Introduction to mathematical statistics ゼミ (第 1b，2b 回)

担当: 伊藤真道

未定
(おそらく講義しないので各自読んでおいてください)

1 Chap.1. Probability and Distribution

1.10 Important Inequalities / 重要な不等式

この節では，期待値に関するいくつかの有名な不等式について論ずる．ここで，論じた不等式は以降の章で利用されている．

定理 1.1. X を確率変数とし，m を正整数とし，$E[X^m]$ の存在を仮定する．もし，$k \leq m$ なる正整数 k が存在するならば，$E[X^k]$ が存在する．

証明 1.1. ここでは連続の場合についてのみ考える．$f(x)$ を X の確率密度関数とすると．

$$\begin{aligned}
E[X^k] \leq E[|X|^k] &= \int_{-\infty}^{\infty} |x|^k f(x)dx \\
&= \int_{|x|\leq 1} |x|^k f(x)dx + \int_{|x|>1} |x|^k f(x)dx \\
&\leq \int_{|x|\leq 1} f(x)dx + \int_{|x|>1} |x|^m f(x)dx \\
&\leq \int_{-\infty}^{\infty} f(x)dx + \int_{-\infty}^{\infty} |x|^m f(x)dx \\
&\leq 1 + E[|X|^m] < \infty
\end{aligned}$$

となるため，$E[X^k]$ は存在する

¡note¿

定理 1.10.2(マルコフの不等式)，定理 1.10.3(チェビシェフの不等式) に関しては，第 1 回の伊藤の資料を参照．

定義 1.1. 凸関数 *(convex function)*，狭義凸 *(strictly convex)* $-\infty \le a < b \le \infty$ なる区間 (a,b) 上で定義された関数が，凸関数であるとは，任意の $x, y \in (a,b), 0 < \gamma < 1$ に対して，

$$\phi[\gamma x + (1-\gamma)y] \le \gamma\phi(x) + (1-\gamma)\phi(y)$$

が成り立つことを言う．もし，上の不等式が，狭義 *(strict)* な大小関係 *(=左辺と右辺が”<”の大小関係となること)* ならば，関数 ϕ は狭義凸 *(strictly convex)* であると言う．

¡note¿

ϕ の一階微分，二階微分 (それぞれ ϕ', ϕ'' で表す) が存在するとすると，以下の定理が成り立つ

定理 1.2. もし ϕ が (a,b) 上で微分可能ならば，

1. 任意の $a < x < y < b$ に対して，$\phi'(x) \le \phi'(y)$ が成り立つとき，ϕ は凸である
2. 任意の $a < x < y < b$ に対して，$\phi'(x) < \phi'(y)$ が成り立つとき，ϕ は狭義凸である

さらに，ϕ が (a,b) 上で二階微分可能ならば，

1. $a < x < b$ なる任意の x について $\phi''(x) \ge 0$ が成り立つとき，ϕ は凸である
2. $a < x < b$ なる任意の x について $\phi''(x) > 0$ が成り立つとき，ϕ は狭義凸である

証明は，例えば Hewitt and Stromberg(1965) などを参照．この凸性から，大変重要な確率不等式が得られる．

¡note¿

定理 1.3. イェンセンの不等式 *(Jensen's Inequality)* もし，ϕ が，ある開区間 I 上で凸であり，なおかつ，X が有限の期待値を持ち，台 *(*=関数の値が *0* とならない点の集合，ここで関数とは確率変数のことを指す．*)* が区間 I に含まれているような確率変数であるとするならば，

$$\phi[E(X)] \leq E[\phi(X)]$$

となる．もし，ϕ が狭義凸ならば，X が定数確率変数でない限り，不等号は，狭義なものとなる．

証明 1.2. 以下の証明では，ϕ が二階の導関数を持つと仮定するが，一般的な証明では，凸性のみが仮定されている．$\phi(x)$ を $\mu = E[X]$ の周りで *2* 次までのテイラー展開を行うと，

$$\phi(x) = \phi(\mu) + \phi'(x)(x-\mu) + \frac{\phi''(\xi)(x-2)^2}{2}$$

となる．ここで，ξ は x と μ の間にあるとする．上の関数の右辺第三項は非負であるため，

$$\phi(x) \geq \phi(\mu) + \phi'(\mu)(x-\mu)$$

を得る．ここで，両辺の X に関する期待値をとることで，定理の不等式が得られる．この不等式は，任意の $x \in (a,b)$ について $\phi''(x) > 0$ であり，X が定数でないとき，狭義である．

