Approximation of Student's t-distribution.

Gagarine Yaikhom

I. Introduction

Things cannot be known or controlled perfectly. There is going to be some error in measurement. To make progress, we must accept this reality and design our experiments such that we know the limit to how much error we are willing to tolerate. For instance, if we were weighing a bag of sand, we are likely to tolerate a higher error in measurement than, say, if we were measuring something like gold. What this means is that the degree of error we are willing to accept depends on the task at hand. In statistics, that is the first important decision we must make. The decision of how much error we are willing to tolerate during estimation or inference.

Let us consider another use case. Imagine we are a company that builds computers. Although we want to all computers to work, few of the computers are defective. The question is, when do we decide that the process needs investigating to determine the cause of the defect? How many defects are acceptable before we are required to do any investigation? Of course, we would like to fix all defects, however, investigations cost resources. Hence, we must decide whether an investigation is worth investing on. For instance, if we found out that 1 out of 1000 computers produced are defective, the detect rate is 0.1%. Perhaps, in this case, we might decide that it is not worth an investigation. However, if we found that 100 out fo 1000 computers were defective, then it has a 10% defect rate. In this case, we might lean towards investigating the process as a 10% defect rate is likely to outweigh the cost of an investigation. We make such decisions when we deal with statistical estimation and inference.

In statistics, we are trying to say somthing about a class of objects or phenomena in general based on our observations of a specific subset of that class. For instance, we might be interested in knowing the average height of men. To do this, we measure instances of the feature or property for a subset of the class that we are interested in. For instance, we could measure the height of few of our male friends. Such a measurement subset would be referred to as a sample. Now, we might wonder why we can't just go and measure all men. That way we will know for sure. While it may be possible to do this for some classes, it can be impractical and expensive to collect data this way for a lot of the classes we are interested in. For some classes, it might even be impossible for humans to accomplish. Furthermore, by selecting the sample carefully, we are able to infer the required information to a degree of certainty we have decided is acceptable for our purpose. This is the most interesting part of statistical inference, which we shall discuss shortly.

With respect to terminology, the set of all members of a class is referred to as the *population* of that class. For instance, the population of all men. The process of selecting a subset from a population, a *sample*, is referred to as the *sampling process*. Now, the manner of sampling is extremely important, as it can add bias to our results, if not done correctly. For instance, if we sample ten male friend of ours by measuring their heights, we are unlikely to get an average height close to the true average height of the male population. The result will be biased because their locality to me biases the height, as certain factors in how I chose my friends could inadvertently affect the height of the persons I have chosen as friends. This is the reason why we must carry out random sampling from the population, where members of the sample are chosen at random. Of course, perfect random sampling is often not possible, but we try to minimise bias in our results by designing the experiment carefully. For instance, we are more likely to get a better estimate of the average height of men if we randomly sampled from the entire human race. Furthermore, the size of the sample matters too. For instance, you are likely to get a better estimate of the average height if we sampled 100 men instead of just 10. Even better would be to sample 1000 men. Of course, after a certain point, increasing the sample size would result is diminishing returns relative to the effort and resource we put into the exercise. In other words, after a certain sample size, we won't see improvements to the estimate that are practically relevant. When deciding the sample size, one factor that is key is the amount of error we are willing to accept in our estimate. We shall discuss this further in the following sections.

Using statistical methods, there are two types of questions we are interested in asking. We have already eluded to the first of these, which is the task of *parameter estimation*, where we estimate a measurement of a population property based on the sampling of subsets from the population. The second is *hypothesis testing*, where we test whether a proposition about the properties of a population is falsifiable. In other words, hypothesis testing is concerned with the proving of the truthness or falsity of a statement made concerning populations of classes. Since statistics is about uncertainty, one must note that the work "proving" used in the previous sentence is a weaker form of proof based on empirial data. Once again, the degree to which we are willing to accept errors in our judgment determines the process of statistical hypothesis testing. The manner of the proof realised through hypothesis testing is therefore designed to take this acceptable error of judgement into account, hence, we do not prove our hypothesis directly, but instead, show that the alternative condition where the hypothesis is false is unlikely to be valid given the empirical data. For instance, if we wanted to prove a hypothesis, say, "Average height of men is larger than the average height of women for ages between 30 and 40." we would collect samples of both men and women, and show using the data that it was unlikely that the average height of women of ages between

30 and 40 will be larger or greater than the average height of men of the same ages. We shall discuss this further in the following sections, and try to understand why we have to take such a convoluted way of proving.

