Reference: Chapter 6 of Devore(8e) ESTIMATION Maghsoodloo

There are two types of estimators: point and interval estimates. A point estimate, or estimator[1], is a random variable (or statistic) before the sample is drawn, but it becomes a numerical value after the sample is drawn estimating a population parameter [such as $\mu = E(X)$ and $\sigma = \sqrt{V(X)}$]. For example, any one of the statistics, the sample mean, median, mode, and the mid-range $[x_{(1)}+x_{(n)}]/2$ can be used as point estimators of the population mean μ. If the parameter under consideration is the population variance σ^2, then we may use any one of the statistics $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ = CSS/(n – 1) , or the 2nd sample central moment $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = m_2$ as point estimates of σ^2. Further, if the universe is normal, then $(R/d_2)^2$ is an estimator of σ^2, where R is the sample range, $d_2 = E(R/\sigma)$, and the rv $R/\sigma = w$ is called the relative range in the field of QC (Quality Control). Another point estimator of σ^2 of a normal universe is $\hat{\boldsymbol{\sigma}}^{\mathbf{2}} = (IQR/1.3489795)^2$. The question is then how one chooses among several available point estimators of a given process parameter? The obvious answer is to select the estimator that is the most accurate, i.e., whose value comes closest to the true value of the parameter being estimated in the long run, i.e., over all possible samples of size n.

Since a point estimator, $\hat{\theta}$, is a random variable (i.e., it changes randomly from sample to sample), it has a frequency function, and for $\hat{\theta}$ to be an "accurate" estimator, the pdf of $\hat{\theta}$ should be closely concentrated about the parameter θ that $\hat{\theta}$ is estimating. The criterion used to decide which one of the competing estimators ($\hat{\theta}_1$ or $\hat{\theta}_2$) of the parameter θ is better now follows.

---

[1]The term estimator is applied to the random variable, and the word estimate is reserved for the numerical value taken on by the random variable after the sample is drawn and the experimental data have been inserted into the expression for the estimator.

The statistic $\hat{\theta}_1$ is said to be a more accurate estimator of θ than $\hat{\theta}_2$ iff $E[(\hat{\theta}_1 - \theta)^2] < E[(\hat{\theta}_2 - \theta)^2]$. Since $E[(\hat{\theta} - \theta)^2]$ is defined to be the mean square error of $\hat{\theta}$, i.e., $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$, then $\hat{\theta}_1$ is a more accurate estimator than $\hat{\theta}_2$ iff $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$, as depicted in Figure 6.

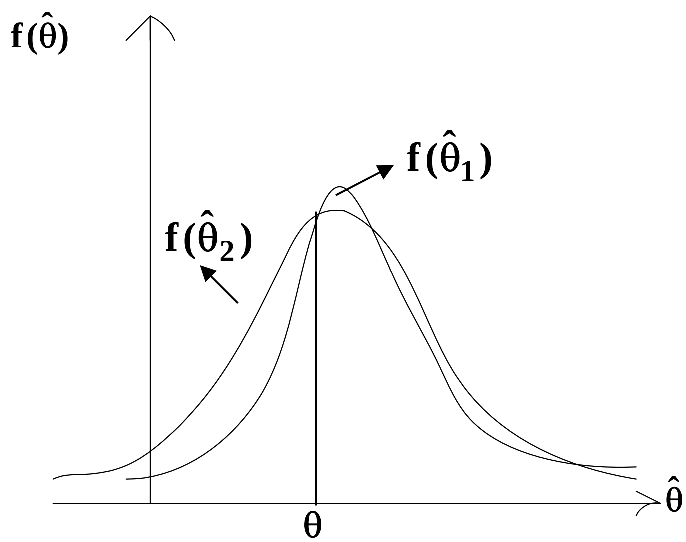

Figure 6. Density Functions (pdfs) of Two Estimators of θ

<u>Example 33</u>. A population parameter, θ, is believed to have the value of 2. Two statistics that can be used to estimate $\theta = 2$ are $\hat{\theta}_1$ and $\hat{\theta}_2$. The frequency functions (or pmfs) of $\hat{\theta}_1$ and $\hat{\theta}_2$ are:

$$p(\hat{\theta}_1) = \begin{cases} 1/8, & \hat{\theta}_1 = 0, 6 \\ 2/8, & \hat{\theta}_1 = 1 \\ 4/8, & \hat{\theta}_1 = 2 \end{cases} \quad \text{and} \quad p(\hat{\theta}_2) = \begin{cases} 2/5, & \hat{\theta}_2 = 1 \\ 2/5, & \hat{\theta}_2 = 2 \\ 1/5, & \hat{\theta}_2 = 3. \end{cases}$$

Compute the mean square error (MSE) for $\hat{\theta}_1$ and $\hat{\theta}_2$, and select the most accurate estimator.

