Probability and Statistics I

25. The Distribution of Sample Totals, Means, and Proportions

6.2 The Distribution of Sample Totals, Means, and Proportions

The Case of a Normal Population Distribution

Distribution of the Sample Mean for a Normal Population

If $X_1, \dots, X_n$ represent a random sample from a normal distribution with mean μ and variance σ^2, mathematically

$$X_i \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2),$$

then[1]

1. $T_n = \sum_{i=1}^n X_i \sim \text{Normal}(n\mu, n\sigma^2)$
2. $\bar{X}_n = \sum_{i=1}^n X_i / n \sim \text{Normal}(\mu, \sigma^2/n)$
3. $Z = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \text{Normal}(0, 1)$

[1] The textbook now uses T_o and $\bar{X}$ with no subscript to denote the total and the mean. I prefer to use T_n and $\bar{X}_n$ to emphasize the sample size.

Example 25.1

According to Burmaster and Murray (1998), the height of men between the ages of 50 and 80 is normally distributed with a mean of 174.20 cm and a variance of 42.36 cm^2.

Suppose that we collect a random sample of 25 men from the population. Let their heights be denoted by $X_1, \ldots, X_{25}$.

a) What is the sampling distribution of the total height, $T_{25} = \sum_{i=1}^{25} X_i$?
b) What is the sampling distribution of the average height, $\bar{X}_{25} = \sum_{i=1}^{25} X_i/25$?
c) What is the sampling distribution of $Z = \frac{(\bar{X}_{25} - 174.20)}{\sqrt{1.69}}$?

6.2 The Distribution of Sample Totals, Means, and Proportions

The Central Limit Theorem

Central Limit Theorem

Let $X_1, \ldots, X_N$ be a random sample from *any* distribution with mean μ and variance $\sigma^2 < \infty$. Then

$$\lim_{n \to \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = P(Z \leq z)$$

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

Central Limit Theorem

Let $X_1, \ldots, X_N$ be a random sample from *any* distribution with mean μ and variance $\sigma^2 < \infty$. Then

$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = P(Z \leq z)$$

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

We say

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \dot{\sim} \text{Normal}(0,1)$$

or

$$\bar{X}_n \dot{\sim} \text{Normal}(\mu, \sigma^2/n).$$

Example 25.2

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed Bernoulli random variables such that

$$P(X_i = 0) = 1 - p \text{ and } P(X_i = 1) = p$$

for all $i = 1, \ldots, n$.

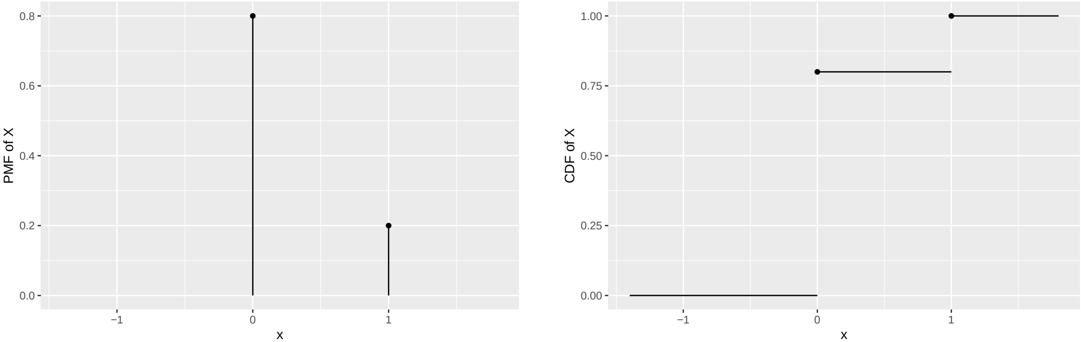
a) What is the pmf of $\bar{X}_n$?

b) What is the approximate cdf of

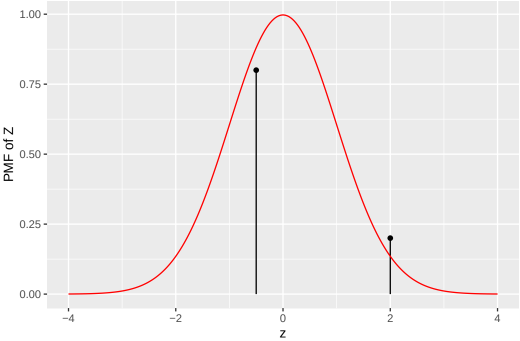
$$Z = \frac{\bar{X}_n - np}{\sqrt{p(1-p)/n}}$$