II. Parameter estimation

Let us now begin with the task of estimating certain properties or features of a population using sampling. Now, any quantifiable property, a feature that we can measure, are referred to as a *population parameter*. For instance, we can weigh a man, or measure his height, and hence, the average weight or height of a man is a population parameter for the class of mean. On the other hand, beauty is an abstract and sibjective property, which is not easily quantifiable. Hence, such qualitative properties are outside the scope of statistical analysis. Of course, one would devise surrogate metrics that pretend to quantify these qualitative properties, however, it is easy to bias these measurements depending on the mechanism used to quantify the property. For instance, an IQ test may abstractly measure the intelligence of humans but perhaps there are several forms of intelligence that are not quantifiable using these IQ tests. In the following, we do not deal with such parameters.

Let us assume that our task is to estimate the average height of men of a certain age. Now, if we could measure the height of all men of the certain age, we don;t need statistics for that. We don't even need to estimate because we will have a single value that is the exact measurement of the average height of men for that given age. In the following, imaging that the entire population of men of a certain age consists of only 10 men. Then, the average height of the population is 161.7 centimetres. With regards to notation, the average of a population property is denoted with μ. Hence, μ = 161.7 centimetres. Now, how do we arrive at this value? We simply sum the heights in the population and this by the number of members in the population. Notationally, we write:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} h_i, \tag{1}$$

where, h_i is the height of the ith man of the N men in the population. In our example, $N = 10$.

For real life situations, this is not possible to calculate. There are billions on men on planet earth, and there is no consistent way to measure the height of every one of them. So, how can we estimate the average height so that the estimated value is as close to the real value? Through sampling, we can measure the heights of few randomly selected men and use this to estimate the height statistically. Let's say we could only measure the heights of three randomly selected men from the population of 10 men. We list three randomly selected samples: $s_0 = \{155, 180, 181\}$, $s_1 = \{155, 160, 169\}$, and $s_2 = \{169, 172, 176\}$. Just like the way we calculated μ using (1), we can similarly calculate the mean for each of these three samples. Notationally, sample means are denoted as $\bar{x}$, so that:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} h_i,, \tag{2}$$

where, h_i is the height of the ith man in the sample, n is the *sample size*, which is the number of members in a sample. In our example, $n = 3$.

Using this, we have three sample means: $\bar{s}_0 = 172$ centimetres, $\bar{s}_1 = 161.33$ centimetres and $\bar{s}_2 = 172.33$ centimetres. Now, if we compare these samples against the population mean, we can see that sample mean $\bar{s}_1$ *underestimates* the population mean. In other words, the value is less than the population mean. On the other hand, the sample means $\bar{s}_0$ and $\bar{s}_2$ *overestimates* the population mean; the estimated value is larger than the population mean. Furthermore, for the two sample that overestimates the population mean, the degree to which they overestimate are not the same. Hence, if we could only choose one sample, we can never be sure whether the sample is overestimating or underestimating the population parameter. The question therefore is, with so much variability between sample estimates, is it even possible to estimate the population parameter using random sampling in a meaningful way? The answer is fortunately, yes, as long as we do not focus on the individual samples, but rather focus on the behaviour of these samples as a whole. In other words, if we know that although individual random samples from a population behave quite differently, they vary relative to one another in a very specific pattern, we can use this pattern to adjust our estimate by applying a correction based on this pattern. This pattern is in fact the distribution of the sample estimates which we shall now discuss.

III. Introduction

We implement the Student's t-distribution [].

Function `Z_26_2_1()` calculates the value of the probability function, $Z(x)$, for the realisation x of the random variable $X \sim \mathcal{N}(0,1)$:

$$Z(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{3}$$

⟨*Functions*⟩≡

```
double Z_26_2_1(double x)
{
    double t = 1.0 / sqrt(2.0 * M_PI);
    return t * exp(-0.5 * x * x);
}
```

Function `P_26_2_16()` implements an approximation for $P(x)$, based on equation 26.2.16 in [1]:

$$P(x) = 1 - Z(x)(a_1 t + a_2 t^2 + a_3 t^3) + \epsilon(x) \tag{4}$$

where,

$$t = \frac{1}{1+px}, \quad p = 0.33267, a_1 = 0.4361836, \quad a_2 = -0.1201676, \text{ and } a_3 = 0.9372980.$$

The domain of $P(x)$ is $0 \leq x < \infty$, and the approximation has an absolute error $|\epsilon(x)| < 1 \times 10^{-5}$. We use Horner's rule to reduce the number of arithmetic operations while evaluating the polynomial. That is, the polynomial $a_1 t + a_2 t^2 + a_3 t^3$ is evaluated as $t(a_1 + t(a_2 + a_3 t))$, reducing the number of operations from 8 (2 additions and 6 multiplications) to 5 (2 additions and 3 multiplications).

