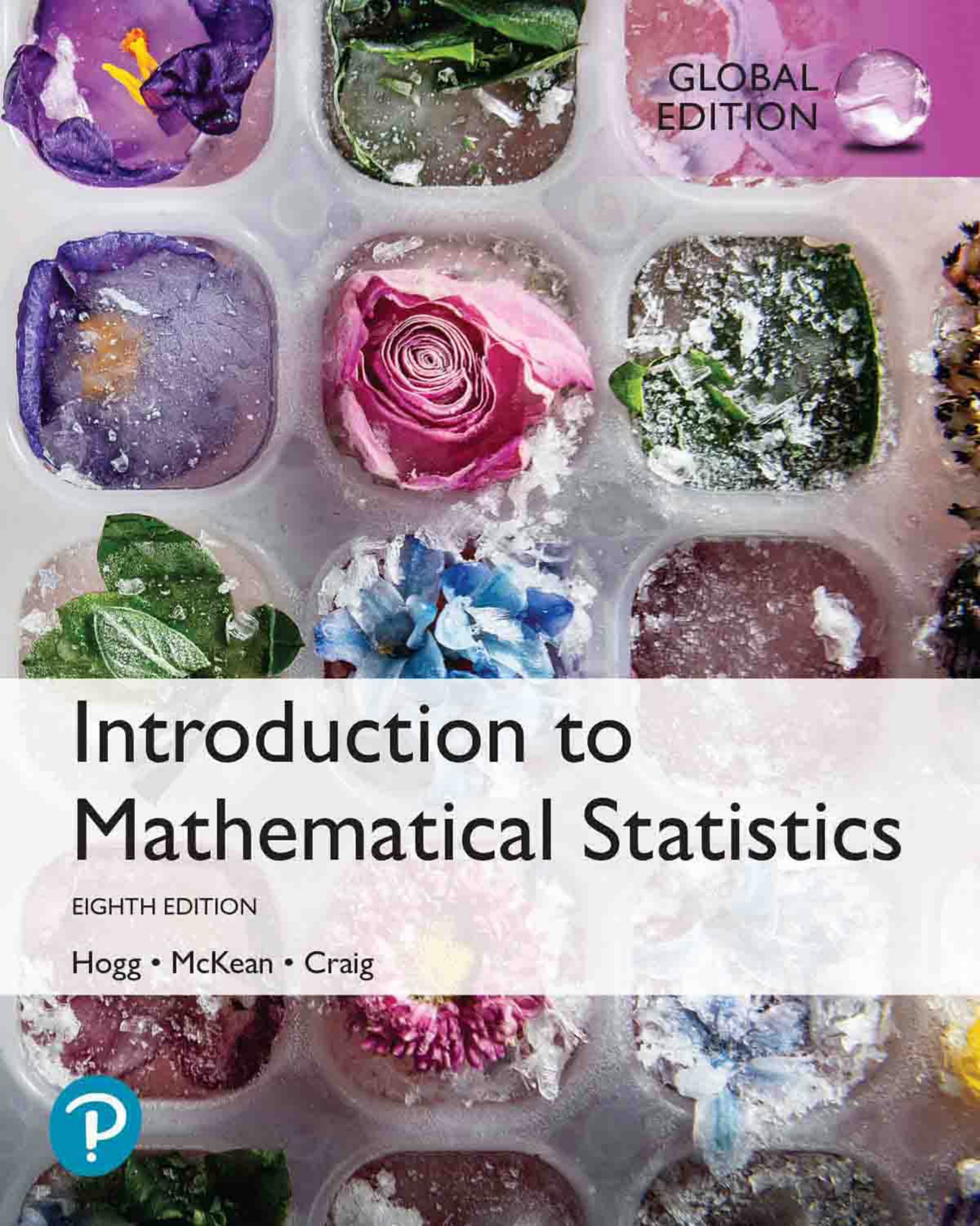

GLOBAL EDITION

Introduction to Mathematical Statistics

EIGHTH EDITION

Hogg • McKean • Craig

Introduction to Mathematical Statistics

Eighth Edition

Global Edition

Robert V. Hogg
University of Iowa

Joseph W. McKean
Western Michigan University

Allen T. Craig
Late Professor of Statistics
University of Iowa

Pearson

In terms of R computation, the command `pf(2.50,3,8)` computes to the value 0.8665 which is the probability $P(F \le 2.50)$ when F has the F-distribution with 3 and 8 degrees of freedom. The 95th percentile of F is `qf(.95,3,8) = 4.066` and the code `x=seq(.01,5,.01);plot(df(x,3,8)~x)` draws a plot of the pdf of this F random variable. Note that the pdf is right-skewed. Before the age of modern computation, tables of the quantiles of F-distributions for selected probabilities and degrees of freedom were used. Table IV in Appendix D displays the 95th and 99th quantiles for selected degrees of freedom. Besides its use in statistics, the F-distribution is used to model lifetime data; see Exercise 3.6.13.

Example 3.6.2 (Moments of F-Distributions)**.** Let F have an F-distribution with r_1 and r_2 degrees of freedom. Then, as in expression (3.6.7), we can write $F = (r_2/r_1)(U/V)$, where U and V are independent χ^2 random variables with r_1 and r_2 degrees of freedom, respectively. Hence, for the kth moment of F, by independence we have

$$E\left(F^k\right) = \left(\frac{r_2}{r_1}\right)^k E\left(U^k\right) E\left(V^{-k}\right),$$

provided, of course, that both expectations on the right side exist. By Theorem 3.3.2, because $k > -(r_1/2)$ is always true, the first expectation always exists. The second expectation, however, exists if $r_2 > 2k$; i.e., the denominator degrees of freedom must exceed twice k. Assuming this is true, it follows from (3.3.8) that the mean of F is given by

