

χ^2 分布, t分布, F分布

自由度, ポアソン分布

意味のわかる統計学 P.260～P.280

2018年11月16日

60216087 木村 祐太

おさらい

推定に用いる統計量(1/2)

- σ が既知のとき, μ を区間推定

「統計量 $T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は, 標準正規分布 $N(0, 1)$ に従う」

- σ が未知のとき, μ を区間推定

「統計量 $T = \frac{\bar{X} - \mu}{\frac{U}{\sqrt{n}}}$ は自由度 $n - 1$ の t 分布に従う」

$$\text{※ } U = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n - 1}}$$

推定に用いる統計量(2/2)

- μ が既知のとき, σ を区間推定

「統計量 $T = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{\sigma^2}$ は,

自由度 n の χ^2 分布に従う」

- μ が未知のとき, σ を区間推定

「統計量 $T = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{\sigma^2}$ は,

自由度 $n - 1$ の χ^2 分布に従う」

検定に用いる統計量(1/2)

- σ が既知のとき, μ を検定

「統計量 $T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は, 標準正規分布 $N(0, 1)$ に従う」

- σ が未知のとき, μ を検定

「統計量 $T = \frac{\bar{X} - \mu}{\frac{U}{\sqrt{n}}}$ は自由度 $n - 1$ の t 分布に従う」

$$\text{※ } U = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n - 1}}$$

検定に用いる統計量(2/2)

- μ が既知のとき, σ を検定

「統計量 $T = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{\sigma^2}$ は,

自由度 n の χ^2 分布に従う」

- μ が未知のとき, σ を検定

「統計量 $T = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{\sigma^2}$ は,

自由度 $n - 1$ の χ^2 分布に従う」

2変量の検定に用いる統計量(1/3)

(2変量の検定における前提)

正規分布Aから抽出したサンプルをX,
正規分布Bから抽出したサンプルをYとする.

- $\sigma_x^2 = \sigma_y^2$ であることを検定

「 $\sigma_x^2 = \sigma_y^2$ のもとで, 統計量 $T = \frac{U_x^2}{U_y^2}$ は,

自由度 $(m - 1, n - 1)$ の F 分布に従う」

2変量の検定に用いる統計量(2/3)

- σ_x^2, σ_y^2 がともに既知のとき, $\mu_x = \mu_y$ であることを検定

「 $\mu_x = \mu_y$ のもとで, 統計量 $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$ は,

標準正規分布 $N(0, 1)$ に従う」

2変量の検定に用いる統計量(3/3)

- σ_x^2 と σ_y^2 が等しいと見なせるとき, $\mu_x = \mu_y$ であることを検定

「 $\mu_x = \mu_y$ のもとで, 統計量 $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) U_{xy}^2}}$ は,

自由度($m + n - 2$)の t 分布に従う」

χ^2 分布

χ^2 分布とは

- 実現値と理論値の違いを調べるための分布
- 標本平均のみで判断できないときに用いる
(変量の散らばり、偏り)

