平均と分散

確率変数に対する極めて重要な２つの数は，**平均** m と **分散** σ^2 である.「期待値」$\mathrm{E}[e]$ は，とりうる誤差 $e_1, e_2, \ldots$ の確率 $p_1, p_2, \ldots$ から求められる（分散 σ^2 は，常に平均からの差として計算される）.

離散的な確率変数に対して，誤差 e_j が確率 p_j で起こるとする（p_j の和は１である）:

$$\textbf{平均}\ \boldsymbol{m} = \mathrm{E}[e] = \sum e_j p_j \qquad \textbf{分散}\ \sigma^2 = \mathrm{E}[(e-m)^2] = \sum (e_j - m)^2 p_j \tag{3}$$

例 1　公正なコインを投げる．その結果は，１（表）か０（裏）である．それらの事象は等しい確率 $p_0 = p_1 = 1/2$ で起こる．その平均は $m = 1/2$ であり，分散は $\sigma^2 = 1/4$ である：

$$\text{平均} = (0)\frac{1}{2} + (1)\frac{1}{2} \qquad \text{分散} = \left(0 - \frac{1}{2}\right)^2 \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \frac{1}{2} = \frac{1}{4}.$$

例 2　**（二項分布）** 公正なコインを N 回投げ，表の出た数を数える．３回投げたとき，表が出る回数は $M = 0, 1, 2,$ または３回である．その確率は $1/8, 3/8, 3/8, 1/8$ である．表が $M = 2$ 回出るのは表表裏，表裏表，裏表表の３通りあるが，表が $M = 3$ 回出るのは表表表しかない.

任意の N について表が M 回出る場合の数は，二項係数「N **から** M **を選ぶ組合せ**」である．すべてのとりうる結果の総数 2^N で割ることで，各 M に対する確率を得る：

$$\begin{matrix}\textbf{N 回投げて} \\ \textbf{M 回表が出る}\end{matrix} \qquad \boldsymbol{p_M} = \frac{1}{2^N}\binom{N}{M} = \frac{1}{2^N}\,\frac{N!}{M!(N-M)!} \qquad \text{確認}\ \frac{1}{2^3}\,\frac{3!}{2!\,1!} = \frac{3}{8} \tag{4}$$

ギャンブルをする人は，無意識のうちにこのことを知っている．確率 p_M の和は $\left(\frac{1}{2} + \frac{1}{2}\right)^N = 1$ である．表が出る回数の平均は $m = N/2$ であり，分散は $\sigma^2 = N/4$ となる．標準偏差 $\sigma = \sqrt{N}/2$ は，平均のまわりの想定される広がりを示す.

二項分布の係数をn枚の札の並べ方として考えてみましょう。最初の札はn枚の札の中から一枚を選ぶので、その選び方はn通りあります。つぎに２番目の札は一枚をすでに使ってしまっているので$n-1$枚の札の中から一枚を選びます。その選び方は$n-1$通りあります。３番目の札も同様に考えるとその選び方は$n-2$通りあります。このように続けていくと最後には一枚の札が残りその選び方は１通りになります。つまり札の並べ方は$n\cdot(n-1)\cdot\cdots\cdot 2\cdot 1$通りとなります。これを$n!$で表します。札は赤と青に色付けされていて、その並べ方を数えるとすると、この数え方では赤と青の札の並べ方を重複して数えてしまっています。そこで赤の札をk枚、青の札を$n-k$枚とすると、赤のk枚分の札と青の$n-k$枚分の札はそれぞれ$k!$通りと$(n-k)!$通り重複して数えてしまっています。それらを調整すると赤と青の札の並べ方は$\frac{n!}{k!(n-k)!}$通りとなります。よって、２項分布のpとnについて確率質量関数は

$$f(x) = {}_nC_x p^x (1-P)^{(n-x)} = \frac{n!}{k!\,(n-x)!} p^x (1-p)^{(n-x)}$$

となります。この分布を用いて、誤差の平均と分散を求めます。

$$E[X] = \sum_{x=0}^{n} x \cdot \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} (1-p)^{n-x} = np$$

$$V[X] = E[X(X-1)] + E[X] - \{E[X]\}^2$$

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1) \cdot \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x} = n(n-1)p^2 \sum_{x=2}^{n-2} \frac{(n-2)!}{(x-2)!\,(n-x)!} p^{x-2} (1-p)^{n-x}$$
$$= n(n-1)p^2$$

$$V[X] = E[X(X-1)] + E[X] - \{E[X]\}^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

