Chapter 14

SAMPLING DISTRIBUTIONS ASSOCIATED WITH THE NORMAL POPULATIONS

Given a random sample $X_1, X_2, ..., X_n$ from a population X with probability distribution $f(x; \theta)$, where θ is a parameter, a *statistic* is a function T of $X_1, X_2, ..., X_n$, that is

$$T = T\left(X_1, X_2, ..., X_n\right)$$

which is free of the parameter θ. If the distribution of the population is known, then sometimes it is possible to find the probability distribution of the statistic T. The probability distribution of the statistic T is called the sampling distribution of T. The joint distribution of the random variables $X_1, X_2, ..., X_n$ is called the distribution of the sample. The distribution of the sample is the joint density

$$f(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

since the random variables $X_1, X_2, ..., X_n$ are independent and identically distributed.

Since the normal population is very important in statistics, the sampling distributions associated with the normal population are very important. The most important sampling distributions which are associated with the normal population are the followings: the chi-square distribution, the student's t-distribution, the F-distribution, and the beta distribution. In this chapter, we only consider the first three distributions, since the last distribution was considered earlier.

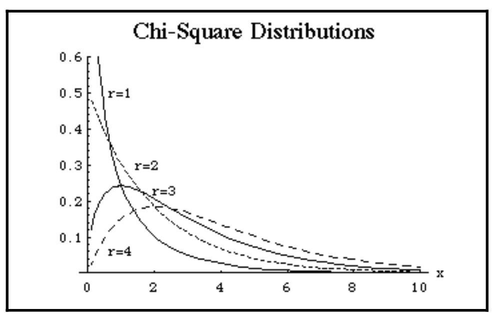
14.1. Chi-square distribution

In this section, we treat the Chi-square distribution, which is one of the very useful sampling distributions.

Definition 14.1. A continuous random variable X is said to have a chi-square distribution with r degrees of freedom if its probability density function is of the form

$$f(x;\, r) = \begin{cases} \frac{1}{\Gamma(\frac{r}{2})\, 2^{\frac{r}{2}}}\, x^{\frac{r}{2}-1}\, e^{-\frac{x}{2}} & \text{if } 0 \le x < \infty \\ 0 & \text{otherwise,} \end{cases}$$

where $r > 0$. If X has chi-square distribution, then we denote it by writing $X \sim \chi^2(r)$. Recall that a gamma distribution reduces to chi-square distribution if $\alpha = \frac{r}{2}$ and $\theta = 2$. The mean and variance of X are r and $2r$, respectively.



Thus, chi-square distribution is also a special case of gamma distribution. Further, if $r \to \infty$, then chi-square distribution tends to normal distribution.

Example 14.1. If $X \sim GAM(1, 1)$, then what is the probability density function of the random variable $2X$?

Answer: We will use the moment generating method to find the distribution of $2X$. The moment generating function of a gamma random variable is given by

$$M(t) = (1 - \theta\, t)^{-\alpha}, \qquad \text{if} \qquad t < \frac{1}{\theta}.$$

Since $X \sim GAM(1,1)$, the moment generating function of X is given by

$$M_X(t) = \frac{1}{1-t}, \qquad t < 1.$$

Hence, the moment generating function of $2X$ is

$$\begin{aligned} M_{2X}(t) &= M_X(2t) \\ &= \frac{1}{1-2t} \\ &= \frac{1}{(1-2t)^{\frac{2}{2}}} \\ &= \text{MGF of } \chi^2(2). \end{aligned}$$

Hence, if X is $GAM(1,1)$ or is an exponential with parameter 1, then $2X$ is chi-square with 2 degrees of freedom.

Example 14.2. If $X \sim \chi^2(5)$, then what is the probability that X is between 1.145 and 12.83?

Answer: The probability of X between 1.145 and 12.83 can be calculated from the following:

$$\begin{aligned} &P(1.145 \leq X \leq 12.83) \\ &\quad = P(X \leq 12.83) - P(X \leq 1.145) \\ &\quad = \int_0^{12.83} f(x)\,dx - \int_0^{1.145} f(x)\,dx \\ &\quad = \int_0^{12.83} \frac{1}{\Gamma\left(\frac{5}{2}\right)\, 2^{\frac{5}{2}}}\, x^{\frac{5}{2}-1}\, e^{-\frac{x}{2}}\,dx - \int_0^{1.145} \frac{1}{\Gamma\left(\frac{5}{2}\right)\, 2^{\frac{5}{2}}}\, x^{\frac{5}{2}-1}\, e^{-\frac{x}{2}}\,dx \\ &\quad = 0.975 - 0.050 \qquad (\text{from } \chi^2 \text{ table}) \\ &\quad = 0.925. \end{aligned}$$

The above integrals are hard to evaluate and thus their values are taken from the chi-square table.

