7.43. In this problem, we envision the sample $Y_1, Y_2, ..., Y_{100}$, where

$$Y_i = \text{height of } i\text{th man (measured in inches)}, \quad i = 1, 2, ..., 100.$$

The population distribution is unknown (at least, it is not provided in the problem), but the population standard deviation is assumed to be $\sigma = 2.5$ inches. We regard $Y_1, Y_2, ..., Y_{100}$ as an iid sample from this unknown population distribution. We want to find

$$P(-0.5 < \overline{Y} - \mu < 0.5).$$

Note that

$$\overline{Y} = \frac{1}{100}\sum_{i=1}^{100} Y_i$$

is the sample mean height of the 100 men and μ is the population mean height. The difference between them is $\overline{Y} - \mu$. Even though the population distribution is unknown, the Central Limit Theorem says the (approximate) sampling distribution of $\overline{Y}$ is

$$\overline{Y} \sim \mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right) \implies \overline{Y} \sim \mathcal{AN}\left(\mu, \frac{2.5^2}{100}\right).$$

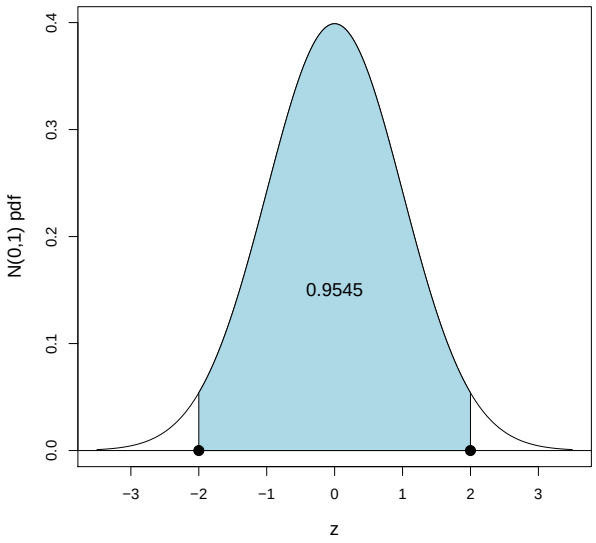
Therefore,

$$P(-0.5 < \overline{Y} - \mu < 0.5) = P\left(-\frac{0.5}{2.5/\sqrt{100}} < \frac{\overline{Y} - \mu}{2.5/\sqrt{100}} < \frac{0.5}{2.5/\sqrt{100}}\right) \approx P(-2 < Z < 2),$$

where $Z \sim \mathcal{N}(0, 1)$. This (approximate) probability is easy to calculate in R:

```
> pnorm(2,0,1)-pnorm(-2,0,1)  #P(-2 < Z < 2)
[1] 0.9544997
```

See the $\mathcal{N}(0, 1)$ pdf shown below:



7.52. In this problem, we envision the sample $Y_1, Y_2, ..., Y_{25}$, where

$$Y_i = \text{resistance of } i\text{th resistor (measured in ohms)}, \quad i = 1, 2, ..., 25.$$

The population distribution is unknown (at least, it is not provided in the problem), but the population mean is assumed to be $\mu = 200$ ohms and the population standard deviation is assumed to be $\sigma = 10$ ohms. We regard $Y_1, Y_2, ..., Y_{25}$ as an iid sample from this unknown population distribution.

(a) In this part, we want to find

$$P(199 < \overline{Y} < 202),$$

where

$$\overline{Y} = \frac{1}{25}\sum_{i=1}^{25} Y_i$$

is the sample mean. Even though the population distribution is unknown, the Central Limit Theorem says the (approximate) sampling distribution of $\overline{Y}$ is

$$\overline{Y} \sim \mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right) \implies \overline{Y} \sim \mathcal{AN}\left(\mu, \frac{10^2}{25}\right).$$