$$E(F) = \frac{r_2}{r_1} r_1 \frac{2^{-1}\Gamma\left(\frac{r_2}{2} - 1\right)}{\Gamma\left(\frac{r_2}{2}\right)} = \frac{r_2}{r_2 - 2}. \tag{3.6.8}$$

If r_2 is large, then $E(F)$ is about 1. In Exercise 3.6.7, a general expression for $E(F^k)$ is derived. ■

3.6.3 Student's Theorem

Our final note in this section concerns an important result for the later chapters on inference for normal random variables. It is a corollary to the t-distribution derived above and is often referred to as Student's Theorem.

Theorem 3.6.1. *Let $X_1, \ldots, X_n$ be iid random variables each having a normal distribution with mean μ and variance σ^2. Define the random variables*

$$\overline{X} = \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i \ \text{ and } \ S^2 = \tfrac{1}{n-1}\textstyle\sum_{i=1}^n (X_i - \overline{X})^2.$$

Then

(a) *$\overline{X}$ has a $N\left(\mu, \frac{\sigma^2}{n}\right)$ distribution.*

(b) *$\overline{X}$ and S^2 are independent.*

(c) *$(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution.*

(d) *The random variable*

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{3.6.9}$$

has a Student t-distribution with $n-1$ degrees of freedom.

Proof: Note that we have proved part (a) in Corollary 3.4.1. Let $\mathbf{X} = (X_1, \ldots, X_n)'$. Because $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ random variables, $\mathbf{X}$ has a multivariate normal distribution $N(\mu\mathbf{1}, \sigma^2\mathbf{I})$, where $\mathbf{1}$ denotes a vector whose components are all 1. Let $\mathbf{v}' = (1/n, \ldots, 1/n) = (1/n)\mathbf{1}'$. Note that $\overline{X} = \mathbf{v}'\mathbf{X}$. Define the random vector $\mathbf{Y}$ by $\mathbf{Y} = (X_1 - \overline{X}, \ldots, X_n - \overline{X})'$. Consider the following transformation:

$$\mathbf{W} = \begin{bmatrix} \overline{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{v}' \\ \mathbf{I} - \mathbf{1}\mathbf{v}' \end{bmatrix} \mathbf{X}. \tag{3.6.10}$$