¡note¿

例 1.4. 調和平均 *(Harmonic Mean)* と幾何平均 *(Geometric Mean)* $\{a_1, \ldots, a_n\}$ を正の数の集合とする．それぞれの $a_1, \ldots, a_n$ に対して，重み $1/n$*(*=確率のようなものと思ってください*)* を与えることで，確率変数 X に対する分布を構成する．この場合，X の平均は算術平均 *(arithmetic* *mean, AM)* であり，$E(X) = n^{-1}\sum_{i=1}^{n} a_i$ で与えられる．$-\log x$ は凸関数である *(*確かめてみて*)* ことから，イェンセンの不等式を利用して，

$$-log(\frac{1}{n}\sum_{i=1}^{n} a_i) \leq E[-\log X] = -\frac{1}{n}\sum_{i=1}^{n}\log a_i = -\log(a_1 a_2 \cdots a_n)^{1/n}$$

となるが，これを変形すると

$$\log(\frac{1}{n}\sum_{i=1}^{n} a_i) \geq \log(a_1 a_2 \cdots a_n)^{1/n}$$

が得られる．さらに，$\log(x)$ の単調性から，

$$(a_1 a_2 \cdots a_n)^{1/n} \leq \frac{1}{n}\sum_{i=1}^{n} a_i)$$

となる．左辺を幾何平均 *(Geometric Mean, GM)* と呼ぶ．以上の不等式は，正の数の有限集合に対して，*GM*≤*AM* となることを表している．次に，a_i を $1/a_i$ で置き換えることを考える．すると，

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i}) \geq (\frac{1}{a_1}\frac{1}{a_2}\cdots\frac{1}{a_n})^{1/n}$$

さらに，両辺の逆数をとって

$$\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i})} \leq (a_1 a_2 \cdots a_n)^{1/n}$$

が得られる．左辺を調和平均 *(Harmonic Mean, HM)* と呼ぶ．以上をまとめると，正の数の有限集合に対して，

$$HM \leq GM \leq AM$$

なる大小関係が成立する．

¡note¿

2 Chap.2. Multivariate Distribution

2.6 Extension to Several Random Variables / 複数の確率変数への拡張

2.6.1 までは，第二回の山本さんの資料と伊藤の資料を参照.

2.6.1 Multivariate Variance-Covariance Matrix / 分散共分散行列

ここでは，確率変数が n 変数である場合を考える．$\mathbf{X} = (X_1, \dots, X_n)'$ を n 次元確率変数ベクトルとする．n 次元確率変数ベクトルに対して，期待値は，要素である確率変数それぞれの期待値を並べたものとして，

$$E[\mathbf{X}] = (E[X_1], \dots, E[X_n])$$

のように定義される．確率変数ベクトルの拡張として，確率変数行列を考える．$m \times n$ 行列 $\mathbf{W}$ が確率変数を要素にもつ行列とする．確率変数行列の期待値は，確率変数ベクトルと同様に，

$$E[\mathbf{W}] = [E(W_{ij})], 1 \leq i \leq m, 1 \leq j \leq n$$

のように定義される．期待値作用素の線型性から，以下の定理が成り立つ．

定理 2.1. $m \times n$ 行列 $\mathbf{W_1}, \mathbf{W_2}$ を確率変数行列とし，A_1, A_2 を $k \times m$ の，B を $n \times l$ の定数行列とする．このとき，

$$E[A_1\mathbf{W_1} + A_2\mathbf{W_2}] = A_1E[\mathbf{W_1}] + A_2E[\mathbf{W_2}]$$
$$E[A_1\mathbf{W_1}B] = A_1E[\mathbf{W_1}]B$$

が成立する．

証明 2.1. 一つ目の等式に関して，左辺の (i, j) 要素は，

$$\begin{aligned} & E[\sum_{s=1}^{m} a_{1_{is}} W_{1_{sj}} + \sum_{s=1}^{m} a_{2_{is}} W_{2_{sj}}] \\ = & \sum_{s=1}^{m} a_{1_{is}} E[W_{1_{sj}}] + \sum_{s=1}^{m} a_{2_{is}} E[W_{2_{sj}}] \\ (= & a_{1i.}E[W_{1.j}] + a_{2i.}E[W_{2.j}] \text{ (ベクトルで書いた場合)}) \end{aligned}$$