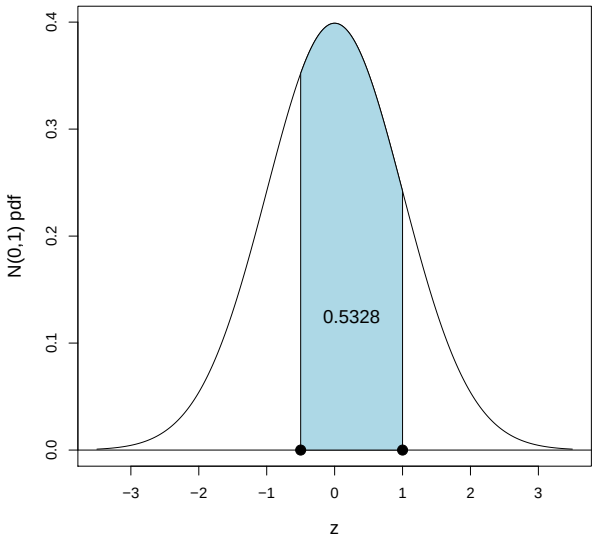
Therefore,

$$P(199 < \overline{Y} < 202) = P\left(\frac{199-200}{10/\sqrt{25}} < \frac{\overline{Y}-200}{10/\sqrt{25}} < \frac{202-200}{10/\sqrt{25}}\right) \approx P(-0.5 < Z < 1),$$

where $Z \sim \mathcal{N}(0,1)$. This (approximate) probability is easy to calculate in R:

```
> pnorm(1,0,1)-pnorm(-0.5,0,1)  #P(-0.5 < Z < 1)
[1] 0.5328072
```

See the $\mathcal{N}(0,1)$ pdf shown below:



(b) In this part, we want to find

$$P(T < 5100),$$

where

$$T = \sum_{i=1}^{25} Y_i$$

is the sample sum of the 25 resistors. Even though the population distribution is unknown, the Central Limit Theorem says the (approximate) sampling distribution of T is

$$T \sim \mathcal{AN}\left(n\mu, n\sigma^2\right) \implies T \sim \mathcal{AN}\left(25(200), 25(10)^2\right).$$

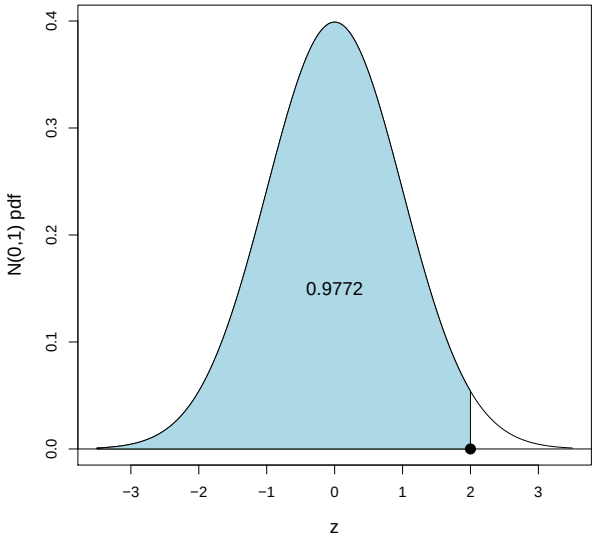
Therefore,

$$P(T < 2100) = P\left(\frac{T - 25(200)}{\sqrt{25(10)^2}} < \frac{5100 - 25(200)}{\sqrt{25(10)^2}}\right) \approx P(Z < 2),$$

where $Z \sim \mathcal{N}(0,1)$. This (approximate) probability is easy to calculate in R:

```
> pnorm(2,0,1)  #P(Z < 2)
[1] 0.9772499
```

See the $\mathcal{N}(0,1)$ pdf shown below:



7.53. In this problem, for part (b), we envision the sample $Y_1, Y_2, ..., Y_{100}$, where

$$Y_i = \text{concentration of } i\text{th air sample (measured in ppm)}, \quad i = 1, 2, ..., 100.$$

The population distribution is unknown (at least, it is not provided in the problem), but the population mean is assumed to be $\mu = 12$ ppm and the population standard deviation is assumed to be $\sigma = 9$ ppm. In part (b), we regard $Y_1, Y_2, ..., Y_{100}$ as an iid sample from this unknown population distribution.

