 2021 年 6 月 30 日作成

名古屋工業大学 先進セラミックス研究センター　井田 隆 2025 年 1 月 10 日更新

統計学的な検定　Statistical test

2020 年から 2021 年にかけてコロナウィルス感染の大規模な拡大が進行し，疫学調査など大規模なデータの収集，PCR (ポリメラーゼ連鎖反応 polymerase chain reaction) 検査による診断，ワクチンの有効性の評価など，統計学的な分析に基づいた意思決定の重要性が，強く印象付けられることになりました。

この記事では，昨今耳にする機会の増えた「統計学的な検定」のことについて，整理を試みます。「統計学的な検定」として「*t* 検定（ティーけんてい）」「χ^2 検定（カイじじょうけんてい）」「F 検定（エフけんてい）」などの語が使われます。

１．Student の t 検定

Student の t 検定 (Student's t -test) （補足 1.A）は，データの統計的な分布が独立であり（データの間に相関がなく）正規分布に従うとみなせる場合に有効な仮説検定法（仮説が正しいかを調べる方法）です。この方法は Student が 1908 年に標本平均のとりうる誤差 (The probable error of a mean) について発表した内容 (Student, 1908) に基づきます。

Student の考え方に従えば，正規分布に従う有限の標本値から計算される標本平均と標本標準偏差から推定される統計分布は，正規分布より一般的に裾の広がった分布になり，正規分布を仮定してしまうと誤差を過小に評価する傾向があります。

その後「Student の t 検定」として，有意水準 (significance level) を例えば 5% と設定し，t 分布表 (t-table) を参照し，有意水準以下の確率でしか成立しない**帰無仮説** (null hypothesis) を棄却して，帰無仮説を否定する意味の「**対立仮説** (alternative hypothesis)」を採用するという使われ方をされました。一方で，現代的な計算システムを利用した t テストでは，あらかじめ有意水準を決めることも， t 分布表を使うことも，帰無仮説と対立仮説をあらかじめ決めておくことも必要ではなくなっています。

Student t 検定は，データが独立な正規分布に従うとき，**検定統計 量** (test statistic) が **Student の t 分布** (t-distribution) に従うことに基づきます。ただし「データが正規分布に従わなければ使ってはいけない」ということではなく，**データ**が正規分布に従わなくても，t 検定をすれば「何かの意味のある結果」は得られます。ただし，この検定のしかたがどのような論理に基づくかを知らなければ，その意味を正しく理解することも，うまく利用することも難しいでしょう。

t 検定の使われ方に「**一標本 t 検定** (one-sample location test)」と「**二標本 t 検定** (two-sample location test)」が存在します。ところが，現代的な計算システムを利用できる場合であっても「二標本 t 検定」は，多くの場合に間違った使われ方がされています。

標本平均 $\overline{X}$，標本平均 $\hat{\sigma}$，要素数 n のとき，一標本 t 検定で**検定統計量** (test statistc) は，

$$t = \frac{\overline{X} - \mu}{\hat{s}} = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}} \tag{1.1}$$

文脈によっては

$$\hat{t} = \frac{\overline{X}}{\hat{s}} = \frac{\overline{X}}{\hat{\sigma}/\sqrt{n}} \tag{1.2}$$

のような式で表されます。式 (1.1) の中の μ は母平均 (population mean) を表します。

式 (1.1) の右辺の分母 $\hat{s}$ は**尺度化パラメータ** (scaling parameter) とも呼ばれ，データの統計的な分布の広がりの大きさも考えに入れて，（広がりが大きい場合でも，小さい場合でも）「t の分布」について共通の数学的な形式（数表や函数表現など）を利用できるようにするためのものと考えれば良いでしょう。

式 (1.1) で表される t の意味する「**検定統計量**」とは，「データに基づいて仮説の検定（テスト）をするために（数表や函数表現などを使いやすくするために尺度化した）統計的な値」というような意味合いのことです。

帰無仮説が正しいのに棄却してしまうこと（誤った対立仮説を採択してしまうこと）は**第一種過誤** (Type I error; error of the first kind) あるいは**偽陽性** (false positive) と呼ばれます。また**帰無仮説**が誤っているのに採択すること（対立仮説が正しいのに棄却してしまうこと）ことは**第二種過誤** (Type II error; error of the second kind) あるいは**偽陰性** (false negative) と呼ばれます（補足 1.B）。

1-1　一標本 t 検定

一標本 t 検定は，データの母平均値推定のために用いられます。一般的には母平均 μ が基準値 μ_{ref} より大きいか・小さいかを調べるために用いられますが，ここでは単純化のために $\mu_{\mathrm{ref}} = 0$ とします。この場合に**検定統計量** t の値は，

$$t = \frac{z}{\hat{s}} = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}} \tag{1.1.1}$$

と表されます。$\mu_{\mathrm{ref}} \neq 0$ の場合についての結果を求めるためには，式 (1.1.1) 中の $\overline{X}$ を $\overline{X} - \mu_{\mathrm{ref}}$ で置き換えます。

基準母平均値を $\mu_{\mathrm{ref}} = 0$ とするとき，「推定平均値 μ が負である ($\mu < 0$)」あるいは「推定平均値が正である ($\mu > 0$)」ことが「帰無仮説」あるいは「対立仮説」となります。

式 (1.1.1) の $\overline{X}$ は，**標本平均** (sample mean) であり，n 個の標本値（実験・調査などで得られた複数の数値）$\{X_1, X_2, \cdots, X_n\}$ があったとき

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{1.1.2}$$

として計算される値です。式 (1.1.1) の $\hat{s}$ には**標本平均の標準誤差** (standard error of the sample mean) と呼ばれる値を使います。$\hat{s}$ は「データから得られた標本平均の値 $\overline{X}$ の誤差として推定される値」の意味を持ちます。$\hat{\sigma}$（シグマ・ハット）（σ はギリシャ文字小文字のシグマ, sigma）は「**標本標準偏差** (sample standard deviation)」です。式 (1.1.1) では

$$\hat{s} = \hat{\sigma}/\sqrt{n} \tag{1.1.3}$$

の関係を使っています。式 (1.1.3) は「標本平均 $\overline{X}$ の分散推定値の期待値 $\langle \hat{s}^2 \rangle$ は，母集団の分散推定値の期待値 $\langle \hat{\sigma}^2 \rangle$ の『n 分の一』である」という関係から導かれます。標本標準偏差 $\hat{\sigma}$ の値は，

$$\hat{\sigma} = \sqrt{\frac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \cdots + \left(X_n - \overline{X}\right)^2}{n - 1}} \tag{1.1.4}$$

あるいは

$$\hat{\sigma} = \sqrt{\frac{n}{n-1}}\sqrt{\frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n} - \left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)^2} \tag{1.1.5}$$

として計算します。

記号 μ（μ はギリシャ文字小文字のミュー, mu）は統計分布の「真の平均」「**母平均** (population mean)」（母集団の平均）を表します。母平均 μ の値は「本来なら，正確には知ることができないはずの値」だということに注意してください。

同じように「真の標準偏差 ＝ **母標準偏差** (population standard deviation)」が σ であるとして，母標準偏差 σ の値も「本来なら，正確には知ることができないはずの値」です。式 (1.1.4) または 式 (1.1.5) で計算される値には，母標準偏差を意味する σ の記号にハット (hat) 記号 ^ を付けて $\hat{\sigma}$ と表しますが，このことには，「標本標準偏差 $\hat{\sigma}$」が「母標準偏差 σ」とは違う値であることをはっきりさせたいという意図があります。

標本値（データ）$\{X_1, X_2, \cdots, X_n\}$ の統計分布が母平均 μ，母標準偏差 σ の**正規分布** (normal distribution) に従えば，**標本平均** (sample mean) $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ の統計分布は平均 μ，標準偏差 $\sigma/\sqrt{n}$ の**正規分布**に従います（補足 1.1.A）。また，

$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ の統計分布は平均 0，標準偏差 1 の正規分布（**標準正規分布**；standard normal distribution）に従います。

確率変数 x が標準正規分布（平均 0, 標準偏差 1 の正規分布）に従うとき，x^2 の統計分布は「**自由度 1 のカイ自乗分布**」(chi squared distribution) に従います（補足 1.1.B）。また**標本分散** (sample variance) あるいは**不偏標本分散** (unbiased sample variance) $\hat{\sigma}^2$ の値を母分散 σ^2 で除し要素数 n を乗じた値 $\hat{s}^2 = n\hat{\sigma}^2/\sigma^2$ は，自由度 $\nu = n-1$ のカイ自乗分布に従います（補足 1.1.C）（補足 1.1.D）（補足 1.1.E）。

$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ が**標準正規分布** (standard normal distribution) に従う場合に，**検定統計量**

$t = \dfrac{\overline{X}-\mu}{\hat{\sigma}/\sqrt{n}}$ は**自由度** (degree of freedom) $\nu = n-1$ （ν はギリシャ文字小文字のニュー, nu）の**スチューデント t 分布** (Student's t- distribution) に従います。スチューデント t 分布の**確率密度函数** (probability density function) は

$$f_{\mathrm{Student}}(t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{\nu}\,\mathrm{B}\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{1.1.6}$$

と表されます。式 (1.1.6) の中の $\Gamma(\alpha)$ は**ガンマ函数** (gamma function) あるいは完全ガンマ函数 (complete gamma function) であり，

$$\Gamma(\alpha) \equiv \int_0^\infty t^{\alpha-1}\,\mathrm{e}^{-t}\,\mathrm{d}t \tag{1.1.7}$$

として定義されます。$\mathrm{B}(a,b)$ は**ベータ函数** (beta function) あるいは完全ベータ函数 (complete beta function) であり，

$$\mathrm{B}(a,b) \equiv \int_0^1 t^{a-1}\,(1-t)^{b-1}\,\mathrm{d}t \tag{1.1.8}$$

として定義されます。

一標本 t 検定で母平均に関する検定をする場合に，標本数 n に対して自由度が $\nu = n-1$ となり，「自由度 ν がデータ数 n より 1 小さい数」になるのは，「標本平均 $\overline{X}$ を計算する時に，データのうちの標本 1 つ分の情報を失うから」と説明される場合が多いようです。

Student t 分布は **Pearson VII 型分布**の標準化された形式の意味を持ち，$\nu = 1$ のとき標準化された**コーシー分布** (Cauchy distribution)，$\nu \to \infty$ のとき標準化された**正規分布** (normal distribution) と一致します。

$\nu = 1$ の時の**Student t 分布**の確率密度函数 (probability density function) は

$$f_{\text{Student}}(t; 1) = \frac{1}{\pi} \left(1 + t^2\right)^{-1} \tag{1.1.8}$$

と表されます。式 (1.1.8) で表される確率密度函数は**ローレンツ型函数** (Lorentzian function) とも呼ばれます（補足 1.1.F）。

$\nu \to \infty$ の時の**Student t 分布**の確率密度函数は

$$f_{\text{Student}}(t; \infty) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \tag{1.1.9}$$

と表され，式 (1.1.9) で表される確率密度函数は，**ガウス型函数** (Gaussian function) とも呼ばれます（補足 1.1.G）（補足 1.1.H）（補足 1.1.I）。

Student t 分布の確率密度函数の形状を Figure 1.1.1 に示します。この確率密度函数が正しく検定統計量 t の統計分布を表す特徴を持つらしいことは，モンテカルロ・シミュレーションによって確認することができます（補足 1.1.J）。

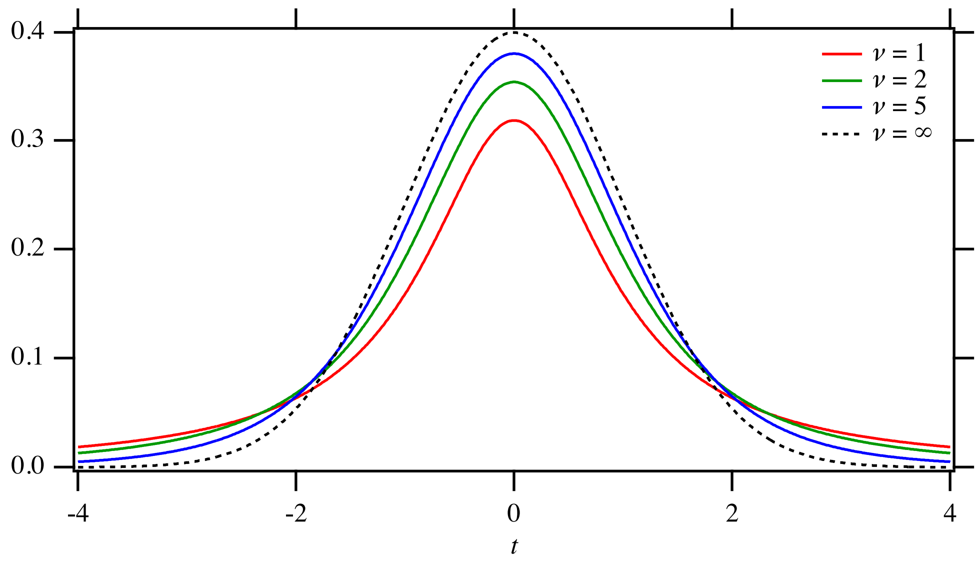

Figure 1.1.1　Student t 分布の確率密度函数。自由度を ν とする。

確率密度函数は「確率そのもの」を表すものではありません。確率密度函数の値 $f(t)$ には，「$\mathrm{d}t \to 0$ のときに『確率変数が t から $t + \mathrm{d}t$ の間の値をとる確率』が $f(t)\,\mathrm{d}t$ と表される」という意味があることに注意してください。

Student t 分布 (Student's t-distribution) の**累積分布函数** (cumulative distribution function) は，$I_x(a, b)$ あるいは $I(a, b; x)$ などと表記される**正則不完全ベータ函数** (regularized incomplete beta function) を使って，

$$F_{\mathrm{Student}}(t;\nu) = \int_{-\infty}^{t} f_{\mathrm{Student}}(u;\nu)\,\mathrm{d}u = \begin{cases} 1 - \dfrac{1}{2} I\left(\dfrac{\nu}{2}, \dfrac{1}{2}; \dfrac{\nu}{t^2+\nu}\right) & [t \geq 0] \\ \dfrac{1}{2} I\left(\dfrac{\nu}{2}, \dfrac{1}{2}; \dfrac{\nu}{t^2+\nu}\right) & [t < 0] \end{cases} \tag{1.1.10}$$

と表されます。正則不完全ベータ函数 $I(a,b;x)$ は

$$I(a,b;x) \equiv \frac{\mathrm{B}(a,b;x)}{\mathrm{B}(a,b)} \tag{1.1.11}$$

として定義されます。$\mathrm{B}(a,b;x)$ は**不完全ベータ函数** (incomplete beta function) であり，

$$\mathrm{B}(a,b;x) \equiv \int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t \qquad [0 < x < 1] \tag{1.1.12}$$

として定義されます。

式 (1.1.10) によって計算される Student t 分布の累積度数分布函数を Figure 1.1.2 に示します。ただし，$\nu \to \infty$ の場合には

$$F_{\mathrm{Student}}(t;\infty) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{t}{\sqrt{2}}\right) \tag{1.1.14}$$

とします。ここで $\mathrm{erf}(x)$ は誤差函数 (error function) であり，

$$\mathrm{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-t^2\right)\,\mathrm{d}t \tag{1.1.15}$$

として定義されます。

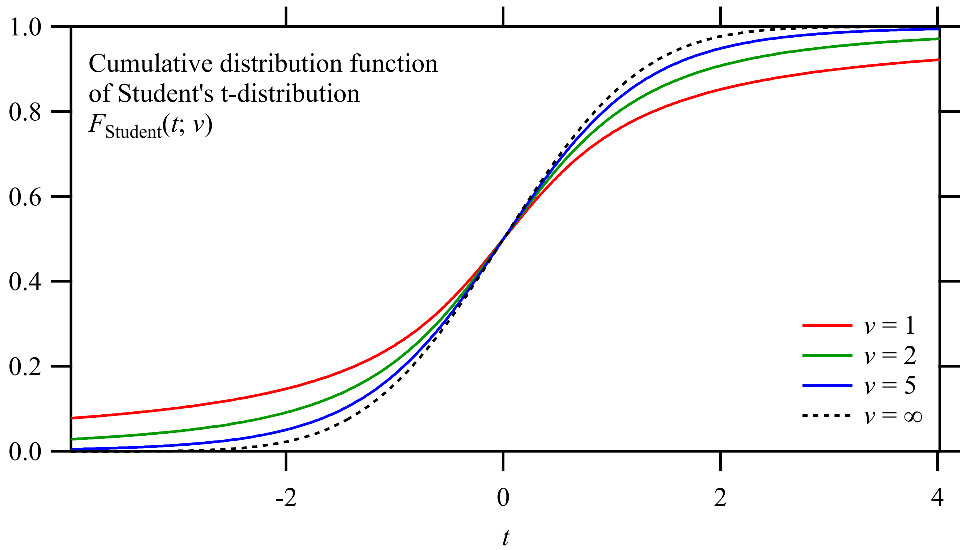

Figure 1.1.2　自由度 $\nu = 1, 2, 5, \infty$ の Student t 分布の累積分布函数

1-2　ベイズ推定の文脈による t 検定の解釈

「伝統的な t 検定の考え方・表現」には，わかりにくく受け入れにくいところも多いようです。ここでは「**ベイズ推定** Bayesian inference の考え方をあてはめて，解釈を試みます。

検定統計量 t の**真値**を t_{true} （ティー・トゥルー）として，「$t_{\mathrm{true}} = 0$ のときにデータから計算される統計量 t の統計分布」の確率密度函数が $f_{\mathrm{Student}}(t;\nu)$ と表されるとします。もし $t_{\mathrm{true}} \neq 0$ であれば，「データから計算される統計量 t 」の統計分布の確率密度函数は $f_{\mathrm{Student}}\left(t - t_{\mathrm{true}};\nu\right)$ と表されます。真値の統計分布の確率密度函数が $f_{\mathrm{true}}(t_{\mathrm{true}})$ と表されるなら，統計量 t の分布の密度函数 $f(t)$ は

$$f(t) = \int_{-\infty}^{\infty} f_{\mathrm{true}}\left(t_{\mathrm{true}}\right) f_{\mathrm{Student}}\left(t - t_{\mathrm{true}}\right)\, \mathrm{d}t_{\mathrm{true}} \tag{1.2.1}$$

と表されます。この関係を「検定統計量 t の分布は，密度函数 $f_{\mathrm{true}}(t)$ で表される確率分布と，密度函数 $f_{\mathrm{Student}}(t;\nu)$ で表される確率分布の**畳込** (convolution) として表される」と言い換えることもできます。

現実のデータから得られた検定統計量が $t = t_{\mathrm{obs}}$ （ティー・オブス）という値（obs は「観測された」 observed の意味）だとして，真値の統計分布の確率密度 $f_{\mathrm{true}}(t_{\mathrm{true}})$ は求められるでしょうか？「真値は測定や観測とは無関係に，絶対的に正しい値として存在しているはずなので，確率論的に解釈しようとすること自体が間違っている」と思う人もいます。そのような考え方は**頻度主義** frequentism と呼ばれ，そのように考える人は**頻度主義者** frequentist と呼ばれる場合があります。

「**ベイズ推定** Bayesian inference」の考え方では「『実験や観測，調査によって得られたデータ』から『真値を確率論的・統計学的に推定することはできる』」とします。もしそのことを受け入れられなければ，実験や観測，調査などをすることに「意味」を持たせにくくなるでしょう。

ベイズ推定 (Bayesian inference) の文脈に従って，以下のように言い換えることにします。

検定統計量の真値 t_{true} の**事前確率** (prior probability) の密度函数が $f_{\mathrm{prior}}(t_{\mathrm{true}})$ と表されるとします。この函数は式 (1.2.1) で $f_{\mathrm{true}}(t_{\mathrm{true}})$ と表した函数と同じことです。

検定統計量の真値が t_{true} という値をとったときに検定統計量の観測値が t_{obs} という値を取る「**条件付き確率** (conditional probability)」，**尤度函数**（likelihood function) の密度函数は $f_{\mathrm{likelihood}}\left(t_{\mathrm{obs}} \mid t_{\mathrm{true}}\right) = f_{\mathrm{Student}}\left(t_{\mathrm{obs}} - t_{\mathrm{true}};\nu\right)$ と表されます。

検定統計量が t_{obs} という値を取る「**エビデンス**（証拠） (evidence)」の確率，**周辺 尤度** (marginal likelihood) の密度函数を $f_{\mathrm{evidence}}(t_{\mathrm{obs}})$ とします。

観測された検定統計量が t_{obs} という値を取ったときに，検定統計量の真値が t_{true} という値を取る（と推定される）条件付き確率，**事後確率** (posterior probability) の密度函数を $f_{\mathrm{posterior}}(t_{\mathrm{true}} \mid t_{\mathrm{obs}})$ と表します。

検定統計量の真値が t_{true} という値を取り，それと同時に観測された検定統計量が t_{obs} という値を取る確率，**同時確率** (joint probability) の密度 $f_{\mathrm{joint}}(t_{\mathrm{true}}, t_{\mathrm{obs}})$ は

$$f_{\mathrm{joint}}(t_{\mathrm{true}}, t_{\mathrm{obs}}) = f_{\mathrm{likelihood}}\left(t_{\mathrm{obs}} \mid t_{\mathrm{true}}\right) f_{\mathrm{prior}}(t_{\mathrm{true}}) \tag{1.2.2}$$

と表されることと，

$$f_{\mathrm{joint}}(t_{\mathrm{true}}, t_{\mathrm{obs}}) = f_{\mathrm{posterior}}\left(t_{\mathrm{true}} \mid t_{\mathrm{obs}}\right) f_{\mathrm{evidence}}(t_{\mathrm{obs}}) \tag{1.2.3}$$

とも表されることから，**事後確率** (posterior probability) $f_{\mathrm{posterior}}(t_{\mathrm{true}} \mid t_{\mathrm{obs}})$ の密度函数は

$$f_{\mathrm{posterior}}\left(t_{\mathrm{true}} \mid t_{\mathrm{obs}}\right) = \frac{f_{\mathrm{likelihood}}(t_{\mathrm{obs}} \mid t_{\mathrm{true}}) f_{\mathrm{prior}}(t_{\mathrm{true}})}{f_{\mathrm{evidence}}(t_{\mathrm{obs}})} \tag{1.2.4}$$

と書けます。式 (1.2.2) と式 (1.2.3)，あるいは式 (1.2.4) で表される関係は「**ベイズの定理** (Bayes' theorem)」と呼ばれます。

また**エビデンス**の確率密度 $f_{\mathrm{evidence}}(t_{\mathrm{obs}})$ は

$$f_{\mathrm{evidence}}(t_{\mathrm{obs}}) = \int_{-\infty}^{\infty} f_{\mathrm{likelihood}}\left(t_{\mathrm{obs}} \mid t_{\mathrm{true}}\right) f_{\mathrm{prior}}\left(t_{\mathrm{true}}\right)\, \mathrm{d}t_{\mathrm{true}} \tag{1.2.5}$$

と書けます。式 (1.2.4) と式 (1.2.5) から，**事後確率** (posterior probability) $f_{\mathrm{posterior}}(t_{\mathrm{true}} \mid t_{\mathrm{obs}})$ の密度は

$$f_{\mathrm{posterior}}\left(t_{\mathrm{true}} \mid t_{\mathrm{obs}}\right) = \frac{f_{\mathrm{likelihood}}(t_{\mathrm{obs}} \mid t_{\mathrm{true}}) f_{\mathrm{prior}}(t_{\mathrm{true}})}{\int_{-\infty}^{\infty} f_{\mathrm{likelihood}}(t_{\mathrm{obs}} \mid t_{\mathrm{true}}) f_{\mathrm{prior}}(t_{\mathrm{true}})\, \mathrm{d}t_{\mathrm{true}}} \tag{1.2.6}$$

と書きなおせます。

式 (1.2.6) で表されるように，一般的に**ベイズ推定** Bayesian inference では，**尤度函数** (likelihood function) の密度 $f_{\mathrm{likelihood}}\left(t_{\mathrm{obs}} \mid t_{\mathrm{true}}\right)$ がわかれば，「仮定した**事前確率** (prior probability) の密度函数 $f_{\mathrm{prior}}(t_{\mathrm{true}})$ から，**事後確率** (posterior probability) の密度函数 $f_{\mathrm{posterior}}(t_{\mathrm{true}} \mid t_{\mathrm{obs}})$ が導かれる」という関係を使います。

尤度函数 (likelihood function) の密度 $f_{\mathrm{likelihood}}\left(t_{\mathrm{obs}} \mid t_{\mathrm{true}}\right)$ が Student t 分布の密度函数 $f_{\mathrm{Student}}(t; \nu)$ を使えば

$$f_{\mathrm{likelihood}}\left(t_{\mathrm{obs}} \mid t_{\mathrm{true}}\right) = f_{\mathrm{Student}}\left(t_{\mathrm{obs}} - t_{\mathrm{true}}; \nu\right) \tag{1.2.7}$$

と表せることから，**事後確率** (posterior probability) の密度 $f_{\mathrm{posterior}}\left(t_{\mathrm{true}} \mid t_{\mathrm{obs}}\right)$ を表す式 (1.2.6) は

$$f_{\text{posterior}}\left(t_{\text{true}} \,|\, t_{\text{obs}}\right) = \frac{f_{\text{Student}}(t_{\text{obs}} - t_{\text{true}}; \nu) f_{\text{prior}}(t_{\text{true}})}{\int_{-\infty}^{\infty} f_{\text{Student}}(t_{\text{obs}} - t_{\text{true}}; \nu) f_{\text{prior}}(t_{\text{true}})\, \mathrm{d}t_{\text{true}}} \tag{1.2.8}$$

と書き直せます。

そして「**ベイズ推定** Bayesian inference」と呼ばれることのうちのトリッキー (tricky) かもしれないことに，「『事前情報』がまったくない場合には，**事前確率** (prior probability) の分布は**一様分布** (uniform distribution) と仮定する」ことがあります。つまり $f_{\text{prior}}(t_{\text{true}}) = c$ （c は t_{true} によらない定数）とします。確率変数が有界 (bounded) でなければ函数 $f_{\text{prior}}(t_{\text{true}})$ は普通の意味では規格化できないことになりますが，$c \to 0$ とすれば，辻褄を合わせることはできます。

一様事前分布仮定をおけば，**事後確率** (posterior probability) を表す式 (1.2.8) は

$$f_{\text{posterior}}\left(t_{\text{true}} \,|\, t_{\text{obs}}\right) = \frac{f_{\text{Student}}(t_{\text{obs}} - t_{\text{true}}; \nu)}{\int_{-\infty}^{\infty} f_{\text{Student}}(t_{\text{obs}} - t_{\text{true}}; \nu)\, \mathrm{d}t_{\text{true}}} = f_{\text{Student}}(t_{\text{obs}} - t_{\text{true}}; \nu) \tag{1.2.9}$$

となり，Student t 分布の密度函数 $f_{\text{Student}}\,(t; \nu)$ を使って表すことができます。式 (1.2.7) と式 (1.2.9) とを比べればわかるように，「一様事前分布仮定」をおけば，式 (1.2.9) で表される事後確率 (posterior probability) は，式 (1.2.7) で表される尤度函数 (likelihood function) と「同一の数式」で表されます。このことも「ベイズ推定」と呼ばれることをわかりにくくする要因の一つかもしれません。

Student t 分布が「検定統計量の真値 t_{true} が $t_{\text{true}} = 0$ であるとした場合の観測検定統計量 t_{obs} の分布」である一方で，一様事前分布仮定をおいたベイズ推定の考え方では，式 (1.2.9) で示すように，「観測値 t_{obs}を前提として，推定される真値 t_{true} の分布」も Student t 分布を使って表されます。