例 3 （**ポアソン分布**）非常に偏ったコイン（$p << \frac{1}{2}$ と小さい）を何度も投げる（N が大きい）．**積** $\lambda = pN$ **を不変とする**．毎回，裏の確率は高く，$1-p$ である．したがって，N 回投げて表が 1 回も出ない（毎回裏が出る）確率 p_0 は，$(1-p)^N = (1-\lambda/N)^N$ である．N を大きくすると，これは $e^{-\lambda}$ へと近づく．非常に偏ったコインを N 回投げて j 回表が出る確率 p_j は，$\lambda = pN$ を用いた式となる：

ポアソン確率 $\quad p_j = \dfrac{\lambda^j}{j!}e^{-\lambda} \quad$ **平均** $m = \lambda \quad$ **分散** $\sigma^2 = \lambda \qquad (5)$

ポアソンは，めったに起きない（低い確率 p）事象を，長い時間 T にわたって数えるのに応用した．そのとき，$\lambda = pT$ である．

$$P(X=x) = {}_nC_x\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-x}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}\frac{\lambda^x}{x!}$$

ここで$\lim_{k\to\infty}\left(1+\frac{a}{k}\right)^k = e^a$を用いた。

つぎに平均と分散は

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

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

$$\mathrm{E[V]} = E[X(X-1)] + E[X] - (E[X])^2 = \lambda$$

連続確率変数では，$p_1, p_2, \ldots$ の代わりに確率**密度**関数 $p(x)$ をとる．「x と $x+dx$ の間の事象が起こる確率は $p(x)\,dx$ である」．確率の総和は $\int p(x)\,dx = 1$ であり，いずれかの事象が必ず起きる．和が積分に変わる：

$$\textbf{平均 } m = \textbf{期待値} = \int xp(x)\,dx \qquad \textbf{分散 } \sigma^2 = \int (x-m)^2 p(x)\,dx. \qquad (6)$$

確率密度関数 $p(x)$ (pdf と呼ばれる) の極めて重要な例は**正規分布** $\mathbf{N}(0,\sigma)$ である．対称性よりその平均は零である．その分散は σ^2 である：

正規（ガウス） $\quad p(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma}\,e^{-x^2/2\sigma^2} \quad$ ただし $\quad \displaystyle\int_{-\infty}^{\infty} p(x)\,dx = 1. \qquad (7)$

$p(x)$ のグラフは，有名な釣鐘型の曲線である．$-\sigma$ から σ までの $p(x)$ の積分は，標本値の平均からの距離が標準偏差 σ より小さくなる確率である．これはおよそ 2/3 である．MATLAB の **randn** では，$\sigma = 1$ の正規分布が用いられる．

中心極限定理により，この正規分布 $p(x)$ はいたるところに現れる．中心極限定理は，ある確率分布（例えば二項分布）を持つ多数の独立な試行の平均が，$N \to \infty$ としたときに正規分布に近づく，というものである．変位することで $m=0$ とでき，拡大縮小することで $\sigma = 1$ とできる．

$$\textbf{正規化された，表の出る数} \quad x = \frac{M - \text{平均}}{\sigma} = \frac{M - N/2}{\sqrt{N}/2} \longrightarrow \text{正規分布 } \mathbf{N}(0,1).$$

共分散行列

ここで，異なる m 個の実験を同時に行う．それらは独立かもしれないし，それらの間に相関があるかもしれない．各計測 $\boldsymbol{b}$ は，m 個の要素からなるベクトルとなる．それらの要素は m 個の実験の結果 b_i である．

