Usage of `get_onesmpltea_tab.m` (onesmpltea1-kato.pdf)

Tsuyoshi Kato

February 10, 2020

1 Example

In statistically assessing the difference among two statistical values, the one sample t-test is often employed. In the author's research group, a matlab code `get_onesmpltea_tab.m` is used to perform the one sample t-test. In this note, the usage of this function is described. The sample code used in this section is `demo011_01.m` which is available from [1]

観測値の差を統計的に評価するとき，One Sample t-Test がよく用いられている．著者の研究グループでは `get_onesmpltea_tab.m` を使って One Sample t-Test を行っている．本稿では，`get_onesmpltea_tab.m` の使い方を解説する．本節で用いられているサンプルコードは `demo011_01.m` であり，[1] からダウンロードできる．

Now let us suppose that performance test is done $(n = 5)$ times independently for $(m = 3)$ different methods. Then, the number of the quantitative values representing the performance is 15. Assume that the following quantitative performance values are given:

いま，例として，$(m = 3)$ 種類の方法の性能評価のため，$(n = 5)$ 回試験したとする．すると，性能を表す定量値が各方法につき $(n = 5)$ 個，合計 $(m \cdot n = 15)$ 個得られることになる．次のように得られたとする．

$$\texttt{accmat} := \begin{bmatrix} (\boldsymbol{x}^{(1)})^\top \\ (\boldsymbol{x}^{(2)})^\top \\ (\boldsymbol{x}^{(3)})^\top \end{bmatrix} = \begin{bmatrix} 0.4567 & 0.5288 & 0.6189 & 0.5175 & 0.4412 \\ 0.3334 & 0.3854 & 0.4962 & 0.4813 & 0.7183 \\ 0.6125 & 0.7191 & 0.6327 & 0.6726 & 0.5864 \end{bmatrix} \tag{1}$$

The average over five trials for each method is given by:

おのおのの方法の平均は次のようになる：

$$\bar{x}^{(1)} := \left\langle \boldsymbol{x}^{(1)}, \mathbf{1} \right\rangle / n = 0.5126, \qquad \bar{x}^{(2)} := \left\langle \boldsymbol{x}^{(2)}, \mathbf{1} \right\rangle / n = 0.4829, \qquad \bar{x}^{(3)} := \left\langle \boldsymbol{x}^{(3)}, \mathbf{1} \right\rangle / n = 0.6447 \tag{2}$$

The Matlab function `get_onesmpltea_tab.m` assesses the statistical hypothetical test for the following three hypotheses:

Matlab 関数 `get_onesmpltea_tab.m` を使うと，次の 3 つの仮説を統計的に評価ができる：

- Is Method 1 better than Method 2 ?
- Is Method 3 better than Method 2 ?
- Is Method 3 better than Method 1 ?

For this table `accmat`, doing `ptab = get_onesmpltea_tab(accmat);` yields the following P-value table `ptab`:

この表 `accmat` に対して，`ptab = get_onesmpltea_tab(accmat);` を実行すると，P 値の表 `ptab` が次のように得られる：

$$\texttt{ptab} := \begin{bmatrix} +1 & +0.3611 & +0.0038 \\ +0.3611 & +1 & +0.0511 \\ +0.0038 & +0.0511 & +1 \end{bmatrix} \tag{3}$$

The table `ptab` is always a symmetric matrix. Each entry in `ptab` represents the following:

`ptab` はつねに対称行列になる．この表にある各要素はつぎのことを表している．

- `ptab(1,2)` is the P-value of the null hypothesis that Method 1 is not better than Method 2.
- `ptab(2,3)` is the P-value of the null hypothesis that Method 3 is not better than Method 2.