Solution. MSE($\hat{\theta}_1$) = E [$(\hat{\theta}_1 - \theta)^2$] = $(0 - 2)^2$ (1/8) + $(1 - 2)^2$(2/8) + $(2 - 2)^2$(4/8) + $(6 - 2)^2$(1/8) = 11/4 = 2.75

Similarly, E [$(\hat{\theta}_2 - \theta)^2$] = 0.60 $\rightarrow$ MSE ($\hat{\theta}_2$) < MSE ($\hat{\theta}_1$) $\rightarrow$ Thus, $\hat{\theta}_2$ is a much more accurate estimator of θ than $\hat{\theta}_1$ iff θ = 2.

PROPERTIES OF POINT ESTIMATORS

An estimator is said to be consistent iff the $\lim_{n \to \infty} \hat{\theta} = \theta$. If the population is finite of size N, then $\hat{\theta}$ is consistent iff lim $\hat{\theta} = \theta$ as n$\rightarrow$ N. For example, $\overline{x}$ is a consistent estimator of μ for both finite and infinite populations because lim $\overline{x}$ = μ as n $\rightarrow$ N. However, S^2 is not a consistent estimator of σ^2 for finite populations but S^2 becomes consistent as N $\rightarrow$ ∞.

One of the most important property of a point estimator is the amount of bias in the estimator. The amount of bias in a point estimator is defined as

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta = E(\hat{\theta} - \theta)$$

And, therefore, $\hat{\theta}$ is an unbiased estimator iff E($\hat{\theta}$) = θ (i.e., B($\hat{\theta}$) = 0). Students make the common mistake that an unbiased estimator is one whose observed value is equal to the parameter it is estimating. This is completely false! To illustrate, we compute the amount of bias for the PD's of the **Example 33** where θ = 2:

$$E(\hat{\boldsymbol{\theta}}_1) = \frac{6}{8} + \frac{2}{8} + \frac{8}{8} = 2.0 = \theta \longrightarrow B(\hat{\boldsymbol{\theta}}_1) = E(\hat{\boldsymbol{\theta}}_1) - \theta = 2 - 2 = 0.$$

Since E ($\hat{\theta}_2$) = 1.80, then B($\hat{\theta}_2$) = – 0.20, i.e., $\hat{\boldsymbol{\theta}}_2$ is a biased estimator of θ, while $\hat{\boldsymbol{\theta}}_1$ is an unbiased estimator of θ. The difference $(\hat{\theta} - \theta)$ is called the amount of error in estimation or the inaccuracy of the estimator $\hat{\theta}$, while $E(\hat{\theta} - \theta)$ is the amount of bias in the estimator $\hat{\theta}$, i.e., $E(\hat{\theta} - \theta)$ = $E(\hat{\theta})$ – θ = B($\hat{\theta}$).

We now develop a relationship between the accuracy of an estimator $\hat{\theta}$, measured by its MSE, and the amount of bias in $\hat{\theta}$. Let E $(\hat{\theta})$ = m = the mean of $\hat{\theta}$; then $B(\hat{\theta}) = m - \theta$, and the $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - m) + (m - \theta)]^2 = E[(\hat{\theta} - m) + B]^2 = E[(\hat{\theta} - m)^2 + 2B(\hat{\theta} - m) + B^2] \rightarrow MSE(\hat{\theta}) = E(\hat{\theta} - m)^2 + 0 + E(B^2) =$

$$V(\hat{\theta}) + B^2 = [E(\hat{\theta}^2) - E^2(\hat{\theta})] + B^2 \quad (32)$$

Eq. (32) clearly shows that $MSE(\hat{\theta}) = V(\hat{\theta})$ iff $\hat{\theta}$ is an unbiased estimator of θ. For the PD's of Example 33, $V(\hat{\theta}_1) = MSE(\hat{\theta}_1) = 2.75$ because $\hat{\theta}_1$ was an unbiased estimator, but $V(\hat{\theta}_2) = E(\hat{\theta}_2^2) - E^2(\hat{\theta}_2) = 19/5 - (9/5)^2 = 0.560$ so that $MSE(\hat{\theta}_2) = 0.560 + (-0.20)^2 = 0.60$, as was also obtained at the top of page 81. This example should illustrate that a biased estimator is not necessarily the least accurate estimator. In fact, $\hat{\theta}_2$ is a more accurate estimator than $\hat{\theta}_1$ iff $\theta < 7.3750$.