のようにかける．これを，全ての (i, j) の組み合わせについて考えることで，一つ目の等式が成立することがわかる． □

二つ目の等式に関して，左辺の (i, j) 要素は，

$$\begin{aligned} & E[\sum_{h=1}^{n} \sum_{s=1}^{m} a_{1_{is}} W_{1_{sh}} b_{hj}] \\ = & \sum_{h=1}^{n} \sum_{s=1}^{m} a_{1_{is}} E[W_{1_{sh}}] b_{hj} \end{aligned}$$

のようにかける．これを全ての $(i,j), 1 \leq i \leq k, 1 \leq j \leq l$ について考えることで，二つ目の等式が成り立つことがわかる． $\square$

¡note¿

$\mathbf{X} = (X_1, \ldots, X_n)'$ が n 次元確率変数ベクトルであり，その要素の分散が存在する，つまり，$\sigma_i^2 = V[X_i] < \infty$ とする．平均 $\mu = E[\mathbf{X}]$ を用いて，分散共分散行列を以下のように定義する．

$$Cov(\mathbf{X}) = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)'] = [\sigma_{ij}]$$

$Cov(\mathbf{X})$ の対角成分は $\sigma_{ii} = \sigma_i^2 = Var(X_i)$ であり，非対角成分は $\sigma_{ij} = Cov(X_i, X_j), i \neq j$ である．

定理 2.2. $\mathbf{X} = (X_1, \ldots, X_n)'$ が n 次元確率変数ベクトルであり，その要素の分散が存在する，つまり，$\sigma_i^2 = V[X_i] < \infty$ とする．A を $m \times n$ の定数行列とする．この時，

$$\begin{aligned} Cov(\mathbf{X}) &= E[\mathbf{X}\mathbf{X}'] - \mu\mu' \\ Cov(A\mathbf{X}) &= ACov(\mathbf{X})A' \end{aligned}$$

が成立する．

証明 2.2. 一つ目の等式について，

$$\begin{aligned} Cov(\mathbf{X}) &= E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)'] \\ &= E[\mathbf{X}\mathbf{X}' - \mu\mathbf{X}' - \mathbf{X}\mu' + \mu\mu'] \\ &= E[\mathbf{X}\mathbf{X}'] - \mu E[\mathbf{X}'] - E[\mathbf{X}]\mu' + \mu\mu' \\ &= E[\mathbf{X}\mathbf{X}'] - \mu\mu' \end{aligned}$$

$\square$

二つ目の等式に関して，

$$\begin{aligned} Cov(A\mathbf{X}) &= E[(A\mathbf{X} - A\mu)(A\mathbf{X} - A\mu)'] \\ &= E[A(\mathbf{X} - \mu)(\mathbf{X} - \mu)'A'] \\ &= AE[(\mathbf{X} - \mu)(\mathbf{X} - \mu)']A' = ACov(\mathbf{X})A' \end{aligned}$$

$\square$

¡note¿

全ての分散共分散行列は，**半正定値** *(positive semi-definite)* 行列である．つまり，$a'Cov(\mathbf{X})a \geq 0, \forall a \in \mathbb{R}^n$ である．ある $n \times 1$ ベクトル a を用いて，$Y = a'\mathbf{X}$ のような確率変数を構成する．このとき，分散の非負性から，

$$0 \leq V(Y) = V(a'\mathbf{X}) = a'Cov(\mathbf{X})a$$

となり，$Cov(\mathbf{X})$ の半正定値性が示された．
¡note¿

2.7 Transformations for Several Random Variables / 複数の確率変数の変換

本節では，n 変数の場合の変数変換について説明する．変数変換とは，単に置換積分のことなので，練習したければ，微分積分の教科書の演習を解けば良いのではないのでしょうか．さて，以下の形式の積分を考える．

$$\int \cdots \int_A f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots , dx_n$$

ここで，$A \subset \mathcal{S}$，$\mathcal{S}$ は確率変数 $X_1, \ldots, X_n$ の取りうる値の空間である．様々な理由で，$x_1, \ldots, x_n$ を

$$y_1 = u_1(x_1, \ldots, x_n),\ \ldots,\ y_n = u_n(x_1, \ldots, x_n)$$

のように変換してから積分することがある．ここで，$y_1,\dots,y_n$ の空間を $\mathcal{T}$，その部分空間を B とする．

様々な理由の例

- $Y = X_1 + X_2$ などの分布を計算したい
- 変換した方が積分計算が楽になることがある (数 3 の極座標変換とか?)

