

本日の課題。代表的な確率分布曲線を作る。(20230518)

本日作る、確率分布曲線は、ポアソン分布、カイ二乗分布 (χ^2 分布)、t分布、F分布の4つです。それぞれの分布が、何を意味しているのかをしっかりと知っておくために、作ってみるのですが、途中の計算は、解析学的な微積分の知識が必要になります。また、ガンマー関数 (Γ 関数) ベータ関数 (β 関数) という、あまり聞きなれない超越関数 (有限個の項数の四則演算では数値を計算できない (代数的には $\lim_{n \rightarrow \infty}$ という極限記号を使わないと表現できない内容の関数、例えば e や π など) が出てきますが、あまり気にしないでそういう超越関数なのだと思っておいてください。その確率分布曲線の意味を理解するために、何をしているのかがわかれば十分です。

まず、ポアソン分布を作ります。今日の大きな目的は t 分布と F 分布つくることですから、その本流からは離れるのですが、まれに起こる事象 (0 が多い事象) の回数の統計は二項分布では取り扱えないので、これを先にやります。次に、 χ^2 分布をやります。これは、 χ^2 二乗検定が、正規性や等分散性がない場合にも使える、頑健な検定なので、 χ^2 分布を作るという実用的な目的もあるのですが、それ以上に、 χ^2 分布が期待値とデータの距離の確率分布なので、t 分布 (2つの分布の平均値の差の期待値を 0 として差の観測値が得られる可能性の分布) や F 分布 (2つの分布の分散の比の期待値を 1 としてその期待値から観測値が得られる可能性の分布) なので、 χ^2 分布を作らないと t 分布も F 分布も作れないからです。

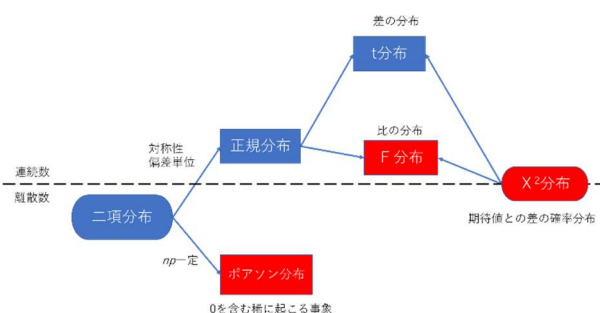


図1. 確率分布の関係性

図1に、今日つくる確率分布の関係性を示しました。二項分布は回数 (回数) のような離散数の確率分布で完璧ではありませんが、 n が大きくなると対称性があります (対称性のあるものを青、ないものを赤で示しました)。二項分布を無理数を含めた連続数に拡張したものが正規分布で、その際、平均値を中心に、偏差を

横の単位として書き表します。この導出は、先週やりました。分散が同じ二つの正規分布の平均値の差の確率分布をカイ二乗分布を使って作ったのは t 分布。同じ二つの正規分布の分散の比の観測値の確率分布をカイ二乗分布を使って作ったのが F 分布です。

ポアソン分布

ポアソン分布とは、稀におこる現象がある条件の中で、 k 回起こる確率を表した離散確率

分布曲線です。つまり、あらかじめ何かの情報で、ある条件でそれが起こる確率を期待値として知っていて、それが実際にその条件で何回（ k ）起きるか、 k の確率を表す曲線です。稀に起こるといのが一つのポイントで、ポアソン分布は一種の近似計算だから、頻繁の起こる現象には使えないのです。頻繁に起こる現象にポアソン分布を使わなければならない理由もありません。期待値を知っているならば、何も統計解析する必要はないだろうという意見もありそうです。しかし、世の中はそういうものではなくて、例えば、稀にしか起こらない地震が、今後一種間の内に、3回以上起こる可能性はどのくらいあるかとか、今後10年間で3人以上の人が、熊に食べられる可能性はどのくらいあるかとか、いろいろ、使える場面はいろいろあります。ポアソン分布という名前は、それを考えた数学者、シメオン・ドニ・ポアソンの名前に由来します。私は、しばらく、ポアソンをフランス語の魚だと思っていて、確かに魚はめったに釣れないなと思っていました。しかし、めったに釣れないのは私だけで、上手な人は、結構連れますから、ポアソン分布は魚に由来していません。ポアソン分布の例でとして、紹介されるのは、多くの場合、ある一定時間に起こる現象に関するものが多く、実際、ポアソン分布の応用例として最初の報告されたのは、軍隊で、一定期間中に馬に蹴られて死ぬ人の数の予測だったそうです（聞きかじりの知識で本当かどうか知りません。）。しかし、時間的に起こる確率でなくても、ポアソン分布的なものはあります。私が知っているのは、プランクトンの計数です。今は、様々なタイプのパーティクル・カウンターがありますから、一定量の液体の中の粒子の数を数えることは簡単です。形態の識別をしてくれる機械もあります。昔は、顕微鏡下で観察して、形態を識別して、数取器でカウントしました。計数板というものを使います。

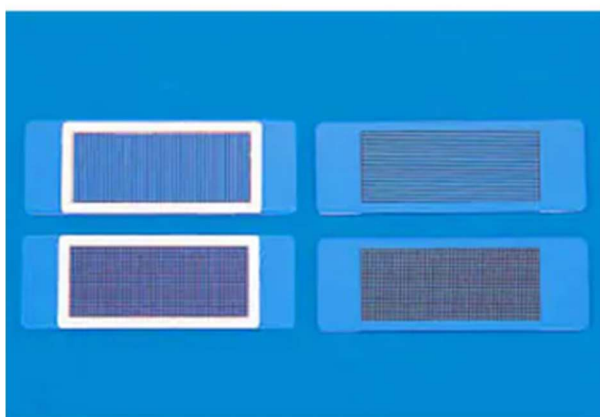


図2 界線スライドグラス

様々な係数板がありますが、図2に示したものは、モノタローで売っている界線スライドグラスです。1mmメッシュで深さ1mmです。これだと、1区画の体積は、 0.001mm^3 です。プランクトンの大きさや種類にもよりますが、出現数の少ないプランクトンの種類だと、全く出現しない区画がたくさんできます。かといって、1ml以上調べようとするると、1000区画もカウントしなければなりません。ポアソン分布にしたがうと考えて、最頻値を期待値として、それを1000倍して1ml当たりの出現量の推定値とした方が良さそうです。なお、界線スライドグラ

スの値段はバカ高いので、自作した方が良いでしょう。ただし、もっと細かい、細胞数のカウントなどだと、さすがに。血球など細胞の計数盤などを買わざるを得ないでしょう。

では、ポアソン分布を作ります。要は、二項分布の変形です。二項分布は

$$W(k) = \binom{n}{k} p^k q^{(n-k)}$$

ですが、ポアソン分布らしく、 $W(k)$ ではなくて、 $P(k)$ と書きます。

$$P(k) = \binom{n}{k} p^k q^{(n-k)}$$

ある事象が k 回起きる確率です。

$$P(k) = \frac{n!}{(n-k)!k!} p^k q^{(n-k)} \quad (p + q = 1)$$

$$P(k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)}$$

$k!$ には p も n も含まれないので、これを外に取り出して、次の3つの部分に分ける

$$P(k) = \frac{1}{k!} AB$$

$$A = \frac{n!}{(n-k)!} p^k$$

$$B = (1-p)^{(n-k)}$$

AとBについて、 $n \rightarrow \infty$ の値を考える。

$$A = \frac{n!}{(n-k)!} p^k$$

$$= n(n-1) \cdots (n-k+1) p^k$$

ここで無理やり、 $(np)^k$ を作ります

$$= \frac{n(n-1) \cdots (n-k+1)}{n^k} n^k p^k$$

$$= \frac{n}{n} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (np)^k$$

$$= \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (np)^k$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0, \quad \lim_{n \rightarrow \infty} \frac{k-1}{n} = 0$$

だから

$$\lim_{n \rightarrow \infty} A = \frac{n!}{(n-k)!} p^k = (np)^k$$

二項分布では平均値（期待値 μ ）は、 $\mu = np$ だから

$$\lim_{n \rightarrow \infty} A = \mu^k$$

B について、

$$B = (1 - p)^{(n-k)}$$

$\mu = np$ を使って、 p を消して n 、 k の式に変形。

$$\begin{aligned} B &= \left(1 - \frac{\mu}{n}\right)^{(n-k)} \\ &= \left(1 - \frac{\mu}{n}\right)^{\left(\frac{\mu^n}{\mu} - k\right)} \\ &= \frac{\left(1 - \frac{\mu}{n}\right)^{\frac{\mu^n}{\mu}}}{\left(1 - \frac{\mu}{n}\right)^k} \end{aligned}$$

$$\text{分母} \left(1 - \frac{\mu}{n}\right)^k \text{ については、} \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right) = 1$$

