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入る。年齢と身長と体重を測るかもしれない（N 人についてM=3 回測る）。各実験はそれぞれ平均値を持っている。そこで、M 個の平均値を含むベクトル m=(m1,m2,m3)を用意する。これらは年齢と身長・体重の標本平均かもしれません。あるいは、m1,m2,m3 は、既知の確率に基づく年齢、身長、体重の期待値である可能性もあります。

行列は分散を見るときに関わってきます。各実験は、標本分散 Si2、または平均からの距離の 2 乗に基づく期待値 σi2=E[(xi-mi)2]を持つことになります。これらの M 個の数値 σ12,...,σm2 は「分散-共分散行列」の主対角線上に位置します。ここまでで、M 個の並列実験の間には何の関係もありません。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していない。

子供の年齢と身長・体重 (a,h,w) を測定すると、その結果は強い相関がある。一般に年長の子どもは身長が高く、体重も重い。平均値 ma, mh, mw が既知であるとする。そして、σa2, σ_h2, σaw2 が年齢、身長、体重の個別分散である。新しい数値は、年齢と身長の関連を測定する共分散 σah2 である。

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

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す。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 ＝a と身長 ＝h を持つランダムな子供の確率：両方を同時に持つ

p_{ij}=実験 1 が xi を生成し、実験 2 が yi を生成する確率

実験 1（年齢）の平均が m1 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 xi と身長 yj のすべての組に注目する。そのペアの共同確率 pij を掛ける。

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

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

この「共同確率 pij」の考え方を計算するために、まず、2 つの小さな例から始める。
例 1 2 枚のコインを別々に裏返す。1 が表で 0 が裏の場合、結果は(1,1)か(1,0)か(0,1)か(0,0)になることがある。
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、確率を掛け合わせます。
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します。可能性があるのは(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 p10 と p01 は 0 である。(1,0)と(0,1)は、コインがくっつくので起こりません：頭か尾の両方があります。

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

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、(x1,y1)=(head, heads) と (x1,y2)=(heads, tails) から始まる各対 (xi,yj) の確率 pij を示しています。行の和 p1,p2、列の和 P1,P2、そして総和=1 であることに注目しよう。

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

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

それらの和 p1,p2、P1,P2 が結合確率行列 P の周辺分布である。

p1=p11+p12=コイン 1 から表が出る確率(コイン 2 は表か裏が出る)
P1=p11+p21=コイン 2 から表が出る確率（コイン 1 は表にも裏にもなる）

例 1 では､独立した確率変数を示しました｡すべての確率 pij は π×pj に等しい(この例では 1/2×1/2 で pij=1/4 となる)。この場合、共分散シグマ 12 はゼロになります。1 枚目のコインの表か裏かは、2 枚目のコインについての情報を与えない。

Zero covariance σ_{12} for independent trials	$V=\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}$	= **diagonal covariance matrix V.**

独立した実験では、式 2）において、すべての pij が(pi)(pj)に等しいので、sighma12=0 となる。

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

例 3 接着されたコインは､完全な相関関係を示しています｡一方が表ならもう一方も表である｡共分散シグマ 12 は 0 からシグマ 1 のシグマ 2 倍まで動く。これはシグマ 12 の取りうる最大の値である。ここでは、(1/2)(1/2)=σ12=(1/4)であることが、別の計算で確認された。

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

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、σ12 は -σ1 σ2 と σ1 σ2 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×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(Xi-barX)(Xi-barX)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。

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