黒木玄 Gen Kuroki
@genkuroki

mathtod.online/@h_okumura/1186…

分散の計算で必ず n で割り、$n-1$ で割る話をしないのか。それだと、初歩的な分散の推定で一番面白い部分をネグっている感じ。

不偏分散の話は始めて聞くと非常に面白い話だと思う。

私は表現論をちょっと知っているので、不偏分散の話は

「$\mathrm{gl}(n)$ と $\mathrm{sl}(n)$ のCartan部分代数の違いと同じ」

と理解しています。

$\mathrm{gl}(n)$ のCartan部分代数のc正規直交基底 E_{ii} をトレースが0の対角行列に射影すると

$$E_{ii} - M,$$
$$M = \frac{1}{n}\sum_{i=1}^n E_{ii}$$

になります。 E_{ij} は (i,j) 成分だけが 1 で他が 0 の行列(所謂行列単位)。

2017年05月16日 14:11 · Web · 1 · 6 · Webで開く

黒木玄 Gen Kuroki @genkuroki on May 16
続き

$$\sum_{i=1}^n \mathrm{trace}(E_{ii}^2) = n$$

で、

$$M = \frac{1}{n}\sum_{i=1}^n E_{ii} = \frac{1}{n}E$$

のとき

$$\sum_{i=1}^n \mathrm{trace}((E_{ii}-M)^2) = n-1$$

です。後者は

$$\sum_{i=1}^{n}(E_{ii}-M)^2 = \frac{n-1}{n}E$$

のトレースなので $n-1$ になります。

ここで計算して出て来た n と $n-1$ の違いは、「不偏分散ではどうして $n-1$ で割るのか」の話と本質的に同じ構造になっています。

黒木玄 **Gen Kuroki** @genkuroki on May 16

E_{ii} を $E_{ii}-(1/n)E$ に射影することによって、長さ(ノルム)の二乗が $1-1/n=(n-1)/n$ 倍に小さくなることと、不偏分散で $n-1$ で割る話は本質的に同じです。

本当は、行列の話にしなくてもよくて、$\mathbb{R}^n$ の標準基底の σ 倍 σe_i を座標の総和が 0 の部分空間に射影する様子を想像しても良いです。 σe_i のノルムの2乗は σ^2 で、射影した結果のノルムの2乗は $((n-1)/n)\sigma^2$ になります。

黒木玄 **Gen Kuroki** @genkuroki on May 16

$E[XY]$ は X と Y の内積だと思ってよいという話がわかっていれば、ベクトルに関する視覚的なイメージに不偏分散では $n-1$ で割らなければいけない理由が帰着してします。

わけのわからない理由で $n-1$ で割るのではなく、直観的イメージに基いて当然そうなるべきだということがわかる。

n 次元は絵が描けないので 3 次元くらいの絵を描いて納得するしかないのですが。

統計学の易しい教科書は線形代数を知らなくてもよいように書かれているせいで、結果的にわけのわからない計算に帰着してしまうケースが多いように思えます。

黒木玄 **Gen Kuroki** @genkuroki on May 16

もう一度、強調しておきますが、$E[X]$ や $E[XY]$ は推定量ではなく、母集団分布で決まる量です。

そして、X_k が独立同分布な確率変数列のとき、標本平均を

$$M=\frac{X_1+\cdots+X_n}{n}$$

と書いたときの

$$S^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-M)^2$$

は分散の推定量(所謂不偏分散)です。

推定量である標本平均 M と不偏分散 S^2 が確率変数になっていることに注意して下さい。

確率変数として扱う理由は統計学ではサンプルを取り直すごとに推定量がゆらぐからです。そのゆらぎがどうなっているかを理解しないと、推定量がどれだけ信頼できるかわからなくなってしまいます。

$E[M]$ や $E[S^2]$ は母集団分布で決まる量になります。それらはそれぞれ母集団の平均と分散に一致します。 S^2 の定義で n ではなく、$n-1$ で割らなければいけないという話は最初聞いたときには誰もが「あれ？それはどうしてだ？」と興味を持つような話題だと思います。

