BST 675 – Fall 2011 – Dr. Charnigo

Unit V: Sampling Distributions

a. Motivating case study #1: Where do the familiar hypothesis test and confidence interval for a population mean come from?

Beginning with Unit V of BST 675 and continuing into BST 676 next semester, you will notice a subtle shift in emphasis. Instead of asking questions like "What is the probability that $\mathbf{X}$ will do such and such if $\mathbf{X}$ is governed by a known parameter μ?", we will be asking questions like "What is our best guess for an unknown parameter μ governing $\mathbf{X}$ if we observe $\mathbf{X}$ do such and such?"

So, with that in mind, let us recall a situation considered in your introductory methods course. For some $n \in \{2, 3, \ldots\}$, we observe $\mathbf{X} := (X_1, X_2, \ldots, X_n)^T$ with $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where "iid" is shorthand for "independently and identically distributed" and $N(\mu, \sigma^2)$ is shorthand for "normal with mean μ and variance σ^2". Since we speak of a mean μ and a variance σ^2 rather than a mean vector and a covariance matrix, the context makes clear that $N(\mu, \sigma^2)$ represents the marginal distribution of X_1 — yet not only X_1 but also X_2 and X_3, etc.

The problem, of course, is that neither μ nor σ^2 is known. Can we use the observed $\mathbf{X}$ to make good guesses (or, in the formal language of statisticians, "inferences") about μ and σ^2?

In fact, we can, and you learned how to do so in your introductory methods course. Let

$$\bar{X} := \sum_{i=1}^{n} X_i/n \quad \text{and} \quad S^2 := \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1),$$

quantities instantly recognizable to you as the "sample mean" and the "sample variance".

As an aside, note that $\bar{X}$ and S^2 are actually random variables. While completely obvious now (if for no other reason than that I used capital letters), this may have created some confusion in your introductory methods course, since there the sample mean and sample variance were initially introduced to you as numbers calculated from data sets rather than as random variables. When I teach STA 580, I try to make the distinction explicit by referring to $\bar{X}$ and S^2 as "random conceptualizations" of the sample mean and sample variance, at least until students become comfortable with the idea that a number calculated from a data set is a realization of a random variable.

Returning to the question of how to make a good guess for μ, a natural choice is $\bar{X}$. First, $\bar{X}$ is an unbiased estimator of μ in the sense that the expected value of the estimator equals the target, $E[\bar{X}] = \mu$. Second, $\bar{X}$ does not have unnecessarily large variance. In fact, you can now easily verify that $Var[\bar{X}] = \sigma^2/n$, a fact you might have been asked to take for granted in your introductory methods course. In contrast, X_1 is also an unbiased estimator of μ but one whose variance is unnecessarily large, since $\sigma^2 > \sigma^2/n$.

Yet, making a good guess for μ entails more than observing $\bar{X}$. In particular, you may want to construct a 95% confidence interval for μ. The formula

$$\bar{X} \pm t_{n-1,0.975}\sqrt{S^2/n} \tag{1}$$

is undoubtedly familiar, but where does it come from?

You might have been told in your introductory methods course that the quantity

$$(\bar{X} - \mu)/\sqrt{S^2/n} \tag{2}$$

has the T distribution on $(n-1)$ degrees of freedom — something symmetric about 0 and vaguely bell-shaped but with thicker tails than a standard normal distribution — and that upper quantiles of T distributions — that is, numbers of the form $t_{n-1,1-\alpha/2}$ for $\alpha \in \{0.01, 0.05, 0.10\}$ and $n \in \{2, 3, \ldots, 31\}$ — are tabulated in the back of a textbook. As such,

$$\begin{aligned}
& P\left(-t_{n-1,1-\alpha/2} \le (\bar{X} - \mu)/\sqrt{S^2/n} \le t_{n-1,1-\alpha/2}\right) \\
= \; & P\left(-t_{n-1,1-\alpha/2}\sqrt{S^2/n} \le \bar{X} - \mu \le t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right) \\
= \; & P\left(-\bar{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n} \le -\mu \le -\bar{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right) \\
= \; & P\left(\bar{X} - t_{n-1,1-\alpha/2}\sqrt{S^2/n} \le \mu \le \bar{X} + t_{n-1,1-\alpha/2}\sqrt{S^2/n}\right),
\end{aligned}$$

from which formula (1) is immediate.