B 分子については、ネイピア数 e が t 次のような超越関数であることを知っていれば、

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

μ を一つの単位数と考えて、 $-\frac{\mu}{n}$ を $\frac{1}{n}$ と書き換えて、

$$\left(1 - \frac{\mu}{n}\right)^{\frac{\mu^n}{\mu}} = e^\mu$$

となります。解法はネイピア数を知っていれば簡単です。ただ、これを思いつくのは、ポアソン分布が指数関数になっていることを知っているからです。私もそうでした（無理に指数関数を作ろうとした。）、何の事前情報もなしにこの解放を思いつく人は少ないでしょう（数学に慣れている人を除く）。とにかく、必要ならば、ブログの統計学 III-3-2.（ネイピア数）の説明を読めばよいので、知っていることにして

$$\lim_{n \rightarrow \infty} B = \frac{\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^{\frac{\mu^n}{\mu}}}{\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^k} = \frac{e^\mu}{1} = e^\mu$$

3 つを合わせて、

$$\lim_{n \rightarrow \infty} P(k) = \frac{1}{k!} \lim_{n \rightarrow \infty} A \lim_{n \rightarrow \infty} B = \frac{1}{k!} \mu^k e^\mu$$

となります。期待値 λ の事象が、ある制限内で起こるということを無限回繰り返した時に、描かれる確率分布はこのように表せます。この計算では、二項分布の平均値の期待値として μ という記号を使ってきましたが、ポアソン分布の場合、その期待値は、何らかの情報によって外から与えられる（例えば 10 年間で熊に喰われたか人の数）ので、 μ を λ に書き換えて

$$P(k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

と書きます。ネイピア数の計算の仕方はいくつかあるのですが、 $e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$ を知っていれば、後はちょっとしたテクニックを使った式の書き換えだけで導出できます。

ポアソン分布では、平均値に依存する形で分散が変化します。2項分布だって分散が平均値によって変わります。ポアソン分布は np が一定で、 p が小さくなれば n が大きくなる。 n が大きくなれば分布が平均値を中心に尖ってくるのだから、当たり前だろうと言われればそれまでです。しかし、完全なポアソン分布では、 k の分散の値は平均の値と同じになります。この性質は、データの分散がポアソンのかそうでないかを考えるときに重要です。その証明をします。二乗の期待値から期待値の二乗を引くと分散になるという、2項分布で使った式から出発します。

$$V(k) = E(k^2) - (E(k))^2$$

ポアソン分布は二項分布の極限ですから、この式は成り立ちます。

ポアソン分布ではデータが k である確率は、

$$P(k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

ですから、 k^2 の期待値は

$$E_{k^2} = \sum_{k=0}^{\infty} k^2 P(k) = \sum_{k=0}^{\infty} k^2 \frac{1}{k!} \lambda^k e^{-\lambda}$$

ここから、1段階ずつ k を減らした無限総和の式を作っていきます。

$$\begin{aligned} E_{k^2} &= \sum_{k=0}^{\infty} k^2 \frac{1}{k!} \lambda^k e^{-\lambda} \\ &= \sum_{k=0}^{\infty} k \frac{1}{(k-1)!} \lambda^k e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{(k-1)+1}{(k-1)!} \lambda^k e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{k-1}{(k-1)!} \lambda^k e^{-\lambda} + \sum_{k=0}^{\infty} \frac{1}{(k-1)!} \lambda^k e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{k-1}{(k-1)!} \lambda^k e^{-\lambda} + \lambda \sum_{k=0}^{\infty} \frac{1}{(k-1)!} \lambda^{k-1} e^{-\lambda} \end{aligned}$$

第2項は $k = k - 1$ とした時の確率の総和だから1。

$$E_{k^2} = \sum_{k=0}^{\infty} \frac{k-1}{(k-1)!} \lambda^k e^{-\lambda} + \lambda$$

第1項について同じ手法を繰り返して、

$$\begin{aligned} & \sum_{k=0}^{\infty} \frac{k-1}{(k-1)!} \lambda^k e^{-\lambda} \\ &= \sum_{k=0}^{\infty} \frac{1}{(k-2)!} \lambda^k e^{-\lambda} \\ &= \lambda^2 \sum_{k=0}^{\infty} \frac{1}{(k-2)!} \lambda^{k-2} e^{-\lambda} \end{aligned}$$

シグマの中は $k = k - 1$ の時の牡蛎率の総和だから 1、したがって

$$E_{k^2} = \lambda^2 + \lambda$$

そもそも

$$E_k = \lambda$$

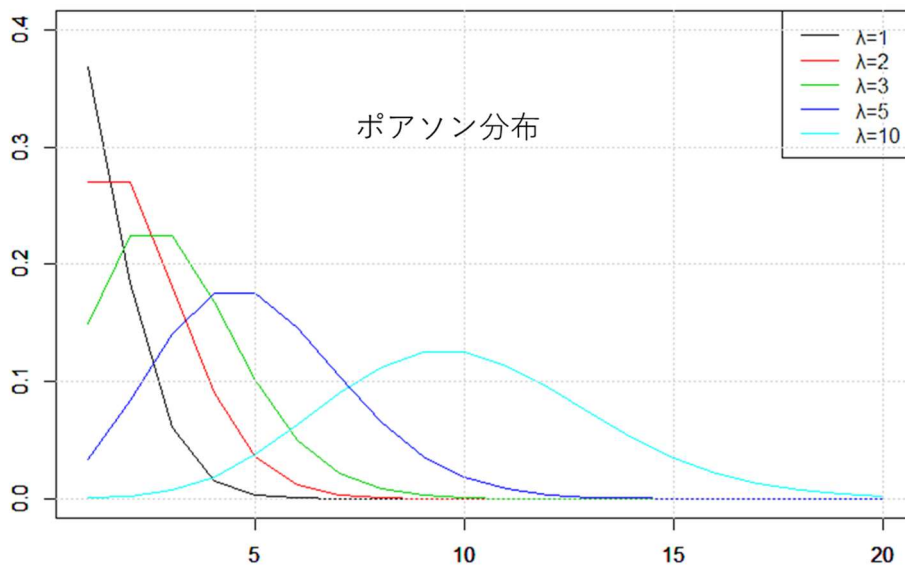
だから

$$(E_k)^2 = \lambda^2$$

したがって

$$\begin{aligned} V_{(k)} &= E_{(k^2)} - (E_{(k)})^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

となります。ポアソン分布では、分散は期待値と同じになります。これを分散が平均値と等しくなると覚えてもまあ良いのかなと思います。



カイ二乗分布
正規分布で

$$N_{(x)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\frac{x-\mu}{\sigma}$ というのは、標準化で、 x と μ の距離を偏差を基準とした値に変換するということです。こうして標準化された正規分布

$$N_{(0,1)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}}$$

を標準正規分布と言います。カイ二乗とは、正規分布の指数部分 $\left(\frac{x-\mu}{\sigma}\right)^2$ のことで、 χ^2 と書きます。 χ の二乗ではなくて、 χ^2 、これ一つで、一つの変数を表します。何故、エックスではなくてカイを使うのかはよく知りません。多分、エックスに由来するが、それから導かれたカイ二乗という値ということだと思います。

$P_{(x)}$ が正規分布にしたがうとき、 $P_{(\chi^2)}$ がしたがう確率分布曲線を作るというのが、これからの作業です。カイ二乗値は自由度によって違います。 χ^2_1 、 χ^2_2 のように下付きの添え字で、自由度を表します。

まず、自由度1のカイ二乗の確率分布です。これは簡単です。そのまま正規分布の式を書き換えれば良いのですから。

$$P_{(x)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\Phi_{(x)} = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

ですが、 $x^2 = (-x)^2$ ですから、反対側にも同じものがある、0を中心にこれを折り返して、重ね合わせます。

$$\Phi_{(x)} = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

二乗なので、横方向に引き伸ばさなければなりません。確率はこの累積確率の微分値だから、 $x^2 = \chi^2$ と変換して、これを微分します。

$$\Phi_{(x)} = 2 \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

この定積分で、 $\int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \left[\int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right]_0^x$ の0を入れたものは定数部分したらなくなるから微分の結果はもとにもどって

$$P_{(x)} = 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$P(\chi^2_1) = \frac{d\Phi(x)}{d\chi^2}$$

$\chi^2 = x^2$ の変換を行うので

$$\frac{d\chi^2}{dx} = 2x = 2\sqrt{\chi^2}$$

$$d\chi^2 = 2\sqrt{\chi^2}dx$$

$$P(\chi^2_1) = \frac{d\Phi(x)}{d\chi^2} = \frac{d\Phi(x)}{2dx} = P(x) \frac{1}{2\sqrt{\chi^2}} = 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{\chi^2}{2}} \frac{1}{2\sqrt{\chi^2}} = \frac{1}{\sqrt{2\pi\chi^2_1}} e^{-\frac{\chi^2_1}{2}}$$

自由度1であることをしっかり明記して

$$P(\chi^2_1) = \frac{1}{\sqrt{2\pi\chi^2_1}} e^{-\frac{\chi^2_1}{2}} = \frac{1}{\sqrt{2\pi}} \chi^2_1^{-\frac{1}{2}} e^{-\frac{\chi^2_1}{2}}$$