Example 14.3. If $X \sim \chi^2(7)$, then what are values of the constants a and b such that $P(a < X < b) = 0.95$?

Answer: Since

$$0.95 = P(a < X < b) = P(X < b) - P(X < a),$$

we get

$$P(X < b) = 0.95 + P(X < a).$$

We choose $a = 1.690$, so that

$$P(X < 1.690) = 0.025.$$

From this, we get

$$P(X < b) = 0.95 + 0.025 = 0.975$$

Thus, from chi-square table, we get $b = 16.01$.

The following theorems were studied earlier in Chapters 6 and 13 and they are very useful in finding the sampling distributions of many statistics. We state these theorems here for the convenience of the reader.

Theorem 14.1. If $X \sim N(\mu, \sigma^2)$, then $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$.

Theorem 14.2. If $X \sim N(\mu, \sigma^2)$ and $X_1, X_2, ..., X_n$ is a random sample from the population X, then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n).$$

Theorem 14.3. If $X \sim N(\mu, \sigma^2)$ and $X_1, X_2, ..., X_n$ is a random sample from the population X, then

$$\frac{(n-1)\, S^2}{\sigma^2} \sim \chi^2(n-1).$$

Theorem 14.4. If $X \sim GAM(\theta, \alpha)$, then

$$\frac{2}{\theta}\, X \sim \chi^2(2\alpha).$$

Example 14.4. A new component is placed in service and n spares are available. If the times to failure in days are independent exponential variables, that is $X_i \sim EXP(100)$, how many spares would be needed to be 95% sure of successful operation for at least two years ?

Answer: Since $X_i \sim EXP(100)$,

$$\sum_{i=1}^{n} X_i \sim GAM(100, n).$$

Hence, by Theorem 14.4, the random variable

$$Y = \frac{2}{100} \sum_{i=1}^{n} X_i \sim \chi^2(2n).$$

We have to find the number of spares n such that

$$\begin{aligned}
0.95 &= P\left(\sum_{i=1}^{n} X_i \geq 2 \text{ years} \right) \\
&= P\left(\sum_{i=1}^{n} X_i \geq 730 \text{ days} \right) \\
&= P\left(\frac{2}{100} \sum_{i=1}^{n} X_i \geq \frac{2}{100} 730 \text{ days} \right) \\
&= P\left(\frac{2}{100} \sum_{i=1}^{n} X_i \geq \frac{730}{50} \right) \\
&= P\left(\chi^2(2n) \geq 14.6 \right). \\
2n &= 25 \qquad\qquad (\text{from } \chi^2 \text{ table})
\end{aligned}$$

Hence $n = 13$ (after rounding up to the next integer). Thus, 13 spares are needed to be 95% sure of successful operation for at least two years.

Example 14.5. If $X \sim N(10, 25)$ and $X_1, X_2, ..., X_{501}$ is a random sample of size 501 from the population X, then what is the expected value of the sample variance S^2 ?

Answer: We will use the Theorem 14.3, to do this problem. By Theorem 14.3, we see that

$$\frac{(501-1)\, S^2}{\sigma^2} \sim \chi^2(500).$$

Hence, the expected value of S^2 is given by

$$\begin{aligned}
E\left[S^2\right] &= E\left[\left(\frac{25}{500}\right) \left(\frac{500}{25}\right) S^2 \right] \\
&= \left(\frac{25}{500}\right) E\left[\left(\frac{500}{25}\right) S^2 \right] \\
&= \left(\frac{1}{20}\right) E\left[\chi^2(500)\right] \\
&= \left(\frac{1}{20}\right) 500 \\
&= 25.
\end{aligned}$$

14.2. Student's t-distribution

Here we treat the Student's t-distribution, which is also one of the very useful sampling distributions.