平均 m_i からの距離をとると，各誤差 $e_i = b_i - m_i$ の**平均は零**である．2 つの誤差 e_i と e_j が**独立**（それらの間に相関がない）ならば，それらの積 e_ie_j も平均が零となる．しかし，それらの計測がほぼ同じ時間に同じ観測者によってなされたならば，それらの誤差 e_i と e_j は同じ符号や同じ大きさを持つ傾向がある．**m 個の実験の誤差は相関を持ちうる**．積 e_ie_j を（それらの確率）p_{ij} で重みづけすると**共分散** $\sigma_{ij} = \sum\sum p_{ij}e_ie_j$ になる．$e_i^2 p_{ii}$ の和は分散 σ_i^2 である：

共分散 $\quad \boldsymbol{\sigma_{ij} = \sigma_{ji} = \mathrm{E}[e_ie_j] = (e_i \times e_j)}$ **の期待値．** (8)

これが，**共分散行列** $\boldsymbol{\Sigma}$ の (i,j) 要素と (j,i) 要素である．(i,i) 要素は $\sigma_{ii} = \sigma_i^2$ である．

例 4 （多変量正規分布） m 個の確率変数では，確率分布関数は $p(x)$ から $p(\boldsymbol{b}) = p(b_1, \ldots, b_m)$ になる．平均が零である正規分布は，ある正の数 σ^2 によって定まっていた．$p(\boldsymbol{b})$ は，$m \times m$ の正定値行列 $\boldsymbol{\Sigma}$ によって定まる．これは共分散行列であり，その行列式は $|\boldsymbol{\Sigma}|$ である：

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-x^2/2\sigma^2} \quad \text{が} \quad p(\boldsymbol{b}) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}}\, e^{-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}/2} \quad \text{となる}$$

$p(\boldsymbol{b})$ の m 次元空間全体での積分は 1 である．$\boldsymbol{bb}^{\mathrm{T}}p(\boldsymbol{b})$ の積分は $\boldsymbol{\Sigma}$ である．

指数 $-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}/2$ を扱うには，$\boldsymbol{\Sigma}$ の固有値と正規直交する固有ベクトルを用いるのがよい（**ここで線形代数が登場する**）．$\boldsymbol{\Sigma} = Q\Lambda Q^{\mathrm{T}} = Q\Lambda Q^{-1}$ であるとき，$\boldsymbol{b}$ を $Q\boldsymbol{c}$ で置き換えることで $p(\boldsymbol{b})$ を m 個の 1 次元正規分布に分ける：

$$\exp\left(-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}/2\right) = \exp\left(-\boldsymbol{c}^{\mathrm{T}}\Lambda^{-1}\boldsymbol{c}/2\right) = \left(e^{-c_1^2/2\lambda_1}\right)\cdots\left(e^{-c_m^2/2\lambda_m}\right).$$

行列式は $|\boldsymbol{\Sigma}|^{1/2} = |\Lambda|^{1/2} = (\lambda_1 \cdots \lambda_m)^{1/2}$ である．$-\infty < c_i < \infty$ における積分はいずれも 1 次元となり，ここで $\lambda = \sigma^2$ である．任意の線形変換を行っても（ここでは $\boldsymbol{c} = Q^{-1}\boldsymbol{b}$）多変量正規分布が維持されるという素晴らしい事実に注意せよ．

さらに，$\boldsymbol{b}$ から $\boldsymbol{z}$ への変換に $\sqrt{\Lambda}$ を含めることで，変数 $= 1$ とすることもできる：

標準正規 $\quad \boldsymbol{b} = \sqrt{\Lambda}Q\boldsymbol{z}$ により $p(\boldsymbol{b})d\boldsymbol{b}$ が $p(\boldsymbol{z})d\boldsymbol{z} = \dfrac{e^{-\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}/2}}{(2\pi)^{m/2}}d\boldsymbol{z}$ **に変わる**

適切な重み行列 W によって $A\boldsymbol{x} = \boldsymbol{b}$ が $WA\boldsymbol{x} = W\boldsymbol{b}$ に対する通常の最小 2 乗法となることがわかる．$W\boldsymbol{b}$ を標準正規分布 $\boldsymbol{z}$ としたいので，W は $\sqrt{\Lambda}Q$ の逆行列となる．それよりも大事なのは，$\boldsymbol{C = W^{\mathrm{T}}W}$ **が** $\boldsymbol{Q\Lambda Q^{\mathrm{T}}}$ **の逆行列であり，それが** $\boldsymbol{\Sigma}$ **であることだ．**

