Distributions: T, F, and Normal Approximation

The purpose of both of these distributions is to allow for inferences about μ and σ in an unknown distribution. Both are quotients of known distributions.

Preliminaries

Sample Mean: Let $Y_1, \dots, Y_n$ be a random, independent sample from a distribution with mean μ and variance σ^2. Then,

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \text{Sample Mean}$$

is a distribution with mean $\overline{\mu} = \mu$ and variance $\overline{\sigma}^2 = \frac{\sigma^2}{n}$. If the underlying distribution is a normal distribution, then $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is a *standard* normal distribution.

Sample Variance: The *sample variance* is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2. \qquad \text{Sample Variance}$$

It is important to note that the sample variance is found for samples drawn from a distribution; for population standard deviation/variance, we use n instead of $n-1$ in the denominator.

When Y_i is a normal distribution, then $\frac{(n-1)S^2}{\sigma^2}$ is a χ^2 distribution with $n-1$ df — S^2 and $\overline{Y}$ are independent.

Definition of T Distribution

Let Z be a standard normal distribution, W be χ^2 with ν df, and Z and W be independent. Then,

$$T = \frac{Z}{\sqrt{W/\nu}}$$

has a T distribution with ν df.

Creating a T Distribution: Let Y_i be sampled from a normal distribution with mean μ and standard deviation σ.

Then, $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is a standard normal distribution, and $W = \frac{(n-1)S^2}{\sigma^2}$ is χ^2 with $n-1$ df.

So,

$$\begin{aligned} T &= \frac{Z}{\sqrt{W/(n-1)}} \\ &= \frac{(\overline{Y}-\mu)\sqrt{n}}{\sigma}\sqrt{\frac{(n-1)\sigma^2}{S^2}} \\ &= \frac{(\overline{Y}-\mu)\sqrt{n}}{S} \end{aligned}$$

has a T distribution with $n-1$ df.

T **Distribution:** Let $Y_1, \dots, Y_6$ be samples from a normal distribution with unknown μ, σ. Estimate $P(|\overline{Y} - \mu| < (2S/\sqrt{n}))$.

Thus, we have

$$\begin{aligned} P\left(|\overline{Y} - \mu| \leq \frac{2S}{\sqrt{n}}\right) &= P\left(-2 \leq \frac{\sqrt{n}(\overline{Y} - \mu)}{S} \leq 2\right) \\ &= P(-2 \leq T \leq 2) \end{aligned}$$

Thus, for $n = 6$, we have that our random variable T has 5 df. By looking at a T distribution table, we can find that $P \approx 0.9$. We can also use R.

Definition of F Distribution

Let W_1 and W_2 be independent χ^2 distributions with ν_1 and ν_2 df respectively. Then, the F distribution with ν_1 numerator df and ν_2 denominator df is found as follows:

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

Simplifying an F Distribution: Let n_1 samples be drawn from normal distribution with mean μ_1 and variance σ_1^2, and n_2 samples be drawn from normal distribution with mean μ_2 and variance σ_2^2. Both distributions are independent.

From each of these samples, we find the sample variance, and create χ^2 distributions with their respective df.

$$\begin{aligned} W_1 &= \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \\ W_2 &= \frac{(n_2 - 2)S_2^2}{\sigma_2^2} \end{aligned}$$

Therefore, we have

$$\begin{aligned} F &= \frac{W_1/(n_1 - 1)}{W_2/(n_2 - 1)} \\ &= \frac{(n_1 - 1)S_1^2}{\sigma_1^2(n_1 - 1)} \frac{\sigma_2^2(n_2 - 1)}{(n_2 - 1)S_2^2} \\ &= \frac{\sigma_2^2}{\sigma_1^2} \frac{S_1^2}{S_2^2} \end{aligned}$$

as an F distribution with $n_1 - 1$ numerator df and $n_2 - 1$ denominator df.

Applying the F Distribution: Let $n_1 = 6$ and $n_2 = 10$ be two samples from independent normal distributions with the same σ^2. Find b such that $P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0.95$.

$$\frac{S_1^2}{S_2^2} = \frac{S_1^2/\sigma^2}{S_2^2/\sigma^2}$$

The given F distribution has 5 numerator df and 9 denominator df. Therefore, we want to find $0.95 = P(F_{5,9} < b)$, or find the 0.95 quantile; in R, we find this with the `qt` function.

Normal Approximation of Binomial

Recall that a binomial distribution Y with n trials and p probability of success has probabilities found below:

$$P(Y \leq \ell) = \sum_{k=0}^{\ell} \binom{n}{k} p^k (1-p)^{n-k}.$$

For very large n, this sum is hard to calculate. We could approximate with the Poisson distribution, but this still requires a lot of calculations and large factorial values. Instead, we will try the following:

$$\begin{aligned} X_i &= \begin{cases} 1 & i \text{ trial success} \\ 0 & i \text{ trial failure} \end{cases} \\ E(X_i) &= p \\ E(X_i^2) &= p \\ V(X_i) &= p(1-p) \\ \overline{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{Y}{n} \\ E(\overline{X}) &= p \\ V(\overline{X}) &= \frac{p(1-p)}{n} \end{aligned}$$

By the Central Limit Theorem, we approximate $\overline{X}$ as a normal distribution with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

Alternatively, we can create, for large fixed n, $Y = n\overline{X}$ with mean np and standard deviation $\sqrt{np(1-p)}$.