You may also want to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ for some real number μ_0 (perhaps 0). In this case, you can define a test statistic

$$(\bar{X} - \mu_0)/\sqrt{S^2/n} \tag{3}$$

and reject the null hypothesis at significance level 0.05 if the absolute value of the test statistic exceeds $t_{n-1,0.975}$.

This testing procedure "works" for two reasons. First, if H_0 is true, then (3) coincides with (2), so the probability of incorrectly rejecting H_0 is indeed 0.05. Second, if H_0 is false, then (3) does not coincide with (2), and the probability of correctly rejecting H_0 is greater than 0.05. Hence, rejection of H_0 is more likely when H_0 is false, an eminently reasonable requirement for a testing procedure referred to as "unbiasedness" (not to be confused with "unbiasedness" of an estimator). Students in introductory methods courses tend to "get" the first point but not the second point. In particular, stating that (3) has a T distribution on $(n-1)$ degrees of freedom is actually not correct, unless one qualifies the statement with an assumption that H_0 is true!

So, what remains to be resolved in this motivating case study? We have not actually stated what a T distribution is, much less established that (2) abides a T distribution. We will attend to these items later in Unit V.

b. Motivating case study #2: Where do the familiar (?) hypothesis test and confidence interval for a population variance come from?

Continuing from the first motivating case study, what if we also want to make a guess for σ^2? A natural choice may seem to be

$$\widehat{\sigma}^2 := \sum_{i=1}^{n}(X_i - \bar{X})^2/n, \tag{4}$$

and indeed $\widehat{\sigma}^2$ is the so-called maximum likelihood estimator of σ^2 when $X_1, \ldots, X_n$ are normally distributed. However, many people use S^2 instead of

$\widehat{\sigma}^2$. Of course, for large n, there is little difference between S^2 and $\widehat{\sigma}^2$. However, for small n, the difference is noticeable.

Many people prefer S^2 because of its unbiasedness, $E[S^2] = \sigma^2$. Interestingly, the unbiasedness of S^2 is a general property, not one that requires normality of $X_1, \ldots, X_n$. Indeed, writing out $X_i - \bar{X} = \sum_{j=1}^n c_{ij} X_j$ with $c_{ij} = 1_{\{i=j\}} - 1/n$, we have

$$(X_i - \bar{X})^2 = \left(\sum_{j=1}^n c_{ij} X_j \right)^2 = \sum_{j=1}^n c_{ij} X_j \sum_{k=1}^n c_{ik} X_k = \sum_{j=1}^n \sum_{k=1}^n c_{ij} c_{ik} X_j X_k.$$

If $j = k$, then

$$E[X_j X_k] = E[X_j^2] = Var[X_j] + (E[X_j])^2 = \sigma^2 + \mu^2,$$

while if $j \neq k$, then

$$E[X_j X_k] = E[X_j] E[X_k] = \mu^2.$$

As such,

$$\begin{aligned}
E[(X_i - \bar{X})^2] &= \sum_{j=1}^n \sum_{k=1}^n c_{ij} c_{ik} E[X_j X_k] \\
&= \sum_{j=1}^n \sum_{k=j} c_{ij} c_{ik} (\sigma^2 + \mu^2) + \sum_{j=1}^n \sum_{k \neq j} c_{ij} c_{ik} \mu^2 \\
&= \sigma^2 \sum_{j=1}^n \sum_{k=j} c_{ij} c_{ik} + \mu^2 \sum_{j=1}^n \sum_{k=1}^n c_{ij} c_{ik} \\
&= \sigma^2 \sum_{j=1}^n (c_{ij})^2 + \mu^2 \sum_{j=1}^n c_{ij} \sum_{k=1}^n c_{ik} \\
&= \sigma^2 \left[\sum_{j=i} (c_{ij})^2 + \sum_{j \neq i} (c_{ij})^2 \right] + \mu^2 \times 0 \times 0 \\
&= \sigma^2 [(1 - 1/n)^2 + (n-1)(-1/n)^2] \\
&= \sigma^2 (n-1)/n,
\end{aligned}$$

from which unbiasedness of S^2 follows since

$$E[S^2] = (n-1)^{-1} \sum_{i=1}^n E[(X_i - \bar{X})^2] = (n-1)^{-1} n \sigma^2 (n-1)/n = \sigma^2.$$