となります。

自由度2の場合を考えます。ある現象 x とある現象 y があって、どちらも正規分布に従う場合のことです。 $x \sim N(0,1)$ and $y \sim N(0,1)$. 互いに独立なのだから、それが同時の起こる確率は掛け算になります。

$$P(x,y) = P(x)P(y)$$

混乱しないように一つは $\chi^2_{1;x}$ 、 $\chi^2_{1;y}$ と書くことにします。下付きの添え字の1は自由度1の意味です。求める確率分布の変数は $\chi^2_{2;x+y}$ となります。これを使って

$$P(\chi^2_{2;x+y})$$

の公式を作ります。

$$\int_0^\infty P(\chi^2_{2;x+y}) = 2 \int_0^\infty \int_0^\infty P(\chi^2_{1;x}) P(\chi^2_{1;y}) d\chi^2_{1;x} d\chi^2_{1;y} = 1$$

として、積分係数を決めればよいわけですが、このまま積分するのは難しそうです。二回の積分の、積分方向が $\chi^2_{1;x} + \chi^2_{1;y}$ と並行する線の方と、それに直交する線にすれば、重積分できそうです。 $\chi^2_{1;x}$ と $\chi^2_{1;y}$ を $\frac{\pi}{4}$ 回転させて、足し合わせれば良さそうです。 $O_1(x,y)$, という直交座標系を極座標変換して回転し、得られた $P_1(r, \theta_1)$ という極座標径系の式を、座標変換して変換して回転し、 $O_2(z,w)$ という直交座標に変換するという操作をします。

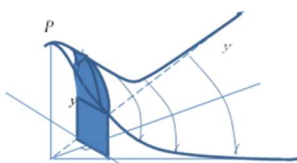


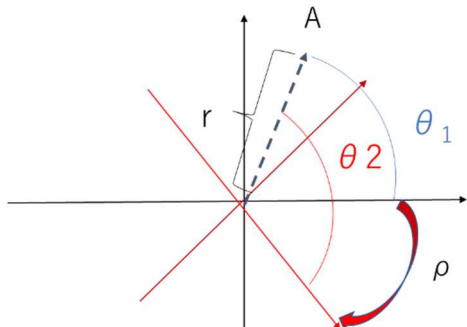
図4. 重積分の方向

直交座標系 $O_1(x,y)$ の点 $A(x,y)$ を極座標系で表すと $O_1(x,y) \rightarrow P_1(r, \theta_1)$

$$x = r \cos \theta_1$$

$$y = r \sin \theta_1$$

直交座標系 $O_2(z,w)$ の点 $A(z,w)$ を極座標系で表すと



$$O_2(z, w) \rightarrow P_2(r, \theta_2)$$

$$z = r \cos \theta_2$$

$$w = r \sin \theta_2$$

$$\theta_2 = \rho + \theta_1$$

という関係になります。

図5. 直交座標系 $O_1(x,y)$ で記述された点Aを回転した直交座標系 $O_2(z,w)$ で記述する作業

この関係を利用して、 ρ と θ_1 で w

$$z = r \cos(\rho + \theta_1) = r(\cos \rho \cos \theta_1 - \sin \rho \sin \theta_1) = x \cos \rho - y \sin \rho$$

$$w = r \sin(\rho + \theta_1) = r(\sin \rho \cos \theta_1 + \cos \rho \sin \theta_1) = x \sin \rho + y \cos \rho$$

$$\rho = \frac{\pi}{4}$$

$$\sin \frac{\pi}{4} = \frac{1}{\sqrt{2}}, \quad \cos \frac{\pi}{4} = \frac{1}{\sqrt{2}}$$

ですから、

$$z = r \cos(\rho + \theta_1) = x \cos \rho - y \sin \rho = \frac{1}{\sqrt{2}}x - \frac{1}{\sqrt{2}}y = \frac{1}{\sqrt{2}}(x - y)$$

$$w = r \sin(\rho + \theta_1) = x \sin \rho + y \cos \rho = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y = \frac{1}{\sqrt{2}}(x + y)$$

となります。z軸は x と y の差の $\sqrt{2}$ 分の1、 w 軸は x と y の和の $\sqrt{2}$ 分の1です。

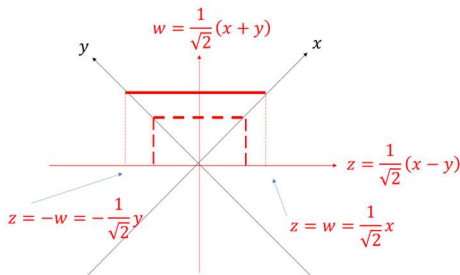


図6. 重積分の俯瞰図

図6はこの重積分を上から俯瞰してみた図です。第一段階で、z軸に平行しw軸と直交する平面で積分（赤い破線や太線で表現した平面上の面積を求める。積分範囲は $-w \rightarrow w$ ）し、それをw軸上で0から無限大まで積分します（重ね合わせる）。

このような方針で、 $\int_0^\infty P(\chi^2_{2;x+y})$ を解いていきます。

$$\int_0^\infty P(\chi^2_{2;x+y}) d\chi^2_{2;x+y} = 2 \int_0^\infty \int_0^\infty P(\chi^2_{1;x}) P(\chi^2_{1;y}) d\chi^2_{1;x} d\chi^2_{1;y}$$

ですから、

まず、不定積分 $\iint P(\chi^2_{1:x}) P(\chi^2_{1:y}) d\chi^2_{1:x} d\chi^2_{1:y}$ についてです。

$$P(\chi^2_{1:x}) = \frac{1}{\sqrt{2\pi\chi^2_{1:x}}} e^{-\frac{\chi^2_{1:x}}{2}}$$

$$P(\chi^2_{1:y}) = \frac{1}{\sqrt{2\pi\chi^2_{1:y}}} e^{-\frac{\chi^2_{1:y}}{2}}$$

$$\iint P(\chi^2_{1:x}) P(\chi^2_{1:y}) d\chi^2_{1:x} d\chi^2_{1:y} = \iint \frac{1}{\sqrt{2\pi\chi^2_{1:x}}} e^{-\frac{\chi^2_{1:x}}{2}} \frac{1}{\sqrt{2\pi\chi^2_{1:y}}} e^{-\frac{\chi^2_{1:y}}{2}} d\chi^2_{1:x} d\chi^2_{1:y}$$

$$= \frac{1}{2\pi} \iint \frac{1}{\sqrt{\chi^2_{1:x}\chi^2_{1:y}}} e^{-\frac{(\chi^2_{1:x} + \chi^2_{1:y})}{2}} d\chi^2_{1:x} d\chi^2_{1:y}$$

$$= \frac{1}{2\pi} \iint \frac{1}{\sqrt{\chi^2_{1:x}\chi^2_{1:y}}} e^{-\frac{\chi^2_{2:x+y}}{2}} d\chi^2_{1:x} d\chi^2_{1:y}$$

$$\chi^2_{1:x} = x, \quad \chi^2_{1:y} = y$$

$$w = \frac{1}{\sqrt{2}}(x + y) = \frac{1}{\sqrt{2}}\chi^2_{2:x+y}$$

$$\chi^2_{2:x+y} = \sqrt{2}w$$

$$z = \frac{1}{\sqrt{2}}(x - y)$$

だから、

$$x = \frac{1}{\sqrt{2}}(w + z)$$

$$y = \frac{1}{\sqrt{2}}(w - z)$$

$$xy = \frac{1}{2}(w^2 - z^2)$$

$$\frac{dx}{dw} = \frac{1}{\sqrt{2}}, \quad \frac{dx}{dz} = \frac{1}{\sqrt{2}}, \quad \frac{dy}{dw} = \frac{1}{\sqrt{2}}, \quad \frac{dy}{dz} = -\frac{1}{\sqrt{2}}$$

ですが、必要なのは $dx = \frac{1}{\sqrt{2}}dw$ と $dy = -\frac{1}{\sqrt{2}}dz$ ですが、 y 軸上で z を w に変換しなければなり

ませんが y 軸上では、 $z = -w$ だから、 $dz = -dw$ です。これを頭に入れておいて、これらを使って、

$$\frac{1}{2\pi} \iint \frac{1}{\sqrt{\chi^2_{1:x} \chi^2_{1:y}}} e^{-\frac{(\chi^2_{1:x} + \chi^2_{1:y})}{2}} d\chi^2_{1:x} d\chi^2_{1:y}$$

を変換する。

$$\begin{aligned} &= \frac{1}{2\pi} \iint \frac{1}{\sqrt{xy}} e^{-\frac{(x+y)}{2}} dx dy \\ &= \frac{1}{2\pi} \iint \frac{1}{\sqrt{\frac{1}{2}(w^2 - z^2)}} e^{-\frac{\sqrt{2}w}{2}} \frac{1}{\sqrt{2}} dw \frac{-1}{\sqrt{2}} dz \\ &= -\frac{1}{2\sqrt{2}\pi} \iint \frac{1}{\sqrt{(w^2 - z^2)}} e^{-\frac{w}{\sqrt{2}}} dw dz \end{aligned}$$

積分範囲を与えて、内側の積分を定積分として求めると

$$= -\frac{1}{2\sqrt{2}\pi} \iint_D \frac{1}{\sqrt{(w^2 - z^2)}} e^{-\frac{w}{\sqrt{2}}} dw dz$$

