12.2 共分散行列と同時確率

線形代数はM個の異なる実験を一度に行うときに登場する。N 人について年齢と身長と体重をM=3 回測るとする。各実験はそれぞれについて平均値を計算できる。そこで、M 個の平均値を含むベクトル$m=(m_1,m_2,m_3)$を考える。これらは年齢と身長と体重の標本平均かもしれないし、既知の確率に基づく年齢、身長、体重の期待値かもしれない。

行列は分散の計算に用いる。各実験について、標本分散S_i^2、または平均からの距離の 2 乗に基づく期待値$\sigma_i^2=E[(x_i-m_i)^2$が計算できる。これらのM個の数値 $\sigma_i^2,\cdots,\sigma_m^2$ は「分散-共分散行列」の主対角線上に位置している。ここまでで、M個の並列実験の間には何の関係もない。これらは異なる確率変数を測定しているが、実験は必ずしも独立ではない。

子供の年齢(a)と身長(h)と体重(w)を測定すると、その結果には強い相関がある。一般に年長の子どもは身長が高く、体重も重い。平均値m_a,m_hは既知であるとする。そして、$\sigma_a^2,\sigma_h^2,\sigma_w^2$は年齢、身長、体重の分散である。$\sigma_{ah}$は、年齢と身長の関連を測定する共分散である。

$$\boxed{\textbf{Covariance}\quad \boldsymbol{\sigma_{ah}} = \mathbf{E}\,[(\textbf{age} - \textbf{mean age})\ (\textbf{height} - \textbf{mean height})].} \qquad (1)$$

この定義はよく見ておく必要がある。σ_{ah} を計算するためには、各年齢の確率と各身長の確率を知るだけでは不十分である。各組（年齢と身長）の結合確率を知らなければならない。これは、年齢と身長が結びついているからである。

p_{ah}= 年齢が a で身長が h の無作為な子供の確率
p_{ij}=実験 1 の結果がx_iで、実験 2 の結果がy_jの確率

実験 1（年齢）の平均がm_1であったとする。実験 2（身長）の平均は m_2 である。実験 1 と実験 2 の間の式(1)の共分散は、年齢x_iと身長y_jのすべての組に注目する。そのペアの同時確率p_{ij}を掛ける。

Expected value of
$(x-m_1)(y-m_2)$

$$\boxed{\textbf{Covariance}\quad \boldsymbol{\sigma_{12}} = \sum_{\textbf{all}\ i,j}\sum \boldsymbol{p_{ij}(x_i-m_1)(y_j-m_2)}} \qquad (2)$$

この「同時確率p_{ij}」を計算するために、まず、例を 2 つ見てみる。

例 1. 2 枚のコインを別々に投げる。1 が表で 0 が裏の場合、結果は(1,1)か(1,0)か(0,1)か(0,0)のどれかである。この 4 つの結果は、すべて確率$(1/2)^2=1/4$となる。
独立した実験では、確率を掛け合わせる。
p_{ij}= (i,j)の確率 =(iの確率)×(jの確率)

例 2 コインを同じ向きで接着する。結果は(1,1)と(0,0)だけである。これらの確率は 1/2 と 1/2 である。確率p_{10}とp_{01}は 0 である。(1,0)と(0,1)は、コインが接着されているので起こらない：表が出るか裏が出るかのどちらかしかない。

$$\textbf{Joint probability matrices for Examples 1 and 2} \quad \boldsymbol{P} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \quad \text{and} \quad \boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

P をうまく行列表記にするために、もう少し詳しく説明しよう。この行列は、(x_1, y_1)=(head, heads) と (x_1, y_2))=(heads, tails) から始まる各対 (x_i, y_j)) の確率 p_{ij} を示している。行の和p_1, p_2、列の和P_1, P_2、そして総和=1 であることに注目しよう。

$$\textbf{Probability matrix} \quad P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad \begin{matrix} p_{11} + p_{12} = \boldsymbol{p_1} \\ p_{21} + p_{22} = \boldsymbol{p_2} \end{matrix} \begin{pmatrix} \text{first} \\ \text{coin} \end{pmatrix}$$

$$\text{(second coin) column sums} \quad \boldsymbol{P_1} \quad \boldsymbol{P_2} \qquad 4 \text{ entries add to } 1$$

それらの和p_1, p_2, P_1, P_2が同時確率行列Pの周辺分布である。

$p_1 = p_{11} + p_{12}$ =コイン 1 で表が出る確率(コイン 2 は表か裏が出る)
$P_1 = p_{11} + p_{21}$ =コイン 2 で表が出
る確率(コイン 1 は表か裏が出る)

例 1 は、独立した確率変数である。すべての確率p_{ij}は$p_i \times P_j$に等しい(この例では 1/2×1/2 でp_{ij}=1/4 となる)。この場合、共分散σ_{12}はゼロとなる。1 枚目のコインの表か裏かは、2 枚目のコインの結果とは関係ない。

$$\boxed{\textbf{Zero covariance } \boldsymbol{\sigma_{12}} \textbf{ for independent trials} \quad \boldsymbol{V} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \textbf{diagonal covariance matrix } \boldsymbol{V}.}$$