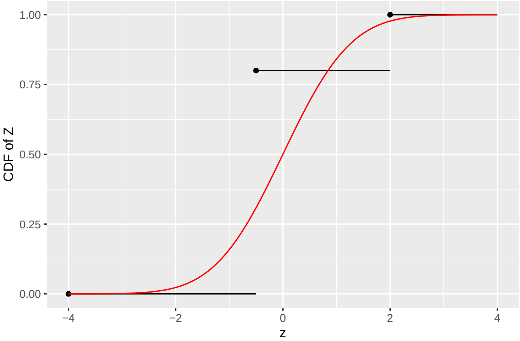
when n is large?

PMF & CDF of X



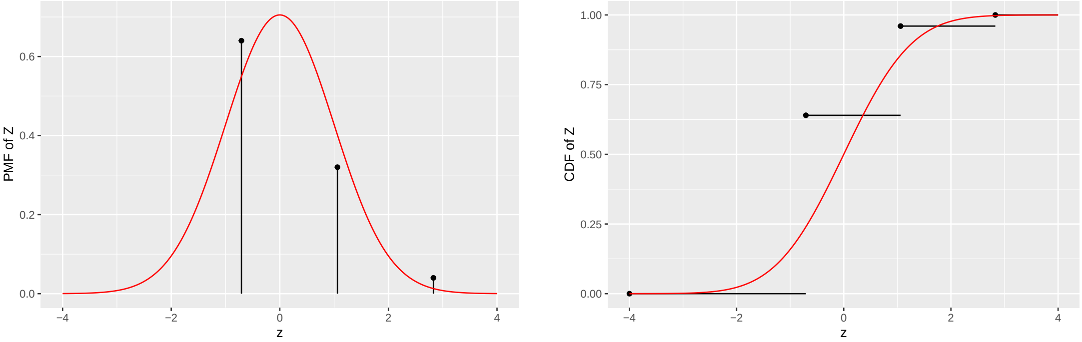
PMF & CDF of $Z = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 2$



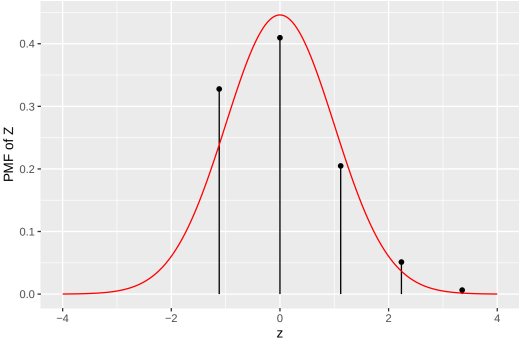


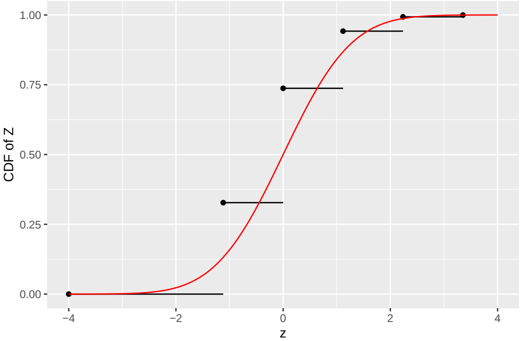
Central Limit Theorem

PMF & CDF of $Z = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 5$

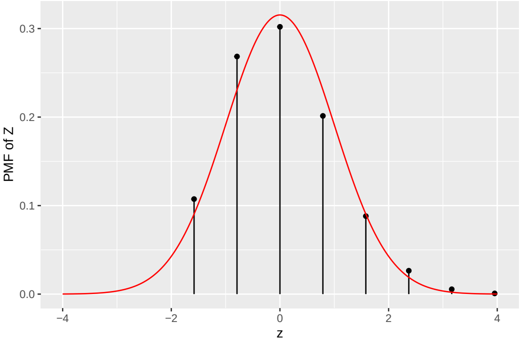


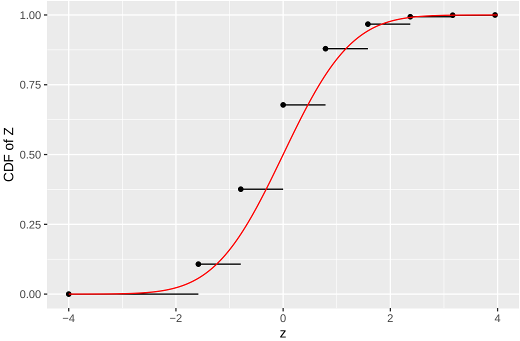
PMF & CDF of $Z = \frac{\sqrt{n}\left(\bar{X}_n - p\right)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 10$