まとめ 誤差が独立のとき，$\boldsymbol{\Sigma}$ は対角行列 $\mathbf{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ である．普通はこうである．等式 $A\boldsymbol{x} = \boldsymbol{b}$ に対する適切な重み w_i は，$1/\sigma_1, \ldots, 1/\sigma_m$ である（これにより，すべての分散が 1 に等しくなる）．重み付き最小 2 乗法の中央に位置する行列 $C = W^{\mathrm{T}}W$ は，まさに $\boldsymbol{\Sigma}^{-1}$ である：

重み付き最小 2 乗法 $\quad A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}$ (9)

重みをこのように選ぶことで，$A\boldsymbol{x} = \boldsymbol{b}$ を信頼性が等しく誤差が独立な最小 2 乗法の問題 $WA\boldsymbol{x} = W\boldsymbol{b}$ にする．通常の等式 $(WA)^{\mathrm{T}}WA\widehat{\boldsymbol{x}} = (WA)^{\mathrm{T}}W\boldsymbol{b}$ は，式 (9) と同じである．

このバイアスのない最適な線形予測 $\widehat{\boldsymbol{x}}$ を発見したのはガウスである．$\widehat{\boldsymbol{x}}$ は $\boldsymbol{x}-\widehat{\boldsymbol{x}}$ の平均が零であることからバイアスがなく，式 (9) より線形であり，$\boldsymbol{x}-\widehat{\boldsymbol{x}}$ の共分散が最小であることから最適である．($\boldsymbol{b}$ の誤差ではなく，$\widehat{\boldsymbol{x}}$ の誤差に対する）共分散は重要である：

最適な $\widehat{\boldsymbol{x}}$ の共分散 $\quad P=\mathrm{E}\left[(\boldsymbol{x}-\widehat{\boldsymbol{x}})(\boldsymbol{x}-\widehat{\boldsymbol{x}})^{\mathrm{T}}\right]=\left(A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A\right)^{-1}. \qquad (10)$

例 5　独立で等しく信頼できる 10 人の医者が，あなたの心拍数を 10 回計測した．それぞれの b_i の平均誤差は零であり，それぞれの分散は σ^2 である．そのとき，$\boldsymbol{\Sigma}=\sigma^2 I$ である．10 個の等式 $x=b_i$ により，要素がすべて 1 である 10×1 の行列 A が 10 個作られる．最適な予測 $\widehat{x}$ は，10 個の b_i の平均である．**その平均値の分散 $\widehat{x}$ が次の数 P である**：

$$P=(A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A)^{-1}=\sigma^2/10 \qquad \text{したがって，平均をとることで分散が小さくなる．}$$

行列 $P=(A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A)^{-1}$ は，実験の結果 $\widehat{\boldsymbol{x}}$ にどのくらいの信頼性があるかを示す (問題 6). P は，実際の実験における b_i には依存しない．それらの b_i は確率分布を持つ．各実験によって，標本 $\boldsymbol{b}$ から $\widehat{\boldsymbol{x}}$ の値が得られる．

$\boldsymbol{\Sigma}$ が小ければ入力 $\boldsymbol{b}$ がより信頼できるのと同様に，P が小さければ出力 $\widehat{\boldsymbol{x}}$ がより信頼できる．重要な公式 $P=(A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A)^{-1}$ によって，それらの共分散が関連づけられる．

$$f_{\mathbf{X}}(x_1,\ldots,x_k)=\frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}$$

2変量の場合 [編集]

