統計学　計算ノート

2022 年 6 月 11 日版

馬場真哉

https://logics-of-blue.com/

更新履歴

2022 年 06 月 11 日：『統計学　計算ノート』の第 1 版を公開

はじめに

いわゆる「統計学入門」で扱う計算を、なるべく式変形を端折らずに解説します。高校理系卒業レベル（ほとんどは高校 2 年生レベル）で理解できる内容だと思います。

当方の新刊『Python で学ぶあたらしい統計学の教科書 第 2 版』の内容ともある程度対応していますので、拙著の副読本としてもお使いいただけると思います。

なお、本資料は統計学の入門的な内容をすでに学んだ方を読者として想定しています。統計学の基本的な考え方を全く知らないという方は、先に『Python で学ぶあたらしい統計学の教科書 第 2 版』など、初等的な統計学の入門書を読まれておくことをおすすめします。

証明のために、さまざまな公式や定理を利用します。証明を読むだけでなく、公式や定理を使えるようになっておくと、中級以上の統計学の教科書を読み進める際に役に立つと思います。

ただ、式変形にばかりこだわっていても統計学は理解できないので、実践的なデータ分析と数式の変形はバランスよく勉強するのがおすすめです。難しいと感じたら、計算は後回しでもいいかもしれません。プログラミングをしながら手を動かしてデータ分析を学ぶというやり方が向いている人もいます。もちろん数式から入る方が向いている人もいます。

数理的な側面は、興味を持った時に時間をかけてゆっくりと勉強するのがおすすめです。一朝一夕では何ともならないですので、気長にどうぞ。

証明の際には下記の書籍を参考にしました。上から順に少しずつ難易度が上がるイメージです。

- 倉田博史・星野崇宏. (2009). 入門統計解析. 新世社
- 改訂版 日本統計学会公式認定　統計検定 2 級対応　統計学基礎. 東京図書
- 山田作太郎・北田修一. (2004). 生物統計学入門. 成山堂書店
- 古賀弘樹. (2018). 一段深く理解する 確率統計. 森北出版
- 松井秀俊・小泉和之（竹村彰通 編）. (2019). 統計モデルと推測. 講談社
- 山本 拓. (1995). 計量経済学 (新経済学ライブラリ). 新世社
  - 2022 年 3 月に第 2 版が出たようです（参照したのは初版）
- 鈴木武・山田作太郎. (1996). 数理統計学－基礎から学ぶデータ解析－. 内田老鶴圃

※このノートは不定期に更新される可能性があります。また、著者個人が作成したものであり、所属する組織の意見ではありません。誤りやご指摘がありましたら、ブログにコメントをいただけますと幸いです。

定義集

初等的な統計学の入門書でしばしば登場する定義をここで整理します。証明を読み進める際に定義を忘れてしまった方はこちらを参照してください。

記号

X, Y　確率変数

x, y　実現値

μ　母平均（期待値）

σ^2　母分散

平均値（標本平均）

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{0-1}$$

期待値

離散型の確率分布　$$E(X) = \sum_{i=1}^{n} f(x_i) \cdot x_i \tag{0-2}$$

連続型の確率分布　$$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \, dx \tag{0-3}$$

標本分散

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{0-4}$$

不偏分散

$$u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{0-5}$$

確率変数の分散

一般的な定義　$V(X) = E[(X-\mu)^2]$　(0-6)

離散型の確率分布　$V(X) = \sum_{i=1}^{n} f(x_i)\cdot(x_i-\mu)^2$　(0-7)

連続型の確率分布　$V(X) = \int_{-\infty}^{\infty} f(x)\cdot(x-\mu)^2\,dx$　(0-8)

標本共分散

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{0-9}$$

確率変数の共分散

$$C(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \tag{0-10}$$

条件付き確率分布

$$P\left(X = x_i \middle| Y = y_j\right) = \frac{P\left(X = x_i, Y = y_j\right)}{P\left(Y = y_j\right)} \tag{0-11}$$

独立（下記の 2 つは同義）

$$P\left(X = x_i, Y = y_j\right) = P(X = x_i) \cdot P\left(Y = y_j\right) \tag{0-12}$$

$$P\left(X = x_i \middle| Y = y_j\right) = P(X = x_i) \tag{0-13}$$

基本公式

証明の際、大いに参照する公式を整理します。なお、連番を振る都合上、定理や公式といった用語の使い分けはしません。すべて「公式」と呼び統一的に連番を振ります。

公式をひたすら覚えるのは無駄なので、必要に応じて参照するような使い方にしてください。

０．Σ記号に関する公式

Σ記号の公式を整理します。下記の公式は、特に明示しないで参照します。

添え字がつかない場合

$$\sum_{i=1}^{n} x = n \cdot x \tag{1-1}$$

Σ記号の中身がx_iではなくxであることに注意してください。添え字iが無くなっています。この場合はxはずっと同じ値のまま変化しません。ただしn回足し合わせるという作業は行われるので、$x + x + x + \cdots$をn回繰り返して、結果は$n \cdot x$となります。

Σ記号の中の足し算

以下のようなΣ記号の中の足し算は、Σ記号の外側に出すことができます。

$$\begin{aligned} \sum_{i=1}^{n}(x_i + y_i) &= x_1 + y_1 + x_2 + y_2 + \cdots + x_n + y_n \\ &= (x_1 + \cdots + x_n) + (y_1 + \cdots + y_n) \qquad (1\text{-}2) \\ &= \sum_{i=1}^{n}(x_i) + \sum_{i=1}^{n}(y_i) \end{aligned}$$

1 行目でΣ記号の中身を展開し、2 行目でxとyで並び替えています。

以下のような分解もしばしば登場します。定数aには添え字iがついていないので、これは$n \cdot a$となります。

$$\sum_{i=1}^{n}(a+x_i) = n \cdot a + \sum_{i=1}^{n} x_i \tag{1-3}$$

Σ記号の中の掛け算

以下のようなΣ記号の中の掛け算は、Σ記号の外側に出すことができます。

$$\begin{aligned}\sum_{i=1}^{n} a \cdot x_i &= a \cdot x_1 + a \cdot x_2 + \cdots + a \cdot x_n \\ &= a \cdot (x_1 + \cdots + x_n) \\ &= a \sum_{i=1}^{n} x_i\end{aligned} \tag{1-4}$$

Σ記号の中の割り算

割り算も、掛け算と同様です。

$$\begin{aligned}\sum_{i=1}^{n}(x_i \div a) &= \left(\sum_{i=1}^{n} x_i\right) \div a \\ &= \frac{1}{a}\sum_{i=1}^{n} x_i\end{aligned} \tag{1-5}$$

2重Σ記号の掛け算

やや難易度が高いです。公式7-3の証明の際に登場します。

2重Σ記号の掛け算は、分解できます。一般的には以下のようになります。

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j = \left[\sum_{i=1}^{m} a_i\right] \times \left[\sum_{j=1}^{n} b_j\right] \tag{1-6}$$

$m = n = 2$の場合のみ示します。

$$\begin{aligned}\sum_{i=1}^{2}\sum_{j=1}^{2} a_i b_j &= a_1 b_1 + a_1 b_2 + a_2 b_1 + a_2 b_2 \\ &= (a_1 + a_2) \times (b_1 + b_2) \\ &= \left[\sum_{i=1}^{2} a_i\right] \times \left[\sum_{j=1}^{2} b_j\right]\end{aligned} \tag{1-7}$$

1. 平均値に関する公式

公式 1-1 平均値の変換公式

公式 1-1　平均値の変換公式

a, bを任意の定数とする。データをx_iとし、$y_i = ax_i + b$とする。$\bar{x}$をx_iの平均値とし、$\bar{y}$をy_iの平均値とすると、以下が成り立つ。

$$\bar{y} = a\bar{x} + b \tag{1-8}$$

データを加減乗除してから平均値をとっても、平均値に対して加減乗除しても、結果は変わりません。便利なので頻繁に登場します。

証明■■■

式変形して証明します。

$$\begin{aligned}
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\
&= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) \\
&= a\frac{1}{n}\sum_{i=1}^{n}(x_i) + \frac{1}{n}\sum_{i=1}^{n}(b) \\
&= a\bar{x} + \frac{1}{n}nb \\
&= a\bar{x} + b
\end{aligned} \tag{1-9}$$

以上により公式 1-1 が示されました。

■■■証明終わり

公式 1-2 平均値に関する計算公式①

公式 1-2　平均値に関する計算公式①

データをx_iとし。$\bar{x}$をx_iの平均値とする。以下が成り立つ。

$$\sum_{i=1}^{n}(x_i-\bar{x})=0 \tag{1-10}$$

平均値はデータの重心位置なのだと考えると、重心からの差分の合計値は、プラマイがゼロとなるのは自然ですね。自明として証明を略する教科書がほとんどですが、念のため述べておきます。公式 1-2 は明示せずに使うことがあります。

証明■■■

式変形して証明します。定義とΣ記号の性質から以下が成り立ちます。

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})&=\sum_{i=1}^{n}(x_i)-\sum_{i=1}^{n}(\bar{x})\\&=\sum_{i=1}^{n}(x_i)-n\bar{x}\end{aligned} \tag{1-11}$$

ここでやや技巧的ですが、右辺第一項に$n\cdot 1/n=1$をかけて整理します。

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})&=n\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i)-n\bar{x}\\&=n\bar{x}-n\bar{x}\\&=0\end{aligned} \tag{1-12}$$

以上により公式 1-2 が示されました。

■■■証明終わり

公式 1-3 平均値に関する計算公式②

公式 1-3　平均値に関する計算公式②

データをx_i, y_iとし。$\bar{x}$をx_iの平均値と、$\bar{y}$をy_iの平均値とする。以下が成り立つ。

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i \tag{1-13}$$

公式 1-3 の左辺は、共分散の分子です。共分散は実はと言うと$\bar{y}$を使わずに計算できます。どちらかというと技巧的な公式です。

証明■■■

式変形して証明します。

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n}[x_i y_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y}] \\ &= \sum_{i=1}^{n}[(x_i - \bar{x})y_i - (x_i - \bar{x})\bar{y}] \\ &= \sum_{i=1}^{n}[(x_i - \bar{x})y_i] - \sum_{i=1}^{n}[(x_i - \bar{x})\bar{y}] \\ &= \sum_{i=1}^{n}[(x_i - \bar{x})y_i] - \bar{y}\sum_{i=1}^{n}(x_i - \bar{x})\end{aligned} \tag{1-14}$$

ここで公式 1-2 より、第 2 項は 0 となります。

以上により公式 1-3 が示されました。

■■■証明終わり

2. 期待値に関する公式

公式 2-1　期待値の変換公式

> **公式 2-1　期待値の変換公式**
>
> a, bを任意の定数とする。確率変数をXとする。以下が成り立つ。
>
> $$E(aX + b) = aE(X) + b \tag{1-15}$$

平均値と同じく、何となく成り立ちそうな公式ですね。期待値の線形性と呼ぶこともあります。

証明■■■

関数$g(X) = aX + b$と置きます。すると$E(aX + b)$は$E[g(X)]$と表記できます。上記を利用し、離散型の確率分布の場合と、連続型の確率分布の場合で別々に式変形して示します。

○離散型の確率分布の場合

確率質量関数を$f(x_i)$とすると、期待値の定義から$E[g(X)]$は以下のように計算できます。期待値の定義式においてxの代わりに$g(x_i)$を置いただけです。

$$E[g(X)] = \sum_{i=1}^{n} f(x_i) \cdot g(x_i) \tag{1-16}$$

上記の結果を変形します。

$$\begin{aligned} E[g(X)] &= \sum_{i=1}^{n} f(x_i) \cdot g(x_i) \\ &= \sum_{i=1}^{n} f(x_i) \cdot (ax_i + b) \\ &= \sum_{i=1}^{n} [f(x_i) \cdot ax_i] + \sum_{i=1}^{n} [f(x_i) \cdot b] \\ &= a\sum_{i=1}^{n} [f(x_i) \cdot x_i] + b\sum_{i=1}^{n} [f(x_i)] \end{aligned} \tag{1-17}$$