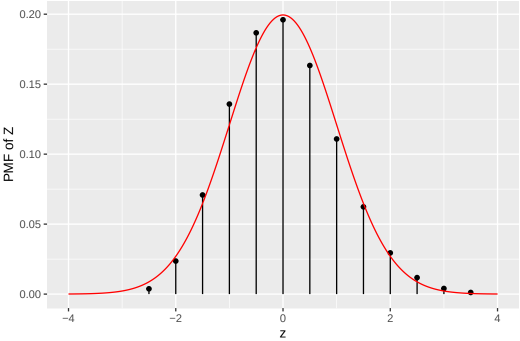


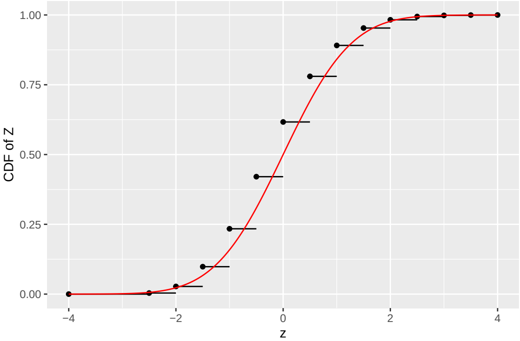
PMF & CDF of $Z = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 25$



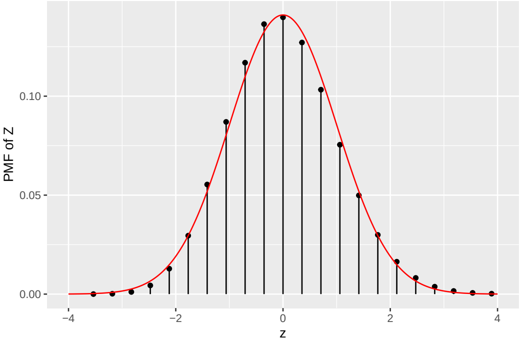


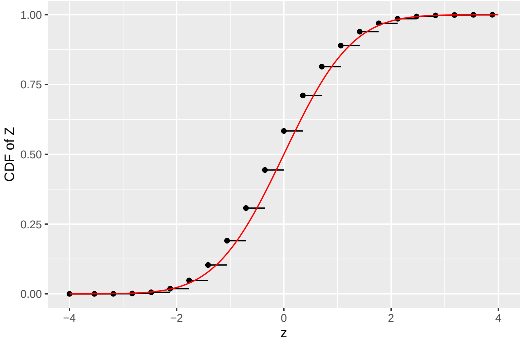
PMF & CDF of $Z = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 50$



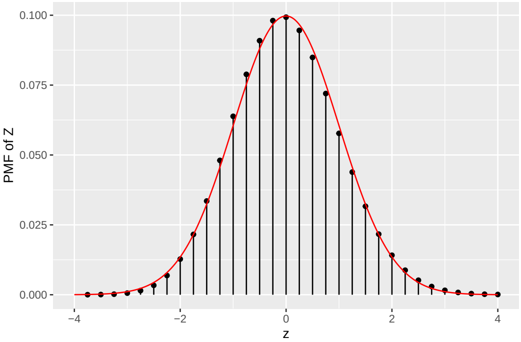


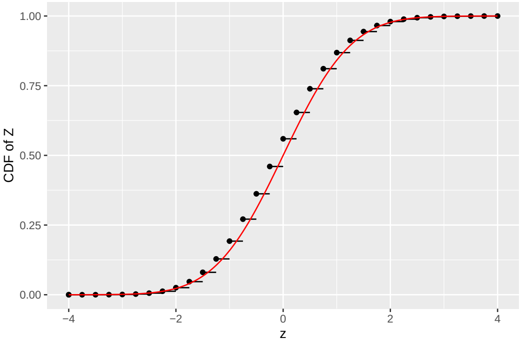
PMF & CDF of $Z = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 100$