2次元で非退化の場合（*k* = rank(Σ) = 2）、ベクトル [*X Y*]'（右肩のダッシュは転置を表す）の確率密度関数は、

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)$$

となる。ここで *ρ* は *X* と *Y* の相関係数であり、$\sigma_X>0$ かつ $\sigma_Y>0$ である。このとき、

$$\boldsymbol{\mu}=\begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}$$

2次元のときは、多変量正規分布であるための同値な条件として挙げた最初の方は、やや緩められる：

> 可算無限通りの X と Y の線型結合がどれも正規分布に従うならば、ベクトル [X Y]' は2変量正規分布に従う[6]。

2変数の場合の等高線を *x*,*y*-平面にプロットすると楕円になる。相関係数 *ρ* が大きくなっていくとき、楕円は次の直線：

$$y(x)=\mathrm{sgn}(\rho)\frac{\sigma_Y}{\sigma_X}(x-\mu_X)+\mu_Y.$$

の方向に向かって押しつぶされていく。この背景として、この式の sgn(*ρ*) （"sgn" は符号関数）を *ρ* に取り換えたものは、*X* の値が与えられたときの *Y* の最良線形不偏予測量（英語版）（best linear unbiased prediction）になっているという性質がある[7]。

主成分分析

以降の段落で，データ行列 A から有益な情報を発見することについて示す．n 個の標本の m 個の性質（m 個の特徴）について計測したとする．それらは，n 人の学生の m の講義の成績（1つの行が各講義を表し，1つの列が各学生を表す）でもよい．各行について，その平均を引くことで，平均は零となる．データから最も多くの情報が得られるような，**講義の組合せ** および／または **学生の組合せ** を探そう．

情報とは「でたらめさからの距離」であり，それは**分散**によって測られる．講義の成績の分散が大きければ，分散が小さい場合よりもより多くの情報を持つ．

鍵となる行列の考え方は，特異値分解 $A = U\Sigma V^{\mathrm{T}}$ である．$A^{\mathrm{T}}A$ と AA^{T} に再び戻ろう．なぜならば，それらの単位固有ベクトルは，V の特異ベクトル $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ と U の特異ベクトル $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ であるからだ．対角行列 Σ（共分散ではない）の特異値は降順に並んでおり，σ_1 が最も重要である．m 個の講義を $\boldsymbol{u}_1$ の要素によって重みづけすると，その成績が最も重要な「重要講義」または「固有講義」が得られる．

例 6　成績 A, B, C, F の価値が $\mathbf{4, 2, 0, -6}$ 点であったとする．各講義および各学生がそれぞれ1つずつ成績をとったとき，全体の平均は零である．以下に成績行列 A を示す．その行列の零空間には $(1,1,1,1)$ が含まれる（階数は3である）．計算結果が整数となるよう，A の特異値分解を $2U$ と $\Sigma/4$ と $(2V)^{\mathrm{T}}$ の積とする．σ は $12, 8, 4$ である：

$$\begin{bmatrix} -6 & 2 & 0 & 4 \\ 0 & 4 & -6 & 2 \\ 4 & 0 & 2 & -6 \\ 2 & -6 & 4 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & & \\ & 2 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

行（講義）を $\boldsymbol{u}_1 = \frac{1}{2}(-1,-1,1,1)$ で重みづけすると，**固有講義**を得る．列（学生）を $\boldsymbol{v}_1 = \frac{1}{2}(1,-1,1,-1)$ で重みづけすると，**固有学生**を得る．成績行列のうち，その1つの講義と一人の学生によって「説明される」割合は $\sigma_1^2/(\sigma_1^2+\sigma_2^2+\sigma_3^2) = 9/14$ である．特異値分解における σ は，分散 σ^2 である．

この固有講義は入試委員長が探しているものだと思う．体育の成績がすべて同じであれば，A のその行はすべて零になり，体育は固有講義には含まれない．おそらく解析は固有講義に含まれるであろうが，解析をとらなかった学生はどうなるか？**欠損データ**（行列 A の穴）の問題は，社会科学や国勢調査や実験における統計で非常に難しい問題である．

遺伝子発現データ　遺伝子や遺伝子の組合せの機能を決定することは，遺伝学の中心的課題である．どの遺伝子の組合せがどの特性を与えるか？どの遺伝子の不具合がどの病気に影響するか？

今では，遺伝子発現データを信じられないほど高速に求めることができる．遺伝子のマイクロアレイは1つのアフィメトリクスチップ上に詰められ，1つの個体（一人の人）の数万の遺伝子を測定できる．遺伝的データの理解（**バイオインフォマティクス**）は，線形代数の大規模な応用となったのである．