Figure 1.2.1 に「伝統的な t 検定」で使われる「**片側検定** (one-sided test; one-tailed test)」の考え方を示します。検定統計量 $\hat{t}$ が正 ($\hat{t} > 0$) の値を取った時に，「片側検定」の場合には「**上側確率** upper probability」

$$P(t > \hat{t}\,) = \int_{\hat{t}}^{\infty} f_{\text{Student}}(t; \nu)\, \mathrm{d}t = 1 - F_{\text{Student}}(\hat{t}; \nu) \tag{1.2.10}$$

を P 値として計算します。そして $P(t > \hat{t}\,)$ の値が例えば 5 % 以下だったとしたら，「有意水準 5 % で帰無仮説は棄却された」と言います。

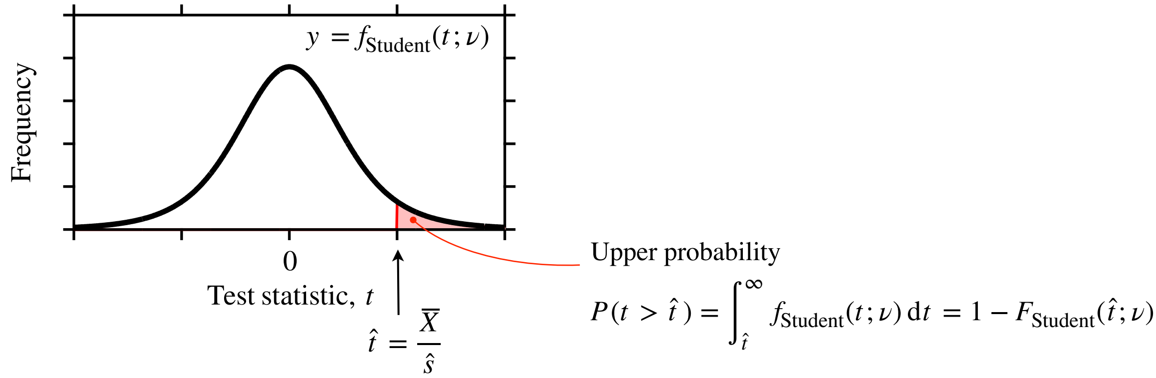

Figure 1.2.1　片側検定 (one-sided test) での検定統計量 t と，Student t 分布の確率密度関数 (probability density function) $f_{\mathrm{Student}}(t; \nu)$，上側確率 (upper probability) $P(t > t_0)$ の関係。

このとき棄却される帰無仮説は「どのような意味の仮説」で，5 % という数字が確率を意味するものであれば，その 5% とは「どのような確率」を意味するものでしょうか？このことは，ベイズ推定の考え方を取り入れないと，理解しにくいようです。

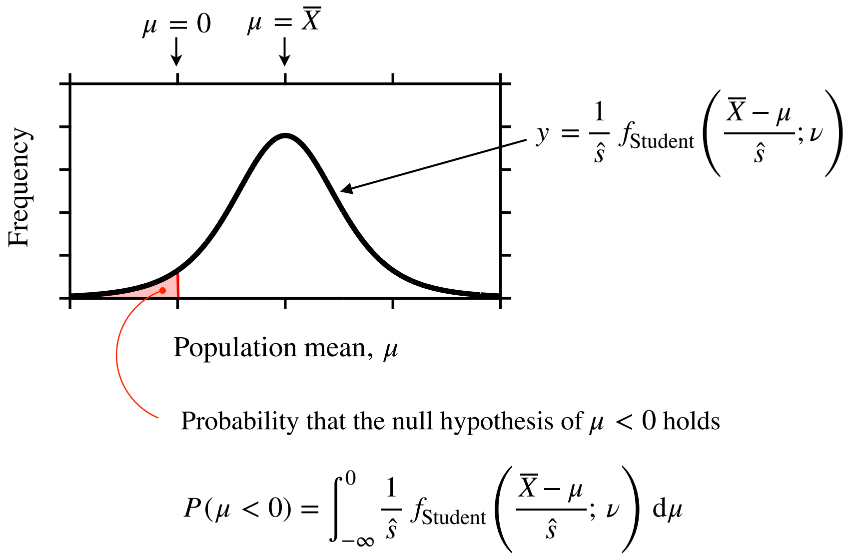

Figure 1.2.2　片側 t 検定 (one-sided t-test) のベイズ推定の文脈での解釈。Figure 1.2.1 で示したグラフに尺度パラメータ $\hat{s}$ を使った尺度変換をしてから，水平方向に $\overline{X}$ 位置をずらして左右反転させた図形が「データに基づいて推定される真の母平均値 μ の統計分布」（事後確率）という意味になる。

ベイズ推定の文脈で式 (1.2.9) のように表される「**事後確率** (posterior probability)」の確率密度函数を Figure 1.2.2 に示します。片側 t 検定での「上側確率」は「推定された母平均値 μ が 0 以下になる確率」に相当することがわかります。

医薬品の効果についての検定の場合であれば，このように片側検定に相当するベイズ推定の考え方から「効果の真値として推定される値がゼロ以下なら無効，ゼロ以上なら有効」として「無効率は 5% 以下」あるいは「有効率は 95% 以上」などと言うと考えれば良いでしょう。

医薬品の効果に関する t 検定の説明で，「『効果がない』ことを帰無仮説とする検定が行われる」との文言を多くみますが，その表現は正確ではありません。片側 t 検定を使う場合には「『負の効果がある』ことを帰無仮説とする」と言う方が正確です。医薬品に「負の効果がある」というのは，「その医薬品を投与した方が，プラセボ（偽薬）(placebo) を投与するより治癒が遅くなる（退院や通院に必要な日数が多くなる）」と言うようなことであり，それはゼロでない確率で出現しうる現象です。

医薬品の効果に関する検定には，「**両側検定** (two-sided test; two-tailed test)」と呼ばれる方法の用いられる例が多かったようです（青木, 1982）。

Figure 1.2.3 に，「両側検定」の（無意味な）考え方を示します。検定統計量 $\hat{t}$ が正 $(\hat{t} > 0)$ の値を取った時に，「両側検定」の場合でも式 (1.2.9) に示した「**上側確率** upper probability」$P(t > \hat{t}\,)$ が

$$P(t > \hat{t}\,) = \int_{\hat{t}}^{\infty} f_{\text{Student}}(t;\nu)\,\mathrm{d}t = 1 - F_{\text{Student}}(\hat{t};\nu) \tag{1.2.9}$$

を計算することは変わりません。検定統計量の分布が Student t 分布であると仮定すれば，以下の形式で表現される「**下側確率** lower probability」$P(t < -\,\hat{t}\,)$

$$P(t < -\,\hat{t}\,) = \int_{-\infty}^{-\hat{t}} f_{\text{Student}}(t;\nu)\,\mathrm{d}t = F_{\text{Student}}(-\hat{t};\nu) \tag{1.2.10}$$

は上側確率と等しく，$P(t < -\,\hat{t}\,) = P(t > \hat{t}\,)$ の関係は常に成立します。「両側 t 検定」の場合に起こることは，実質的には「有意水準」あるいは「P 値」を「２倍にするだけ」のことです。

「それでも両側検定をしようとする人」は，$P(t > \hat{t}\,) + P(t < -\,\hat{t}\,)$ の値が 5 % 以下のときに，「有意水準 5 % で帰無仮説は棄却された」と言います。

このとき棄却される帰無仮説は「どのような意味の仮説」で，たとえば 5 % という P 値が確率を意味するものであれば「どのような確率」を意味するものでしょうか？

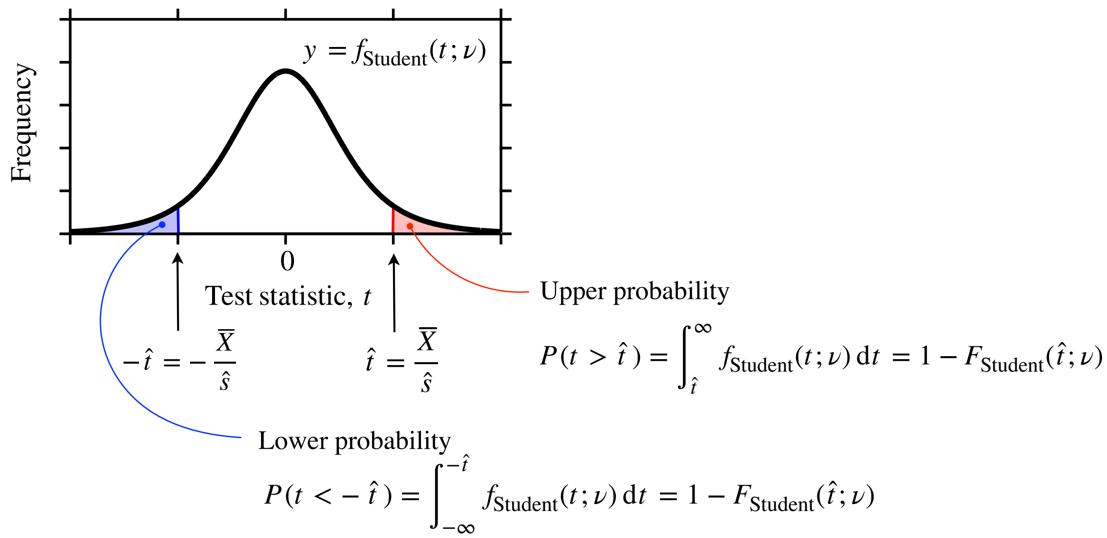

Figure 1.2.3　両側検定 (two-sided test) での検定統計量 t と，Student t 分布の確率密度函数 (probability density function) $f_{\mathrm{Student}}(t; \nu)$，上側確率 (upper probability) $P(t > \hat{t})$，下側確率 (lower probability) $P(t < -\hat{t})$ の関係。

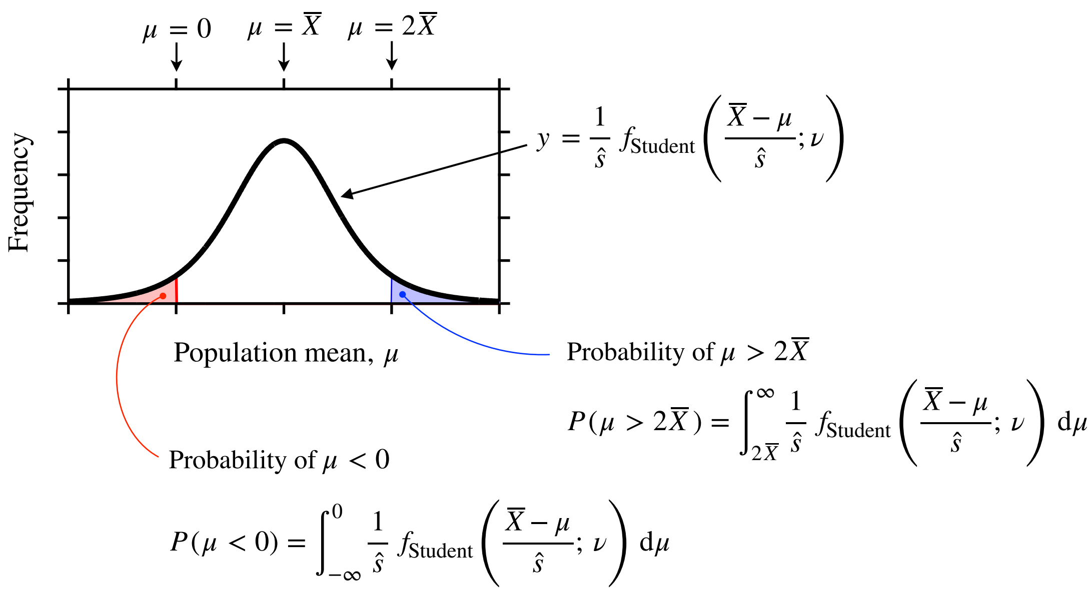

Figure 1.2.4　両側 t 検定 (two-sided t-test) のベイズ推定の文脈での解釈。Figure 1.2.3 で示したグラフを左右反転させ，水平方向に $\overline{X}$ 位置をずらした図形が「データに基づいて推定される真値の統計分布」（事後確率）という意味になる。ピンクで塗られた図形の面積が両側 t 検定での「上側確率」，薄青で塗られた図形の面積が両側 t 検定での「下側確率」に対応する。

ベイズ推定の文脈で，式 (1.2.7) のように表される「**事後確率** (posterior probability)」の確率密度函数と上側確率，下側確率との関係を Figure 1.2.4 に示します。両側 t 検定での「上側確率」が「推定された母平均値が 0 以下の値になる確率」であることには変わりありませんが，「下側確率」は「推定された母平均値が $2\overline{X}$ 以上の値になる確率」に相当します。

医薬品の効果に関して両側 t 検定を行うことをベイズ推定の文脈で解釈すれば「『負の効果がある』か『代表的な効果の２倍以上の効果がある』ことを帰無仮説とする」ことになります。「代表的な効果の２倍以上の効果が現れる」ことも一定の確率で出現する現象ですが，両側 t 検定を用いる人は，そのことも棄却の対象（帰無仮説）に含めてしまいます。例えば「そのように劇的な効果が現れるというのは，（これも一定の割合で出現する）虚偽報告・捏造によるかもしれない」という考え方によるかもしれませんが，それは「まったく文脈の異なること」を「ごちゃ混ぜ」にして，他の人を心理的な混乱状態に誘導しようとしているだけのように見えます。

「**両側検定**」と呼ばれることに，普通に「意味のあること」をしているように見える例は，事実上存在しません。しかし，後述するように「二標本 t 検定」と呼ばれる方法を使うと採択（認可）の基準が不当に甘くなりすぎるので，「両側検定」により採択基準を厳しく設定することに実際には意味のある例があるのかもしれません。（補足 1.2.A）

そのこととは別のこととして「違わない・同じ」ことを帰無仮説とすると，必然的に論理的な破綻を招くことになります。「$A < B$ を帰無仮説とするときに『上側検定』を使い，$A > B$ を帰無仮説をとするときに『下側検定』を使う。$A = B$ を帰無仮説とするときには『両側検定』を使う」のように書かれる場合も多いのですが，この文言が決定的に間違っていることは，本来ならば中学生でもわかるはずのことです。「連続的で特異性 (singularity) を持たない統計分布」であれば「違わない ($A = B$) 確率」は必ずゼロになり，「違う ($A \neq B$) 確率」は必ず 1 になります。一標本 t 検定で「$A = 0$ を帰無仮説とする」とするのも，二標本 t 検定で「$A = B$ を帰無仮説とする」のも無意味です。

1−3　Python 言語と SciPy ライブラリを使った t テスト

1−3−1　コンピュータを使ったステューデント t テストの考え方

現在でも書籍やインターネット等で「統計学的な検定」について得られる情報には，「コンピュータを使えなかった時代の話」の影響が強く残ります。ここではコンピュータを使って，現代的な方法で「ステューデント t テスト（t 検定のようなこと）」をする手順についてまとめます。

データの母平均 μ が，ある基準値（参照値）$\mu_{\rm ref}$ より大きいか，小さいかを調べる目的で「一標本ステューデント t テスト」をすることについて考えます。母平均 μ が正か負かを調べる場合には，参照値 $\mu_{\rm ref}$ としてゼロを用います。

尺度化参照値 $t_{\rm ref}$ と検定統計量 (test statistic) t とを

$$t_{\rm ref} = \frac{\overline{X} - \mu_{\rm ref}}{\hat{s}} \quad \Leftrightarrow \quad \mu_{\rm ref} = \overline{X} - \hat{s}\, t_{\rm ref}$$

$$t = \frac{\overline{X} - \mu}{\hat{s}} \quad \Leftrightarrow \quad \mu = \overline{X} - \hat{s}\, t = \mu_{\rm ref} - \hat{s}\left(t - t_{\rm ref}\right)$$

とします。

「**上側 P 値** (upper-sided P value)」$P(t > t_{\mathrm{ref}})$ は

$$P(t > t_{\mathrm{ref}}) = \int_{t_{\mathrm{ref}}}^{\infty} f_{\mathrm{Student}}(t;\, \nu)\, \mathrm{d}t$$

と表され，

$$\begin{aligned} \mathrm{d}t &= -\frac{\mathrm{d}\mu}{\hat{s}} \\ t \ &: \ t_{\mathrm{ref}} \ \rightarrow \ \infty \\ \mu \ &: \ \mu_{\mathrm{ref}} \ \rightarrow \ -\infty \end{aligned}$$

の関係から

$$P(t > t_{\mathrm{ref}}) = \int_{-\infty}^{\mu_{\mathrm{ref}}} \frac{1}{\hat{s}} f_{\mathrm{Student}}\left(\frac{\overline{X} - \mu}{\hat{s}};\, \nu\right) \mathrm{d}\mu = P(\mu < \mu_{\mathrm{ref}})$$

とも書けて，上側 P 値は「$\mu < \mu_{\mathrm{ref}}$ となる確率」の意味にもなります。また，

$$P(t > t_{\mathrm{ref}}) = 1 - F_{\mathrm{Student}}(t_{\mathrm{ref}}; \nu) = \begin{cases} \dfrac{1}{2} I\left(\dfrac{\nu}{2}, \dfrac{1}{2}; \dfrac{\nu}{t_{\mathrm{ref}}^2 + \nu}\right) & [t_{\mathrm{ref}} \geq 0] \\ 1 - \dfrac{1}{2} I\left(\dfrac{\nu}{2}, \dfrac{1}{2}; \dfrac{\nu}{t_{\mathrm{ref}}^2 + \nu}\right) & [t_{\mathrm{ref}} < 0] \end{cases} \tag{1.3.1.1}$$

と表されます。ここで函数 $I(a, b; x)$ は正則不完全ベータ函数 (regularized incomplete beta function) ですが，例えば Python 言語と SciPy ライブラリを使えば `scipy.special.betainc(a,b,x)` として参照できます。

「上側 P 値」$P(t > t_{\mathrm{ref}})$ を使うとき，帰無仮説は「統計量 t が尺度化参照値 t_{ref} より大きい値をとる」こと，「母平均 μ が 参照値 μ_{ref} より小さい値をとる」ことに対応します。

標本平均 $\overline{X}$ と標本標準偏差 $\hat{\sigma}$，自由度 ν（必要があれば有限の参照値 μ_{ref}）の値が決まれば，**P 値**が確定します。ステューデント t テスト（t 検定）のためにコンピュータを使えるようになってからは，「ステューデント t テストをしたら，（「有意水準」と「棄却／採択」について「書いてはいけないわけではない」が）P 値は必ず示す」ことが常識的になりました。コンピュータを使えなかった過去の時代の人は，「…という帰無仮説をたてて t 検定をした結果，有意水準 5% で棄却された」のように記述していました。現代の t テストでは，例えば「…という帰無仮説をたてて t テストをした結果，P 値が 4.3% になった」のように記述されます。

両側検定 (two-sided test) を用いるとき「上側 P 値」は $P(t > |t_{\mathrm{ref}}|)$ と表され，その解釈は混乱します。参照統計量 $t_{\mathrm{ref}} = \dfrac{\overline{X} - \mu_{\mathrm{ref}}}{\hat{s}}$ が正の場合（標本平均 $\overline{X}$ が参照値 μ_{ref} より大きい場合）に「母平均 μ が参照値 μ_{ref} より小さい値をとる確率 $P\left(\mu < \mu_{\mathrm{ref}}\right)$」を表すことは変

わりませんが，参照統計量 t_{ref} が負の場合（標本平均 $\overline{X}$ が参照値 μ_{ref} より小さい場合）には

$$t > -t_{\mathrm{ref}} \quad \Rightarrow \quad \frac{\overline{X}-\mu}{\hat{s}} > -\frac{\overline{X}-\mu_{\mathrm{ref}}}{\hat{s}} \quad \Rightarrow \quad \mu < 2\overline{X}-\mu_{\mathrm{ref}}$$

という条件を満たす確率「母平均 μ が標本平均 $\overline{X}$ の２倍と参照値 μ_{ref} の差より小さい値をとる確率 $P\left(\mu < 2\overline{X}-\mu_{\mathrm{ref}}\right)$」 を表すことになります。

下側(したがわ) P(ピー) 値(ち) (lower(ロウワ)-sided(サイデド) *P*(ピー) value(ヴァリュー)) を $P(t < -|t_{\mathrm{ref}}|)$ と表したとき，その解釈は混乱します。$t_{\mathrm{ref}} = \dfrac{\overline{X}-\mu_{\mathrm{ref}}}{\hat{s}}$ の値が負であれば

$$P(t < -|t_{\mathrm{ref}}|) = P(t < t_{\mathrm{ref}}) = \int_{-\infty}^{t_{\mathrm{ref}}} f_{\mathrm{Student}}(t;\,\nu)\,\mathrm{d}t = F_{\mathrm{Student}}(t_{\mathrm{ref}};\nu)$$
$$= \int_{\mu_{\mathrm{ref}}}^{\infty} \frac{1}{\hat{s}} f_{\mathrm{Student}}\left(\frac{\overline{X}-\mu}{\hat{s}};\,\nu\right)\mathrm{d}\mu \tag{1.3.1.2}$$

のように表され「母平均 μ が μ_{ref} より大きい確率 $P\left(\mu > \mu_{\mathrm{ref}}\right)$」を意味することになりますが，$t_{\mathrm{ref}} = \dfrac{\overline{X}-\mu_{\mathrm{ref}}}{\hat{s}}$ の値が正の場合には $t = \dfrac{\overline{X}-\mu}{\hat{s}} \quad \Leftrightarrow \quad \mu = \mu_{\mathrm{ref}} - \hat{s}\left(t - t_{\mathrm{ref}}\right)$

$$\begin{aligned} \mathrm{d}t &= -\frac{\mathrm{d}\mu}{\hat{s}} \\ t &: \ -\infty \ \rightarrow \ -t_{\mathrm{ref}} \\ \mu &: \ \infty \ \rightarrow \ 2\overline{X}-\mu_{\mathrm{ref}} \end{aligned}$$

の関係から

$$P(t < -|t_{\mathrm{ref}}|) = P(t < -t_{\mathrm{ref}}) = \int_{-\infty}^{-t_{\mathrm{ref}}} f_{\mathrm{Student}}(t;\,\nu)\,\mathrm{d}t$$
$$= \int_{2\overline{X}-\mu_{\mathrm{ref}}}^{\infty} \frac{1}{\hat{s}} f_{\mathrm{Student}}\left(\frac{\overline{X}-\mu}{\hat{s}};\,\nu\right)\mathrm{d}\mu \tag{1.3.1.2}$$

のようになり，「母平均 μ が $2\overline{X}-\mu_{\mathrm{ref}}$ より大きい確率」を意味することになります

両側 P(ピー) 値(ち) (two-sided(サイデド) *P*(ピー) value(ヴァリュー)) $P(t < -|t_{\mathrm{ref}}| \text{ or } t > |t_{\mathrm{ref}}|)$ は，

$$\begin{aligned} &P(t < -|t_{\mathrm{ref}}| \text{ or } |t_{\mathrm{ref}}| < t) \\ &= P(t < -|t_{\mathrm{ref}}|) + P(|t_{\mathrm{ref}}| < t) \\ &= 2P(t < -|t_{\mathrm{ref}}|) \\ &= 2P(t > |t_{\mathrm{ref}}|) \\ &= I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t_{\mathrm{ref}}^2+\nu}\right), \end{aligned} \tag{1.3.1.4}$$

のように表現され「$t_{\mathrm{ref}} > 0$ のときの上側 P 値の２倍」「$t_{\mathrm{ref}} < 0$ のときの下側 P 値の２倍」とも言えます。

「両側 P 値（two-sided P value）」を「母平均 μ のとりうる値」に対応づければ，$t_{\mathrm{ref}} > 0$ ($\overline{X} > \mu_{\mathrm{ref}}$)のとき「$\mu < \mu_{\mathrm{ref}}$ または $\mu > 2\overline{X} - \mu_{\mathrm{ref}}$となる確率」を意味し，$t_{\mathrm{ref}} < 0$ のとき「$\mu > \mu_{\mathrm{ref}}$ または $\mu < 2\overline{X} - \mu_{\mathrm{ref}}$となる確率」を意味します

「両側 P 値（two-sided P value）」$P(t < -|t_{\mathrm{ref}}| \text{ or } t > |t_{\mathrm{ref}}|)$ の値をそのまま使うことは，基本的には不合理です。

コンピュータを使えなかった過去の時代には，検定統計量が正の値だった場合に有意水準 c として帰無仮説が棄却されるのは，

$$F(t) = \int_{-\infty}^{t} f(u)\,\mathrm{d}u > 1 - c \tag{1.3.1.4}$$

の場合，あるいは $1 - F(t) - c < 0$ の場合でした。つまり，

$$\frac{1}{2} I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t^2 + \nu}\right) - c < 0 \tag{1.3.1.5}$$

の時に帰無仮説が棄却 (reject) され，

$$\frac{1}{2} I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t^2 + \nu}\right) - c \geq 0 \tag{1.3.1.6}$$

の場合に帰無仮説は棄却 (reject) できなかったことになります。

1−3−2　Python 言語と SciPy ライブラリを使った t テスト

ここではコンピュータのプログラミングをするために良く使われる Python 言語と，Python 言語から呼び出すことのできる科学技術計算ライブラリである SciPy ライブラリを使います。

SciPy ライブラリの `special` モジュール (module) には，科学技術計算で用いられることの多い特殊函数 (special function) が多く収められており，正則不完全ベータ函数 (regularized incomplete beta function) $I(a, b; x)$は “`scipy.special.betainc(a,b,x)`” という名前の函数（メソッド method）として利用できて，その逆函数 $I^{-1}(a, b; y)$ は “`scipy.special.betaincinv(a,b,y)`” という名前の函数（メソッド method）として利用できます。

また，SciPy ライブラリの `stats` モジュールには，統計解析に利用できるメソッドも提供されており，「一標本 t 検定 (one-sample t-test)」のためのメソッド（函数）として

```
stats.ttest_1samp(a, popmean[,axis, nan_policy, alternative])
```

が提供されます。基本的な使い方では，第一引数の `a` に 配列型の**「標本値（データ）」** $\{X_1, X_2, \cdots, X_n\}$ を入力し，第二引数の `popmean` に float 型（浮動小数点型）の母平均基準値 μ_{ref} を入力します。第二引数の `popmean` は「母平均」(population mean) という意味で

はないということが混乱を招くかもしれません。函数の戻り値 (return value) は，**統計量** (`statistic`) t の値と両側検定としたときの **P 値** (`p-value`)（有意確率）の値とを合わせた Python タプル (tuple) です。

`ttest_1samp()` メソッドは，デフォルト（省略時設定）では両側検定に対応する P 値（両側 P 値）が出力されることに注意する必要があります。

オプション引数で `alternative='greater'` を明示的に指定すれば，「**P 値**」として，上側確率の値が得られます。（補足 1.3.2.A）。

1–3–3　t 検定での言い回し・有意水準

一標本に限らず「 t 検定の文脈」では，例えば「5% の有意水準で帰無仮説が棄却された」"The null hypothesis is rejected for a significance level of 5%" などと言う表現がとられます。この表現の意味することは，単純化すれば「帰無仮説の正しい確率は 5% 以下である」（ほぼ間違っている）というようなことです（補足 1.3.3.A）。

検定統計量 t の統計分布（t 分布）の確率密度函数 $f(t)$ は式 (1.1.6)：

$$f_{\mathrm{Student}}(t;\nu)=\frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\dfrac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{1.1.6}$$