Definition 14.2. A continuous random variable X is said to have a t-distribution with ν degrees of freedom if its probability density function is of the form

$$f(x;\,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\,\nu}\;\Gamma\left(\frac{\nu}{2}\right)\,\left(1+\frac{x^2}{\nu}\right)^{\left(\frac{\nu+1}{2}\right)}}, \qquad -\infty < x < \infty$$

where $\nu > 0$. If X has a t-distribution with ν degrees of freedom, then we denote it by writing $X \sim t(\nu)$.

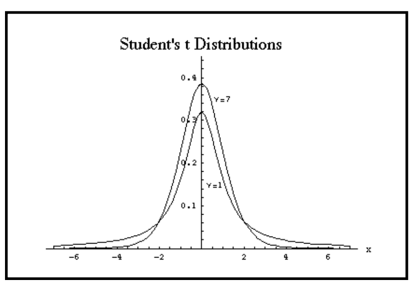
The t-distribution was discovered by W.S. Gosset (1876-1936) of England who published his work under the pseudonym of student. Therefore, this distribution is known as Student's t-distribution. This distribution is a generalization of the Cauchy distribution and the normal distribution. That is, if $\nu = 1$, then the probability density function of X becomes

$$f(x;\,1) = \frac{1}{\pi\,(1+x^2)} \qquad -\infty < x < \infty,$$

which is the Cauchy distribution. Further, if $\nu \to \infty$, then

$$\lim_{\nu\to\infty} f(x;\,\nu) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty,$$

which is the probability density function of the standard normal distribution. The following figure shows the graph of t-distributions with various degrees of freedom.



Example 14.6. If $T \sim t(10)$, then what is the probability that T is at least 2.228 ?

Answer: The probability that T is at least 2.228 is given by

$$\begin{aligned} P(T \geq 2.228) &= 1 - P(T < 2.228) \\ &= 1 - 0.975 \qquad \text{(from t − table)} \\ &= 0.025. \end{aligned}$$

Example 14.7. If $T \sim t(19)$, then what is the value of the constant c such that $P\left(|T| \leq c\right) = 0.95$?

Answer:

$$\begin{aligned} 0.95 &= P\left(|T| \leq c\right) \\ &= P\left(-c \leq T \leq c\right) \\ &= P\left(T \leq c\right) - 1 + P\left(T \leq c\right) \\ &= 2\,P\left(T \leq c\right) - 1. \end{aligned}$$

Hence

$$P\left(T \leq c\right) = 0.975.$$

Thus, using the t-table, we get for 19 degrees of freedom

$$c = 2.093.$$

Theorem 14.5. If the random variable X has a t-distribution with ν degrees of freedom, then

$$E[X] = \begin{cases} 0 & \text{if} \quad \nu \geq 2 \\ DNE & \text{if} \quad \nu = 1 \end{cases}$$

and

$$Var[X] = \begin{cases} \frac{\nu}{\nu-2} & \text{if} \quad \nu \geq 3 \\ DNE & \text{if} \quad \nu = 1, 2 \end{cases}$$

where DNE means does not exist.

Theorem 14.6. If $Z \sim N(0,1)$ and $U \sim \chi^2(\nu)$ and in addition, Z and U are independent, then the random variable W defined by

$$W = \frac{Z}{\sqrt{\frac{U}{\nu}}}$$

has a t-distribution with ν degrees of freedom.

Theorem 14.7. If $X \sim N(\mu, \sigma^2)$ and $X_1, X_2, ..., X_n$ be a random sample from the population X, then

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1).$$

Proof: Since each $X_i \sim N(\mu, \sigma^2)$,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Thus,

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

Further, from Theorem 14.3 we know that

$$(n-1)\,\frac{S^2}{\sigma^2} \sim \chi^2(n-1).$$

Hence

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)\,S^2}{(n-1)\,\sigma^2}}} \sim t(n-1) \qquad \text{(by Theorem 14.6).}$$

This completes the proof of the theorem.

Example 14.8. Let X_1, X_2, X_3, X_4 be a random sample of size 4 from a standard normal distribution. If the statistic W is given by

$$W = \frac{X_1 - X_2 + X_3}{\sqrt{X_1^2 + X_2^2 + X_3^2 + X_4^2}},$$

then what is the expected value of W ?