ここで、期待値の定義より$\sum_{i=1}^{n} f(x_i) \cdot x_i = E(x)$です。また、確率質量関数の定義から$\sum_{i=1}^{n} f(x_i) = 1$です。よって、以下が成り立ちます。

$$E[g(X)] = aE(x) + b \tag{1-18}$$

以上により、離散型の確率分布において期待値の線形性が示されました。

○連続型の確率分布の場合

確率密度関数を$f(x)$とすると、期待値の定義から$E[g(X)]$は以下のように計算できます。期待値の定義式においてxの代わりに$g(x)$を置いただけです。

$$E[g(X)] = \int_{-\infty}^{\infty} f(x) \cdot g(x)\, dx \tag{1-19}$$

上記の結果を変形します。

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} f(x) \cdot g(x)\, dx \\ &= \int_{-\infty}^{\infty} f(x) \cdot (ax + b)\, dx \\ &= \int_{-\infty}^{\infty} f(x) \cdot ax\, dx + \int_{-\infty}^{\infty} f(x) \cdot b\, dx \\ &= a\int_{-\infty}^{\infty} f(x) \cdot x\, dx + b\int_{-\infty}^{\infty} f(x)\, dx \end{aligned} \tag{1-20}$$

ここで、期待値の定義より$\int_{-\infty}^{\infty} f(x) \cdot x\, dx = E(x)$です。また、確率密度関数の定義から$\int_{-\infty}^{\infty} f(x)\, dx = 1$です。よって、以下が成り立ちます。

$$\begin{aligned} E[g(X)] &= a\int_{-\infty}^{\infty} f(x) \cdot x\, dx + b\int_{-\infty}^{\infty} f(x)\, dx \\ &= aE(x) + b \end{aligned} \tag{1-21}$$

以上により、連続型の確率分布において期待値の線形性が示されました。

以上により、期待値の線形性が示されました。

■■■証明終わり

離散型の確率分布でも、連続型の確率分布でも、ほぼ同じように示せます。そのため、多くの教科書ではどちらか片方だけを示します。本資料でもそれにならい、今後はどちらかの証明を省略することがあります。

3. 標本分散に関する公式

公式 3-1　標本分散の計算公式

公式 3-1　標本分散の計算公式

データをx_iとする。$\bar{x}$をx_iの平均値と、s_x^2をx_iの標本分散とする。以下が成り立つ。

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left[x_i^2\right] - (\bar{x})^2 \tag{1-22}$$

計算を簡単にするためにしばしば登場する、技巧的な公式です。

証明■■■

式変形して示します。

$$\begin{aligned} s_x^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[x_i^2 - 2x_i\bar{x} + (\bar{x})^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[x_i^2\right] - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n}\left[x_i\right] + \frac{1}{n}n(\bar{x})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[x_i^2\right] - 2(\bar{x})^2 + (\bar{x})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[x_i^2\right] - (\bar{x})^2 \end{aligned} \tag{1-23}$$

以上により、公式 3-1 が示されました。

■■■証明終わり

公式 3-2　標本分散の変換公式

公式 3-2　分散の変換公式

a, bを任意の定数とする。データをx_iとし、$y_i = ax_i + b$とする。s_x^2をx_iの標本分散とし、s_y^2をy_iの標本分散とする。以下が成り立つ。

$$s_y^2 = a^2 s_x^2 \tag{1-24}$$

平均値と違って、bは無視されます。変化しない定数が加わっても、ばらつきは大きくなりませんので。

また、分散は 2 乗する計算が入るので、a^2倍されることに注意が必要です。

証明■■■

式変形して示します。分散の定義とy_iの定義から以下が示せます。

$$\begin{aligned} s_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\{(ax_i + b) - \bar{y}\}^2 \end{aligned} \tag{1-25}$$

ここで、公式 1-1 より$\bar{y} = a\bar{x} + b$であることを利用します。

$$\begin{aligned} s_y^2 &= \frac{1}{n}\sum_{i=1}^{n}\{(ax_i + b) - (a\bar{x} + b)\}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\{ax_i + b - a\bar{x} - b\}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\{a(x_i - \bar{x})\}^2 \\ &= a^2\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= a^2 s_x^2 \end{aligned} \tag{1-26}$$

以上により、公式 3-2 が示されました。

■■■証明終わり

4. 確率変数の分散に関する公式

公式 4-1　分散の計算公式

公式 4-1　分散の計算公式

確率変数をXとする。$E(X)$をXの平均値と、$V(X)$をXの分散とする。以下が成り立つ。

$$V(X) = E(X^2) - \{E(X)\}^2 \tag{1-27}$$

標本分散とほぼ同じ公式が 2 個続きます。

証明■■■

式変形して示します。分散の定義から$V(X) = E[(X-\mu)^2]$であることより以下が示せます。なお$\mu = E(X)$です。

$$\begin{aligned} V(X) &= E[(X-\mu)^2] \\ &= E[X^2 - 2X\mu + \mu^2] \end{aligned} \tag{1-28}$$

公式 2-1（期待値の変換公式）を利用します。

$$\begin{aligned} V(X) &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2E(X)E(X) + \{E(X)\}^2 \\ &= E(X^2) - \{E(X)\}^2 \end{aligned} \tag{1-29}$$

以上により、公式 4-1 が示されました。

■■■証明終わり

公式 4-2 分散の変換公式

公式 4-2　分散の変換公式

a, bを任意の定数とする。確率変数をXとし、$Y = aX + b$とする。$V(X)$をXの分散とし、$V(Y)$をYの分散とする。以下が成り立つ。

$$V(Y) = a^2 V(X) \tag{1-30}$$

証明■■■

式変形して示します。分散の定義より以下が示せます。

$$\begin{aligned} V(Y) &= E[\{Y - E(Y)\}^2] \\ &= E[\{(aX + b) - E(Y)\}^2] \end{aligned} \tag{1-31}$$

ここで公式 2-1 より、$E(Y) = E(aX + b) = aE(X) + b$であることから、以下のように変形できます。

$$\begin{aligned} V(Y) &= E[\{(aX + b) - (aE(X) + b)\}^2] \\ &= E[\{aX + b - aE(X) - b\}^2] \\ &= E\left[\left\{a\left(X - E(X)\right)\right\}^2\right] \\ &= a^2 \cdot E[\{X - E(X)\}^2] \\ &= a^2 V(X) \end{aligned} \tag{1-32}$$

以上により、公式 4-2 が示されました。

■■■証明終わり

5. 標本共分散に関する公式

公式 5-1 標本共分散の計算公式

公式 5-1　標本共分散の計算公式

データをx_i, y_iとする。$\bar{x}$をx_iの平均値と、$\bar{y}$をy_iの標本平均と、$Cov(x, y)$をx_i, y_iの共分散とする。以下が成り立つ。

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}[x_i y_i] - \bar{x}\bar{y} \tag{1-33}$$

計算を簡単にするための技巧的な公式です。

証明■■■

式変形して示します。

$$\begin{aligned} Cov(x, y) &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n}\sum_{i=1}^{n}[x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\bar{y}] \\ &= \frac{1}{n}\sum_{i=1}^{n}[x_i y_i] - \bar{x}\frac{1}{n}\sum_{i=1}^{n}[y_i] - \bar{y}\frac{1}{n}\sum_{i=1}^{n}[x_i] + \frac{1}{n}n\bar{x}\bar{y} \qquad (1\text{-}34) \\ &= \frac{1}{n}\sum_{i=1}^{n}[x_i y_i] - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} \\ &= \frac{1}{n}\sum_{i=1}^{n}[x_i y_i] - \bar{x}\bar{y} \end{aligned}$$

以上により、公式 5-1 が示されました。

■■■証明終わり

公式 5-2 標本共分散と標本分散の関係に関する公式

公式 5-2　標本共分散と標本分散の関係に関する公式

データをx_iとする。s_x^2をx_iの標本分散とする。以下が成り立つ。

$$Cov(x,x)=s_x^2 \tag{1-35}$$

共分散と分散の関係を見るうえで大切な公式です。証明の際にもしばしば登場します。

証明■■■

式変形して示します。

$$\begin{aligned} Cov(x,x) &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x}) \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ &= s_x^2 \end{aligned} \tag{1-36}$$

以上により、公式 5-2 が示されました。

■■■証明終わり

公式 5-3 標本共分散の変換公式

公式 5-3　標本共分散の変換公式

a, b, c, dを任意の定数とする。データをx_i, y_iとする。

$w_i = ax_i + b$と$z_i = cy_i + d$とする。$Cov(w, z)$をw_i, z_iの共分散とする。以下が成り立つ。

$$Cov(w, z) = ac \cdot Cov(x, y) \tag{1-37}$$

共分散は平均値で差し引く処理が入るので、定数b, dの影響は無視されます。

証明■■■

式変形して示します。

$$\begin{aligned} Cov(w, z) &= \frac{1}{n}\sum_{i=1}^{n}(w_i - \bar{w})(z_i - \bar{z}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{(ax_i + b) - \bar{w}\}\{(cy_i + d) - \bar{z}\} \end{aligned} \tag{1-38}$$

ここで、公式 1-1 より、$\bar{w} = a\bar{x} + b$であり、$\bar{z} = c\bar{y} + d$ですので、以下のように変形できます。

$$\begin{aligned} Cov(w, z) &= \frac{1}{n}\sum_{i=1}^{n}\{(ax_i + b) - (a\bar{x} + b)\}\{(cy_i + d) - (c\bar{y} + d)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b - a\bar{x} - b)(cy_i + d - a\bar{y} - d) \\ &= \frac{1}{n}\sum_{i=1}^{n}\{a(x_i - \bar{x})\}\{c(y_i - \bar{y})\} \\ &= ac\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ &= ac \cdot Cov(x, y) \end{aligned} \tag{1-39}$$

以上により、公式 5-3 が示されました。

■■■証明終わり

6. 確率変数の共分散に関する公式

公式 6-1 共分散の計算公式

公式 6-1　共分散の計算公式

確率変数をX, Yとする。$E(X)$をXの平均値と、$V(X)$をXの分散と、$C(X, Y)$をX, Yの共分散とする。以下が成り立つ。

$$C(X, Y) = E(XY) - E(X)E(Y) \tag{1-40}$$

標本共分散とほぼ同じ公式が 3 個続きます。

証明■■■

式変形して示します。なお$\mu_X = E(X)$であり、$\mu_Y = E(Y)$です。

$$\begin{aligned} C(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \end{aligned} \tag{1-41}$$

ここで、後ほど紹介する確率変数の和に関する公式 7-1 より、和の期待値は期待値の和となります。よって、以下のように変形できます。

$$\begin{aligned} C(X, Y) &= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned} \tag{1-42}$$

以上により、公式 6-1 が示されました。

■■■証明終わり

公式 6-2 共分散と分散の関係に関する公式

公式 6-2　共分散と分散の関係に関する公式

確率変数をXとする。$V(X)$をXの分散とする。以下が成り立つ。

$$C(X,X) = V(X) \tag{1-43}$$

証明■■■

式変形して示します。なお$\mu = E(X)$です。

$$\begin{aligned} C(X,X) &= E[(X-\mu)(X-\mu)] \\ &= E[(X-\mu)^2] \\ &= V(X) \end{aligned} \tag{1-44}$$

以上により、公式 6-2 が示されました。

■■■証明終わり

公式 6-3 共分散の変換公式

公式 6-3　共分散の変換公式

a, b, c, dを任意の定数とする。確率変数をX, Yとする。

$W = aX + b$と$Z = cY + d$とする。$C(W, Z)$をW, Zの共分散とする。以下が成り立つ。

$$C(W, Z) = ac \cdot C(X, Y) \tag{1-45}$$

証明■■■

式変形して示します。なお$\mu_W = E(W)$であり、$\mu_Z = E(Z)$です

$$\begin{aligned} C(W, Z) &= E[(W - \mu_W)(Z - \mu_Z)] \\ &= E[\{(aX + b) - \mu_W\}\{(cY + d) - \mu_Z\}] \end{aligned} \tag{1-46}$$