(a) If the population mean is $\mu = 12$ and the population standard deviation is $\sigma = 9$, then the population distribution cannot be normal. Concentrations must be positive, and a concentration of 0 is only 1.5 standard deviations below the mean. The $\mathcal{N}(12, 9^2)$ distribution would allow for a substantial portion of the measurements (about 7%) to be negative, which does not make sense.

(b) In this part, we want to find

$$P(\overline{Y} > 14),$$

where

$$\overline{Y} = \frac{1}{100}\sum_{i=1}^{100} Y_i$$

is the sample mean of the 100 air sample concentrations. Even though the population distribution is unknown, the Central Limit Theorem says the (approximate) sampling distribution of $\overline{Y}$ is

$$\overline{Y} \sim \mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right) \implies \overline{Y} \sim \mathcal{AN}\left(12, \frac{9^2}{100}\right).$$

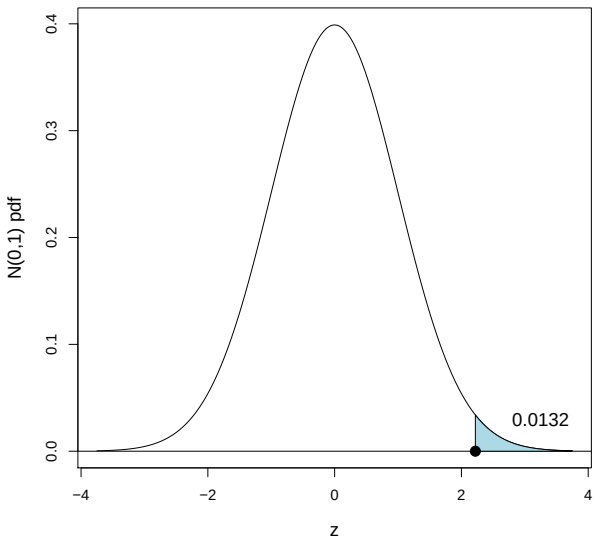
Therefore,

$$P(\overline{Y} > 14) = P\left(\frac{\overline{Y} - 12}{9/\sqrt{100}} > \frac{14 - 12}{9/\sqrt{100}}\right) \approx P(Z > 2.22),$$

where $Z \sim \mathcal{N}(0, 1)$. This (approximate) probability is easy to calculate in R:

```
> 1-pnorm(2.22,0,1)  #P(Z > 2.22)
[1] 0.01320938
```

See the $\mathcal{N}(0, 1)$ pdf shown below:



7.58. We have independent random samples:

- $X_1, X_2, ..., X_n$ is an iid sample from a population with mean μ_1 and variance σ_1^2
- $Y_1, Y_2, ..., Y_n$ is an iid sample from a population with mean μ_2 and variance σ_2^2.

Consider the new random variables

$$W_i = X_i - Y_i, \quad i = 1, 2, ..., n,$$

that is, W_i is the difference of X_i and Y_i, for $i = 1, 2, ..., n$. Note that

$$E(W_i) = E(X_i - Y_i) = E(X_i) - E(Y_i) = \mu_1 - \mu_2$$

and

$$V(W_i) = V(X_i - Y_i) = V(X_i) + V(Y_i) - 2\underbrace{\text{Cov}(X_i, Y_i)}_{=\,0} = \sigma_1^2 + \sigma_2^2.$$

Note that $\text{Cov}(X_i, Y_i) = 0$ because the samples are assumed to be independent. Therefore, $W_1, W_2, ..., W_n$ are iid random variables with mean $\mu = \mu_1 - \mu_2$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Provided $\sigma^2 < \infty$, applying the CLT directly yields

$$U_n = \frac{\overline{W} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $n \to \infty$. However, note that U_n algebraically equals

$$\begin{aligned}
\frac{\overline{W} - \mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^n W_i - \mu}{\sigma/\sqrt{n}} &= \frac{\frac{1}{n}\sum_{i=1}^n (X_i - Y_i) - \mu}{\sigma/\sqrt{n}} \\
&= \frac{(\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n Y_i) - \mu}{\sigma/\sqrt{n}} \\
&= \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}/\sqrt{n}} = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}.
\end{aligned}$$

Therefore,

$$U_n = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $n \to \infty$, as claimed.