Answer: Since $X_i \sim N(0,1)$, we get

$$X_1 - X_2 + X_3 \sim N(0,3)$$

and

$$\frac{X_1 - X_2 + X_3}{\sqrt{3}} \sim N(0,1).$$

Further, since $X_i \sim N(0,1)$, we have

$$X_i^2 \sim \chi^2(1)$$

and hence

$$X_1^2 + X_2^2 + X_3^2 + X_4^2 \sim \chi^2(4)$$

Thus,

$$\frac{\frac{X_1 - X_2 + X_3}{\sqrt{3}}}{\sqrt{\frac{X_1^2+X_2^2+X_3^2+X_4^2}{4}}} = \left(\frac{2}{\sqrt{3}}\right) W \sim t(4).$$

Now using the distribution of W, we find the expected value of W.

$$\begin{aligned} E\left[W\right] &= \left(\frac{\sqrt{3}}{2}\right) E\left[\frac{2}{\sqrt{3}} W\right] \\ &= \left(\frac{\sqrt{3}}{2}\right) E\left[t(4)\right] \\ &= \left(\frac{\sqrt{3}}{2}\right) 0 \\ &= 0. \end{aligned}$$

Example 14.9. If $X \sim N(0,1)$ and X_1, X_2 is random sample of size two from the population X, then what is the 75^{th} percentile of the statistic $W = \frac{X_1}{\sqrt{X_2^2}}$?

Answer: Since each $X_i \sim N(0,1)$, we have

$$\begin{aligned} X_1 &\sim N(0,1) \\ X_2^2 &\sim \chi^2(1). \end{aligned}$$

Hence

$$W = \frac{X_1}{\sqrt{X_2^2}} \sim t(1).$$

The 75^{th} percentile a of W is then given by

$$0.75 = P(W \leq a)$$

Hence, from the t-table, we get

$$a = 1.0$$

Hence the 75^{th} percentile of W is 1.0.

Example 14.10. Suppose $X_1, X_2, ..., X_n$ is a random sample from a normal distribution with mean μ and variance σ^2. If $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $V^2 =$

$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$, and X_{n+1} is an additional observation, what is the value of m so that the statistics $\frac{m(\overline{X}-X_{n+1})}{V}$ has a t-distribution.

Answer: Since

$$\begin{aligned}
&X_i \sim N(\mu,\sigma^2)\\
\Rightarrow\ &\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\\
\Rightarrow\ &\overline{X}-X_{n+1} \sim N\left(\mu-\mu,\ \frac{\sigma^2}{n}+\sigma^2\right)\\
\Rightarrow\ &\overline{X}-X_{n+1} \sim N\left(0,\ \left(\frac{n+1}{n}\right)\sigma^2\right)\\
\Rightarrow\ &\frac{\overline{X}-X_{n+1}}{\sigma\sqrt{\frac{n+1}{n}}} \sim N(0,1)
\end{aligned}$$

Now, we establish a relationship between V^2 and S^2. We know that

$$\begin{aligned}
(n-1)\,S^2 &= (n-1)\,\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i-\overline{X})^2\\
&= \sum_{i=1}^{n}(X_i-\overline{X})^2\\
&= n\ \left(\frac{1}{n}\ \sum_{i=1}^{n}(X_i-\overline{X})^2\right)\\
&= n\,V^2.
\end{aligned}$$

Hence, by Theorem 14.3

$$\frac{n\,V^2}{\sigma^2} = \frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2(n-1).$$

Thus

$$\left(\sqrt{\frac{n-1}{n+1}}\right)\ \frac{\overline{X}-X_{n+1}}{V} = \frac{\frac{\overline{X}-X_{n+1}}{\sigma\sqrt{\frac{n+1}{n}}}}{\sqrt{\frac{\frac{n\,V^2}{\sigma^2}}{(n-1)}}} \sim t(n-1).$$

Thus by comparison, we get

$$m = \sqrt{\frac{n-1}{n+1}}.$$

14.3. Snedecor's F-distribution

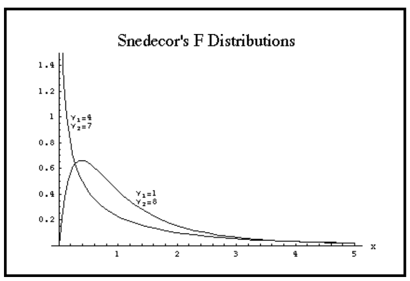
The next sampling distribution to be discussed in this chapter is Snedecor's F-distribution. This distribution has many applications in mathematical statistics. In the analysis of variance, this distribution is used to develop the technique for testing the equalities of sample means.