ここで、公式 2-1 より、$\mu_W = aE(X) + b$であり、$\mu_Z = cE(Y) + d$ですので、以下のように変形できます。

$$\begin{aligned} C(W, Z) &= E[\{(aX + b) - (aE(X) + b)\}\{(cY + d) - (cE(Y) + d)\}] \\ &= E[(aX + b - aE(X) - b)(cY + d - cE(Y) - d)] \\ &= E\Big[\big\{a\big(X - E(X)\big)\big\}\big\{c\big(Y - E(Y)\big)\big\}\Big] \\ &= ac \cdot E[\{X - E(X)\}\{Y - E(Y)\}] \\ &= ac \cdot C(X, Y) \end{aligned} \tag{1-47}$$

以上により、公式 6-3 が示されました。

■■■証明終わり

公式 6-4 独立な確率変数における共分散に関する公式

公式 6-4　独立な確率変数における共分散に関する公式

独立な確率変数をX,Yとする。以下が成り立つ。

$$C(X,Y)=0 \tag{1-48}$$

証明■■■

公式 6-1 と、後ほど紹介する公式 7-3「独立な確率変数の積の期待値に関する公式」を利用して、式変形して示します。

公式 6-1 より以下が成り立ちます。

$$C(X,Y)=E(XY)-E(X)E(Y) \tag{1-49}$$

ここで、公式 7-3 より、独立な確率変数X,Yにおいて$E(XY)=E(X)E(Y)$です。よって独立な確率変数X,Yにおいて$E(XY)-E(X)E(Y)=0$となります。

以上により、公式 6-4 が示されました。

■■■証明終わり

7. 確率変数の和と積に関する公式

公式 7-1 確率変数の和の期待値に関する公式

公式 7-1　確率変数の和の期待値に関する公式

確率変数をX, Yとする。以下が成り立つ。

$$E(X+Y)=E(X)+E(Y) \tag{1-50}$$

期待値の線形性と共に頻繁に登場する公式です。直観的に成り立ちそうに思えますが、同時確率分布の知識を要求されます。

証明■■■

離散型の確率分布を対象にして、式変形して示します。連続型についても同様に示せます。

確率変数Xの実現値をx_iと、Yの実現値をy_jとします。$E(X+Y)$は以下のように計算されます。

$$E(X+Y)=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_i+y_j)\cdot P(X=x_i, Y=y_j) \tag{1-51}$$

確率変数が 2 つになると、Σ記号が 2 つ重なることに注意してください。今回の例ではΣ記号の順番は入れ替え可能です。これを利用して右辺を 2 つに分けます。

$$\begin{aligned} E(X+Y)&=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_i+y_j)\cdot P(X=x_i, Y=y_j) \\ &=\sum_{i=1}^{m}\sum_{j=1}^{n}\{x_i\cdot P(X=x_i, Y=y_j)\}+\sum_{i=1}^{m}\sum_{j=1}^{n}\{y_j\cdot P(X=x_i, Y=y_j)\} \\ &=\sum_{i=1}^{m}x_i\sum_{j=1}^{n}\{P(X=x_i, Y=y_j)\}+\sum_{j=1}^{n}y_j\sum_{i=1}^{m}\{P(X=x_i, Y=y_j)\} \end{aligned} \tag{1-52}$$

ここで周辺化の公式を使って、同時確率分布を周辺分布にします。

$$E(X+Y)=\sum_{i=1}^{m}x_i\cdot P(X=x_i)+\sum_{j=1}^{n}y_j\cdot P\big(Y=y_j\big)$$

$$=E(X)+E(X) \tag{1-53}$$

以上により、公式 7-1 が示されました。

■■■証明終わり

一般的に、確率変数の列$X_1, X_2, \dots X_n$において以下が成り立ちます。

$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n) \tag{1-54}$$

仮に$X_1, X_2, \dots X_k$が、期待値μの独立で同一な確率分布に従うとするならば、$E(X_1+X_2+\cdots+X_n)= n\mu$となります。

公式 7-2 確率変数の和の分散に関する公式

公式 7-2　確率変数の和の分散に関する公式

確率変数をX, Yとする。以下が成り立つ。

$$V(X+Y) = V(X) + V(Y) + 2C(X, Y) \tag{1-55}$$

期待値と同じく$V(X+Y) = V(X) + V(Y)$と言いたいところなのですが、後ろに$2C(X, Y)$がくっつくことに注意が必要です。

証明する前にイメージ的な説明をします。共分散$C(X, Y)$が正の大きな値をとるということは「Xが大きな値をとると、Yもそれに引きずられて大きな値を取りやすい」ことになります。逆も同じで「Xが小さな値をとると、Yもそれに引きずられて小さな値を取りやすい」です。そのため、共分散が正の大きな値をとると「相手に引きずられて大きく動いてしまう」ため、単なる分散の和よりも大きくなります。

逆も同じですね。負の共分散を持つ場合は「自分の変動を打ち消すように相手が動く」ので分散の和は小さくなります。

証明■■■

式変形して示します。

μ_XをXの期待値と、μ_YをYの期待値とすると、公式 7-1 より、$E(X+Y)=\mu_X+\mu_Y$となります。これを利用すると、$X+Y$の分散$V(X+Y)$は以下のように計算できます。

$$V(X+Y)=E[\{(X+Y)-(\mu_X+\mu_Y)\}^2] \tag{1-56}$$

上記の結果を変形します。

$$\begin{aligned} V(X+Y) &= E[\{(X-\mu_X)+(Y-\mu_Y)\}^2] \\ &= E[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)] \\ &= E[(X-\mu_X)^2]+E[(Y-\mu_Y)^2]+2E[(X-\mu_X)(Y-\mu_Y)] \\ &= V(X)+V(Y)+2C(X,Y) \end{aligned} \tag{1-57}$$

以上により、公式 7-2 が示されました。

■■■証明終わり

公式 7-3 独立な確率変数の積の期待値に関する公式

公式 7-3　独立な確率変数の積の期待値に関する公式

独立な確率変数をX, Yとする。以下が成り立つ。

$$E(XY) = E(X)E(Y) \tag{1-58}$$

独立である場合には、「積の期待値$E(XY)$」は「期待値の積$E(X)E(Y)$」となります。独立であるといろいろな計算が楽になるということの一例と言えるかと思います。良くも悪くも、独立を仮定すると計算は楽です。本当にその仮定が満たされているかどうかは別として。

証明■■■

離散型の確率分布を対象にして、式変形して示します。連続型についても同様に示せます。$E(XY)$は以下のように定義されます。

$$E(XY) = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_i y_j) \cdot P(X = x_i, Y = y_j) \tag{1-59}$$

独立の定義から$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$が成り立つことを利用して、式を変形します。

$$\begin{aligned}
E(XY) &= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_i y_j) \cdot P(X = x_i) \cdot P(Y = y_j) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i \cdot P(X = x_i) \cdot y_j \cdot P(Y = y_j) \\
&= \left[\sum_{i=1}^{m} x_i \cdot P(X = x_i)\right] \cdot \left[\sum_{j=1}^{n} y_j \cdot P(Y = y_j)\right] \\
&= E(X)E(Y)
\end{aligned} \tag{1-60}$$

以上により、公式 7-3 が示されました。

■■■証明終わり

公式 7-4 独立な確率変数の和の分散に関する公式

公式 7-4　独立な確率変数の和の分散に関する公式

独立な確率変数をX, Yとする。以下が成り立つ。

$$V(X+Y) = V(X) + V(Y) \tag{1-61}$$

独立だと、公式 7-2 において後ろにくっついていた$2C(X, Y)$が無くなります。

証明■■■

公式 7-2 を再掲します。

$$V(X+Y) = V(X) + V(Y) + 2C(X, Y) \tag{1-62}$$

公式 6-4 より、独立な確率変数X, Yにおいて$C(X, Y) = 0$となります。よって独立な確率変数X, Yにおいて$V(X+Y) = V(X) + V(Y)$となります。

以上により、公式 7-4 が示されました。

■■■証明終わり

公式 7-5 確率変数の和の共分散に関する公式

公式 7-5　確率変数の和の共分散に関する公式

a,b,c,dを任意の定数とする。確率変数をX,Y,W,Zとする。以下が成り立つ。

$$C(aX+bY,cW+dZ) = ac\cdot C(X,W)+ad\cdot C(X,Z)+bc\cdot C(Y,W)+bd\cdot C(Y,Z) \tag{1-63}$$

入門書ではあまり見かけませんが､英語版の Wikipedia などでは解説があります[URL: https://en.wikipedia.org/wiki/Covariance]｡「共分散の双線形性」とも呼ばれるようです。

証明■■■

式変形して示します。

$$C(aX+bY, cW+dZ)$$

$$= E[\{(aX+bY)-E(aX+bY)\}\{(cW+dZ)-E(cW+dZ)\}] \tag{1-64}$$

公式 2-1 と公式 7-1 より、$E(aX+bY)=aE(X)+bE(Y)$であり、$E(cW+dZ)=cE(W)+dE(Z)$ですので、以下のように変形できます。

$$C(aX+bY, cW+dZ)$$

$$= E\Big[\Big\{(aX+bY)-\big(aE(X)+bE(Y)\big)\Big\}\Big\{(cW+dZ)-\big(cE(W)+dE(Z)\big)\Big\}\Big]$$

$$= E[\{aX+bY-aE(X)-bE(Y)\}\{cW+dZ-cE(W)+dE(Z)\}]$$

$$= E\Big[\Big\{a\big(X-E(X)\big)+b\big(Y-E(Y)\big)\Big\}\Big\{c\big(W-E(W)\big)+d\big(Z-E(Z)\big)\Big\}\Big]$$

$$= E\Big[a\big(X-E(X)\big)\cdot c\big(W-E(W)\big)+a\big(X-E(X)\big)\cdot d\big(Z-E(Z)\big)$$

$$+b\big(Y-E(Y)\big)\cdot c\big(W-E(W)\big) \tag{1-65}$$

$$+b\big(Y-E(Y)\big)\cdot d\big(Z-E(Z)\big)\Big]$$

$$= ac\cdot E\Big[\big(X-E(X)\big)\big(W-E(W)\big)\Big]+ad\cdot E\Big[\big(X-E(X)\big)\big(Z-E(Z)\big)\Big]$$

$$+bc\cdot E\Big[\big(Y-E(Y)\big)\big(W-E(W)\big)\Big]$$

$$+bd\cdot E\Big[\big(Y-E(Y)\big)\big(Z-E(Z)\big)\Big]$$

$$= ac\cdot C(X,W)+ad\cdot C(X,Z)+bc\cdot C(Y,W)+bd\cdot C(Y,Z)$$

以上により、公式 7-5 が示されました。

■■■証明終わり

公式 7-5 にはバリエーションがいくつかあるので紹介します。
まずは、確率変数の数が少ないバージョンです。

$$C(aX + bY, cW) = ac \cdot C(X, W) + \ bc \cdot C(Y, W) \tag{1-66}$$

続いて、確率変数が多いバージョンです。

$$C\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_j a_i \cdot C\left(X_i, Y_j\right) \tag{1-67}$$

代表的な確率分布とその特徴

　初等的な統計学の入門書で、ほぼ必ず登場する確率分布に絞って解説します。ただし、正規分布は少し難しいので省略します。

1. 一様分布（離散型）

実現値が$x_1, x_2, \dots, x_n$であり、起こり得る結果の種類数がnである一様分布の確率質量関数を$\mathrm{U}(X|n)$とします。一様分布の確率質量関数は以下の通りです。

$$\mathrm{U}(X|n) = \frac{1}{n} \tag{2-1}$$

<u>一様分布の確率質量関数の和が 1 になることの証明</u>

確率質量関数の和が 1 になることを示します。

証明■■■

式変形して示します。

$$\begin{aligned} \sum_{i=1}^{n} f(x_i) &= \sum_{i=1}^{n} \frac{1}{n} \\ &= 1 \end{aligned} \tag{2-2}$$

以上により、確率質量関数の和が 1 になることが示せました

■■■証明終わり

一様分布（離散型）の期待値

$X \sim \mathrm{U}(X|n)$である確率変数Xの期待値は以下の通りです。

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2-3}$$

証明■■■

式変形して示します。

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} f(x_i) \cdot x_i \\ &= \sum_{i=1}^{n} \frac{1}{n} \cdot x_i \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned} \tag{2-4}$$