図6に示したように、この積分範囲Dは $-w \leq z \leq w$ で、 $dz = -dw$

$$= -\frac{1}{2\sqrt{2}\pi} \int_{-w}^w \frac{1}{\sqrt{(w^2 - z^2)}} e^{-\frac{w}{\sqrt{2}}} dw dz$$

内側の定積分の積分範囲の $-w$ 、 w は z の値だから、 $\int_{-w}^w \frac{1}{\sqrt{\frac{1}{2}(w^2 - z^2)}} dz$ を先に計算して、 w の関

数とする。

$$\begin{aligned} &\int_{-w}^w \frac{1}{\sqrt{\frac{1}{2}(w^2 - z^2)}} dz \\ &= \int_{-w}^w \frac{1}{\sqrt{(w^2 - z^2)}} dz \end{aligned}$$

$\frac{z}{w} = -u$ と変換する。(y軸上では $z = -w$ だからマイナスを付けておいた方が楽)

$$\frac{du}{dz} = -\frac{1}{w}$$

$$dz = -w du$$

だから

$$\begin{aligned} &= \int_{-w}^w \frac{1}{\sqrt{(w^2 - z^2)}} dz \\ &= \int_{-w}^w \frac{1}{w \sqrt{\left(1 - \left(\frac{z}{w}\right)^2\right)}} dz \end{aligned}$$

$$\begin{aligned}
&= \int_{-1}^1 \frac{1}{w\sqrt{(1-u^2)}}(-w)du \\
&= - \int_{-1}^1 \frac{1}{\sqrt{(1-u^2)}} du
\end{aligned}$$

$u = \sin \theta$ とおく

$$\frac{du}{d\theta} = \cos \theta$$

$$\sqrt{(1-u^2)} = \sqrt{1 - \sin^2 \theta} = \sqrt{\cos^2 \theta} = \cos \theta$$

だから

$$\begin{aligned}
&= - \int_{-1}^1 \frac{1}{\sqrt{(1-u^2)}} du \\
&= - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{\cos \theta} \cos \theta d\theta = - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} du \\
&= - \left(\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right) = -\pi
\end{aligned}$$

結論として、

$$\begin{aligned}
&\iint P(\chi^2_{1:x}) P(\chi^2_{1:y}) d\chi^2_{1:x} d\chi^2_{1:y} \\
&= -\frac{1}{2\sqrt{2}\pi} \iint_{-w}^w \frac{1}{\sqrt{(w^2 - z^2)}} e^{\frac{-w}{\sqrt{2}}} dw dz \\
&= -\frac{1}{2\sqrt{2}\pi} \int -\pi e^{\frac{-w}{\sqrt{2}}} dw \\
&= \frac{1}{2\sqrt{2}} \int e^{\frac{-w}{\sqrt{2}}} dw
\end{aligned}$$

となるが

$$w = \frac{1}{\sqrt{2}} \chi^2_{2:x+y}$$

を使って、カイ二乗の式にすると、

$$\begin{aligned}
&\frac{dw}{d\chi^2_{2:x+y}} = \frac{1}{\sqrt{2}} \\
&= \frac{1}{2\sqrt{2}} \int e^{\frac{-w}{\sqrt{2}}} dw = \frac{1}{2\sqrt{2}} \int e^{\frac{-\frac{1}{\sqrt{2}}\chi^2_{2:x+y}}{\sqrt{2}}} \frac{1}{\sqrt{2}} d\chi^2_{2:x+y} \\
&= \frac{1}{4} \int e^{\frac{-\chi^2_{2:x+y}}{2}} d\chi^2_{2:x+y}
\end{aligned}$$

$\frac{\chi^2_{2:x+y}}{2} = t$ とすると

$$\frac{d\chi^2_{2:x+y}}{2} = dt$$

$$\frac{1}{4} \int e^{-\frac{\chi^2_{2:x+y}}{2}} d\chi^2_{2:x+y} = \frac{1}{4} \int e^{-t} 2dt = \frac{1}{2} \int e^{-t} dt$$

これを $t = 0$ から無限大で定積分すると、

$$\frac{1}{2} \int_0^\infty e^{-t} dt = \frac{1}{2} ((-e^{-\infty}) - (-e^0)) = \frac{1}{2} (0 + 1) = \frac{1}{2}$$

となります。

$$\begin{aligned} \int_0^\infty P(\chi^2_{2:x+y}) &= 2 \int_0^\infty \int_0^\infty P(\chi^2_{1:x}) P(\chi^2_{1:y}) d\chi^2_{1:x} d\chi^2_{1:y} \\ &= 2 \times \frac{1}{2} = 1 \end{aligned}$$

で確率の総和が1になることが確認できます。ということで、

$$\int P(\chi^2_{2:x+y}) d\chi^2_{2:x+y} = \frac{1}{2} \int e^{-\frac{\chi^2_{2:x+y}}{2}} d\chi^2_{2:x+y} =$$

$\chi^2_{2:x+y}$ までの累積確率は

$$\int_0^{\chi^2_{2:x+y}} P(\chi^2_{2:x+y}) d\chi^2_{2:x+y} = \frac{1}{2} \int_0^{\chi^2_{2:x+y}} e^{-\frac{\chi^2_{2:x+y}}{2}} d\chi^2_{2:x+y}$$

これを微分すれば、積分記号の中だけになって

$$P(\chi^2_{2:x+y}) = \frac{1}{2} e^{-\frac{\chi^2_{2:x+y}}{2}}$$

これで、自由度2の時の χ^2 分布曲線を作ることが出来ます。このやり方で、自由度dの時の χ^2 分布曲線を作ることが出来ますが、あまりに手間がかかって面倒です。法則を見つけて、一般化した公式を作るべきでしょう。

今のところ、わかっているのは自由度1の時と、2の時です。

$\varphi = 1$ の時 (φ は自由度)

$$P(\chi^2_1) = \frac{1}{\sqrt{2\pi\chi^2_1}} e^{-\frac{\chi^2_1}{2}} = \frac{1}{\sqrt{2\pi}} \chi^2_1^{-\frac{1}{2}} e^{-\frac{\chi^2_1}{2}}$$

$\varphi = 2$ の時

$$P(\chi^2_2) = \frac{1}{2} e^{-\frac{\chi^2_2}{2}} = \frac{1}{2} \chi^2_2^{\frac{2}{2}-1} e^{-\frac{\chi^2_2}{2}}$$

$\varphi = 3$ の時、さすがに2例だけでは法則が見つけられないので、 $P(\chi^2_3)$ は今までと同じ方法で作ります。

$$\int_0^\infty P(\chi^2_3) d\chi^2_3 = 2 \int_0^\infty \int_0^\infty P(\chi^2_1) P(\chi^2_2) d\chi^2_1 d\chi^2_2$$

$\iint P(\chi^2_1) P(\chi^2_2) d\chi^2_1 d\chi^2_2$ について

$$P(\chi^2_1) = \frac{1}{\sqrt{2\pi\chi^2_1}} e^{-\frac{\chi^2_1}{2}}$$

$$P(\chi^2_2) = \frac{1}{\sqrt{2\pi\chi^2_2}} e^{-\frac{\chi^2_2}{2}}$$

$$\iint P(\chi^2_1) P(\chi^2_2) d\chi^2_1 d\chi^2_2 = \iint \frac{1}{\sqrt{2\pi\chi^2_1}} e^{-\frac{\chi^2_1}{2}} \frac{1}{\sqrt{2\pi\chi^2_2}} e^{-\frac{\chi^2_2}{2}} d\chi^2_1 d\chi^2_2$$

$$= \frac{1}{2\sqrt{2\pi}} \iint \frac{1}{\sqrt{\chi^2_1}} e^{-\frac{(\chi^2_1 + \chi^2_2)}{2}} d\chi^2_1 d\chi^2_2$$

$$\chi^2_{1:x} = x, \quad \chi^2_{2:y} = y$$

$$w = \frac{1}{\sqrt{2}}(x + y) = \frac{1}{\sqrt{2}}\chi^2_{3:x+y}$$

$$\chi^2_{3:x+y} = \sqrt{2}w$$

$$z = \frac{1}{\sqrt{2}}(x - y)$$

だから、

$$x = \frac{1}{\sqrt{2}}(w + z)$$

$$y = \frac{1}{\sqrt{2}}(w - z)$$

$$\frac{dx}{dw} = \frac{1}{\sqrt{2}}, \quad \frac{dx}{dz} = \frac{1}{\sqrt{2}}, \quad \frac{dy}{dw} = \frac{1}{\sqrt{2}}, \quad \frac{dy}{dz} = -\frac{1}{\sqrt{2}}$$