Because $\mathbf{W}$ is a linear transformation of multivariate normal random vector, by Theorem 3.5.2 it has a multivariate normal distribution with mean

$$E[\mathbf{W}] = \begin{bmatrix} \mathbf{v}' \\ \mathbf{I} - \mathbf{1}\mathbf{v}' \end{bmatrix} \mu\mathbf{1} = \begin{bmatrix} \mu \\ \mathbf{0}_n \end{bmatrix}, \tag{3.6.11}$$

where $\mathbf{0}_n$ denotes a vector whose components are all 0, and covariance matrix

$$\begin{aligned} \mathbf{\Sigma} &= \begin{bmatrix} \mathbf{v}' \\ \mathbf{I} - \mathbf{1}\mathbf{v}' \end{bmatrix} \sigma^2\mathbf{I} \begin{bmatrix} \mathbf{v}' \\ \mathbf{I} - \mathbf{1}\mathbf{v}' \end{bmatrix}' \\ &= \sigma^2 \begin{bmatrix} \frac{1}{n} & \mathbf{0}_n' \\ \mathbf{0}_n & \mathbf{I} - \mathbf{1}\mathbf{v}' \end{bmatrix}. \end{aligned} \tag{3.6.12}$$

Because $\overline{X}$ is the first component of $\mathbf{W}$, we can also obtain part (a) by Theorem 3.5.1. Next, because the covariances are 0, $\overline{X}$ is independent of $\mathbf{Y}$. But $S^2 = (n-1)^{-1}\mathbf{Y}'\mathbf{Y}$. Hence, $\overline{X}$ is independent of S^2, also. Thus part (b) is true.

Consider the random variable

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

Each term in this sum is the square of a $N(0,1)$ random variable and, hence, has a $\chi^2(1)$ distribution (Theorem 3.4.1). Because the summands are independent, it follows from Corollary 3.3.1 that V is a $\chi^2(n)$ random variable. Note the following identity:

$$\begin{aligned} V &= \sum_{i=1}^{n} \left(\frac{(X_i - \overline{X}) + (\overline{X} - \mu)}{\sigma} \right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma} \right)^2 + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \\ &= \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \right)^2. \end{aligned} \tag{3.6.13}$$

By part (b), the two terms on the right side of the last equation are independent. Further, the second term is the square of a standard normal random variable and, hence, has a $\chi^2(1)$ distribution. Taking mgfs of both sides, we have

$$(1-2t)^{-n/2} = E\left[\exp\{t(n-1)S^2/\sigma^2\}\right](1-2t)^{-1/2}. \tag{3.6.14}$$

Solving for the mgf of $(n-1)S^2/\sigma^2$ on the right side we obtain part (c). Finally, part (d) follows immediately from parts (a)–(c) upon writing T, (3.6.9), as

$$T = \frac{(\overline{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)S^2/(\sigma^2(n-1))}}. \quad \blacksquare$$

EXERCISES

3.6.1. Let T have a t-distribution with 10 degrees of freedom. Find $P(|T| > 2.228)$ from either Table III or by using R.

3.6.2. Let T have a t-distribution with 14 degrees of freedom. Determine b so that $P(-b < T < b) = 0.90$. Use either Table III or by using R.

3.6.3. Let T have a t-distribution with $r > 4$ degrees of freedom. Use expression (3.6.4) to determine the kurtosis of T. See Exercise 1.9.15 for the definition of kurtosis.

3.6.4. Using R, plot the pdfs of the random variables defined in parts (a)–(e) below. Obtain an overlay plot of all five pdfs, also.

(a) X has a standard normal distribution. Use this code:
```
x=seq(-6,6,.001); plot(dnorm(x)~x).
```

(b) X has a t-distribution with 1 degree of freedom. Use the code:
```
lines(dt(x,1)~x,lty=2).
```

(c) X has a t-distribution with 5 degrees of freedom.
```
lines(dt(x,5)~x,lty=2).
```

(d) X has a t-distribution with 15 degrees of freedom.
```
lines(dt(x,15)~x,lty=2).
```

(e) X has a t-distribution with 25 degrees of freedom.
```
lines(dt(x,25)~x,lty=2).
```

3.6.5. Using R, investigate the probabilities of an "outlier" for a t-random variable and a normal random variable. Specifically, determine the probability of observing the event $\{|X| \geq 2\}$ for the following random variables:

(a) X has a standard normal distribution.

(b) X has a t-distribution with 1 degree of freedom.

(c) X has a t-distribution with 3 degrees of freedom.

(d) X has a t-distribution with 10 degrees of freedom.

(e) X has a t-distribution with 30 degrees of freedom.

