不偏分散は，どうして $n-1$ で割るの？ *

浜田　宏

登場人物

青葉（あおば）: 大学生。数学が苦手

花京院（かきょういん）：大学生。数学が好き。スタンドは使えない。

花京院は作業机の隅で、本を読んでいる。

広い机の上には、論文のコピーと計算用紙が散乱している。

青葉は授業課題のリポートを書くために、パソコンのモニタに向かっている。

数理行動科学研究室の分析室には、青葉と花京院の 2 人だけがいた。

「ねえ、花京院くん。いま時間ある？」花京院の読書が一段落したのを見計らって、青葉は声をかけた。

「うん、まあ暇と言えば暇だし、暇でないと言えば暇じゃない」花京院は、壁の時計で時間を確認した。

「どっちなのよ」

「ごめん、暇だった。なに？」

「えっとね、統計学の授業で《不偏分散》って、習ったでしょ……。よくわかんないから教えてほしいなと思って」

---

* ver1.0. 2022 年 6 月 20 日

「不偏分散？ なにそれ」

「ほら、あの $n-1$ で割るやつだよ」

「ああ、この推定量のことかな」

花京院は机のうえの計算用紙に式を書いた。

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

「そうそう。これこれ」

「これ、不偏分散っていう名前なんだ」

「いま自分で式を書いたじゃん。名前覚えてないの？」と青葉が聞いた。

「うーん、式はなんとなく覚えてるけど、名前まで覚えてなかった」

「この式って、すごく変じゃない？」

「30 分」

「え？」

「ちゃんと説明するのに、30 分」

そう言うと花京院はパソコンの前に座ると、キーボードをたたき始めた。

1　青葉の疑問

「君は不偏分散のどこが変だと思うの？」

「えーっとね、n 個の項を足してるのに、$n-1$ で割るところ」

「そうだね。おかしいね」

「でしょ？ 花京院君だって、おかしいと思うでしょ？」

「そうだね。最初に見たときは、なんでだろうと思ったよ。でも理由を考えたら納得したよ」

「理由あるんだ」

「ちゃんと理由があるんだよ。実験して確かめてみよう」

「実験？」

花京院は，キーボードでなにやら謎のコードを書くと、ヒストグラム

を一つ画面に描いた。

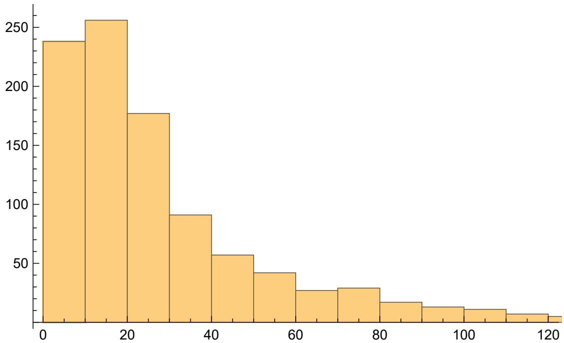

図 1　全体の分布

「これをね、観察する対象の《全体》としよう。1000 個の数値をヒストグラムにまとめたものだよ」

「わかった。なんのデータなの？」

「いや、特に意味はないよ。テキトーにコンピュータに作らせたから」

「なにか意味がないと、イメージ湧かないよー」

「しょうがないな。じゃあ柿ピー」

「は？」

「大体 0〜120 くらいで分布しているから、1 日に食べる柿ピーのグラム数ってことにしよう。柿ピーの小袋って 30g なんだよ」花京院は机のうえに置いてあったお菓子の袋をひとつ手に取った。

「わかった。じゃあこれは《1 日に食べる柿ピーのグラム数》のデータだね」

「そういうこと。この 1000 人は M 町に住む成人男性全員の数としよう。全員ってところが重要だよ。この《全体》のデータからまず平均値を計算してみよう。1000 人分のデータを

$$x_1, x_2, \ldots, x_{1000}$$

で表すと、その平均は

$$\frac{1}{1000}\sum_{i=1}^{1000} x_i$$

で計算できる。実際に計算してみると 31.91 だから、1 人平均 32g ほど柿ピーを食べることが分かる。」

「意外と少ないんだね。私なら 2 袋はいけるよ」

「もちろん、そういう人もいるだろう。そういう食べる量の《バラつき》を表すのが分散だったね。《全体》データから分散を計算してみよう。さきほど計算した平均値を μ とおけば分散は