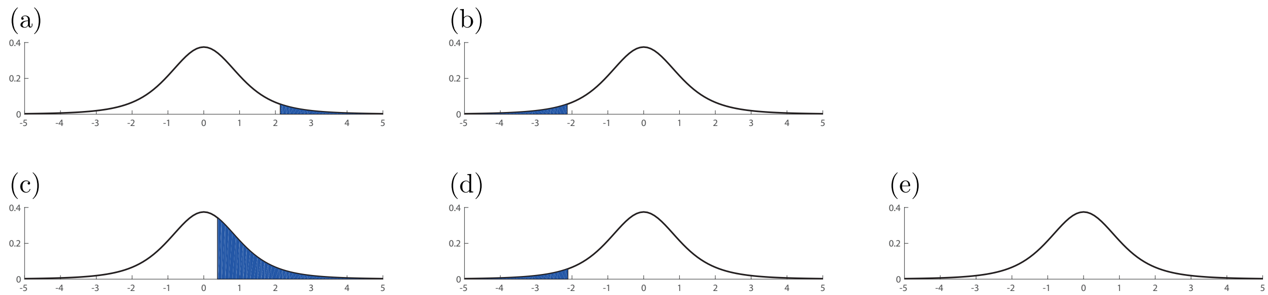

Figure 1: Student t distribution.

- `ptab(3,1)` is the P-value of the null hypothesis that Method 3 is not better than Method 1.

If we set the significance level to 0.01, each non-diagonal entry in `ptab` is compared with 0.01 to determine the significance. The procedure concludes that the difference between Method 1 and Method 2 is not significant, neither is the difference between Method 2 and Method 3. Meanwhile, we can say that Method 3 is better than Method 1.

たとえば，有意水準を 0.01 とすると，各要素を 0.01 より大きいか小さいか比較することにより，方法 1 が方法により優れているとはいえず，また，方法 3 が方法 2 より優れているとはいえない．一方，方法 3 が方法 1 より優れているといえる．

2 Algorithm

In this section, what the Matlab function `get_onesmpltea_tab.m` is doing is described. Every diagonal entry in `ptab` is always one. Below, the procedure of computing a non-diagonal entry `ptab(1,2)` is written. The other non-diagonal entries are computed in a similar way.

Matlab function `get_onesmpltea_tab.m` の内部で行っている計算を説明する．P 値の表 `ptab` の対角成分は常に 1 である．非対角成分として `ptab(1,2)` の計算方法をのべる．ただし，ほかの非対角成分も同様な計算方法である．

Suppose that we are given quantitative performance values of Method 1 and Method 2, $\boldsymbol{x}^{(1)} = \left[x_1^{(1)}, \ldots, x_n^{(1)}\right]^\top$ and $\boldsymbol{x}^{(2)} = \left[x_1^{(2)}, \ldots, x_n^{(2)}\right]^\top$. The difference between them is computed:

方法 1 および方法 2 の性能評価値がそれぞれ $\boldsymbol{x}^{(1)} = \left[x_1^{(1)}, \ldots, x_n^{(1)}\right]^\top$ および，$\boldsymbol{x}^{(2)} = \left[x_1^{(2)}, \ldots, x_n^{(2)}\right]^\top$ が得られていたとする．この量から差分

$$\boldsymbol{d} = [d_1, \ldots, d_n]^\top := \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)} \tag{4}$$

Then, sample mean and Bessel-corrected standard deviation of n values, $d_1, \ldots, d_n$, denoted by $\bar{d}$ and s_u, and then the t-value are computed as:

を計算する．差分の標本平均 $\bar{d}$， Bessel-corrected 標準偏差 s_u そして t 値を計算する：

$$\bar{d} := \langle \mathbf{1}, \boldsymbol{d} \rangle / n, \qquad s_\mathrm{u} := \sqrt{\frac{\|\boldsymbol{d}\|^2 - \bar{d}^2/n}{n-1}}, \qquad t := \frac{\bar{d}}{s_\mathrm{u}/\sqrt{n}}. \tag{5}$$

The P-value is computed:

P 値を計算する．

$$p := \int_{-\infty}^{-|t|} \mathcal{T}_{n-1}(x) dx. \tag{6}$$

where $\mathcal{T}_\nu$ is the probabilistic density function of Student t distribution with ν degrees of freedom.

ただし，$\mathcal{T}_\nu$ は自由度 ν の t 分布の密度関数を表す．

The value of p is `ptab(1,2)`. Similarly, the other non-diagonal entries are computed.