<u>Exercise 46.</u> (a) For the finite population of size N = 4 with elements {2, 4, 4′, 6}, compute the population mean μ and variance σ^2; note that 2 elements have the same value of 4. Note that R_{X_1,X_2} = {22, 24, 24′, 26, 42, 44, 44′, 4′6, 4′2, 4′4, 4′4′, 4′6, 62, 64, 64′, 66}, and the corresponding $R_{\bar{x}}$ = {2, 3, 3, 4, 3, 4, 4 , 5, 3, 4, 4, 5, 4, 5, 5, 6}.

(b) Obtain the PD's of $\bar{\mathbf{x}}$ and S^2 for a random sample of size n = 2 (with replacement). (c) Compute the amounts of bias in $\bar{X}$ (as an estimator of μ), and in S^2 as a point estimator of σ^2. (d) Compute the $MSE(\bar{\mathbf{x}})$ and $MSE(S^2)$. (e) Obtain the pmf of S and use it to compute its amount of bias and its MSE.

<u>Exercise 46 con't.</u> Secondly, consider an infinite population (not necessarily a normal population) with parameters μ and σ^2, and a random sample of size of n > 1 is drawn. (f) Show that $E(\bar{\mathbf{x}}) = \mu$. (g) Use the identity

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \sum_{i=1}^{n} [(x_i - \mu) - (\bar{x} - \mu)]^2 \equiv \sum_{i=1}^{n}\left[(x_i - \mu)^2 - 2(\bar{x} - \mu)(x_i - \mu) + (\bar{x} - \mu)^2\right] =$$

$\ldots \equiv \sum (x_i - \mu)^2 \; - n(\overline{x} - \mu)^2$, to prove that E ($S^2$) = σ^2 for all infinite populations.

The relative efficiency (REL-EFF) of $\hat{\theta}_1$ to $\hat{\theta}_2$ is defined as MSE($\hat{\theta}_2$)/MSE($\hat{\theta}_1$). As an example, consider a random sample of size n = 6 from a population with unknown mean μ and unknown variance σ^2. Recall that for any simple random sample, E(x_i) = μ and V(x_i) = σ^2, i = 1, 2, …, n. Let $\hat{\theta}_1 = (x_1 + x_2 + x_3 + x_4 + x_5 + x_6)/6$ and $\hat{\theta}_2 = (2x_1 + x_4 - x_6)/2$. Then, E ($\hat{\theta}_1$) = E ($\hat{\theta}_2$) = μ but the MSE ($\hat{\theta}_1$) = $\sigma^2/6$, MSE($\hat{\theta}_2$) = $1.5\sigma^2$ so that the REL-EFF of $\hat{\theta}_1$ to $\hat{\theta}_2$ = ($1.5\sigma^2$) / (σ^2/ 6) = 900 %. For the example 33, the Rel-Eff of $\hat{\theta}_2$ to $\hat{\theta}_1$ is 2.75/0.6 = 458.333%.

Exercise 47. Work Exercises 1, 4, 7, 8, 12, 13, and 15 on pages 252 -255 of Devore (8e).

Methods of Obtaining Point Estimators

The definition of bias does not generally lend itself to a method of obtaining a point estimate of a parameter θ. Devore covers two methods of obtaining a point estimate: (1) The method of Moments, (2) The method of Maximum Likelihood (ML) Estimation. We will cover only the 1st method and refer the interested reader to pp. 257-263 of Devore (8e) for the method of ML Estimation, even if MLE has widespread applications.

THE METHOD OF MOMENTS

Consider a population with k unknown parameters $\theta_1, \theta_2, ..., \theta_k$ and a pmf or pdf $f(x; \theta_1, \theta_2, ..., \theta_k)$. Recall that the Exponential density has only one parameter (λ), the Uniform has two (a & b), the Normal has two (μ and σ), while the Gamma density has two parameters (n and λ). If the exponential represents TTF (Time-to-Failure) distribution, it may have a 2nd parameter, namely, the minimum-life. To obtain Moment estimators of k parameters, one must go through the following 3 steps:

(1) Draw a random sample of size n, denoted by $X_1, X_2, ..., X_n$, from the frequency function $f(x; \theta_1, \theta_2, ..., \theta_k)$. The sample values, after the sample has been drawn are no longer rvs, and will be denoted by small letters $x_1, x_2, ..., x_n$.