で表されるように $f(t)=f(-t)$ の関係の成立する偶函数 (even function) です。

累積分布函数 (cumulative distribution function) は

$$F_{\mathrm{Student}}(t;\nu)=\int_{-\infty}^{t} f_{\mathrm{Student}}(u;\nu)\,\mathrm{d}u \tag{1.3.3.1}$$

と表現されます。累積分布函数 $F_{\mathrm{Student}}(t;\nu)$ は，**検定統計量** t が増えれば単調に増加する函数であり， $t\to-\infty$ のとき $F_{\mathrm{Student}}(t;\nu)\to 0$， $t\to\infty$ のとき $F_{\mathrm{Student}}(t;\nu)\to 1$ となります。

正の**検定統計量** t の値に対して，「 t が $-t_0$ より小さい値を取る確率」と「 t が t_0 より大きい値を取る確率」は等しく，この関係を

$$\int_{-\infty}^{-t_0} f_{\mathrm{Student}}(u;\nu)\,\mathrm{d}u=\int_{t_0}^{\infty} f_{\mathrm{Student}}(u;\nu)\,\mathrm{d}u \tag{1.3.3.2}$$

あるいは

$$F_{\mathrm{Student}}(-t_0;\nu)=1-F_{\mathrm{Student}}(t_0;\nu) \tag{1.3.3.3}$$

と書けます。「**検定統計量** t の値が t_{c} より大きくなる確率」の値

$$P(t>t_{\mathrm{c}})=F_{\mathrm{Student}}(-t_{\mathrm{c}};\nu)=1-F_{\mathrm{Student}}(t_{\mathrm{c}};\nu) \tag{1.3.3.4}$$

は「**P 値** (p-value)」「有意確率」などと呼ばれます「**有意 水準** (significant level)」とは「有意確率より大きくてキリの良い数（1%, 2%, 5%, ... など）を選ぶこと」と考えても良いでしょう。

逆に「**有意水準**」の値 c を決めたときに，t_c の値には「**検定統計 量** t の値が t_c より大きくなる確率が c となるような値」という意味付けがされます。この t_c の値が「臨界値 (critical value)」と呼ばれることもあります。

1–4　t 分布表をつくる

1–4–1　t 分布表を作る考え方

古い「統計学の教科書」や古い「web ページ」などに，「t 検定をするときには t 分布表を使う」と書かれている場合が多いでしょう。懐古趣味 (nostalgia) を持つ人は，古い教科書の指示にしたがって「コンピュータを使えなかった時代の人は，どのように t 検定をしていたか」を経験してみても良いと思います（補足 1.4.1.A）。

ここでは片側検定の有意水準として 0.05, 0.025, 0.005 の 3 通り（「両側検定」の有意水準 0.10, 0.05, 0.01 と同じ），自由度 ν としては 1 から 10 までの値に対するt 分布表を作ることにします。

「コンピュータを使える現代人」は「コンピュータを使った t テスト」をすれば良く，実際には「t 分布表を使った t 検定」をする必要はありません。ただし t テスト（t 検定）と呼ばれることについての理解のしかたや計算のしかた，解釈のしかたが間違っていないかを確認するために，この程度の小規模な t 分布表を実際に作成して「古い統計学の教科書」に記載されている値と矛盾していないか確認することには，意味があります。一方で，これ以上大規模な表を作っても，そのこと自体にはあまり意味がないかもしれません。

t 分布表を作るために必要な式は，式 (1.3.3.4) に示した

$$c = 1 - F_\mathrm{Student}(t_\mathrm{c}; \nu) \tag{1.4.1.1}$$

と，式 (1.1.10) に示した関係から

$$F_\mathrm{Student}(t_\mathrm{c}; \nu) = \int_{-\infty}^{t_\mathrm{c}} f_\mathrm{Student}(t; \nu)\,\mathrm{d}t = 1 - \frac{1}{2} I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t_\mathrm{c}^2 + \nu}\right) \tag{1.4.1.2}$$

導かれる

$$c = \frac{1}{2} I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t_\mathrm{c}^2 + \nu}\right) \quad \Rightarrow \quad t_\mathrm{c} = \left\{\nu\left[\frac{1}{I^{-1}(\nu/2, 1/2; 2c)} - 1\right]\right\}^{1/2} \tag{1.4.1.3}$$

という数式だけです。

式 (1.4.1.3) 中の $I^{-1}(a,b;y)$ は正規化不完全ベータ函数 (regularized incomplete beta function) $I(a,b;x)$ の逆函数で,

$$y = I(a,b;x) \quad \Leftrightarrow \quad x = I^{-1}(a,b;y) \tag{1.4.1.4}$$

の関係が成立するとします。

1−4−2　Python 言語と SciPy ライブラリを使った t 分布表のつくりかた

ここではコンピュータのプログラミングをするために良く使われる Python 言語と，Python 言語から呼び出すことができる科学技術計算ライブラリである SciPy ライブラリを使って「t 分布表」を作ることにします。

SciPy ライブラリの `special` モジュールには，科学技術計算で用いられることの多い特殊函数 (special function) が多く収められており，正則不完全ベータ函数 $I(a,b;x)$ は `scipy.special.betainc(a,b,x)` という名前のメソッドとして，逆正則不完全ベータ函数 $I^{-1}(a,b;y)$ は `scipy.special.betaincinv(a,b,y)` という名前のメソッドとして呼び出せます。

以下のような Python コードを書いて動作させ，有意水準 (level of significance) c と自由度 (degree of freedom) ν を指定したときに，棄却・採択の境界線となる検定統計量 (test statistic) t の臨界値 (critical value) t_{c} の値が得られるかを調べます。

［t 分布表を作成する Python プログラム t_dist.py］

```
import numpy as np
import scipy.special as sp
def calc_tc(nu, c): # nu: degree of freedom; c: confidence level
    tc = np.sqrt(nu*(1/sp.betaincinv(nu/2,0.5,2*c)-1))
    return tc

print("  c = 0.05   c = 0.025  c = 0.005")
for nu in range(1, 11):
    line = "{:10.4f}".format(calc_tc(nu,0.05))
    line += " {:10.4f}".format(calc_tc(nu,0.025))
    line += " {:10.4f}".format(calc_tc(nu,0.005))
    print(line)
```

[t_dist.py の実行と出力]

```
...$ python3 t_dist.py
  c = 0.05  c = 0.025  c = 0.005
    6.3138    12.7062    63.6567
    2.9200     4.3027     9.9248
    2.3534     3.1824     5.8409
    2.1318     2.7764     4.6041
    2.0150     2.5706     4.0321
    1.9432     2.4469     3.7074
    1.8946     2.3646     3.4995
    1.8595     2.3060     3.3554
    1.8331     2.2622     3.2498
    1.8125     2.2281     3.1693
```

インターネットなどで検索すれば見つかる「t 分布表」と比較して数値に矛盾がなければ，ここでの t 検定についての考え方に大きな間違いはなさそうとわかります。

1-5　二標本 t 検定

1-5-1　二標本 t 検定の考え方

二標本　t　検定 (two-sample t- test) は，「二つの標本（データ）が**独立** (independent) とみなせるとき，二つの標本の**母平均**の大小に関するテストをする場合に用いられる」とされています。

二つの標本集団 $A = \{A_1, \cdots, A_{n_A}\}$，$B = \{B_1, \cdots, B_{n_B}\}$ が母平均を μ_A, μ_B，母分散を σ_A^2, σ_B^2 とする独立な正規分布に従うとします。それぞれの標本の要素数を n_A, n_B，標本平均を $\overline{A}, \overline{B}$，それぞれの標本からの母分散の推定値（標本分散）を $\hat{\sigma}_A^2, \hat{\sigma}_B^2$ とすれば，それぞれの標本平均値の誤差の推定値（標準誤差）は $\hat{s}_A = \hat{\sigma}_A / \sqrt{n_A}$，$\hat{s}_B = \hat{\sigma}_B / \sqrt{n_B}$ となります。

標本平均 $\overline{A}$ と母平均 μ_A の差 $\overline{A} - \mu_A$ の統計分布は，尺度化パラメータ $\hat{s}_A$，自由度 $\nu_A = n_A - 1$ のステューデント t 分布に従い，標本平均 $\overline{B}$ と母平均 μ_B の差 $\overline{B} - \mu_B$ の統計分布は，尺度化パラメータ $\hat{s}_B$，自由度 $\nu_B = n_B - 1$ のステューデント t 分布に従います。

「標本群 A, B が存在しているときに，母分散の差 $\Delta\mu = \mu_A - \mu_B$ がどのような統計分布に従うかを表す確率密度函数 $f(\Delta\mu)$ は，

$$f(\Delta\mu) = \frac{1}{\hat{s}_A \hat{s}_B} \int_{-\infty}^{\infty} f_{\text{Student}}\left(\frac{\Delta\mu + t - \overline{A} + \overline{B}}{\hat{s}_A}; \nu_A\right) f_{\text{Student}}\left(\frac{t}{\hat{s}_B}; \nu_B\right) \mathrm{d}t \qquad (1.5.1.1)$$

のように表現されます。

二組の標本の母平均（の推定値）の差の統計分布を求めることは，単純化して考えれば，「ステューデント t 分布の『畳込』 (convolution) あるいは『（相互）相関』((mutual) correlation) によって表される統計分布」を求めることと同じことです。一般的にはその統計分布は（特別な場合を除いて）「ステューデント t 分布」にはなりません。

たとえば $\Delta\mu < 0$ （$A < B$）となる確率 $P(\Delta\mu < 0)$ を求めるためには，

$$P(\Delta\mu < 0) = \frac{1}{\hat{s}_A\hat{s}_B}\int_{-\infty}^{0}\int_{-\infty}^{\infty} f_{\text{Student}}\left(\frac{\Delta\mu + y - \overline{A} + \overline{B}}{\hat{s}_A};\nu_A\right) f_{\text{Student}}\left(\frac{y}{\hat{s}_B};\nu_B\right) \mathrm{d}y\,\mathrm{d}\Delta\mu$$

$$= \frac{1}{\hat{s}_B}\int_{-\infty}^{\infty}\left[\frac{1}{\hat{s}_A}\int_{-\infty}^{0} f_{\text{Student}}\left(\frac{\Delta\mu + y - \overline{A} + \overline{B}}{\hat{s}_A};\nu_A\right)\mathrm{d}\Delta\mu\right] f_{\text{Student}}\left(\frac{y}{\hat{s}_B};\nu_B\right)\mathrm{d}y$$

$$= \frac{1}{\hat{s}_B}\int_{-\infty}^{\infty}\left[\frac{1}{\hat{s}_A}\int_{-\infty}^{y-\overline{A}+\overline{B}} f_{\text{Student}}\left(\frac{u}{\hat{s}_A};\nu_A\right)\mathrm{d}u\right] f_{\text{Student}}\left(\frac{y}{\hat{s}_B};\nu_B\right)\mathrm{d}y$$

$$= \frac{1}{\hat{s}_B}\int_{-\infty}^{\infty} F_{\text{Student}}\left(\frac{y - \overline{A} + \overline{B}}{\hat{s}_A}\right) f_{\text{Student}}\left(\frac{y}{\hat{s}_B};\nu_B\right)\mathrm{d}y \qquad (1.5.1.2)$$

と表現される式を使った計算をしなければならないはずです。このような積分で表される数式を数値的に解くための方法は知られています。（補足 1.5.1.A）（補足 1.5.1.B）

しかし，以下に述べるように，このような考え方と異なる考え方で「二標本 t 検定」の行われる場合が多いようです。

1−5−2　独立な二標本の母分散が等しいとみなせる場合の t 検定

「独立な二つの標本集団 $A = \{A_1, \cdots, A_{n_A}\}$，$B = \{B_1, \cdots, B_{n_B}\}$ の母平均は異なっているかもしれないが，母分散が等しいと仮定できる」として，以下のような処理の行われる場合があります。この処理のしかたのことが「二標本ステューデント t 検定」のように呼ばれることもあります。

標本平均の差 $\left(\overline{A} - \overline{B}\right)$ の統計分布の分散の推定値 $\hat{\sigma}^2$ を

$$\hat{\sigma}^2 = \frac{\left(n_A - 1\right)\hat{\sigma}_A^2 + \left(n_B - 1\right)\hat{\sigma}_B^2}{n_A + n_B - 2} \qquad (1.5.2.1)$$

とします。この仮定は，「全体の分散を二つのグループの標本分散の重み付きの平均として計算する」という発想に基づいていると推測されます。

以下の式で $\hat{t}$ を計算します。

$$\hat{t} = \frac{\overline{A} - \overline{B}}{\hat{s}} = \frac{\overline{A} - \overline{B}}{\hat{\sigma}\sqrt{1/n_A + 1/n_B}} \qquad (1.5.2.2)$$

二つの集団の母平均を μ_A, μ_B として，テスト統計量 t を

$$t = \frac{\overline{A} - \overline{B} - \Delta\mu}{\hat{s}} = \frac{\overline{A} - \overline{B} - \Delta\mu}{\hat{\sigma}\sqrt{1/n_A + 1/n_B}} \qquad (1.5.2.3)$$

$$\Delta\mu = \mu_A - \mu_B \tag{1.5.2.4}$$

と表します。

テスト統計量 t の分布が自由度 $\nu = n_A + n_B - 2$ のステューデント t 分布に従うと考えます。そのように考えれば，二標本 t 検定と一標本 t 検定との論理的な構造は似てきます。式(1.5.2.2) で計算される $\hat{t}$ の値が正(せい)のとき，ステューデント t 分布を仮定して「上側(うえがわ) P 値」を求めれば，それが「二つの標本の母平均が $\mu_A < \mu_B$ となる確率」に対応すると考えられているようです。

このときに「両側 P 値」はどのような確率を意味するでしょうか？一標本 t 検定の場合と同じように「$\mu_A - \mu_B < 0$ または $\mu_A - \mu_B > 2\left(\overline{A} - \overline{B}\right)$ となる確率」という意味になるはずです。「２標本 t 検定で両側確率が充分に小さければ，２つの標本の母平均の等しい確率が高くなる」ということ自体は間違っていません。しかし，それは「検定統計量が正の時に上側確率（上側棄却確率）が充分に小さければ，２つの母平均の等しい確率が高くなる」ことと同じです。「確率」という考え方が普通の意味で使えるのは片側検定 (one-sided test）だけで，棄却(ききゃく)できる帰無仮説は標本平均の大小に関することでしかありません。

「コンピュータを使った t テスト」は「コンピュータを使えなかった時代の t 検定」より，作業が楽(らく)というだけでなく，どちらかと言えば極端な間違いを起こしにくく，「第一種過誤（偽陽性）の出現確率」が直接求められるように見える利点があります。

また「コンピュータを使った t テスト」では「帰無仮説／対立仮説」を立てる必要もなく，「有意水準」を決めなくてよく，「両側検定にするか片側（上側・下側）検定にするか」について考える必要もありません。

このことについては，自分で作業をして確かめるのが良いでしょう（補足 1.5.2.A）。

Python 言語で利用できる SciPy ライブラリの `stats` モジュールで提供される `stats.ttest_ind(a,b[,axis,…])` を用いれば，デフォルト（省略時）設定では「母分散（母標準偏差）の等しい二つの独立な (independent) 標本の平均の大小についての t テストを行うことができるとされてます。

ただし，以上の考え方に根本的な誤りが含まれているらしいことは単純な例で証明することができます。（補足 1.5.2.B）

1−5−3　独立な二標本の母分散が等しいとは限らない場合の t テスト（t 検定）

「独立な二つの標本集団 $A = \{A_1, \cdots, A_{n_A}\}$，$B = \{B_1, \cdots, B_{n_B}\}$ の母平均も母分散も異なっているかもしれない場合」に使われる方法で，**ウェルチの t(てぃー) テスト**（t(てぃー) 検定(けんてい)）(Welch's t -test)(ウェルチス ティー テスト) とも呼ばれます。

以下の値 $\hat{t}$ を計算します。

$$\hat{t} = \frac{\overline{A} - \overline{B}}{\hat{s}} \tag{1.5.3.1}$$

とここで尺度パラメータ $\hat{s}$ は,

$$\hat{s} = \sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}} \tag{1.5.3.2}$$

とします。ただし，それぞれの標本からの母分散の推定値（不偏分散）を $\hat{\sigma}_A^2, \hat{\sigma}_B^2$ とします。ウェルチの t 検定では，自由度を

$$\nu = \frac{\left(\dfrac{\hat{\sigma}_A^2}{n_A} + \dfrac{\hat{\sigma}_B^2}{n_B}\right)^2}{\dfrac{\left(\hat{\sigma}_A^2/n_A\right)^2}{n_A - 1} + \dfrac{\left(\hat{\sigma}_B^2/n_B\right)^2}{n_B - 1}} \tag{1.5.3.3}$$

として計算すると言われる場合があります。母分散が等しいことを前提としないウェルチの t 検定には，自由度 ν が整数になるとは限らないという特徴があります。

Python 言語で利用できる SciPy ライブラリの `stats` モジュールで提供される `stats.ttest_ind(a, b [, axis, …])` では，等分散 (equal variance) を仮定するかを指定するオプションのパラメータ `equal_var` に `False`（偽）を指定すれば「母分散（母標準偏差）が等しいとは限らない独立な (independent) 二つの標本の平均の大小についての t テスト（ウェルチの t テスト）を行うことができるとされています（補足 1.5.3.A）。

「コンピュータを使った t テスト」では，自由度 ν が**整数**でも**非整数**でも**計算コスト**が大きく変わるわけではなく，自由度 ν を式 (1.5.3.3) で計算して，P 値を式 (1.3.1.1) または式 (1.3.1.2) で計算するというだけです。この点でも「コンピュータを使った t テスト」は，「t 分布表」を使う「t 検定」と，基本的な考え方自体は本来は同じことなのだとしても，実際の「使い勝手」は，かなり違うことがわかります。

2. カイ自乗検定

カイ自乗検定 (chi-squared test) はカイ二乗検定，χ^2 検定（χ はギリシャ小文字のカイ chi）などとも書かれ，検定統計量が（ステューデント t 分布でなく）カイ自乗分布 (chi-square distribution) に従うと仮定します。

自由度 ν のカイ自乗分布の確率密度関数 $f_{\chi^2}(x; \nu)$ は

$$
f_{\chi^2}(x;\nu) = \begin{cases} \dfrac{1}{2\Gamma(\nu/2)} \left(\dfrac{x}{2}\right)^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{x}{2}} & [x > 0] \\ 0 & [\text{otherwise}] \end{cases} \tag{2.1}
$$

のように表されます。

自由度 ν のカイ自乗分布は，ν 個の独立な標準正規分布 (standard normal distribution) に従う確率変数の自乗の和の分布（畳込）と理解することができます（補足 2.A）。

ピアソンのカイ自乗検定は，複数のカテゴリー・データ (categorical data) の組に対して，観測される違いが偶然生じることが「どの程度ありえそうか」を見積もるために用いられます。

Python 言語で利用できる SciPy ライブラリの `stats` モジュールで提供されるカイ自乗テストのためのメソッド `stats.chisquare(f_obs[, f_exp=None, ddof=0, axis=0])` では，カテゴリー・データ (categorical data) の観測頻度 (observed frequency) `f_obs` が，与えられた予想頻度 (expected frequency) `f_exp` より低いことを帰無仮説とする確率の計算が行われます。予想頻度 (expected frequency) `f_exp` を省略した場合には，予想頻度として一様分布が仮定されます。`stats.chisquare(…)` メソッドは，検定統計量 `statistic` と P 値 `pvalue` を出力します。カイ自乗検定の P 値は上側確率（上側棄却域）に相当する値です。

3．F 検定

F 検定 (F-test) は，検定統計量が F 分布 (F distribution) に従うと仮定する統計的な検定です。F 分布の確率密度函数 (probability density function) $f_{\mathrm{F}}(x;d_1,d_2)$ は，

$$
f_{\mathrm{F}}(x;d_1,d_2) = \frac{1}{\mathrm{B}\left(d_1/2, d_2/2\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2}x\right)^{-\frac{d_1+d_2}{2}} \tag{3.1}
$$

と表されます。ここで $\mathrm{B}(a,b)$ はベータ函数 (beta function) です。式 (3.1) 中の d_1 と d_2 は，二つの自由度に相当するパラメータに対応します。F 分布の累積分布函数 (cumulative distribution function) は，正則不完全ベータ函数 (regularized incomplete beta function) $I(a,b;x)$ を使って，

$$
F_{\mathrm{F}}(x;d_1,d_2) = I\left(\frac{d_1}{2}, \frac{d_2}{2}; \frac{d_1 x}{d_1 x + d_2}\right) \tag{3.2}
$$

と表されます。

F 検定は，二組のデータの分散あるいは標準偏差が等しいことを帰無仮説とする検定にも用いられると言われます。このタイプの統計的な検定のようなことが**分散分析** (analysis of variance; ANOVA) と呼ばれる場合があります。

二組のデータ $\{X_1, \cdots, X_m\}$，$\{Y_1, \cdots, Y_n\}$ がそれぞれ独立同分布 (independent and identically distributed) の標本データであり，母平均と母分散は異なる値をとるかもしれないとします。それぞれの標本平均が

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{m} X_i \tag{3.3}$$

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{3.4}$$

であり，標本分散が

$$S_X^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(X_i - \overline{X}\right)^2 \tag{3.5}$$

$$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 \tag{3.6}$$

であるとします。検定統計量

$$F = \frac{S_X^2}{S_Y^2} \tag{3.7}$$

は「等分散」を仮定できれば自由度 $m-1$ と $n-1$ を持つ F 分布に従うはずであり，そうでなければ，真の分散の比によって尺度化された F 分布になると言われます。検定統計量 F の値が極端に大きい場合と極端に小さい場合には，等分散仮定は棄却されます。t 検定とは異なり，F 検定をする人たちには「$F > 1$ となるように検定統計量をとり，上側 P 値を計算する」慣習があるようです。

F 検定は本来「分散が等しいか等しくないか」を調べるためのものではないことには，注意すべきでしょう。

「t 検定をする前に F 検定をすると良い」と言われる場合があります。コンピュータを使えなかった時代には，たとえば「等分散」を帰無仮説とした F 検定の P 値が，「かなり大きい値」になれば，それを根拠として「等分散を仮定してステューデント t 検定をしてもよい」と思われていたかもしれません。

例えば，F 検定の結果 P 値が 5% を超える値になり「5% の有意水準では等分散性が棄却されなかった」場合に「だから等分散性を仮定して良い」と言い出す人をときどきみかけます。しかし「5%以上の確率で成立すること」は「あるともいえないし，ないともいえない」ようなことです。F 検定は等分散性を棄却するためには使えますが，等分散性を採用するためには使えません。

現代はコンピュータを使えるので「等分散性」を仮定すること自体にあまり意味はありません。

Python 言語で利用できる SciPy ライブラリの `stats` モジュールで提供される `stats.f_oneway()` では，カテゴリー・データ (categorical data) の観測頻度 (observed frequency) `f_obs` が，予想頻度 (expected frequency) `f_exp` と一致することを帰無仮説とする F 検定が行われます。検定統計量 `statistic` と P 値 `pvalue` が出力されます。

複数データの母平均に関する検定が目的であれば，「分散が等しいと仮定できなければ，ウェルチの t テスト（t 検定）をすれば良い」とする場合もあるようですが，全体的に無意味な検定法とも言えます。

4．その他の検定

4-1　シャピロ・ウィルク検定

シャピロ・ウィルク検定 (Shapiro-Wilk test) は，統計分布の正規性を評価する目的で用いられると言われる検定の一つです。シャピロ・ウィルクの検定統計量は

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{4.1}$$

と表されます。ここで $x_{(i)}$ は標本 $\{x_1, x_2, \cdots, x_n\}$ のうち i 番目の順序統計量 (order statistic)，$\bar{x} = (x_1 + \cdots + x_n)/n$ は標本平均とします。係数 a_i は

$$(a_1, \cdots, a_n) = \frac{\mathbf{m}^{\mathrm{T}}\mathbf{V}^{-1}}{\left(\mathbf{m}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{m}\right)^{1/2}}$$

によって与えられ，$\mathbf{m}$ は，

$$\mathbf{m} = \left(m_1, \cdots, m_n\right)^{\mathrm{T}}$$

として，標準正規分布からサンプリングされた独立同分布のランダム変数 (independent and identically distributed random variables) の順序統計量の期待値からなり，$\mathbf{V}$ はこの順序統計量の分散共分散行列 (covariance matrix) です。

SciPy ライブラリの `stats` モジュールで提供される `stats.shapiro()` メソッドは，データが正規分布に従うことを帰無仮説とするシャピロ・ウィルク検定をするもので，検定統計量 `statistic` と P 値 `pvalue` が出力されます。P 値が極端に小さい値であれば「正規性」は棄却されますが，どの程度大きければ正規性を仮定できるかは，必ずしもはっきりとはしません。

統計分布の正規性を評価するための検定法には，シャピロ・ウィルク検定の他に**アンダーソン・ダーリング検定** (Anderson-Darling test），**コルモゴロフ・スミルノフ検定** (Kolmogorov-Smirnov test) があり，SciPy ライブラリの `stats` モジュールでは，それぞれ `stats.anderson()` メソッド，`stats.kstest()` メソッドとして提供されています。

4-2　分散分析

バートレット検定 (Bartlett test) は，分散分析の目的で用いられる統計的な検定の一つで，カイ自乗検定と似た性格を持ちます。要素数 n_i，標本分散 S_i $(i = 1,2,\cdots,k)$ の k 標本があったとき，バートレットの検定統計量は

$$\chi^2 = \frac{(N-k)\ \ln S_p^2 - \sum_{i=1}^{k}(n_i-1)\ln S_i^2}{1+\dfrac{1}{3(k-1)}\left(\displaystyle\sum_{i=1}^{k}\frac{1}{n_i-1}-\frac{1}{N-k}\right)} \tag{4.1}$$

と表されます。ここで $N = \displaystyle\sum_{i=1}^{k} n_i$ として， $S_p^2 = \dfrac{1}{N-k}\displaystyle\sum_{i=1}^{k}(n_i-1)\,S_i^2$ は合併分散推定値 (pooled estimate for the variance) とします。

SciPy ライブラリの `stats` モジュールで提供される `stats.bartlett()` は，複数のデータについてバートレット検定をするもので，検定統計量 `statistic` と P 値 `pvalue` が出力されます。P 値が極端に小さい値であれば「等分散仮定」は棄却されますが，どの程度大きければ等分散を仮定できるかは，必ずしもはっきりとはしません。

ルビーン検定 (Levene's test) も，分散分析の目的で用いられる統計的な検定の一つで，検定統計量は分散分析を目的とした F 検定量に等しく，

$$W = \frac{(N-k)\sum_{i=1}^{k}N_i(Z_{i\cdot}-Z_{\cdot\cdot})^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Z_{ij}-Z_{i\cdot})^2} \tag{4.2}$$

と表されます。ただし， k は比較する群の数，N は全群の総観測数，N_i は第 i 群の観測数，Y_{ij} は第 i 群の j 番目の変数の値で，$\overline{Y}_{i\cdot}$ を第 i 群の平均値として，$Z_{ij} = |Y_{ij} - \overline{Y}_{i\cdot}|$，$Z_{\cdot\cdot} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{N_i} Z_{ij}$ は全ての Z_{ij} の平均値，$Z_{i\cdot} = \displaystyle\sum_{j=1}^{N_i} Z_{ij}$ は第 i 群の Z_{ij} の平均値とします。

SciPy ライブラリの `stats` モジュールで提供される `stats.levene()` は，複数のデータのルビーン検定をするもので，検定統計量 `statistic` と P 値 `pvalue` が出力されます。P 値が極端に小さい値であれば「等分散仮定」は棄却されますが，どの程度大きければ等分散を仮定できるかは，必ずしもはっきりとはしません。