この値 p が `ptab(1,2)` となる．同様にして，対称行列 `ptab` のすべての非対角成分を計算する．

3 Theory

The procedure of the statistical hypothesis test described in the previous section is based on the following theorem.

前節で述べた検定方法は，次の理論に基づいている．

Theorem 1. Assume that n observations $d_1, \dots, d_n$ are drawn from a normal distribution $\mathcal{N}(0, \sigma^2)$. Let $\boldsymbol{d} := [d_1, \dots, d_n]^\top$. Let

$$\bar{d} := \langle \mathbf{1}, \boldsymbol{d} \rangle / n, \qquad s_{\mathrm{u}} := \sqrt{\frac{\|\boldsymbol{d}\|^2 - \bar{d}^2/n}{n-1}}, \qquad t := \frac{\bar{d}}{s_{\mathrm{u}}/\sqrt{n}}. \tag{7}$$

Then the density function of the random variable t is given by $\mathcal{T}_{n-1}$.

If $d_1, \dots, d_n$ are i.i.d. and drawn from a zero-mean normal distribution, the distribution of the t-value is described with the solid line in Figure 1a. This plot suggests that the t value is likely to occur around zero, and the t value occurs more rarely in the area more distant from zero.

$d_1, \dots, d_n$ が i.i.d で平均 0 の正規分布と仮定すると，統計量 t の分布は図 1a の実線のような形になる．図からわかるように，0 付近が最も生起しやすく，0 から離れれば離れるほど，生起しにくくなることが分かる．

Let us look back at the example used in Section 1. Recall that the difference, denoted by $\boldsymbol{d}^{(1,2)}$ in this section, is given by:

これらを踏まえて，1 節の例に戻ることにする．差 $\boldsymbol{d}^{(1,2)}$ は次のようになっていた：

$$\boldsymbol{d}^{(1,2)} := \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)} = \begin{bmatrix} +0.1233 & +0.1434 & +0.1227 & +0.0361 & -0.2772 \end{bmatrix} \tag{8}$$

It can be observed that the performances of Method 1 are higher than those of Method 2 in the first four trials, although in the last trial Method 1 gets a worse performance than Method 2. The average of the difference is positive:$\bar{d}^{(1,2)} = +0.0297$. Now let us discuss whether we can say that the difference is statistically significant.

すなわち，最初の４回の試行では，方法１のほうが高く，最後の１回だけ方法２のほうが高かった．平均すると $\bar{d}^{(1,2)} = +0.0297$ と正の値になった．これより，方法１のほうが有意に高いといえるか考えてみる．

The corresponding t-value is $t^{(1,2)} = 0.3761$. As shown in the solid curve in Figure 1a, the density at $t^{(1,2)} = 0.3761$ is still high, suggesting that the t-value is possible enough even if the population mean is zero.

対応する t 値は $t^{(1,2)} = 0.3761$ となった．図 1a の実線の分布をみると，$t^{(1,2)} = 0.3761$ の値の密度は高く，真の平均が 0 であっても十分に起きうるずれといえそうである．

In the one-sided one-sample t-test, assessment of the statistical significance is performed by checking if the t-value is in the blue interval of Figure 1a.

One Sample t-test の片側検定では，図 1a の青い領域に入っているか否かで，真の平均が 0 でも偶然起きうるか，ほとんど起きないのか評価する．

The area of the blue interval is 0.05 if the significant level is set to 0.05. The probability of entering the blue interval is 5%. If the t-value falls in the blue interval, we can say that the event has occurred although the probability is only 5%, allowing us to reject the null hypothesis that the population mean is zero.

いま，有意水準を 0.05 としたとすると，青い領域の面積は 0.05 である．この青い領域に入る確率は 5% になる．青い領域に t 値が入ったとしたら，5% しか起きないような低い確率のことが起きてしまったとして，めったに起きないことが起こった，と解釈して，真の平均が 0 という仮説を棄却する．

The blue interval in Figure 1a is $[2.132, +\infty)$. The null hypothesis that the population mean is zero cannot be rejected because $t^{(1,2)} = 0.3761 \notin [2.132, +\infty)$.