What if we want to create a 95% confidence interval for σ^2? A naive attempt by a student in an introductory methods course is

$$S^2 \pm t_{n-1,0.975} \text{ something },$$

but in fact the way to go is

$$\left[(n-1)S^2/\chi^2_{n-1,0.975}, (n-1)S^2/\chi^2_{n-1,0.025}\right]. \tag{5}$$

Formula (5) is different from formula (1) in that S^2 is adjusted multiplicatively rather than additively to obtain the confidence limits. Students often confuse which quantile — $\chi^2_{n-1,0.975}$ or $\chi^2_{n-1,0.025}$ — belongs to which confidence limit — lower or upper. Can you think of an obvious way to help a student remember?

Of course, we are interested in where (5) comes from. I can tell you to take as a starting point that $(n-1)S^2/\sigma^2$ has the chi-square distribution on $(n-1)$ degrees of freedom, which permits you to mimic the logic at the bottom of page 2. However, that is not a complete answer. While you do know from your study of gamma distributions what the chi-square distribution is, you do not yet have any proof that $(n-1)S^2/\sigma^2$ has the chi-square distribution on $(n-1)$ degrees of freedom. We will attend to that later in Unit V.

Before leaving this motivating case study, let us document the corresponding testing procedure. With $H_0 : \sigma^2 = \sigma_0^2$ and $H_1 : \sigma^2 \neq \sigma_0^2$, we reject H_0 at level 0.05 if either

$$(n-1)S^2/\sigma_0^2 > \chi^2_{n-1,0.975} \quad \text{or} \quad (n-1)S^2/\sigma_0^2 < \chi^2_{n-1,0.025}.$$

The former contingency suggests that S^2 is too large for us to believe that $\sigma^2 = \sigma_0^2$, while the latter contingency suggests that S^2 is too small for us to believe that $\sigma^2 = \sigma_0^2$.

Finally, a similar caveat applies to this testing procedure as to the testing procedure on page 3, namely that the test statistic follows the indicated reference distribution only if H_0 is true.

c. Random samples and linear combinations

Consider an effectively infinite population whose individuals are described by some quantitative characteristic.

Suppose there exists a continuous function f such that, for any real number x, the proportion of individuals in the population with characteristic less than or equal to x is given by $\int_{-\infty}^{x} f(t)\, dt$.

Suppose also that we randomly select n individuals from the population, in such a way that any two groups of n individuals have the same probability of being selected. In particular, every individual in the population has the same probability of being selected. This is called a simple random sample.

Let the random variables $X_1, \ldots, X_n$ denote the characteristics of the n individuals in the simple random sample. Then $X_1, \ldots, X_n$ will be iid with common marginal probability density function f. In particular, X_1 will have marginal probability density function f. Therefore, statements about proportions in the whole population can be cast as statements about probabilities for a single individual randomly selected from that population.

We often employ this interpretation in methods courses. For example, whenever we say that a population is normal, what we mean is that there exist numbers $\mu \in \mathbb{R}, \sigma \in (0, \infty)$ such that $\int_{-\infty}^{x} (2\pi\sigma^2)^{-1/2} \exp[-(t-\mu)^2/(2\sigma^2)]\, dt$ gives the proportion of individuals in the population with characteristic less than or equal to x. However, we usually represent a normal population to undergraduates as one for which the random variables $X_1, \ldots, X_n$ arising from a simple random sample are normally distributed. Then we draw a picture of a bell curve to describe the probabilistic behavior of X_1.

