第7回：多次元の確率分布

尚　晋
大学院経済学研究科　助教

2025年5月27日

今日のポイント

1. 多次元の確率分布

1 多次元の確率分布

現実の現象は 1個の確率変数で表されるような単一のものではなくて，多くの確率変数を必要とする．これらの相互関係を表すのに，同時確率分布，条件付確率分布が用いられる．とりわけ重要なのは，「無相関」と［独立」の考え方である．というのは，統計データが，いろいろな意味で互いに無関係独立と仮定されることが多いからである．

1.1 同時確率分布と周辺確率分布

離散型確率変数の同時確率分布 一般に，二つの確率変数X, Yがあるとし，これを2次元のベクトル(X, Y)として表そう. 2変数を同時に考えるのは，それらの間にたがいに関係があると考えているからである．とりあえずは，X, Yは離散型とする.$X = x$であり同時に$Y = y$である確率

$$P(X = x, Y = y) = f(x, y) \tag{1}$$

を，2次元確率変数(X, Y)の同時確率分布 joint probability distributionといい，$f(x, y)$は同時確率関数 joint probability functionという.$f(x, y)$は，1次元のときと同じく

$$f(x, y) \geq 0 \quad \text{かつ} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} f(x, y) = 1 \tag{2}$$

を満たさなければならない.

同時確率分布は上の表のようにまとめることができる.

離散型確率変数の周辺確率分布 $p_{i\cdot}$は，Y がどの値をとるかに依存せず，X が x_i をとる確率である．これを確率変数 X の**周辺確率分布** (marginal distribution) という．同様に $p_{\cdot j}$

表 1.1 同時確率分布

X ＼ Y	y_1	y_2	$\cdots$	y_m	計
x_1	p_{11}	p_{12}	$\cdots$	p_{1m}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2m}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nm}	$p_{n\cdot}$
計	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot m}$	1

を確率変数 Y の周辺確率分布という．同時確率分布から、X, Yの単独の確率分布は

$$f_x(x) = P(X = x_i) = p_{i\cdot} = \sum_{j=1}^{m} p_{ij} \tag{3a}$$

$$f_y(y) = P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{n} p_{ij} \tag{3b}$$

で求められる．周辺にあるから、それぞれ X, Y の**周辺確率分布** marginal probability distributionという．$f_x(x), f_y(y)$を周辺確率関数 marginal probability functionという．確率の総和が 1となることから，$\sum_{j=1}^{m} p_{i\cdot} = \sum_{i=1}^{n} p_{\cdot j} = 1$

例1：次の同時分布の、X とYそれぞれの周辺分布$f_x(x), f_y(y)$は？

X ＼ Y	$Y = 8$	$Y = 9$	$f_x(x)$
$X = 1$	0.1	0.1	0.2
$X = 2$	0.2	0.3	0.5
$X = 3$	0.1	0.2	0.3
$f_y(y)$	0.4	0.6	

Xの周辺分布$f_x(x)$は定義通りに求める：

$$f_x(x) = \begin{cases} 0.2 & \text{for } x = 1 \\ 0.5 & \text{for } x = 2 \\ 0.3 & \text{for } x = 3 \end{cases}$$

同じで、Yの周辺分布$f_y(y)$も定義通りに求める：

$$f_y(y) = \begin{cases} 0.4 & \text{for } y = 8 \\ 0.6 & \text{for } y = 9 \end{cases}$$

連続型確率変数の同時確率密度関数 X, Yが連続型の確率変数のときには，$f(x, y)$は 2 次元の確率密度関数である。 $a < X < b$かつ$c < Y < d$となる確率$P(a < X < b, c < Y < d)$が同時確率密度関数 (joint probability density function)$f(x, y)$を用いて

$$P(a < X < b, c < Y < d) = \int_c^d \int_a^b f(x, y)dx\,dy \tag{4}$$

で与えられる．

連続型確率変数の周辺確率密度関数 X, Yが連続型の確率変数のときには，周辺確率密度関数 (marginal probability density function) は，同時確率密度関数から

