

6 Inference Based On Normal Distribution

*ENSY SILVER*¹

Wednesday 2nd September, 2020

¹Thanks to my family, my friend and freedom.

1 Introduction

Finding probability distributions to describe—and, ultimately, to predict—empirical data is one of the most important contributions a statistician can make to the research scientist. Already we have seen a number of functions playing that role.

2 Estimator of variance σ^2

Suppose that a random sample of n measurements, Y_1, Y_2, \dots, Y_n , is to be taken on a trait that is thought to be normally distributed, the objective being to draw an inference about the underlying pdf's true mean, μ . If the variance σ^2 is known, we already know how to proceed.

In practice, though, the parameter σ^2 is seldom known, so the ratio $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ can not be calculated. The usual estimator for the population variance,

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

whose deduction is at [2].

Historically, many early practitioners of statistics felt that replacing σ with S had, in fact, no effect on the distribution of the Z ratio. Sometimes they were right. If the sample size is very large (which was not an unusual state of affairs in many of the early applications of statistics), the estimator S is essentially a constant and for all intents and purposes equal to the true σ .

The ratio $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ is called the **Student t distribution**.

3 Chi square distribution

To introduce the Student t distribution, we first introduce the **chi square distribution**.

Theorem 3.1. *Let $U = \sum_{j=1}^m Z_j^2$, where $\{Z_j\}$ are independent standard normal random variables. Then U has a gamma distribution with $r = \frac{m}{2}$ and $\lambda = \frac{1}{2}$. That is,*

$$f_U(u) = \frac{1}{2^{m/2} \Gamma(\frac{m}{2})} u^{(m/2)-1} e^{-u/2}, \quad u \geq 0$$

which we called the chi square distribution with m degrees of freedom.

Theorem 3.2. *Let Y_1, \dots, Y_n be a random sample from a normal distribution with mean μ and variance σ^2 . Then*

1. S^2 and \bar{Y} are independent.
2. $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2$ has a chi square distribution with $n-1$ degrees of freedom.

6 Inference Based On Normal Distribution STUDENT T DISTRIBUTION

Proof. The proof is at [1]. □

Suppose that U and V are independent chi square random variables with n and m degrees of freedom, respectively. A random variable of the form $\frac{V/m}{U/n}$ is said to have an F distribution with m and n degrees of freedom.

Theorem 3.3. Suppose $F_{m,n} = \frac{V/m}{U/n}$ denotes an F random variable with m and n degrees of freedom. The pdf of $F_{m,n}$ has the form

$$f_F(w) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2} w^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right) (n + mw)^{(m+n)/2}}$$

4 Student t distribution

Now, we introduce the formal definition of **Student t distribution** derived from *chi square distribution*.

Theorem 4.1. Let Z be a standard normal random variable and let U be a chi square random variable independent of Z with n degrees of freedom. The Student t ratio with n degrees of freedom is denoted T_n , where

$$T_n = \frac{Z}{\sqrt{U/n}}$$

We review the Student t distribution $\frac{\bar{X}-\mu}{\sqrt{S/n}}$, and we easily get that it is a Student t distribution with $n - 1$ degrees of freedom.

The pdf of T_n is given by the theorem.

Theorem 4.2. The pdf of a Student t random variable with n degrees of freedom is given by

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right) \left(1 + \frac{t^2}{n}\right)^{(n+1)/2}}$$

BTW, the estimation of μ of a Student t distribution is the same as the case of normal distribution. The only difference is the table.

In the case that X is not normally distributed, we may use computer to calculate the integral.

4.1 Estimation of σ^2

Since we have know that

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (Y_k - \bar{Y})^2$$

which satisfies *chi square distribution*, the confidence interval of σ is derived from the equation

$$P \left[\chi_{\alpha/2, n-1}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{1-\alpha/2, n-1}^2 \right] = 1 - \alpha$$

where $\chi_{p,n}^2$ denotes that in the chi square distribution with n degrees of freedom, $P(x \leq \chi_{p,n}^2) = p$.

References

- [1] jld. *Independence of meean and estimator*. <https://stats.stackexchange.com/questions/344960/showing-s2-and-overliney-are-independent-seeking-a-solution-to-this-tex>. May 2018.
- [2] Wikipedia. *Bias of an estimator*. https://en.wikipedia.org/wiki/Bias_of_an_estimator. June 2020.