これらを使って、 w と z で書かれた式にして、最終的に z を消す

$$\iint \frac{1}{\sqrt{\chi^2_1}} e^{-\frac{(\chi^2_1 + \chi^2_2)}{2}} d\chi^2_1 d\chi^2_2$$

$$= \iint \frac{1}{\sqrt{x}} e^{-\frac{(x+y)}{2}} dx dy$$

$x = \frac{1}{\sqrt{2}}(w + z)$, $w = \frac{1}{\sqrt{2}}(x + y)$, $dx = \frac{1}{\sqrt{2}}dw$, $dy = -\frac{1}{\sqrt{2}}dz$ を使って

$$\begin{aligned}
&= -\frac{1}{2} \iint \frac{1}{\sqrt{\frac{1}{\sqrt{2}}(w+z)}} e^{\frac{-w}{\sqrt{2}}} dw dz \\
&= -2^{\frac{1}{4}} \iint \frac{1}{\sqrt{(w+z)}} e^{\frac{-w}{\sqrt{2}}} dw dz
\end{aligned}$$

積分記号の前の定数がうるさいから、後で考えることにして

$$\begin{aligned}
&\iint \frac{1}{\sqrt{(w+z)}} e^{\frac{-w}{\sqrt{2}}} dw dz \\
&= \iint \frac{1}{w^{\frac{1}{2}} \sqrt{\left(1 + \left(\frac{z}{w}\right)\right)}} e^{\frac{-w}{\sqrt{2}}} dw dz
\end{aligned}$$

前例に倣って $\frac{z}{w} = -u$ という変換を入れる。 $dz = -wdu$ だから

$$\begin{aligned}
&= \iint \frac{1}{w^{\frac{1}{2}} \sqrt{(1-u)}} w e^{\frac{-w}{\sqrt{2}}} dw du \\
&= \iint \frac{1}{\sqrt{(1-u)}} w^{\frac{1}{2}} e^{\frac{-w}{\sqrt{2}}} dw du
\end{aligned}$$

w できている部分と u で出来ている部分に分かれたから、積分範囲を与えて、 u で出来ている部分を定積分して、先に計算します。 u は 0 から 1 で変動します。

$$\begin{aligned}
&= \iint \frac{1}{\sqrt{(1-u)}} w^{\frac{1}{2}} e^{\frac{-w}{\sqrt{2}}} dw du \\
&\int \left(\int_0^1 \frac{1}{\sqrt{(1-u)}} du \right) w^{\frac{1}{2}} e^{\frac{-w}{\sqrt{2}}} dw
\end{aligned}$$

括弧の中の積分は、それなりの値になるでしょう。これを定数として外に取り出すと、

$$B \int w^{\frac{1}{2}} e^{\frac{-w}{\sqrt{2}}} dw = B \int \chi^2_3 \frac{1}{2} e^{\frac{-\chi^2_3}{2}} d\chi^2_3$$

という形になります。元に戻って、

$$\int_0^\infty P(\chi^2_3) d\chi^2_3 = A_3 \int_0^\infty \chi^2_3 \left(\frac{3}{2}-1\right) e^{\frac{-\chi^2_3}{2}} d\chi^2_3 = 1$$

として、 A_3 を決めれば良いのですが、ここではそんなこと知る必要もありません。それはそれとして、

$$\int_0^{\chi^2_3} P(\chi^2_3) d\chi^2_3 = A_3 \int_0^{\chi^2_3} \chi^2_3 \frac{1}{2} e^{\frac{-\chi^2_3}{2}} d\chi^2_3 = 1$$

を微分したものが $P(\chi^2_3) d\chi^2_3$ ですから、

$$P(\chi^2_3) = A_3 \chi^2_3 \frac{1}{2} e^{\frac{-\chi^2_3}{2}}$$

です。

$\varphi = 1$ の時 (φ は自由度)

$$P(\chi^2_1) = \frac{1}{\sqrt{2\pi\chi^2_1}} e^{-\frac{\chi^2_1}{2}} = \frac{1}{\sqrt{2\pi}} \chi^2_1^{-\frac{1}{2}} e^{-\frac{\chi^2_1}{2}} = \frac{1}{\sqrt{2\pi}} \chi^2_1^{\left(\frac{\varphi}{2}-1\right)} e^{-\frac{\chi^2_1}{2}}$$

$\varphi = 2$ の時

$$P(\chi^2_2) = \frac{1}{2} e^{-\frac{\chi^2_2}{2}} = \frac{1}{2} \chi^2_2^0 e^{-\frac{\chi^2_2}{2}} = \frac{1}{2} \chi^2_2^{\left(\frac{\varphi}{2}-1\right)} e^{-\frac{\chi^2_2}{2}}$$

$\varphi = 3$ の時

$$P(\chi^2_3) = A_3 \chi^2_3^{\frac{1}{2}} e^{-\frac{\chi^2_3}{2}} = A_3 \chi^2_3^{\left(\frac{\varphi}{2}-1\right)} e^{-\frac{\chi^2_3}{2}}$$

となっているので

$\varphi = n$ の時

$$P(\chi^2_n) = A_n \chi^2_n^{\left(\frac{n}{2}-1\right)} e^{-\frac{\chi^2_n}{2}}$$

となりそうな気がします。しかし、今のところはそんな気がするだけです、 A_n が何だかわかりません。とにかくそう掛けることにします。わかっているのは、

$$A_1 = \frac{1}{\sqrt{2\pi}}, A_2 = \frac{1}{2}$$

です。もし、そのように書けるのならば

$$\int_0^\infty P(\chi^2_n) d\chi^2_n = A_n \int_0^\infty \chi^2_n^{\left(\frac{n}{2}-1\right)} e^{-\frac{\chi^2_n}{2}} d\chi^2_n = 1$$

ですから、

$$A_n = \frac{1}{\int_0^\infty \chi^2_n^{\left(\frac{n}{2}-1\right)} e^{-\frac{\chi^2_n}{2}} d\chi^2_n}$$

この分母は無限積分が含まれていて、超越関数です。これは、ガンマ関数という超越関数を知っていれば、簡単に書き表せます。

ガンマ関数は大数学者、オイラーが考えた超越関数で、次のように定義されています。

$$\Gamma(z) = \int_0^\infty t^{(z-1)} e^{-t} dt$$

$\int_0^\infty \chi^2_n^{\left(\frac{n}{2}-1\right)} e^{-\frac{\chi^2_n}{2}} d\chi^2_n$ をガンマ関数を使って書き換えます。

$$\frac{\chi^2_n}{2} = t, \quad \frac{n}{2} = z$$

と書き換えますから、 $d\chi^2_n = 2dt$

$$\int_0^\infty \chi^2_n \left(\frac{n-1}{2}\right) e^{-\frac{\chi^2_n}{2}} d\chi^2_n = \int_0^\infty (2t)^{(z-1)} e^{-t} 2dt$$

$$= 2^z \int_0^\infty t^{(z-1)} e^{-t} dt$$

$\int_0^\infty t^{(z-1)} e^{-t} dt = \Gamma(z)$ だから

$$\int_0^\infty \chi^2_n \left(\frac{n-1}{2}\right) e^{-\frac{\chi^2_n}{2}} d\chi^2_n = 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) = 2^n \Gamma\left(\frac{n}{2}\right)$$

$A_n = \frac{1}{\int_0^\infty \chi^2_n \left(\frac{n-1}{2}\right) e^{-\frac{\chi^2_n}{2}} d\chi^2_n}$ に戻ると、

$$A_n = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

ちなみに、

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma(1) = 1$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$$

です。

$$A_1 = \frac{1}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} = \frac{1}{\sqrt{2}\sqrt{\pi}}, A_2 = \frac{1}{2^{\frac{2}{2}} \Gamma(1)} = \frac{1}{2^{\frac{2}{2}} \cdot 1} = \frac{1}{2}$$

を確認してください。いずれにしても

$$P(\chi^2_n) = \frac{\chi^2_n \left(\frac{n-1}{2}\right) e^{-\frac{\chi^2_n}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

です。

ここで、使った χ^2 分布の導出の仕方はあまり一般的ではありません。ネットで調べてもこのやり方をしている例はありませんでした。この導出法は私が自分で考えたのですが、あまり洗練された方法ではありませんね。ネットや教科書で使われている導出の仕方は、先に Γ 関数を与えてそこから χ^2 の牡蛎率密度関数を作ったり、高等数学を使って、一気に自由度 n のカイ二乗分布を作っています。その方がわかりやすいという人もいるでしょう。私は高等数学がわからないから、積分方向を工夫して、ゴリゴリと初等の解析数学の手法で導出しました。それが正しかと言われると自信がありません。少し気になることがあります。 A_n を先に定積分して取り出すことが可能だとしたことです。最も気になるのは、 A_n がいつでも外に取り出せる形で存在するのかという問題です。取り出せるのは自明だろうという見方もあるかもしれませんが、積分範囲の決め方も強引かもしれません。気にしすぎかもしれませんが、何処かで確認しようと思います。導出法を解説した理由は、導出法