(2) Equate the 1st k origin moments of the population, μ'_r, r = 1, 2, …, k, to those of the sample origin moments m'_r, respectively, and place a hat (ˆ) on the top of all population parameters, i.e., compute E(X) and set it equal to $\bar{x}$; secondly, because $\mu'_2 = E(X^2)$, then equate $m'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ to $\mu'_2 = E(X^2)$; similarly set $\mu'_3 = E(X^3)$ equal to the 3rd sample origin moment $m_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^3$ but put hats on all the parameters in the system ok k equations with k unknowns.

(3) Then solve the above system of k equations in k unknowns obtained in step 2 simultaneously for $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$. As practice, work thru Examples 6.12- 6.14 on pp. 256-257 of Devore (8e). Note that Devore uses the notation $1/\beta = \lambda$ for the Gamma density function. Further, in the Step 2 above, the moment estimators can also be obtained by equating the population central moments of X, $\mu_k = E[(X-\mu)^k]$, to those of the sample central moments $m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$ for k = 2, 3, 4, ….

Example 34. Suppose the rv, X, is $N(\mu, \sigma^2)$, where both parameters μ and σ^2 are unknown and thus have be estimated from a random sample of size n. Since the number of unknown parameters is k = 2 and the 1st two origin moments of any population are $\mu'_1 = E(X) = \mu$ and $\mu'_2 = E(X^2)$, then by definition the 1st two origin moments of the sample are $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = m'_1$ and $\mu'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$. These imply that $\hat{\mu} = \bar{x}$ and $\mu'_2 = E(X^2)$ must be set equal to $m'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$. However, $\sigma^2 = \mu_2 = \mu'_2 - (\mu'_1)^2$

implies that $\hat{\sigma}^2 = m_2' - (\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right] = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i\right] =$

$\frac{1}{n}\left[\sum_{i=1}^{n} x_i(x_i - \bar{x})\right] = \frac{1}{n}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] = m_2$. Note that the moment estimator of μ is unbiased because the $E(\bar{x}) = \mu$, while the moment estimator of σ^2 is biased because $E(\hat{\sigma}^2) = E\left(\frac{1}{n}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]\right) = E\left[\frac{(n-1)S^2}{n}\right] = (n - 1)\sigma^2/n$, which shows that the amount of bias in $\hat{\sigma}^2 = \frac{1}{n}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] = m_2$ is equal to $-\sigma^2/n$.

<u>Exercise 48.</u> (a) Work Exercises 22(a) on page 264 of Devore (8e). (b) Suppose X is distributed as U(a, b), obtain the moment estimators of the parameters a and b if a random sample of n gave $\bar{x} = 50.00$ and $\hat{\sigma}^2 = m_2' - (m_1')^2 = 5.00$. Hint: For part (b), first use the fact that $m_1' \equiv \bar{x}$ and $m_2' \equiv \frac{1}{n}\sum_{i=1}^{n} x_i^2$ in order to obtain the value of the sample 2nd origin moment $m_2' \equiv \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 2505$. Or more directly, set $E(X) = (a+b)/2$ equal to $\bar{x}$ and $\mu_2 = (b-a)^2/12$ equal $m_2 = m_2' - (m_1')^2$.

SAMPLING DISTRIBUTIONS (SMDs) of STATISTICS From NORMAL PARENT POPULATIONS

Throughout this section we will be sampling a normal universe and will consider statistics from the sample that will have either a normal, a Chi-square, a Student's t, or the Fisher's F distribution. Due to lack of time, we will defer the Fisher's F distribution to STAT 3610.

(i) Statistics with the Normal Distribution

Suppose $X \sim N(\mu, \sigma^2)$ and a random sample of size n has been drawn with sample values $x_1, x_2, \ldots, x_n$. Since the statistic $\overline{x} = (\sum_{i=1}^{n} x_i)/n$ is a LC of NID rvs, then $V(\overline{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_x^2$ $= \frac{n\sigma_x^2}{n^2}$ so that $\overline{x}$ is also $N(\mu, \sigma_x^2/n)$.

(ii) Statistics with the Chi-Square Distribution (χ^2)

A random variable, W, is said to follow a chi-square (χ^2) distribution with ν degrees of freedom (*df*) iff its pdf at point w is given by

$$f(w) = C w^{(\frac{\nu}{2}-1)} e^{-w/2}, \qquad 0 \le w < \infty .$$

Since $\int_0^{\infty} f(w)\,dw$ must equal to 1(or 100% probability), it can be shown the sole value of the normalizing constant C that makes f(w) a pdf is $C = [2^{\nu/2}\Gamma(\nu/2)]^{-1}$, where Γ represents the gamma function. Note that the χ^2 distribution is a special case of the Gamma density function with $n = \nu/2$ and Poisson occurrence rate-parameter of $\lambda = 1/2$.