3.6.6. In expression (3.4.13), the normal location model was presented. Often real data, though, have more outliers than the normal distribution allows. Based on Exercise 3.6.5, outliers are more probable for t-distributions with small degrees of freedom. Consider a location model of the form

$$X = \mu + e,$$

where e has a t-distribution with 3 degrees of freedom. Determine the standard deviation σ of X and then find $P(|X - \mu| \geq \sigma)$.

3.6.7. Let F have an F-distribution with parameters r_1 and r_2. Assuming that $r_2 > 2k$, continue with Example 3.6.2 and derive the $E(F^k)$.

3.6.8. Let F have an F-distribution with parameters r_1 and r_2. Using the results of the last exercise, determine the kurtosis of F, assuming that $r_2 > 8$.

3.6.9. Let F have an F-distribution with parameters r_1 and r_2. Argue that $1/F$ has an F-distribution with parameters r_2 and r_1.

3.6.10. Suppose F has an F-distribution with parameters $r_1 = 5$ and $r_2 = 10$. Using only 95th percentiles of F-distributions, find a and b so that $P(F \leq a) = 0.05$ and $P(F \leq b) = 0.95$, and, accordingly, $P(a < F < b) = 0.90$.
Hint: Write $P(F \leq a) = P(1/F \geq 1/a) = 1 - P(1/F \leq 1/a)$, and use the result of Exercise 3.6.9 and R.

3.6.11. Let $U = |W|/\sqrt{V/r}$, where the independent variables W and V are, respectively, normal with mean zero and variance 1 and chi-square with r degrees of freedom. Show that U^2 has an F-distribution with parameters $r_1 = 1$ and $r_2 = r$.

3.6.12. Show that the t-distribution with $r = 1$ degree of freedom and the Cauchy distribution are the same.

3.6.13. Let F have an F-distribution with $2r$ and $2s$ degrees of freedom. Since the support of F is $(0, \infty)$, the F-distribution is often used to model time until failure (lifetime). In this case, $Y = \log F$ is used to model the log of lifetime. The $\log F$ family is a rich family of distributions consisting of left- and right-skewed distributions as well as symmetric distributions; see, for example, Chapter 4 of Hettmansperger and McKean (2011). In this exercise, consider the subfamily where $Y = \log F$ and F has 2 and $2s$ degrees of freedom.

(a) Obtain the pdf and cdf of Y.

(b) Using R, obtain a page of plots of these distributions for $s = .4, .6, 1.0, 2.0, 4.0, 8$. Comment on the shape of each pdf.

(c) For $s = 1$, this distribution is called the **logistic** distribution. Show that the pdf is symmetric about 0.

3.6.14. Show that

$$Y = \frac{1}{1 + (r_1/r_2)W},$$

where W has an F-distribution with parameters r_1 and r_2, has a beta distribution.

3.6.15. Let $X_1,\ X_2$ be iid with common distribution having the pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Show that $Z = X_1/X_2$ has an F-distribution.

3.6.16. Let $X_1,\ X_2$, and X_3 be three independent chi-square variables with $r_1,\ r_2$, and r_3 degrees of freedom, respectively.

(a) Show that $Y_1 = X_1/X_2$ and $Y_2 = X_1 + X_2$ are independent and that Y_2 is $\chi^2(r_1 + r_2)$.

(b) Deduce that

$$\frac{X_1/r_1}{X_2/r_2} \quad \text{and} \quad \frac{X_3/r_3}{(X_1 + X_2)/(r_1 + r_2)}$$

are independent F-variables.

3.7 *Mixture Distributions

Recall the discussion on the contaminated normal distribution given in Section 3.4.1. This was an example of a mixture of normal distributions. In this section, we extend this to mixtures of distributions in general. Generally, we use continuous-type notation for the discussion, but discrete pmfs can be handled the same way.