練習問題 8.6

1　時刻 $t = 0, 1, 2$ における3つの独立な測定値 $1, 2, 4$ があり，その分散 $\sigma_1^2, \sigma_2^2, \sigma_3^2$ が $1, 1, 2$ であるとき，それらに最も近い直線 $Ct + D$ を求めよ．重み $w_i = 1/\sigma_i$ を用いよ．

この問題の重みの役割を理解するには、３つのはかりを用いて、おもりの重さを量ることを考えればよい。そして、はかりには信頼度があって、２つのはかりは同じ信頼度で１つは信頼度が他と比較して悪い場合を想定すればよい。

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix},\ A^T C A\hat{x} = A^T C b,$$

$$C = W^T W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$

$$\begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0.5\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\\1 & 2\end{bmatrix}\hat{x} = \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0.5\end{bmatrix}\begin{bmatrix}1\\2\\4\end{bmatrix}$$

$$\begin{bmatrix}2.5 & 2\\2 & 3\end{bmatrix}\begin{bmatrix}C\\D\end{bmatrix} = \begin{bmatrix}5\\6\end{bmatrix}, C = \frac{6}{7}, D = \frac{10}{7}$$

2 問題 1 において，3 番目の測定が**完全に信頼できない**と仮定する．分散 σ_3^2 が無限大になる．そのとき，最適な直線は ____ を用いない．1 番目と 2 番目の点を通る直線を求め，$\boldsymbol{Ax} = \boldsymbol{b}$ の最初の 2 つの等式を正確に解け．

３つ目のはかりの精度が非常に割るので、$W = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & \frac{1}{\infty}\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$

３番目の観測値は用いないので、

$A = \begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}, b = \begin{bmatrix}1\\2\end{bmatrix},\ A^T A\hat{x} = A^T b,$

$\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}\hat{x} = \begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\hat{x} = \begin{bmatrix}3\\2\end{bmatrix}$, C=1, D=1

3 問題 1 において，3 番目の測定が**完全に信頼できる**ものと仮定する．分散 σ_3^2 は零に近づく．すると，最適な直線は 3 番目の点を厳密に通る．最初の 2 つの誤差の 2 乗和を最小化するような直線を選べ．

３番目の点は確実に通るので、それを切片とするように解けばよい。または、$W = \begin{bmatrix}\frac{1}{\infty} & 0 & 0\\0 & \frac{1}{\infty} & 0\\0 & 0 & 1\end{bmatrix} = \begin{bmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{bmatrix}$

$b = Ct + D$とすると、３番目の点を必ず通るようにするには

$4 = 2C + D \therefore b - 4 = C(t-2)$

$A = \begin{bmatrix}-2\\-1\\0\end{bmatrix}, b = \begin{bmatrix}-3\\-2\\0\end{bmatrix},\ A^T A\hat{x} = A^T b,$

$[-2 \quad -1 \quad 0]\begin{bmatrix}-2\\-1\\0\end{bmatrix}\hat{x} = [-2 \quad -1 \quad 0]\begin{bmatrix}-3\\-2\\0\end{bmatrix}, 5\hat{x} = 8$, C=8/5, D=4/5

4 1 枚の公正なコインを投げる（0 または 1）とき，その平均は $m = 1/2$ であり分散は $\sigma^2 = 1/4$ である．これは例 1 であった．2 回投げたときの合計について，その平均は $m = 1$ である．結果 0,1,2 とその確率を用いて，その分散 σ^2 を計算せよ．

コインを２回投げたとき、結果が 0,1,2 となるのは、裏・裏、(表・裏、裏・表)、表・表の結果が出たときであり、これは n=2,p=0.5 としたときの二項分布なので、分散は

$$\frac{1}{4}(0-1)^2 + 2\cdot\frac{1}{4}(1-1)^2 + \frac{1}{4}(2-1)^2 = \frac{1}{2} = n(p(1-p))$$

と求められる。

5 投げた結果を足すのではなく，それらを2つの独立な実験とする．結果は $(0,0),(1,0),(0,1)$ または $(1,1)$ である．共分散行列 $\boldsymbol{\Sigma}$ を求めよ．