例題

あなたは，ある製造系民間企業に勤務する社員です。新しい商品の紹介・販売促進を目的として，新しい企画も取り入れて工夫を凝らして開催したイベントの参加者アンケートの結果と，前回の同様なイベントの参加者アンケートの結果とを比較して，新しいイベントが効果的であったかを分析することになりました。どのように分析すれば良いか，インターネットで検索したところ，以下のような情報が得られました。

(1) このような場合には「効果がなかった」ことを帰無仮説とする二標本 t 検定を行えば良い。

(2) t 検定のためには，データが正規分布に従う必要がある。そのためシャピロ・ウィルク検定 (Shapiro-Wilk test) かコルモゴロフ・スミルノフ検定 (Kolmogorov-Smirnov test) などの方法により正規性についての検定をすれば良い。

(3) 二つのデータの分散が等しいかを検定するために F 検定を用いる。

(4) 二つのデータの分散が等しいとみなせる場合にはステューデントの t 検定 (Student's t-test) を行う。

(5) 二つのデータの分散が等しいとみなせない場合にはウェルチの t 検定 (Welch's t-test) を行う。

(6) このような場合には両側検定をするべきで，片側検定をするべきではない。

これらの情報の中には，いくつか誤った内容が含まれているようです。どこがどのように誤っているでしょうか？

解答例

本文に示した通り，独立な二標本の平均の大小についての検定を行うためには，「二標本 t 検定」は原則的に使えません。このときには二つの Student t 分布の畳込の計算をしなければいけないはずです。

また，「効果がなかった」ことを帰無仮説とする検定を行うのではなく，「新しいイベントは前のイベントに比べて効果がないか，悪い効果（負の効果）があった」という帰無仮説をたてて「二標本についての検定」を行うべきです。

「t 検定のためにはデータが正規分布に従う必要がある」とは限りません。シャピロ・ウィルク検定をしてもかまいませんが，シャピロ・ウィルク検定は「正規分布に従う」ことを帰無仮説とする検定なので，例えば「有意水準 5% で『正規分布に従う』とする帰無仮説が棄却されなかった」として「だから正規分布に従うとして良い」わけではなく「P 値が 5% 以上になったので，正規分布に従わないとは言えない」という意味にしかなりません。それは実質的には「ほぼ無意味なこと」です。F 検定についても同様で，F 検定は「分散が等しいこと」を帰無仮説とする検定なので，「有意水準 5% で『分散が等しい』とする帰無仮説が棄却されなかった」として，それは「分散が等しくないとは言えない」という程度の意味にしかなりません。

「二つのデータの分散が等しいとみなせる場合にはステューデント t 検定をすれば良い」のは主に「コンピュータを使えなかった時代の人」の発想です。「コンピュータを使える現代の人」は，「二標本データの比較のためにはウェルチの t 検定をすれば良い」とします。ただし，その考え方にも根本的な間違いが含まれます。

「両側検定をするべきである」という人の中には「『違いがない』ことを帰無仮説として『違いがある』ことを対立仮説とするのであれば，片側検定ではなく，両側検定をするべきである」というように，ひょっとすると「もっともらしく聞こえる」かもしれないけれど，よく考えれば「まったく不合理な内容」のことを言う人が，実際に多いのが現実です。検定統計量 t を計算した時点で，それが正（プラス; plus）の値なのか負（マイナス; minus）の値なのかは「見ればわかる」でしょう。検定統計量 t が正の値なら「真値は 0 以下である」という帰無仮説をたて，検定統計量 t が負の値なら「真値は 0 以上である」という帰無仮説をたてて「片側検定」をすれば「まだまし」に聞こえるはずです。

SciPy ライブラリの `stats` モジュールに含まれる一標本 t 検定用の `scipy.stats.ttest_1samp(…)` メソッドと二標本 t 検定用の `scipy.stats.ttest_ind(…)` メソッドが，いずれもデフォルト（省略時設定）では「両側検定 P 値」を出力する仕様となっていることも，混乱を招くかもしれません。でもその本当の意

味を知っているユーザであれば，「その仕様の方が，上側か下側を指定する必要がないので，便利で使いやすい」という面もあるかもしれません。

補足

（補足 1.A）Student の t 分布（↩）

Student は，英国（アイルランド）ギネス醸造所（じょうぞうしょ）の醸造（じょうぞう）技術者であった William Sealy Gosset のペンネーム (Student, 1908) で，この人物は「現代統計学の創始者（そうししゃ）」とも呼ばれます。当時のギネス醸造所（じょうぞうしょ）の経営者は，従業員である技術者に学術論文の投稿を奨励（しょうれい）していたが，本名（ほんみょう）と違うペンネームを使って投稿をさせており，普通には「学生」と言う意味の Student と言う言葉をペンネームとして使ったというところには，英国流のユーモアのようなものが感じられます。「ギネス醸造所（じょうぞうしょ）ではビールの品質を評価するために，統計学を使っている」ということを，ライバル会社に知られたくなかったので「身元が明らかにならないようにペンネームで投稿するように指示をしていた」と言われています。（↩）

（補足 1.B）第一種と第二種の過誤，偽陽性と偽陰性（↩）

第一種の過誤のことが「あわて者（もの）のしくじり」，第二種の過誤のことが「ぼんやり者（もの）のしくじり」のように表現されることがあります。典型的な帰無仮説検定のパターンでは，第一種過誤は「誤っていることを，正しいと思い込んでしまう」ことに対応し，第二種過誤は「正しいことを誤っていると思い込んでしまう」ことに対応するので，「うまい表現」のように思います。

治療のための新薬・感染予防のための新しいワクチンが開発されたときに，安全性と有効性を正しく評価することは重要です。投薬・接種による副作用 (side effects)・副反応 (side reactions) は，ゼロではない確率で現れることであり，危険な副作用・副反応が現れる場合もあります。「コンピュータを使えなかった時代」には医薬品が「有効ではない」とする「帰無仮説」をたてて「有意水準」を設定して t 検定をし，一定の基準で（例えば有意水準 5%で）帰無仮説を棄却できれば「副作用・副反応のリスクを考慮しても，充分に有効と判断できるので，その医薬品・ワクチンの使用を認可する」と言うようにしていたと考えれば良いでしょう。

1980 年代以降コンピュータ利用が普及し，「その医薬品は有効ではない」とする「帰無仮説（きむかせつ）」をたてれば「効果がないのに効果があると誤って判断する第一種過誤」の発生する確率「P（ピー） 値（ち）」(P（ピー）-value)（無効率）が直接求められるようになりました。そして，例えば「P（ピー） 値（ち）」（無効率）が $p = 4.0\,\%$ になったとすれば，有効率 (efficacy)（エフィカスィ） は $1 - p = 96.0\,\%$ ということになります。

ところが，当然のように「その有効率 $1 - p$ の値はどの程度信頼できるの (confident)（コンフィデント） か？」という疑問が生じます。普通の t 検定の手順で「P（ピー） 値（ち）」を求めたとしても，その疑問に応（こた）えるのは実際には困難です。そこで，「正しい有効率 $1 - p$」が，かなり高い確率（普通は 95%とする）で含まれると推定される区間が「信頼区間 (confidence interval)（コンフィデンス インターヴァル）（95% 信頼区間）」として併記されるようになりました (*e.g.*, Flechner and Tseng, 2011)。

「95% 信頼区間 (confidence interval)」とは，「100 回の調査のうち 95 回の調査で真値（しんち） (true value) をカバー (cover) するような区間」と意味付けられるそうです (*e.g.*, du Prel *et al.*, 2009)。100 回の調査のそれぞれに「見

積もられた有効率についての区間」を設定します。100 回の調査のうち 95 回は「設定された有効率の区間」の中に真値が含まれるが，5 回は「設定された有効率の区間」の中に真値が含まれないとします。「『そのような区間』を 95% 信頼区間と呼ぶ」ということです。具体的にどのように区間を設定するかは別として，そのようにすれば「正確な有効率の値が求められているとは言えない場合」には信頼区間が広くなり，「有効率をかなり正確に求められている場合」には信頼区間が狭くなるので，「有効率の評価結果の信頼性」についての情報が得られます。

最近では「信頼区間 (confidence interval) 95%」と基本的には同じことを「信用区間 (credible interval) 95%」と表記することが主流になっています。「信用区間 (credible interval) 95%」とは，ベイズ推定 (Bayesian inference) の考え方に基づいて，「事後確率 (posterior probability) の 95% を含む範囲」と意味付けられます。

例えばファイザー (Pfeizer) 社の開発した対 Covid-19 ウイルス用ワクチン BNT162b2 は「有効率 95%」「95%（95% 信用区間 (credible interval), 90.3–97.6%）」と 2020 年に報告されました (Polack *et al.*, 2020)。これは P 値から見積もられた「有効率」が 95 % であり，『正しい有効率の値が 90.3 – 97.6 % の間にある』と推定される確率は 95 % である」という意味合いのことです。（↩）

（補足 1.1.A）標本平均の統計分布（↩）

標本値（データ）$\{X_1, X_2, \cdots, X_n\}$ の統計分布が母平均 μ，母 標準 偏差 σ の**正規分布** (normal distribution) に従うとします。このとき $(X_1 + X_2 + \cdots + X_n)$ の平均（1 階キュムラント）は $n\mu$，分散（2 階キュムラント）は $n\sigma^2$，3 階以上のキュムラントはすべて 0 になります。**標本平均** (sample mean) $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ の平均は $\dfrac{n\mu}{n} = \mu$ となり，分散は $\dfrac{n\sigma^2}{n^2} = \dfrac{\sigma^2}{n}$，3 階以上のキュムラントはすべて 0 になります。標本平均 $\overline{X}$ は，標準 偏差 $\dfrac{\sigma}{\sqrt{n}}$ の**正規分布**に従います。（↩）

（補足 1.1.B）自由度 1 のカイ自乗分布（↩）

確率変数 x の統計分布が平均 0，標準 偏差 1 の正規分布，**標準 正規分布** (standard normal distribution) に従うとします。**標準 正規分布**の確率密度函数 (probability density function) $f_{\rm SN}(x)$ は

$$f_{\rm SN}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{1.1.B.1}$$

と表されます。

かりに確率変数 X の統計分布が「平均 μ（μ はギリシャ小文字のミュー mu），標準 偏差 σ（σ はギリシャ小文字のシグマ sigma）の**正規分布** (standard normal distribution)」に従うとすれば，確率変数 $X - \mu$ の統計分布は「平均 0，標準 偏差 σ の**正規分布**」に従い，確率変数 $\dfrac{X-\mu}{\sigma}$ の統計分布は「平均 0，標準 偏差 1 の**正規分布**」（**標準正規分布**）に従います。

このことは「確率変数 x に定数 a を加えた確率変数 $x + a$ の統計分布は，一階キュミュラント（平均）が a 増えるだけで，他の階数のキュミュラントは変化しない」こと，「確率変数 x に定数 b を掛けた確率変数 bx の統計分布では，k 階キュミュラントの値が b^k 倍になる」という「一般的に成立する数学的な関係」から導けます。このことを「キュミュラントの尺度化法則 (scaling rules for cumulants)」として知っておくと良いでしょう。

確率変数 x が標準正規分布に従うとしたときに，$x^2 = y$ はどのような統計分布に従うでしょうか？この問題を解く場合には，ディラックのデルタ函数 (Dirac delta function) $\delta(x)$を使って，確率変数 y の統計分布の確率密度函数を $f_{\chi^2}(y;1)$ と表すことにして，

$$f_{\chi^2}(y;1) = \int_{-\infty}^{\infty} \delta\left(y - x^2\right) f_{\mathrm{SN}}(x)\ \mathrm{d}x \tag{1.1.B.2}$$

と表されるということから出発すると，間違えにくいでしょう。

ディラックのデルタ函数 $\delta(x)$ が任意の函数 $f(x)$ との積の積分について

$$\int_a^b \delta(x) f(x)\,\mathrm{d}x = \begin{cases} f(0) & [a < 0 < b] \\ -f(0) & [b < 0 < a] \\ 0 & [\text{otherwise}] \end{cases} \tag{1.1.B.3}$$

という関係を持つことを使います。

はじめに，式 (1.1.B.2) の右辺の被積分函数が，x については偶函数 (even function) であることから，

$$f_{\chi^2}(y;1) = 2\int_0^{\infty} \delta\left(y - x^2\right) f_{\mathrm{SN}}(x)\ \mathrm{d}x \tag{1.1.B.4}$$

と書き直します。つぎに，デルタ函数 $\delta(\cdots)$ の変数の部分 $y - x^2$ をまとめて他の変数におきかえます。ここでは

$$y - x^2 \equiv -u \quad \Rightarrow \quad x^2 = y + u \quad \Rightarrow \quad x = \sqrt{u + y} \tag{1.1.B.5}$$

として，

$$\mathrm{d}x = \frac{\mathrm{d}u}{2\sqrt{u+y}} \tag{1.1.B.6}$$

$$\begin{array}{cccccc} x & : & 0 & \rightarrow & \infty \\ u & : & -y & \rightarrow & \infty \end{array} \tag{1.1.B.7}$$

の関係から，式 (1.1.B.4) を以下のように変形します。

$$\begin{aligned} f_{\chi^2}(y;1) &= 2\int_{-y}^{\infty} \delta(-u) f_{\mathrm{SN}}\left(\sqrt{u+y}\right) \frac{\mathrm{d}u}{2\sqrt{u+y}} = 2\int_{-y}^{\infty} \delta(u) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u+y}{2}\right) \frac{\mathrm{d}u}{2\sqrt{u+y}} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-y}^{\infty} \delta(u) \frac{1}{\sqrt{u+y}} \exp\left(-\frac{u+y}{2}\right) \mathrm{d}u \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \begin{array}{ll} \dfrac{1}{\sqrt{y}} \exp\left(-\dfrac{y}{2}\right) & [y > 0] \\ 0 & [y \le 0] \end{array} \right\} = \begin{cases} \dfrac{1}{2\sqrt{\pi}} \left(\dfrac{y}{2}\right)^{-\frac{1}{2}} \exp\left(-\dfrac{y}{2}\right) & [y > 0] \\ 0 & [y \le 0] \end{cases} \end{aligned} \tag{1.1.B.8}$$

式 (1.1.B.8) の最後の式変形は，自由度 (degree of freedom) k のカイ自乗分布 (chi-squared distribution) の確率密度函数 $f_{\chi^2}(x;k)$ が一般的に

$$f_{\chi^2}(x;k) = \begin{cases} \dfrac{1}{2\,\Gamma\left(\dfrac{k}{2}\right)} \left(\dfrac{x}{2}\right)^{\frac{k}{2}-1} \exp\left(-\dfrac{x}{2}\right) & [x > 0] \\ 0 & [x \le 0] \end{cases} \tag{1.1.B.9}$$

と表現されることに形式を合わせるためのものです。式 (1.1.B.9) に現れる函数 $\Gamma(x)$ は，ガンマ函数 (Gamma function) あるいは完全ガンマ函数 (complete Gamma function) と呼ばれる特殊函数 (special function) で，

$$\Gamma(x) \equiv \int_0^{\infty} t^{x-1}\,\mathrm{e}^{-t}\,\mathrm{d}t \tag{1.1.B.10}$$

と定義されます。

デルタ函数を使えば，式 (1.1.B.8) のように「$f_{\chi^2}\left(y;1\right)$ の値は $y \leq 0$ の場合には 0 になる」というような「場合わけ」まで含めて，あまり何も考えずに機械的に正しい結果を導けます。印刷物でも電子書類でも，過去の「教科書的なもの」では「$f_{\chi^2}\left(y;1\right)$ の値は $y \leq 0$ の場合には 0 とする」ことが「省略されている」場合が多いので，自力でデータ解析のためのコーディングをするときには注意した方が良いかもしれません。（↩）

（補足 1.1.C）自由度 2 のカイ自乗分布（↩）

標準正規分布に従う二つの**標本値** $\{x_1, x_2\}$ の自乗 $\{x_1^2, x_2^2\}$ の統計分布は，いずれも自由度 1 のカイ自乗分布に従います。さらに，二つの標本値の自乗の和

$$x \equiv x_1^2 + x_2^2 \tag{1.1.C.1}$$

として定義される確率変数 x の統計分布は，「標本値の自乗 x_1^2 の統計分布」と「標本値の自乗 x_2^2 の統計分布」の**畳込** (convolution) として理解されます。「**自由度 2 のカイ自乗分布**」(chi squared distribution for the degree of freedom of two) は，２つの標準正規分布に従う確率変数の自乗の和として表される確率変数の確率分布を意味するとも言えますし，「**自由度 1 のカイ自乗分布**」の**畳込** (convolution)として理解することもできます。標準正規分布の**確率密度函数を** $f_{\mathrm{SN}}(x)$ **とし，自由度 2 のカイ自乗分布**」分布の確率密度函数を $f_{\chi^2}(x;2)$ と表すことにすれば

$$\begin{aligned}
f_{\chi^2}(x;2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta\left(x - x_1^2 - x_2^2\right) f_{\mathrm{SN}}\left(x_1\right) f_{\mathrm{SN}}\left(x_2\right)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta\left(x - x_1^2 - x_2^2\right) \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{0}^{\infty} \delta\left(x - x_1^2 - x_2^2\right) \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \mathrm{d}x_1\,\mathrm{d}x_2 \\
&\underset{\uparrow}{=} \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{x_2^2-x}^{\infty} \delta(u) \exp\left(-\frac{u + x - x_2^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \frac{\mathrm{d}u}{2\sqrt{u + x - x_2^2}}\,\mathrm{d}x_2
\end{aligned} \tag{1.1.C.2}$$

$$x_1^2 + x_2^2 - x \equiv u$$

$$\Rightarrow\ x_1 = \sqrt{u + x - x_2^2}$$

$$\mathrm{d}x_1 = \frac{\mathrm{d}u}{2\sqrt{u + x - x_2^2}}$$

$$\begin{array}{ccccc}
x_1 & : & 0 & \to & \infty \\
u & : & x_2^2 - x & \to & \infty
\end{array}$$

$$
= \frac{1}{\pi} \int_{-\infty}^{\infty} \begin{Bmatrix} \dfrac{1}{2\sqrt{x - x_2^2}} \exp\left(-\dfrac{x}{2}\right) & \left[x_2^2 - x < 0\right] \\ 0 & \left[x_2^2 - x \geq 0\right] \end{Bmatrix} \mathrm{d}x_2
$$

$$
= \frac{1}{\pi} \exp\left(-\frac{x}{2}\right) \begin{Bmatrix} \displaystyle\int_0^{\sqrt{x}} \frac{\mathrm{d}x_2}{\sqrt{x - x_2^2}} & [x > 0] \\ 0 & [x \leq 0] \end{Bmatrix}
$$

$$
\underset{\substack{x_2 \equiv \sqrt{x}\sin\theta \\ \mathrm{d}x_2 = \sqrt{x}\cos\theta\,\mathrm{d}\theta \\ x_2 \;:\; 0 \;\to\; \sqrt{x} \\ \theta \;:\; 0 \;\to\; \pi/2}}{\underset{\uparrow}{=}} \frac{1}{\pi} \exp\left(-\frac{y}{2}\right) \begin{Bmatrix} \displaystyle\int_0^{\pi/2} \mathrm{d}\theta & [y > 0] \\ 0 & [y \leq 0] \end{Bmatrix}
$$

$$
= \begin{cases} \dfrac{1}{2} \exp\left(-\dfrac{x}{2}\right) & [x > 0] \\ 0 & [x \leq 0] \end{cases} \tag{1.1.C.3}
$$

となります。この函数 $f_{\chi^2}(x; 2)$ の形状は**裁断指数函数** (truncated exponential function) と呼ばれることがあります。（↩）

（補足 1.1.D）自由度 ν のカイ自乗分布（↩）

標準正規分布に従う ν 個（ν はギリシャ小文字のニュー，nu）の**標本値** $\{x_1, x_2, \cdots, x_\nu\}$ の自乗 $\{x_1^2, x_2^2, \cdots, x_\nu^2\}$ の統計分布のそれぞれが，いずれも自由度 1 のカイ自乗分布に従うとします。確率変数 x が

$$
x \equiv x_1^2 + x_2^2 + \cdots + x_\nu^2 \tag{1.1.D.1}
$$

と定義されるとき，その統計分布は「標本値の自乗 x_1^2 の統計分布」の ν 重の**畳込** (convolution) として理解されます。「**自由度 ν のカイ自乗分布**」(chi squared distribution for the degree of freedom of nu) は，ν 個の標準正規分布に従う確率変数の自乗の和として表される確率変数の確率分布を意味するとも言えますし，「**自由度 1 のカイ自乗分布**」の ν 重の畳込 (convolution) として理解することもできます。標準正規分布の**確率密度函数**を $f_{\mathrm{SN}}(x)$ とし，「**自由度 ν のカイ自乗分布**」(chi squared distribution for the degree of freedom of nu) が，$f_{\chi^2}(x; \nu)$ と表されるとすると，

$$
f_{\chi^2}(x; \nu) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta\left(x - \sum_{j=1}^{\nu} x_j^2\right) f_{\mathrm{SN}}(x_1) \cdots f_{\mathrm{SN}}(x_\nu)\, \mathrm{d}x_1 \cdots \mathrm{d}x_\nu
$$

$$
= \cdots
$$

$$
= \begin{cases} \dfrac{1}{2\,\Gamma(\nu/2)} \left(\dfrac{x}{2}\right)^{\frac{\nu}{2}-1} \exp\left(-\dfrac{x}{2}\right) & [x > 0] \\ 0 & [x \leq 0] \end{cases} \tag{1.1.D.2}
$$

となります。ここで $\Gamma(\alpha)$ はガンマ函数で

$$\Gamma(\alpha) \equiv \int_0^\infty t^{\alpha-1}\,\mathrm{e}^{-t}\,\mathrm{d}t \tag{1.1.D.3}$$

として定義されますが，変数の値が 1/2 以上の整数または半整数の場合には，

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},\ \Gamma(1) = 1,\ \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2},\ \Gamma(2) = 1,\ \Gamma\left(\frac{5}{2}\right) = \frac{3\sqrt{\pi}}{4},\ \Gamma(3) = 2\cdots$$

などの値をとり，$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ と $\Gamma(1) = 1$，$\Gamma(z) = (z-1)\,\Gamma(z-1)$ の漸化式で値が決まると考えることもできます。

自由度 ν のカイ自乗分布の確率密度函数は式 (1.1.D.2) に示すように

$$f_{\chi^2}(x;\nu) = \begin{cases} \dfrac{1}{2\,\Gamma(\nu/2)}\left(\dfrac{x}{2}\right)^{\frac{\nu}{2}-1}\exp\left(-\dfrac{x}{2}\right) & [x > 0] \\ 0 & [x \le 0] \end{cases} \tag{1.1.D.2}$$

と表されます。この函数のフーリエ変換 (Fourier transform) は，

$$\mathfrak{F}_{\chi^2}\left(k;\nu\right) = \int_{-\infty}^\infty f_{\chi^2}(x;\nu)\,\mathrm{e}^{2\pi\mathrm{i}kx}\,\mathrm{d}x = \frac{1}{(1-4\pi\,\mathrm{i}k\nu)^{\nu/2}}$$

と書けます。

また自由度 ν のカイ自乗分布の確率密度函数 $f_{\chi^2}(x;\nu)$ をガンマ分布の確率密度函数の一般形：

$$f_\Gamma\left(x;\gamma,\alpha\right) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\,\gamma}\left(\dfrac{x}{\gamma}\right)^{\alpha-1}\exp\left(-\dfrac{x}{\gamma}\right) & [x > 0] \\ 0 & [x \le 0] \end{cases} \tag{1.1.D.4}$$

と比較すると $f_{\chi^2}(x;\nu) = f_\Gamma\left(x;2,\dfrac{\nu}{2}\right)$ に相当します。

ガンマ分布の確率密度函数のフーリエ変換は

$$\begin{aligned}
\mathfrak{F}_\Gamma\left(k;\gamma,\alpha\right) &= \int_{-\infty}^\infty f_\Gamma(x;\gamma,\alpha)\,\mathrm{e}^{2\pi\mathrm{i}kx}\,\mathrm{d}x = \frac{1}{\Gamma(\alpha)\,\gamma}\int_0^\infty \left(\frac{x}{\gamma}\right)^{\alpha-1}\exp\left(-\frac{x}{\gamma}\right)\,\mathrm{e}^{2\pi\mathrm{i}kx}\,\mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha)\,\gamma}\int_0^\infty \left(\frac{x}{\gamma}\right)^{\alpha-1}\exp\left(2\pi\,\mathrm{i}k\,x - \frac{x}{\gamma}\right)\,\mathrm{d}x = \frac{1}{\Gamma(\alpha)\,\gamma}\int_0^\infty \left(\frac{x}{\gamma}\right)^{\alpha-1}\exp\left[\left(2\pi\,\mathrm{i}k - 1/\gamma\right)x\right]\,\mathrm{d}x \\
&= \left[\frac{1}{\Gamma(\alpha)\,\alpha}\left(\frac{x}{\gamma}\right)^{\alpha}\exp\left[\left(2\pi\,\mathrm{i}k - \frac{1}{\gamma}\right)x\right]\right]_0^\infty \\
&\qquad - \frac{1}{\Gamma(\alpha)\,\alpha}\left(2\pi\,\mathrm{i}k - \frac{1}{\gamma}\right)\int_0^\infty \left(\frac{x}{\gamma}\right)^{\alpha}\exp\left[\left(2\pi\,\mathrm{i}k - \frac{1}{\gamma}\right)x\right]\,\mathrm{d}x \\
&= -\frac{\Gamma(\alpha+1)\,\gamma}{\Gamma(\alpha)\,\alpha}\left(2\pi\,\mathrm{i}k - \frac{1}{\gamma}\right)\frac{1}{\Gamma(\alpha+1)\,\gamma}\int_0^\infty \left(\frac{x}{\gamma}\right)^{\alpha}\exp\left[\left(2\pi\,\mathrm{i}k - \frac{1}{\gamma}\right)x\right]\,\mathrm{d}x \\
&= -\gamma\left(2\pi\,\mathrm{i}k - \frac{1}{\gamma}\right)\mathfrak{F}(k;\gamma,\alpha+1) = \left(1 - 2\pi\,\mathrm{i}k\gamma\right)\mathfrak{F}(k;\gamma,\alpha+1)
\end{aligned}$$

の関係から

$$\mathfrak{F}_\Gamma\left(k;\gamma,\alpha\right) = \left(1 - 2\pi\,\mathrm{i}k\gamma\right)\mathfrak{F}(k;\gamma,\alpha+1)$$

$$\Rightarrow \mathfrak{F}(k;\gamma,\alpha+1)=\frac{\mathfrak{F}(k;\gamma,\alpha)}{1-2\pi\,\mathrm{i}k\gamma}$$

として漸化式の成立することが確認できます。

(↩)

（補足 1.1.E）不偏標本分散の自由度（↩）

標本値（データ） $\{X_1, X_2, \cdots, X_n\}$ がどのような統計分布に従うかとは関係なく，その**不偏標本分散** $\hat{\sigma}^2$ は

$$\hat{\sigma}^2=\frac{\left(X_1-\overline{X}\right)^2+\cdots+\left(X_n-\overline{X}\right)^2}{n-1} \tag{1.1.E.1}$$

として計算されます。

標本値 $\{X_1, X_2, \cdots, X_n\}$ の**独立性** (independence) を仮定できるとき，母平均が μ，母標準偏差が σ なら，

$$\left\langle \left(X_i-\mu\right)\left(X_j-\mu\right)\right\rangle=\begin{cases}\sigma^2 & [i=j]\\ 0 & [i\neq j]\end{cases} \tag{1.1.E.2}$$