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy \tag{5a}$$

$$f_y(y) = \int_{-\infty}^{\infty} f(x,y)dx \tag{5b}$$

で得られる．

Y の値にかかわらず，Xが$a < X < b$となる確率$P(a < X < b)$は同時確率密度関数および周辺確率密度関数を用いて

$$P(a < X < b) = \int_{-\infty}^{\infty}\int_{a}^{b} f(x,y)dx\,dy \tag{6a}$$

$$= \int_{a}^{b} f_x(x)dx \tag{6b}$$

で与えられる．

同時確率密度関数$f(x,y)$は

$$f(x,y) \geq 0 \quad \text{かつ} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\,dy = 1 \tag{7}$$

を満たさなければならない．同様に，周辺確率密度関数$f_x(x), f_y(y)$は

$$f_x(x) \geq 0, \quad \int_{-\infty}^{\infty} f_x(x)dx = 1 \tag{8a}$$

$$f_y(y) \geq 0, \quad \int_{-\infty}^{\infty} f_y(y)dy = 1 \tag{8b}$$

を満たさなければならない．

1.2　期待値、共分散と相関係数

期待値(離散型) 離散型確率変数の X, Y の確率分布が表1.1 のように与えられていたとする．このとき X, Yの関数$g(X,Y)$の期待値は

$$E(g(X,Y)) = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i,y_j)f(x_i,y_j) \tag{9}$$

で定義される．ただし$f(x,y)$は X, Yの同時確率関数である．

$g(X,Y)$が Yに依存しない場合，つまり$g(X,Y) = g(X)$である場合、

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i)f(x_i,y_j) \\ &= \sum_{i=1}^{n} g(x_i)\sum_{j=1}^{m} f(x_i,y_j) \\ &= \sum_{i=1}^{n} g(x_i)f_x(x_i) \end{aligned} \tag{10}$$

が成り立つ。ただし$f_x(x_i)$は Xの周辺確率関数である．同じで$E(g(Y)) = \sum_{j=1}^{m} g(y_i)f_y(y_i)$も成り立つ。即ち、2変数あっても，1変数のみに依存する関数の期待値を求める場合には，周辺分布を用いて期待値を計算すれば良い．

期待値(連続型) 連続型確率変数X,Yの関数$g(X,Y)$の期待値は

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \tag{11}$$

で定義される．

$g(X,Y)$が Yに依存しない場合，つまり$g(X,Y)=g(X)$である場合、

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty} g(x)\int_{-\infty}^{\infty} f(x,y)dydx \\ &= \int_{-\infty}^{\infty} g(x)f_x(x)dx \end{aligned} \tag{12}$$

が成り立つ。ただし，$f(x,y)$はX,Yの同時確率密度関数，$f_x(x)$はXの周辺確率密度関数である．

和の期待値の性質:

- a,b は定数で、$E(aX+bY)$の期待値は、

$$\begin{aligned} E(aX+bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)f(x,y)dxdy \\ &= a\int_{-\infty}^{\infty} x[\int_{-\infty}^{\infty} f(x,y)dy]dx + b\int_{-\infty}^{\infty} y[\int_{-\infty}^{\infty} f(x,y)dx]dy \\ &= a\int_{-\infty}^{\infty} xf_x(x)dx + b\int_{-\infty}^{\infty} yf_y(y)dy \\ &= aE(X)+bE(Y) \end{aligned} \tag{13}$$

共分散 X,Yの共分散covarianceは

$$Cov(X,Y) = E\{(X-\mu_x)(Y-\mu_y)\} \qquad (\mu_x = E(X), \mu_y = E(Y)) \tag{14}$$

で定義される。なお、共分散 $Cov(X,Y)$ は同時確率分布で

$$Cov(X,Y) = \begin{cases} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i-\mu_x)(y_j-\mu_y)f(x_i,y_j) & \text{離散型} \\ = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)dxdy & \text{連続型} \end{cases}$$