$$\frac{1}{1000}\sum_{i=1}^{1000}(x_i-\mu)^2$$

だね。計算した結果は 1287.11 だよ」

「これは n で割ったね」

「そう。いま計算したのは《全体》の分散。これを《母集団》の分散ともいう。もし観察対象全てのデータが手元にあるのなら、分散を計算するときただ n で割ればいい」

「うん。それは分かった。じゃあ、いつ $n-1$ で割るの？」

「標本から《全体》の分散を推定するときだよ」

「ちょっとなに言ってるかわからない」

「実例で示そう」

2　標本をつかった推定

「いま《全体》から 20 個だけデータをランダムに抜き出したとしよう。たとえばこんな感じ」

$$65.06, 11.44, 13.08, \ldots, 145.86$$

「うん」

「省略して書いたけど、いまここに 20 個の数値がある。この 20 個を使って平均値を計算してみよう。全部足して 20 で割るだけだよ。結果は

$$\frac{65.06+11.44+13.08+\cdots+145.86}{20}=28.46$$

だよ。20 個取り出して平均をとったから一般的な式で書くと

$$\frac{x_1 + x_2 + \cdots + x_{20}}{20} = 28.46$$

だよ。これは標本の平均値で、記号で $\bar{x}$ と書くことにしよう。ここまではいいかな？」

「うん、まあ平均値の計算くらい分かるよ」

「さらにこの標本平均値を使って、不偏分散を計算する[*1]。つまり

> 全ての値から標本平均値 28.46 を引いて 2 乗した数を全部足し、最後に 19 で割る。

という操作をする。こんな感じだ。

$$\begin{aligned}&\{(65.06 - 28.46)^2 + (11.44 - 28.46)^2 + \cdots + (145.86 - 28.46)^2\}/19 \\ &= 1183.32\end{aligned}$$

比較のために $n = 20$ で割った数も計算しておこう。

$$\begin{aligned}&\{(65.06 - 28.46)^2 + (11.44 - 28.46)^2 + \cdots + (145.86 - 28.46)^2\}/20 \\ &= 1124.15\end{aligned}$$

となる」

「うーん、計算方法は分かるよ。でもどうして $n-1$ で割ったり、n で割ったりするの？」

「$n-1$ で割った推定値と、n で割った推定値を比べてみよう。どっちが《全体》の分散に近い？」

「えーっと、さっき計算した《全体》の分散は、1287.11 だったね。ということは、

$$\begin{array}{cl} n-1 & \text{で割った場合} : 1183.32 \\ n & \text{で割った場合} : 1124.15 \end{array}$$

[*1] 不偏分散という確率変数（推定量）の実現値（推定値）を計算する、とも表現できます

だから、$n-1$ で割った場合のほうが，本当の値に近いよ」

「そのとおり。実験の結果、標本から《全体》の分散を推測するときに、n で割るより $n-1$ で割った《不偏分散》を使うほうが、本当の値に近い値になった」

「でも、いまのは偶然かもしれないじゃん」

青葉は怪しそうに計算結果を見比べた。

「いいねえ。その注意深さ。じゃあ計算を繰り返してみよう。手続きはこうだよ」

1. 《全体》から 20 個の標本を抜き出す
2. 標本から「n で割る推定値」と「$n-1$ で割る推定値」を計算する
3. この操作を 100 回繰り返した後でそれぞれの推定値の平均を求める
4. どちらが本当の値に近いかを比較する