7.75. In this problem, we envision the sample $Y_1, Y_2, ..., Y_{64}$, where

$$Y_i = \begin{cases} 1, & i\text{th voter favors bond issue} \\ 0, & \text{otherwise.} \end{cases}$$

We regard $Y_1, Y_2, ..., Y_{64}$ as an iid sample from a Bernoulli(p) population distribution, where the pollster believes $p = 0.20$. In this problem, we are supposed to assume $p = 0.20$, consistent with the pollster's belief. Define the sample proportion

$$\widehat{p} = \frac{1}{64}\sum_{i=1}^{64} Y_i,$$

that is, the proportion of voters in the sample who favor the bond issue. We are being asked to calculate

$$P(-0.06 < \widehat{p} - 0.20 < 0.06).$$

From the CLT (applied to sample proportions), we have

$$\widehat{p} \sim \mathcal{AN}\left(p, \frac{p(1-p)}{n}\right) \implies \widehat{p} \sim \mathcal{AN}\left(0.20, \frac{0.20(1-0.20)}{64}\right).$$

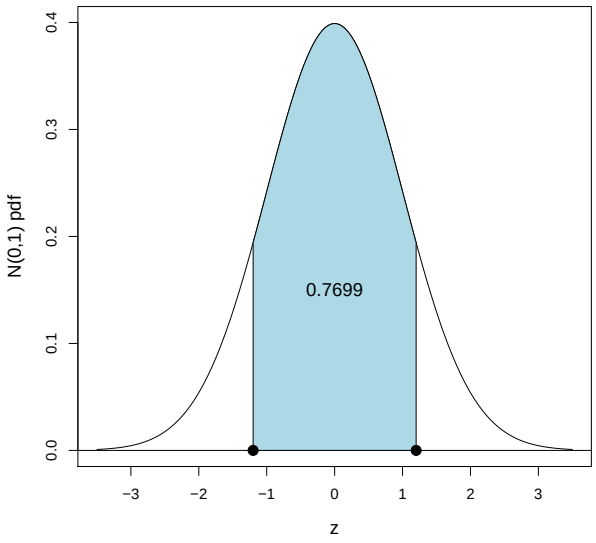
Therefore,

$$\begin{aligned}
P(-0.06 < \widehat{p} - 0.20 < 0.06) &= P\left(\frac{-0.06}{\sqrt{\frac{0.20(1-0.20)}{64}}} < \frac{\widehat{p} - 0.20}{\sqrt{\frac{0.20(1-0.20)}{64}}} < \frac{0.06}{\sqrt{\frac{0.20(1-0.20)}{64}}}\right) \\
&\approx P(-1.2 < Z < 1.2),
\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$. This (approximate) probability is easy to calculate in R:

```
> pnorm(1.2,0,1)-pnorm(-1.2,0,1)  #P(-1.2 < Z < 1.2)
[1] 0.7698607
```

See the $\mathcal{N}(0,1)$ pdf shown below:



7.80. In this problem, we envision the sample $Y_1, Y_2, ..., Y_{100}$, where

$$Y_i = \begin{cases} 1, & i\text{th resident younger than median (31 years)} \\ 0, & \text{otherwise.} \end{cases}$$

We regard $Y_1, Y_2, ..., Y_{100}$ as an iid sample from a Bernoulli($p = 0.5$) population distribution; i.e., if the median of the population is $\phi_{0.5} = 31$, then, by definition, half of the population is younger than 31 years and half of the population is older than 31 years. Define the sample sum

$$T = \sum_{i=1}^{100} Y_i,$$

that is, the number of residents in the sample who are younger than 31 years. We are being asked to calculate

$$P(T \geq 60).$$

We can calculate this probability exactly and also approximately by using the CLT (the question only asks for the approximate answer).