PMF & CDF of $Z = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}$ when $p = 0.2$ and $n = 100$





Example 25.3

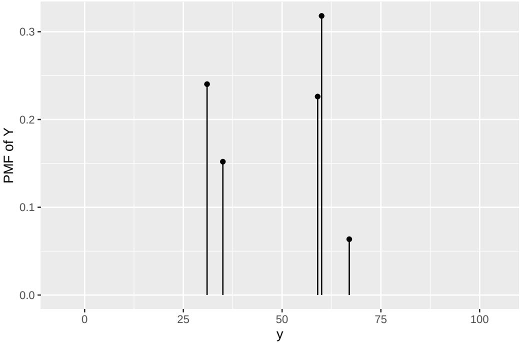
Consider the random variable Y with pmf

y	31	35	59	60	67
p	0.24	0.15	0.23	0.32	0.06

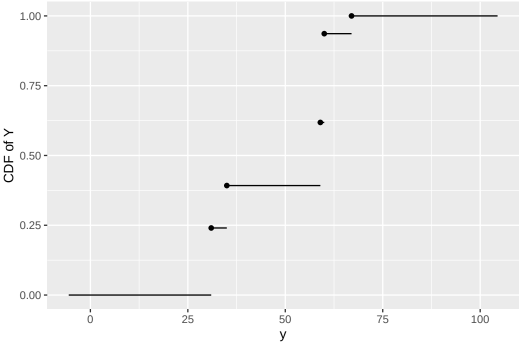
The mean and variance of Y are

$$\mu = E(Y) = 49.45 \text{ and } \sigma^2 = V(Y) = 189.13.$$

PMF & CDF of Y

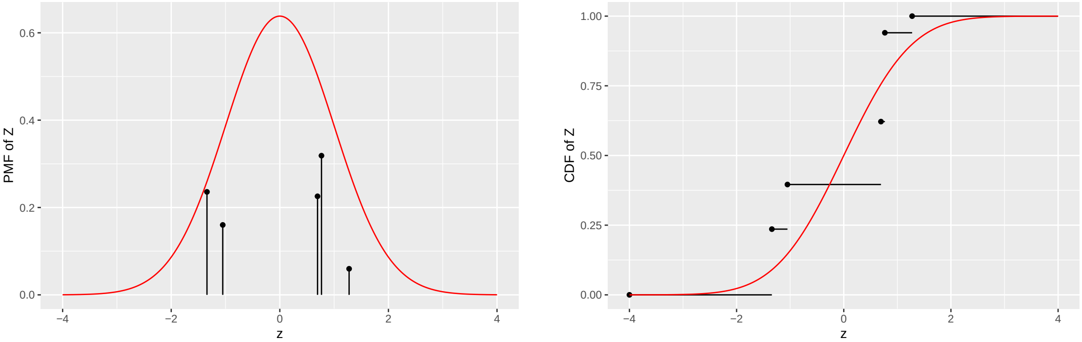


PMF of Y
0.3
0.2
0.1
0.0
0
25
50
75
100
y



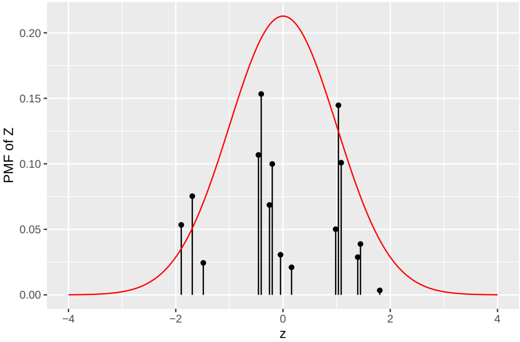
CDF of Y
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
y