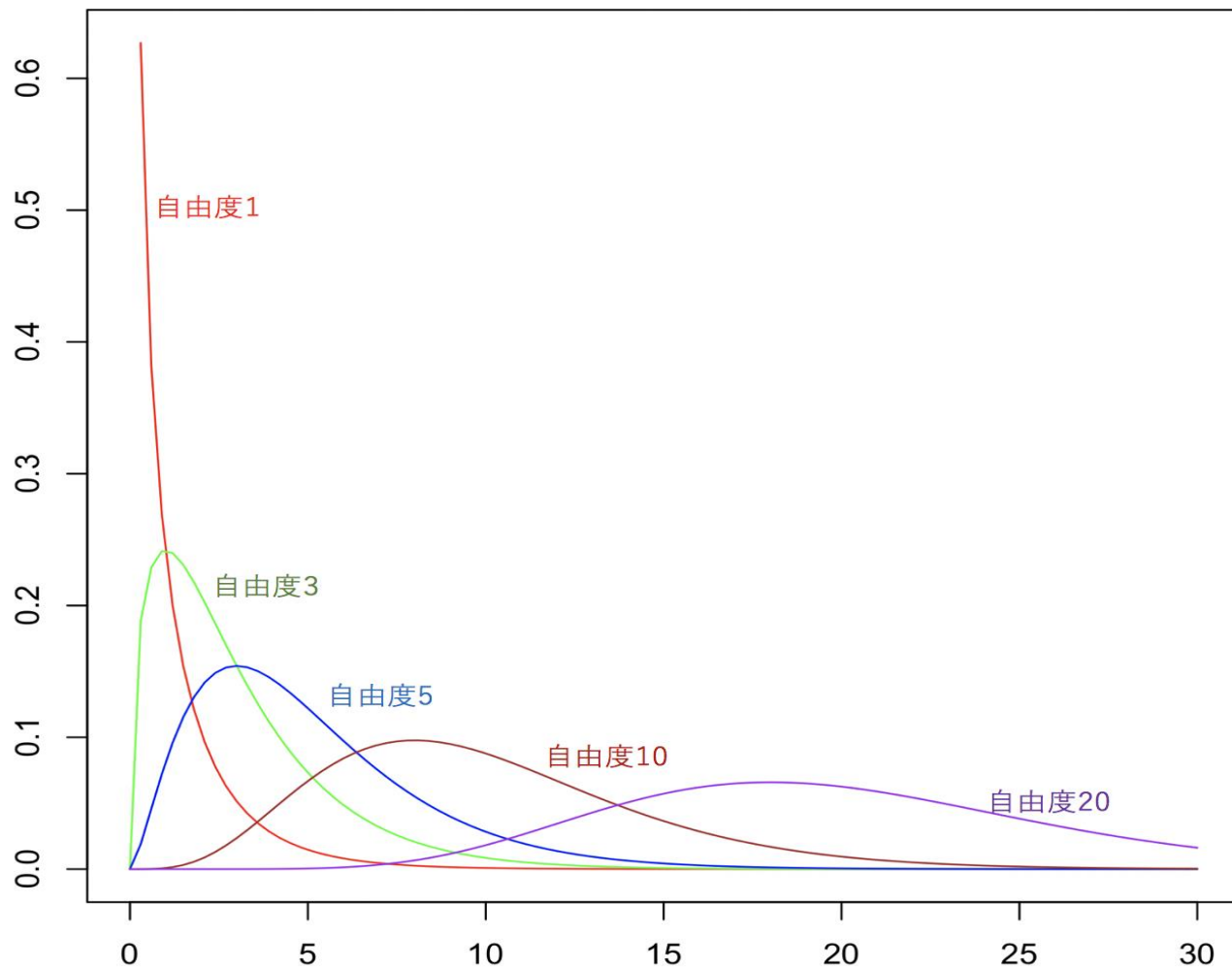
χ^2 分布の定義

- $Z_1, Z_2, \dots, Z_n$ が互いに独立な標準正規分布 $N(0, 1)$ に従うとする.

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

- この X の分布を, 自由度 n の χ^2 分布という.

χ^2 分布のグラフ



推定の確認(1/6)

- μ が既知のとき, σ を区間推定

「統計量 $T = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{\sigma^2}$ は,

自由度 n の χ^2 分布に従う」

推定の確認(2/6)

- 統計量 $T = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{\sigma^2}$

$$\begin{aligned} T &= \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2}{\sigma^2} \\ &= \frac{(X_1 - \mu)^2}{\sigma^2} + \frac{(X_2 - \mu)^2}{\sigma^2} + \dots + \frac{(X_n - \mu)^2}{\sigma^2} \\ &= \left(\frac{X_1 - \mu}{\sigma} \right)^2 + \left(\frac{X_2 - \mu}{\sigma} \right)^2 + \dots + \left(\frac{X_n - \mu}{\sigma} \right)^2 \\ &= Z_1^2 + Z_2^2 + \dots + Z_n^2 \end{aligned}$$