Exercise 49. (a) Show that the modal point of a χ^2 distribution occurs at $\nu - 2$ iff $\nu > 2$. (b) Show that $E(\chi_\nu^2) = \nu$ and $V(\chi^2) = 2\nu$.

The graph of $f(\chi^2)$ with $\nu = 6$ df is given in Figure 7. The corresponding pdf is given by $f(\chi_6^2 = w) = (1/16)w^2e^{-w/2}$, $0 \le w < \infty$. Table A.7 on Appendix page A-11 provides the percentage points of χ^2 for *df* from 1 to 40. For example, from Table A.7, the $P(\chi_6^2 \ge 12.592) = 0.05$. Therefore, 12.592 is the 5 percentage point (or the 95th percentile) of χ^2 with 6 *df*, i.e., $12.592 = \chi^2_{0.05,6}$ and similarly $\chi^2_{0.90,6} = 2.204$, i.e., the 10^{th}

percentile of χ_6^2 is equal to 2.204, which is also called the 90th percentage point of χ_6^2, i.e., the Pr($\chi_6^2 \geq$ 2.204) = 0.90.

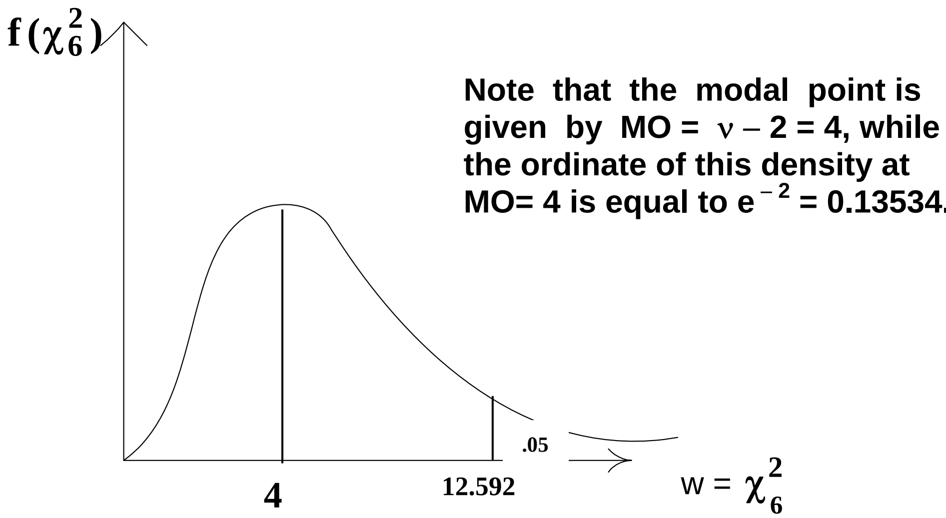

Figure 7. The pdf of χ_6^2

Exercise 50. Obtain the 5 and 2.5 percentage points of χ^2 with 1, 2 and 3 *df* from Table A.7 on page A-11. (b) Graph the pdfs of χ^2 with ν = 1, 2, and 3 *df*. Recall that $\Gamma(1/2) = \sqrt{\pi}$ and Γ (n + 1) = nΓ(n). (c) Verify the 5 percentage point of χ_2^2 by integrating the corresponding pdf.

THE ADDITIVE PROPERTY OF χ^2

Suppose Z ~ N(0,1). Then it can be proven that Z^2 follows a χ^2 distribution with 1 *df*. For example, the Pr ($|Z|$ > 1.959964) = 0.05, or Pr ($|Z|^2$ > 1.959964^2) =

$P(Z^2 > 3.84146) = 0.05$. Table A.7 verifies that indeed $3.84146 = \chi^2_{0.05,1}$ because Z^2 follows a χ^2 distribution with 1 *df*, although Devore has an error in the 3rd decimal.

Now, let $Z_1, Z_2 \ldots, Z_n$ be NID (0, 1). Then $\sum_{i=1}^{n} Z_i^2$ follows a χ^2 distribution with n *df*. Since by definition, *df* is the number of independent rvs and $\sum_{i=1}^{n} Z_i^2$ contains n independent rvs, then $\sum_{i=1}^{n} Z_i^2$ has exactly n *df* (not n – 1 as sometimes misunderstood).