Suppose that we have k distributions with respective pdfs $f_1(x), f_2(x), \ldots, f_k(x)$, with supports $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_k$, means $\mu_1, \mu_2, \ldots, \mu_k$, and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, with positive mixing probabilities $p_1, p_2, \ldots, p_k$, where $p_1 + p_2 + \cdots + p_k = 1$. Let $\mathcal{S} = \cup_{i=1}^k \mathcal{S}_i$ and consider the function

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \cdots + p_k f_k(x) = \sum_{i=1}^{k} p_i f_i(x), \quad x \in \mathcal{S}. \tag{3.7.1}$$

Note that $f(x)$ is nonnegative and it is easy to see that it integrates to one over $(-\infty, \infty)$; hence, $f(x)$ is a pdf for some continuous-type random variable X. Integrating term-by-term, it follows that the cdf of X is:

$$F(x) = \sum_{i=1}^{k} p_i F_i(x), \quad x \in \mathcal{S}, \tag{3.7.2}$$

where $F_i(x)$ is the cdf corresponding to the pdf $f_i(x)$. The mean of X is given by

$$E(X) = \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} x f_i(x)\, dx = \sum_{i=1}^{k} p_i \mu_i = \overline{\mu}, \tag{3.7.3}$$

a weighted average of $\mu_1, \mu_2, \ldots, \mu_k$, and the variance equals

$$\begin{aligned}\operatorname{var}(X) &= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} (x-\overline{\mu})^2 f_i(x)\,dx \\ &= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} [(x-\mu_i)+(\mu_i-\overline{\mu})]^2 f_i(x)\,dx \\ &= \sum_{i=1}^{k} p_i \int_{-\infty}^{\infty} (x-\mu_i)^2 f_i(x)\,dx + \sum_{i=1}^{k} p_i(\mu_i-\overline{\mu})^2 \int_{-\infty}^{\infty} f_i(x)\,dx,\end{aligned}$$

because the cross-product terms integrate to zero. That is,

$$\operatorname{var}(X) = \sum_{i=1}^{k} p_i\sigma_i^2 + \sum_{i=1}^{k} p_i(\mu_i-\overline{\mu})^2. \tag{3.7.4}$$

Note that the variance is not simply the weighted average of the k variances, but it also includes a positive term involving the weighted variance of the means.

Remark 3.7.1. It is extremely important to note these characteristics are associated with a mixture of k distributions and have nothing to do with a linear combination, say $\sum a_i X_i$, of k random variables. ■

For the next example, we need the following distribution. We say that X has a **loggamma** pdf with parameters $\alpha > 0$ and $\beta > 0$ if it has pdf

$$f_1(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{-(1+\beta)/\beta}(\log x)^{\alpha-1} & x > 1 \\ 0 & \text{elsewhere.}\end{cases} \tag{3.7.5}$$

The derivation of this pdf is given in Exercise 3.7.1, where its mean and variance are also derived. We denote this distribution of X by $\log\Gamma(\alpha,\beta)$.

Example 3.7.1. Actuaries have found that a mixture of the loggamma and gamma distributions is an important model for claim distributions. Suppose, then, that X_1 is $\log\Gamma(\alpha_1,\beta_1)$, X_2 is $\Gamma(\alpha_2,\beta_2)$, and the mixing probabilities are p and $(1-p)$. Then the pdf of the mixture distribution is

$$f(x) = \begin{cases} \frac{1-p}{\beta_2^{\alpha_2}\Gamma(\alpha_2)} x^{\alpha_2-1}e^{-x/\beta_2} & 0 < x \le 1 \\ \frac{p}{\beta_1^{\alpha_1}\Gamma(\alpha_1)}(\log x)^{\alpha_1-1}x^{-(\beta_1+1)/\beta_1} + \frac{1-p}{\beta_2^{\alpha_2}\Gamma(\alpha_2)}x^{\alpha_2-1}e^{-x/\beta_2} & 1 < x \\ 0 & \text{elsewhere.}\end{cases} \tag{3.7.6}$$

Provided $\beta_1 < 2^{-1}$, the mean and the variance of this mixture distribution are

$$\begin{aligned}\mu &= p(1-\beta_1)^{-\alpha_1} + (1-p)\alpha_2\beta_2 & (3.7.7)\\ \sigma^2 &= p[(1-2\beta_1)^{-\alpha_1} - (1-\beta_1)^{-2\alpha_1}] \\ &\quad + (1-p)\alpha_2\beta_2^2 + p(1-p)[(1-\beta_1)^{-\alpha_1} - \alpha_2\beta_2]^2; & (3.7.8)\end{aligned}$$

see Exercise 3.7.3. ■