を覚えるためではありません。導出法を覚えるならば、もっとエレガントな導出法を覚えてください。ネットで探せばあります。わざわざ、積分にこだわったのは、カイ二乗が正規分布を土台に置いた確率分布だということを実感してもらうためです。統計解析を2つに分けて、パラメトリックな統計解析と、ノンパラメトリックな解析法があり χ^2 分析はノンパラメトリックな統計解析に分類している例が多くあります。 χ^2 分析がノンパラメトリックな統計解析に多く使われるのは事実ですが、パラメトリックなデータにも使えます。そもそも、パラメトリックとは何かという問題があります。教科書によれば、パラメトリックなデータとは、特定のデータ分布にしたがうと想定できるデータで、たとえば、t検定では、比較対象される2つのデータ群が、同じ分散を持った正規分布に従うことを前提に、検定が行われます。分析以前に正規性、等分散性が問われます。 χ^2 分析では、比較する群の等分散性は問われていませんが、全体として正規分布していることを前提としています。私は、 χ^2 分析がノンパラメトリックのデータに特化した分析法だとは思っていません。ただ、頑健性が強く、いろいろなものに使えるということだと思います。ここで確認したように χ^2 分布は、正規分布を前提にしています。だからこそ、t分布やF分布も、正規分布とカイ二乗分布から作られるのです。一般に、等分散性や正規性が高ければ、統計検定の検出感度が高くなります。つまり、わずかな違いでも、有意差とされることが多いのです。反対に、いろいろの藻に使える頑健な統計分析法は、検出感度が低いと言えます。しかし、それでも違いがあると結論されれば、その結論の信頼性は高いでしょう。ですから、t検定やF検定で出来るものを、感度の低い統計分析法で検定することは、お勧めできないのです（有意差を見逃す可能性がある。）。

χ^2 分布は分散比の検定だという話。

χ^2 検定では、 χ^2 の値を次の式で計算します。

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}$$

f_i : 観測値、 e_i : 期待値

分子の方はカイ二乗値の定義から十分納得できますが、分母が何故期待値なのかわかりません。これについて考えます。統計分析の実務としては、こんなことを考える必要はありませんが、統計的な確率とは何かを理解する上では、この考察は重要です。しっかり理解してください。

カイ二乗値の分散について考えます。 $V(x) = E(x^2) - E(x)^2$ という分散の計算式を使いたいので、 $E(z^2)$ を求めます。

期待値というのは値と確率の積の総和ですから、 z^k の期待値は

$$E(z^k) = \int_0^{\infty} z^k P(z) dz$$

$P(z)$ が二項分布ならば

$$E(z^k) = \int_0^\infty z^k \frac{z^{\frac{n}{2}-1} e^{-\frac{z}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} dz$$

となります。変形していきます。

$$\begin{aligned} E(z^k) &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty z^k z^{\frac{n}{2}-1} e^{-\frac{z}{2}} dz \\ &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty z^{\frac{n}{2}+k-1} e^{-\frac{z}{2}} dz \end{aligned}$$

$\frac{z}{2} = t$ とする。 $dz = 2dt$

$$\begin{aligned} E(z^k) &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty (2t)^{\frac{n}{2}+k-1} e^{-t} 2dt \\ &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty 2^{\frac{n}{2}+k-1} 2(t)^{\frac{n}{2}+k-1} e^{-t} dt \\ &= \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty 2^{\frac{n}{2}+k} (t)^{\frac{n}{2}+k-1} e^{-t} dt \\ &= \frac{2^{\frac{n}{2}+k}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty (t)^{\frac{n}{2}+k-1} e^{-t} dt \end{aligned}$$

$$\int_0^\infty (t)^{\frac{n}{2}+k-1} e^{-t} dt = \Gamma\left(\frac{n}{2} + k\right)$$

$$E(z^k) = \frac{2^{\frac{n}{2}+k} \Gamma\left(\frac{n}{2} + k\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

$$E(z^0) = \frac{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} = 1$$

$$\begin{aligned} E(z^1) &= \frac{2^{\frac{n}{2}+1} \Gamma\left(\frac{n}{2} + 1\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \\ &= \frac{2 \Gamma\left(\frac{n}{2} + 1\right)}{\Gamma\left(\frac{n}{2}\right)} \end{aligned}$$

ガンマ関数の計算 $\Gamma(a+1) = a\Gamma(a)$ これを使う

$$= \frac{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = n$$

$$E(z^1) = n$$

$$E(z^2) = \frac{2^{\left(\frac{n}{2}+2\right)} \Gamma\left(\frac{n}{2}+2\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

ガンマ関数の計算 $\Gamma(a+1) = a\Gamma(a)$ これを 2 回使う

$$E(z^2) = \frac{2^{\left(\frac{n}{2}+2\right)} \left(\frac{n}{2}+1\right) \Gamma\left(\frac{n}{2}+1\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{2^{\left(\frac{n}{2}+2\right)} \left(\frac{n}{2}+1\right) \left(\frac{n}{2}\right) \Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{2^{\left(\frac{n}{2}+2\right)} \left(\frac{n+2}{2}\right) \left(\frac{n}{2}\right) \Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{2^{\left(\frac{n}{2}\right)} (n+2) (n) \Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} = n(n+2)$$

分散の計算式の戻って

$$\begin{aligned} V_{(z)} &= E_{(z^2)} - E_{(z)}^2 \\ &= n(n+2) - n^2 \\ &= 2n \end{aligned}$$

となります。つまり χ^2 分布では、分散 $V_{(x)} = 2n$ は期待値 $E_{(z)} = n$ の 2 倍なのです。

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}$$

f_i : 観測値、 e_i : 期待値

という式は

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{\frac{1}{2}\sigma^2}$$

f_i : 観測値、 σ^2 : 分散

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{\left(\frac{\sigma}{\sqrt{2}}\right)^2} = \left(\frac{f_i - e_i}{\frac{\sigma}{\sqrt{2}}}\right)^2$$

と考えれば、期待値と観測値との差を偏差の $\sqrt{2}$ 分の 1 で標準化し、その総和の値を χ^2 分布

確率密度曲線に当てはめて、そのような値が得られる確率を論じているのです。つまり、 χ^2 分布も確率分布モデルを使った分散分析なのです。

III-2-6. Student の t 分布

t 分布は student の t 検定に使われる確率分布です。Student の t 検定は、データの差の有意性の検定です。ですから、t 分布とは正規分布すると仮定されるあるデータと、あるデータの平均値の差の分布のことです。この確率分布は正規分布に似ていて、実際自由度が十分大きければ正規分布に近似できます。左右対称というところも正規分布に似ています。正規分布と大きく違うところは、自由度によって分布が変わるところです。この点は χ^2 分布です。というのも、この確率分布が正規分布と χ^2 分布の合成によってできているからです。

これを発見したのはビール会社ギネスの技師だった W.Gosset ですが、Student の t 分布という名前がついているのは、会社から論文投稿を禁じられていたために、彼が Student という筆名でその発見を論文化したからです。おそらく、彼の発見の動機は、今、私たちが抱いている疑問と同じだったろうと思います。つまり、「正規分布することが想定されるデータについて、データから得られた平均値がどのくらい正規分布の期待値（平均値・中央値）に近いのかは、母集団の標準偏差(σ)を尺度にして、標準化して、標準正規分布 $N(0,1)$ の分布の中でデータの平均値がどの位置にあるのかを考えれば良いというのはわかるにしても、そもそも母集団の標準偏差(σ)を知らないのだから、標準化することができません。データから得られるのは標本分散から推定した母分散の推定値だから確率的に変動する。その変動をどう読みこむのか。」という疑問です。

我々の疑問を数式的に表し、それをどのように解決すればよいかを考えます。まず、我々は正規分布というものの存在を認めています。u が標準正規分布 $N(0,1)$ に従うのならば、その確率は

$$N_{(0,1)} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{u^2}{2}}$$

次はuをどう求めるかですが、その作業が標準化と言われる作業で、実際のデータ、期待値、分散（偏差）が使われます。これは習った通り。

$$u = \frac{x - \mu}{\sigma}$$

なのですが、我々が疑問としているのは、「この式で使っているのは正規分布している本来の理想的なデータの期待値=平均値(μ)と分散 (σ^2) なのだから、実際のデータしか知らない私たちが、そんなものは知るわけがないだろう。」ということです。仮に、母集団の平均を知っていることにして、実際に得られた値と真の平均値との差を考えられる偏差を基準値としてその平方和を求めると、