図 1a の青い区間は $[2.132, +\infty)$ である．$t^{(1,2)} = 0.3761 \notin [2.132, +\infty)$ なので，真の平均が 0 という仮説を棄却できない．

Next let us compare Method 2 with Method 3. The performance difference is given by:

今度は方法２と方法３の比較を考えてみる．５回の試行それぞれにおける性能の差は次のようになる．

$$\boldsymbol{d}^{(2,3)} := \boldsymbol{x}^{(2)} - \boldsymbol{x}^{(3)} = [-0.2791, -0.3337, -0.1365, -0.1913, +0.132]^\top \tag{9}$$

The average difference is negative:$\bar{d}^{(2,3)} = -0.1617$. We now discuss whether Method 2 is inferior to Method 3. The t-value is $t^{(2,3)} = -1.9973$. The interval from the left which has the area under the curve being 0.05 is $(-\infty, -2.132]$. This interval is shown in Figure 1b. Falling in the interval realizes with probability 0.05. If the event occurs, the null hypothesis of the population mean being zero is rejected. The t-value, $t^{(2,3)} = -1.9973$, is not in the critical region $(-\infty, -2.132]$, implying that the null hypothesis is not rejected.

差の平均は $\bar{d}^{(2,3)} = -0.1617$ と負の値になる．では，方法2は方法3より劣っているといえるか考えよう．t 値は $t^{(2,3)} = -1.9973$ となる．t 分布において，左側から面積が 0.05 になる区間は $(-\infty, -2.132]$ である．図 1b はこの区間を示している．この区間に入っていれば，5% しか起きないほど珍しいことが起きたとして，真の平均が 0 という仮説を棄却できる．$t^{(2,3)} = -1.9973$ は区間 $(-\infty, -2.132]$ に入っていないので，真の平均が 0 という仮説を棄却できない．

Lastly, comparison of Method 3 and Method 1 is considered. The differences in the five trials are given by:

最後に方法3と方法1の比較を考える．5回の試行それぞれにおける性能の差は次のようになる．

$$\boldsymbol{d}^{(3,1)} := \boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)} = [+0.1558, +0.1903, +0.0138, +0.1551, +0.1452]^\top \tag{10}$$

The average difference is $\bar{d}^{(3,1)} = 0.1320$, resulting that the corresponding t-value satisfies $t^{(3,1)} = 4.3249 \in [2.132, +\infty)$, and thereby it can be concluded that the performance difference is significantly large enough to reject the null hypothesis that the population mean is zero.

差の平均は $\bar{d}^{(3,1)} = 0.1320$，対応する t 値は $t^{(3,1)} = 4.3249 \in [2.132, +\infty)$ になるので，真の平均が 0 という仮説を棄却できるほど，性能差の平均は大きいと結論づけられる．

The values of `ptab` show the so-called P-values. When the average difference is positive (i.e. $\bar{d} > 0$), then the P-value is the definite integral of $T_n(x)$ in the interval $[t, +\infty)$. When the average difference is negative, then the P-value is the definite integral of $T_{n-1}(x)$ in $(-\infty, t]$.

`ptab` の値は，P 値と呼ばれる値を示している．差の平均 $\bar{d} > 0$ ならば，区間 $[t, +\infty)$ で $T_{n-1}(x)$ を定積分した値，差の平均 $\bar{d} < 0$ ならば，区間 $(-\infty, t]$ で $T_{n-1}(x)$ を定積分した値になる．

In the example of Section 1, the values of `ptab(1,2)`, `ptab(2,3)`, `ptab(3,1)` are the blue areas in Figure 1c,d,e.

1 節の例では，`ptab(1,2)`，`ptab(2,3)`，`ptab(3,1)` は，図 1c,d,e の青い部分の面積である．

The assessment of the statistical significance is done by comparing the P-value with the significant level.

この P 値と有意水準を比較することで，帰無仮説を棄却できるかできないか判断することができる．

References

[1] http://www.net-machine.net/tsattach1/kato/200210/prot490-demo011-onesmpltea.2002102218.tgz.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, USA, 2006.