■■■証明終わり

一様分布（離散型）の分散

$X\sim \mathrm{U}(X|n)$である確率変数Xの分散は以下の通りです。

$$V(X)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{2-5}$$

証明■■■

式変形して示します。

$$\begin{aligned} V(X)&=\sum_{i=1}^{n}f(x_i)\cdot(x_i-\mu)^2 \\ &=\sum_{i=1}^{n}\frac{1}{n}\cdot(x_i-\mu)^2 \\ &=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned} \tag{2-6}$$

■■■証明終わり

2. 一様分布（連続型）

確率変数Xがとりうる範囲がα以上β以下である一様分布の確率密度関数を$\mathrm{U}(X|\alpha,\beta)$とします。一様分布の確率密度関数は以下の通りです。

$$\mathrm{U}(X|\alpha,\beta) = \frac{1}{\beta - \alpha} \tag{2-7}$$

一様分布の確率密度関数の積分値が 1 になることの証明

確率密度関数の積分値が 1 になることを示します。

証明■■■

式変形して示します。

$$\begin{aligned}
\int_{\alpha}^{\beta} f(x)\,dx &= \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha}\,dx \\
&= \left[\frac{1}{\beta - \alpha} \cdot x\right]_{\alpha}^{\beta} \\
&= \frac{1}{\beta - \alpha} \cdot \beta - \frac{1}{\beta - \alpha} \cdot \alpha \\
&= \frac{\beta - \alpha}{\beta - \alpha} \\
&= 1
\end{aligned} \tag{2-8}$$

以上により、確率密度関数の積分値が 1 になることが示せました

■■■証明終わり

一様分布（連続型）の期待値

$X \sim \mathrm{U}(X|\alpha,\beta)$である確率変数$X$の期待値$E(X)$は以下の通りです。

$$E(X) = \frac{\beta+\alpha}{2} \tag{2-9}$$

証明■■■

式変形して示します。

$$\begin{aligned} E(X) &= \int_{\alpha}^{\beta} f(x) \cdot x \, dx \\ &= \int_{\alpha}^{\beta} \frac{1}{\beta-\alpha} \cdot x \, dx \\ &= \left[\frac{1}{\beta-\alpha} \cdot \frac{x^2}{2}\right]_{\alpha}^{\beta} \\ &= \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} \\ &= \frac{(\beta-\alpha)(\beta+\alpha)}{2(\beta-\alpha)} \\ &= \frac{\beta+\alpha}{2} \end{aligned} \tag{2-10}$$

■■■証明終わり

一様分布（連続型）の分散

$X \sim \mathrm{U}(X|\alpha, \beta)$である確率変数$X$の分散$V(X)$は以下の通りです。

$$V(X) = \frac{(\beta - \alpha)^2}{12} \tag{2-11}$$

証明■■■

$E(X) = (\beta + \alpha)/2$であることと、公式 4-1 を利用します。

$$\begin{aligned} V(X) &= E(X^2) - \{E(X)\}^2 \\ &= \int_{-\infty}^{\infty} \frac{1}{\beta - \alpha} \cdot x^2 \, dx - \left(\frac{\beta + \alpha}{2}\right)^2 \\ &= \left[\frac{1}{\beta - \alpha} \cdot \frac{x^3}{3}\right]_{\alpha}^{\beta} - \left(\frac{\beta + \alpha}{2}\right)^2 \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} - \frac{\alpha^2 + 2\alpha\beta + \beta^2}{4} \end{aligned} \tag{2-12}$$

ここで因数分解の公式から$\beta^3-\alpha^3=(\beta-\alpha)(\alpha^2+\alpha\beta+\beta^2)$であることを利用してさらに式変形します。

$$
\begin{aligned}
V(X) &= \frac{\beta^3-\alpha^3}{3(\beta-\alpha)}-\frac{\alpha^2+2\alpha\beta+\beta^2}{4} \\
&= \frac{(\beta-\alpha)(\alpha^2+\alpha\beta+\beta^2)}{3(\beta-\alpha)}-\frac{\alpha^2+2\alpha\beta+\beta^2}{4} \\
&= \frac{\alpha^2+\alpha\beta+\beta^2}{3}-\frac{\alpha^2+2\alpha\beta+\beta^2}{4} \\
&= \frac{4(\alpha^2+\alpha\beta+\beta^2)-3(\alpha^2+2\alpha\beta+\beta^2)}{12} \\
&= \frac{\alpha^2-2\alpha\beta+\beta^2}{12} \\
&= \frac{(\beta-\alpha)^2}{12}
\end{aligned}
\tag{2-13}
$$

■■■証明終わり

というわけで連続型の一様分布においては$V(X)=(\beta-\alpha)^2/12$となりました。微妙に大変な計算だったと思います。

今回の計算過程を覚える必要はありません。ましてや連続型の一様分布の分散の結果を暗記しても統計学の理解は深まらないと思います。

ただし、定義通りに計算すれば、さまざまな確率変数の期待値や分散を計算できるな、という実感は持ってもらえると嬉しいです。

以下、いろんな確率分布に従う確率変数の期待値と分散を計算しますが、「まぁそんなもんかな～」で最初はいいと思います。必要になった時に参照しましょう。

3. 二項分布

成功確率p、試行回数をnとした二項分布の確率質量関数を$\mathrm{Bin}(X|n,p)$とします。二項分布の確率質量関数は以下の通りです。

$$\mathrm{Bin}(X|n,p) = {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \tag{2-14}$$

離散型の確率変数ではx_iと添え字iを使うことが多かったです。ところで二項分布の場合、成功回数xは 0 以上で試行回数n以下であるため$x = 0,1,2,\dots,n$となります。

これからは添え字iを省略し、確率質量関数として$f(x) = {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x}$と表記します。

二項分布を理解するために必要な公式

階乗

階乗は以下の通り計算します。

$$x! = x \cdot (x-1) \cdot (x-2) \cdot \ldots \cdot 2 \cdot 1 \tag{2-15}$$

順列組み合わせの公式

順列組合せの公式は下記の通りです。

$$ {}_n\mathrm{C}_x = \frac{n!}{x! \cdot (n-x)!} \tag{2-16}$$

二項定理

二項分布の名前の由来にもなった二項定理は下記の通りです。

$$(p+q)^n = \sum_{x=0}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot q^{n-x} \tag{2-17}$$

例えば$n = 2$なら、馴染み深い下記の結果が得られます。

$$\begin{aligned} (p+q)^2 &= \sum_{x=0}^{2} {}_2\mathrm{C}_x \cdot p^x \cdot q^{2-x} \\ &= {}_2\mathrm{C}_0 \cdot p^0 \cdot q^{2-0} + {}_2\mathrm{C}_1 \cdot p^1 \cdot q^{2-1} + {}_2\mathrm{C}_2 \cdot p^2 \cdot q^{2-2} \\ &= q^2 + 2pq + p^2 \end{aligned} \tag{2-18}$$

二項分布の確率質量関数の和が 1 になることの証明

確率質量関数の和が 1 になることを示します。

証明■■■

二項定理において$q = 1 - p$を代入すると、二項定理の右辺が、二項分布の確率質量関数に等しいことがわかります。

よって、以下が成り立ちます。

$$\begin{aligned}\sum_{x=0}^{n} f(x) &= \sum_{x=0}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \\ &= \{p + (1-p)\}^n \\ &= 1^n \\ &= 1\end{aligned} \tag{2-19}$$

以上により、確率質量関数の和が 1 になることが示せました

■■■証明終わり

二項分布の期待値

$X \sim \mathrm{Bin}(X|n,p)$である確率変数$X$の期待値$E(X)$は下記の通りです。

$$E(X) = np \tag{2-20}$$

証明■■■

式変形して示します。

期待値の定義から以下が成り立ちます。

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} f(x) \cdot x \\ &= \sum_{x=0}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \cdot x \end{aligned} \tag{2-21}$$

また、0 に何をかけても 0 なので、上記の計算から$x = 0$のときは無視できます。

$$E(X) = \sum_{x=1}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \cdot x \tag{2-22}$$

ところで、階乗の定義から以下が成り立ちます。

$$a! = a \cdot (a-1)! \tag{2-23}$$

よって$n!$と$x!$を変形することで$x \cdot {}_n\mathrm{C}_x$は以下のように変形できます。

$$\begin{aligned} x \cdot {}_n\mathrm{C}_x &= x \cdot \frac{n!}{x! \cdot (n-x)!} \\ &= x \cdot \frac{n \cdot (n-1)!}{x \cdot (x-1)! \cdot (n-x)!} \\ &= \frac{n \cdot (n-1)!}{(x-1)! \cdot (n-x)!} \\ &= n \cdot {}_{n-1}\mathrm{C}_{x-1} \end{aligned} \tag{2-24}$$

この結果を式(2-22)に代入します。

$$
\begin{aligned}
E(X) &= \sum_{x=1}^{n} n \cdot {}_{n-1}\mathrm{C}_{x-1} \cdot p^{x} \cdot (1-p)^{n-x} \\
&= n \cdot \sum_{x=1}^{n} {}_{n-1}\mathrm{C}_{x-1} \cdot p^{x} \cdot (1-p)^{n-x}
\end{aligned}
\tag{2-25}
$$

ここで$p^x = p^{x-1} \cdot p$より以下が成り立ちます。

$$
\begin{aligned}
E(X) &= n \cdot \sum_{x=1}^{n} {}_{n-1}\mathrm{C}_{x-1} \cdot p^{x-1} \cdot p \cdot (1-p)^{n-x} \\
&= np \cdot \sum_{x=1}^{n} {}_{n-1}\mathrm{C}_{x-1} \cdot p^{x-1} \cdot (1-p)^{n-x}
\end{aligned}
\tag{2-26}
$$

ここで$y = x-1, m = n-1$と置くと、以下のように変形されます。

$$
E(X) = np \cdot \sum_{y=0}^{m} {}_{m}\mathrm{C}_{y} \cdot p^{y} \cdot (1-p)^{m-y} \tag{2-27}
$$

ここで、二項分布の確率質量関数がΣ記号の中に登場したことに気が付くでしょうか。二項分布の確率質量関数は$\mathrm{Bin}(X|n,p) = {}_{n}\mathrm{C}_{x} \cdot p^{x} \cdot (1-p)^{n-x}$です。$x$の代わりに$y$を、$n$の代わりに$m$を代入すれば、二項分布そのものですね。よって、確率質量関数の和は 1 になることから、Σ記号の中身は 1 になります。

よって以下が示せました。

$$
E(X) = np \tag{2-28}
$$

■■■証明終わり

二項分布の分散

$X \sim \mathrm{Bin}(X|n,p)$である確率変数$X$の分散$V(X)$は下記の通りです。

$$V(X) = np(1-p) \tag{2-29}$$

証明■■■

式変形して示します。公式 4-1 を利用します。ここで、技巧的ですが、期待値の線形性より$E(X^2 - X) = E(X^2) - E(X)$を利用して、分散を以下のように変形します。

$$\begin{aligned} V(X) &= E(X^2) - \{E(X)\}^2 \\ &= E[X(X-1)] + E(X) - \{E(X)\}^2 \end{aligned} \tag{2-30}$$

ここで、二項分布における$E[X(X-1)]$を計算します。**以下の計算の流れは期待値の計算とほぼ同じです。**多くの教科書では（面倒くさいから）省略されますが、ここではすべて示します。

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^{n} f(x) \cdot x \cdot (x-1) \\ &= \sum_{x=0}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \cdot x \cdot (x-1) \end{aligned} \tag{2-31}$$

ここで、$x = 0$のときと$x = 1$のときは、xまたは$(x-1)$が 0 になるので、Σ記号の中身が 0 になることに注意してください。よって、xは 2 以降だけを考えます。

$$E[X(X-1)] = \sum_{x=2}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \cdot x \cdot (x-1) \tag{2-32}$$

階乗の定義から以下が成り立ちます。

$$a! = a \cdot (a-1) \cdot (a-2)! \tag{2-33}$$

よって$n!$と$x!$を変形することで$x \cdot (x-1) \cdot {}_n\mathrm{C}_x$は以下のように変形できます。

$$
\begin{aligned}
x \cdot (x-1) \cdot {}_n\mathrm{C}_x &= x \cdot (x-1) \cdot \frac{n!}{x! \cdot (n-x)!} \\
&= x \cdot (x-1) \cdot \frac{n \cdot (n-1) \cdot (n-2)!}{x \cdot (x-1) \cdot (x-2)! \cdot (n-x)!} \\
&= \frac{n \cdot (n-1) \cdot (n-2)!}{(x-2)! \cdot (n-x)!} \\
&= n \cdot (n-1) \cdot {}_{n-2}\mathrm{C}_{x-2}
\end{aligned}
\tag{2-34}
$$