Definition 14.3. A continuous random variable X is said to have a F-distribution with ν_1 and ν_2 degrees of freedom if its probability density function is of the form

$$f(x;\, \nu_1, \nu_2) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\left(1+\frac{\nu_1}{\nu_2}\,x\right)^{\left(\frac{\nu_1+\nu_2}{2}\right)}} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise,} \end{cases}$$

where $\nu_1,\, \nu_2 > 0$. If X has a F-distribution with ν_1 and ν_2 degrees of freedom, then we denote it by writing $X \sim F\,(\nu_1, \nu_2)$.

The F-distribution was named in honor of Sir Ronald Fisher by George Snedecor. F-distribution arises as the distribution of a ratio of variances. Like, the other two distributions this distribution also tends to normal distribution as ν_1 and ν_2 become very large. The following figure illustrates the shape of the graph of this distribution for various degrees of freedom.



The following theorem gives us the mean and variance of Snedecor's F-distribution.

Theorem 14.8. If the random variable $X \sim F\,(\nu_1, \nu_2)$, then

$$E[X] = \begin{cases} \frac{\nu_2}{\nu_2-2} & \text{if } \quad \nu_2 \geq 3 \\ DNE & \text{if } \quad \nu_2 = 1, 2 \end{cases}$$

and

$$Var[X] = \begin{cases} \frac{2\,\nu_2^2\,(\nu_1+\nu_2-2)}{\nu_1\,(\nu_2-2)^2\,(\nu_2-4)} & \text{if } \quad \nu_2 \geq 5 \\ DNE & \text{if } \quad \nu_2 = 1, 2, 3, 4. \end{cases}$$

Here DNE means does not exist.

Example 14.11. If $X \sim F(9,10)$, what $P(X \geq 3.02)$? Also, find the mean and variance of X.

Answer:

$$\begin{aligned} P(X \geq 3.02) &= 1 - P(X \leq 3.02) \\ &= 1 - P(F(9,10) \leq 3.02) \\ &= 1 - 0.95 \qquad \text{(from F} - \text{table)} \\ &= 0.05. \end{aligned}$$

Next, we determine the mean and variance of X using the Theorem 14.8. Hence,

$$E(X) = \frac{\nu_2}{\nu_2 - 2} = \frac{10}{10-2} = \frac{10}{8} = 1.25$$

and

$$\begin{aligned} Var(X) &= \frac{2\,\nu_2^2\,(\nu_1 + \nu_2 - 2)}{\nu_1\,(\nu_2 - 2)^2\,(\nu_2 - 4)} \\ &= \frac{2\,(10)^2\,(19-2)}{9\,(8)^2\,(6)} \\ &= \frac{(25)\,(17)}{(27)\,(16)} \\ &= \frac{425}{432} = 0.9838. \end{aligned}$$

Theorem 14.9. If $X \sim F(\nu_1, \nu_2)$, then the random variable $\frac{1}{X} \sim F(\nu_2, \nu_1)$.

This theorem is very useful for computing probabilities like $P(X \leq 0.2439)$. If you look at a F-table, you will notice that the table start with values bigger than 1. Our next example illustrates how to find such probabilities using Theorem 14.9.

Example 14.12. If $X \sim F(6,9)$, what is the probability that X is less than or equal to 0.2439 ?

Answer: We use the above theorem to compute

$$\begin{aligned} P\left(X \leq 0.2439\right) &= P\left(\frac{1}{X} \geq \frac{1}{0.2439}\right) \\ &= P\left(F(9,6) \geq \frac{1}{0.2439}\right) \qquad \text{(by Theorem 14.9)} \\ &= 1 - P\left(F(9,6) \leq \frac{1}{0.2439}\right) \\ &= 1 - P\left(F(9,6) \leq 4.10\right) \\ &= 1 - 0.95 \\ &= 0.05. \end{aligned}$$

The following theorem says that F-distribution arises as the distribution of a random variable which is the quotient of two independently distributed chi-square random variables, each of which is divided by its degrees of freedom.