$SS = v$ として、

$$v = \sum_{i=1}^m \left(\frac{y_i - \mu}{\sigma_{se}} \right)^2$$

です。 σ_{se} と添え字を付けたのは、まだ、 y_i が何だかわからないからです。とりあえず、

something estimated だと覚えておいて下さい。そこれをよく見ると、 σ_{se} が外から与えた期待値ならば、第一の式の総和記号の中は χ^2 の定義式で、 v は χ^2 分布します。このこの場合、平均値という形で合計の値が縛られている のだから、 χ^2 の値を-1 個目まで決めた後の最後の値は自動的に決まります。だから全体の自由度は $m-1$ です。つまり v は自由度 $m-1$ の χ^2 分布します。この u と v は互いに独立なのです（関係ない。 u が変化しても v は変化しない。反対に v が変化して u の値に変化はない。という意味です。これが W.Gosset のやった最大の発見なのです。だから、2つの変数を合成した変数を作り、その変数の確率を2つの確率の積として計算できるのです。）。私は、この2つが独立かと問われると、一瞬、言葉に詰まってわからなくなります。でも落ちつて考えると、確かに独立ですね。そこで、2つの関数を関係づけるために、この2つの変数からできる合成関数を考えます。母集団の分散 σ は平方和を $(n-1)$ で割ったものが推定になるというのをやりましたが。この場合は、 μ と σ_{se} を既知のものとして、外から与えているので、自由度は m です。

$$v = \sum_{i=1}^m \left(\frac{x_i - \mu}{\sigma_{se}} \right)^2 = \frac{1}{\sigma_{se}^2} \sum_{i=1}^m (x_i - \mu)^2$$

$$s_{se}^2 = \frac{1}{n} \sum_{i=1}^m (x_i - \mu)^2$$

だから

$$s^2 = \frac{1}{n} v \sigma_{se}^2$$

$$s = \sqrt{\frac{v}{n}} \sigma_{se}$$

$$\frac{s}{\sigma_{se}} = \sqrt{\frac{v}{n}}$$

$u = x - \mu \sigma$ は正規分布しますが、問題は、分母つまり単位となる物差しの長さが、データからもとめた偏差と真の偏差との間で違っているということです。その解決として「実際のデータ から求めた値に、実際の値から得た偏差と真の偏差の比を掛けたものを、合成変数として、その値になる確率を考える。」というのが W.Gosset の提案内容です。 v は χ^2 分布し、 u は正規分布して、互いに独立ですから、2つを合成した関数変数の確率は両者の確率の積です。

具体的には、合成変数 ($t:t$ 値) は次の式です。

$$t = \frac{y - \mu}{\sigma_{se}} \cdot \frac{\sigma}{s}$$

$$t = u \sqrt{\frac{n}{v}}$$

確率変数 u と v が同時にそれぞれ独立にある値をとって、 t の値が決まるので、 u になる確率 $W(u)$ と v になる確率 $P(v)$ の積が t になる確率 $S(t)$ です。

$$S(t) = W(u)P(v)$$

確率の総和は次の重積分で計算できて、その値は1です（確率の総和は1）。

$$\int_{-\infty}^{\infty} \int_0^{\infty} W(u)P(v) dudv = 1$$

これを t と s の式に変換します。その際、 $t = u\sqrt{\frac{n}{v}}$ からヤコビアンをもとめます。ヤコビアンとは、正規分変数の変換によって生ずる。係数のようなものです。ブログに解説を載せておきました。わからなければ読んでください。

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_0^{\infty} W(u)P(v) dudv \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} J(u, v/t, s)W(u)P(v) dt ds \end{aligned}$$

v から s の変換ですが、 s は v そのものでも良さそうだから、 $s = v$ として、

$$\frac{du}{dt} = \frac{\sqrt{v}}{\sqrt{n}}, \quad \frac{dv}{ds} = 0, \quad \frac{dv}{dt} = \frac{-2nu^2}{t^{-3}}, \quad \frac{dv}{ds} = 1$$

$$J(u, v/t, s) = \begin{vmatrix} \frac{\sqrt{v}}{\sqrt{n}} & 0 \\ \frac{-2nu^2}{t^{-3}} & 1 \end{vmatrix} = \frac{\sqrt{v}}{\sqrt{n}}$$

$$W(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{u^2}{2}}$$

$$P(v) = P\left(\chi^2_{\varphi}\right) = \frac{\chi^2_{\varphi}^{\frac{(\varphi-1)}{2}} e^{-\frac{\chi^2_{\varphi}}{2}}}{2^{\frac{\varphi}{2}} \Gamma\left(\frac{\varphi}{2}\right)}$$

なのですが、今、考えているのは、自由度は外から与えた平均値と分散を与えた時の確率分布ですから、自由度は $m-1$ ではなくて m なのです。W.Gosset が、考えた t 分布は、2つの群の平均値の差が、0 という帰無仮説を立てて、実測された二つの群の差から t 値を計算し、帰無仮説から導かれる t 値=0 という期待値に対して、実測された t 値がどのような確率で存在するかという確率分布です。だから、自由度は1です。

$$P(v) = \frac{\frac{-1}{v} e^{-\frac{v^2}{2}}}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)}$$

$\Gamma(\cdot)$ はガンマ関数（ガンマ関数についてはカイ二乗分布を作るときに説明しました。これらを使って、以下の重積分を変形します。

$$\int_{-\infty}^{\infty} \int_0^{\infty} W(u)P(v) dudv$$

$$\int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{u^2}{2}} v^{\frac{-1}{2}} e^{-\frac{v^2}{2}} \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} dudv$$

$$\int_{-\infty}^{\infty} \int_0^{\infty} \frac{\sqrt{v}}{\sqrt{n}\sqrt{2\pi}\sigma} e^{-\frac{u^2}{2}} v^{\frac{-1}{2}} e^{-\frac{v^2}{2}} \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} dsdt$$

$$\int_{-\infty}^{\infty} \int_0^{\infty} \frac{1}{\sqrt{2n\pi}\sigma} e^{-\frac{u^2}{2}} \frac{e^{-\frac{v^2}{2}}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} dsdt$$

胃かかなり長い計算になるので、省略します。興味があれば、ブログを読んでください。少し誤りがあるので、近々、修正しておきます。

最終的に、

$$\int_{-\infty}^{\infty} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})\left(\frac{t^2}{n^2}+1\right)^{\frac{n}{2}+1}} dt = 1$$

となりますが

個々の t の値をとる確率は、 $-\infty$ から t の値までの定積分を微分すれば良いので、積分記号の中の間数になって、

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

です。この複雑な式は、 β 関数を使えば、簡略化して表現できます。 β 関数はガンマ関数の積で以下の式で表せます。

$$\beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$ ですから

$$\beta\left(\frac{n}{2}, \frac{1}{2}\right) = \frac{\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2}+\frac{1}{2})} = \frac{\pi\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2}+\frac{1}{2})}$$

で

$$S(t) = \frac{1}{\sqrt{n\pi}\beta\left(\frac{t^2}{n^2}+1\right)^{\frac{n}{2}+1}}$$

となります。一般に student の t 値の確率密度関数として紹介されているのは、この式です。

話を元に戻します。

$$v = \sum_{i=1}^m \left(\frac{y_i - \mu}{\sigma_{se}} \right)^2$$

のところで、 y_i は何だかわからないとしましたが、今はわかります。 y_i は t 値です。 μ はその期待値で0です。つまり、グループ A とグループ B の平均値の差です。 μ は0なんだから、 $\left(\frac{y_i - \mu}{\sigma_{se}} \right) = \left(\frac{y_A - y_B}{\sigma_{se}} \right)$ です。

A の平均値 y_A が期待値0だとすれば、 y_B は期待値0の周辺で確率的に変動するでしょう。

また、B の平均値 y_B が期待値0だとすれば、 y_A は期待値0の周辺で確率的に変動するでしょう。これは、真の平均値がデータから得られた平均値の周りで変動するときの積率でもあります。データから推定された平均値 m が真の平均値の周りで変動するときの積率を平均誤差と言うということは、すでに説明しました。平均誤差は $\frac{\sigma}{\sqrt{n}}$ です。つまり、2つの群の

平均値から推定した、t 値の分散は $\frac{\sigma^2}{n}$ で偏差は $\frac{\sigma}{\sqrt{n}}$ です。 $\sigma_{se} = \frac{\sigma}{\sqrt{n}}$ つまり、 σ_{se} (Something estimated)は S tandard error だという話です。t 値の定義式を書き換えておきます。

$$t = \frac{y - \mu}{\sigma_{se}} \cdot \frac{\sigma}{s} = \frac{y - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sigma}{s} = \frac{y - \mu}{\frac{s}{\sqrt{n}}}$$

sはサンプル集団の偏差

ポイントを要約すると、

t 分布は、正規分布する分散の等しい二つのデータ群の平均値の差から、t 値を作って、その値がカイ二乗分布するとして、求めた確率分布で、データ群の差を標準化するために、データ群から求めた標準誤差で差を割ったもので差を割るということです。

F 分布

F 分布は分散の比を確率変数とした確率分布です。t 分布は正規分布する確率と χ^2 分布する確率の積の確率分布でしたが、F 分布は χ^2 分布する 2 つの確率の積の分布です。比の分布という考え方は χ^2 分布するところにも出てきました。今までやってきた、確率の組み合わせと同じ作業をします。違っているのは分散比 (F) が新たな確率変数になるということです。具体的には、 χ^2 分布する 2 つの変数があり、それらの分散を、 $\frac{z_1}{n_1}$ 、

$\frac{z_2}{n_2}$ としたとき (n は自由度、データ数ではないので注意)、

$$F = \frac{\frac{z_1}{n_1}}{\frac{z_2}{n_2}}$$

を F 比として、その確率分布を考えるということです。確率変動するものと確率変動する物の比の確率というのは少し悩ましいのですが、両者が独立していて、一方が何らかの値をとった時に他方が独立的に変動するのだから、両者の確率の積で変動すると考えれば良いでしょう。それぞれは、ある期待値からのへだたりだから、 χ^2 分布します。