Exact calculation: We know $T \sim b(n = 100, p = 0.5)$. Therefore, we can calculate $P(T \geq 60)$ exactly as follows:

$$P(T \geq 60) = \sum_{t=60}^{100} \binom{100}{t} (0.5)^t (0.5)^{100-t} \approx 0.0284.$$

This calculation is carried out in R as follows:

```
> 1-pbinom(59,100,0.5)  #P(T >= 60)
[1] 0.02844397
```

CLT approximation: From the CLT, we know

$$T = \sum_{i=1}^{n} Y_i \sim \mathcal{AN}(np, np(1-p)) \implies T = \sum_{i=1}^{100} Y_i \sim \mathcal{AN}(50, 25)$$

because

$$\begin{aligned} np &= 100(0.5) = 50 \\ np(1-p) &= 100(0.5)(1-0.5) = 25. \end{aligned}$$

Therefore, we can calculate $P(T \geq 60)$ approximately as follows:

$$P(T \geq 60) = P\left(\frac{T-50}{\sqrt{25}} \geq \frac{60-50}{\sqrt{25}}\right) \approx P(Z \geq 2),$$

where $Z \sim \mathcal{N}(0,1)$. This (approximate) probability is easy to calculate in R:

```
> 1-pnorm(2,0,1)  #P(T >= 60) approximated by using P(Z >= 2)
[1] 0.02275013
```

7.87. In this problem, we envision the sample $Y_1, Y_2, ..., Y_{100}$, where

$$Y_i = \begin{cases} 1, & i\text{th customer waits longer than 10 minutes} \\ 0, & \text{otherwise.} \end{cases}$$

We regard $Y_1, Y_2, ..., Y_{100}$ as an iid sample from a Bernoulli(p) population distribution. What is p? We are given the waiting time (say W) for each customer follows an exponential distribution with mean $\beta = 10$ minutes. Therefore,

$$p = P(W > 10) = 1 - P(W \leq 10) = 1 - \underbrace{(1 - e^{-10/10})}_{\exp(10)\text{ cdf}} = e^{-1} \approx 0.3679.$$

Define the sample sum

$$T = \sum_{i=1}^{100} Y_i,$$

that is, the number of customers in the sample that wait longer than 10 minutes. We are being asked to calculate

$$P(T \geq 50).$$

We can calculate this probability exactly and also approximately by using the CLT (the question does not specify which one it wants).

Exact calculation: We know $T \sim b(n = 100, p = 0.3679)$. Therefore, we can calculate $P(T \geq 50)$ exactly as follows:

$$P(T \geq 50) = \sum_{t=50}^{100} \binom{100}{t} (0.3679)^t (1-0.3679)^{100-t} \approx 0.0047.$$

This calculation is carried out in R as follows:

```
> 1-pbinom(49,100,0.3679)  #P(T >= 50)
[1] 0.004713515
```

CLT approximation: From the CLT, we know

$$T = \sum_{i=1}^{n} Y_i \sim \mathcal{AN}(np, np(1-p)) \implies T = \sum_{i=1}^{100} Y_i \sim \mathcal{AN}(36.8, 23.25)$$

because

$$\begin{aligned} np &= 100(0.3679) = 36.8 \\ np(1-p) &= 100(0.3679)(1-0.3679) \approx 23.25. \end{aligned}$$

Therefore, we can calculate $P(T \geq 50)$ approximately as follows:

$$P(T \geq 50) \approx P\left(\frac{T - 36.8}{\sqrt{23.25}} \geq \frac{60 - 36.8}{\sqrt{23.25}}\right) \approx P(Z \geq 2.74),$$

where $Z \sim \mathcal{N}(0,1)$. This (approximate) probability is easy to calculate in R:

```
> 1-pnorm(2.74,0,1)  #P(T >= 50) approximated by using P(Z >= 2.74)
[1] 0.003071959
```

7.94. In this problem, we envision the sample $Y_1, Y_2, ..., Y_5$, where

$$Y_i = \text{repair cost for } i\text{th machine}, \quad i = 1, 2, ..., 5.$$

The population distribution is exponential with mean 20; i.e., $Y \sim$ exponential(20). We regard $Y_1, Y_2, ..., Y_5$ as an iid sample from an exponential(20) population distribution.