∴ 自由度 n の χ^2 分布に従う

推定の確認(3/6)

- μ が未知のとき, σ を区間推定

「統計量 $T = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{\sigma^2}$ は,

自由度 $n - 1$ の χ^2 分布に従う」

推定の確認(4/6)

- 統計量 $T = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{\sigma^2}$

$$\begin{aligned}\bar{X} &= \frac{X_1 + X_2 + \dots + X_n}{n} \\ &= \frac{(\sigma Z_1 + \mu) + (\sigma Z_2 + \mu) + \dots + (\sigma Z_n + \mu)}{n} \quad (\because Z = \frac{X - \mu}{\sigma}) \\ &= \sigma \frac{Z_1 + Z_2 + \dots + Z_n}{n} + \mu \\ &= \sigma \bar{Z} + \mu \\ \therefore \bar{Z} &= \frac{\bar{X} - \mu}{\sigma}\end{aligned}$$

推定の確認(5/6)

$$\begin{aligned} Z - \bar{Z} &= \frac{X - \mu}{\sigma} - \frac{\bar{X} - \mu}{\sigma} \\ &= \frac{X - \bar{X} - \mu + \mu}{\sigma} \\ &= \frac{X - \bar{X}}{\sigma} \end{aligned}$$

推定の確認(6/6)

$$\begin{aligned} T &= \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{\sigma^2} \\ &= \frac{(X_1 - \bar{X})^2}{\sigma^2} + \frac{(X_2 - \bar{X})^2}{\sigma^2} + \dots + \frac{(X_n - \bar{X})^2}{\sigma^2} \\ &= \left(\frac{X_1 - \bar{X}}{\sigma}\right)^2 + \left(\frac{X_2 - \bar{X}}{\sigma}\right)^2 + \dots + \left(\frac{X_n - \bar{X}}{\sigma}\right)^2 \\ &= (Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 + \dots + (Z_n - \bar{Z})^2 \end{aligned}$$

再生性より, $Z - \bar{Z}$ もまた標準正規分布

∴ 自由度 $n - 1$ の χ^2 分布に従う

t分布

t分布とは

- 標本平均と不偏分散から母平均を推定
- 十分なサンプルがないときに用いる

t分布の定義

- Y, Z が互いに独立な確率変数とする.
- Y は標準正規分布 $N(0, 1)$ に従い, Z は自由度 n の χ^2 分布に従うものとする.

$$X = \frac{Y}{\sqrt{\frac{Z}{n}}}$$

- X の分布を、自由度 n の t 分布という.

(再掲載)

- σ が未知のとき, μ を区間推定

「統計量 $T = \frac{\bar{X} - \mu}{\frac{U}{\sqrt{n}}}$ は自由度 $n - 1$ の t 分布に従う」

$$\times U = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}}$$

(再掲載)

- σ_x^2 と σ_y^2 が等しいと見なせるとき, $\mu_x = \mu_y$ であることを検定

「 $\mu_x = \mu_y$ のもとで, 統計量 $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) U_{xy}^2}}$ は,

自由度 $(m + n - 2)$ の t 分布に従う」

F分布

F分布とは

- 2つのサンプルの分散の違いを調べるための分布
- 自由度, 不偏分散を用いる

F分布の定義

- Y, Z が互いに独立な確率変数とする.
- Y は自由度 m の χ^2 分布に従い, Z は自由度 n の χ^2 分布に従うものとする.

$$X = \frac{\left(\frac{Y}{m}\right)}{\left(\frac{Z}{n}\right)}$$

- X の分布を、自由度 n の F 分布という.

(再掲載)

(2変量の検定における前提)

正規分布Aから抽出したサンプルをX,
正規分布Bから抽出したサンプルをYとする.

- $\sigma_x^2 = \sigma_y^2$ であることを検定

「 $\sigma_x^2 = \sigma_y^2$ のもとで, 統計量 $T = \frac{U_x^2}{U_y^2}$ は,

自由度 $(m - 1, n - 1)$ の F 分布に従う」

自由度

自由度とは(1/2)

- 自由に動かすことのできる変数の個数
- <例>

変数 x, y, z が存在.