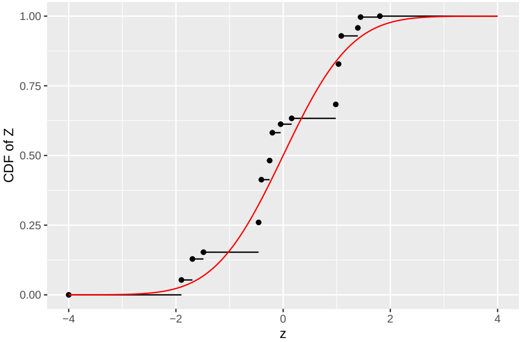
PMF & CDF of $Z = \frac{\sqrt{n}\left(\bar{Y}_n - \mu\right)}{\sigma}$ when $n = 1$



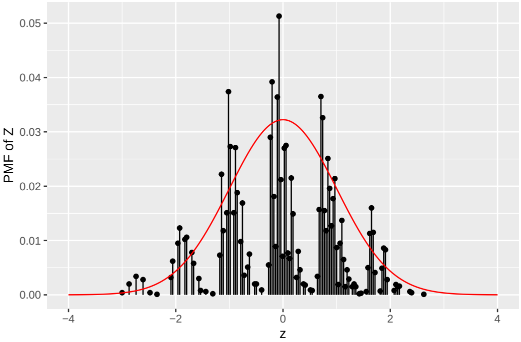
Central Limit Theorem

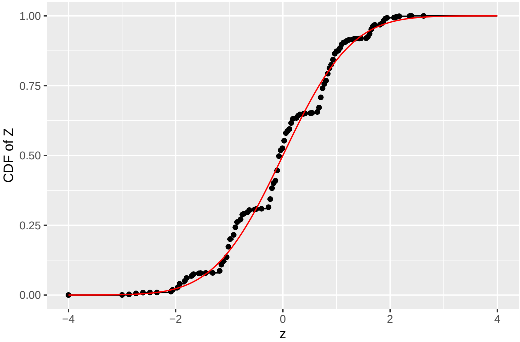
PMF & CDF of $Z = \frac{\sqrt{n}\left(\bar{Y}_n - \mu\right)}{\sigma}$ when $n = 2$





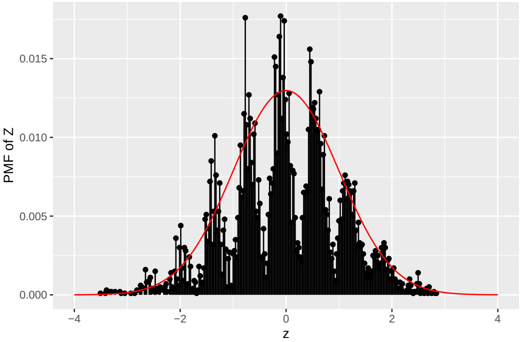
PMF & CDF of $Z = \frac{\sqrt{n}\left(\bar{Y}_n - \mu\right)}{\sigma}$ when $n = 5$

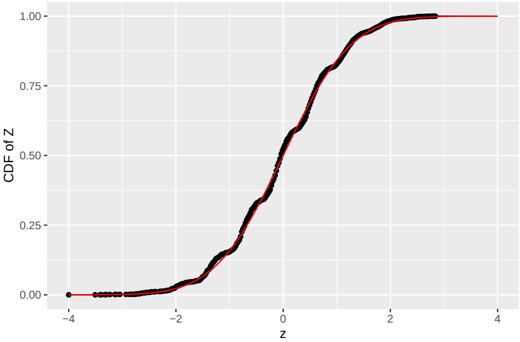




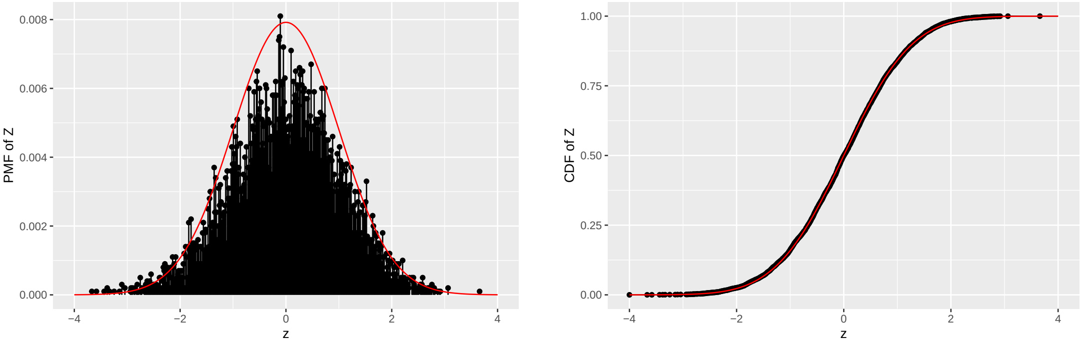
Central Limit Theorem

PMF & CDF of $Z = \frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}$ when $n = 10$