独立した実験では、式 2) において、すべてのp_{ij}が$p_i \times P_j$に等しいので、σ_{12}=0 となる。

$$\sigma_{12} = \sum_i \sum_j (p_i)(p_j)(x_i - m_1)(y_j - m_2) = \left[\sum_i (p_i)(x_i - m_1)\right]\left[\sum_j (p_j)(y_j - m_2)\right] = [\mathbf{0}][\mathbf{0}].$$

$$\sigma_{12} = [p_1(x_1 - m_1) + p_2(x_2 - m_1)][P_1(y_1 - m_2) + P_2(y_2 - m_2)] =$$

$$\left[\frac{1}{2}\cdot\left(1 - \frac{1}{2}\right) + \frac{1}{2}\cdot\left(0 - \frac{1}{2}\right)\right]\left[\frac{1}{2}\cdot\left(1 - \frac{1}{2}\right) + \frac{1}{2}\cdot\left(0 - \frac{1}{2}\right)\right] = \left[\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\left(-\frac{1}{2}\right)\right]\left[\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\left(-\frac{1}{2}\right)\right] = 0 \cdot 0$$

$$\sigma_{11} = [p_1(x_1 - m_1) + p_2(x_2 - m_1)][p_1(y_1 - m_2) + p_2(y_2 - m_1)]$$
$$= p_1(x_1 - m_1)(P_1|p_1)(y_1 - m_2) + p_2(x_2 - m_1)(P_1|p_2)(y_1 - m_2) + p_1(x_1 - m_1)(P_2|p_1)(y_2 - m_2) + p_2(x_2 - m_1)(P_2|p_2)(y_2 - m_2)$$
$$= p_1(x_1 - m_1)1(y_1 - m_2) + p_2(x_2 - m_1)0(y_1 - m_2) + p_1(x_1 - m_1)0(y_2 - m_2) + p_2(x_2 - m_1)1(y_2 - m_2)$$
$$= \left[\frac{1}{2}\cdot\left(1 - \frac{1}{2}\right)^2 + \frac{1}{2}\cdot\left(0 - \frac{1}{2}\right)^2\right] = \left[\frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2}\cdot\frac{1}{4}\right] = 1/4$$

例 3 接着されたコインは、完全な相関関係を示している。一方が表ならもう一方も表である。共分散σ_{12}は 0 から$\sigma_1 \times \sigma_2$倍まで動く。これはσ_{12}の取りうる最大の値である。ここでは、(1/2)(1/2)= σ_{12}=(1/4)であることが、別の計算で確認された。

$$\textbf{Means} = \frac{\mathbf{1}}{\mathbf{2}} \qquad \boldsymbol{\sigma_{12}} = \frac{\mathbf{1}}{\mathbf{2}}\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{2}\right) + \mathbf{0} + \mathbf{0} + \frac{\mathbf{1}}{\mathbf{2}}\left(0 - \frac{1}{2}\right)\left(0 - \frac{1}{2}\right) = \frac{\mathbf{1}}{\mathbf{4}}$$

$$\sigma_{12} = p_1(x_1 - m_1)(P_1|p_1)(y_1 - m_2) + p_2(x_2 - m_1)(P_1|p_2)(y_1 - m_2) + p_1(x_1 - m_1)(P_2|p_1)(y_2 - m_2) + p_2(x_2 - m_1)(P_2|p_2)(y_2 - m_2)$$
$$= p_1(x_1 - m_1)(P_1|p_1)(y_1 - m_2) + p_2(x_2 - m_1)0(y_1 - m_2) + p_1(x_1 - m_1)0(y_2 - m_2) + p_2(x_2 - m_1)(P_2|p_2)(y_2 - m_2)$$
$$= p_1(x_1 - m_1)(P_1|p_1)(y_1 - m_2) + 0 + 0 + p_2(x_2 - m_1)(P_2|p_2)(y_2 - m_2) = \left[\frac{1}{2}\cdot\frac{1}{2}^2 + 0 + 0 + \frac{1}{2}\cdot\left(-\frac{1}{2}\right)^2\right] = \frac{1}{4}$$

コイン１の表と裏は、接着したコイン２の表と裏の完全な情報を提供します。

Glued coins give largest possible covariances
Singular covariance matrix: determinant = 0

$$V_{\text{glue}} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

常に$\sigma_1^2\sigma_2^2 \geq (\sigma_{12})^2$.。したがって、$\sigma_{12}$は -$\sigma_1\sigma_2$と$\sigma_1\sigma_2$の間にある。行列$V$は正定値（この特異なコインの糊付けの場合、V は正半定値）である。これらは、M 個の実験に対する$M \times M$個の共分散行列V すべてに関する重要な事実である。

N 回の実験から得られた標本共分散行列Sは確実に半正定値であることに注意してください。すべての新しい標本 X=(age, height, weight)は標本平均$\bar{X}$(ベクトル)に寄与しています。各ランク１項$(X_i - \bar{X})(X_i - \bar{X})^T$は正の半正定値であり、我々はただ行列 S に到達するために追加します。
S には確率はなく、実際の結果のみです。

$$\boxed{\overline{\boldsymbol{X}} = \frac{X_1 + \cdots + X_N}{N} \qquad \boldsymbol{S} = \frac{(X_1 - \overline{X})(X_1 - \overline{X})^{\mathrm{T}} + \cdots + (X_N - \overline{X})(X_N - \overline{X})^{\mathrm{T}}}{N-1}} \quad (3)$$