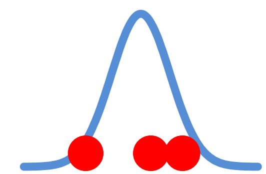
F分布

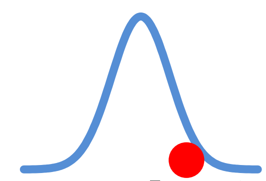
複数の変数を含む2群
(分散の比を考える)

カイ二乗分布



複数の変数
(分散を考える)

標準正規分布

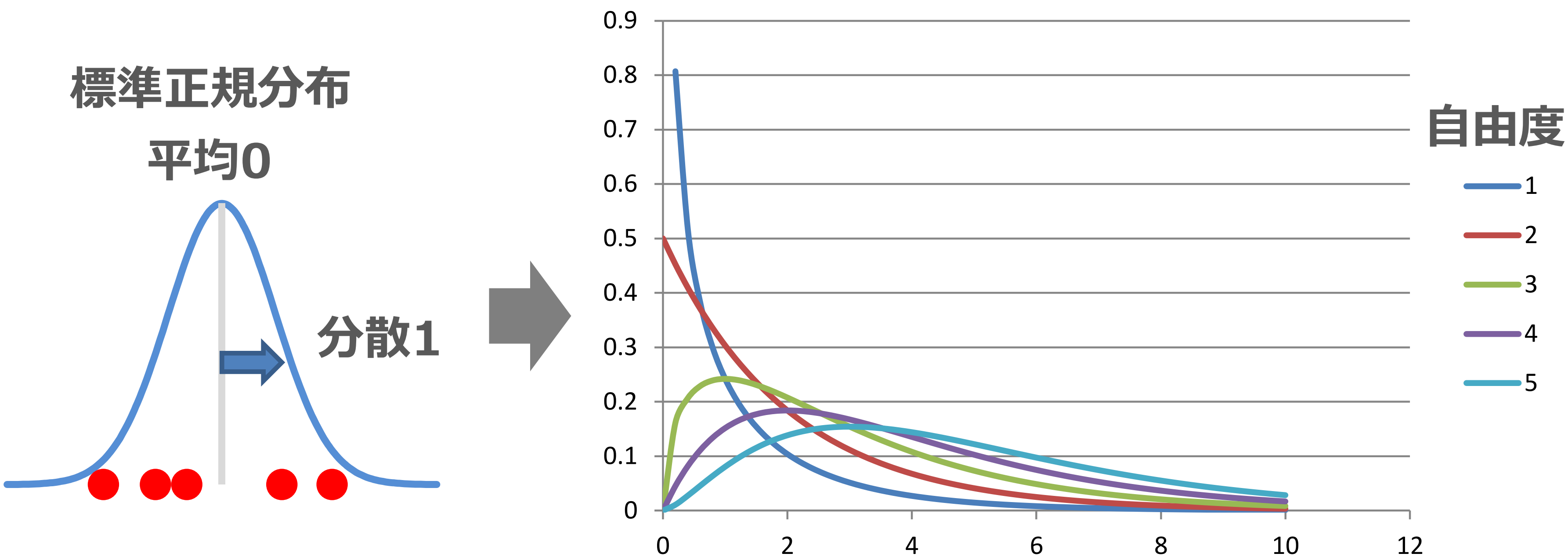


ひとつの変数

by 櫻井

カイ二乗分布

標準正規分布に従った独立した変数がいくつかあるとき、その二乗和が従う分布



カイ二乗分布の性質

正規分布$N(\mu, \sigma^2)$に従ったk個の変数x_iについて、偏差（平均からの差）の平方和と分散の比は、自由度kのカイ二乗分布に従う

$$\chi^2 = \frac{\sum_{i=1}^{k}(x_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{k}\left(\frac{x_i - \mu}{\sigma}\right)^2$$