Consider the following practice problems regarding a simple random sample governed by probability density function $f(x; \theta) := \theta x^{\theta-1} 1_{\{0<x<1\}}$ and cumulative distribution function $F(x; \theta) := x^{\theta} 1_{\{0<x<1\}} + 1_{\{x \geq 1\}}$, where $\theta \in (0, \infty)$.

1. Fix $b \in (0, 1)$. What is the probability that all of $X_1, \ldots, X_n$ exceed b?

2. What is the probability that none of $X_1, \ldots, X_n$ exceed b?

3. What is the probability that exactly one of $X_1, \ldots, X_n$ exceeds b?

4. What is the probability that exactly m $(0 < m < n)$ of $X_1, \ldots, X_n$ exceed b?

5. Now suppose that $f(x;\theta)$ and $F(x;\theta)$ had not been specified explicitly. Can you still come up with answers to questions 1 through 4? Also note that, with a little massaging, the answers to questions 1 and 2 tell you the distributions of the minimum and maximum observations in a simple random sample.

Suppose we are confronted with the problem of deriving the probabilistic behavior of a linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ when $X_1, \ldots, X_n$ are iid with known marginal distribution and $c_1, \ldots, c_n$ are real constants. Since we can easily determine the probabilistic behavior of $V_i := c_iX_i$ $(1 \le i \le n)$ given the probabilistic behavior of X_i, the problem simplifies to deriving the probabilistic behavior of $V_1 + V_2 + \cdots + V_n$ when $V_1, \ldots, V_n$ are independent (though not necessarily identically distributed, in case not all of the c_i are the same) with known marginal distributions.

We can sometimes solve the problem by appealing to moment generating functions (Cf. Written Assignment 5, BST 675, Fall 2010). For example, a sum of independent normal random variables is normal. However, appealing to moment generating functions does not always work. In particular, we become stuck if the moment generating function of $V_1+V_2+\cdots+V_n$ is not recognizable or, worse yet, $V_1,\ldots,V_n$ do not have finite moment generating functions in a neighborhood of 0.

The Central Limit Theorem (to be studied in BST 676) and its generalizations can be used, in many cases, to approximate the distribution of $V_1+V_2+\cdots+V_n$ for medium to large n. But what if n is small or, as is the case for $V_1,\ldots,V_n \overset{iid}{\sim} \pi^{-1}/(1+v^2)$, the Central Limit Theorem does not apply? Then we have one more avenue at our disposal, the so-called convolution formula.

For integer-valued discrete random variables the convolution formula is

$$f_{V_1+V_2}(v)=\sum_{w\in\mathbb{Z}} f_{V_1}(w)f_{V_2}(v-w),$$

while for continuous random variables the convolution formula is

$$f_{V_1+V_2}(v)=\int_{\mathbb{R}} f_{V_1}(w)f_{V_2}(v-w)\ dw.$$

The convolution formula can be derived in the continuous setting using the bivariate transformation formula; this is left to you as an exercise. The convolution formula can be derived in the discrete setting using the law of total probability with $A:=\{V_1+V_2=v\}$ and $B_w:=\{V_1=w\}$ for $w\in\mathbb{Z}$,

$$P(A)=\sum_{w\in\mathbb{Z}} P(A|B_w)P(B_w)$$

or

$$\begin{aligned}P(V_1+V_2=v) &= \sum_{w\in\mathbb{Z}} P(V_1+V_2=v\mid V_1=w)P(V_1=w)\\ &= \sum_{w\in\mathbb{Z}} P(V_2=v-w\mid V_1=w)P(V_1=w)\\ &= \sum_{w\in\mathbb{Z}} P(V_2=v-w)P(V_1=w).\end{aligned}$$

The convolution formula as presented above seems to work only for $n=2$. What can be done for $n>2$?

d. Resolution of Motivating Case Study #2: the chi-square pivot for a normal population

Since resolution of the first motivating case study requires a theoretical result from the second motivating case study, we resolve the second motivating case study before returning to the first motivating case study.

Our present goal is to show that, for $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with $n \in \{2, 3, \ldots\}$,

$$(n-1)S^2/\sigma^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2 \sim \chi^2_{n-1}.$$

Since

$$\sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

for $Z_i := (X_i - \mu)/\sigma$, showing that

$$\sum_{i=1}^{n} (Z_i - \bar{Z})^2 \sim \chi^2_{n-1} \tag{6}$$

for $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0, 1)$ will suffice.