の関係が成立します。ここで $\langle x\rangle$ は x の期待値 (expectation value) を表します。

式 (1.1.E.2) の母平均 μ を標本平均 $\overline{X}$ に置き換えると

$$\begin{aligned}
\left\langle \left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)\right\rangle &= \left\langle \left(X_i-\frac{1}{n}\sum_{k=1}^{n}X_k\right)\left(X_j-\frac{1}{n}\sum_{l=1}^{n}X_l\right)\right\rangle\\
&= \left\langle \left[\left(X_i-\mu\right)-\frac{1}{n}\sum_{k=1}^{n}\left(X_k-\mu\right)\right]\left[\left(X_j-\mu\right)-\frac{1}{n}\sum_{l=1}^{n}\left(X_l-\mu\right)\right]\right\rangle\\
&= \left\langle \left(X_i-\mu\right)\left(X_j-\mu\right)\right\rangle-\frac{1}{n}\sum_{l=1}^{n}\left\langle \left(X_i-\mu\right)\left(X_l-\mu\right)\right\rangle\\
&\qquad -\frac{1}{n}\sum_{k=1}^{n}\left\langle \left(X_k-\mu\right)\left(X_j-\mu\right)\right\rangle+\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\left\langle \left(X_k-\mu\right)\left(X_l-\mu\right)\right\rangle\\
&= \begin{Bmatrix}\sigma^2 & [i=j]\\ 0 & [i\neq j]\end{Bmatrix}-\frac{1}{n}\sum_{l=1}^{n}\begin{Bmatrix}\sigma^2 & [i=l]\\ 0 & [i\neq l]\end{Bmatrix}-\frac{1}{n}\sum_{k=1}^{n}\begin{Bmatrix}\sigma^2 & [k=j]\\ 0 & [k\neq j]\end{Bmatrix}+\frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\begin{Bmatrix}\sigma^2 & [k=l]\\ 0 & [k\neq l]\end{Bmatrix}\\
&= \begin{Bmatrix}\sigma^2 & [i=j]\\ 0 & [i\neq j]\end{Bmatrix}-\frac{\sigma^2}{n}-\frac{\sigma^2}{n}+\frac{1}{n^2}\sum_{k=1}^{n}\sigma^2\\
&= \begin{Bmatrix}\sigma^2 & [i=j]\\ 0 & [i\neq j]\end{Bmatrix}-\frac{\sigma^2}{n}=\begin{cases}\dfrac{(n-1)\sigma^2}{n} & [i=j]\\ -\dfrac{\sigma^2}{n} & [i\neq j]\end{cases}
\end{aligned} \tag{1.1.E.3}$$

となります。式 (1.1.E.3) から，特に $i=j$ のとき

$$\left\langle \left(X_i-\overline{X}\right)^2\right\rangle=\frac{(n-1)\,\sigma^2}{n} \tag{1.1.E.4}$$

となることに注意してください。

線形代数 (linear algebra) の文脈では $\left\langle \left(X_i-\mu\right)\left(X_j-\mu\right)\right\rangle$ を i 行 j 列の行列要素とする $n\times n$ 行列のランクは n ですが，$\left\langle \left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)\right\rangle$ を i 行 j 列の行列要素とする $n\times n$ 行列のランクは $n-1$ になります。行あるいは列が互いにすべて独立ではなく，**線形 従属** (linear dependence) の関係が一組含まれるからです。このようなことは**階数落ち（ランク落ち）** (rank deficiency) とも呼ばれます。

他の説明のしかたもありますが，いずれにしても $\left\{X_1-\overline{X}, X_2-\overline{X}, \cdots, X_n-\overline{X}\right\}$ は自由度を一つ失って，$n-1$ の自由度を持つデータとなります。

例えば $n=2$ のとき**標本値（データ）**は $\{X_1, X_2\}$ とあらわされ，標本平均は $\overline{X}=\dfrac{X_1+X_2}{2}$ です。**不偏標本分散** $\hat{\sigma}^2$ は

$$\hat{\sigma}^2=\left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2=\frac{\left(X_1-X_2\right)^2}{4}+\frac{\left(X_2-X_1\right)^2}{4}=\frac{\left(X_2-X_1\right)^2}{2} \qquad (1.1.E.4)$$

として計算されます。

$$\left\langle \left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)\right\rangle=\begin{cases}\dfrac{(n-1)\,\sigma^2}{n} & [i=j] \\ -\dfrac{\sigma^2}{n} & [i\neq j]\end{cases}$$

の関係から，$\left\langle \left(X_1-\overline{X}\right)\left(X_1-\overline{X}\right)\right\rangle=\left\langle \left(X_2-\overline{X}\right)\left(X_2-\overline{X}\right)\right\rangle=\dfrac{\sigma^2}{2}$，$\left\langle \left(X_1-\overline{X}\right)\left(X_2-\overline{X}\right)\right\rangle$
$=\left\langle \left(X_2-\overline{X}\right)\left(X_1-\overline{X}\right)\right\rangle=-\dfrac{\sigma^2}{2}$

$\left\langle \left(X_i-\overline{X}\right)\left(X_j-\overline{X}\right)\right\rangle$ を i 行 j 列の行列要素とする 2×2 行列は

$$\begin{pmatrix}\sigma^2/2 & -\sigma^2/2 \\ -\sigma^2/2 & \sigma^2/2\end{pmatrix}=\sigma^2\begin{pmatrix}1/2 & -1/2 \\ -1/2 & 1/2\end{pmatrix}=\sigma^2\begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & 0\end{pmatrix}\begin{pmatrix}1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}$$

のように，本質的には 1×1 行列の意味しかない $\begin{pmatrix}1 & 0 \\ 0 & 0\end{pmatrix}$ の線形変換として表されます。幾何学図形に当て嵌めれば，二次元の図形を別の二次元の図形に変換する行列ではなく，二次元の図形を「ある直線の上」に投影 (project) するような変換です。 (↩)

（補足 1.1.F）自由度 1 の Student t 分布の確率密度函数（↩）

ここでは標本数 2（自由度 1）の場合に検定統計量 t の分布が Cauchy 分布（自由度 1 の Student t 分布）に従うことを導きます。

二つの標本値 $\{X_1, X_2\}$ が独立で母平均 μ，母標準偏差 σ の正規分布に従うとすれば，標本平均 $\overline{X}=\dfrac{X_1+X_2}{2}$ は，平均 μ，標準誤差 $\dfrac{\sigma}{\sqrt{2}}$ の正規分布に従うと考えられ，確率変数 $\bar{x}=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{2}}$ の分布は，平均 0，標準偏差 1 の正規分布，標準正規分布に従い，確率密度函数は

$$f_{\mathrm{SN}}(\overline{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\overline{x}^2}{2}\right) \tag{1.1.F.1}$$

と表（あらわ）されるはずです。

母平均 μ と母標準偏差 σ は未知なのですが，$x_1 = \dfrac{X_1 - \mu}{\sigma/\sqrt{2}}$，$x_2 = \dfrac{X_2 - \mu}{\sigma/\sqrt{2}}$ として仮想的に標準化された標本値 $\{x_1, x_2\}$ の標本分散に相当する値 $\hat{s}^2 = \dfrac{\hat{\sigma}^2}{\sigma^2/2}$ は，

$$\hat{s}^2 = 2\left[\frac{x_1^2 + x_2^2}{2} - \left(\frac{x_1 + x_2}{2}\right)^2\right] = \frac{2x_1^2 + 2x_2^2 - x_1^2 - 2x_1x_2 - x_2^2}{2} = \frac{x_1^2 - 2x_1x_2 + x_2^2}{2}$$

$$= \frac{(x_1 - x_2)^2}{2} \tag{1.1.F.2}$$

と書けます。x_1, x_2 のいずれも密度函数が $f_{\mathrm{SN}}(x)$ と表される標準正規分布に従うと考えられることから，$\hat{s}^2$ の統計分布の確率密度函数を $f_{\chi^2}(\hat{s}^2; 1)$ と表すことにすれば，ディラックのデルタ函数 $\delta(x)$ を使って，

$$f_{\chi^2}(\hat{s}^2; 1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta\left(\hat{s}^2 - \frac{(x_1 - x_2)^2}{2}\right) f_{\mathrm{SN}}(x_1)\, f_{\mathrm{SN}}(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta\left(\hat{s}^2 - \frac{(x_1 - x_2)^2}{2}\right) \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \mathrm{d}x_1\, \mathrm{d}x_2 \tag{1.1.F.3}$$

のように表されます。以下の変数変換：

$$\left\{\begin{aligned} \frac{x_1 - x_2}{\sqrt{2}} &\equiv y_1 \\ \frac{x_1 + x_2}{\sqrt{2}} &\equiv y_2 \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} x_1 &= \frac{y_1 + y_2}{\sqrt{2}} \\ x_2 &= \frac{-y_1 + y_2}{\sqrt{2}} \end{aligned}\right\} \tag{1.1.F.4}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(\frac{y_1 + y_2}{\sqrt{2}}, \frac{-y_1 + y_2}{\sqrt{2}}\right) \mathrm{d}y_1\, \mathrm{d}y_2 \tag{1.1.F.5}$$

を施（ほどこ）せば，

$$f_{\chi^2}(\hat{s}^2; 1) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta\left(\hat{s}^2 - y_1^2\right) \exp\left(-\frac{y_1^2 + y_2^2}{2}\right) \mathrm{d}y_1\, \mathrm{d}y_2$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \delta\left(\hat{s}^2 - y_1^2\right) \exp\left(-\frac{y_1^2}{2}\right) \mathrm{d}y_1 \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{y_2^2}{2}\right) \mathrm{d}y_2$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \delta\left(\hat{s}^2 - y_1^2\right) \exp\left(-\frac{y_1^2}{2}\right) \mathrm{d}y_1$$

$$= \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} \delta\left(y_1^2 - \hat{s}^2\right) \exp\left(-\frac{y_1^2}{2}\right) \mathrm{d}y_1 \tag{1.1.F.6}$$

となります。さらに

$$y_1^2 - \hat{s}^2 \equiv u \quad \Rightarrow \quad y_1 = \sqrt{u + \hat{s}^2} \tag{1.1.F.7}$$

$$\mathrm{d}y_1 = \frac{\mathrm{d}u}{2\sqrt{u + \hat{s}^2}} \tag{1.1.F.8}$$

$$
\begin{aligned}
y_1 &: \quad 0 \quad \rightarrow \quad \infty \\
u &: \; -\hat{s}^2 \; \rightarrow \quad \infty
\end{aligned}
\tag{1.1.F.9}
$$

から，

$$
f_{\chi^2}\left(\hat{s}^2;1\right) = \frac{2}{\sqrt{2\pi}} \int_{-\hat{s}^2}^{\infty} \delta\left(u\right) \exp\left(-\frac{u+\hat{s}^2}{2}\right) \frac{\mathrm{d}u}{2\sqrt{u+\hat{s}^2}} \tag{1.1.F.10}
$$

$$
= \left\{ \begin{matrix} \frac{1}{\sqrt{2\pi}} \left(\hat{s}^2\right)^{-\frac{1}{2}} \exp\left(-\frac{\hat{s}^2}{2}\right) & \left[\hat{s}^2 \geq 0\right] \\ 0 & \left[\hat{s}^2 < 0\right] \end{matrix} \right\}
$$

$$
= \left\{ \begin{matrix} \frac{1}{2\sqrt{\pi}} \left(\frac{\hat{s}^2}{2}\right)^{-\frac{1}{2}} \exp\left(-\frac{\hat{s}^2}{2}\right) & \left[\hat{s}^2 \geq 0\right] \\ 0 & \left[\hat{s}^2 < 0\right] \end{matrix} \right. \tag{1.1.F.11}
$$

のように表(あらわ)されます。式 (1.1.F.11) で表される函数 $f_{\chi^2}\left(\hat{s}^2;1\right)$ は「自由度 1 のカイ自乗分布」と呼ばれる統計分布の確率密度函数です。

式 (1.1.F.11) を導出(どうしゅつ)するために「母平均 μ と母標準偏差 σ が既知である」ことも「標本数が 2 の場合に自由度が 1 になる」ことも前提としておらず，「結果的にそうなった」ことに注意してください。「標本数が 2 の場合に自由度が 1 になる」ことには何かの意味があり，そのことには色々な説明のしかたはあるでしょうけれど，その意味はわからなくても正しい答えは導けます。

標準化された標本値 $\{x_1, x_2\}$ の標本分散に相当する値 $\hat{s}^2$ の統計分布が式 (1.1.F.11) で表されるということなので，標準化された標本平均の不偏標準偏差に相当(そうとう)する $\hat{s}$ の分布は，$\hat{s} \geq 0$ で非ゼロ値をとるとして，

$$
f_1\left(\hat{s}\right) = \frac{\mathrm{d}\hat{s}^2}{\mathrm{d}\hat{s}} f_{\chi^2}\left(\hat{s}^2;1\right) = \left\{ \begin{matrix} \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{\hat{s}^2}{2}\right) & \left[\hat{s} \geq 0\right] \\ 0 & \left[\hat{s} < 0\right] \end{matrix} \right. \tag{1.1.F.12}
$$

と表されます。「分散 $\hat{s}^2$」が「自由度 1 のカイ自乗分布」に従うのなら，「標準偏差 $\hat{s}$」の確率密度函数は「『標準正規分布の確率密度函数の変数が正の部分』の 2 倍」になることは，見落としやすいかもしれません。このことに気づけば「自由度 1 のステューデント t 分布」の確率密度函数(かんすう)を導(みちび)くことは難しくありません。

このとき検定統計量 $t \equiv x/\hat{s}$ の分布は，ディラックのデルタ函数 $\delta(x)$ を使って，

$$
\begin{aligned}
f_{\mathrm{Student}}\left(t;1\right) &= \int_0^{\infty} \int_{-\infty}^{\infty} \delta\left(t - \frac{x}{\hat{s}}\right) f_{\mathrm{SN}}\left(x\right) f_1\left(\hat{s}\right) \,\mathrm{d}x\,\mathrm{d}\hat{s} \\
&= \frac{1}{\pi} \int_0^{\infty} \int_{-\infty}^{\infty} \delta\left(t - \frac{x}{\hat{s}}\right) \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{\hat{s}^2}{2}\right) \mathrm{d}x\,\mathrm{d}\hat{s}
\end{aligned}
\tag{1.1.F.13}
$$

と表されます。積分変数 x を

$$
\frac{x}{\hat{s}} - t \equiv u \quad \Rightarrow \quad x = \hat{s}\left(u+t\right) \tag{1.1.F.14}
$$

$$
\mathrm{d}x = \hat{s}\,\mathrm{d}u \tag{1.1.F.15}
$$

として変数 u に変換すれば，

$$
\begin{aligned}
f_{\text{Student}}(t;1) &= \frac{1}{\pi}\int_0^\infty \int_{-\infty}^\infty \delta(u)\exp\left[-\frac{\hat{s}^2(u+t)^2}{2}\right]\exp\left(-\frac{\hat{s}^2}{2}\right)\left(\hat{s}\,\mathrm{d}u\right)\,\mathrm{d}\hat{s} \\
&= \frac{1}{\pi}\int_0^\infty \hat{s}\exp\left(-\frac{\hat{s}^2 t^2}{2}\right)\exp\left(-\frac{\hat{s}^2}{2}\right)\,\mathrm{d}z \\
&= \frac{1}{\pi}\int_0^\infty \hat{s}\exp\left[-\frac{(1+t^2)\,\hat{s}^2}{2}\right]\,\mathrm{d}\hat{s} \\
&= \frac{1}{\pi}\left[-\frac{1}{1+t^2}\exp\left[-\frac{(1+t^2)\,\hat{s}^2}{2}\right]\right]_{\hat{s}=0}^{\infty} \\
&= \frac{1}{\pi\left(1+t^2\right)}
\end{aligned}
\tag{1.1.F.16}
$$

となり，「自由度 1 の Student t 分布」は，確率密度函数が Lorentz（ローレンツ）型函数で表される Cauchy（コーシー）分布に従うことがわかります。 (↩)

（補足 1.1.G）自由度 2 の Student t 分布の確率密度函数（↩）

自由度 2 の場合には，三つの標本値 $\{X_1, X_2, X_3\}$ が独立で母平均 μ，母標準偏差 σ の正規分布に従うとすれば，標本平均 $\overline{X} = \dfrac{X_1 + X_2 + X_3}{2}$ は，平均 μ，標準誤差 $\dfrac{\sigma}{\sqrt{3}}$ の正規分布に従い，確率変数 $x = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{3}}$ の分布の確率密度函数が

$$
f_{\text{SN}}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)
\tag{1.1.G.1}
$$

と表されます。

一方で，$x_j = \dfrac{X_j - \mu}{\sigma/\sqrt{3}}$ として標準化された標本値 $\{x_1, x_2, x_3\}$ の標本分散に相当する値 $\hat{s}^2 = \dfrac{\hat{\sigma}^2}{\sigma^2/3}$ は，

$$
\hat{s}^2 = \frac{\hat{\sigma}^2}{\sigma^2/3} = \frac{(X_1-\overline{X})^2 + (X_2-\overline{X})^2 + (X_3-\overline{X})^2}{2(\sigma^2/3)} = \frac{x_1^2 + x_2^2 + x_3^2}{2}
$$

と表されることから，$2\hat{s}^2$ の統計分布は自由度 2 のカイ自乗分布に従い，確率密度函数は

$$
f_{\chi^2}(2\hat{s}^2;2) = \frac{1}{2}\exp\left(-\hat{s}^2\right)
\tag{1.1.G.2}
$$

と表されます（ここでは実際の計算はせずに，公式をあてはめました）。このとき標準化された標本平均の不偏標準偏差に相当する $\hat{s}$ の分布は，$\hat{s} \geq 0$ で非ゼロ値をとるとして，

$$
f_2\left(\hat{s}\right) = \frac{\mathrm{d}(\hat{s}^2)}{\mathrm{d}\hat{s}} f_{\chi^2}(2\hat{s}^2;2) = 2\hat{s}\exp\left(-\hat{s}^2\right)
\tag{1.1.G.3}
$$

と表されます。検定統計量 $t \equiv x/\hat{s}$ の分布は，

$$
\begin{aligned}
f_{\text{Student}}(t;2) &= \int_0^\infty \int_{-\infty}^\infty \delta\left(t - \frac{x}{\hat{s}}\right) f_{\text{SN}}(x)\, f_2\left(\hat{s}\right)\,\mathrm{d}x\,\mathrm{d}\hat{s} \\
&= \frac{2}{\sqrt{2\pi}}\int_0^\infty \int_{-\infty}^\infty \delta\left(t - \frac{x}{\hat{s}}\right)\exp\left(-\frac{x^2}{2}\right)\hat{s}\exp\left(-\hat{s}^2\right)\,\mathrm{d}x\,\mathrm{d}\hat{s}
\end{aligned}
\tag{1.1.G.4}
$$

と表されます。積分変数 x を

$$
\begin{aligned}
&\frac{x}{\hat{s}} - t \equiv u \quad \Rightarrow \quad x = \hat{s}\,(u+t) \\
&\mathrm{d}x = \hat{s}\,\mathrm{d}u \\
&x \;:\; -\infty \;\to\; \infty \\
&u \;:\; -\infty \;\to\; \infty
\end{aligned}
$$

として変数 u に変換すれば

$$
\begin{aligned}
f_{\text{Student}}(t;2) &= \frac{2}{\sqrt{2\pi}} \int_0^\infty \int_{-\infty}^\infty \delta(u) \exp\left[-\frac{\hat{s}^2 (u+t)^2}{2}\right] \hat{s} \exp\left(-\hat{s}^2\right) \left(\hat{s}\,\mathrm{d}u\right) \mathrm{d}\hat{s} \\
&= \frac{2}{\sqrt{2\pi}} \int_0^\infty \hat{s}^2 \exp\left(-\frac{\hat{s}^2 t^2}{2}\right) \exp\left(-\hat{s}^2\right) \mathrm{d}\hat{s} \\
&= \frac{2}{\sqrt{2\pi}} \int_0^\infty \hat{s}^2 \exp\left[-\left(1+\frac{t^2}{2}\right)\hat{s}^2\right] \mathrm{d}\hat{s} \\
&= \frac{1}{\sqrt{2\pi}} \left\{ \left[-\frac{\hat{s}}{1+t^2/2} \exp\left[-\left(1+\frac{t^2}{2}\right)\hat{s}^2\right] \right]_0^\infty + \frac{1}{1+t^2/2} \int_0^\infty \exp\left[-\left(1+\frac{t^2}{2}\right)\hat{s}^2\right] \mathrm{d}\hat{s} \right\} \\
&= \frac{1}{\sqrt{2\pi}} \left(1+\frac{t^2}{2}\right)^{-1} \int_0^\infty \exp\left[-\left(1+\frac{t^2}{2}\right)\hat{s}^2\right] \mathrm{d}\hat{s} \\
&= \frac{1}{\sqrt{2\pi}} \left(1+\frac{t^2}{2}\right)^{-1} \frac{\sqrt{\pi}}{2\sqrt{1+t^2/2}} \\
&= \frac{1}{2\sqrt{2}} \left(1+\frac{t^2}{2}\right)^{-3/2} \qquad (1.1.\mathrm{G}.5)
\end{aligned}
$$

となります。（↩）

（補足 1.1.H）自由度 ν の Student t 分布の確率密度函数（↩）

自由度 ν の場合には，個数 $n = \nu + 1$ の標本値 $\{X_1, \cdots, X_n\}$ が独立で母平均 μ，母標準偏差 σ の正規分布に従うとすれば，標本平均 $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n}$ は，平均 μ，標準誤差 $\dfrac{\sigma}{\sqrt{n}}$ の正規分布に従い，確率変数 $x = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ の統計分布の確率密度函数が

$$
f_{\mathrm{SN}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad (1.1.\mathrm{H}.1)
$$

に従います。

一方で，$x_j = \dfrac{X_j - \mu}{\sigma/\sqrt{n}}$ として標準化された標本値 $\{x_1, \cdots, x_n\}$ の標本分散に相当する値 $\hat{s}^2 = \dfrac{\hat{\sigma}^2}{\sigma^2/n}$ は，

$$
\hat{s}^2 = \frac{\hat{\sigma}^2}{\sigma^2/n} = \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{(n-1)(\sigma^2/n)}
$$

と表されます。

$\hat{s}^2$ の統計分布は自由度 $\nu = n - 1$ のカイ自乗分布に従い，その確率密度函数は

$$f_{\chi^2}\left(\hat{s}^2;\nu\right)=\frac{1}{2\,\Gamma(\nu/2)}\left(\frac{\hat{s}^2}{2}\right)^{\nu/2-1}\exp\left(-\frac{\hat{s}^2}{2}\right) \tag{1.1.H.2}$$

と表されます。このとき標準化された標本平均の不偏標準偏差に相当（そうとう）する $\hat{s}$ の分布は，$\hat{s}\geq 0$ で非ゼロ値をとるとして，

$$\begin{aligned}f_{\hat{s}}\left(\hat{s}\right)&=\frac{\mathrm{d}(\hat{s}^2)}{\mathrm{d}\hat{s}}f_{\chi^2}\left(\nu\,\hat{s}^2;\nu\right)=\frac{2\,\hat{s}}{2\,\Gamma(\nu/2)}\left(\frac{\hat{s}^2}{2}\right)^{\nu/2-1}\exp\left(-\frac{\hat{s}^2}{2}\right)\\&=\frac{\hat{s}^{\nu-1}}{2^{\nu/2-1}\,\Gamma(\nu/2)}\exp\left(-\frac{\hat{s}^2}{2}\right)=\frac{n^{\nu/2-1/2}\hat{\sigma}^{\nu-1}}{2^{\nu/2-1}\,\Gamma(\nu/2)\sigma^{\nu-1}}\exp\left(-\frac{n\,\hat{\sigma}^2}{2\sigma^2}\right)\end{aligned} \tag{1.1.H.3}$$

と表されます。$\hat{\sigma}$ の分布はの確率密度函数は，

$$\begin{aligned}f_{\hat{\sigma}}\left(\hat{\sigma}\right)&=\frac{\mathrm{d}(\sqrt{n}\,\hat{\sigma}/\sigma)}{\mathrm{d}\hat{\sigma}}f_{\hat{s}}\left(\hat{s}\right)=\frac{n^{\nu/2}\hat{\sigma}^{\nu-1}}{2^{\nu/2-1}\,\Gamma(\nu/2)\sigma^{\nu}}\exp\left(-\frac{\hat{s}^2}{2}\right)\\&=\frac{n^{n/2-1/2}\hat{\sigma}^{n-2}}{2^{n/2-3/2}\,\Gamma(n/2-1/2)\sigma^{n-1}}\exp\left(-\frac{n\,\hat{\sigma}^2}{2\sigma^2}\right)\end{aligned} \tag{1.1.H.4}$$

と表され，このとき検定統計量 $t\equiv x/\hat{s}$ の分布は，

$$\begin{aligned}f_{\mathrm{Student}}\left(t;\nu\right)&=\int_0^\infty\int_{-\infty}^\infty\delta\left(t-\frac{x}{\hat{s}}\right)f_{\mathrm{SN}}\left(x\right)f_\nu\left(\hat{s}\right)\,\mathrm{d}x\,\mathrm{d}\hat{s}\\&=\frac{\nu^{\nu/2}}{\sqrt{2\pi}\,2^{\nu/2-1}\,\Gamma(\nu/2)}\int_0^\infty\int_{-\infty}^\infty\delta\left(t-\frac{x}{\hat{s}}\right)\exp\left(-\frac{x^2}{2}\right)\hat{s}^{\nu-1}\exp\left(-\frac{\nu\,\hat{s}^2}{2}\right)\,\mathrm{d}x\,\mathrm{d}\hat{s}\end{aligned} \tag{1.1.H.5}$$

と表されます。

$$\begin{aligned}&\frac{x}{\hat{s}}-t\equiv u\quad\Rightarrow\quad x=\hat{s}\left(u+t\right)\\&\mathrm{d}x=\hat{s}\,\mathrm{d}u\end{aligned}$$

から

$$\begin{aligned}f_{\mathrm{Student}}\left(t;\nu\right)&=\frac{\nu^{\nu/2}}{\sqrt{2\pi}\,2^{\nu/2-1}\,\Gamma(\nu/2)}\int_0^\infty\int_{-\infty}^\infty\delta\left(u\right)\exp\left[-\frac{\hat{s}^2\left(u+t\right)^2}{2}\right]\hat{s}^{\nu-1}\exp\left(-\frac{\nu\,\hat{s}^2}{2}\right)\left(\hat{s}\,\mathrm{d}u\right)\,\mathrm{d}\hat{s}\\&=\frac{\nu^{\nu/2}}{\sqrt{2\pi}\,2^{\nu/2-1}\,\Gamma(\nu/2)}\int_0^\infty\hat{s}^{\nu}\exp\left(-\frac{\hat{s}^2t^2}{2}\right)\exp\left(-\frac{\nu\,\hat{s}^2}{2}\right)\,\mathrm{d}\hat{s}\\&=\frac{\nu^{\nu/2}}{\sqrt{2\pi}\,2^{\nu/2-1}\,\Gamma(\nu/2)}\int_0^\infty\hat{s}^{\nu}\exp\left[-\frac{\nu}{2}\left(1+\frac{t^2}{\nu}\right)\hat{s}^2\right]\,\mathrm{d}\hat{s}\end{aligned} \tag{1.1.H.6}$$