この結果を代入します。

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=2}^{n} {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \cdot x \cdot (x-1) \\
&= n \cdot (n-1) \sum_{x=2}^{n} {}_{n-2}\mathrm{C}_{x-2} \cdot p^x \cdot (1-p)^{n-x}
\end{aligned}
\tag{2-35}
$$

ここで$p^x = p^{x-2} \cdot p^2$より以下が成り立ちます。

$$
\begin{aligned}
E[X(X-1)] &= n \cdot (n-1) \cdot \sum_{x=2}^{n} {}_{n-2}\mathrm{C}_{x-2} \cdot p^{x-2} \cdot p^2 \cdot (1-p)^{n-x} \\
&= n \cdot (n-1) \cdot p^2 \cdot \sum_{x=2}^{n} {}_{n-2}\mathrm{C}_{x-2} \cdot p^{x-2} \cdot (1-p)^{n-x}
\end{aligned}
\tag{2-36}
$$

ここで$y = x-2, m = n-2$と置くと、以下のように変形されます。

$$
E[X(X-1)] = n \cdot (n-1) \cdot p^2 \cdot \sum_{y=0}^{m} {}_m\mathrm{C}_y \cdot p^y \cdot (1-p)^{m-y} \tag{2-37}
$$

Σ記号の中身が二項分布の確率質量関数となりましたので、Σ記号の中身は1になります。

よって以下が示せました。

$$E[X(X-1)] = n \cdot (n-1) \cdot p^2 \tag{2-38}$$

よって分散は以下のように計算されます。

$$\begin{aligned} V(X) &= E[X(X-1)] + E(X) - \{E(X)\}^2 \\ &= n \cdot (n-1) \cdot p^2 + np - (np)^2 \\ &= (np)^2 - np^2 + np - (np)^2 \\ &= np(1-p) \end{aligned} \tag{2-39}$$

■■■証明終わり

4. ポアソン分布

強度をλとしたポアソン分布の確率質量関数を$\mathrm{Pois}(x|\lambda)$とします。ポアソン分布の確率質量関数は以下の通りです。

$$\mathrm{Pois}(x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{2-40}$$

$x = 0,1,2,\ldots$であり、上限が無いのが二項分布とは違いますね。

ポアソン分布を理解するために必要な公式

ネイピア数

$$e = \lim_{t \to 0}(1+t)^{\frac{1}{t}} \tag{2-40}$$

ポアソン分布の導出

ポアソン分布は、$p \to 0, n \to \infty$という条件下の二項分布で$np = \lambda$とした結果だとみなすことができます。日本語で書くと「成功確率が限りなく 0 に近いが、試行回数が限りなく大きい二項分布」です。

証明■■■

二項分布の確率質量関数を、式変形して示します。階乗の定義を利用します。分子の$n!$と分母の$(n-x)!$の共通部分を約分するイメージです。

$$
\begin{aligned}
\mathrm{Bin}(X|n,p) &= {}_n\mathrm{C}_x \cdot p^x \cdot (1-p)^{n-x} \\
&= \frac{n!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x} \\
&= \frac{(n-x+1)(n-x+2)\dots(n-1)n}{x!} \cdot p^x \cdot (1-p)^{n-x}
\end{aligned}
\tag{2-41}
$$

$(n-x+1)(n-x+2)\dots(n-1)n$は、$x$回の掛け算が行われていることに注意してください。

ここでポアソン分布の仮定から$np = \lambda$より$p = \lambda/n$を代入して、階乗の部分を整理します。

$$
\begin{aligned}
\mathrm{Bin}(X|n,p) &= \frac{(n-x+1)(n-x+2)\dots(n-1)n}{x!} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{(n-x+1)(n-x+2)\dots(n-1)n}{x!} \cdot \frac{1}{n^x} \cdot \lambda^x \cdot \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{\left(1-\frac{x-1}{n}\right)\left(1-\frac{x-2}{n}\right)\dots\left(1-\frac{1}{n}\right)1}{x!} \cdot \lambda^x \cdot \left(1-\frac{\lambda}{n}\right)^{n-x}
\end{aligned}
\tag{2-42}
$$

上記の結果は、見づらいので以下をまとめてAと表記します。

$$
A = \left(1-\frac{x-1}{n}\right)\left(1-\frac{x-2}{n}\right)\dots\left(1-\frac{1}{n}\right)1 \tag{2-43}
$$

ここで、以下が成り立つことに注意してください。

$$\lim_{n\to\infty}\left(1-\frac{x-1}{n}\right)\left(1-\frac{x-2}{n}\right)\dots\left(1-\frac{1}{n}\right)1=\lim_{n\to\infty}A=1 \tag{2-44}$$

続いて$(1-\lambda/n)^{n-x}$を変形します。$a^{b+c}=a^b a^c$であり$a^{bc}=(a^b)^c$であることに注意します。

$$\begin{aligned}
\mathrm{Bin}(X|n,p)&=\frac{A}{x!}\cdot\lambda^x\cdot\left(1-\frac{\lambda}{n}\right)^{n-x}\\
&=\frac{A}{x!}\cdot\lambda^x\cdot\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}\\
&=\frac{A}{x!}\cdot\lambda^x\cdot\left\{\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right\}^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-x}\\
&\xrightarrow[n\to\infty]{}\frac{1}{x!}\cdot\lambda^x\cdot e^{-\lambda}\cdot(1-0)^{-x}\\
&=\frac{e^{-\lambda}\lambda^x}{x!}
\end{aligned} \tag{2-45}$$

なお、4 行目の計算には以下を利用します。すなわち$t=-\lambda/n$としたとき、$n\to\infty$で$t\to 0$であるので

$$\lim_{t\to 0}(1+t)^{\frac{1}{t}}=e \tag{2-46}$$

以上により、ポアソン分布が導出できました。

■■■証明終わり

ポアソン分布の確率質量関数の和が 1 になることの証明

確率質量関数の和が 1 になることを示します。

確率変数は 0 以上の整数をとるので、以下を示すことになります。

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = 1 \tag{2-47}$$

証明■■■

式変形して示します。まずはポアソン分布の和を変形します。

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!} \tag{2-48}$$

ここでマクローリン展開を利用します。マクローリン展開は以下を指します。ただし$f^{(k)}$は関数$f(x)$のk階微分です。

$$f(x) = \sum_{k=0}^{\infty} f^{(k)}(0)\frac{x^{k}}{k!} \tag{2-49}$$

なお、マクローリン展開は高校数学で勉強しないので、後ほど補足します。

ここで、$f(x) = e^{x}$としたとき、e^{x}はどれだけ微分しても常にe^{x}なので、以下が成り立ちます。

$$e^{x} = \sum_{k=0}^{\infty} e^{0}\frac{x^{k}}{k!} = \sum_{k=0}^{\infty} \frac{x^{k}}{k!} \tag{2-50}$$

よってポアソン分布の和は以下のように計算されます。

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda}e^{\lambda} = 1 \tag{2-51}$$

以上により、確率質量関数の和が 1 になることが示せました。

■■■証明終わり

マクローリン展開の初歩

マクローリン展開はテイラー展開の特殊な形です。テイラー展開についての証明付きの正確な説明は、例えば「中井悦司. (2018). 技術者のための基礎解析学 機械学習に必要な数学を本気で学ぶ. 翔泳社」などを参照してください。ここでは直観に訴えた簡単な解説にとどめます。

マクローリン展開は、関数を変換（あるいは近似）する別の関数を作るのに、とても便利です。

例えば関数$f(x)$が、扱いの難しい関数だったとします。けれども、この関数の微分が簡単にできるなら、マクローリン展開を使って「そのままだと扱いが難しいが、扱いが簡単な関数に変換（あるいは近似）」できるかもしれません。e^xなどはその最たるものと言えます。ややこしいe^xという代物を単なる足し算に変換することで、ポアソン分布の証明が簡単にできました。

ここでは説明の簡単のため、「扱いが簡単な関数」を対象にしてマクローリン展開について紹介しますが、本来の使い道を忘れないようにしてください。

ここからマクローリン展開の本論に移ります。

テイラー展開でもマクローリン展開でもそうですが、2 階微分以降はやや難しくなります。1 階微分までなら解釈は難しくありません。

kを∞まで大きくするのではなく、ちょっと小さめの 1 までで止めておくことにします。この場合のマクローリン展開は以下の通りです。

$$
\begin{aligned}
f(x) &= \sum_{k=0}^{1} f^{(k)}(0)\frac{x^k}{k!} \\
&= f(0)\frac{x^0}{0!} + f^{(1)}(0)\frac{x^1}{1!} \\
&= f(0) + f^{(1)}(0)x
\end{aligned}
\tag{2-52}
$$

ここで$f(x) = ax + b$という簡単な関数を置いて考えます。

ここで、意地悪な先生がいて、元の関数$f(x)$を教えてもらえないのだとします。けれ

ども$f(0) = b$である、そして関数を$x = 0$で微分した結果が$f^{(1)}(0) = a$であることは教えてくれました。

このとき、関数が一次関数であれば、$f(0)$の値が切片であり、$f^{(1)}(0)$の値が傾きであることがわかりますね。そのため傾きと切片から元の関数$f(x)$を復元できるわけです。

マクローリン展開でいうと以下のようになります。

$$
\begin{aligned}
f(x) &= f(0) + f^{(1)}(0)x \\
&= b + ax
\end{aligned}
\tag{2-53}
$$

基本的にはこれだけの話です。

なお、1 次関数だとこれで終了なのですが、2 次関数が相手だと、1 階微分の結果だけを使うのが心許なくなります。

例えば$f(x) = ax^2 + b$だと$f(0) = b$です。また$f^{(1)}(x) = 2ax$より$f^{(1)}(0) = 0$です。今回の$f(x)$だと、傾きを 0 だと評価してしまいます。二次関数のように傾きがxの値によってコロコロ変わるという場合には、2 階微分した結果も取り入れる必要があります。

2 階微分した結果は$f^{(2)}(x) = 2a$より$f^{(2)}(0) = 2a$です。よって、マクローリン展開をすれば、微分係数から元の関数が復元できるのがわかります。

$$
\begin{aligned}
f(x) &= \sum_{k=0}^{2} f^{(k)}(0)\frac{x^k}{k!} \\
&= f(0)\frac{x^0}{0!} + f^{(1)}(0)\frac{x^1}{1!} + f^{(2)}(0)\frac{x^2}{2!} \\
&= f(0) + f^{(1)}(0)x + f^{(2)}(0)\frac{x^2}{2} \\
&= b + 0 + 2a\frac{x^2}{2} \\
&= b + ax^2
\end{aligned}
\tag{2-54}
$$

グネグネと曲がった複雑な関数を変換するためには、微分する階数を増やして対応することになります。それでマクローリン展開の公式が登場します。

ところで、微分する位置を$x = 0$にこだわる必要は無いですね。$x = a$に一般化したものがテイラー展開と呼ばれます。大体同じような公式なので略します。

ちなみに、マクローリン展開の公式でx^kが出てきているのは大事なポイントです。$|x| < 1$ならば、x^kはkが大きくなると 0 に近づくので無視できます。そのため、$|x|$が 0 に近い場合は、2 階や 3 階程度の微分で、それなりに関数をうまく近似できることがあります。

難しい関数をそのまま相手にせずに、簡単な関数に変換（あるいは近似）してから作業するというのはよく登場するテクニックなので、名前くらいは憶えておいてください。マクローリン展開とその一般系であるテイラー展開のお話でした。

ポアソン分布の期待値

※　以下の計算は二項分布の期待値と分散の計算とよく似ているので、比較すると勉強になると思います。

$X\sim\mathrm{Pois}(X|\lambda)$である確率変数$X$の期待値$E(X)$は以下の通りです。

$$E(X)=\lambda \tag{2-40}$$

証明■■■

式変形して示します。期待値の定義から以下が成り立ちます。

$$\begin{aligned} E(X)&=\sum_{x=0}^{\infty}f(x)\cdot x\\ &=\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^{x}}{x!}\cdot x \end{aligned} \tag{2-41}$$