先ほどの $x_1,\dots,x_n$ の空間 $\mathcal{S}$ から $y_1,\dots,y_n$ の空間 $\mathcal{T}$ への変換 $u_1(\cdot),\cdots,u_n(\cdot)$ の *1* 対 *1* 対応の逆変換

$$x_1 = w_1(y_1,\dots,y_n),\ \dots,\ x_n = w_n(y_1,\dots,y_n)$$

を考える．$w_1,\cdots,w_n$ の一階の導関数が連続であり，空間 $\mathcal{S}$ と空間 $\mathcal{T}$ の面積比 (ここ，雑な表現です．厳密には違います) を表すヤコビアン *(Jacobian)* と呼ばれる $n\times n$ の行列式 (determinant)

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

が $\mathcal{T}$ で 0 でないとする．このとき，冒頭の積分は，

$$\begin{aligned} &\int\cdots\int_A f(x_1,x_2,\dots,x_n)dx_1dx_2\cdots,dx_n \\ =&\int\cdots\int_B f[w_1(y_1,y_2,\dots,y_n),w_2(y_1,y_2,\dots,y_n),\dots,w_n(y_1,y_2,\dots,y_n)]|J|dy_1dy_2\cdots,dy_n \end{aligned}$$

のように変換できる．$X_1,\dots,X_n$ の同時密度関数が $f(x_1,\dots,x_n)$ の時,$Y_1 = u_1(X_1,\dots,X_n),\dots,Y_n = u_n(X_1,\dots,X_n)$ の同時密度関数は $(y_1,\dots,y_n)\in\mathcal{T}$ 上で，

$$g(y_1,\dots,y_n) := f(w_1(y_1,\dots,y_n),\dots,w_n(y_1,\dots,y_n))|J|$$

で与えられる．また，$(y_1,\dots,y_n)\in\mathcal{T}$ 以外の領域では 0 である．

抽象的な話題になったので，具体例として，以下の簡単な問題を解いてみてください．

例 **2.1.** $0<x<1, 0<y<1$ の時，$u = x+y,\ v = y$ と変換するときの面積比 *(=ヤコビアン)* を求めよ．変換後の u,v の領域についても求めよ．

[以下，解答用の空白]

次に，一対一の逆変換の構成が容易でないケースについて紹介する．X がコーシー分布 (Cauchy distribution) の密度関数

$$f(x) = \frac{1}{\pi(1+x^2)},\ -\infty < x < \infty$$

を持つとし，$Y = X^2$ とする．このとき，X の空間は $\mathcal{S} := \{x| -\infty < x < \infty\}$ であり，Y の空間は $\mathcal{T} := \{y|0 \leq y < \infty\}$ である．この変換は，$y = 0$ を除いて，ある一つの y の値に対して，2 つの $x \in \mathcal{S}$ が対応しているため，**一対一対応ではない**．このようなケースには，一対一の対応にするために X の空間を複数の排反な空間に分割することで対処する．例えば，今回のケースでは，$\mathcal{S} := \{x| -\infty < x < \infty, x \neq 0\}$，$f(0) = 0$ と再定義し，$A_1 := \{x| -\infty < x < 0\}$，$A_2 := \{x|0 < x < \infty\}$ のように分割する．こうすると，x が正の場合には，逆変換 $x = -\sqrt{y}$ が $A_1, \mathcal{T}$ を一対一に結びつけ，$x = \sqrt{y}$ が $A_2, \mathcal{T}$ を一対一に結びつける．ここで，$A_3 := \{x|x = -\sqrt{y}, y \in B\} \subset A_1, A_4 := \{x|x = \sqrt{y}, y \in B\} \subset A_2$ と定義すると，$Y \in B$ の確率は，

$$\begin{aligned} P(Y \in B) &= P(X \in A_3) + P(X \in A_4) \\ &= \int_{A_3} f(x)dx + \int_{A_4} f(x)dx \end{aligned}$$

となる．領域 A_3 から B への変換でのヤコビアン J_1 は $dx = -1/(2\sqrt{y})$，領域 A_4 から B への変換でのヤコビアン J_2 は $dx = 1/(2\sqrt{y})$ であることから，

$$\begin{aligned} P(Y \in B) &= \int_B f(-\sqrt{y})| - \frac{1}{2\sqrt{y}}|dy + \int_B f(\sqrt{y})|\frac{1}{2\sqrt{y}}|dy \\ &= \int_B [f(-\sqrt{y}) + f(\sqrt{y})]\frac{1}{2\sqrt{y}}dy \end{aligned}$$