黒木玄 **Gen Kuroki** @genkuroki on May 16

線形代数と不偏分散の関係への補足

ポイントは「自由度」がどうして一つ減って n から $n-1$ になるかです。

定理：$e_1,\dots,e_n$ は n 次元実ユークリッド空間 V の正規直交基底であり、W は V の m 次元部分空間であるとする。このとき、e_i の W への直交射影を $\tilde{e}_i$ と書くと、

$$\sum_{i=1}^n ||\tilde{e}_i||^2 = m.$$

不偏分散のケースはこの定理の $m=n-1$ の場合になっています。

「自由度」が減る仕組みは直交射影。

黒木玄 **Gen Kuroki** @genkuroki on May 16

続き

証明：W の正規直交基底 $w_1,\dots,w_m$ を取ると, e_i の W への直交射影 $\tilde{e}_i$ は

$$\tilde{e}_i = \sum_{k=1}^m (w_k, e_i) w_k$$

と書ける. ゆえに

$$\begin{aligned}
&\sum_{i=1}^n ||\tilde{e}_i||^2 \\
&= \sum_{i=1}^n \sum_{k=1}^m (w_k, e_i)^2 \\
&= \sum_{k=1}^m ||w_k||^2 = m.
\end{aligned}$$

2つ目の等号で $e_1,\dots,e_n$ が全体の空間 V の正規直交基底であることを使った. q.e.d.

より一般に

$$\sum_{i=1}^n (a, e_i)(e_i, b) = (a, b)$$

が成立している。

量子力学の教科書には

$$\sum_{i=1}^{n} |e_i\rangle\langle e_i| = 1$$

のように書いてあることが多い.

黒木玄 **Gen Kuroki** @genkuroki on May 16

続き

先の線形代数における初歩的結果を独立同分布確率変数達 $X_1, \dots, X_n$ に適用するとどうなるでしょうか?

簡単のため X_k の平均は 0 で分散は $\sigma^2 = 1$ に標準化されている場合を考えます。一般の場合はその場合に帰着します。このとき、

$$E[X_i X_j] = \delta_{ij}$$

なので、X_i 達はそれらで張られる実ベクトル空間 V の正規直交基底になっています。内積は $E[XY]$ で与えておきます。

標本平均 M を

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

と定め、V の $n-1$ 次元部分空間 W を M と直交する V のベクトル全体の空間と定めます。このとき、X_i の W への直交射影は $X_i - M$ になります。実際、$E[MX_i]$ も $E[MM]$ も $1/n$ になるので、それらの差を取ることによって、$E[M(E_i - M)] = 0$ となり、$E_i - M \in W$ となることがわかります。

続く

黒木玄 **Gen Kuroki** @genkuroki on May 16

続く

不偏分散 U^2 の定義は

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

なのですが、先の定理より、右辺の和の部分の期待値は W の次元の $n-1$ になります。ゆえに

$$E[U^2] = 1 = \sigma^2.$$

すなわち、不偏分散の期待値は母集団分布の分散に一致します($\sigma^2 \neq 1$ の場合も同様)。これで $n-1$ で割って不偏分散を定義するメリットが判明しました。

黒木玄 **Gen Kuroki** @genkuroki on May 16

まるで教科書だな。

黒木玄 **Gen Kuroki** @genkuroki on May 20
続き。

不偏分散ではなく、不偏共分散でも同じことになることも説明しておきます。正規直交基底をすべて双対基底に置き換えて議論するだけ。

定理: V は体 K 上の n 次元ベクトル空間であるとし, その双対空間を V^* と表わす. M, N はそれぞれ V, V^* の m 次元部分空間であり, $M \times N$ へのペアリングの制限は非退化であると仮定する. このとき直和分解

$$V = M \oplus N^\perp,$$
$$V^* = N \oplus M^\perp$$

が成立している. この直和分解に関する $x \in V$ の $N^\perp$ への射影を $p(x)$ と書き, $y \in V^*$ の $M^\perp$ への射影を $q(y)$ と書くことにする. このとき V, V^* の双対基底 x_i, y_i について

$$\sum_{i=1}^n \langle p(x_i), q(y_i) \rangle = n - m.$$

すなわち「自由度」が m 減る.

黒木玄 **Gen Kuroki** @genkuroki on May 20
ここまでの抽象化に耐えられる人は楽々証明できると思うので証明は略. 双対基底について

$$\sum_{i=1}^n |x_i\rangle\langle y_i| = 1$$

が成立することを使うだけ.

$E[X_i Y_j] = \delta_{ij}$ のとき

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n},$$
$$\overline{Y} = \frac{Y_1 + \cdots + Y_n}{n}$$

とおくと, $E[\overline{X}\,\overline{Y}] = 1/n \neq 0$ なので,

$$V = \sum_{i=1}^n \mathbb{R} X_i,$$
$$V^* = \sum_{i=1}^n \mathbb{R} Y_i,$$
$$M = \mathbb{R}\overline{X},$$
$$N = \mathbb{R}\overline{Y}$$

とおくと, $m = 1$ の場合の上の定理を使えて

$$E\left[\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})\right] = n - 1.$$

直接計算した方が易しいですね. 😂

mathtod.online powered by Mastodon