となります。以下の変数変換

$$\frac{\nu}{2}\left(1+\frac{t^2}{\nu}\right)\hat{s}^2\equiv u \tag{1.1.H.7}$$

を用いれば，

$$\hat{s}=2^{1/2}\nu^{-1/2}\left(1+\frac{t^2}{\nu}\right)^{-1/2}u^{1/2} \tag{1.1.H.8}$$

$$\mathrm{d}\hat{s}=2^{1/2}\nu^{-1/2}\left(1+\frac{t^2}{\nu}\right)^{-1/2}\frac{u^{-1/2}\,\mathrm{d}u}{2}=\frac{\nu^{-1/2}}{\sqrt{2}}\left(1+\frac{t^2}{\nu}\right)^{-1/2}\frac{\mathrm{d}u}{u^{1/2}} \tag{1.1.H.9}$$

から，スチューデント t 分布の確率密度函数は

$$f_{\text{Student}}(t;\nu) = \frac{\nu^{\nu/2}}{\sqrt{2\pi}\,2^{\nu/2-1}\,\Gamma(\nu/2)} \int_0^\infty 2^{\nu/2}\nu^{\nu/2}\left(1+\frac{t^2}{\nu}\right)^{-\nu/2} u^{\nu/2}\,\exp(-u)\,\frac{\nu^{-1/2}}{\sqrt{2}}\left(1+\frac{t^2}{\nu}\right)^{-1/2}\frac{\mathrm{d}u}{u^{1/2}}$$

$$= \frac{1}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \int_0^\infty u^{\nu/2-1/2}\,\exp(-u)\,\mathrm{d}u$$

$$= \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\dfrac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{1.1.H.10}$$

となります。（↩）

（補足 1.1.I）スチューデント t 分布と他の統計分布との関係（↩）

スチューデント t 分布 (Student's t - distribution) は，**ピアソン VII 型分布** (Pearson Type VII distribution) の「**標準化された形式**」(standardized formula) と見ることもできます。ピアソン VII 型分布の確率密度函数は，**ピアソン VII 型函数** Pearson Type VII function)とも呼ばれ，**ベータ函数** $\mathrm{B}(x,y)$ を使って

$$f_{\mathrm{P}7}(x;\lambda,\alpha,m) = \frac{1}{\alpha\mathrm{B}\left(m-\dfrac{1}{2},\dfrac{1}{2}\right)}\left[1+\left(\frac{x-\lambda}{\alpha}\right)^2\right]^{-m} \tag{1.1.I1}$$

と表されたり，

$$f_{\mathrm{P}7}(x;w,\mu) = \frac{\Gamma(\mu)}{\sqrt{\pi}\Gamma\left(\mu-\dfrac{1}{2}\right)w}\left(1+\frac{x^2}{w^2}\right)^{-\mu} \tag{1.1.I.2}$$

と言う表現 (Gupta, 1998) や，他の表現 (Hall, Jr., *et al.*, 1977; Young and Wiles, 1982) の用いられる例もあります。

自由度 2 のスチューデント t 分布の確率密度函数に相当する

$$f_{\chi^2}(t;2) = \frac{1}{2\sqrt{2}}\left(1+\frac{t^2}{2}\right)^{-3/2} \tag{1.1.I.3}$$

の形の函数が**中間ローレンツ型函数** (intermediate Lorentzian function)，自由度 3 のスチューデント t 分布の確率密度函数に相当する

$$f_{\chi^2}(t;3) = \frac{2}{\sqrt{3}\pi}\left(1+\frac{t^2}{3}\right)^{-2} \tag{1.1.I.4}$$

の形を持つ函数が**修正ローレンツ型函数** (modified Lorentzian function) と呼ばれる場合もあります (Young and Wiles, 1982)。（↩）

（補足 1.1.J）Student t 分布の確率密度函数，モンテカルロ・シミュレーション（←）

以下の３つのことを前提として t 検定が行われると考えても良いでしょう。

（１）標本平均 $\overline{X}$ の統計分布は，平均 μ，分散 $\dfrac{\sigma^2}{n}$ の正規分布 (normal distribution) に従う。

（2） $\dfrac{s^2(n-1)}{\sigma^2} = \chi^2$ は自由度 (degree of freedom) $n-1$ の χ^2 分布（カイじじょうぶんぷ）に従う。

（3） $Z = \overline{X} - \mu$ と $s = \dfrac{\hat{\sigma}}{\sqrt{n}}$ とは独立である。

そうだとして，これらのことからすぐに「ステューデント t 分布」を導くのは難しいでしょう。

ここではコンピュータを使ったモンテ・カルロ法 (Monte Carlo method)（マンティー カーロウ メソッド）で，本当に検定統計量 t の統計分布が「ステューデント t 分布」を満たすように見えるのかを調べてみます。

一般的に，自由度 $\nu = n-1$ のときに，標本点数は n 個です。標準正規分布に従う乱数 $X_1, X_2, \cdots, X_n$ を生成し，標本平均 $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ を計算します。標本標準偏差を $\hat{\sigma} = \sqrt{\dfrac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \cdots + \left(X_n - \overline{X}\right)^2}{n-1}}$ として計算するためには，繰り返し計算を２回しなければいけないことと，要素数 n の配列 $\left\{X_1, X_2, \cdots, X_n\right\}$ を準備しなければいけないことから，そのようにせずに済（す）む $\hat{\sigma} = \sqrt{\dfrac{n}{n-1}}\sqrt{\dfrac{X_1^2 + X_2^2 + \cdots + X_n^2}{n} - \left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right)^2}$ の形式で計算します。標本平均の標準偏差推定値（標準誤差）を $s = \dfrac{\hat{\sigma}}{\sqrt{n}}$ とします。検定統計量 $t = \dfrac{\overline{X}}{s}$ の値を計算し，適当な幅で区切った t の値のビン (bin) ごとに出現頻度のヒストグラム (histogram) を描かせます。

1 000 000 （百万（ひゃくまん））回の試行（しこう）を行い，400 ビンのヒストグラムを描（えが）かせました。結果を Figure 1.1.J.1 に示します。比較をしやすくするために，式 (1.1.5) を使って計算された Student t 分布（Figure 1.1.1）を再掲（さいけい）します。

モンテカルロ・シミュレーションの結果（Figure 1.1.J.1）は，Figure 1.1.1 に示した Student t 分布の形状とほぼ一致することがわかります。

1 000 000 （百万）回 の試行をすると言っても，要素数 1 000 000 （百万）の配列（メモリ領域）を確保する必要はなく，400 ビンのヒストグラムを描かせるために必要な配列の要素数は 400 です。ここではグラフ描画アプリケーション（Wavemetrics 社 Igor Pro ver. 8）のマクロ言語でモンテカルロ・シミュレーションのためのコーディングをしました。Figure 1.1.J.1 のヒストグラムを描かせるのに，自由度 $\nu = 100$ では約 1 億回乱数を発生することになりますが，それでも計算が完了するまでにかかった時間は約 8.6 s でした。Igor Pro マクロのコードを以下に示します。

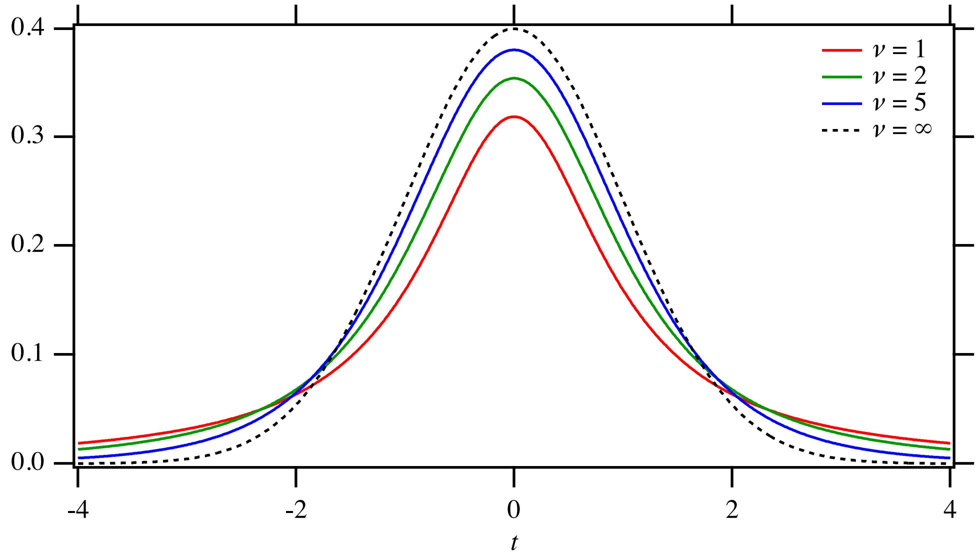

Figure 1.1.1　Student t 分布の確率密度函数。自由度を ν とする。（再掲）

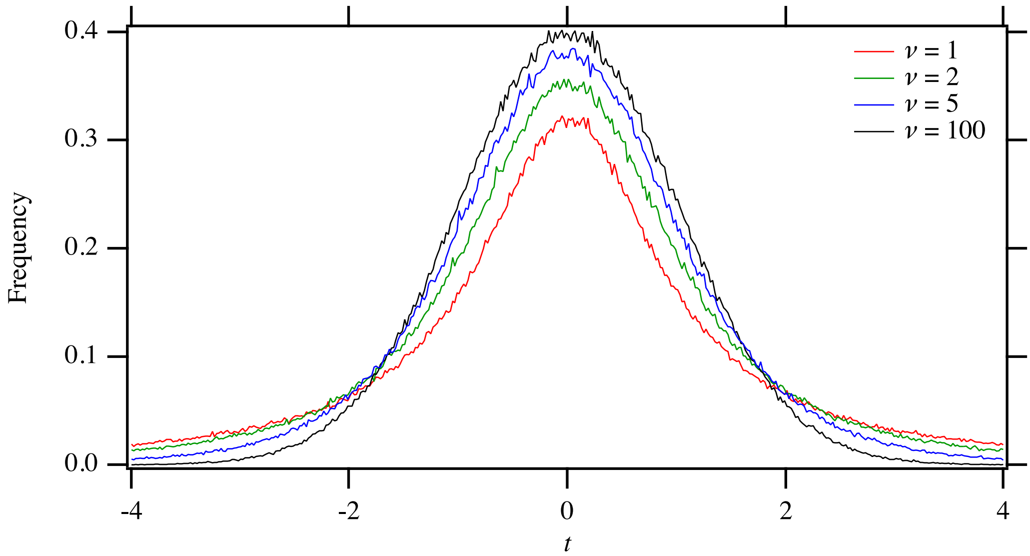

Figure 1.1.J.1　検定統計量 t の分布のモンテカルロ・シミュレーション（百万標本セット，400 ビン）

[Student t 分布のモンテカルロ・シミュレーションを行う Igor Pro マクロのコード]

```
//********************************
// Mote Carlo simulation for
// Student's t-distribution
// coded by T. Ida, 08 Jul. 2021
//********************************
Function MakeHistogram(df, sWaveName)
	variable df; // Degree of freedom
	string sWaveName; // Destination wave name
	variable nSample = df + 1; // Number of samples
	variable vHalfW = 4.0; // Half width for display area
	variable nBin = 400; // Number of bins
	variable vBinW = 2*vHalfW/nBin; // Width of bin
	variable vLeft = -vHalfW + 0.5*vBinW; // (Shift by +half step)
	MAKE/D/O/N=(nBin) $sWaveName; // Make a wave for frequency
	WAVE wFreq = $sWaveName; // Pointer for the frequency wave
	SetScale/P X (vLeft),(vBinW),"",wFreq; // Set abscissa scale for the wave
	wFreq = 0; // Reset the frequency
	variable nTrial = 1000000; // Number of trials
	variable iTrial; // Counter for trials
	for (iTrial = 0; iTrial < nTrial; iTrial += 1)
		variable iSample; // Counter for sample values
		variable vS1 = 0, vS2 = 0; // Reset sums
		For (iSample = 0; iSample < nSample; iSample += 1)
			variable vRnd = gnoise(1); // Generate a normal random number
			vS1 += vRnd; // Add the value to vS1
			vS2 += vRnd^2; // Add the squared value to vS2
		EndFor; // (iSample = 0; iSample < nSample; iSample += 1)
		variable vSampleMean = vS1 / nSample; // Sample mean
		variable vSampleVar; // Sample variance
		vSampleVar = vS2 / nSample - vSampleMean^2; // Biased sample variance
		vSampleVar *= nSample / (nSample - 1); // Unbiased sample variance
		variable vSampleStd = sqrt(vSampleVar); // Sample std (chi_hat)
		variable vStdErr = vSampleStd/sqrt(nSample); // Standard error (s)
		variable t = vSampleMean / vStdErr; // Test statistic (t)
		variable iBin; // Index of bin
		iBin = round((t + vHalfW)/vBinW); // Find the bin to be assigned
		If ((0 <= iBin) %& (iBin < nBin))
			wFreq[iBin] += 1; // Count frequency
		EndIf; // ((0 <= index) %& (index < nBin))
	EndFor; // (iTrial = 0; iTrial < nTrial; iTrial += 1)
	wFreq /= vBinW * nTrial; // Calculate probability density
End Function; // MakeHistogram(df, sWaveName)
```

(↩)

（補足 1.2.A）両側検定について（↩）

特に「医学・薬学」の分野では，「**両側検定**をするべき」という文言（もんごん）を多く見ます（*c.f.* 永田, 1996, Altman, 1999, 丹後, 2013, Armitage, 2011, 浜田, 1999）。「**片側検定**」を「**両側検定**」に置き換えれば，**第一種過誤**（有効でない医薬品・ワクチン・治療法を有効と勘違いしてしまうこと）の出現確率を 1/2 にすることはできますが，**第二種過誤**（有効な医薬品・ワクチン・治療法を無効と勘違いしてしまうこと）の出現確率を 2 倍にします。（↩）

（補足 1.3.2.A）SciPy ライブラリ `scipy.stats.ttest_1samp()` の使用例（↩）

SciPy ライブラリの `stats.ttest_1samp()` メソッドを利用して，一標本 t 検定を試（こころ）みます。ここでは Python インタープリタのインタラクティブ・ユーザー・インターフェスを使うことにします。

はじめに，乱数（ランダム ナンバーズ） (random numbers) を使って平均値 50，標準偏差 10の正規分布（せいきぶんぷ） (normal（ノーマル） distribution（ディストリビューション）) にしたがう 100 標本のデータを一組作ってみます。ここでは `stats.norm.rvs()` 函数（かんすう）を使います。

［乱数を利用したテストデータの作成，Python インタープリタでの指定］

```
>>> import numpy as np
>>> from scipy import stats
>>> rvs = stats.norm.rvs(loc=50, scale=10, size=100)
```

ここでは `stats.norm.rvs()` 函数（かんすう）の引数（ひきすう） `loc=50` (location) で平均 50，`scale=10` で標準偏差 10を指定し，`size=100` で 100 標本点のデータを一次元 NumPy 配列 `rvs` として出力するよう指示しています。

［テストデータの内容の確認］

```
>>> rvs
array([39.37112101, 42.65777518, 51.78769476, 40.51518578, 53.39914632,
       56.40734135, 62.5311289 , 51.01165136, 39.20558351, 54.27488574,
       51.98254918, 38.84428037, 56.50505084, 52.95886257, 63.75415962,
       38.0303666 , 35.99411362, 37.7567901 , 58.0685928 , 43.22119157,
       34.37611903, 32.45342762, 49.44403389, 50.55918466, 49.47280203,
       45.65095524, 47.35333301, 55.89209185, 41.98283523, 37.77347701,
       48.65106808, 55.48648519, 59.45953231, 50.32934   , 65.25318289,
       34.34704793, 61.28539786, 37.25466441, 50.37489447, 50.91267812,
       47.04671962, 61.64205938, 55.82089844, 74.89241664, 45.24095406,
       47.02855637, 65.47958679, 56.78670491, 68.11637444, 44.61758141,
       73.69925204, 55.96417461, 52.50537696, 51.14949382, 38.35238909,
       50.87053938, 40.80289696, 42.44736765, 50.4275475 , 52.74176496,
       67.02019812, 59.51386747, 50.7117099 , 48.82668764, 31.13429586,
       41.60196398, 48.21016066, 52.83986113, 59.34931082, 53.51397197,
       49.78485044, 54.31486904, 23.17425458, 54.88999518, 38.35677884,
       50.77971115, 43.36948629, 42.19408506, 52.96057433, 45.20001115,
       35.08305628, 41.22150972, 61.67982236, 37.11304266, 56.99029506,
       40.83530186, 34.01945876, 36.03457118, 54.17970817, 32.24621509,
       59.05674804, 41.5357389 , 55.4745467 , 62.04105861, 40.38658418,
       41.78441869, 50.23625498, 39.27906067, 39.56362075, 49.02050009])
```

このようにして作成した `rvs` の標本平均と標本標準偏差の値を確認するために，`numpy.mean()` メソッドと `numpy.std()` メソッドを使うことができます。

［テストデータの標本標準偏差と標準偏差推定値の確認］

```
>>> np.mean(rvs)
48.77750833393443
>>> np.std(rvs,ddof=1)
9.893895989958118
```

ここでは標本平均の値が 48.8，標本標準偏差の値が 9.9という結果になりました。 `numpy.std()` メソッドの引数（ひきすう） `ddof` は “delta degree of freedom” の略とされており，`ddof=1` を指定すれば普通の意味での標本標準偏差が計算される仕様になっています。「標本平均の標準偏差推定値（標準誤差）」は $10/\sqrt{100} = 1.0$ と見積もられるので，標本平均 (sample mean) の値が母平均（ぼへいきん） (population mean) の値 50 から「1.0 かその 2 倍程

度ずれること」は普通のことです。また，このときに小数点以下第一位以下の数字は「あまり意味がない」ということにも注意してください。

かりに帰無仮説を「母平均が 50 以下の値である」としたときに，この帰無仮説についての t 検定を行った結果を最も手早く表示するには，例えば以下のようにします。

［「母平均が 50 以下である」とする帰無仮説についての t 検定（デフォルト指定）］

```
>>> stats.ttest_1samp(rvs, popmean=50)
Ttest_1sampResult(statistic=-1.235601897681509, pvalue=0.21953058122797478)
```

`scipy.stats.ttest_1samp()` メソッドの出力は Python tuple であり，検定統計量 (test statistic) `statistic` とデフォルト（省略時設定）として両側 P 値 (two-sided P-value) が `pvalue` として出力されます。検定統計量 (`statistic`) 出力が，以下の式：

$$t_{\mathrm{ref}} = \frac{\overline{X} - \mu_{\mathrm{ref}}}{\hat{\sigma}/\sqrt{n}}\,, \tag{1.3.2.A.1}$$

で計算される値であることは，以下のように確認できます。

［“scipy.stat.ttest_1samp()” メソッドの “statistic” 出力の確認］

```
>>> (np.mean(rvs)-50)*np.sqrt(100)/np.std(rvs,ddof=1)
-1.235601897681509
```

デフォルト指定での `scipy.stats.ttest_1samp()` メソッドの “`pvalue`” 出力が以下の式：

$$2P(t > |t_{\mathrm{ref}}|) = 2P(t < -|t_{\mathrm{ref}}|) = I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t_{\mathrm{ref}}^2 + \nu}\right), \tag{1.3.2.A.2}$$

で表されるような「両側 P 値」であることは，以下のように確認できます。

［“scipy.stat.ttest_1samp()” メソッドの “pvalue” 出力の確認］

```
>>> from scipy import special as sp
>>> tref,p=stats.ttest_1samp(rvs,popmean=50)
>>> sp.betainc(99/2,1/2,99/(tref**2+99))
0.21953058122797467
```

検定統計量 (`statistic`) 出力の値が負なので，伝統的な t 検定での「帰無仮説」は「母平均 μ が参照値 μ_{ref} より大きいこと ($\mu > \mu_{\mathrm{ref}}$)」あるいは同じ意味を持つ “$t < t_{\mathrm{ref}}$” ということになり「下側 P 値」が計算されます。この下側 P 値は

$$P(\mu > \mu_{\mathrm{ref}}) = P(t < t_{\mathrm{ref}}) = \frac{1}{2}\, I\left(\frac{\nu}{2}, \frac{1}{2}; \frac{\nu}{t_{\mathrm{ref}}^2 + \nu}\right), \tag{1.3.2.A.3}$$

と表され，例えば以下のようにして数値を示すこともできます。

［“scipy.stat.ttest_1samp()” のデフォルト出力から片側 P 値を得る］

```
>>> p/2
0.10976529061398739
```

ここではこの確率 $P(\mu > \mu_{\mathrm{ref}})$ が 0.110 (11.0%) とみなせます。下側 P 値の表示を明示的に指定するために，オプション引数の “`alternative`” を指定することもできます。

[“scipy.stat.ttest_1samp()” での下側 P 値表示の指定]

```
>>> stats.ttest_1samp(rvs,popmean=50,alternative='less')
Ttest_1sampResult(statistic=-1.235601897681509, pvalue=0.10976529061398739)
```

“alternative” は対立仮説 (alternative hypothesis) に由来すると思われます。ただし，P 値は帰無仮説の成立する確率に相当することなど，「混乱してしまいやすそう」なことも残っており，オプション引数の “alternative” を明示的に指定すれば「わかりやすくなる」ということではないかもしれません。

「95% 信頼区間」(95% confidence interval) は「母平均が 95% 以上の確率で含まれる区間」と解釈しても良いでしょう（「それは間違いだ」と言う人もいます）。母平均 μ の確率分布が平均 $\overline{X}$，尺度パラメータ $\hat{s}$，自由度 ν の（尺度化された）Student t 分布に従うとすれば，母平均推定での「95% 信頼区間」の下限 $\mu_{\rm L}$ と上限 $\mu_{\rm U}$ は以下のように求められます。$c = 0.05$ とします。

$$P(\mu < \mu_{\rm L}) = P(\mu > \mu_{\rm U}) = \frac{c}{2} \ \Rightarrow\ P(t > t_{\rm c}) = P(t < -t_{\rm c}) = \frac{c}{2},$$

$$t_{\rm c} = \frac{\overline{X} - \mu_{\rm L}}{\hat{s}} = -\frac{\overline{X} - \mu_{\rm U}}{\hat{s}} \ \Rightarrow\ \begin{array}{l} \mu_{\rm L} = \overline{X} - \hat{s}\, t_{\rm c} = \overline{X} - \dfrac{\hat{\sigma}\, t_{\rm c}}{\sqrt{n}} \\ \mu_{\rm U} = \overline{X} + \hat{s}\, t_{\rm c} = \overline{X} + \dfrac{\hat{\sigma}\, t_{\rm c}}{\sqrt{n}} \end{array}$$

$$F_{\rm Student}(-t_{\rm c};\nu) = 1 - F_{\rm Student}(t_{\rm c};\nu) = \frac{c}{2} \ \Rightarrow\ I\left(\frac{\nu}{2},\frac{1}{2};\frac{\nu}{t_{\rm c}^2+\nu}\right) = c$$

$$\Rightarrow\ \frac{\nu}{t_{\rm c}^2+\nu} = I^{-1}\left(\frac{\nu}{2},\frac{1}{2};c\right) \ \Rightarrow\ \frac{t_{\rm c}^2}{\nu} + 1 = \frac{1}{I^{-1}(\nu/2,1/2;c)} \ \Rightarrow\ t_{\rm c}^2 = \nu\left[\frac{1}{I^{-1}(\nu/2,1/2;c)} - 1\right]$$

$$\Rightarrow\ t_{\rm c} = \left\{\nu\left[\frac{1}{I^{-1}(\nu/2,1/2;c)} - 1\right]\right\}^{1/2},$$

ここで，$I^{-1}(a,b;y)$ は正則不完全ベータ函数 (regularized incomplete beta function) $I(a,b;x)$ の逆函数であり，SciPy ライブラリから “scipy.special.betaincinv(a,b,y)” として呼び出すことができます。下限 $\mu_{\rm L}$ と上限 $\mu_{\rm U}$ を求めて，一標本 t 検定で結果の正しさを確かめるには，例えば以下のようにします。

[母平均推定での 95% 信頼区間の求め方]

```
>>> import numpy as np
>>> from scipy import special as sp
>>> from scipy import stats
>>> t_c=np.sqrt(99*(1/sp.betaincinv(99/2,1/2,0.05)-1))
>>> muL=np.mean(rvs)-np.std(rvs,ddof=1)*t_c/np.sqrt(100)
>>> muU=np.mean(rvs)+np.std(rvs,ddof=1)*t_c/np.sqrt(100)
>>> muL,muU
(46.81434471988365, 50.740671947985206)
>>> stats.ttest_1samp(rvs,popmean=muL,alternative='greater')
Ttest_1sampResult(statistic=1.9842169515864179, pvalue=0.024999999999999984)
>>> stats.ttest_1samp(rvs,popmean=muU,alternative='less')
Ttest_1sampResult(statistic=-1.9842169515864179, pvalue=0.024999999999999984)
```

ここでは母平均推定についての「95% 信頼区間」は [46.8, 50.7] という結果になりました。

(↩)

（補足 1.3.3.A）ステューデント t 検定の文脈の理解のしかた（↩）

ステューデント t 検定では「5% の有意水準で帰無仮説が棄却された」 “The null hypothesis is rejected for a significance level of 5%” と言うような表現のとられるのが普通のことでした。

もう少し平たい言い方をすれば，例えば一標本 t 検定の場合（正確にはデータの統計的な確率分布が正規分布に従う場合に）「推定値 μ がゼロまたは負の値である確率は 5% 以下と見積もられた」という意味合いで使われます。このときに「5% の有意水準で帰無仮説が棄却された」と言う人が「何を言いたいのか」と言うと「推定値 μ がゼロまたは負の値であるという仮説は，95% 以上程度の確率で**間違っている**」というようなことです。

「それならそう言えば良いのに」と思われるかもしれません。しかし，データの確率分布が正規分布に従わなかったとしても，「t 検定により 5% の有意水準で帰無仮説が棄却された」という表現は「間違い・嘘」にはなりません。「データの確率分布が正規分布に従うかは確認できないが，t 検定をしたら，5% の有意水準で帰無仮説が棄却される結果となった」という表現も成立します。この場合に「5%」という数値は，「正しく確率を表す数値」ではありませんが，例えば「統計学的に分析した結果，帰無仮説は成立しそうにないことがわかった」などと言われるより，説得力もありますし，意味のある情報が付け加えられています。

また「データの確率分布が正規分布に従うこと」について，「そのように仮定することを正当化するための充分な根拠がある」場合も確かにあるのでしょうけれど，「そうでない場合」も実際には多いのが現実です。「データが厳密に正規分布に従うのでなければ t 検定は意味がない」と言うのであれば「t 検定は，ほとんどの場合に意味がない」ことになりますが，そのようなことはありません。「その現象であれば，本当の統計分布が正規分布からどのようにずれるか知っている」人なら，「その結果をどちらの方向に修正すれば正しい結論に近づけそうだ」とわかる場合もあります。わからない人にとっては「意味のないこと」かもしれませんが，わかる人にとっては「意味のないこと」ではありません。

ステューデント t 検定を実験的な研究で用いる多くの場合に，「**帰無仮説**」とは，否定される・棄却されることを前提として意図的に「たてる」ものであり，「**対立仮説**」は，採用されることを前提として意図的に「たてる」ものと言われます（*e.g.*, 青木, 1982）。確かに，そうでなければ，わざわざ実験をしたり，データの収集・統計解析などをする必要がありません。「学会」のように「研究者」と呼ばれる職種の人たちの集まりでも「実験結果が理論とよく合いました」というタイプのことを言う人が少なくないのですが「それなら，そもそも実験などをする必要はなかったのでないですか？」と質問されることになるかもしれません。