We want to find the constant c that satisfies

$$P\left(\sum_{i=1}^{5} Y_i > c\right) = 0.05.$$

What is the sampling distribution of the sample sum $T = \sum_{i=1}^{5} Y_i$? Recall the population mgf of $Y \sim$ exponential(20) is

$$m_Y(t) = \frac{1}{1-20t},$$

for $t < 1/20$. Therefore, the mgf of the sum is

$$m_T(t) = [m_Y(t)]^5 = \left(\frac{1}{1-20t}\right)^5,$$

for $t < 1/20$. We recognize this as the mgf of a gamma random variable with shape $\alpha = 5$ and scale $\beta = 20$. Because mgfs are unique, we know $T = \sum_{i=1}^{5} Y_i \sim$ gamma$(5, 20)$. Therefore, we want to find

$$c = \text{95th percentile (0.95 quantile) of a gamma(5,20) distribution.}$$

We can find this easily in R:

```
> qgamma(0.95,5,1/20)
[1] 183.0704
```

7.96. In this problem, we envision the sample $Y_1, Y_2, ..., Y_{40}$, where

$$Y_i = \text{proportion of impurity for } i\text{th iron ore sample}, \quad i = 1, 2, ..., 40.$$

The population distribution is described by the pdf

$$f_Y(y) = \begin{cases} 3y^2, & 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is the pdf of $Y \sim \text{beta}(3, 1)$. We regard $Y_1, Y_2, ..., Y_{40}$ as an iid sample from a beta$(3, 1)$ distribution. We want to find

$$P(\overline{Y} > 0.7),$$

where

$$\overline{Y} = \frac{1}{40}\sum_{i=1}^{40} Y_i$$

is the sample mean of the 40 impurity measurements. We will approximate this probability by using the CLT. We need to know the population mean μ and the population variance σ^2. Using what we know about the beta distribution (CH 4), we have

$$\begin{aligned} \mu &= \frac{3}{3+1} = 0.75 \\ \sigma^2 &= \frac{3(1)}{(3+1)^2(3+1+1)} = \frac{3}{80} = 0.0375. \end{aligned}$$

The Central Limit Theorem says the (approximate) sampling distribution of $\overline{Y}$ is

$$\overline{Y} \sim \mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right) \implies \overline{Y} \sim \mathcal{AN}\left(0.75, \frac{0.0375}{40}\right).$$

Therefore,

$$P(\overline{Y} > 0.7) = P\left(\frac{\overline{Y} - 0.75}{\sqrt{0.0375/40}} > \frac{0.7 - 0.75}{\sqrt{0.0375/40}}\right) \approx P(Z > -1.63),$$

where $Z \sim \mathcal{N}(0, 1)$. This (approximate) probability is easy to calculate in R:

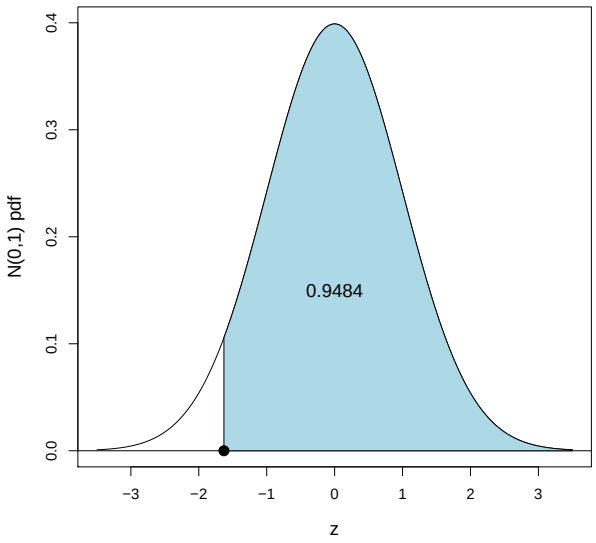
```
> 1-pnorm(-1.63,0,1)  #P(Z>-1.63)
[1] 0.9484493
```

See the $\mathcal{N}(0, 1)$ pdf shown at the top of the next page.

7.97. In this problem, we assume $X_1, X_2, ..., X_n$ are iid $\chi^2(1)$; i.e., the population distribution is $\chi^2(1)$. Recall the sampling distribution of the sample sum $Y = \sum_{i=1}^n X_i \sim \chi^2(n)$. To remember why this is true, recall that

$$m_Y(t) = [m_X(t)]^n = \left[\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}\right]^n = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}.$$

We recognize this as the mgf of a χ^2 random variable with n degrees of freedom. Because mgfs are unique, it must be true that $Y \sim \chi^2(n)$.



