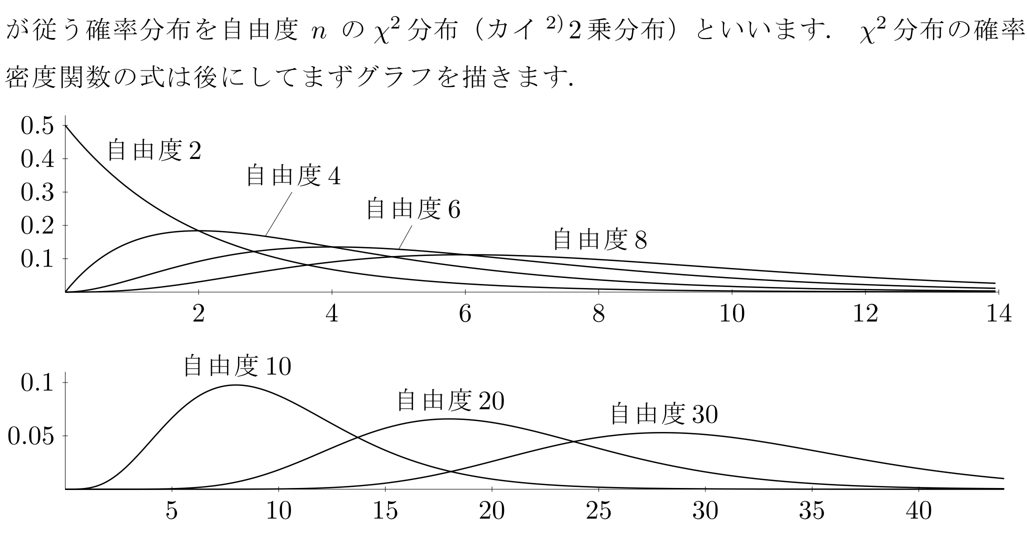
§9.1　χ^2 分布

　正の自然数 n 及び標準正規分布 $\mathrm{N}(0,1)$ に従う互いに独立な n 個の確率変数 Z_1, $Z_2, Z_3, \ldots, Z_n$ に対して，確率変数

$$Z_1^2 + Z_2^2 + Z_3^2 + \cdots + Z_n^2 = \sum_{k=1}^{n} Z_k^2$$

が従う確率分布を自由度 n の χ^2 分布（カイ[2] 2乗分布）といいます． χ^2 分布の確率密度関数の式は後にしてまずグラフを描きます．



χ^2 分布の確率密度関数のグラフ

　次の定理が χ^2 分布を用いる議論の元になります． その説明は後に回します．

定理 9.1.1　2 以上の自然数 n に対して n 個の確率変数 $X_1, X_2, X_3, \cdots, X_n$ は総て同じ正規分布に従い，互いに独立であるとする． それらの標準偏差を σ とおく． X_1, $X_2, X_3, \cdots, X_n$ の平均を表す確率変数 $\overline{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ 及び偏差平方和を表す確率変数 $S = \sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$ に対して，確率変数 $\dfrac{S}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う．

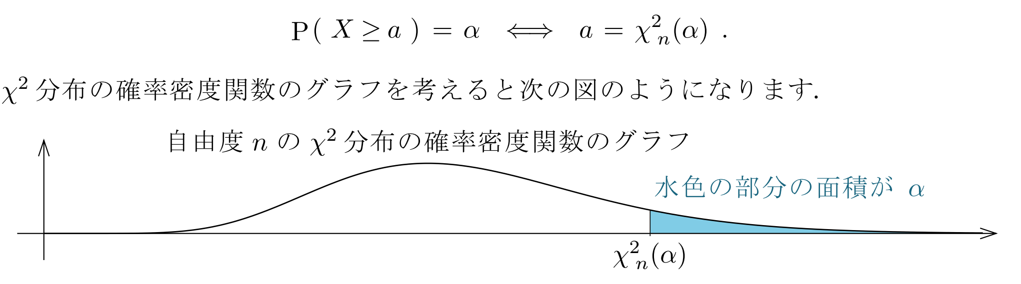
　この定理を標本分布に適用します． 母標準偏差が σ （ $\sigma > 0$ ）である正規母集団から無作為抽出された大きさ n の標本の標本確率変数 $X_1, X_2, X_3, \cdots, X_n$ は，総て同じ正規分布に従い，互いに独立で，それらの標準偏差は σ です． 前述の定理より，標本確率変数 $X_1, X_2, X_3, \cdots, X_n$ の平均を表す確率変数 $\overline{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ 及び X_1, $X_2, X_3, \cdots, X_n$ の偏差平方和を表す確率変数 $S = \sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$ に対して，確率変数 $\dfrac{S}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従います．