大学生の学生実験であれば「理論通りの結果になった。よかった。よかった」と思うのでも良いでしょうけれど，プロの研究者は「従来の理論」や「従来の予測」「従来の方法」などを否定するために実験をします。大学の教員にはプロの研究者に近い人と，そうでもない人とがいると思いますが，なるべく早い時期に「プロフェッショナルな研究者がする研究とは，どのようなことか」を知っておくと良いでしょう。

もちろん「5% の有意水準で帰無仮説は棄却されなかった」という「残念な結果」になる場合も，実際には多いでしょう。そのときに一標本 t 検定の結果として言えることがあるとすれば，例えば「推定値 μ が正の値である確率は 5% 程度以上と見積もられた」と言うことと同じです。「正しい確率が 5% 程度以上」と言われても，正しいのか間違っているのか「まったくわからなかった」のと「ほぼ同じこと」ですから，「そんなことをわざわざ言う輩はいない」はずです（実際には少なからずいます）。

「5% 以上の確率」とは，例えば [1] から [36] までのスロットと [0] と [00] の 38 スロットを持つルーレットで「[0] か [00] のスロットに玉が入る確率 $2/38 \approx 0.053$」と同程度です。「確率が 5% 以上あるので，[0] と [00] のスロットにだけ賭け続ける人」はいるかもしれませんが，普通ではないでしょう。

「5% の有意水準で帰無仮説は棄却されなかった」という「残念な結果」になった場合でも，そのことに意味がないわけではありません。「もっと多くの回数の実験をし，多くのデータを集めるべきだった」「帰無仮説・対立仮説の立て方が良くなかった」「有意水準を厳しく設定しすぎた」「実験計画の立て方が良くなかった」「そもそも何を言おうとするか決めていなかった」などの反省をするために，良い経験になるでしょう。そして，次からは「何を言おうとするか」「どのような帰無仮説・対立仮説をたてるか」をはっきりとさせ「そのために必要になるデータの規模・データを集めるのに必要な時間・労力・経費」を予測します。そして実現不可能と予想するならあきらめれば良いでしょうし，実現可能と予想するのであれば，それに合わせて実験計画を立てるようにすれば良いのです。このような研究のしかたが「**前向きな研究**」と表現される場合もあります（青木, 1982）。あるいは「実現可能と予想される範囲で棄却できそうな帰無仮説とそれに対する対立仮説をたてる」ことにします。

そのようなことをするのは「作為的」(artificial) 「意図的」 (intentional) すぎるので，するべきではないと思う人もいるかもしれません。一方で，「t 検定」とは，「数学の問題」と言うより，「修辞学」 (rhetoric) （説得力を強化する表現技術）的な技法の一つと割り切ってしまえば，むしろわかりやすくなる場合が多いと思います。 (↩)

（補足 1.4.1.A）有意水準と有意確率，t 分布表（↩）

コンピュータ利用がそれほど一般的ではなかった 1980 年代でも「『有意水準』と言うような『人によって選び方が違うような曖昧な数値』だけでなく，P 値 (p-value) （有意確率）（第一種過誤の出現確率）も示すべきだ」という意見はありました（青木, 1982）。しかし，その時点では， Student t 分布の積分値を正確に計算できる計算システムを利用することは難しい場合も多かったかもしれませんし，「『t 分布表』を使うしかないので『有意水準』を使うのもやむを得ない」という状況だったかもしれません。

現在は「コンピュータを使えば『t 分布表』を使う必要もないし，『有意水準』は『数値を切り上げてキリの良い数として表現する以外の意味はない』」でしょう。「今でも『t 分布表』『有意水準』を使うなどというのは，時代錯誤だ」という印象を持たれるかもしれません。

ただし，コンピュータを使うのでも「メモリ上に数表 (numerical table) を保持して参照することによって，数値計算を高速化する計算システム設計」は存在し，大量のデータに対して高速に t 検定をするシステムが必要になる場合があれば，「t 分布表」を使うタイプのコーディングをするかもしれませんから，「t 分布表にはまったく意味がない」というほどではありません。

また，人によっては，文章で説明されなくても，数表を眺めて数字の変化の挙動・規則性から本質的な意味を見出す場合もあるでしょう。実験装置を設計・製作したり，適切な実験計画を立てたり，高効率・高精度な解析プログラムを開発するために必要になる「数感覚 (number sense)」「量感覚 (quantity sense)」のようなものが，数表を眺めることから得られることもあります。そのような「感覚 (sense)」は，「論理 (logics)」や「言語を用いたコミュニケーション (linguistic communication) の技術」とは「まったく違うもの」と思って良いくらいです。「t 分布表」が，教育・学習の目的で有効な場合があること自体は間違っていないだろうと思われます。 (↩)

（補足 1.5.1.A）畳込の原始関数の数値的な解法（↩）

式 (1.5.1.1) に示したように

$$f(\Delta\mu) = \frac{1}{\hat{s}_A \hat{s}_B} \int_{-\infty}^{\infty} f_{\text{Student}}\left(\frac{\Delta\mu + t - \overline{A} + \overline{B}}{\hat{s}_A}; \nu_A\right) f_{\text{Student}}\left(\frac{t}{\hat{s}_B}; \nu_B\right) \mathrm{d}t \tag{1.5.1.A.1}$$

と表されるような「畳込」(convolution) （正確には「相関」(correlation)；ただしこの場合はどちらも同じこと）に対して，式 (1.5.1.2) に示したように

$$P(\Delta\mu < 0) = \frac{1}{\hat{s}_A \hat{s}_B} \int_{-\infty}^{0} \int_{-\infty}^{\infty} f_{\text{Student}}\left(\frac{\Delta\mu + y - \overline{A} + \overline{B}}{\hat{s}_A}; \nu_A\right) f_{\text{Student}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \mathrm{d}y \, \mathrm{d}\Delta\mu \tag{1.5.1.A.2}$$

で表されるような「畳込」の積分を数値的な方法で解くことについて考えます。

$$\begin{aligned}
P(\Delta\mu < 0) &= \frac{1}{\hat{s}_B} \int_{-\infty}^{\infty} \left[\frac{1}{\hat{s}_A} \int_{-\infty}^{0} f_{\text{Student}}\left(\frac{\Delta\mu + y - \overline{A} + \overline{B}}{\hat{s}_A}; \nu_A\right) \mathrm{d}\Delta\mu\right] f_{\text{Student}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \mathrm{d}y \\
&= \frac{1}{\hat{s}_B} \int_{-\infty}^{\infty} \left[\frac{1}{\hat{s}_A} \int_{-\infty}^{y-\overline{A}+\overline{B}} f_{\text{Student}}\left(\frac{u}{\hat{s}_A}; \nu_A\right) \mathrm{d}u\right] f_{\text{Student}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \mathrm{d}y \\
&= \frac{1}{\hat{s}_B} \int_{-\infty}^{\infty} F_{\text{Student}}\left(\frac{y - \overline{A} + \overline{B}}{\hat{s}_A}\right) f_{\text{Student}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \mathrm{d}y \\
&= \frac{1}{\hat{s}_B} \int_{-\infty}^{\infty} F_{\text{Student}}\left(\frac{y - \overline{A} + \overline{B}}{\hat{s}_A}\right) f_{\text{Student}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \mathrm{d}y
\end{aligned} \tag{1.5.1.A.3}$$

のように式を変形し， 変数変換

$$\begin{aligned}
\xi &= F_{\text{Student}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \quad \Leftrightarrow \quad y = \hat{s}_B \, F^{-1}_{\text{Student}}\left(\xi; \nu_B\right) \\
\mathrm{d}\xi &= \frac{1}{\hat{s}_B} f_{\text{Studnet}}\left(\frac{y}{\hat{s}_B}; \nu_B\right) \mathrm{d}y \\
y &: \; -\infty \; \rightarrow \; \infty \\
\xi &: \; 0 \; \rightarrow \; 1
\end{aligned}$$

を用いれば，

$$P(\Delta\mu < 0) = \int_0^1 F_{\text{Student}}\left(\frac{y - \overline{A} + \overline{B}}{\hat{s}_A}\right) \mathrm{d}\xi \tag{1.5.1.A.4}$$

$$y = \hat{s}_B \, F^{-1}_{\text{Student}}\left(\xi; \nu_B\right) \tag{1.5.1.A.5}$$

と表されます。

関数 $F_{\text{Student}}(t;\nu)$ は，式 (1.1.10) に示したように

$$F_{\text{Student}}(t;\nu) = \int_{-\infty}^{t} f_{\text{Student}}(u;\nu)\,\mathrm{d}u = \begin{cases} 1 - \dfrac{1}{2} I\left(\dfrac{\nu}{2}, \dfrac{1}{2}; \dfrac{\nu}{t^2+\nu}\right) & [t \geq 0] \\ \dfrac{1}{2} I\left(\dfrac{\nu}{2}, \dfrac{1}{2}; \dfrac{\nu}{t^2+\nu}\right) & [t < 0] \end{cases} \tag{1.5.1.A.6}$$

で表され，関数 $F_{\text{Student}}(t;\nu)$ の逆関数 $F^{-1}_{\text{Student}}(u;\nu)$ は，

$$F^{-1}_{\text{Student}}(u;\nu) = \begin{cases} \left\{\nu\left[\dfrac{1}{I^{-1}(\nu/2,1/2;2-2u)}-1\right]\right\}^{1/2} & \left[\dfrac{1}{2} \le u < 1\right] \\ -\left\{\nu\left[\dfrac{1}{I^{-1}(\nu/2,1/2;2u)}-1\right]\right\}^{1/2} & \left[0 < u < \dfrac{1}{2}\right] \end{cases} \tag{1.5.1.A.7}$$

と表されます。式 (1.5.1.A.4) と式 (1.5.1.A.5) で表される積分は，中点法 (mid-point method) による数値積分では，

$$P(\Delta\mu < 0) \approx \frac{1}{n}\sum_{i=0}^{n-1} F_{\text{Student}}\left(\frac{y_i - \overline{A} + \overline{B}}{\hat{s}_A}\right) \tag{1.5.1.A.8}$$

$$y_i = \hat{s}_B\, F^{-1}_{\text{Student}}\left(\xi_i;\, \nu_B\right) \tag{1.5.1.A.9}$$

$$\xi_i = \frac{i+0.5}{n} \tag{1.5.1.A.10}$$

と表され，ガウス・ルジャンドル法 (Gauss-Legendre quadrature) による数値積分では，

$$P(\Delta\mu < 0) \approx \frac{1}{2}\sum_{i=0}^{n-1} w_i\, F_{\text{Student}}\left(\frac{y_i - \overline{A} + \overline{B}}{\hat{s}_A}\right) \tag{1.5.1.A.8}$$

$$y_i = \hat{s}_B\, F^{-1}_{\text{Student}}\left(\xi_i;\, \nu_B\right) \tag{1.5.1.A.9}$$

$$\xi_i = \frac{1+x_i}{2} \tag{1.5.1.A.10}$$

$$F(x) = F_{\text{Student}}\left(\frac{x}{\hat{s}_A};\nu_A\right)$$

$$G(x) = F_{\text{Student}}\left(\frac{x}{\hat{s}_B};\nu_B\right)$$

とします。

次に，単純な事例で式 (1.5.1.A.8)–(1.5.1.A.10) で表される計算手法がどのように振る舞うかを調べます。。

$A = \{0,1\}$，$B = \{1,2\}$ とします。標本平均は $\overline{A} = 1/2$，$\overline{B} = 3/2$ であり，標本標準偏差は $\hat{\sigma}_A = \sqrt{[(1/2)^2 + (1/2)^2]/(2-1)} = 1/\sqrt{2}$，$\hat{\sigma}_B = 1/\sqrt{2}$，標本平均の標準偏差（標準誤差）は $\hat{s}_A = \hat{\sigma}_A/\sqrt{2} = 1/2$，$\hat{s}_B = 1/2$ となります。このとき母平均の推定値 μ_A と μ_B の統計分布を表す確率密度函数，$f_A(\mu_A)$ と $f_B(\mu_B)$ は

$$f_A(\mu_A) = \frac{1}{\hat{s}_A} f_{\text{Student}}\left(\frac{\mu_A - \overline{A}}{\hat{s}_A};1\right) \tag{1.5.1.A.11}$$

$$f_B(\mu_B) = \frac{1}{\hat{s}_B} f_{\text{Student}}\left(\frac{\mu_B - \overline{B}}{\hat{s}_B};1\right) \tag{1.5.1.A.12}$$

$$f_{\text{Student}}(x;1) = \frac{1}{\pi}\left(1+x^2\right)^{-1} \tag{1.5.1.A.13}$$

と表され，母平均の差 $\Delta\mu \equiv \mu_A - \mu_B$ の統計分布の確率密度函数 $f_{A-B}(\Delta\mu)$ は

$$f_{A-B}(\Delta\mu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\Delta\mu - \mu_A + \mu_B) f_A(\mu_A) f_B(\mu_B)\, \mathrm{d}\mu_A\, \mathrm{d}\mu_B = \int_{-\infty}^{\infty} f_A(\Delta\mu + \mu_B) f_B(\mu_B)\, \mathrm{d}\mu_B$$

$$
= \frac{1}{\pi^2 \hat{s}_A \hat{s}_B} \int_{-\infty}^{\infty} \left[1 + \frac{\left(\Delta\mu + \mu_B - \overline{A}\right)^2}{\hat{s}_A^2} \right]^{-1} \left[1 + \frac{\left(\mu_B - \overline{B}\right)^2}{\hat{s}_B^2} \right]^{-1} \mathrm{d}\mu_B
$$

$$
= \frac{1}{\pi \left(\hat{s}_A + \hat{s}_B\right)} \left[1 + \frac{\left(\Delta\mu - \overline{A} + \overline{B}\right)^2}{\left(\hat{s}_A + \hat{s}_B\right)^2} \right]^{-1} \tag{1.5.1.A.14}
$$

と表されます（補足 1.5.1.A.1）。

式 (1.5.1.A.14) で表される函数は，尺度化パラメータ $\hat{s}_A + \hat{s}_B$により尺度化された「自由度 1 のステューデント t 分布の確率密度函数」とも言えます。ここでの具体的な数値 $\overline{A} = 1/2$，$\overline{B} = 3/2$，$\hat{s}_A = 1/2$，$\hat{s}_B = 1/2$ をあてはめれば，

$$
f_{A-B}(\Delta\mu) = \frac{1}{\pi} \left[1 + \left(\Delta\mu + 1\right)^2 \right]^{-1}
$$

と表され，累積度数分布函数は

$$
F_{A-B}(\Delta\mu) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\Delta\mu + 1\right)
$$

となります。$\Delta\mu < 0$ となる確率は $P(\Delta\mu < 0) = F_{A-B}(0) = 1/2 + \arctan(1)/\pi = 1/2 + 1/4 = 3/4 = 0.75$ です。

一方で，式 (1.5.1.A.8)–(1.5.1.A.10) の形式を使う場合に，以下のような Python コードを実行すれば計算結果が得らます。

[二標本の母平均の比較をする Python コード（確認用） test151a.py]

```
import numpy as np
from scipy import special

# Two-sided P-value
def twosp(t,nu):
   return special.betainc(nu/2,0.5,nu/(t**2+nu))
# Cumulative distribution function of Students t-distribution
def cdf(t,nu):
   ans=np.where(t>0,1-0.5*twosp(t,nu),0.5*twosp(-t,nu))
   return ans

# Inverse function of two-sided P-value
def twosp_inv(u,nu):
   return np.sqrt(nu*(1/special.betaincinv(nu/2,0.5,u)-1))
# Inverse function of c.d.f. of Students t-distribution
def cdf_inv(u,nu):
   ans=np.where(u>0.5,twosp_inv(2-2*u,nu),-twosp_inv(2*u,nu))
   return ans

def cdf0(a,b,N):
   # a,b: sample groups (NumPy array)
   # N: sampling points for numerical integral
   n_a = a.size
   mean_a = np.mean(a)
   s_a = np.std(a,ddof=1) / np.sqrt(n_a)
   n_b = b.size
   mean_b = np.mean(b)
   s_b = np.std(b,ddof=1) / np.sqrt(n_b)
   xi_i = np.linspace(0.5/N,(N+0.5)/N,N,endpoint=False)
   # print(xi_i)
   t_i = s_b * cdf_inv(xi_i,n_a-1)
   term = cdf((t_i-mean_a+mean_b)/s_a,n_a-1)
   ans = np.sum(term) / N
   return ans

a=np.array([0,1])
b=np.array([1,2])
print('N = 10, P = ',cdf0(a,b,10))
print('N = 100, P = ',cdf0(a,b,100))
```

[test151a.py の実行と出力]

```
...$ python3 test151a.py
N = 10, P =  0.750994394782359
N = 100, P =  0.7500000000000001
```

確かに代数的な手法で求められた値 0.75 には，数値積分のための標本点数 N を増やすほど近づく傾向が表れていますが，この方法では $\left|\overline{A}-\overline{B}\right| \gg \hat{s}_A$ の場合には計算がうまくいかないことも予想されます。

（補足 1.5.1.A.1）ローレンツ型函数の畳込（↩）

ローレンツ型函数 (Lorentzian function) は，

$$f_{\text{Lorentz}}(x;w) = \frac{1}{\pi w}\left(1+\frac{x^2}{w^2}\right)^{-1} \tag{1.5.1.A.1.1}$$

と表されます。そのフーリエ変換 $\mathfrak{F}_{\text{Lorentz}}(k;w)$ は

$$\mathfrak{F}_{\mathrm{Lorentz}}(k\,;w) = \int_{-\infty}^{\infty} f_{\mathrm{Lorentz}}(x\,;w)\,\mathrm{e}^{2\pi\mathrm{i}kx}\,\mathrm{d}x = \exp\left(-2\pi\,w\,|\,k\,|\right) \tag{1.5.1.A.1.2}$$

と表されます。二つのローレンツ型函数$f_{\mathrm{Lorentz}}(x\,;w_1)$ と $f_{\mathrm{Lorentz}}(x\,;w_2)$ の畳込

$$f_{\mathrm{Lorentz}}(x\,;w_1) * f_{\mathrm{Lorentz}}(x\,;w_1) = \int_{-\infty}^{\infty} f_{\mathrm{Lorentz}}(x-y\,;w)\,f_{\mathrm{Lorentz}}(y\,;w)\,\mathrm{d}y \tag{1.5.1.A.1.3}$$

のフーリエ変換は，畳込定理から

$$\mathfrak{F}_{\mathrm{Lorentz}}(k\,;w_1)\,\mathfrak{F}_{\mathrm{Lorentz}}(k\,;w_2) = \exp\left[-2\pi\,(w_1+w_2)\,|\,k\,|\right] \tag{1.5.1.A.1.4}$$

と表され，二つのローレンツ型函数$f_{\mathrm{Lorentz}}(x\,;w_1)$ と $f_{\mathrm{Lorentz}}(x\,;w_2)$ の畳込は，その逆フーリエ変換として

$$\begin{aligned}
f_{\mathrm{Lorentz}}(x\,;w_1) * f_{\mathrm{Lorentz}}(x\,;w_1) &= \int_{-\infty}^{\infty} \mathfrak{F}_{\mathrm{Lorentz}}(k\,;w_1)\,\mathfrak{F}_{\mathrm{Lorentz}}(k\,;w_2)\,\mathrm{e}^{-2\pi\mathrm{i}kx}\,\mathrm{d}k \\
&= \int_{-\infty}^{\infty} \exp\left[-2\pi\,(w_1+w_2)\,|\,k\,|\right]\mathrm{e}^{-2\pi\mathrm{i}kx}\,\mathrm{d}k \\
&= \frac{1}{\pi\left(w_1+w_2\right)}\left[1+\frac{x^2}{\left(w_1+w_2\right)^2}\right]^{-1}
\end{aligned} \tag{1.5.1.A.1.5}$$

のように導かれます。（↩）

（補足 1.5.1.A.2）

一般的に

$$h(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\,\mathrm{d}y \tag{1.5.1.A.1}$$

で表される積分を解くために，

$$\begin{aligned}
&\xi = \frac{F(x)-F(0)}{x}G\left(\frac{x[F(x)-F(x-y)]}{F(x)-F(0)}\right) \quad\Leftrightarrow\quad \frac{x\xi}{F(x)-F(0)} = G\left(\frac{x[F(x)-F(x-y)]}{F(x)-F(0)}\right) \\
&\Leftrightarrow\quad G^{-1}\left(\frac{x\xi}{F(x)-F(0)}\right) = \frac{x[F(x)-F(x-y)]}{F(x)-F(0)} \\
&\Leftrightarrow\quad \frac{F(x)-F(0)}{x}G^{-1}\left(\frac{x\xi}{F(x)-F(0)}\right) = F(x)-F(x-y) \\
&\Leftrightarrow\quad F(x-y) = F(x) - \frac{F(x)-F(0)}{x}G^{-1}\left(\frac{x\xi}{F(x)-F(0)}\right) \\
&\Leftrightarrow\quad x-y = F^{-1}\left(F(x) - \frac{F(x)-F(0)}{x}G^{-1}\left(\frac{x\xi}{F(x)-F(0)}\right)\right) \\
&\Leftrightarrow\quad y = x - F^{-1}\left(F(x) - \frac{F(x)-F(0)}{x}G^{-1}\left(\frac{x\xi}{F(x)-F(0)}\right)\right) \\
&\mathrm{d}\xi = f(x-y)\,g\left(\frac{x[F(x)-F(x-y)]}{F(x)-F(0)}\right) \\
&y \;:\; -\infty \;\to\; \infty \\
&\xi \;:\; \xi_{\mathrm{min}} \;\to\; \xi_{\mathrm{max}}
\end{aligned}$$

$$
\xi_{\min} = \frac{F(x) - F(0)}{x} G\left(\frac{x[F(x) - F(\infty)]}{F(x) - F(0)}\right)
$$

$$
\xi_{\max} = \frac{F(x) - F(0)}{x} G\left(\frac{x[F(x) - F(-\infty)]}{F(x) - F(0)}\right)
$$

のように積分変数を y から ξ に変換した置換積分を施します。

$$
h(x) = \int_{\xi_{\min}}^{\xi_{\max}} \frac{g(y)}{g\left(y'\right)} \, \mathrm{d}\xi
$$

$$
y = x - F^{-1}\left(F(x) - \frac{F(x) - F(0)}{x} y'\right)
$$

$$
y' = G^{-1}\left(\frac{x\xi}{F(x) - F(0)}\right)
$$

として畳込積分の数値計算をすれば，比較的効率良く高精度の値が得られることが知られています。

$$
y' = \frac{x\left[F(x) - F(x - y)\right]}{F(x) - F(0)}
$$

の関係から，$y = 0$ のときに $y' = 0$，$y = x$ のときに $y' = x$ となり，被積分函数 $\dfrac{g(y)}{g(y')}$ が 1 に近い滑らかな函数になるので，比較的少ない計算回数で高い精度の計算結果が得られます。

中点法を用いる場合には

$$
h(x) \approx \frac{\xi_{\max} - \xi_{\min}}{n} \sum_{i=0}^{n-1} \frac{g(y_i)}{g(y_i')} \tag{1.5.1.A.2}
$$

$$
y_i = x - F^{-1}\left(F(x) - F_x y_i'\right) \tag{1.5.1.A.3}
$$

$$
y_i' = G^{-1}\left(\frac{\xi_i}{F_x}\right) \tag{1.5.1.A.4}
$$

$$
\xi_i = \xi_{\min} + \frac{i + 0.5}{n}(\xi_{\max} - \xi_{\min}) \tag{1.5.1.A.5}
$$

$$
\xi_{\min} = F_x \, G\left(\frac{F(x) - F(\infty)}{F_x}\right) \tag{1.5.1.A.6}
$$

$$
\xi_{\max} = F_x \, G\left(\frac{F(x) - F(-\infty)}{F_x}\right) \tag{1.5.1.A.7}
$$

$$
F_x = \begin{cases} f(0) & \left[x = 0\right] \\ \dfrac{F(x) - F(0)}{x} & \left[x \neq 0\right] \end{cases} \tag{1.5.1.A.8}
$$

とします。

ガウス・ルジャンドル法 (Gauss-Legendre quadrature) を用いる場合には

$$
h(x) \approx \frac{\xi_{\max} - \xi_{\min}}{2} \sum_{i=0}^{n-1} \frac{w_i \, g(y_i)}{g(y_i')} \tag{1.5.1.A.9}
$$

$$
y_i = x - F^{-1}\left(F(x) - F_x y_i'\right) \tag{1.5.1.A.10}
$$

$$
y_i' = G^{-1}\left(\frac{\xi_i}{F_x}\right) \tag{1.5.1.A.11}
$$

$$\xi_i = \frac{1-x_i}{2}\xi_{\min} + \frac{1+x_i}{2}\xi_{\max} \tag{1.5.1.A.12}$$

$$\xi_{\min} = F_x\,G\left(\frac{F(x)-F(\infty)}{F_x}\right) \tag{1.5.1.A.13}$$

$$\xi_{\max} = F_x\,G\left(\frac{F(x)-F(-\infty)}{F_x}\right) \tag{1.5.1.A.14}$$

$$F_x = \begin{cases} f(0) & \left[x=0\right] \\ \dfrac{F(x)-F(0)}{x} & \left[x \neq 0\right] \end{cases} \tag{1.5.1.A.15}$$

とします。ただし $\{x_i\}$ は n のルジャンドル多項式 $P_n(x)$ のゼロ点（ノード）の位置であり，$[-1,1]$ の範囲の値をとります。$\{w_i\}$ は対応する重みで

$$w_i = \frac{2}{(1-x_i^2)[P_n'(x_i)]^2} = \frac{2(1-x_i^2)}{(n+1)^2[P_{n+1}(x_i)]^2}$$

とします。$w_1 + \cdots + w_n = 2$ の関係があります。

（補足 1.5.1.B）畳込積分の数値的な解法，修正版（↩）

函数 $f(x), g(x)$ がいずれも原点 $x=0$ にピークを持つ左右対称な確率密度函数であり，累積分布函数

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \tag{1.5.1.B.1}$$

$$G(x) = \int_{-\infty}^{x} g(t)\,\mathrm{d}t \tag{1.5.1.B.2}$$

も，その逆函数 $F^{-1}(y)$，$G^{-1}(y)$ の計算の仕方も知られているとします。

$$h(x) = \int_{-\infty}^{\infty} f(x-y)\,g(y)\,\mathrm{d}y \tag{1.5.1.B.3}$$

のような数式で表される「畳込」(convolution) を数値積分で計算するための積分変数変換（$y \to \xi$）として，

$$\xi = \frac{F(x)-F(0)}{x}\,G\left(\frac{x\left[F(x)-F(x-y)\right]}{F(x)-F(0)}\right)$$

$$\Leftrightarrow \quad y = x - F^{-1}\left(F(x) - \frac{F(x)-F(0)}{x}G^{-1}\left(\frac{x\xi}{F(x)-F(0)}\right)\right)$$

とする考え方があります。(Ida, 1998)。この変数変換を用いれば

$$\mathrm{d}\xi = f(x-y)\,g\left(\frac{x\left[F(x)-F(x-y)\right]}{F(x)-F(0)}\right)\mathrm{d}y$$

となります。

$$\begin{array}{ccccc} y & : & -\infty & \to & \infty \\ \xi & : & \xi_{\min} & \to & \xi_{\max} \end{array}$$

$$\xi_{\min} = \frac{F(x)-F(0)}{x}\,G\left(\frac{x\left[F(x)-F(\infty)\right]}{F(x)-F(0)}\right)$$

$$\xi_{\max} = \frac{F(x) - F(0)}{x} G\left(\frac{x\left[F(x) - F(-\infty)\right]}{F(x) - F(0)}\right)$$