また、0 に何をかけても 0 なので、上記の計算から$x=0$のときは無視できます。

$$E(X)=\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^{x}}{x!}\cdot x \tag{2-42}$$

ところで、階乗の定義から以下が成り立ちます。

$$a!=a\cdot(a-1)! \tag{2-43}$$

また、$\lambda^{x}=\lambda^{x-1}\cdot\lambda$より以下が成り立ちます

$$\begin{aligned} E(X)&=\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^{x-1}\cdot\lambda}{x\cdot(x-1)!}\cdot x\\ &=\lambda\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \end{aligned} \tag{2-44}$$

ここで$y = x - 1$と置くと、以下のように変形されます。

$$E(X) = \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} \tag{2-45}$$

ここで、ポアソン分布の確率質量関数がΣ記号の中に登場したことに気が付くでしょうか。ポアソン分布の確率質量関数において、xの代わりにyを代入すれば、ポアソン分布そのものですね。よって、確率質量関数の和は 1 になることから、Σ記号の中身は 1 になります。

よって以下が示せました。

$$E(X) = \lambda \tag{2-46}$$

■■■証明終わり

ポアソン分布の分散

$X \sim \mathrm{Pois}(X|\lambda)$である確率変数$X$の分散$V(X)$は以下の通りです。

$$V(X) = \lambda \tag{2-40}$$

証明■■■

式変形して示します。

技巧的ですが、期待値の線形性より$E(X^2 - X) = E(X^2) - E(X)$を利用して、分散を以下のように変形します。

$$\begin{aligned} V(X) &= E(X^2) - \{E(X)\}^2 \\ &= E[X(X-1)] + E(X) - \{E(X)\}^2 \end{aligned} \tag{2-41}$$

ここで、ポアソン分布における$E[X(X-1)]$を計算します。**以下の計算の流れは期待値の計算と同じ（さらに言うと、二項分布の計算と同じ）です**。

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^{\infty} f(x) \cdot x \cdot (x-1) \\ &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} \cdot x \cdot (x-1) \end{aligned} \tag{2-42}$$

ここで、$x = 0$のときと$x = 1$のときは、xまたは$(x-1)$が 0 になるので、Σ記号の中身が 0 になることに注意してください。よって、xは 2 以降だけを考えます。

$$E[X(X-1)] = \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} \cdot x \cdot (x-1) \tag{2-43}$$

階乗の定義から以下が成り立ちます。

$$a! = a \cdot (a-1) \cdot (a-2)! \tag{2-44}$$

また、$\lambda^x = \lambda^{x-2} \cdot \lambda^2$より以下が成り立ちます

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2} \cdot \lambda^2}{x \cdot (x-1) \cdot (x-2)!} \cdot x \cdot (x-1) \\
&= \lambda^2 \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!}
\end{aligned}
\tag{2-45}
$$

ここで$y = x - 2$と置くと、以下のように変形されます。

$$
E[X(X-1)] = \lambda^2 \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} \tag{2-46}
$$

ここで、ポアソン分布の確率質量関数がΣ記号の中に登場したので、Σ記号の中身は 1 になります。よって$E[X(X-1)] = \lambda^2$です。

よって分散は以下のように計算されます。

$$
\begin{aligned}
V(X) &= E[X(X-1)] + E(X) - \{E(X)\}^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}
\tag{2-47}
$$

よって$V(X) = \lambda$と示せました。

■■■証明終わり

統計的推定

　標本から計算される統計量に関する計算を紹介します。扱う内容が広いですが、初等的な教科書の多くが扱う内容だけに絞ります。

1. 標本平均の特性

サンプルサイズがnである標本を、確率変数$X_1, X_2, \dots, X_n$とします。標本は期待値μ、分散σ^2の独立で同一な確率分布に従うと仮定します。標本平均は以下のように計算されます。

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3-1}$$

標本が確率変数なので、標本平均$\bar{X}$も確率変数です。

なお、独立で同一な確率分布に従うことを$X \sim i.i.d$とも表記しますが、i.i.d に従うことを以下では仮定していることに注意してください。

<u>標本平均の期待値</u>

標本$X_1, X_2, \dots, X_n$は期待値μ、分散σ^2の独立で同一な確率分布に従うと仮定します。標本平均の期待値$E(\bar{X})$は以下の通りです。

$$E(\bar{X}) = \mu \tag{3-2}$$

証明■■■

公式 7-1 より、以下が成り立ちます。

$$E(X_1 + X_2 + \cdots + X_n) = E\left(\sum\nolimits_{i=1}^{n} X_n\right) = n\mu \tag{3-3}$$

公式 2-1 より以下が成り立ちます。

$$E(aX) = a \cdot E(X) \tag{3-4}$$

よって$\bar{X}$の期待値は以下のように示せます。

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n} n\mu \\ &= \mu \end{aligned} \tag{3-5}$$

■■■証明終わり

標本平均の分散

標本$X_1, X_2, \dots, X_n$は期待値μ、分散σ^2の独立で同一な確率分布に従うと仮定します。標本平均の分散$V(\bar{X})$は以下の通りです。

$$V(\bar{X}) = \frac{\sigma^2}{n} \tag{3-6}$$

証明■■■

公式 7-4 より以下が成り立ちます。

$$\begin{aligned} V(X_1 + X_2 + \cdots + X_n) &= V\left(\sum\nolimits_{i=1}^{n} X_n\right) \\ &= n\sigma^2 \end{aligned} \tag{3-7}$$

公式 4-2 より以下が成り立ちます。

$$V(aX) = a^2 \cdot V(X) \tag{3-8}$$

よって$\bar{X}$の分散は以下のように示せます。

$$\begin{aligned} V(\bar{X}) &= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2} n\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned} \tag{3-9}$$

■■■証明終わり

2. 標本分散の特性

サンプルサイズがnである標本を、確率変数$X_1, X_2, \ldots, X_n$とします。標本は期待値μ、分散σ^2の独立で同一な確率分布に従うと仮定します。標本分散S^2と不偏分散U^2は以下のように計算されます。

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{3-10}$$

$$U^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}) \tag{3-11}$$

標本が確率変数なので、標本分散S^2も不偏分散U^2も確率変数です。

なお、独立で同一な確率分布に従うことを$X \sim i.i.d$とも表記しますが、i.i.d に従うことを以下では仮定していることに注意してください。

<u>標本分散の期待値</u>

標本$X_1, X_2, \dots, X_n$は期待値μ、分散σ^2の独立で同一な確率分布に従うと仮定します。標本分散S^2の期待値$E(S^2)$は以下の通りです。

$$E(S^2) = \frac{n-1}{n}\sigma^2 \tag{3-12}$$

一方、不偏分散U^2の期待値$E(U^2)$は以下の通りです。

$$E(U^2) = \sigma^2 \tag{3-13}$$

標本分散は母分散と比べると小さくなるバイアスがあり、不偏分散はそのバイアスがありません。

証明■■■

標本分散を計算する際、母平均μではなく、標本平均$\bar{X}$を使っているというのが大きなポイントです。そこで「確率変数Xとその標本平均$\bar{X}$の差異$(X_i - \bar{X})^2$」を「母平均からの正しい差異$(X_i - \mu)^2$」と、「標本平均から母平均の差異$(\bar{X} - \mu)^2$」の 2 つに分解します。

期待値の定義と標本分散の定義から以下が成り立ちます。

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\ &= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \end{aligned} \tag{3-14}$$

ここで、母平均μと標本平均$\bar{X}$の違いが分かるように、以下の恒等式を利用します。右辺に$\mu - \mu = 0$を加えただけです。

$$\begin{aligned} X_i - \bar{X} &= X_i - \bar{X} + \mu - \mu \\ &= (X_i - \mu) - (\bar{X} - \mu) \end{aligned} \tag{3-15}$$

ここで$X_i-\mu$が母平均との正しい差異であり、$\bar{X}-\mu$が標本平均を母平均の代わりに使ったことがもたらす差異となります。

偏差平方和$\sum_{i=1}^{n}(X_i-\bar{X})^2$を、上記の恒等式を利用して展開します。

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\bar{X})^2 &= \sum_{i=1}^{n}\{(X_i-\mu)-(\bar{X}-\mu)\}^2\\ &= \sum_{i=1}^{n}\{(X_i-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2\}\\ &= \sum_{i=1}^{n}\{(X_i-\mu)^2\}-2\sum_{i=1}^{n}\{(X_i-\mu)(\bar{X}-\mu)\}+\sum_{i=1}^{n}\{(\bar{X}-\mu)^2\}\\ &= \sum_{i=1}^{n}\{(X_i-\mu)^2\}-2(\bar{X}-\mu)\sum_{i=1}^{n}\{(X_i-\mu)\}+n(\bar{X}-\mu)^2\end{aligned} \tag{3-16}$$

ここで、標本平均の定義から$\sum_{i=1}^{n}\{(X_i-\mu)\}$について以下が成り立ちます。

$$\begin{aligned}\sum_{i=1}^{n}\{(X_i-\mu)\} &= \sum_{i=1}^{n}\{X_i\}-\sum_{i=1}^{n}\{\mu\}\\ &= n\frac{1}{n}\sum_{i=1}^{n}\{X_i\}-\sum_{i=1}^{n}\{\mu\}\\ &= n\bar{X}-n\mu\\ &= n(\bar{X}-\mu)\end{aligned} \tag{3-17}$$

よって$\sum_{i=1}^{n}(X_i-\bar{X})^2$は以下となります。

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i-\bar{X})^2 &= \sum_{i=1}^{n}\{(X_i-\mu)^2\}-2(\bar{X}-\mu)n(\bar{X}-\mu)+n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}\{(X_i-\mu)^2\}-2n(\bar{X}-\mu)^2+n(\bar{X}-\mu)^2 \\
&= \sum_{i=1}^{n}\{(X_i-\mu)^2\}-n(\bar{X}-\mu)^2
\end{aligned}
\tag{3-18}
$$

「母平均からの正しい差異$(X_i-\mu)^2$」と、「標本平均から母平均の差異$(\bar{X}-\mu)^2$」の 2 つに分解できました。

上記の結果を利用して、$E(S^2)$を式変形します。

$$
\begin{aligned}
E(S^2) &= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}\{(X_i-\mu)^2\}-n(\bar{X}-\mu)^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}\{(X_i-\mu)^2\}\right]-\frac{1}{n}E[n(\bar{X}-\mu)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2]-E[(\bar{X}-\mu)^2]
\end{aligned}
\tag{3-19}
$$

ここで$E[(X_i-\mu)^2]=\sigma^2$です。一方で$E[(\bar{X}-\mu)^2]$は標本平均の分散であり$E[(\bar{X}-\mu)^2]=V(\bar{X})=\sigma^2/n$となります。

これを利用すると以下のように整理できます。

$$E(S^2) = \frac{1}{n} n\sigma^2 - \frac{\sigma^2}{n}$$

$$= \sigma^2 - \frac{\sigma^2}{n} \tag{3-20}$$

$$= \frac{n-1}{n}\sigma^2$$

以上から、標本分散S^2の期待値は、母分散σ^2の$(n-1)/n$倍になっていることがわかります。母分散を過小評価しているわけです。そのため、標本分散を$n/(n-1)$倍した不偏分散は、母分散の不偏推定量となります。

■■■証明終わり

回帰分析

１．単回帰モデルの仮定

回帰モデルの標準的仮定

証明に必要な仮定を明確にするため、モデル式を定義します。

応答変数をy_iと、説明変数をx_iとします。β_0, β_1を単回帰モデルの係数とします。一般的な単回帰モデルは以下のように表記されます。

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{4-1}$$

ここでε_iは**誤差項**や**かく乱項**と呼びます。

単回帰モデルでは、以下を仮定します。これを**標準的仮定**と呼ぶこともあります。

- 説明変数x_iは定数である（確率変数ではないということです）
- 誤差項ε_iは確率変数である（正規分布に従うとは限りません）
- 誤差項の期待値は 0 である。すなわち$E(\varepsilon_i) = 0$が成り立つ
- 誤差項の分散は一定である。$V(\varepsilon_i) = \sigma^2$とする。
- 誤差項は互いに無相関である。すなわち$C\left(\varepsilon_i, \varepsilon_j\right) = 0$が成り立つ。