Example 35. A random sample of size n = 2 is drawn from a population with pdf $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. (a) Compute the Pr ($X_1^2 + X_2^2 \le 1$).

Solution. Since X_1 and X_2 are NID (0, 1), then the sampling distribution of $X_1^2 + X_2^2$ follows a χ^2 with 2 *df*. As a result the Pr ($X_1^2 + X_2^2 \le 1$) = $P(\chi_2^2 \le 1)$ = $\int_0^1 (1/2) e^{-w/2}\, dw = \left[-e^{-w/2} \right]_0^1$ = 0.39347. (b) We now verify the 10th percentage point of χ_2^2 , whose value (to 3 decimals) is listed in Table A.7 on p. A-11, to be 4.605. We must require that $\int_0^{\chi^2_{0.10,2}} (1/2)\, e^{-w/2}\, dw$ = 0.90. By completing this integration, we obtain $\chi^2_{0.10,2}$ = 4.6052.

Exercise 51. (a) Verify the 5 percentage point of χ_4^2. (b) A random sample of size n = 11 is drawn from a N(100, 36). Compute the Pr [$\sum_{i=1}^{11} (x_i - \mu)^2 > 164.70$]. (c)

Compute the Pr [$\sum_{i=1}^{11}(x_i - 100)^2$ > 708.30]. (d) Determine α such that the P (12.549 ≤ χ^2_{10} ≤ $\chi^2_{\alpha,10}$) = 0.20. ANS: (b) 0.95, (c) 0.05.

THE SMD OF S^2 FROM A NORMAL UNIVERSE

Recall that the sampling distribution (SMD) of $\bar{\mathbf{x}}$ from a N(μ, σ^2) universe is normal with E($\bar{x}$) = μ and V($\bar{x}$) = σ^2/n. First, the sampling distribution of S^2 is not normal but by some algebraic manipulation, shown below, its sampling distribution when multiplied by the scaling factor of (n –1)/σ^2 follows a χ^2 with (n –1) *df* (not n *df*). Clearly,

$$(n-1)S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 = CSS = \sum_{i=1}^{n}\left[(x_i - \mu) - (\bar{x} - \mu)\right]^2 = [\sum(x_i - \mu)^2] - n(\bar{x} - \mu)^2.$$

$$\text{Thus, } (n-1)S^2/\sigma^2 = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 - \frac{(\bar{x} - \mu)^2}{\sigma^2 / n} \sim \chi^2_n - \chi^2_1 \sim \chi^2_{df=n-1}.$$

The above developments clearly show that the sampling distribution of the statistic (n –1)S^2/σ^2 = CSS/σ^2 follows a χ^2 with ν = (n –1) *df*.

<u>Exercise 51 (con't)</u>. (e) Compute the Pr ($S^2 \geq 57.5532$), where n = 11 and σ^2 = 36. (f) Compute the Pr [$\sum_{i=1}^{11}(x_i - \bar{x})^2 \leq 737.388$]. (g) Compute the Pr ($\sum_{i=1}^{11} x_i^2 \leq$ 107600.612) given that the value of $\bar{X}$ was computed to be 98.6.

Finally, it can be shown (thru integration) that the 1st four moments of χ^2_ν are

$\mu = \mu'_1 = \nu$, $\mu_2 = \sigma^2 = 2\nu$, $\mu_3 = 8\mu = 8\nu$, and $\mu_4 = 48\mu + 12\nu^2$. Hence the skewness $\alpha_3 = \sqrt{8/\nu}$, $\alpha_4 =$ 3 + (12/ν), and the kurtosis is equal to β_4 = 12/ν. These 1st four moments show that as $\nu \to \infty$, the distribution of χ^2_ν approaches a normal curve with mean ν and variance 2ν. For $\nu > 100$, a better approximation is that the distribution of $\sqrt{2\chi^2_\nu}$ approaches a N($\sqrt{2\nu-1}$, 1). However, my experience shows that this last approximation is fairly accurate for $\nu > 100$ only if we use a mean of $\sqrt{2\nu-(1/2)}$ for E($\sqrt{2\chi^2_\nu}$) at the tails, where $0 < \alpha \le 0.05$, instead of the recommended mean of $\sqrt{2\nu-1}$ in statistical literature. From Matlab we obtain the exact value of $\chi^2_{0.05,100}$ = 124.3421134, and this author has verified that at $\nu = 100$ *df*, the estimation error using a mean of $\sqrt{200-1}$ is equal to –0.285826, while if we use my recommended mean of $\sqrt{200-1/2}$ the estimation error is equal to –0.00669398.