「いいかな？」

「OK」

「それじゃあ計算するよ。100 回ほど標本抽出と計算を繰り返してみよう」花京院は、計算用のコードを書き足した。

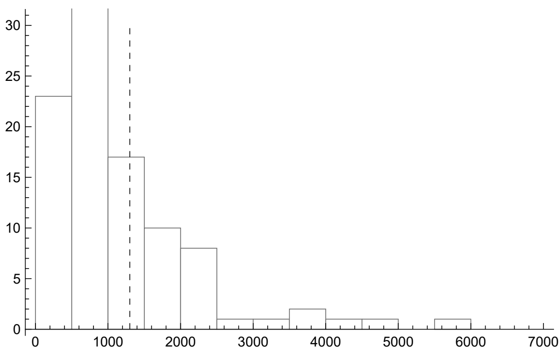

図 2　n で割った場合の結果

「まずこれが n で割った場合。100 回分の計算結果のヒストグラムだ。

点線の位置が本当の値だよ」

「うーん、まあ微妙だね」

「次いこう」

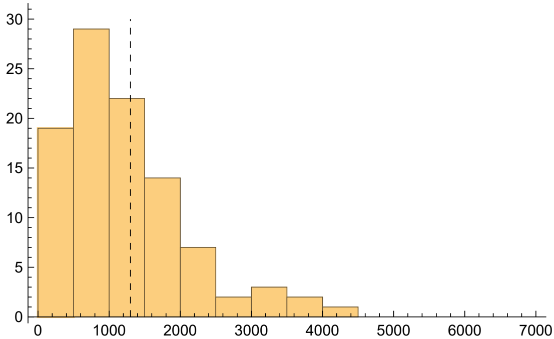

図 3　$n-1$ で割った場合の結果

「これが $n-1$ で割った場合」

「うーん、これも結構ズレてるね」

「二つ並べてみよう」

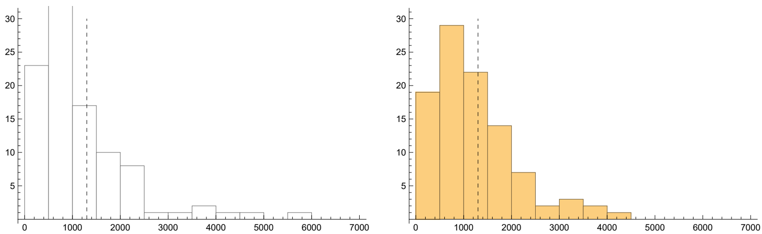

図 4　計算結果の比較。左が n、右が $n-1$ で割った場合の結果

「うーん、微妙だけど、右の方が本当の値に近い数値を推測できてるのかなあ」

「ちょっと見た目では分からないね。それぞれの平均を計算して、本当の値と比較してみよう

$$
\begin{aligned}
n\text{ で割った推定値の平均} &: 1184.31 \\
n-1\text{ で割った推定値の平均} &: 1279.28 \\
\text{本当の値} &: 1287.11
\end{aligned}
$$

この結果を見ると、

> 標本から《全体》の分散を推測するときに、n で割るより $n-1$ で割った《不偏分散》を使うほうが、平均的に本当の値に近い値になる

ことが分かる」

「ちょっと待ってよ。まだなんかしっくりこないなー」

青葉は計算結果を見比べながら、つぶやいた。

「どこが？」

「計算の結果がこうなることは分かったよ。でもどういえばいいのかな……。どうしてそうなるのか、まだ分からない」

「ふむ」

「《全体》の分散を計算する式は

$$\frac{1}{1000}\sum_{i=1}^{1000}(x_i-\mu)^2$$

でしょ。標本の分散を

$$\frac{1}{20}\sum_{i=1}^{20}(x_i-\bar{x})^2$$

っていう似たような式で計算すると、どうしてズレちゃうの？」

「そうだね。それも実験して確かめてみよう」

3　確率変数で表す

「先ほどの実験で確かめたことは標本を使って

$$\frac{1}{20}\sum_{i=1}^{20}(x_i-\bar{x})^2$$

と

$$\frac{1}{19}\sum_{i=1}^{20}(x_i - \bar{x})^2$$

という式で《全体》の分散を推定すると、19 で割ったほうが平均的に本当の値に近いってことだった」

「そうだね」

「ここで標本として取り出した数値

$$x_1, x_2, \ldots, x_n$$

は、観測するまで値が分からないけど、《全体》の中にある数値であることは決まっているという特徴に注目する」

「まあそうだね」

「このような特徴は、確率変数という道具で表すことができる」

「確率変数？」

「とりあえずサイコロを想像してほしい。サイコロは振るまで値が決まらないよね。標本も同じように観測するまでは、どの値になるかわからない。でも《全体》の中にある数値が出てくることは決まってるし、《全体》の中で観測しやすい数値は、やっぱり標本でも観測しやすい。そこで《全体》の分布を確率変数 X で表し、標本を

