

### III.2-5 カイ二乗分布

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

という正規分布の式で $\frac{x-\mu}{\sigma}$ という部分は、標準化で、 $x$ と $\mu$ 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

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

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、 $\chi^2$ と書きます。 $\chi$ の二乗ではなくて、 $\chi^2$ 、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、これからする作業です。カイ二乗値は自由度によって違います。 $\chi^2_1$ 、 $\chi^2_2$ 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って、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とって

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

$$P_{(\chi^2_1)} = \frac{d\Phi_{(x)}}{d\chi^2}$$

$\chi^2 = x^2$ の変換を行うので

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

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。  $O_1(x, y)$ ,

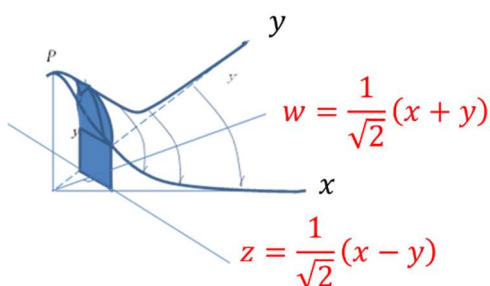


図20. 重積分のイメージ

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、  $O_2(z, w)$  という直交座標系に変換するという操作をします。

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

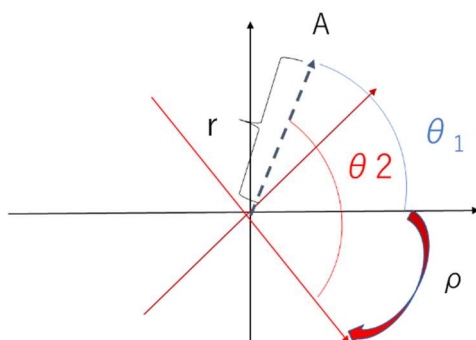


図20-2. 直交座標系 $O_1(x, y)$ で記述された点Aを回転した直交座標系 $O_2(z, w)$ で記述する作業

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

$$z = r \cos \theta_2$$

$$w = r \sin \theta_2$$

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

という関係になります。

この関係を利用して、 $\rho$ と $\theta_1$ 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、

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

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

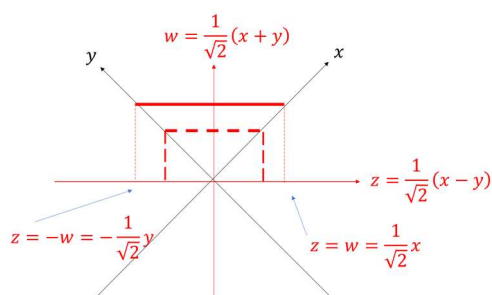


図20-3. 重積分の俯瞰図

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

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

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

$u = \sin \theta$  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、

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

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

$\frac{\chi^2_{2:x+y}}{2} = t$  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。

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

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  $\chi^2$  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  $\chi^2$  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 ( $\varphi$  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

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

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

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

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

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

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

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

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

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

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

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

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

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

という形になります。元に戻って、

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

として、 $A_{\frac{1}{2}}$  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、

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

を微分したものが  $P(\chi^2_{\frac{1}{2}}) d\chi^2_{\frac{1}{2}}$  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 ( $\varphi$ 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し、 $A_n$ が何だかわかりません。とにかくそう書けることにします。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、 $d\chi^2_n = 2dt$

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

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

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

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

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

です。

ここで、使った $\chi^2$ 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 $\Gamma$ 関数を与えてそこから $\chi^2$ 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 $\chi^2$ 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。 $\chi^2$ 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。 $\chi^2$ 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、 $\chi^2$ 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 $\chi^2$ 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。論理的な信頼性が低いわけではない）。

$\chi^2$ 分布は分散比の検定だという話。

$\chi^2$ 検定では、 $\chi^2$ 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、 $E(x^2)$ 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

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

となります。変形していきます。

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

$\frac{z}{2} = t$  とする。  $dz = 2dt$

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

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

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

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

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

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

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  $\chi^2$  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、 $\sigma^2$ : 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 $\chi^2$ 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、 $\chi^2$ 分布も確率分布モデルを使った分散分析なのです。

この解説は、途中でおかしなことをしています。積分変数の変換のところで、

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

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

としています。。どう考えたって、天下りの的に

$$w = x + y$$

$$z = x - y$$

として、解析上は何の問題もないでしょう。 $w$ と $z$ の直交性の問題も何とかなるでしょう

(必要ならばそこで回転の議論をして納得させることも可能です)。わざわざ、 $\frac{1}{\sqrt{2}}$ を付けた

のは、直交的なものを、45度回転するイメージと一致させて、積分を感覚的に納得しやすくするためでもあるのですが、最後の、 $\chi^2$ 値のところで、期待値で割るのが、偏差で割っている、 $\chi^2$ 値が分散比だという感覚を実感させたかったからです。(蛇足)