Exercise 52. (a) Use the properties of χ^2 to show that if X is N (μ, σ^2), then $E(S^2) = \sigma^2$ and $V(S^2) = 2\sigma^4/(n-1)$. (b) Further, show that for a normal universe E(S) = $c_4\sigma$, where $c_4 = \sqrt{\frac{2}{n-1}} \times \Gamma(n/2)/\Gamma(\frac{n-1}{2})$ lies within the open interval (0, 1). As a result, for a Laplace-Gaussian process $V(S) = E(S^2) - [E(S)]^2 = \sigma^2 - (c_4\sigma)^2 = (1 - c_4^2)\sigma^2$, $0 < c_4 < 1$ and for a positive integer $\Gamma(n) = (n-1)!$.

(iii) THE (W. S. Gosset's) STUDENT'S t-DISTRIBUTION

A continuous rv, T, has the (W. S. Gosset's) Student's t-distribution with ν *df* iff its pdf is given by

$$f(t) = C(1 + t^2/\nu)^{-(\nu+1)/2}, \qquad -\infty < t < \infty.$$

Again the normalizing constant C has to be evaluated in such manner that $\int_{-\infty}^{\infty} f(t)\,dt = 1$.

This requirement leads to $C = \Gamma(\frac{\nu+1}{2})/[\Gamma(\nu/2)\sqrt{\pi\nu}\,]$. It can be shown that $E(T_\nu) = 0$ for all $\nu > 1$, $V(T_\nu) = \nu/(\nu-2)$ for all $\nu > 2$, $\mu_3 = 0$ so that the skewness is zero for all $\nu > 3$, and μ_4 = $E[(T_\nu)^4] = \nu^2 \Gamma(5/2)\Gamma(\frac{\nu}{2}-2)/[\Gamma(1/2)\Gamma(\nu/2)]$ for all $\nu > 4$. Note that the origin moments μ'_r of the t distribution exist only for $\nu > r$. The skewness $\alpha_3 = 0$ for all $\nu > 3$ and the kurtosis $\beta_4 = \alpha_4 - 3 = 3(\nu - 2)/(\nu - 4) - 3 = \frac{6}{\nu-4}$ for all $\nu > 4$. These moments show that the sampling distribution of a T_ν rv approaches normality with mean 0 and variance 1 as $\nu \to \infty$ because the skewness and kurtosis of a unit (or standard) normal rv are both zero. The t-distribution is always symmetrical about the median value of 0 for all $\nu \geq 1$. The graph of f(t) is given in Figure 8 for a given df $\nu \geq 1$. The median = the mean for all $\nu >$ 1. Further, the inflection points of T_ν occur at $\pm\sqrt{\nu/(\nu+2)}$.

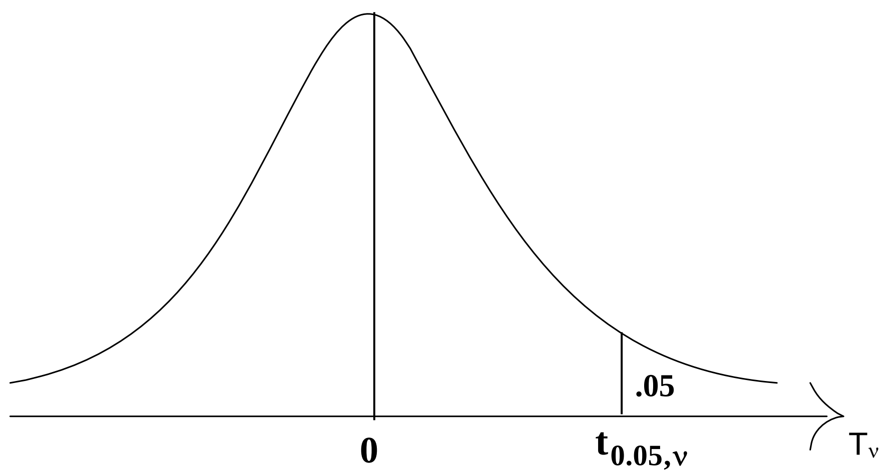

Figure 8. The Graph of the (W. S. Gosset) Student's t-pdf

It should be clear that by symmetry of all t distributions, $t_{1-\alpha,\nu} = -t_{\alpha,\nu}$. Table A.5 on p. A-9 of Devore's 8th edition gives the percentage points of the t distribution for $\nu = 1$ to $\nu =$ 120 *df*. This table shows that the 5 percentage point of the Student's t distribution with 10 *df* is equal to 1.812, i.e., $t_{0.05,10} = 1.812$, which implies that the $\Pr(|T_{10}| \geq 1.812) = 0.10$ due to symmetry. Further, from the Table A.5 we observe that $t_{0.025,30} = 2.042$, i.e., the $\Pr(T_{30} \leq 2.042) = 0.975$, and that $t_{0.975,30} = -2.042$, and so on.