Step #1. Suppose that Y has the standard normal distribution. Then $W := Y^2$ has the chi-square distribution on one degree of freedom. To see this, let $w \in (0, \infty)$ and note that the cumulative distribution function of W is

$$\begin{aligned} P(W \le w) &= P(Y^2 \le w) \\ &= P(-\sqrt{w} \le Y \le \sqrt{w}) \\ &= 2P(0 \le Y \le \sqrt{w}) \\ &= 2\int_0^{\sqrt{w}} (2\pi)^{-1/2} \exp[-y^2/2]\, dy. \end{aligned}$$

Hence, the probability density function of W is, for $w \in (0, \infty)$,

$$2(2\pi)^{-1/2} \exp[-w/2] \frac{d}{dw}\sqrt{w} = \frac{w^{1/2-1}}{2^{1/2}\Gamma[1/2]} \exp[-w/2].$$

Above we have used the well-known fact that the gamma function evaluated at $1/2$ equals $\sqrt{\pi}$. This fact can be verified by writing $\Gamma[1/2] = \int_0^\infty t^{1/2-1} \exp[-t]\, dt$, making the substitution $u := \sqrt{2t}$, and then evaluating the resulting integral using the kernel method.

Step #2. Suppose that Y_1 has the chi-square distribution on k degrees of freedom (k a positive integer) and that Y_2, independent of Y_1, has the chi-square distribution on 1 degree of freedom. Then $Y_1 + Y_2$ has the chi-square distribution on $(k+1)$ degrees of freedom, as can be verified using moment generating functions (Cf. Final Examination, BST 675, Fall 2010).

Step #3. Consider what happens when $n = 2$. We have

$$\begin{aligned}
& \left(Z_1 - 0.5\sum_{i=1}^{2} Z_i\right)^2 + \left(Z_2 - 0.5\sum_{i=1}^{2} Z_i\right)^2 \\
= \; & (0.5Z_1 - 0.5Z_2)^2 + (0.5Z_2 - 0.5Z_1)^2 \\
= \; & 0.25(Z_1 - Z_2)^2 + 0.25(Z_1 - Z_2)^2 \\
= \; & 0.5(Z_1 - Z_2)^2 \\
= \; & \left(\frac{Z_1 - Z_2}{\sqrt{2}}\right)^2 .
\end{aligned}$$

Since

$$Z_1 - Z_2 \sim N(0, 2),$$

we have

$$(Z_1 - Z_2)/\sqrt{2} \sim N(0, 1)$$

and therefore, by Step #1,

$$\left(\frac{Z_1 - Z_2}{\sqrt{2}}\right)^2 = \left(Z_1 - 0.5\sum_{i=1}^{2} Z_i\right)^2 + \left(Z_2 - 0.5\sum_{i=1}^{2} Z_i\right)^2 \sim \chi_1^2.$$

Step #4. Suppose that, for a positive integer $k \geq 2$, the relation (6) is true when $n = k$. We will now show that, under this supposition, (6) is also true when $n = (k+1)$. Since Step #3 verified that (6) is true when $n = 2$, this will imply (6) for arbitrary n. (This reasoning is called "mathematical induction".)

Let us introduce the shorthand

$$\bar{Z}_k := k^{-1}\sum_{i=1}^{k} Z_i \quad \text{and} \quad \bar{Z}_{k+1} := (k+1)^{-1}\sum_{i=1}^{k+1} Z_i = \frac{k\bar{Z}_k + Z_{k+1}}{(k+1)}.$$