Theorem 14.10. If $U \sim \chi^2(\nu_1)$ and $V \sim \chi^2(\nu_2)$, and the random variables U and V are independent, then the random variable

$$\frac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}} \sim F\left(\nu_1, \nu_2\right).$$

Example 14.13. Let $X_1, X_2, ..., X_4$ and $Y_1, Y_2, ..., Y_5$ be two random samples of size 4 and 5 respectively, from a standard normal population. What is the variance of the statistic $T = \left(\frac{5}{4}\right) \frac{X_1^2+X_2^2+X_3^2+X_4^2}{Y_1^2+Y_2^2+Y_3^2+Y_4^2+Y_5^2}$?

Answer: Since the population is standard normal, we get

$$X_1^2 + X_2^2 + X_3^2 + X_4^2 \sim \chi^2(4).$$

Similarly,

$$Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2 \sim \chi^2(5).$$

Thus

$$\begin{aligned} T &= \left(\frac{5}{4}\right) \frac{X_1^2 + X_2^2 + X_3^2 + X_4^2}{Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2} \\ &= \frac{\frac{X_1^2+X_2^2+X_3^2+X_4^2}{4}}{\frac{Y_1^2+Y_2^2+Y_3^2+Y_4^2+Y_5^2}{5}} \\ &= T \sim F(4,5). \end{aligned}$$

Therefore

$$\begin{aligned} Var(T) &= Var[\,F(4,5)\,] \\ &= \frac{2\,(5)^2\,(7)}{4\,(3)^2\,(1)} \\ &= \frac{350}{36} \\ &= 9.72. \end{aligned}$$

Theorem 14.11. Let $X \sim N(\mu_1,\, \sigma_1^2)$ and $X_1, X_2, ..., X_n$ be a random sample of size n from the population X. Let $Y \sim N(\mu_2,\, \sigma_2^2)$ and $Y_1, Y_2, ..., Y_m$ be a random sample of size m from the population Y. Then the statistic

$$\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n-1,\, m-1),$$

where S_1^2 and S_2^2 denote the sample variances of the first and the second sample, respectively.

Proof: Since,

$$X_i \sim N(\mu_1, \sigma_1^2)$$

we have by Theorem 14.3, we get

$$(n-1)\,\frac{S_1^2}{\sigma_1^2} \sim \chi^2(n-1).$$

Similarly, since

$$Y_i \sim N(\mu_2, \sigma_2^2)$$

we have by Theorem 14.3, we get

$$(m-1)\,\frac{S_2^2}{\sigma_2^2} \sim \chi^2(m-1).$$

Therefore

$$\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{\frac{(n-1)\,S_1^2}{(n-1)\,\sigma_1^2}}{\frac{(m-1)\,S_2^2}{(m-1)\,\sigma_2^2}} \sim F(n-1,\, m-1).$$

This completes the proof of the theorem.

Because of this theorem, the F-distribution is also known as the variance-ratio distribution.

14.4. Review Exercises

1. Let $X_1, X_2, ..., X_5$ be a random sample of size 5 from a normal distribution with mean zero and standard deviation 2. Find the sampling distribution of the statistic $X_1 + 2X_2 - X_3 + X_4 + X_5$.

2. Let X_1, X_2, X_3 be a random sample of size 3 from a standard normal distribution. Find the distribution of $X_1^2 + X_2^2 + X_3^2$. If possible, find the sampling distribution of $X_1^2 - X_2^2$. If not, justify why you can not determine it's distribution.

3. Let X_1, X_2, X_3 be a random sample of size 3 from a standard normal distribution. Find the sampling distribution of the statistics $\frac{X_1+X_2+X_3}{\sqrt{X_1^2+X_2^2+X_3^2}}$ and $\frac{X_1-X_2-X_3}{\sqrt{X_1^2+X_2^2+X_3^2}}$.

4. Let X_1, X_2, X_3 be a random sample of size 3 from an exponential distribution with a parameter $\theta > 0$. Find the distribution of the sample (that is the joint distribution of the random variables X_1, X_2, X_3).

5. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a normal population with mean μ and variance $\sigma^2 > 0$. What is the expected value of the sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$?

6. Let X_1, X_2, X_3, X_4 be a random sample of size 4 from a standard normal population. Find the distribution of the statistic $\frac{X_1+X_4}{\sqrt{X_2^2+X_3^2}}$.