For example, consider $p = 0.5$, $n = 100$, $Y =$ number of successes. To find $P\left(\frac{Y}{n} > 0.55\right)$. By the Central Limit Theorem, this is approximately a normal distribution with mean 0.5 and standard deviation 0.05.

Applying Central Limit Theorem: Let Y be a binomial distribution with $n = 25$ and $p = 0.4$. Then, $\mu = np = 10$, and standard deviation $\sigma = \sqrt{\frac{p(1-p)}{n}} = 5\sqrt{0.24}$.

To find $P(Y \leq 8)$, we can potentially approximate with $P(X \leq 8.5)$ — the reason we use 8.5 instead of 8 is due to the fact that n may not be large enough, a process known as the continuity correction.

Using standardization (or R), we find that this probability is approximately 0.269.

The actual probability $P(Y \leq 8)$ is found as below:

$$\begin{aligned} P(Y \leq 8) &= \sum_{k=0}^{8} \binom{25}{k} (0.4)^k (0.6)^{1-k} \\ &= 0.274 \end{aligned}$$

The normal approximation for the binomial is adequate when $p \pm 3\sqrt{\frac{p(1-p)}{n}} \in (0,1)$. Essentially, the binomial trial needs to have an adequate sample size such that the "spread" is small. This is equivalent to $n \geq 9\frac{\max(p,1-p)}{\min(p,1-p)}$.

Estimators

Let Y be a random variable with an *unknown* distribution.

Parameter: Feature of Y's distribution that are not computable from samples.

Examples of Parameters: μ, σ, m_k', interval $(a, b) \ni P(y \in I) = 0.95$.

Statistic: Random variable that is computable from samples.

Examples of Statistics: sample mean, $\overline{Y}$, sample variance, S^2, $Y_{(i)}$.

Estimator: a statistic intended to approximate a parameter. A point estimator estimates a single value.

Examples of Estimators: $\overline{Y}$ as an estimator for μ, and S^2 as an estimator of σ^2.

Bias and Mean Square Error of Estimators

We want to find θ, a constant parameter of the underlying distribution — $\hat{\theta}$ is a random variable.

If $E(\hat{\theta})$ is close to θ, we can say that $\hat{\theta}$ is a good estimator — more precisely, we define the bias $B(\hat{\theta}) = E(\hat{\theta}) - \theta$, and if $B(\hat{\theta}) = 0$, then $\hat{\theta}$ is an unbiased estimator.

In addition to minimizing bias, to see whether or not an estimator is good requires minimizing the variance of the estimator — the mean squared estimator $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta})^2$. Notice that for an *unbiased* estimator, $\text{MSE}(\hat{\theta}) = V(\hat{\theta})$.

Exercise 8.12: Let θ be the true voltage of some electronic device. The voltage test has results uniformly distributed over $[\theta, \theta + 1]$. There are n tests, $Y_1, \dots, Y_n$. Evaluate $\overline{Y}$ as an estimator for θ.

Solution: Since the voltage is uniformly distributed over $[\theta, \theta + 1]$, we have that Y_i is uniform on $[\theta, \theta + 1]$. Therefore, $E(Y_i) = \theta + 0.5$, and $V(Y_i) = \frac{1}{12}$.

Therefore, $E(\overline{Y}) = \theta + 0.5$, and $V(\overline{Y}) = \frac{1}{12n}$, meaning $\text{MSE}(\hat{\theta}) = \frac{1}{12n} + \frac{1}{4}$.

If we want an unbiased estimator for θ, we take $\hat{\theta} = \overline{Y} - \frac{1}{2}$. Then, $E(\hat{\theta}) = E(\overline{Y}) - E(1/2) - \theta = 0$. By shifting this estimator, our new MSE is $\frac{1}{12n}$.

Example 8.1: We will compare the two estimators of σ^2: sample variance and population variance.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$
$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Solution: Recall $V(X) = E(X^2) - (E(X))^2$. Therefore, $E(X^2) = V(X) + (E(X))^2$.

$$\begin{aligned} E(Y_i^2) &= V(Y_i) + (E(Y_i))^2 \\ &= \sigma^2 + \mu^2 \\ E(\overline{Y}^2) &= V(\overline{Y}) + \left(E(\overline{Y})\right)^2 \\ &= \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

Notice that

$$\begin{aligned}
\sum (Y_i - \overline{Y})^2 &= \sum (Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2) \\
&= \sum Y_i^2 - 2\overline{Y}\sum Y_i + \sum \overline{Y}^2 \\
&= \sum Y_i^2 - 2n\overline{Y}^2 + n\overline{Y}^2 \\
&= \sum_{Y_i}^{2} - n\overline{Y}^2 \\
E\left(\sum (Y_i - \overline{Y})^2\right) &= E(\sum Y_i^2) - nE(\overline{Y}^2) \\
&= n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= (n-1)\sigma^2 \\
B(S'^2) &= \frac{1}{n}(n-1)\sigma^2 - \sigma^2 \\
&= -\frac{1}{n}\sigma^2 \neq 0 \\
B(S^2) &= \frac{1}{n--1}(n-1)\sigma^2 - \sigma^2 \\
&= 0
\end{aligned}$$

S'^2 is known as the *biased sample variance*, while S^2 is the unbiased sample variance.

The standard error $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. If $\hat{\theta}$ is unbiased, then $\sigma_{\hat{\theta}} = \sqrt{\text{MSE}(\hat{\theta})}$