$$X_1, X_2, X_3, \ldots, X_{20}$$

という 20 個の確率変数で表す」

「ちょっとなに言ってるかわからない」

「もう少し具体的に考えてみよう。《全体》データに記録された数値 1000 個が表面に刻印された 1000 面サイコロを想像する。全体からサイズ 20 の標本をとりだす作業は、1000 面サイコロの出目を 20 回分記録することに対応する。その結果が

$$(x_1, x_2, \ldots, x_{20}) = (3.28, 7.65, \ldots, 11.09)$$

だとする。次に 20 回ふった結果が

$$(x_1, x_2, \ldots, x_{20}) = (9.75, 2.12, \ldots, 6.58)$$

だとする。

　このように標本の実現値は観測するたびに異なる値がでてくる。

　ただし全ての値は《全体》の中にある数値で、同じサイコロ X の実現値と見なすことができる。個々の具体的な実現値に対応するサイコロ振りの名前が

$$X_1, X_2, X_3, \ldots, X_{20}$$

というわけ。以下大文字で確率変数、小文字で実現値を表すと決めておく。x_1 は 1 回目のサイコロ振り X_1 の結果で、x_{20} は 1 回目のサイコロ振り X_{20} の結果だよ」

「ふむ。サイコロに喩えるとちょっと分かってきた」

「標本から平均値を計算する式は

$$\frac{x_1 + x_2 + \cdots + x_{20}}{20}$$

だった。このとき

$$\begin{aligned} X_1&\text{というサイコロの目が } x_1 \\ X_2&\text{というサイコロの目が } x_2 \\ &\vdots \\ X_{20}&\text{というサイコロの目が } x_n \end{aligned}$$

という対応になっているので、確率変数を使えば

$$\frac{X_1 + X_2 + \cdots + X_{20}}{20}$$

という抽象的な演算として表現できる。

　これは、《複数の確率変数を、しかじかの演算によって 1 つの確率変数として合成しますよ》という計算方法だ。この操作を記号 $\bar{X}$ で表すことにする。標本に含まれるデータを n とおけば

$$\frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

だ。標本を使って計算するので《標本平均》と呼ばれることが多い。この《標本平均》は平均という名前がついているけど、

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

という、ひとかたまりで確率変数なんだよ。だから 1 つの定まった値じゃなくて、確率的にいろいろな値をとる。つまり分布をもっている」

「うーん、ややこしくなってきた。標本平均なのに 1 つの値じゃないって、どういうこと？」

「《全体》の平均値を考えてみよう。これは 1 つしか存在しない。1000 個のデータを足して 1000 で割った数だから、必ず一つの定まった値となる。その値を記号 μ で表すと約束した」