なお、教科書によって標準的仮定の書き方がやや異なります。例えば係数の一致性の証明のために$n \to \infty$において$\sum_{i=1}^{n}(x_i - \bar{x})^2 \to \infty$を標準的仮定に加える教科書もあるようです。ただ、上記仮定はほとんどの場合成り立っていると考えられるので、今回は省略します。

また、標準的仮定において誤差項が正規分布に従うと記載している教科書もありますが、ここでは正規線形モデルと区別するために入れていません。

標準的仮定から成り立つこと

標準的仮定からε_iについて成り立つことを整理します。

$E(\varepsilon_i) = 0$より

$$V(\varepsilon_i) = E\left[\left(\varepsilon_i - E(\varepsilon_i)\right)^2\right] = E\left(\varepsilon_i^2\right) = \sigma^2 \qquad \text{(4-2)}$$

よって$E\left(\varepsilon_i^2\right) = \sigma^2$です。

$E(\varepsilon_i) = 0$より、$i \neq j$において

$$C\left(\varepsilon_i, \varepsilon_j\right) = E\left[\left(\varepsilon_i - E(\varepsilon_i)\right)\left(\varepsilon_j - E\left(\varepsilon_j\right)\right)\right] = E\left(\varepsilon_i \varepsilon_j\right) = 0 \qquad \text{(4-3)}$$

よって$i \neq j$において$E\left(\varepsilon_i \varepsilon_j\right) = 0$です。
$E\left(\varepsilon_i \varepsilon_j\right)$は、$i = j$なら$\sigma^2$で、$i \neq j$なら 0 という事実は、たまに利用します。

標準的仮定からy_iについて成り立つことを整理します。なお、以下の 3 つの性質は倉田・星野(2009)の定理 9.1 を参考にしています。参照する頻度が高い性質だと思います。

y_iの期待値は、回帰直線上にある

$$E(y_i) = \beta_0 + \beta_1 x_i \qquad \text{(4-4)}$$

証明■■■
式変形して示します。$\beta_0 + \beta_1 x_i$が定数であることに注意して公式 2-1 を適用します。

$$\begin{aligned} E(y_i) &= E(\beta_0 + \beta_1 x_i + \varepsilon_i) \\ &= \beta_0 + \beta_1 x_i + E(\varepsilon_i) \qquad \text{(4-5)} \\ &= \beta_0 + \beta_1 x_i \end{aligned}$$

以上により、$E(y_i) = \ \beta_0 + \beta_1 x_i$が示されました。
■■■証明終わり

y_iの分散は、誤差項の分散と等しい

$$V(y_i) = \sigma^2 \tag{4-6}$$

証明■■■

式変形して示します。$\beta_0 + \beta_1 x_i$が定数であることに注意して公式 4-2 を適用します。

$$\begin{aligned} V(y_i) &= V(\beta_0 + \beta_1 x_i + \varepsilon_i) \\ &= V(\varepsilon_i) \\ &= \sigma^2 \end{aligned} \tag{4-7}$$

以上により、$V(y_i) = \sigma^2$が示されました。

■■■証明終わり

y_iは互いに無相関

$i \neq j$において

$$C(y_i, y_j) = 0 \tag{4-8}$$

証明■■■

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$であり、$E(y_i) = \beta_0 + \beta_1 x_i$であることを利用して、式変形して示します。

$$\begin{aligned} & C(y_i, y_j) \\ &= E[\{y_i - E(y_i)\}\{y_j - E(y_j)\}] \\ &= E[\{(\beta_0 + \beta_1 x_i + \varepsilon_i) - (\beta_0 + \beta_1 x_i)\}\{(\beta_0 + \beta_1 x_j + \varepsilon_j) - (\beta_0 + \beta_1 x_j)\}] \\ &= E[\varepsilon_i \varepsilon_j] \\ &= 0 \end{aligned} \tag{4-9}$$

以上により、$C(y_i, y_j) = 0$が示されました。

■■■証明終わり

正規線形モデルの仮定

正規線形モデルでは、誤差項が独立で同一な正規分布に従っていると仮定します。そのため、誤差項は$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$となります。

正規分布の性質から以下が成り立ちます。

$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 \times x_i, \sigma^2) \tag{4-10}$$

２．最小二乗推定量

最小二乗推定量の導出

最小二乗法による係数の推定方法について解説します。モデルは下記とします。β_0, β_1が推定すべき係数です。

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{4-11}$$

残差平方和は下記の通りです。

$$\mathrm{RSS}(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{4-12}$$

ただし応答変数がy_iで、説明変数がx_iです。

偏微分

$\mathrm{RSS}(\beta_0, \beta_1)$を係数で偏微分して 0 になるときの係数を調べます。

$$\begin{gathered} \frac{\partial \mathrm{RSS}(\beta_0, \beta_1)}{\partial \beta_0} = 0 \\ \frac{\partial \mathrm{RSS}(\beta_0, \beta_1)}{\partial \beta_1} = 0 \end{gathered} \tag{4-13}$$

微分の際には合成関数の微分の公式を使うのが簡単です。すなわち一般的に下記が成り立ちます。

$$\{g(f(x))\}' = g'(f(x))f'(x) \tag{4-14}$$

今回の事例では$g(x) = x^2$です。よってβ_0について以下が成り立ちます

$$\begin{aligned}\frac{\partial \mathrm{RSS}(\beta_0, \beta_1)}{\partial \beta_0} &= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-1) \\ &= \sum_{i=1}^{n} 2(-y_i + \beta_0 + \beta_1 x_i) \\ &= 2\left[-\sum_{i=1}^{n}[y_i] + n\beta_0 + \beta_1 \sum_{i=1}^{n}[x_i]\right]\end{aligned} \tag{4-15}$$

同様にβ_1について以下が成り立ちます。

$$\begin{aligned}\frac{\partial \mathrm{RSS}(\beta_0, \beta_1)}{\partial \beta_1} &= \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i)(-x_i) \\ &= \sum_{i=1}^{n} 2\left(-x_i y_i + \beta_0 x_i + \beta_1 x_i^2\right) \\ &= 2\left[-\sum_{i=1}^{n}[x_i y_i] + \beta_0 \sum_{i=1}^{n}[x_i] + \beta_1 \sum_{i=1}^{n}[x_i^2]\right]\end{aligned} \tag{4-16}$$

よって、以下の連立方程式を解くことになります。なお、方程式は両辺を 2 で除した結果を載せています。見やすさのために式(a)と式(b)と呼ぶことにします。

$$\begin{aligned}-\sum_{i=1}^{n}[y_i] + n\beta_0 + \beta_1 \sum_{i=1}^{n}[x_i] = 0 \quad (a) \\ -\sum_{i=1}^{n}[x_i y_i] + \beta_0 \sum_{i=1}^{n}[x_i] + \beta_1 \sum_{i=1}^{n}[x_i^2] = 0 \quad (b)\end{aligned} \tag{4-17}$$

係数$\boldsymbol{\beta_0}$の推定

式(a)を変形します。

$$-\sum_{i=1}^{n}[y_i]+n\beta_0+\beta_1\sum_{i=1}^{n}[x_i]=0$$

$$\begin{aligned} n\beta_0 &= \sum_{i=1}^{n}[y_i]-\beta_1\sum_{i=1}^{n}[x_i] \\ \beta_0 &= \frac{1}{n}\sum_{i=1}^{n}[y_i]-\beta_1\frac{1}{n}\sum_{i=1}^{n}[x_i] \\ \beta_0 &= \bar{y}-\hat{\beta}_1\bar{x} \end{aligned} \tag{4-18}$$

よって$\beta_0=\bar{y}-\hat{\beta}_1\bar{x}$が示せました。

係数$\boldsymbol{\beta_1}$の推定

続いて式(b)を変形します。

$$\begin{aligned} -\sum_{i=1}^{n}[x_iy_i]+\beta_0\sum_{i=1}^{n}[x_i]+\beta_1\sum_{i=1}^{n}[x_i^2] &= 0 \\ \beta_0\sum_{i=1}^{n}[x_i]+\beta_1\sum_{i=1}^{n}[x_i^2] &= \sum_{i=1}^{n}[x_iy_i] \end{aligned} \tag{4-19}$$

ここで$\beta_0=\bar{y}-\hat{\beta}_1\bar{x}$を代入します。

$$\begin{aligned} (\bar{y}-\hat{\beta}_1\bar{x})\sum_{i=1}^{n}[x_i]+\beta_1\sum_{i=1}^{n}[x_i^2] &= \sum_{i=1}^{n}[x_iy_i] \\ \bar{y}\sum_{i=1}^{n}[x_i]-\hat{\beta}_1\bar{x}\sum_{i=1}^{n}[x_i]+\beta_1\sum_{i=1}^{n}[x_i^2] &= \sum_{i=1}^{n}[x_iy_i] \end{aligned} \tag{4-20}$$

やや技巧的ですが、1 項目と 2 項目に$n \cdot (1/n) = 1$をかけて整理します。

$$n\bar{y}\frac{1}{n}\sum_{i=1}^{n}[x_i] - n\hat{\beta}_1\bar{x}\frac{1}{n}\sum_{i=1}^{n}[x_i] + \beta_1\sum_{i=1}^{n}[x_i^2] = \sum_{i=1}^{n}[x_i y_i]$$

$$n\bar{x}\bar{y} - n\hat{\beta}_1(\bar{x})^2 + \beta_1\sum_{i=1}^{n}[x_i^2] = \sum_{i=1}^{n}[x_i y_i] \tag{4-21}$$

$$\hat{\beta}_1\left(\sum_{i=1}^{n}[x_i^2] - n(\bar{x})^2\right) = \sum_{i=1}^{n}[x_i y_i] - n\bar{x}\bar{y}$$

両辺をnで割ってから整理します。2 行目から 3 行目の変形では、公式 3-1 と公式 5-1 を利用します。

$$\hat{\beta}_1\left(\frac{1}{n}\sum_{i=1}^{n}[x_i^2] - (\bar{x})^2\right) = \frac{1}{n}\sum_{i=1}^{n}[x_i y_i] - \bar{x}\bar{y}$$

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}[x_i y_i] - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n}[x_i^2] - (\bar{x})^2} \tag{4-22}$$

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(x,y)}{s_x^2}$$

よって、最小二乗推定量は以下のようになることが示せました。

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(x,y)}{s_x^2} \tag{4-23}$$

なお、$\hat{\beta}_1$の計算において、分子と分母をともにnで割っています。もちろん分子と分母をともにnで割るのをやめても結果は変わりません。この場合は以下のように表記します。

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(x, y)}{s_x^2} = \frac{\mathrm{SS}_{xy}}{\mathrm{SS}_{xx}}$$

$$\mathrm{SS}_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{4-24}$$

$$\mathrm{SS}_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

回帰直線が応答変数と説明変数の平均値を通ることの証明

$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$を証明します。$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$を変形することですぐ示せます。

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_0 + \hat{\beta}_1 \bar{x} &= \bar{y} \end{aligned} \tag{4-25}$$

回帰係数の期待値

回帰係数$\hat{\beta}_0, \hat{\beta}_1$の期待値と分散を求めます。先に期待値を求めてから、分散を求めます。

なお、以下の結果を利用します。

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\mathrm{Cov}(x, y)}{s_x^2} = \frac{\mathrm{SS}_{xy}}{\mathrm{SS}_{xx}} \end{aligned} \tag{4-26}$$

係数$\hat{\beta}_1$の期待値

定義より以下が成り立ちます。

$$E(\hat{\beta}_1) = E\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \tag{4-27}$$

ここで、確率変数はy_iだけであることに注意すると、以下が成り立ちます。

$$\begin{aligned} E(\hat{\beta}_1) &= \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sum_{i=1}^{n}(x_i - \bar{x})\,E(y_i - \bar{y}) \\ &= \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sum_{i=1}^{n}(x_i - \bar{x})\,\{E(y_i) - E(\bar{y})\} \end{aligned} \tag{4-28}$$