> **定理** 9.1.2　母標準偏差が σ （ $\sigma > 0$ ）である正規母集団から無作為抽出した大きさ n （ $n \geq 2$ ）の正規標本の偏差平方和を表す確率変数を S とおくと，確率変数 $\dfrac{S}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う．

　正の自然数 n に対して，確率変数 X は自由度 n の χ^2 分布に従うとします． 証明は略しますが， $0 < \alpha < 1$ である各実数 α に対して， $\mathrm{P}(\,X \geq k\,) = \alpha$ となる実数 k が唯一つ定まります． この実数 k を，自由度 n の χ^2 分布の上側 α 点といい，$\chi^2{}_n(\alpha)$ と書き表します． つまり次のことが成り立ちます： $0 < \alpha < 1$ である各実数 α 及び各実数 a について，

$$\mathrm{P}(\,X \geq a\,) = \alpha \iff a = \chi^2{}_n(\alpha) \ .$$

χ^2 分布の確率密度関数のグラフを考えると次の図のようになります．



自由度 n の χ^2 分布の上側 α 点

更に， $0 < \alpha < 1$ である各実数 α 及び各実数 a について，

$$\begin{aligned} \mathrm{P}(\,X \leq a\,) = \alpha &\iff 1 - \mathrm{P}(\,X \geq a\,) = \alpha \iff \mathrm{P}(\,X \geq a\,) = 1 - \alpha \\ &\iff a = \chi^2{}_n(1-\alpha) \ . \end{aligned}$$

定理 9.1.3　確率変数 X は自由度 n の χ^2 分布に従うとする． $0 < \alpha < 1$ である各実数 α 及び各実数 a について，

$$\mathrm{P}(\,X \geq a\,) = \alpha \iff a = \chi^2{}_n(\alpha) \ ,$$
$$\mathrm{P}(\,X \leq a\,) = \alpha \iff a = \chi^2{}_n(1-\alpha) \ .$$

　正の自然数 n と実数 α （ $0 \leq \alpha \leq 1$ ）とに対する自由度 n の χ^2 分布の上側 α 点 $\chi^2{}_n(\alpha)$ の値を表にしたものを χ^2 分布表といいます．

$\boxed{\text{例題 9.1}}$　母標準偏差が σ （ $\sigma > 0$ ）である正規母集団から無作為に非復元抽出した大きさ 30 の正規標本の偏差平方和を表す確率変数 S に対して，確率変数 $\dfrac{S}{\sigma^2}$ を考える．

(1)　$\mathrm{P}\left(\dfrac{S}{\sigma^2} \geq a\right) = 5\%$ である実数 a を求める．

(2)　$\mathrm{P}\left(\dfrac{S}{\sigma^2} \leq b\right) = 1\%$ である実数 b を求める．

　確率変数 $\dfrac{S}{\sigma^2}$ は自由度 29 の χ^2 分布に従う．

(1)

$$\mathrm{P}\left(\frac{S}{\sigma^2} \geq a\right) = 0.05 \iff a = \chi^2{}_{29}(0.05) \fallingdotseq 42.56 \ .$$

(2)

$$\mathrm{P}\left(\frac{S}{\sigma^2} \leq b\right) = 0.01 \iff b = \chi^2{}_{29}(0.99) \fallingdotseq 14.26 \ .$$

$\boxed{\text{終}}$

$\boxed{\text{問題 9.1}}$　母標準偏差が σ （ $\sigma > 0$ ）である正規母集団から無作為に非復元抽出した大きさ 20 の正規標本の不偏分散を表す確率変数 U に対して，確率変数 $\dfrac{S}{\sigma^2}$ を考えます．