それぞれの確率を

$$P(z_1) = \frac{\chi^2_n \left(\frac{n_1-1}{2}\right) e^{-\frac{\chi^2}{2} \frac{n_1}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n_1}{2}\right)}$$

$$P(z_2) = \frac{\chi^2_n \left(\frac{n_2-1}{2}\right) e^{-\frac{\chi^2}{2} \frac{n_2}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n_2}{2}\right)}$$

$$P(F) = P(z_1)P(z_2)$$

ということで、F で積分できるように F と直交する平面を考えて、その面積を求めれば良い ということで、今までやっていたことと同じです。f と直交する平面をどう考えるのがポイントです。求積法としての積分の考え方に慣れていれば、こういう考え方が自然にできます。F 比を f という変数うにして

$$f = \frac{\frac{z_1}{n_1}}{\frac{z_2}{n_2}} = \frac{z_1 n_2}{n_1 z_2}$$

$$z_1 n_2 = f z$$

$$z_2 n_1 = z$$

と置きます。これなら、f の定義そのままだし、f と z は互いに独立しています。

$0 \leq z_1 \leq \infty, 0 \leq z_2 \leq \infty, 0 \leq f \leq \infty, 0 \leq z \leq \infty$ なので確率総和 = 1 の式は

$$\int_0^\infty P(f)df = \int_0^\infty \int_0^\infty P(z_1)P(z_2)dz_1dz_2 = 1$$

となります。

$$\iint P(z_1)P(z_2)dz_1dz_2$$

積分変数を変換します。

$$\frac{dz_1}{df} = \frac{z}{n_2}$$

$$\frac{dz_1}{dz} = \frac{f}{n_2}$$

$$\frac{dz_2}{df} = 0$$

$$\frac{dz_2}{dz} = \frac{1}{n_1}$$

だからヤコビアンは

$$J(z_1, z_2/f, z) = \begin{vmatrix} \frac{z}{n_2} & \frac{f}{n_2} \\ 0 & \frac{1}{n_1} \end{vmatrix} = \frac{z}{n_1 n_2}$$

$$\begin{aligned} \iint P(z_1)P(z_2)dz_1dz_2 &= \iint P(f)P(z) \cdot \frac{z}{n_1 n_2} dz_1 dz_2 \\ &= \iint \frac{z_1^{\frac{n_1}{2}-1}}{2^{\frac{n_1}{2}} \Gamma\left(\frac{n_1}{2}\right)} e^{-\frac{z_1}{2}} \frac{z_2^{\frac{n_2}{2}-1}}{2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{z_2}{2}} \cdot \frac{z}{n_1 n_2} dz_1 dz_2 \\ &= \iint \frac{\left(\frac{fz}{n_2}\right)^{\frac{n_1}{2}-1}}{2^{\frac{n_1}{2}} \Gamma\left(\frac{n_1}{2}\right)} e^{-\frac{fz}{2}} \frac{\left(\frac{z}{n_1}\right)^{\frac{n_2}{2}-1}}{2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{z}{2}} \cdot \frac{z}{n_1 n_2} dz df \\ &= \int \left(\int \frac{\left(\frac{fz}{n_2}\right)^{\frac{n_1}{2}-1}}{2^{\frac{n_1}{2}} \Gamma\left(\frac{n_1}{2}\right)} e^{-\frac{fz}{2}} \frac{\left(\frac{z}{n_1}\right)^{\frac{n_2}{2}-1}}{2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{z}{2}} \cdot \frac{z}{n_1 n_2} dz \right) df \\ &= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \Gamma\left(\frac{n_1}{2}\right)} \int \left(\frac{z}{n_2}\right)^{\frac{n_1}{2}-1} e^{-\frac{fz}{2}} \left(\frac{z}{n_1}\right)^{\frac{n_2}{2}-1} e^{-\frac{z}{2}} \cdot \frac{z}{n_1 n_2} dz \right) df \\ &= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \Gamma\left(\frac{n_1}{2}\right)} \int \left(\frac{z}{n_2}\right)^{\frac{n_1}{2}-1} e^{-\frac{fz}{2}} \left(\frac{z}{n_1}\right)^{\frac{n_2}{2}-1} e^{-\frac{z}{2}} \cdot \frac{z}{n_1 n_2} dz \right) df \end{aligned}$$

$$\begin{aligned}
&= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} n_2^{\frac{n_1}{2}} n_2^{\frac{n_1}{2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} \int z^{\frac{n_1}{2}-1} z^{\frac{n_1}{2}-1} e^{-\frac{fz}{2}} e^{-\frac{z}{2}} \cdot zdz \right) df \\
&= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} \int z^{\frac{n_1+n_2}{2}-2} e^{-\frac{(fz+z)}{2}} \cdot zdz \right) df \\
&= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} \int z^{\frac{n_1+n_2}{2}-1} e^{-\frac{(f+1)z}{2}} \cdot dz \right) df
\end{aligned}$$

$w = \left(\frac{f}{n_2} + \frac{1}{n_1}\right)z$ とおきかえると、 $z = \frac{2}{\left(\frac{f}{n_2} + \frac{1}{n_1}\right)}w$ だから

$$\begin{aligned}
&= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} \int \left(\frac{2}{\left(\frac{f}{n_2} + \frac{1}{n_1}\right)}w\right)^{\frac{n_1+n_2}{2}-1} e^{-w} \frac{2}{\left(\frac{f}{n_2} + \frac{1}{n_1}\right)} \cdot dw \right) df \\
&= \int \left(\frac{(f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} \int \left(\frac{2}{\left(\frac{f}{n_2} + \frac{1}{n_1}\right)}w\right)^{\frac{n_1+n_2}{2}} e^{-w} \cdot dw \right) df \\
&= \int \left(\frac{2^{\frac{n_1+n_2}{2}} (f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right)} \int \left(\frac{1}{\left(\frac{f}{n_2} + \frac{1}{n_1}\right)}w\right)^{\frac{n_1+n_2}{2}} e^{-w} \cdot dw \right) df \\
&= \int \left(\frac{2^{\frac{n_1+n_2}{2}} (f)^{\frac{1}{n_1}-1}}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right) \left(\frac{f}{n_2} + \frac{1}{n_1}\right)^{\frac{n_1+n_2}{2}}} \int w^{\frac{n_1+n_2}{2}} e^{-w} \cdot dw \right) df
\end{aligned}$$

となりますが、 w の関数で表された

$$\int w^{\frac{n_1+n_2}{2}} e^{-w} \cdot dw$$

だけを無限積分すると、以下のような、ガンマ関数です、

$$\int_0^{\infty} w^{\frac{n_1+n_2}{2}} e^{-w} \cdot dw = \Gamma\left(\frac{n_1+n_2}{2}\right)$$

ですから、

$$\int P(f)df = \int \left(\frac{2^{\frac{n_1+n_2}{2}} (f)^{\frac{1}{n_1}-1} \Gamma\left(\frac{n_1+n_2}{2}\right)}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right) \left(\frac{f}{n_2} + \frac{1}{n_1}\right)^{\frac{n_1+n_2}{2}}} \right) df$$

となりますが、 $P(f)$ を $f=0$ から $f=f$ を積分したものを、 f で微分したものが、 f の確率密度関数ですから、

$$P(f) = \frac{2^{\frac{n_1+n_2}{2}} (f)^{\frac{1}{n_1}-1} \Gamma\left(\frac{n_1+n_2}{2}\right)}{2^{\frac{n_1}{2}} \left(\frac{1}{n_2}\right)^{\frac{1}{n_1}} \left(\frac{1}{n_1}\right)^{\frac{1}{n_2}} \Gamma\left(\frac{n_1}{2}\right) 2^{\frac{n_2}{2}} \Gamma\left(\frac{n_2}{2}\right) \left(\frac{f}{n_2} + \frac{1}{n_1}\right)^{\frac{n_1+n_2}{2}}}$$

となります。この式を整理すれば、もっと簡単に書き表せますが、そこまでやるのはこの講義の主旨ではないので、結論として、 β 関数を使って書き表したFの確率密度関数を示します。

$$P(f) = \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}}{\beta(n_1, n_2)} \cdot \frac{f^{\frac{1}{n_1}-1}}{(n_1 f + n_2)^{\frac{n_1+n_2}{2}}}$$

こんな面倒なことをしてF関数を導出する必要は全くありません。覚えたとしても、一生使うことはないでしょう。ここでわざわざ計算過程を描いたのは、途中で Γ 関数が出てきて、答えが出てしまうところが面白かったからです。知識として、頭にとどめておかななくてはならないのは、**F比はデータから求めた分散の比で、分母も分子も χ^2 分布するという**ことです。ですから、**F比の確率密度は、分母・分子の確率の積になります。**

また、最後に示した式からわかるように、F分布は、**分母分子の自由度で変わります。**

F比を使った検定は、結構たくさんあります。もっとも、重要なのは、分散はいくつかの部分分散に分けることが出来て、それを使って、多要因分散分析が出来ます。そうした、王王的な展開を考えると、現時点で大切なことは、F分布が何の確率分布なのか、どのように牡蠣率密度関数が作られているのかを理解することです。