「そうだったね」

「でも $\bar{X}$ は、$X_1, X_2, X_3, \ldots, X_{20}$ という標本の実現値によって、値が異なる。だから分布を持つ」

「ああなるほど。1 回目にとりだした 20 個の値と、2 回目にとりだした 20 個の値は違うから、$\bar{X}$ の実現値は毎回違う値になるんだ」

「そういうこと。

$$\frac{1}{n}\sum_{i=1}^{n} X_i$$

は、これ自体が確率変数だってことがポイントだよ」

4　確率変数としての不偏分散

「次に《標本平均（確率変数）》をつかい、《標本分散》を定義する。

$$《標本分散》= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

これは君が自然だと思う、《全体》の分散の推定値を確率変数で表したものだよ」

「そうそう。n で割ってるから、こっちのほうが自然な気がする」

「一方、$n-1$ で割る不自然なやつを《不偏分散》という名前で定義する。

$$\text{《不偏分散》} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

」

「うん。これは不自然だよね。どこから $n-1$ が出てきたのか分からない」

「この《標本分散》や《不偏分散》は確率変数かな？ それともただの決まった数かな？」

「うーん、どっちかな。$\bar{X}$ は確率変数だったけど、それをさらに足し合わせるとどうなるのかな」青葉は首をかしげた。

「じゃあ、ヒントを出そう。かく……」

「かく？」

「り……つ……」

「りつ？」

「へん……す……」

「確率変数？ ヒントの出しかた下手すぎるでしょ」

「確率変数同士を足すと、その結果も確率変数になるんだよ。これで準備は整った。先ほど僕らが実験でやった計算は、

$$\text{標本分散}: \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

や

$$\text{不偏分散}: \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

という確率変数の期待値を近似的に求める作業に対応している。期待値とは確率変数の平均のことだよ。実際に《標本分散》という**確率変数**の期待値を計算してみよう」

一般的に、確率変数 X の期待値を

$$E[X] = \sum_{i=1}^{m} P(X = x_i)x_i \quad \text{離散的な場合}$$
$$E[X] = \int_{-\infty}^{\infty} xf(x)dx \quad \text{連続的な場合}$$

と定義する[*2]。

《全体》のデータは 1000 個あるから、その期待値は

$$E[X] = \sum_{i=1}^{1000} P(X = x_i)x_i$$

で

$$P(X = x_i) = \frac{1}{1000}$$

だから

$$E[X] = \sum_{i=1}^{1000} P(X = x_i)x_i = \sum_{i=1}^{1000} \frac{1}{1000}x_i = \frac{1}{1000}\sum_{i=1}^{1000} x_i$$

となり、《全体》の平均値は、《全体》に対応する確率変数 X の期待値と完全に一致することがわかる。

また確率変数 X の分散の定義を

$$V[X] = E[(X - E[X])^2]$$

とする。

以下では《全体》を表す確率変数 X の平均と分散を具体的に

$$E[X] = \mu$$
$$V[X] = \sigma^2$$

とおく。また標本 $X_1, X_2, \ldots, X_n$ は X と同一であり、独立であるとする。これは同じサイコロを振る操作を想像すれば自然な仮定だね。

次にいくつか補題を導入するよ。補題は証明の論理構造を明確にするための命題だよ。

*2 $\sum_{i=1}^{m} P(X = x_i) = 1$ を仮定します。$x_1, x_2, \ldots, x_m$ は、全ての実現値を表しています

補題 1.

$$E[X^2] = V[X] + E[X]^2$$

証明.

$$\begin{aligned}
V[X] &= E[(X - E[X])^2] \\
&= E[X^2 - 2XE[X] + E[X]^2] && \text{展開する} \\
&= E[X^2] - 2E[X]E[X] + E[E[X]^2] && \text{定数を出す} \\
&= E[X^2] - 2E[X]^2 + E[X]^2 && \text{まとめる} \\
&= E[X^2] - E[X]^2
\end{aligned}$$

言い換えると

$$E[X^2] = V[X] + E[X]^2$$

□

補題 2.

$$E[X_i^2] = V[X] + E[X]^2$$

証明. 定義により $X_1, X_2, \ldots, X_n$ は《全体》の分布 X と同じだから、任意の i について

$$E[X] = E[X_i], \quad V[X] = V[X_i]$$

が成立する。 □

補題 3.

$$E[\bar{X}^2] = \frac{1}{n}V[X] + E[X]^2$$

証明. 補題 1 の X を $\bar{X}$ に置き換える。

$$E[\bar{X}^2] = V[\bar{X}] + E[\bar{X}]^2$$

次に

$$V[\bar{X}] = V\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$$

$$
\begin{aligned}
&= \frac{1}{n^2} V\left[X_1 + X_2 + \cdots + X_n\right] \\
&= \frac{1}{n^2}\left\{V[X_1] + V[X_2] + \cdots + V[X_n]\right\} \\
&= \frac{1}{n^2} nV[X] = \frac{1}{n}V[X]
\end{aligned}
$$

および

$$
\begin{aligned}
E[\bar{X}] &= E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\
&= \frac{1}{n}E\left[X_1 + X_2 + \cdots + X_n\right] \\
&= \frac{1}{n}\left\{E[X_1] + E[X_2] + \cdots + E[X_n]\right\} \\
&= \frac{1}{n}nE[X] = E[X]
\end{aligned}
$$