となる．よって Y の密度関数は，

$$g(y) = \frac{1}{2\sqrt{y}}[f(-\sqrt{y}) + f(\sqrt{y})],\ y \in \mathcal{T}$$

であり，これと，$f(x)$ がコーシー分布であることに注意すると，

$$g(y) = \begin{cases} \frac{1}{\pi(1+y)\sqrt{y}} & 0 < y < \infty \\ 0 & elsewhere \end{cases}$$

となることがわかる．これは自由度 1 のカイ二乗分布の密度関数である．つまり以上の結果から，

$$\text{コーシー分布の } Y = X^2 \text{変換} = \text{自由度 1 のカイ二乗分布}$$

が示された．**変数変換では一対一対応となるよう，領域を分割することが重要**である．

¡note¿

2.8 Linear Combination of Random Variables / 確率変数の線型結合

$(X_1, \ldots, X_n)'$ をランダム試行の確率変数ベクトルとする．現実世界では，標本平均 $\bar{X} = 1/n \sum_{i=1}^{n} X_i$ や標本分散 $S^2 = 1/n \sum_{i=1}^{n} (X_i - \bar{X})^2$ などのように，確率変数ベクトルの関数 $T = T(X_1, \ldots, X_n)$ が頻繁に興味の対象となる．この節では，これらの変数の $\mathbf{a} = (a_1, \ldots, a_n)'$ による線型結合

$$T = \sum_{i=1}^{n} a_i X_i$$

を考える．

定理 2.3. $T = \sum_{i=1}^{n} a_i X_i$ とする．$E[|X_i|] < \infty,\ i = 1, \ldots, n$ のもとで，

$$E[T] = \sum_{i=1}^{n} a_i E[X_i]$$

である．

証明 2.3. 期待値の線形性から明らかである． □

¡note¿

定理 2.4. $T = \sum_{i=1}^{n} a_i X_i$, $W = \sum_{i=1}^{m} b_i Y_i$ とする．もし，$E[|X_i^2|] < \infty$, $E[|Y_j^2|] < \infty$, $i = 1, \ldots, n, j = 1, \ldots, m$ であるならば，

$$Cov(T, W) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_i)$$

である．

証明 2.4.

$$\begin{aligned}
Cov(T, W) &= E[(T - E[T])(W - E[W])'] \\
&= E[(\sum_{i=1}^{n} a_i (X_i - E[X_i])(\sum_{j=1}^{m} b_j (Y_j - E[Y_j]))'] \\
&= E[\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j (X_i - E[X_i])(Y_j - E[Y_j])'] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j E[(X_i - E[X_i])(Y_j - E[Y_j])'] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)
\end{aligned}$$

□

¡note¿

系 2.1. $T = \sum_{i=1}^{n} a_i X_i$ とする．$E[|X_i|] < \infty,\ i = 1, \ldots, n$ のもとで，

$$Var(T) = Cov(T, T) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{i<j} a_i a_j Cov(X_i, X_j)$$

証明 2.5. 定理 *2.4* にて W を T に置き換えると，

$$\begin{aligned} Var(T) = Cov(T, T) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2 Cov(X_i, X_i) + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{i<j} a_i a_j Cov(X_i, X_j) \end{aligned}$$

□

¡note¿

系 2.2. もし $X_1, \ldots, X_n$ が違いに独立な確率変数であり，有限の分散を持つならば，

$$Var(T) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

証明 2.6. $X_1, \ldots, X_n$ が違いに独立な確率変数であるとき，系 *2.1* で $Cov(X_i, X_j) = 0$ となることから，系 *2.2* をえる． □

¡note¿

もし $X_1, \dots, X_n$ が違いに独立で同一な分布に従う (independent and identically distributed, iid) 確率変数ならば，この変数は，**共通の分布から，サイズ n の無作為標本を構成している**という．¡note¿

定義 2.1. 標本平均と不偏標本分散標本平均 $\bar{X}$，不偏標本分散 S^2 はそれぞれ，以下のように定義される．

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

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

定理 2.5. 標本平均の期待値，分散はそれぞれ，

$$E[\bar{X}] = \mu \tag{3}$$

$$V[\bar{X}] = \frac{\sigma^2}{n} \tag{4}$$

で定義され，不偏分散の期待値は，

$$E[S^2] = \sigma^2$$

となる．

証明 2.7. 簡単 & 統計勉強してる人の間でこれの証明は常識なので，各自証明してください．わからない場合は，「標本平均の期待値，分散」とか，「不偏標本分散の不偏性」とかで検索したら，模範解答がわんさか出てくると思うのでそれ参考にしてください．

¡note¿