２つの独立な実験なので(実験 1 の結果、実験２の結果)とすると、実験１の結果は 1 と 0、実験 2 の結果は 1,0

よって、コインの表と裏が出る確率は同じなので、それぞれの

実験１の結果の平均は$m_1 = 1 \cdot 1/2 + 0 \cdot 1/2 = 1/2$

実験２の結果の平均は$m_2 = 1 \cdot 1/2 + 0 \cdot 1/2 = 1/2$

よってそれぞれの分散は

$$\sigma_1^2 = (1-1/2)^2 \cdot \frac{1}{2} + \left(0-\frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}$$

$$\sigma_2^2 = (1-1/2)^2 \cdot \frac{1}{2} + \left(0-\frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}$$

つぎに共分散は、

$$\sigma_{12} = \left(1-\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\left(1-\frac{1}{2}\right)\cdot\frac{1}{2} + \left(1-\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\left(0-\frac{1}{2}\right)\cdot\frac{1}{2} + \left(0-\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\left(1-\frac{1}{2}\right)\cdot\frac{1}{2} + \left(0-\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\left(0-\frac{1}{2}\right)\cdot\frac{1}{2}$$

$$=\frac{1}{16} - \frac{1}{16} - \frac{1}{16} + \frac{1}{16} = 0$$

よって、共分散行列は$\Sigma = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix}$

6 例 1 において，コインが公正でないとする．表の出る確率を p，裏の出る確率を $1-p$ とする．この分布の平均 m と分散 σ^2 を求めよ．

表が 1，裏が 0 であるから、

$m = 1 \cdot p + 0 \cdot (1-p) = p$

$\sigma^2 = (1-1+p)^2 \cdot p + (0-1+p)^2 \cdot (1-p) = (1-p)^2 \cdot p + p^2 \cdot (1-p) = p^3 - 2p^2 + p + p^2 - p^3 = (1-p)p$

または２項分布のn=1 の場合とすればよい。

7 2つの独立な測定 $x = b_1$ と $x = b_2$ に対して，最適値 $\widehat{x}$ はある重み付き平均 $\widehat{x} = ab_1 + (1-a)b_2$ である．b_1 と b_2 の平均が零であり分散が σ_1^2 と σ_2^2 であるとき，$\widehat{x}$ の分散は $P = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$ となる．$dP/da = 0$ により，$\boldsymbol{P}$ **を最小化する数** $\boldsymbol{a}$ **を求めよ**．重み $w_1 = 1/\sigma_1$ と $w_2 = 1/\sigma_2$ を用いることで，式 (2) の $\widehat{x}$ がこの a から得られることを示せ．$\widehat{x}$ は，本文中で最適だと主張したものである．

$\frac{dP}{da} = 2a\sigma_1^2 - 2\sigma_2^2 - 2a\sigma_2^2 = 0$より、 $a = \frac{\sigma_2^2}{(\sigma_1^2 - \sigma_2^2)}$

$$w_1^2 = \frac{\sigma_1^2}{(\sigma_1^2 - \sigma_2^2)} = 1 - a, \qquad w_2^2 = \frac{\sigma_2^2}{(\sigma_1^2 - \sigma_2^2)} = a$$

$$\hat{x} = \frac{w_1^2 b_1 + w_2^2 b_2}{w_1^2 + w_2^2} = \frac{\sigma_1^2 b_1 + \sigma_2^2 b_2}{(\sigma_1^2 + \sigma_2^2)}$$

$$w_1^2 b_1 + w_2^2 b_2 = \frac{\sigma_1^2 b_1}{(\sigma_1^2 - \sigma_2^2)} + \frac{\sigma_2^2 b_2}{(\sigma_1^2 - \sigma_2^2)} = \frac{\sigma_1^2 b_1 + \sigma_2^2 b_2}{(\sigma_1^2 - \sigma_2^2)}$$

$$\frac{w_1^2 b_1 + w_2^2 b_2}{w_1^2 + w_2^2} = \frac{\sigma_1^2 b_1 + \sigma_2^2 b_2}{(\sigma_1^2 + \sigma_2^2)}$$

