of a particular model of car. Previous data collection of gas mileage gives her confidence that the standard deviation of the gas mileages that she collects will have a standard deviation of 3.3 miles per gallon. How many cars should she sample in order to be at least 95% certain that the difference between the population mean and sample mean is less than 1 mile per gallon?

Cassie's goal is to find the smallest sample size n such that

$$P\big(|\bar{X}-\mu|<1\big) \geq 0.95$$

where $X_1, X_2, \ldots, X_n$ are the values collected from a population with a population mean miles per gallon μ and population standard deviation $\sigma = 3.3$. Using the central limit theorem, the probability can be approximated by

$$\begin{aligned} P\big(|\bar{X}-\mu|<1\big) &= P\big(-1<\bar{X}-\mu<1\big) \\ &= P\left(-\frac{1}{3.3/\sqrt{n}} < \frac{\bar{X}-\mu}{3.3/\sqrt{n}} < \frac{1}{3.3/\sqrt{n}}\right) \\ &\cong P\left(-\frac{\sqrt{n}}{3.3} < Z < \frac{\sqrt{n}}{3.3}\right). \end{aligned}$$

Since this probability must be at least 0.95, and the 97.5th percentile of the standard normal distribution is approximately 1.96, the appropriate sample size n satisfies

$$\frac{\sqrt{n}}{3.3} = 1.96.$$

Solving for n yields $n = 41.8$. In order to achieve the prescribed precision, Cassie should collect the gas mileage on a sample of at least $n = 42$ cars.

The previous four examples have considered sample sizes of $n = 12$, $n = 100$, $n = 16$, and $n = 42$. In all cases, the central limit theorem has been invoked, and it is assumed that the accuracy is improved for larger n because the central limit theorem is an asymptotic result. Is there a general way to determine what n value is appropriate to invoke the central limit theorem? Although some elementary statistics textbooks state that a sample of at least $n = 30$ is necessary, the answer to the question is a bit more nuanced.

Figure 8.10 contains a 3×3 display of graphs of probability density functions of the *sample means* associated with the observations $X_1, X_2, \ldots, X_n$ drawn from three different continuous populations. Axis labels have been suppressed to avoid cluttering up the graphs; the horizontal axis is x

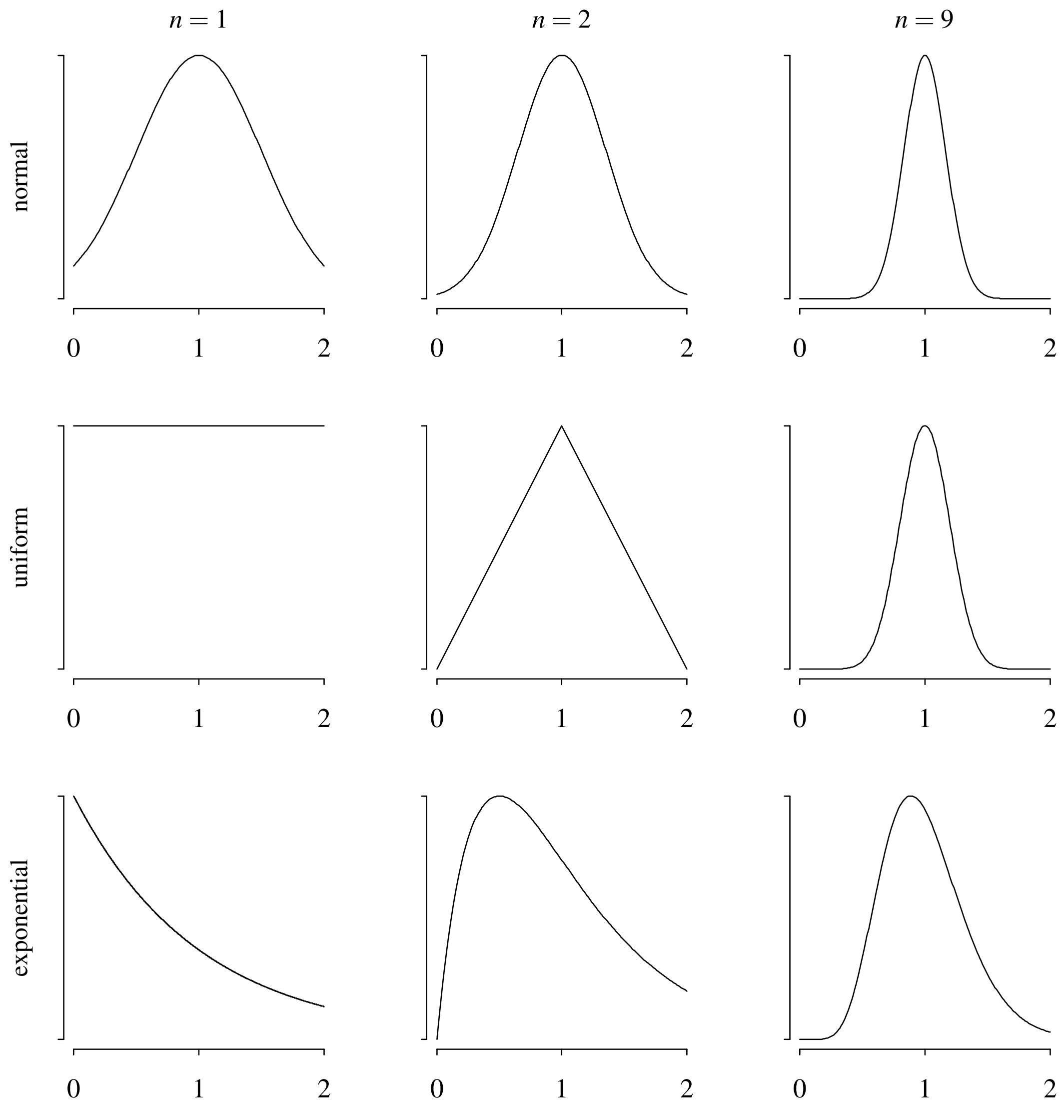

Figure 8.10: Central limit theorem illustrations.

and the vertical axis is $f(x)$ for all graphs. The horizontal axes all range from 0 to 2, and the vertical axes range from 0 to the highest value that the probability density function achieves on $[0, 2]$. The three rows correspond to the probability distributions from which the observations were drawn: normal, uniform, and exponential. More specifically, the sampling distributions are $N(1, 1/4)$, $U(0, 2)$, and exponential(1). The parameters for these distributions were chosen so that each has population mean 1. The population variances for the three population distributions are $1/4$, $1/3$, and 1, respectively. The three columns correspond to sample sizes: $n = 1$, $n = 2$, and $n = 9$. The first column of probability density functions corresponds to the population distribution.