7. Let X_1, X_2, X_3, X_4 be a random sample of size 4 from a standard normal population. Find the sampling distribution (if possible) and moment generating function of the statistic $2X_1^2+3X_2^2+X_3^2+4X_4^2$. What is the probability distribution of the sample?

8. Let X equal the maximal oxygen intake of a human on a treadmill, where the measurement are in milliliters of oxygen per minute per kilogram of weight. Assume that for a particular population the mean of X is $\mu = 54.03$ and the standard deviation is $\sigma = 5.8$. Let $\bar{X}$ be the sample mean of a random sample $X_1, X_2, ..., X_{47}$ of size 47 drawn from X. Find the probability that the sample mean is between 52.761 and 54.453.

9. Let $X_1, X_2, ..., X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. What is the variance of $V^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$?

10. If X is a random variable with mean μ and variance σ^2, then $\mu - 2\sigma$ is called the lower 2σ point of X. Suppose a random sample X_1, X_2, X_3, X_4 is drawn from a chi-square distribution with two degrees of freedom. What is the lower 2σ point of $X_1 + X_2 + X_3 + X_4$?

11. Let X and Y be independent normal random variables such that the mean and variance of X are 2 and 4, respectively, while the mean and variance of Y are 6 and k, respectively. A sample of size 4 is taken from the X-distribution and a sample of size 9 is taken from the Y-distribution. If $P\left(\overline{Y} - \overline{X} > 8\right) = 0.0228$, then what is the value of the constant k ?

12. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with density function

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the distribution of the statistic $Y = 2\lambda \sum_{i=1}^{n} X_i$?

13. Suppose X has a normal distribution with mean 0 and variance 1, Y has a chi-square distribution with n degrees of freedom, W has a chi-square distribution with p degrees of freedom, and W, X, and Y are independent. What is the sampling distribution of the statistic $V = \frac{X}{\sqrt{\frac{W+Y}{p+n}}}$?

14. A random sample $X_1, X_2, ..., X_n$ of size n is selected from a normal population with mean μ and standard deviation 1. Later an additional independent observation X_{n+1} is obtained from the same population. What is the distribution of the statistic $(X_{n+1} - \mu)^2 + \sum_{i=1}^{n}(X_i - \overline{X})^2$, where $\overline{X}$ denote the sample mean?

15. Let $T = \frac{k(X+Y)}{\sqrt{Z^2+W^2}}$, where X, Y, Z, and W are independent normal random variables with mean 0 and variance $\sigma^2 > 0$. For exactly one value of k, T has a t-distribution. If r denotes the degrees of freedom of that distribution, then what is the value of the pair (k, r)?

16. Let X and Y be joint normal random variables with common mean 0, common variance 1, and covariance $\frac{1}{2}$. What is the probability of the event $\left(X + Y \leq \sqrt{3}\right)$, that is $P\left(X + Y \leq \sqrt{3}\right)$?

17. Suppose $X_j = Z_j - Z_{j-1}$, where $j = 1, 2, ..., n$ and $Z_0, Z_1, ..., Z_n$ are independent and identically distributed with common variance σ^2. What is the variance of the random variable $\frac{1}{n} \sum_{j=1}^{n} X_j$?

18. A random sample of size 5 is taken from a normal distribution with mean 0 and standard deviation 2. Find the constant k such that 0.05 is equal to the probability that the sum of the squares of the sample observations exceeds the constant k.

19. Let $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$ be two random sample from the independent normal distributions with $Var[X_i] = \sigma^2$ and $Var[Y_i] = 2\sigma^2$, for $i = 1, 2, ..., n$ and $\sigma^2 > 0$. If $U = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ and $V = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$, then what is the sampling distribution of the statistic $\frac{2U+V}{2\sigma^2}$?

20. Suppose $X_1, X_2, ..., X_6$ and $Y_1, Y_2, ..., Y_9$ are independent, identically distributed normal random variables, each with mean zero and variance $\sigma^2 > 0$. What is the 95^{th} percentile of the statistics $W = \left[\sum_{i=1}^{6} X_i^2\right] / \left[\sum_{j=1}^{9} Y_j^2\right]$?

21. Let $X_1, X_2, ..., X_6$ and $Y_1, Y_2, ..., Y_8$ be independent random samples from a normal distribution with mean 0 and variance 1, and $Z = \left[4\sum_{i=1}^{6} X_i^2\right] / \left[3\sum_{j=1}^{8} Y_j^2\right]$?