It can be shown that a Student's t rv can be generated thru the ratio $Z/\sqrt{\chi^2_\nu/\nu}$,

where Z ~ N (0, 1) and the rv χ^2_ν has a chi-square distribution with ν *df* and independent of Z. To illustrate the application of the t-distribution to sampling statistics, let X be any N (μ, unknown σ^2) rv and consider a random sample of size n from this normal population with unknown mean and unknown variance. Then, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ = CSS/(n – 1) is a point unbiased estimator of σ^2 with ν = n – 1 *df* [recall the constraint $\sum_{i=1}^{n}(x_i - \bar{x}) \equiv 0$].

From the sampling distribution of $\bar{x}$, it follows that $(\bar{x} - \mu)/\sigma_{\bar{x}} = \frac{\bar{x}-\mu}{SE(\bar{x})} = (\bar{x}-\mu)\sqrt{n}/\sigma$ = Z ~ N (0, 1). Recall also that $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ and as a result

$$T_\nu = \frac{Z\sqrt{\nu}}{\sqrt{\chi^2_\nu}} = \frac{[(\bar{x}-\mu\ )\sqrt{n}\ /\sigma]\sqrt{n-1}}{\sqrt{(n-1)S^2/\sigma^2}} = \frac{(\bar{x}-\mu)\sqrt{n}}{S} = \frac{\bar{x}-\mu}{\boldsymbol{se}(\bar{x})}$$

has a Gosset's Student's t-distribution with ν = n –1 *df*.

<u>Bonus HW(5 Points).</u> (a) Suppose X ~ N (μ, σ^2). Determine the sampling distributions of (X –μ)/σ and (X –μ)/S_x .

<u>Example 35.</u> A random sample of size n = 16 is drawn from a N (μ, σ^2) universe. Compute the P ($|\bar{x} - \mu|$ < 1.268) given that the standard deviation for the sample was computed to be 2.380.

<u>Solution</u>. $\Pr(|\bar{x}-\mu| \le 1.268) = \Pr(|\text{Estimation Error}| \le 1.268)$

$$\Pr(-1.268 \le \bar{x}-\mu \le 1.268) = P[\frac{-1.268\sqrt{16}}{2.38} \le \frac{(\bar{x}-\mu)\sqrt{n}}{S} \le \frac{1.268\times 4}{2.38}] =$$

$P[\frac{-1.268\sqrt{16}}{2.38} \le \text{Studentized Error} \le \frac{1.268 \times 4}{2.38}] = P(-2.131 \le T_{15} \le 2.131) =$ 0.95 (see Table A.5 on page A-9 of Devore's 8th edition).

Exercise 53. A random sample of size n = 20 is obtained from a N(μ, σ^2) universe and S^2 was computed to be 0.42014. Compute the Pr that the absolute error in estimation, given by $|\bar{x} - \mu|$, exceeds 0.368. ANS: 0.02.

Exercise 54. A normal population has known mean μ = 10 and unknown variance σ^2. A random sample of size n = 25 is selected from this population resulting in the sample mean of 11.30 and standard deviation of S = 2.6084. How unusual are these sample results? That is, what is the occurrence Pr of the event ($\bar{\mathbf{x}} \ge 11.30$)? ANS: 0.01.

Bonus HW Continued (5 Points). (b) A random sample of size n_x from a N (μ_x, σ^2) universe and one of size n_y from another independent N (μ_y, σ^2) population are drawn, where the two populations have common equal variance σ^2. Show that the random variable

$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} ,$$

where the pooled estimator of σ^2 is given by

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

has a sampling distribution which is the Student's-t with $\nu = n_x + n_y - 2$ *df*.

The only distribution left to study for statistical inference is that of (SIR Ronald A.) FISHER'S F which describes the sampling distribution of $\frac{S_x^2 / \sigma_x^2}{S_y^2 / \sigma_y^2}$ with numerator *df* of $n_x - 1$ and denominator *df* of $n_y - 1$. Confidence interval estimation, which is a branch of statistical inference, will begin in chapter 7.