⟨*Functions*⟩+≡

```
double P_26_2_16(double x)
{
    double  p =  0.33267;
    double  t =  1.0 / (1.0 + p * x);
    double a1 =  0.4361836;
    double a2 = -0.1201676;
    double a3 =  0.9372980;
    double  q =  t * (a1 + t * (a2 + t * a3));
    double  z =  Z_26_2_1(x);
    return (1.0 - z * q);
}
```

Function `P_26_2_17()` implements an approximation for $P(x)$, based on equation 26.2.17 in [1]:

$$P(x) = 1 - Z(x)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \epsilon(x) \tag{5}$$

where,

$$t = \frac{1}{1+px}, \quad p = 0.2316419, b_1 = 0.319381530, \quad b_2 = -0.356563782, b_3 = 1.781477937, \quad b_4 = -1.821255978, \text{ and } b_5 = 1.330274429.$$

The domain of $P(x)$ is also $0 \leq x < \infty$, and the approximation has an absolute error $|\epsilon(x)| < 7.5 \times 10^{-8}$.

⟨*Functions*⟩+≡

```
double P_26_2_17(double x)
{
    double  p =  0.2316419;
    double  t =  1.0 / (1.0 + p * x);
    double b1 =  0.319381530;
    double b2 = -0.356563782;
    double b3 =  1.781477937;
    double b4 = -1.821255978;
    double b5 =  1.330274429;
    double  q =  t * (b1 + t * (b2 + t * (b3 +
                 t * (b4 + t * b5))));
    double  z =  Z_26_2_1(x);
    return (1.0 - z * q);
}
```

Function `t_26_2_22()` determines for probability p the corresponding x_p such that $Q(x_p) = p$.

$$x_p = 1 - \frac{a_0 + a_1 t}{1 + b_1 t + b_2 t^2} + \epsilon(p) \tag{6}$$

where,

$$t = \sqrt{\ln \frac{1}{p^2}}, \quad a_0 = 2.30753, a_1 = 0.27061, \quad b_1 = 0.99229, \text{ and } b_2 = 0.04481.$$

The function's domain is $0 < p \leq .5$, and the approximation has an absolute error $|\epsilon(p)| < 3 \times 10^{-3}$.

⟨*Functions*⟩+≡

```
double t_26_2_22(double p)
{
    double  t = sqrt(log(1.0 / (p * p)));
    double a0 = 2.30753;
    double a1 = 0.27061;
    double b1 = 0.99229;
    double b2 = 0.04481;
    double  q = (a0 + a1 * t) / (1.0 + t * (b1 + t * b2));
    return t - q;
}
```

Function `t_26_2_23()` determines for probability p the corresponding x_p such that $Q(x_p) = p$.

$$x_p = 1 - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \epsilon(p) \qquad (7)$$

where,

$$t = \sqrt{\ln \frac{1}{p^2}},$$
$$c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328,$$
$$d_1 = 1.432788, d_2 = 0.189269, \text{ and } d_3 = 0.001308.$$

The function's domain is $0 < p \leq .5$, and the approximation has an absolute error $|\epsilon(p)| < 4.5 \times 10^{-4}$.

⟨*Functions*⟩+≡

```
double t_26_2_23(double p)
{
    double  t = sqrt(log(1.0 / (p * p)));
    double c0 = 2.515517;
    double c1 = 0.802853;
    double c2 = 0.010328;
    double d1 = 1.432788;
    double d2 = 0.189269;
    double d3 = 0.001308;
    double  q = (c0 + t * (c1 + t * c2)) /
                (1.0 + t * (d1 + t * (d2 + t * d3)));
    return t - q;
}
```

⟨*Test functions*⟩≡

```
void test_P()
{
     double dx = 0.005;
     double x = 0.0;
     for (int i = 0; i < 1000; ++i) {
         printf("%f, %f\n", x, P_26_2_16(x));
         x += dx;
     }
}
```

IV. Main program

Main program for testing the functions.

⟨*Main*⟩≡

```
int main(void) {
    test_P();
    return 0;
}
```

⟨*Standard libraries*⟩≡

```
#include <math.h>
#include <stdio.h>
#include <string.h>
```

⟨*algorithm.c*⟩≡
⟨*Standard libraries*⟩
⟨*Functions*⟩
⟨*Test functions*⟩
⟨*Main*⟩

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions With Formulas, Graphs, and Mathematical Tables*. Number 55 in Applied Mathematics Series. National Bureau of Standards, 1972.

[2] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.