(1)　$\mathrm{P}\left(\dfrac{S}{\sigma^2} \geq a\right) = 1\%$ である実数 a を求めなさい．

(2)　$\mathrm{P}\left(\dfrac{S}{\sigma^2} \leq b\right) = 5\%$ である実数 b を求めなさい．

――――― 定理の説明

　定理 9.1.1 を説明します． μ と σ とは定数で $\sigma > 0$ とします．2 以上の自然数 n に対して，n 個の確率変数 $X_1, X_2, X_3, \ldots, X_n$ は正規分布 $\mathrm{N}(\mu, \sigma^2)$ に従い互いに独立であるとします． n 以下の正の自然数 k に対して，k 個の確率変数 X_1, X_2, $X_3, \ldots, X_k$ の偏差平方和を表す確率変数を S_k とおきます：

$$\begin{aligned} S_k &= \sum_{j=1}^{k}\left(X_j - \frac{1}{k}\sum_{i=1}^{k} X_i\right)^2 = \sum_{j=1}^{k}\left\{{X_j}^2 - \frac{2X_j}{k}\sum_{i=1}^{k} X_i + \frac{1}{k^2}\left(\sum_{i=1}^{k} X_i\right)^2\right\} \\ &= \sum_{j=1}^{k}{X_j}^2 - \frac{2}{k}\sum_{j=1}^{k} X_j \sum_{i=1}^{k} X_i + \frac{1}{k^2}\sum_{j=1}^{k}\left(\sum_{i=1}^{k} X_i\right)^2 \\ &= \sum_{j=1}^{k}{X_j}^2 - \frac{2}{k}\left(\sum_{i=1}^{k} X_i\right)^2 + \frac{1}{k}\left(\sum_{i=1}^{k} X_i\right)^2 \\ &= \sum_{i=1}^{k}{X_i}^2 - \frac{1}{k}\left(\sum_{i=1}^{k} X_i\right)^2 \ . \end{aligned}$$

$k \geq 2$ とします．

$$\begin{aligned} S_k - S_{k-1} &= \sum_{i=1}^{k}{X_i}^2 - \frac{1}{k}\left(\sum_{i=1}^{k} X_i\right)^2 - \sum_{i=1}^{k-1}{X_i}^2 + \frac{1}{k-1}\left(\sum_{i=1}^{k-1} X_i\right)^2 \\ &= {X_k}^2 + \frac{1}{k-1}\left(\sum_{i=1}^{k-1} X_i\right)^2 - \frac{1}{k}\left(\sum_{i=1}^{k} X_i\right)^2 \\ &= {X_k}^2 + \frac{1}{k-1}\left(\sum_{i=1}^{k-1} X_i\right)^2 - \frac{1}{k}\left(\sum_{i=1}^{k-1} X_i + X_k\right)^2 \\ &= {X_k}^2 + \frac{1}{k-1}\left(\sum_{i=1}^{k-1} X_i\right)^2 - \frac{1}{k}\left\{\left(\sum_{i=1}^{k-1} X_i\right)^2 + 2X_k\sum_{i=1}^{k-1} X_i + {X_k}^2\right\} \\ &= \frac{1}{k(k-1)}\left\{\left(\sum_{i=1}^{k-1} X_i\right)^2 - 2(k-1)X_k\sum_{i=1}^{k-1} X_i + (k-1)^2{X_k}^2\right\} \\ &= \frac{1}{k(k-1)}\left\{\sum_{i=1}^{k-1} X_i - (k-1)X_k\right\}^2 \\ &= \left[\frac{1}{\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1} X_i - (k-1)X_k\right\}\right]^2 \ , \end{aligned}$$

σ^2 で割ると

$$\frac{S_k}{\sigma^2} - \frac{S_{k-1}}{\sigma^2} = \left[\frac{1}{\sigma\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1} X_i - (k-1)X_k\right\}\right]^2 \ .$$

確率変数 Z_{k-1} を次のように定めます：

$$Z_{k-1} = \frac{1}{\sigma\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1} X_i - (k-1)X_k\right\} \ .$$