の関係から，

$$P(\Delta\mu < 0) = \int_0^1 \left[1 - F_{\text{Student}}\left(\frac{t - \overline{A} + \overline{B}}{\hat{s}_A}; \nu_A\right)\right] \mathrm{d}\xi \tag{1.5.1.B.6}$$

$$t = \hat{s}_B F^{-1}_{\text{Student}}(\xi; \nu) \tag{1.5.1.B.7}$$

とします。式 (1.5.1.B.6), (1.5.1.B.7) の形式であれば，例えば中点法 (mid-point method) と呼ばれる数値積分の手法を用いて，

$$P(\Delta\mu < 0) \approx \frac{1}{N} \sum_{i=0}^{N-1} \left[1 - F_{\text{Student}}\left(\frac{t_i - \overline{A} + \overline{B}}{\hat{s}_A}; \nu_A\right)\right] \tag{1.5.1.B.8}$$

$$t_i = \hat{s}_B F^{-1}_{\text{Student}}(\xi_i; \nu) \tag{1.5.1.B.9}$$

$$\xi_i = \frac{i + 0.5}{N} \tag{1.5.1.B.10}$$

のように「コンピュータを使って近似計算ができる」形式を導けます。

次に，単純な事例で 事例で式 (1.5.1.B.8)–(1.5.1.B.10) で表される計算手法がどのように振る舞うかを調べます。。

$A = \{0{,}1\}$，$B = \{1, 2\}$ とします。標本平均は $\overline{A} = 1/2$，$\overline{B} = 3/2$ であり，標本標準偏差は $\hat{\sigma}_A = \sqrt{[(1/2)^2 + (1/2)^2]/(2-1)} = 1/\sqrt{2}$，$\hat{\sigma}_B = 1/\sqrt{2}$，標本平均の標準偏差（標準誤差）は $\hat{s}_A = \hat{\sigma}_A/\sqrt{2} = 1/2$，$\hat{s}_B = 1/2$ となります。このとき母平均の推定値 μ_A と μ_B の統計分布を表す確率密度函数，$f_A(\mu_A)$ と $f_B(\mu_B)$ は

$$f_A(\mu_A) = \frac{1}{\hat{s}_A} f_{\text{Student}}\left(\frac{\mu_A - \overline{A}}{\hat{s}_A}; 1\right) \tag{1.5.1.B.11}$$

$$f_B(\mu_B) = \frac{1}{\hat{s}_B} f_{\text{Student}}\left(\frac{\mu_B - \overline{B}}{\hat{s}_B}; 1\right) \tag{1.5.1.B.12}$$

$$f_{\text{Student}}(x; 1) = \frac{1}{\pi}\left(1 + x^2\right)^{-1} \tag{1.5.1.B.13}$$

と表され，母平均の差 $\Delta\mu \equiv \mu_A - \mu_B$ の統計分布の確率密度函数 $f_{A-B}(\Delta\mu)$ は

$$\begin{aligned} f_{A-B}(\Delta\mu) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\Delta\mu - \mu_A + \mu_B) f_A(\mu_A) f_B(\mu_B)\,\mathrm{d}\mu_A\,\mathrm{d}\mu_B = \int_{-\infty}^{\infty} f_A(\Delta\mu + \mu_B) f_B(\mu_B)\,\mathrm{d}\mu_B \\ &= \frac{1}{\pi^2 \hat{s}_A \hat{s}_B} \int_{-\infty}^{\infty} \left[1 + \frac{\left(\Delta\mu + \mu_B - \overline{A}\right)^2}{\hat{s}_A^2}\right]^{-1} \left[1 + \frac{\left(\mu_B - \overline{B}\right)^2}{\hat{s}_B^2}\right]^{-1} \mathrm{d}\mu_B \\ &= \frac{1}{\pi\left(\hat{s}_A + \hat{s}_B\right)} \left[1 + \frac{\left(\Delta\mu - \overline{A} + \overline{B}\right)^2}{\left(\hat{s}_A + \hat{s}_B\right)^2}\right]^{-1} \end{aligned} \tag{1.5.1.B.14}$$

と表されます（補足 1.5.1.A.1）。

式 (1.5.1.B.14) で表される函数は，尺度化パラメータ $\hat{s}_A + \hat{s}_B$により尺度化された「自由度 1 のステューデント t 分布の確率密度函数」とも言えます。ここでの具体的な数値 $\overline{A} = 1/2$，$\overline{B} = 3/2$，$\hat{s}_A = 1/2$，$\hat{s}_B = 1/2$ をあてはめれば，

$$f_{A-B}(\Delta\mu) = \frac{1}{\pi}\left[1 + \left(\Delta\mu + 1\right)^2\right]^{-1}$$

と表され，累積度数分布函数は

$$F_{A-B}(\Delta\mu) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\Delta\mu + 1\right)$$

となります。$\Delta\mu < 0$ となる確率は $P(\Delta\mu < 0) = F_{A-B}(0) = 1/2 + \arctan(1)/\pi = 1/2 + 1/4 = 3/4 = 0.75$ です。

一方で，式 (1.5.1.B.8)–(1.5.1.B.10) の形式を使う場合に，以下のような Python コードを実行すれば計算結果が得られます。 (↩)

（補足 1.5.2.A）二標本ステューデント t テスト（↩）

Python 言語で利用できる SciPy ライブラリの `stats` モジュールで提供される `stats.ttest_ind(a, b [, axis, …])` を用いれば，デフォルト（省略時）設定では「母分散（母標準偏差）の等しい二つの独立な (independent) 標本の平均の大小」についての t テスト（t 検定）を行うことができます。なお，“`stats`” は「統計学」(statistics)，“`ttest`” は「 t テスト」(t -test)，“`ind`” は「独立」（非従属）(independent) の意味です。

ここでは，単純な例として，$A = \{0, 1\}$，$B = \{1, 2\}$ として `stats.ttest_ind(a, b)` の動作について調べることにします。A と B の標本平均 (sample mean) は

$$\overline{A} = \frac{0+1}{2} = 0.5, \tag{1.5.2.A.1}$$

$$\overline{B} = \frac{1+2}{2} = 1.5 \tag{1.5.2.A.2}$$

となります。A と B の標本分散は

$$\hat{\sigma}_A^2 = \frac{(0-0.5)^2 + (1-0.5)^2}{2-1} = 0.5, \tag{1.5.2.A.3}$$

$$\hat{\sigma}_B^2 = \frac{(1-1.5)^2 + (2-1.5)^2}{2-1} = 0.5, \tag{1.5.2.A.4}$$

となります。式 (1.5.2.1) で表されるように $\overline{A} - \overline{B}$ の標本分散を

$$\hat{\sigma}^2 = \frac{\left(n_A - 1\right)\hat{\sigma}_A^2 + \left(n_B - 1\right)\hat{\sigma}_B^2}{n_A + n_B - 2} = \frac{(2-1)\times 0.5 + (2-1)\times 0.5}{2+2-2} = 0.5 \tag{1.5.2.A.5}$$

とします。$\overline{A} - \overline{B}$ の標本標準偏差は $\hat{\sigma} = \sqrt{0.5}$ です。式 (1.5.2.2) で表されるように，検定統計量 t は

$$t = \frac{\overline{A} - \overline{B}}{\hat{\sigma}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{0.5 - 1.5}{\sqrt{0.5}\times\sqrt{\frac{1}{2} + \frac{1}{2}}} = -\sqrt{2} = -1.414\cdots \tag{1.5.2.A.6}$$

となります。自由度は

$$\nu = n_A + n_B - 2 = 2 + 2 - 2 = 2 \tag{1.5.2.A.7}$$

です。式 (1.1.J.3) に示したように，自由度 $\nu = 2$ の Student t 分布の確率密度函数は，中間ローレンツ型函数 (intermediate Lorentzian function) とも呼ばれ，

$$f_{\chi^2}(t;2) = \frac{1}{2\sqrt{2}}\left(1+\frac{t^2}{2}\right)^{-3/2} \tag{1.1.J.3}$$

と表されます。Figure 1.5.2.A.1 に自由度 2 の Student t 分布の確率密度函数とテスト統計量 $t = -\sqrt{2} = -1.414\cdots$ の値との関係を示します。

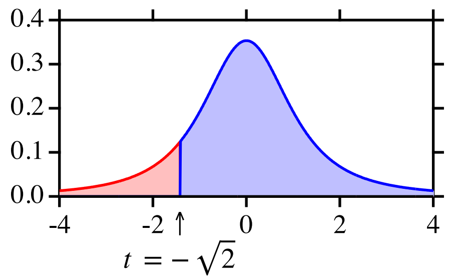

Figure 1.5.2.A.1　自由度 $\nu = 2$ の Student t 分布と検定統計量 $t = -\sqrt{2}$ の位置。ピンク色部分の面積が下側 P 値，薄青色部分の面積が上側 P 値に対応する。

正則不完全ベータ函数 $I(a,b;x)$ を使って，P 値（下側 P 値）は

$$p = 1 - F_{\text{Student}}(t;\nu) = \frac{1}{2}\,I\left(\frac{\nu}{2},\frac{1}{2};\frac{\nu}{t^2+\nu}\right) = \frac{1}{2}\,I\left(\frac{2}{2},\frac{1}{2};\frac{2}{\left(-\sqrt{2}\right)^2+2}\right) = 0.146\cdots \tag{1.5.2.A.8}$$

となります。この計算にも Python と SciPy を使うことができます。

［P 値の計算］

```
>>> import numpy as np
>>> import scipy.special as sp
>>> 0.5*sp.betainc(2/2,1/2,2/((-np.sqrt(2))**2+2))
0.14644660940672624
```

次に二標本 Student t テスト（t 検定）用のメソッド，`stats.ttest_ind(a,b)` の動作について確認します。

［二標本 Student t テスト］

```
>>> import scipy.stats as st
>>> st.ttest_ind([0,1],[1,2])
Ttest_indResult(statistic=-1.414213562373095, pvalue=0.29289321881345254)
```

検定統計量 `statistic` の値は予想通りの値になり，P 値 `pvalue` の値は「下側 P 値として計算した値」のちょうど 2 倍になっているので，`stats.ttest_ind(a,b)` の出力の `pvalue` は「両側 P 値」であることが確認できます。検定統計量 (`statistic`) t が正しく負の値として示されていることにも注目すべきでしょう。「t 分布表を使う t 検定」では「絶対値をとって…」などとしなければいけないので，混乱しやすくなるのですが，コンピュータを使うならそのような心配もする必要がありません。

この結果をどのように解釈したら良いでしょうか？ただし，A の母平均を $\langle A\rangle$，B の母平均を $\langle B\rangle$ とします。正解は「$\langle A\rangle$ が $\langle B\rangle$ より大きい確率は $0.2929/2 \approx 0.146 = 14.6\,\%$」「$\langle A\rangle$ が $\langle B\rangle$ より小さい確率は $1-0.146 = 0.854 = 85.4\,\%$」であり，表現のしかたは，そのどちらでもかまいません。ただし，それ以外の解釈はありえません。「P 値は確率に相当する値だとしても，確率と同一視はできない」という立場をとろうとするなら，少し表現を変えても良いとは思いますが，基本的な解釈は変わりません。

ここでの作業は，「`stats.ttest_ind(a, b)` と言うメソッド」では「第一引数の標本平均から第二引数の標本平均を引く形式で検定統計量が定義されている（だから検定統計量が負の値になった）」あるいは「検定統計量 `statistic` の値が負なので， `pvalue` としては下側 P 値の２倍の値が出力されている」ことを「確認した」意味があります。

「何も考えずにマニュアルどおりにすれば良い」と思われるかもしれませんが，マニュアルには誤記載があるかもしれませんし，ライブラリにはバグがあるかもしれません。それより頻度の多いのは「マニュアルには正しく記載されているのに，それを読む人が誤って解釈してしまう」ことです。SciPy ライブラリは，マニュアルも含めてかなり信頼性は高いのですが，それでもテスト用のデータを使って，最低限のテストをしてから使うようにするべきでしょう。

一方で，特にコンピュータを使うことに関しては，「マニュアルを読む」「考える」ことに無駄な時間を使うより「ためしてみる」ことが有効です。社会実験でも科学実験でも，試行回数を多くするためには必然的に「資金・労力・時間」が必要となります。「コンピュータを使う作業」は「ほぼ無料」「ロボットに作業させるのと同じで人的労力が不要」「機械ロボットと違い，待ち時間は気にならないレベル」という大きな違いがあります。

これ以降の作業では，データを直接見なくても，正解らしい答を導くことができます。

$A = \{0, 1\}$ と $B = \{1, 2\}$ を逆にあてはめ試してみます。

［二標本 Student t テスト（両側検定）］

```
>>> import scipy.stats as st
>>> st.ttest_ind([1,2],[0,1])
Ttest_indResult(statistic=1.414213562373095, pvalue=0.29289321881345254)
```

ここでは「検定統計量が $\overline{B}-\overline{A}$ と同じ符号」で「両側検定」であることが既にわかっているので，「`pvalue` の値の半分が $\overline{B}-\overline{A}$ と同符号の検定統計量に関する上側 P 値になる」ことは，何の予備知識がなくてもわかります。

この結果の解釈は「$\langle B\rangle$ が $\langle A\rangle$ より小さい確率は $0.2929/2 \approx 0.146 = 14.6\,\%$」「$\langle B\rangle$ が $\langle A\rangle$ より大きい確率は $1-0.146 = 0.854 = 85.4\,\%$」のどちらでもかまいません。ただし，それ以外の解釈はありえません。

つぎに「明示的な下側（片側） t テストをしてみます。オプション引数として `alternative='less'` を付けます。

［二標本 Student t テスト（下側検定）］

```
>>> import scipy.stats as st
>>> st.ttest_ind([0,1],[1,2],alternative='less')
Ttest_indResult(statistic=-1.414213562373095, pvalue=0.14644660940672627)
```

「検定統計量が $\overline{A}-\overline{B}$ と同じ符号」で「下側 P 値」が得られているはずです。

この結果の解釈は「$\langle A \rangle$ が $\langle B \rangle$ より大きい確率は 14.6 %」「$\langle A \rangle$ が $\langle B \rangle$ より小さい確率は $1 - 0.146 = 0.854 = 85.4$ %」のどちらでもかまいません。ただし，それ以外の解釈はありえません。

つぎに「明示的な上側（片側） t テストをしてみます。オプション引数として `alternative='greater'` を付けるだけです。

［二標本 Student t テスト（上側検定）］

```
>>> import scipy.stats as st
>>> st.ttest_ind([0,1],[1,2],alternative="greater")
Ttest_indResult(statistic=-1.414213562373095, pvalue=0.8535533905932737)
```

この結果の解釈は「$\langle A \rangle$ が $\langle B \rangle$ より小さい確率は 85.4 %」「$\langle A \rangle$ が $\langle B \rangle$ より大きい確率は $1 - 0.854 = 0.146 = 14.6$ %」のどちらでもかまいません。ただし，それ以外の解釈はありえません。

$\overline{A}$ と $\overline{B}$ の大小関係がすぐにわからなければ，一度 Table 1.5.2.A.2 のような表を作って，その後はその表を参照しても良いでしょう。この大小関係は一標本ステューデント t テストでもウェルチの t テストでも変わりません。

また「両側 t 検定」自体は実質的に無意味なことだとしても，意味を理解して正しく使うのであれば「デフォルトの両側指定」だけを使うのでも「まったく問題ない」こともわかります。

Table 1.5.2.A.2　`scipy.stats.ttest_ind(A, B)` メソッドの両側・上側・下側指定と出力（統計量 `statistic`（スタティスティク）と P 値 `pvalue`（ピーヴァリュー））第一・第二引数母平均推定値の大小の関係

両側／上側／下側指定	統計量 statistic の符号	pvalue の意味	P 値の意味
両側（デフォルト）	正	上側 P 値の 2 倍	$\langle A \rangle < \langle B \rangle$ の確率の 2 倍
両側（デフォルト）	負	下側 P 値の 2 倍	$\langle A \rangle > \langle B \rangle$ の確率の 2 倍
上側		上側 P 値	$\langle A \rangle < \langle B \rangle$ の確率
下側		下側 P 値	$\langle A \rangle > \langle B \rangle$ の確率

「**コンピュータ**を使う **t テスト**」では，「**検定統計量**をどのように定義するか」「**対立仮説**／**帰無仮説**をどのように設定するか」「**両側**（りょうがわ）を使うか**上側**（うえがわ）を使うか**下側**（したがわ）を使うか」など，そのすべてが「どうでも良いこと」と言えます。**検定統計量**がどのように定義されるかは，一回使えばわかります。**対立仮説**／**帰無仮説**は立てる必要がありません。**両側**（りょうがわ）・**上側**（うえがわ）・**下側**（したがわ）のどれを使っても**正解**がわかります。

英語の “Student’s t-test” の “test” の語の意味合いは，日本語の「テスト」あるいは「試（ため）し」に近いので，コンピュータを使うテストとしても，あまり違和感がありません。一方で日本語の「検定」という語には「合否を決める」「棄却か採択かを決める」と言う意味合いが含まれてしまうので，これを「t 検定」と呼ぶのは，むしろ違和感のある言葉の組み合わせになります。「t 分布表を使う t 検定」と「コンピュータを使う t テスト」のように言葉を使い分けても良いかもしれません。（↩）

（補足 1.5.2.B）「二標本ステューデント t テスト」の考え方の過誤（↩）

最も単純な事例で「二標本ステューデント t テスト」の問題点を確認します。

$A = \{0,1\}$，$B = \{1,2\}$ とします。標本平均は $\overline{A} = 1/2$，$\overline{B} = 3/2$ であり，標本標準偏差は $\hat{\sigma}_A = \sqrt{[(1/2)^2 + (1/2)^2]/(2-1)} = 1/\sqrt{2}$，$\hat{\sigma}_B = 1/\sqrt{2}$，標本平均の標準偏差（標準誤差）は $\hat{s}_A = \hat{\sigma}_A/\sqrt{2} = 1/2$，$\hat{s}_B = 1/2$ となります。このとき母平均の推定値 μ_A と μ_B の統計分布を表す確率密度函数，$f_A(\mu_A)$ と $f_B(\mu_B)$ は

$$f_A(\mu_A) = \frac{1}{\hat{s}_A} f_{\text{Student}}\left(\frac{\mu_A - \overline{A}}{\hat{s}_A}; 1\right)$$

$$f_B(\mu_B) = \frac{1}{\hat{s}_B} f_{\text{Student}}\left(\frac{\mu_B - \overline{B}}{\hat{s}_B}; 1\right)$$

$$f_{\text{Student}}(x; 1) = \frac{1}{\pi}\left(1 + x^2\right)^{-1}$$

と表され，母平均の差 $\Delta\mu \equiv \mu_A - \mu_B$ の統計分布の確率密度函数 $f_{A-B}(\Delta\mu)$ は

$$f_{A-B}(\Delta\mu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\Delta\mu - \mu_A + \mu_B) f_A(\mu_A) f_B(\mu_B)\,\mathrm{d}\mu_A\,\mathrm{d}\mu_B = \int_{-\infty}^{\infty} f_A(\Delta\mu + \mu_B) f_B(\mu_B)\,\mathrm{d}\mu_B$$

$$= \frac{1}{\pi^2 \hat{s}_A \hat{s}_B}\int_{-\infty}^{\infty}\left[1 + \frac{\left(\Delta\mu + \mu_B - \overline{A}\right)^2}{\hat{s}_A^2}\right]^{-1}\left[1 + \frac{\left(\mu_B - \overline{B}\right)^2}{\hat{s}_B^2}\right]^{-1}\mathrm{d}\mu_B$$

$$= \frac{1}{\pi\left(\hat{s}_A + \hat{s}_B\right)}\left[1 + \frac{\left(\Delta\mu - \overline{A} + \overline{B}\right)^2}{\left(\hat{s}_A + \hat{s}_B\right)^2}\right]^{-1}$$

と表されます。この函数は「自由度 1 のステューデント t 分布の確率密度函数」を尺度化パラメータ $\hat{s}_A + \hat{s}_B$ により尺度化した函数です。ここでの具体的な数値 $\overline{A} = 1/2$，$\overline{B} = 3/2$，$\hat{s}_A = 1/2$，$\hat{s}_B = 1/2$ をあてはめれば，

$$f_{A-B}(\Delta\mu) = \frac{1}{\pi}\left[1 + \left(\Delta\mu + 1\right)^2\right]^{-1}$$

と表され，累積度数分布函数は

$$F_{A-B}(\Delta\mu) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\Delta\mu + 1\right)$$

となります。$\Delta\mu < 0$ となる確率は $F_{A-B}(0) = 1/2 + \arctan(1)/\pi = 1/2 + 1/4 = 3/4 = 0.75$ です。

1-5-2 節に示した「二標本ステューデント t 検定」では，「自由度 $\nu = n_A + n_B - 2 = 2$ のステューデント t 分布に従う」と仮定する時点で「間違っている」ことがすぐにわかりますが，「$\mu_A < \mu_B$ である確率が 0.854 である」とすることになるので，確率の見積もりも大きく違っています。

1-5-2 節に示した「二標本ステューデント t 検定」は，考え方に根本的な誤りが含まれています。

一方で，1-5-3 節に示した「ウェルチの t 検定」に，$A = \{0,1\}$，$B = \{1,2\}$ の例での数値，$\overline{A} = 1/2$，$\overline{B} = 3/2$，$\hat{\sigma}_A = 1/\sqrt{2}$，$\hat{\sigma}_B = 1/\sqrt{2}$ を当て嵌めれば，尺度化パラメータは

$$\hat{s} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{1}{4} + \frac{1}{4}} = \frac{1}{\sqrt{2}}$$

と見積もられ，自由度は

$$\nu = \frac{\left(\hat{\sigma}_A^2/n_A + \hat{\sigma}_B^2/n_B\right)^2}{\frac{\left(\hat{\sigma}_A^2/n_A\right)^2}{n_A - 1} + \frac{\left(\hat{\sigma}_B^2/n_B\right)^2}{n_B - 1}} = \frac{(1/4 + 1/4)^2}{\frac{(1/4)^2}{2-1} + \frac{(1/4)^2}{2-1}} = \frac{1/4}{1/16} = 4$$

となります。（↩）

（補足 1.5.3.A）ウェルチの t テスト（↩）

Python と SciPy を使って，独立二標本に対するウェルチの t テストをするためには，例えば以下のようにします。

［二組の乱数データの作成，Python インタープリタ（インタラクティブ・ユーザ・インターフェス）での指定］

```
>>> import numpy as np
>>> from scipy import stats
>>> rvs1 = stats.norm.rvs(loc=1.0, scale=1.0, size=50)
>>> rvs2 = stats.norm.rvs(loc=1.1, scale=2.0, size=80)
```

［ウェルチの t 検定］

```
>>> stats.ttest_ind(rvs1, rvs2, equal_var=False)
Ttest_indResult(statistic=-0.10928750412011494, pvalue=0.9131503987868419)
```

この例では両側 P 値 (p-value) がかなり大きな数値になりました。本当は設定した母平均が $\mu_1 = 1.0$ と $\mu_2 = 1.1$ なのに，$\mu_2 < \mu_1$ とする帰無仮説の成立する確率が 41 % 程度もあり，この帰無仮説を棄却することは無理であるという「残念な結果」になったことがわかります。（↩）

（補足 2.A）カイ自乗分布（↩）

標準正規分布 (standard normal distribution) の確率密度関数 $f_{\mathrm{SND}}(x)$ は

$$f_{\mathrm{SND}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{2.A.1}$$

と表されます。このとき，$x^2 = y$ とすれば，y の確率分布の密度関数は

$$\begin{aligned} f_1(y) &= \int_{-\infty}^{\infty} \delta\left(y - x^2\right) f_{\mathrm{SND}}(x)\,\mathrm{d}x = 2\int_0^{\infty} \delta\left(y - x^2\right) f_{\mathrm{SND}}(x)\,\mathrm{d}x \\ &= 2\int_0^{\infty} \delta\left(y - x^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \delta\left(y - x^2\right) \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \end{aligned}$$

$$\underset{\substack{\uparrow \\ y - x^2 \equiv -z \\ \Leftrightarrow x = \sqrt{y+z} \\ \mathrm{d}x = \frac{\mathrm{d}z}{2\sqrt{y+z}}}}{=} \sqrt{\frac{2}{\pi}} \int_y^\infty \delta(-z) \exp\left(-\frac{y+z}{2}\right) \frac{\mathrm{d}z}{2\sqrt{y+z}} = \frac{1}{\sqrt{2\pi}} \int_y^\infty \frac{\delta(-z)}{\sqrt{y+z}} \exp\left(-\frac{y+z}{2}\right) \mathrm{d}x$$

$$\underset{\uparrow}{=} \left\{ \begin{array}{cc} \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right) & [y > 0] \\ 0 & [\text{otherwise}] \end{array} \right\}$$

$$\int_a^b \delta(t) f(t) \mathrm{d}t = \left\{ \begin{array}{cc} f(0) & [a < 0 < b] \\ -f(0) & [b < 0 < b] \\ 0 & [\text{otherwise}] \end{array} \right\}$$

$$= \begin{cases} \frac{1}{2\sqrt{\pi}} \left(\frac{y}{2}\right)^{-1/2} \exp\left(-\frac{y}{2}\right) & [y > 0] \\ 0 & [\text{otherwise}] \end{cases} \tag{2.A.2}$$

と表されます。式 (2.A.2) で表される函数 $f_1(x)$ が「自由度 1 のカイ自乗分布の確率密度函数」と呼ばれます。(↩)

参考文献

Du Prel, J-B., Hommel, G., Röhrig, B., and Blettner, M. (**2009**). "Confidence interval or P-value?," *Dtsch. Arztebl. Int.*, **106**, 335–339. [doi: 10.3238/Arztebl.2009.0335]

Flechner, L. and Tseng, T. Y. (2011). "Understanding results: P-values, confidence intervals, and number need to treat," Indian J. Urol., 27, 532–535. [doi: 10.4103/0970-1591.91447]

Gupta, S. K. (**1998**). "Peak decomposition using Pearson Type VII function," *J. Appl. Crystallogr.*, **31**, 474–476. [doi: 10.1107/S0021889897011047]

Hall, Jr., M. M. (**1977**). "The approximation of symmetric x-ray peaks by Pearson Type VII distribution," *J. Appl. Crystallogr.*, **10**, 66–68. [doi: 10.1107/S0021889877012849]

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D., Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D., Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D., Satrajit Roychoudhury, Ph.D., Kenneth Koury,Ph.D., Ping Li, Ph.D., Warren V. Kalina, Ph.D., David Cooper, Ph.D., Robert W. Frenck, Jr., M.D., Laura L. Hammitt, M.D., Özlem Türeci, M.D., Haylene Nell, M.D., Axel Schaefer, M.D., Serhat Ünal, M.D., Dina B. Tresnan, D.V.M., Ph.D., Susan Mather, M.D., Philip R. Dormitzer, M.D., Ph.D., Uğur Şahin, M.D., Kathrin U. Jansen, Ph.D., and William C. Gruber (**2020**). "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine," *N. Engl. J. Med.*, **383**, 2603–2615. [doi: 10.1056/NEJMoa2034577]

Student (**1908**). "The Probable Error of a Mean," *Biometrika*, **6**, 1–25. (↩)

Young, R. A. and Wiles, D. B. (**1982**). "Profile shape functions in Rietveld refinements," *J. Appl. Crystallogr.*, **28**, 430–438. [doi: 10.1107/S002188988201231X]

青木繁伸 (**1982**).「医学・保健学の例題による統計学」（豊川裕之・柳井晴夫編著）第 7 章，現代数学社

[ISBN-10: 4768701116; ISBN-13: 978-4768701119]