We have

$$\begin{aligned}
&\sum_{i=1}^{k+1}\left(Z_i - \bar{Z}_{k+1}\right)^2 \\
=\ &\sum_{i=1}^{k}\left(Z_i - \bar{Z}_{k+1}\right)^2 + \left(Z_{k+1} - \bar{Z}_{k+1}\right)^2 \\
=\ &\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k + \bar{Z}_k - \bar{Z}_{k+1}\right)^2 + \left(\frac{kZ_{k+1} - k\bar{Z}_k}{k+1}\right)^2 \\
\overset{why?}{=}\ &\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2 + k\left(\bar{Z}_k - \bar{Z}_{k+1}\right)^2 + \left(\frac{kZ_{k+1} - k\bar{Z}_k}{k+1}\right)^2 \\
=\ &\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2 + k\left(\bar{Z}_k - \frac{k\bar{Z}_k + Z_{k+1}}{k+1}\right)^2 + \left(\frac{k}{k+1}\right)^2\left(Z_{k+1} - \bar{Z}_k\right)^2 \\
=\ &\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2 + k\left(\frac{Z_{k+1} - \bar{Z}_k}{k+1}\right)^2 + \left(\frac{k}{k+1}\right)^2\left(Z_{k+1} - \bar{Z}_k\right)^2 \\
=\ &\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2 + \frac{k}{k+1}\left(Z_{k+1} - \bar{Z}_k\right)^2.
\end{aligned}$$

Recall that $\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2 \sim \chi^2_{k-1}$ by the induction supposition. Step #2 will imply that $\sum_{i=1}^{k+1}\left(Z_i - \bar{Z}_{k+1}\right)^2 \sim \chi^2_k$ once we confirm that:

$$\frac{k}{k+1}\left(Z_{k+1} - \bar{Z}_k\right)^2 \sim \chi^2_1 \tag{7}$$

and

$$\frac{k}{k+1}\left(Z_{k+1} - \bar{Z}_k\right)^2 \perp \sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2, \tag{8}$$

where the $\perp$ symbol is shorthand for independence.

Confirmation of (7) is based on the fact that

$$Z_{k+1} - \bar{Z}_k \sim N(0, (k+1)/k)$$

and, as such,

$$\sqrt{k/(k+1)}(Z_{k+1} - \bar{Z}_k) \sim N(0,1).$$

Confirmation of (8) is based on manipulation of the joint probability density function of $Z_1, Z_2, Z_3, \ldots, Z_k, Z_{k+1}$. Employing the multivariate analogue of the bivariate transformation formula, one can manipulate the joint probability density function of $Z_1, Z_2, Z_3, \ldots, Z_k, Z_{k+1}$ to find the joint probability density function of

$$U_1 := \bar{Z}_k, U_2 := Z_2 - \bar{Z}_k, U_3 := Z_3 - \bar{Z}_k, \ldots, U_k := Z_k - \bar{Z}_k, U_{k+1} := Z_{k+1}.$$

(If you are up for a challenge, do so! You will need to use the fact that the determinant of a triangular matrix — that is, a square matrix with either all zeroes above the diagonal or all zeroes below the diagonal — is the product of its diagonal entries.)

The joint probability density function of $U_1, U_2, U_3, \ldots, U_k, U_{k+1}$ factors into the product of the joint probability density function of U_1, U_{k+1} with the joint probability density function of $U_2, U_3, \ldots, U_k$. As such, any function of U_1 and U_{k+1} is independent of any function of $U_2, U_3, \ldots, U_k$.

Clearly

$$\frac{k}{k+1}\left(Z_{k+1} - \bar{Z}_k\right)^2 = \frac{k}{k+1}(U_{k+1} - U_1)^2$$

is a function of U_1 and U_{k+1}. On the other hand,

$$\sum_{i=1}^{k}\left(Z_i - \bar{Z}_k\right)^2 = \left(Z_1 - \bar{Z}_k\right)^2 + U_2^2 + U_3^2 + \cdots + U_k^2.$$

We are done once we verify that this is a function of $U_2, U_3, \ldots, U_k$. This seems obvious except for the $\left(Z_1 - \bar{Z}_k\right)^2$ piece. Any suggestions?

e. Resolution of Motivating Case Study #1: the T pivot for a normal population

Our final goal for Unit V and, indeed, for the BST 675 lecture series is to show that, for $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with $n \in \{2, 3, \ldots\}$,

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim T_{n-1},$$

where T_p is shorthand for the T distribution on $p(> 0)$ degrees of freedom.