カイ二乗検定

	ビール 好き	ビール あんまり
男性	23	12
女性	7	8

二つのカテゴリに関連があるかを調べたい

帰無仮説：
二つのカテゴリは独立である（関連がない）

有意水準：0.05

カイ二乗検定の手順

（１）観測データから、カテゴリーごとに割合を出す

	ビール 好き	ビール あんまり	合計
男性	69	36	105 70%
女性	21	24	45 30%
合計	90 60%	60 40%	150 100%

（２）割合から、カテゴリーが独立な場合の度数（期待度数）を出す

	ビール 好き	ビール あんまり	合計
男性	63	42	105 70%
女性	27	18	45 30%
合計	90 60%	60 40%	150 100%

カイ二乗検定の手順

（3）観測度数と期待度数の差を出す

	ビール好き	ビールあんまり
男性	6	-6
女性	-6	6

（4）その二乗を出す

	ビール好き	ビールあんまり
男性	36	36
女性	36	36

（5）期待度数で割る

	ビール好き	ビールあんまり
男性	36/63 =0.57	36/42 =0.86
女性	36/27 =1.33	36/18 =2

（6）その和を求める

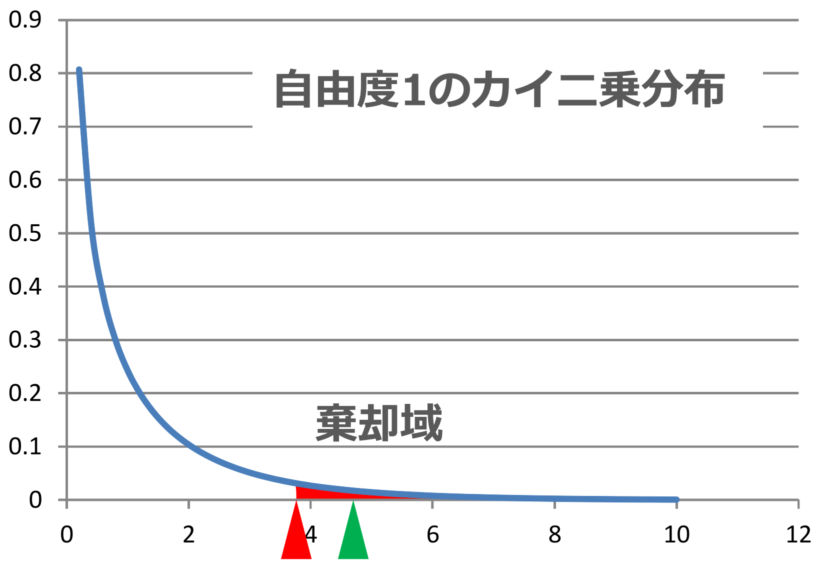
x = 0.57 + 0.86 + 1.33 + 2 = **4.76**

このように求めた値**x**は、カイ二乗分布に**近似できる**。
自由度は、各カテゴリ（性別、ビールの好み）の要素数をそれぞれn_1, n_2とすると、$(n_1-1)*(n_2-1)$。
この例の場合では、(2-1)*(2-1) = 1