→ 自由度3

変数 x, y, z が存在し, x, y, z は

$x + y + z = 1$ という関係式が成り立つ.

→ 自由度2

(理由) x, y の値が決まれば z の値が決まる.

(再掲載)

- μ が既知のとき, σ を区間推定

「統計量 $T = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_n - \mu)^2}{\sigma^2}$ は,

自由度 n の χ^2 分布に従う」

- μ が未知のとき, σ を区間推定

「統計量 $T = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{\sigma^2}$ は,

自由度 $n - 1$ の χ^2 分布に従う」

自由度とは(2/2)

- μ が既知のとき, σ を区間推定
→変量 X に対して制約なし
自由度はサンプルサイズと等しい
- μ が未知のとき, σ を区間推定
→変量 X の標本平均 $\bar{X}$ を用いるため,
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
の関係式が成り立つ
自由度はサンプルサイズに-1した値

ポアソン分布

ポアソン分布とは(1/5)

- 稀にしか起こらない事象を単位時間あたり, 単位空間あたり何回起きるかの統計分布
- 平均発生回数を λ , 発生回数を k とすると

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{で近似できる}$$

- 分布の平均, 分散はともに
 $\lambda (=np (n: \text{個数}, p: \text{発生確率}))$
(マクローリン展開より)

ポアソン分布とは(2/5)

- 二項分布による近似

二項分布: ${}_nC_k p^k (1-p)^{n-k}$, 平均: $pn = \lambda$

$$\begin{aligned} {}nC_k p^k &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{1}{k!} \left(n \frac{\lambda}{n}\right)^k \\ &= 1\left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \end{aligned}$$

$$\lim_{n \rightarrow \infty} {}nC_k p^k = \frac{\lambda^k}{k!}$$

ポアソン分布とは(3/5)

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

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

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

$$\lim_{n \rightarrow \infty} (1-p)^{n-k} = e^{-\lambda}$$

ポアソン分布とは(4/5)