重み付きの最小 2 乗法

m 個の等式 $A\boldsymbol{x}=\boldsymbol{b}$ に重みを付けるには，各 i 番目の等式に重み w_i を掛ける．それらの m 個の重みを対角行列 W に置く．$A\boldsymbol{x}=\boldsymbol{b}$ を $WA\boldsymbol{x}=W\boldsymbol{b}$ によって置き換えるということだ．その方程式に対して，最小 2 乗法を用いることにする．

最小 2 乗法の式 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ は，$(WA)^{\mathrm{T}}WA\widehat{\boldsymbol{x}}=(WA)^{\mathrm{T}}W\boldsymbol{b}$ に変わる．重み付きの最小 2 乗法の $(WA)^{\mathrm{T}}WA$ の中に，行列 $C=W^{\mathrm{T}}W$ がある．

重み付き最小 2 乗法	$\widehat{\boldsymbol{x}}$ に関する $\boldsymbol{n}$ 個の式	$A^{\mathrm{T}}CA\widehat{\boldsymbol{x}}=A^{\mathrm{T}}C\boldsymbol{b}$	の中に $\boldsymbol{C}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$ がある	(1)

$n=1$ および $A=1$ からなる列であるとき，$\widehat{\boldsymbol{x}}$ は平均から重み付き平均に変わる：

最も単純な場合　$\widehat{\boldsymbol{x}}=\dfrac{b_1+\cdots+b_m}{m}$ は $\widehat{\boldsymbol{x}}_W=\dfrac{w_1^2b_1+\cdots+w_m^2b_m}{w_1^2+\cdots+w_m^2}$ に変わる．　(2)

この平均 $\widehat{\boldsymbol{x}}_W$ では，最大の w_i を持つ観測値 b_i に最大の重みを与える．常に，誤差の平均は零であると仮定する（必要があれば平均を引く．したがって，測定値に一方的バイアスはない）．

重み w_i をどのように選べばよいか？ これは，b_i の信頼性に依存する．観測値の分散が σ_i^2 であれば，b_i の 2 乗平均平方根誤差は σ_i である．等式の両辺を $\sigma_1,\ldots,\sigma_m$ で割ると，すべ

よって、m=2 のとき、$w_1^2=\frac{\sigma_1^2}{(\sigma_1^2-\sigma_2^2)}=1-a,\ w_2^2=\frac{\sigma_2^2}{(\sigma_1^2-\sigma_2^2)}=a$

8　$\boldsymbol{\Sigma}^{-1}$ によって正しく重みづけされた最小 2 乗予測は，$\widehat{\boldsymbol{x}}=(A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A)^{-1}A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}$ である．それを $\widehat{\boldsymbol{x}}=L\boldsymbol{b}$ と呼ぶ．$\boldsymbol{b}$ に誤差ベクトル $\boldsymbol{e}$ が含まれるとき，$\widehat{\boldsymbol{x}}$ には誤差 $L\boldsymbol{e}$ が含まれる．出力の誤差 $L\boldsymbol{e}$ の共分散行列は，それらの期待値（平均値）$P=\mathrm{E}\left[(L\boldsymbol{e})(L\boldsymbol{e})^{\mathrm{T}}\right]=L\mathrm{E}\left[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\right]L^{\mathrm{T}}=L\boldsymbol{\Sigma}L^{\mathrm{T}}$ である．**問題：** 積 $L\boldsymbol{\Sigma}L^{\mathrm{T}}$ を計算して，式 (10) で予想されたように，その P が $(A^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}A)^{-1}$ と等しいことを示せ．

9　成績 A, B, C, F の価値を $3,1,-1,-3$ に変える．次の成績行列の特異値分解が，例 5 と同じ $\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{v}_1,\boldsymbol{v}_2$（同じ固有講義）を持つことを示せ．しかしこの場合，A の階数は 2 である．

成績行列　$$A=\begin{bmatrix}3&-1&1&-3\\-1&3&-3&1\\-3&1&-1&3\\1&-3&3&-1\end{bmatrix}$$

注　A の欠損要素を扱う 1 つの方法は，その行列の階数が最小となるように値を埋めることである．統計においては，擬似逆行列 A^+ をよく用いる．($A^{\mathrm{T}}A$ が可逆行列であるとき，その擬似逆行列は正規方程式から得られる左逆行列 $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ に等しい）．