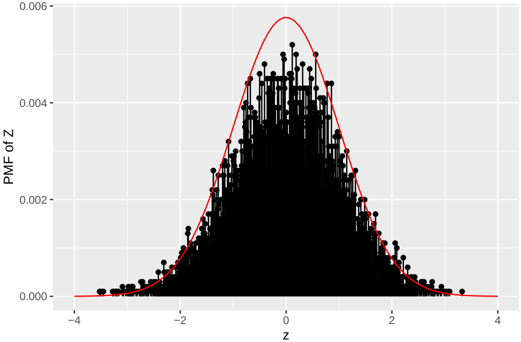


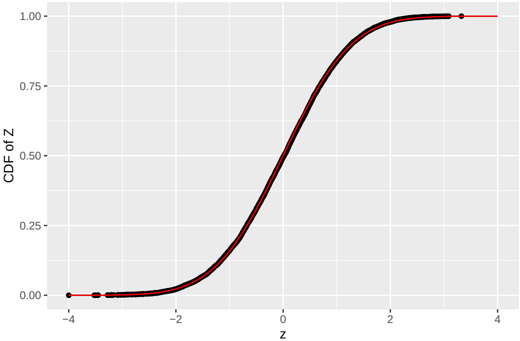


$$\text{PMF \& CDF of } Z = \frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma} \text{ when } n = 25$$



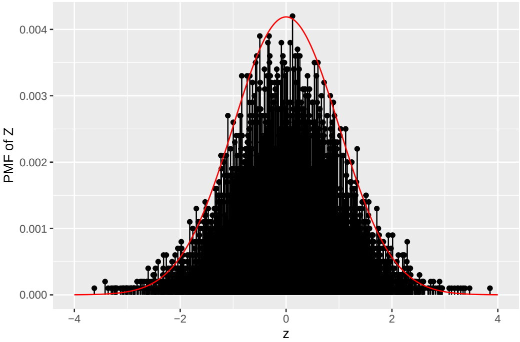
PMF & CDF of $Z = \frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}$ when $n = 50$

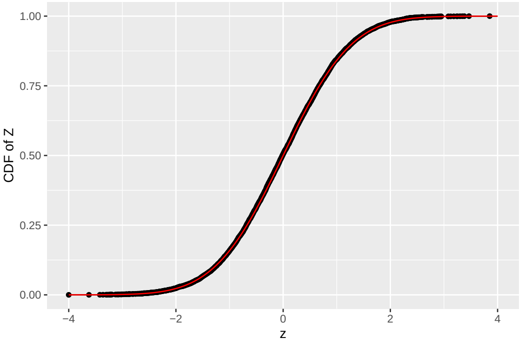




Central Limit Theorem

PMF & CDF of $Z = \frac{\sqrt{n}\left(\bar{Y}_n - \mu\right)}{\sigma}$ when $n = 100$





Example 25.4

The Acme string company produces spools of string advertised to have a length of 100 m. However, the length of string on a randomly selected ball actually has a mean of 101 m and a standard deviation of .2 m.

Approximate the 95-th percentile of the total amount of string in a box containing 50 spools.

Questions?

Exercise 25.1

According to Burmaster and Murray (1998), the log weight in kilograms of men between the ages of 50 and 80 is normally distributed with a mean of 4.41 and variance .46. It can be shown that the weight then follows a log-normal distribution with mean $\mu_W = 91.45$ kg and variance $\sigma_W^2 1970.83$ kg. The pdf is shown on the next slide. The vertical dashed line represents the mean.

Let $W_1, \ldots, W_n$ be a random sample of weights for n men.

a) Describe the shape of the density.

b) Approximate the distribution of $\bar{W}_n$. What conditions need to be satisfied?

c) Explain what the approximation in the previous part means.

d) (Extra) Use the approximation to show that $\lim_{n\to\infty} P(\mu_W - \epsilon < \bar{W}_n < \mu_W + \epsilon) = 1$ for any $\epsilon > 0$. Explain what this means.