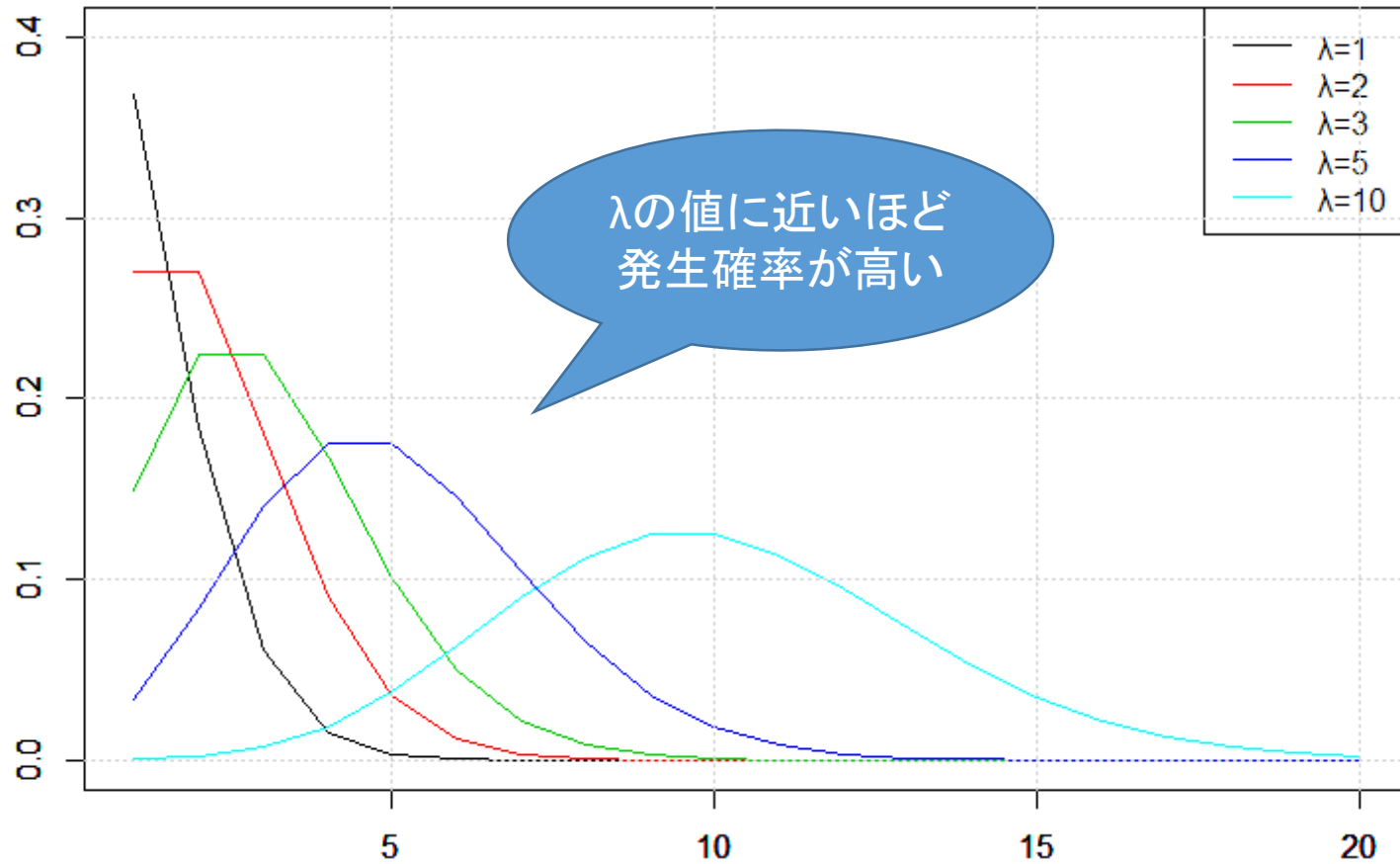
$$\lim_{n \rightarrow \infty} {}_nC_k p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

∴ n が大きくなると、ポアソン分布に近似する.

ポアソン分布とは(5/5)

- λ の値が大きくなると, 平均が λ , 分散が λ の正規分布に近づく

ポアソン分布のグラフ



ポアソン分布の実用例

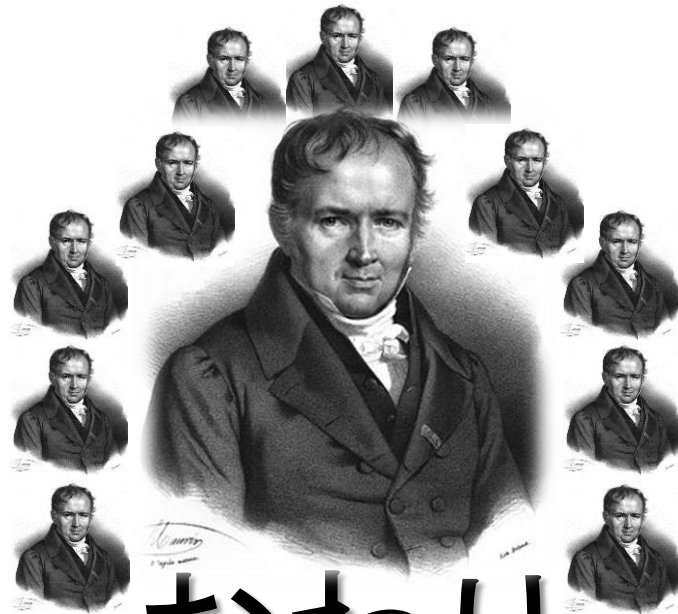
- 1ページ当たりの誤植の個数
- 1日当たりの交通事故の件数
- 生産物に含まれる不良部品の個数

まとめ

まとめ

- χ^2 分布, t分布, F分布の使い分け
- 標本平均を用いた推定, 検定は不偏分散
- 自由度は条件式の数だけ少なくなる
- ポアソンってこんな人





おわり