The T distribution on p degrees of freedom has probability density function

$$f(t; p) = \frac{\Gamma[(p+1)/2]}{\Gamma[p/2]} \frac{1}{\sqrt{p\pi}} \frac{1}{(1 + t^2/p)^{(p+1)/2}}.$$

A few points about this probability density function are worth noting.

First, the case $p = 1$ yields $\pi^{-1}/(1 + t^2)$, demonstrating that the Cauchy distribution with location 0 and scale 1 belongs to the family of T distributions.

Second, we have

$$\begin{aligned}
& \lim_{p \to \infty} \frac{1}{(1 + t^2/p)^{(p+1)/2}} \\
= \ & \lim_{p \to \infty} \frac{1}{(1 + t^2/p)^{p/2}} \\
= \ & \lim_{p \to \infty} \left((1 + t^2/p)^p\right)^{-1/2} \\
= \ & \left(\lim_{p \to \infty} (1 + t^2/p)^p\right)^{-1/2} \\
= \ & \left(\exp[t^2]\right)^{-1/2} \\
= \ & \exp[-t^2/2].
\end{aligned}$$

Hence, the kernel of a T distribution resembles the kernel of the standard normal distribution as the degrees of freedom become large. This provides the justification for the advice given by some textbook authors that a T distribution on more than (say) 200 degrees of freedom may be treated like the standard normal distribution.

Third, for a fixed p we have

$$f(t;p) = O\left(t^{-p-1}\right) \tag{9}$$

as $t \to \infty$, where the "big O" notation indicates approximate proportionality as $t \to \infty$. More precisely, we say that

$$a(t) = O\left(b(t)\right)$$

if there exist a threshold t_0 and a bound $C(> 0)$ such that

$$t > t_0 \quad \text{implies} \quad |a(t)/b(t)| \leq C.$$

(The choices of threshold and bound are not unique. For instance, if t_0 and C work, then so do any numbers greater than t_0 and C.) The importance of (9) is that we can see immediately how many moments of a T distribution will exist finitely, inasmuch as $\int_1^\infty t^{m-p-1}\, dt$ is finite if and only if $m < p$.

Observing that

$$\frac{\bar{X}-\mu}{\sqrt{S^2/n}} = \frac{(\bar{X}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{S^2/n}/\sqrt{\sigma^2/n}} = \frac{(\bar{X}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{[\sum_{i=1}^n (X_i-\bar{X})^2/\sigma^2]/[n-1]}},$$

I propose to accomplish our final goal for Unit V by proving the following general result: Let Z denote a standard normal random variable and W a chi-square random variable on p degrees of freedom, where Z and W are independent. Then $Z/\sqrt{W/p}$ has the T distribution on p degrees of freedom.

The general result will accomplish our final goal for Unit V because $(\bar{X}-\mu)/\sqrt{\sigma^2/n}$ has the standard normal distribution, $\sum_{i=1}^n (X_i-\bar{X})^2/\sigma^2$ has the chi-square distribution on $(n-1)$ degrees of freedom, and these two quantities are independent by an argument similar to that described on page 12 for establishing (8). Here is a brief synopsis of the argument for independence. Put $U_1 := \bar{X}$ and $U_i := X_i - \bar{X}$ for $i \in \{2, \dots, n\}$. Then the joint probability density function of $U_1, U_2, \dots, U_n$ factors into the product of the marginal probability density function of U_1 with the joint probability density function of $U_2, \dots, U_n$. Hence, any function of U_1, such as $(\bar{X}-\mu)/\sqrt{\sigma^2/n}$, is independent of any function of $U_2, \dots, U_n$, such as $\sum_{i=1}^n (X_i-\bar{X})^2/\sigma^2$.