(a) When we discussed the CLT, we learned that sample sums are approximately normally distributed when the sample size n is large; i.e.,

$$Y = \sum_{i=1}^{n} X_i \sim \mathcal{AN}(n\mu, n\sigma^2).$$

Because the population distribution is $\chi^2(1)$, we know that

$$\begin{aligned} \mu &= 1 \\ \sigma^2 &= 2. \end{aligned}$$

Therefore, the CLT says that

$$Y = \sum_{i=1}^{n} X_i \sim \mathcal{AN}(n, 2n),$$

for large n. Standardizing, we have

$$Z_n = \frac{Y-n}{\sqrt{2n}} \sim \mathcal{AN}(0,1) \iff \frac{Y-n}{\sqrt{2n}} \xrightarrow{d} \mathcal{N}(0,1),$$

as $n \to \infty$.

(b) In this part, we envision the sample $Y_1, Y_2, ..., Y_{50}$, where

$$Y_i = \text{length of } i\text{th rod (measured in inches)}, \quad i = 1, 2, ..., 50.$$

The population distribution is $Y \sim \mathcal{N}(6, 0.2)$; i.e., the population mean length is $\mu = 6$ inches and the population variance is $\sigma^2 = 0.2$ (inches)^2. In this part, we regard $Y_1, Y_2, ..., Y_{50}$ as an iid sample from this population distribution.

The cost of handling/repairing the ith rod is given by

$$C_i = 4(Y_i - 6)^2, \quad i = 1, 2, ..., 50.$$

For the 50 rods, the associated costs $C_1, C_2, ..., C_{50}$ are iid with mean

$$\mu_C = E(C) = E[4(Y-6)^2]$$

and variance

$$\sigma_C^2 = V(C) = V[4(Y-6)^2].$$

We need to calculate μ_C and σ_C^2. Calculating $\mu_C = E(C)$ is easy; note that

$$\mu_C = E(C) = E[4(Y-6)^2] = 4E[(Y-6)^2] = 4V(Y) = 4(0.2) = 0.8.$$

Calculating $\sigma_C^2 = V(C)$ is harder. To make it easier, note that

$$Y \sim \mathcal{N}(6, 0.2) \implies \frac{Y-6}{\sqrt{0.2}} \sim \mathcal{N}(0,1) \implies \left(\frac{Y-6}{\sqrt{0.2}}\right)^2 = \frac{(Y-6)^2}{0.2} \sim \chi^2(1).$$

Therefore,

$$V\left(\frac{(Y-6)^2}{0.2}\right) = 2 \implies \frac{1}{(0.2)^2} V[(Y-6)^2] = 2 \implies V[(Y-6)^2] = 2(0.2)^2 = 0.08.$$

Therefore,

$$\sigma_C^2 = V(C) = V[4(Y-6)^2] = 16V[(Y-6)^2] = 16(0.08) = 1.28.$$

Summarizing, $C_1, C_2, ..., C_{50}$ are iid with mean $\mu_C = 0.8$ and variance $\sigma_C^2 = 1.28$, and we want to approximate the probability

$$P\left(\sum_{i=1}^{50} C_i > 48\right).$$

The Central Limit Theorem says the (approximate) sampling distribution of $\sum_{i=1}^{50} C_i$ is

$$\sum_{i=1}^{50} C_i \sim \mathcal{AN}\left(50\mu_C, 50\sigma_C^2\right) \implies \sum_{i=1}^{50} C_i \sim \mathcal{AN}\left(40, 64\right).$$

Therefore,

$$P\left(\sum_{i=1}^{50} C_i > 48\right) = P\left(\frac{\sum_{i=1}^{50} C_i - 40}{\sqrt{64}} > \frac{48-40}{\sqrt{64}}\right) \approx P(Z > 1),$$

where $Z \sim \mathcal{N}(0,1)$. This (approximate) probability is easy to calculate in R:

```
> 1-pnorm(1,0,1)  #P(Z > 1)
[1] 0.1586553
```