より

$$
E[\bar{X}^2] = V[\bar{X}] + E[\bar{X}]^2 = \frac{1}{n}V[X] + E[X]^2
$$

□

これで準備は整った。あとは《標本分散》の期待値を変形していけば証明完了だよ。

$$
E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]
$$

だから

$$
E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]
$$

の部分を先に計算するよ。

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= E\left[\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right] \\
&= E\left[\sum_{i=1}^{n}X_i^2 - 2\bar{X}\sum_{i=1}^{n}X_i + \sum_{i=1}^{n}\bar{X}^2\right] && \text{総和をばらす}
\end{aligned}
$$

$$
\begin{aligned}
&= E\left[\sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right] && \text{総和の計算} \\
&= E\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] && \text{まとめる} \\
&= E\left[\sum_{i=1}^{n} X_i^2\right] - nE[\bar{X}^2] && \text{期待値をばらす} \\
&= \sum_{i=1}^{n} E\left[X_i^2\right] - nE[\bar{X}^2] && \text{期待値をばらす} \\
&= nE\left[X^2\right] - nE[\bar{X}^2] && \text{補題を適用} \\
&= n(V[X] + E[X]^2) - nE[\bar{X}^2] && \text{補題を適用} \\
&= n(V[X] + E[X]^2) - n(\frac{1}{n}V[X] + E[X]^2) && \text{補題を適用} \\
&= n(V[X] + E[X]^2) - (V[X] + nE[X]^2) && \\
&= nV[X] - V[X] = (n-1)V[X] && \text{まとめる} \\
&= (n-1)\sigma^2 && \text{仮定より}
\end{aligned}
$$

まとめると、

$$
E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = (n-1)\sigma^2
$$

であることが分かった。これを使えば《標本分散》の期待値は

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

となることが分かる。

「あれ、なんか変な値になったよ」

「n で割る自然な《標本分散》で全体の分散を推定すると、その平均は

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{n-1}{n}\sigma^2<\sigma^2$$

だから、必ず《全体》の分散 σ^2 よりも小さくなることが分かる」

「ほんとだ」

「推定量の平均値をピッタリ σ^2 に合わせるには、どうしたらいいと思う？」

「うーん、どうすればいいんだろ」青葉は首をひねった。

「《標本分散》の期待値に $(n-1)/n$ の逆数をかけてみよう。ここがおもしろいところだよ」

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

「あ、ピッタリ σ^2 になった」

「左辺は、どういう確率変数の期待値かな？」

「これ、$n-1$ で割ってるよ。…… つまり、《不偏分散》だ」

「そのとおり。不偏分散（確率変数）の期待値は《全体》の分散 σ^2 と一致する。つまり自然な《標本分散》を使って全体の分散を推定すると、平均的には目標の値に一致しない。平均的に目標に一致させるには、推定量を《不偏分散》に修正する必要がある。これが n で割るとズレるけど、$n-1$ で割ると、ピッタリ一致する理由だよ。

さきほどの実験結果は

n で割った推定値の平均 : 1184.31
$n-1$ で割った推定値の平均 : 1279.28
本当の値 : 1287.11

だから、たしかに n で割った推定量は $n-1$ で割った推定量より小さくなっている」

「なるほどー」

「不偏分散のように、期待値が推定の目標である数値と一致する推定量を不偏推定量っていうんだよ。ちなみに不偏だからといって、よい推定量とは限らない。有効性と言って……」

「それはまた今度でいいよ」

「OK。ちょうど 30 分だし、このへんで説明を終わろう。じゃあ柿ピーでも食べようか」

花京院は柿の種の小袋を一つ開封した。

まとめ

Q 不偏分散はどうして $n-1$ で割るの？

A $n-1$ で割る推定量を使うと、その期待値が《全体》の分散と一致するからです。

- パラメータの推定に用いる確率変数（の関数）を推定量といいます。特に期待値が目的の数値と一致する推定量を不偏推定量と言います。
- 標本平均 $\frac{1}{n}\sum_{i=1}^{n} X_i$ や不偏分散 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ は不偏推定量の一種です。
- 確率変数（不偏推定量）としての不偏分散は分布を持っています。標本のデータから計算した具体的な不偏推定値は、不偏推定量の実現値と考えることができます。