To prove the general result, we use the bivariate transformation formula. Put

$$g_1(z,w) := z/\sqrt{w/p} \quad \text{and} \quad g_2(z,w) := w$$

for $z \in \mathbb{R}$ and $w \in (0,\infty)$. Then let

$$U := g_1(Z,W) \quad \text{and} \quad V := g_2(Z,W).$$

We have, for $z \in \mathbb{R}$ and $w \in (0,\infty)$,

$$f_{Z,W}(z,w) = (2\pi)^{-1/2}\exp[-z^2/2]\frac{1}{\Gamma[p/2]2^{p/2}}w^{p/2-1}\exp[-w/2].$$

As such,

$$\begin{aligned} S_{U,V} &= \{(u,v)^T \in \mathbb{R}^2 : \exists (z,w)^T \in \mathbb{R}^2 \text{ with } u = g_1(z,w), v = g_2(z,w), f_{Z,W}(z,w) > 0\} \\ &= \{(u,v)^T \in \mathbb{R}^2 : v > 0\}. \end{aligned}$$

Since

$$Z = U\sqrt{V/p} \quad \text{and} \quad W = V,$$

we take

$$h_1(u,v) := u\sqrt{v/p} \quad \text{and} \quad h_2(u,v) := v.$$

Then

$$Det\begin{bmatrix} \frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v} \\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v} \end{bmatrix} = Det\begin{bmatrix} \sqrt{\frac{v}{p}} & \frac{u}{2\sqrt{vp}} \\ 0 & 1 \end{bmatrix} = \sqrt{v/p}.$$

We have, for $(u,v)^T \in S_{U,V}$,

$$f_{Z,W}(h_1(u,v),h_2(u,v)) = (2\pi)^{-1/2}\exp[-u^2v/(2p)]\frac{1}{\Gamma[p/2]2^{p/2}}v^{p/2-1}\exp[-v/2]$$

and

$$f_{U,V}(u,v) = (2\pi)^{-1/2}\exp[-u^2v/(2p)]\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}v^{(p+1)/2-1}\exp[-v/2].$$

To finish, we still need to obtain the marginal probability density function of U. To do so, we integrate the joint probability density function of U and V in dv. For reasons that will become clear shortly, we put $\alpha := (p+1)/2$ and $\beta := 2/(1+u^2/p)$. We have

$$\begin{aligned}
&\int_0^\infty f_{U,V}(u,v)\;dv \\
=\;& (2\pi)^{-1/2}\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}\int_0^\infty \exp[-u^2v/(2p)]v^{(p+1)/2-1}\exp[-v/2]\;dv \\
=\;& (2\pi)^{-1/2}\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}\int_0^\infty \exp[-v(1+u^2/p)/2]v^{(p+1)/2-1}\;dv \\
=\;& (2\pi)^{-1/2}\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}\int_0^\infty \exp[-v/\beta]v^{\alpha-1}\;dv \\
=\;& (2\pi)^{-1/2}\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}\beta^\alpha\Gamma[\alpha]\int_0^\infty \frac{1}{\beta^\alpha\Gamma[\alpha]}\exp[-v/\beta]v^{\alpha-1}\;dv \\
=\;& (2\pi)^{-1/2}\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}\beta^\alpha\Gamma[\alpha] \\
=\;& (2\pi)^{-1/2}\frac{1}{\Gamma[p/2]2^{p/2}p^{1/2}}2^{(p+1)/2}\frac{1}{(1+u^2/p)^{(p+1)/2}}\Gamma[(p+1)/2] \\
=\;& \frac{\Gamma[(p+1)/2]}{\Gamma[p/2]}\frac{1}{\sqrt{p\pi}}\frac{1}{(1+u^2/p)^{(p+1)/2}}.
\end{aligned}$$

This completes the proof of the general result that $Z/\sqrt{W/p}$ has a T distribution on p degrees of freedom when Z is standard normal and, independently, W is chi-square on p degrees of freedom. The choices

$$Z := (\bar{X}-\mu)/\sqrt{\sigma^2/n} \quad \text{and} \quad W := \sum_{i=1}^n (X_i-\bar{X})^2/\sigma^2$$

then reveal that, for $X_1,\ldots,X_n \stackrel{iid}{\sim} N(\mu,\sigma^2)$ with $n \in \{2,3,\ldots\}$,

$$\frac{\bar{X}-\mu}{\sqrt{S^2/n}} \sim T_{n-1}.$$