ここで$E(y_i) =\ \beta_0 + \beta_1 x_i$を代入します。

$$\begin{aligned} E(\hat{\beta}_1) &= \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sum_{i=1}^{n}(x_i - \bar{x})\,\{(\ \beta_0 + \beta_1 x_i) - (\ \beta_0 + \beta_1 \bar{x})\} \\ &= \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})\beta_1 \\ &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\beta_1 \\ &= \beta_1 \end{aligned} \tag{4-29}$$

よって$E(\hat{\beta}_1) = \beta_1$となり、$\hat{\beta}_1$は不偏推定量であることが示されました。

係数$\hat{\boldsymbol{\beta}}_{\mathbf{0}}$の期待値

定義および$\bar{x}$が定数であることから以下が成り立ちます。

$$\begin{aligned} E(\hat{\beta}_0) &= E(\bar{y} - \hat{\beta}_1 \bar{x}) \\ &= E(\bar{y}) - \bar{x} \cdot E(\hat{\beta}_1) \end{aligned} \tag{4-30}$$

ここで$E(y_i) = \beta_0 + \beta_1 x_i$と、$E(\hat{\beta}_1) = \beta_1$を代入します。

$$\begin{aligned} E(\hat{\beta}_0) &= \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} \\ &= \beta_0 \end{aligned} \tag{4-31}$$

よって$E(\hat{\beta}_0) = \beta_0$となり、$\hat{\beta}_0$は不偏推定量であることが示されました。

係数$\hat{\boldsymbol{\beta}}_1$の分散

定義と公式 1-3 より以下が成り立ちます。

$$\begin{aligned} V(\hat{\beta}_1) &= V\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \\ &= V\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \end{aligned} \tag{4-32}$$

ここで確率変数がy_iだけであり、公式 4-2 に注意すると以下が成り立ちます。

$$V(\hat{\beta}_1) = \frac{1}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\,V(y_i) \tag{4-33}$$

ここで$V(y_i)=\sigma^2$と$\sum_{i=1}^{n}(x_i-\bar{x})^2=\mathrm{SS}_{xx}$を代入します。

$$\begin{aligned} V(\hat{\beta}_1) &= \frac{1}{(\mathrm{SS}_{xx})^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\,\sigma^2 \\ &= \frac{\sigma^2}{(\mathrm{SS}_{xx})^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ &= \frac{\sigma^2}{(\mathrm{SS}_{xx})^2}\mathrm{SS}_{xx} \\ &= \frac{\sigma^2}{\mathrm{SS}_{xx}} \end{aligned} \tag{4-34}$$

係数$\hat{\beta}_0$の分散

$\bar{x}$が定数であり、公式 4-2 と公式 7-2 に注意すると以下が成り立ちます。

$$\begin{aligned} V(\hat{\beta}_0) &= V(\bar{y} - \hat{\beta}_1 \bar{x}) \\ &= V(\bar{y}) - (\bar{x})^2 \cdot V(\hat{\beta}_1) - 2\bar{x} \cdot C(\bar{y}, \hat{\beta}_1) \end{aligned} \tag{4-35}$$

ここで平均値の分散の公式より$V(\bar{y}) = \sigma^2/n$、また$V(\hat{\beta}_1) = \sigma^2/SS_{xx}$であり、そして後ほど示されるように$C(\bar{y}, \hat{\beta}_1) = 0$であるので、以下が成り立ちます。

$$\begin{aligned} V(\hat{\beta}_0) &= \frac{\sigma^2}{n} + (\bar{x})^2 \frac{\sigma^2}{\mathrm{SS}_{xx}} \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x})^2}{\mathrm{SS}_{xx}} \right) \end{aligned} \tag{4-36}$$

以上の結果を整理しておきます。

$$\begin{aligned} V(\hat{\beta}_1) &= \frac{\sigma^2}{\mathrm{SS}_{xx}} \\ V(\hat{\beta}_0) &= \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x})^2}{\mathrm{SS}_{xx}} \right) \end{aligned} \tag{4-37}$$

上記の結果から、誤差項の分散σ^2が大きければ、推定量の分散も大きいことがわかります。これは自然なことです。

推定量の分散の分母にSS_{xx}が入っていることも大きな特徴です。SS_{xx}は説明変数xのばらつきの大きさだと言えます。説明変数がさまざま変化しているときに、推定量の分散は小さくなります。ビールの売り上げy_iを気温x_iから予測する場合は、特定の気温のときだけデータを取得するよりも、いろいろな気温でデータを取得したほうが精度よく推定できることがわかります。こちらも言われてみれば当たり前のことですが、数式でちゃんと証明できました。

$\bar{y}$と$\hat{\beta}_1$の共分散

$C(\bar{y},\hat{\beta}_1)$を求めます。

定義と公式 1-3 から以下となります。

$$\begin{aligned} C(\bar{y},\hat{\beta}_1) &= C\left(\frac{1}{n}\sum_{j=1}^{n} y_j\,, \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \\ &= C\left(\frac{1}{n}\sum_{j=1}^{n} y_j\,, \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \end{aligned} \tag{4-38}$$

ここで確率変数がy_iだけであり、公式 7-5 に注意すると以下が成り立ちます。

$$C(\bar{y},\hat{\beta}_1) = \frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-\bar{x})C(y_i,y_j) \tag{4-39}$$

ここで、標準的仮定から、$i \neq j$において$C(y_i,y_j)=0$であること、そして$i=j$において$C(y_i,y_j)=V(y_i)=\sigma^2$であることを利用すると以下が成り立ちます。

$$\begin{aligned} C(\bar{y},\hat{\beta}_1) &= \frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x})\sigma^2 \\ &= \frac{\sigma^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}(x_i-\bar{x}) \end{aligned} \tag{4-40}$$

ここで公式 1-2 より$\sum_{i=1}^{n}(x_i-\bar{x})=0$なので以下が成り立ちます。

$$C(\bar{y},\hat{\beta}_1) = 0 \tag{4-41}$$

この結果は有用なのでしばしば登場します。

$\hat{\boldsymbol{\beta}}_0$と$\hat{\boldsymbol{\beta}}_1$の共分散

定義と公式 7-5 から以下が成り立ちます。

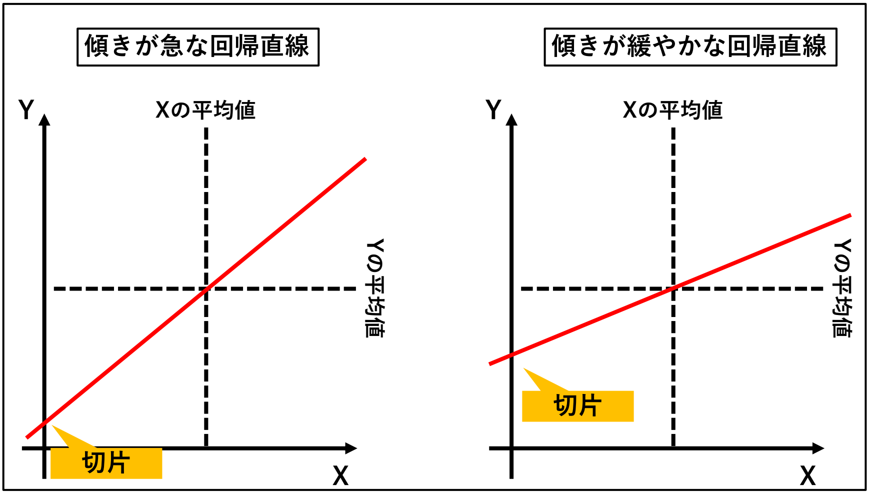
$$\begin{aligned} C(\hat{\beta}_0, \hat{\beta}_1) &= C(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1) \\ &= C(\bar{y}, \hat{\beta}_1) - \bar{x} C(\hat{\beta}_1, \hat{\beta}_1) \end{aligned} \tag{4-42}$$

ここで$C(\bar{y}, \hat{\beta}_1) = 0$であり、$C(\hat{\beta}_1, \hat{\beta}_1) = V(\hat{\beta}_1)$であることから以下が成り立ちます。

$$\begin{aligned} C(\hat{\beta}_0, \hat{\beta}_1) &= -\bar{x} V(\hat{\beta}_1) \\ &= -\bar{x} \frac{\sigma^2}{\mathrm{SS}_{xx}} \end{aligned} \tag{4-43}$$

計算して満足するのではなくて、この結果をしっかりと吟味しましょう。まずは、2 つのパラメータが独立ではないというだけで、ちょっとした発見ですよね。切片$\hat{\beta}_0$と傾き$\hat{\beta}_1$には関連性があるわけです。

$\bar{x}$が正であれば、切片$\hat{\beta}_0$と傾き$\hat{\beta}_1$の共分散は負の値を取ります。回帰直線のグラフをイメージしてもらえればわかるように、「$\bar{x}$と$\bar{y}$を通る直線」を考えた時、傾きが急になると切片は小さな値をとるし、傾きが緩やかだと切片は大きな値を取ります。


傾きが急な回帰直線
Y
Xの平均値
Yの平均値
切片
X
傾きが緩やかな回帰直線
Y
Xの平均値
Yの平均値
切片
X

回帰直線の分散

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$で計算される$\hat{y}_i$の分散を求めます。

$$V(\hat{y}_i) = V\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \tag{4-44}$$

ここで$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$を代入します。

$$\begin{aligned} V(\hat{y}_i) &= V\left(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i\right) \\ &= V\left[\bar{y} - \hat{\beta}_1 (x_i - \bar{x})\right] \end{aligned} \tag{4-45}$$

ここで公式 7-2 と$C\left(\bar{y}, \hat{\beta}_1\right) = 0$から以下が成り立ちます。

$$\begin{aligned} V(\hat{y}_i) &= V(\bar{y}) + (x_i - \bar{x})^2 V\left(\hat{\beta}_1\right) + 2(x_i - \bar{x}) C\left(\bar{y}, \hat{\beta}_1\right) \\ &= V(\bar{y}) + (x_i - \bar{x})^2 V\left(\hat{\beta}_1\right) \\ &= \frac{\sigma^2}{n} + (x_i - \bar{x})^2 \frac{\sigma^2}{\mathrm{SS}_{xx}} \\ &= \sigma^2 \left\{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\mathrm{SS}_{xx}}\right\} \end{aligned} \tag{4-46}$$

$V(\hat{y}_i)$は説明変数x_iによって変化することに注意が必要です。x_iが平均値$\bar{x}$に近ければ$V(\hat{y}_i)$は小さいですが、x_iが平均値$\bar{x}$と離れていると$V(\hat{y}_i)$は大きくなります。

ビールの売り上げと気温の事例でいうと、平均的な気温であるときには売り上げの平均値を推測するのは容易ですが、極端に寒かったり極端に暑かったりする日の売り上げを推測するのは難しいことがわかります。直観にあう結果だと思います。

なお、予測区間を計算する際には、データのばらつきσ^2を加えた$V(\hat{y}_i) + \sigma^2$が利用されます。

Python で学ぶあたらしい統計学の教科書の補足

「Python で学ぶあたらしい統計学の教科書 第 2 版」をより深く理解するために参考となる箇所を紹介します。

第 4 部　確率と確率分布の基本

第 2 章　確率分布の基本

代表的な確率分布とその特徴
→1. 一様分布（離散型）＋　2. 一様分布（連続型）

第 3 章　二項分布

代表的な確率分布とその特徴
→3. 二項分布

第 5 部　統計的推定

第 3 章　母平均の推定

統計的推定
→1. 標本平均の特性

ところで、公式 7-4 は少し難しいので、以下の順番で読むのがおすすめです。

① 公式 7-3（独立な確率変数の積の期待値に関する公式）
② 公式 6-1（共分散の計算公式）
③ 公式 6-4（独立な確率変数における共分散に関する公式）
④ 公式 7-1（確率変数の和の期待値に関する公式）
⑤ 公式 7-2（確率変数の和の分散に関する公式）
⑥ 公式 7-4（独立な確率変数の和の分散に関する公式）

第４章　母分散の推定

統計的推定
→2. 標本分散の特性

第８部　正規線形モデル

第１章　連続型の説明変数を１つもつモデル

書籍本文では正規線形モデルとして単回帰モデルを導入しました。最尤法との兼ね合いで、このように導入したほうがシンプルだからです。

しかし、最小二乗法に関するいくつかの有用な結果は、誤差分布が正規分布以外であっても成り立ちます。