このとき

$$\frac{S_k}{\sigma^2} - \frac{S_{k-1}}{\sigma^2} = {Z_{k-1}}^2 \ .$$

確率変数 $X_1, X_2, X_3, \ldots, X_k$ は正規分布 $\mathrm{N}(\mu, \sigma^2)$ に従い互いに独立なので，

$$\begin{aligned} \mathrm{E}[Z_{k-1}] &= \mathrm{E}\left[\frac{1}{\sigma\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1} X_i - (k-1)X_k\right\}\right] \\ &= \frac{1}{\sigma\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1}\mathrm{E}[X_i] - (k-1)\,\mathrm{E}[X_k]\right\} \\ &= \frac{1}{\sigma\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1}\mu - (k-1)\mu\right\} \\ &= 0 \ , \end{aligned}$$

$$\begin{aligned} \mathrm{V}[Z_{k-1}] &= \mathrm{V}\left[\frac{1}{\sigma\sqrt{k(k-1)}}\left\{\sum_{i=1}^{k-1} X_i - (k-1)X_k\right\}\right] \\ &= \frac{1}{\sigma^2 k(k-1)}\left\{\sum_{i=1}^{k-1}\mathrm{V}[X_i] + (k-1)^2\,\mathrm{V}[X_k]\right\} \\ &= \frac{1}{\sigma^2 k(k-1)}\left\{\sum_{i=1}^{k-1}\sigma^2 + (k-1)^2\sigma^2\right\} \\ &= \frac{1}{\sigma^2 k(k-1)}\{\sigma^2(k-1) + \sigma^2(k-1)^2\} \\ &= 1 \ . \end{aligned}$$

定理 4.1.2 より Z_k は標準正規分布 $\mathrm{N}(0,1)$ に従います． 証明は略しますが確率変数 $Z_1, Z_2, Z_3, \ldots, Z_{n-1}$ は互いに独立です． $\dfrac{S_k}{\sigma^2} - \dfrac{S_{k-1}}{\sigma^2} = {Z_{k-1}}^2$ より，

$$\sum_{k=2}^{n}\left(\frac{S_k}{\sigma^2} - \frac{S_{k-1}}{\sigma^2}\right) = \sum_{k=2}^{n}{Z_{k-1}}^2 \ ,$$

$$\frac{S_2}{\sigma^2} - \frac{S_1}{\sigma^2} + \frac{S_3}{\sigma^2} - \frac{S_2}{\sigma^2} + \frac{S_4}{\sigma^2} - \frac{S_3}{\sigma^2} + \cdots + \frac{S_n}{\sigma^2} - \frac{S_{n-1}}{\sigma^2} = {Z_1}^2 + {Z_2}^2 + {Z_2}^2 + \cdots + {Z_{n-1}}^2 \ ,$$

$S_1 = 0$ なので

$$\frac{S_n}{\sigma^2} = \sum_{k=1}^{n-1}{Z_k}^2 \ .$$

故に確率変数 $\dfrac{S_n}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従います．

――――― χ^2 分布の確率密度関数

　まず Γ 関数といわれる関数を準備します． 区間 $(0, \infty)$ を定義域とする Γ 関数 $\Gamma(x)$ は次のように定義されます：

$$\Gamma(x) = \int_0^{\infty} e^{-t}t^{x-1}dt \quad (\ x > 0\ ) \quad .$$

この Γ 関数 $\Gamma(x)$ について次の公式が成り立ちます：正の自然数 n に対して，

$$\Gamma(n) = (n-1)(n-2)(n-3)\cdots 2\cdot 1 = (n-1)! \ ,$$
$$\Gamma\left(\frac{2n-1}{2}\right) = \frac{(2n-3)(2n-5)(2n-7)\cdots 3\cdot 1}{2^{n-1}}\sqrt{\pi} \ .$$

　正の自然数 n に対して，自由度 n の χ^2 分布の確率密度関数 $f(x)$ は次のようになります：

$$f(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)e^{\frac{x}{2}}} & (\ x \geq 0 \text{ のとき}) \\ 0 & (\ x < 0 \text{ のとき}) \end{cases} \ .$$

---

[2] χ はギリシャ文字で“カイ”と言います．