カイ二乗検定の手順

（７）結論

xの値が棄却域の境界値の外側（3.84 < 4.76, p=0.029 < α）なので、帰無仮説は棄却され、「二つのカテゴリは独立ではない」と判断された。



α=0.05,
境界値3.84

x = 4.76,
p = 0.029

よって、この母集団においては、「性別とビールの好みとの間に何かしらの関連性がある」と結論づけられた。

カイ二乗分布を扱うExcelの関数：
CHISQ.DIST, CHISQ.DIST.RT, CHISQ.INV.RTなど

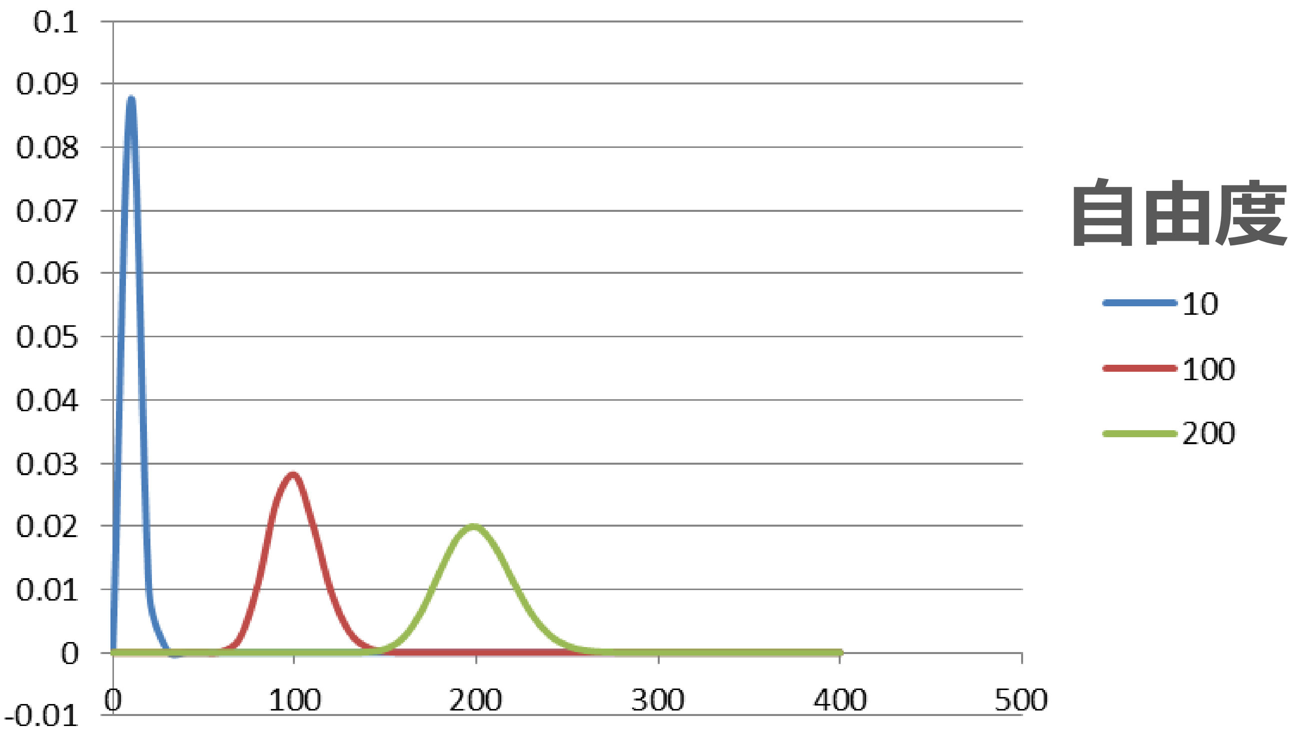
カイ二乗検定の留意点

観測数が少ないとカイ二乗分布への近似ができないので、
その場合はフィッシャーの正確確率検定を行う。

目安：
期待度数が5未満のセルが、全セルの20%以上で存在
する場合、近似が不正確と考えられる
（コクラン・ルール）

期待度数が1未満のセルがあってはならない

カイ二乗分布の性質　その2

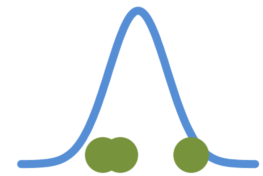
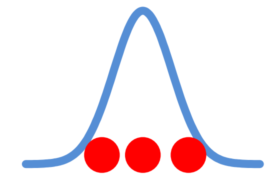


自由度kが大きくなると、

平均値：k
分散：2k

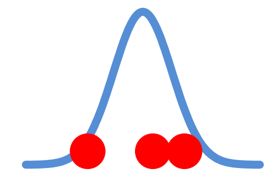
の正規分布に近づいてゆく

F分布

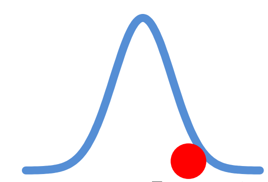
複数の変数を含む2群
（分散の比を考える）

カイ二乗分布



複数の変数
（分散を考える）

標準正規分布



ひとつの変数

by 櫻井

F分布とカイ二乗分布の関係

自由度k_1のカイ二乗分布$\chi^2{}_1$

自由度k_2のカイ二乗分布$\chi^2{}_2$

があるとき、次の値Fは、自由度(k_1, k_2)のF分布に従う

$$F = \frac{\chi^2{}_1/k_1}{\chi^2{}_2/k_2}$$

χ^2分布が
ふたつ!!

F 分布の活用

正規分布N(μ_1,$\sigma^2{}_1$)に従った母集団から得た標本、
標本数：n_1、不偏標本分散：$v^2{}_1$

正規分布N(μ_2,$\sigma^2{}_2$)に従った母集団から得た標本、
標本数：n_2、不偏標本分散：$v^2{}_2$

があるとき、

$$F = \frac{\chi^2{}_1/k_1}{\chi^2{}_2/k_2} = \frac{v^2{}_1/\sigma^2{}_1}{v^2{}_2/\sigma^2{}_2}$$

二つの母集団の分散$\sigma^2{}_1$と$\sigma^2{}_2$が等しいと仮定できる場合は、

$$F = \frac{v^2{}_1}{v^2{}_2}$$

これをF検定で利用している！

F 分布の活用

カイ二乗分布の性質

$$\chi^2 = \frac{\sum_{i=1}^{k}(x_i - \mu)^2}{\sigma^2}$$ 自由度k

この式を変形すると、

不偏標本分散v²になっている！

$$\chi^2 = \frac{k \times \frac{\sum_{i=1}^{k}(x_i - \mu)^2}{k}}{\sigma^2} = \frac{k \times v^2}{\sigma^2}$$

したがって、

$$\frac{\chi^2}{k} = \frac{k \times v^2}{\sigma^2} \times \frac{1}{k} = \frac{v^2}{\sigma^2}$$