と表される。共分散はXとYが，それぞれの平均μ_x,μ_yからたがいに関連しながら，ばらつく程度を表す．$Cov(X,Y)>0$なら，X,Yは大小が同傾向，<0なら反対傾向の関係となる．

分散の加法性$V(X+Y)\neq V(X)+V(Y)$はが必ずしも成立しない。これは，$X+Y$のばらつきには，X,Y単独のばらつきのほかに、相互関連によるばらつき$Cov(X,Y)$が存在し，それを入れてはじめて等号が成立するからである.即ち、$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$。

例えば、株式投資で、A石油の株式とB石油の株式というように，同一業種の株式に同時に投資することは，一般に勧められない．なぜなら，エネルギー危機など共通の経済的要因によって A も B も同傾向に連動するから，$Cov(X,Y)>0$ となり，単独の分散（ばらつき）の和$V(X)+V(Y)$以上にばらつくからである．

共分散の性質:

- $V(aX+bY) = a^2V(X) + 2abCov(X,Y) + b^2V(Y)$

$$\begin{aligned} V(aX+bY) &= E\{[aX+bY-E(aX+bY)]^2\} \\ &= E\{[a(X-E(X))+b(Y-E(Y))]^2\} \\ &= a^2E\{[X-E(X)]^2\} + 2abE\{[X-E(X)][Y-E(Y)]\} + b^2E\{[Y-E(Y)]^2\} \end{aligned} \tag{15}$$

- $Cov(X,Y) = E(XY) - E(X)E(Y)$

$$\begin{aligned} Cov(X,Y) &= E\{(X-E(X))(Y-E(Y))\} \\ &= E\{XY - XE(Y) - E(X)Y + E(X)E(Y)\} \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \end{aligned} \tag{16}$$

補足 1.1 資産運用への応用：ポートフォリオ選択

ポートフォリオ 異なる複数の資産（投資対象）を組み合わせて形成される資産総額を**ポートフォリオ**(portfolio)という。

今、二つの資産 A_X, A_Y でポートフォリオA_W を作る問題を考える。

- A_Xの保有比率をr、A_Yの保有比率を$(1-r)$とおく。各資産の収益率を確率変数 (X,Y)で表せば、このポートフォリオのリターン（収益）は

$$W = rX + (1-r)Y \tag{17}$$

  となる、資産保有者が、比率rを自由に決める。

- 個別資産のリターンX,Yの期待値・分散・共分散を$\mu_X = E(X), \mu_Y = E(Y), \sigma_X = V(X), \sigma_Y = V(Y), \sigma_{XY} = Cov(X,Y)$とおく。
- W の期待値 $E(W)$ を**期待リターン**、分散 $V(W)$ を**リスク**と呼ぶ。
- ポートフォリオ A_Wの期待リターンは$\mu_W = r\mu_X + (1-r)\mu_Y$となる。
- ポートフォリオ A_Wのリスクは$\sigma_W^2 = V(rX+(1-r)Y) = r^2\sigma_X^2 + (1-r)^2\sigma_Y^2 + 2r(1-r)\sigma_{XY}$

資産	A_X	A_Y	ポートフォリオA_W(A_Xの比率はr)
リターン	X	Y	$W = rX + (1-r)Y$
期待リターン	μ_X	μ_Y	$\mu_W = r\mu_X + (1-r)\mu_Y$
リスク	σ_X	σ_Y	$\sigma_W^2 = r^2\sigma_X^2 + (1-r^2)\sigma_Y^2 + 2r(1-r)\sigma_{XY}$

次の基準に基づくポートフォリオ選択を平均・分散基準という。保有比率 r の異なる二つのポートフォリオ A_{W1}（リターン = W1)、A_{W2}（リターン = W2）について

1. **高リターン選好**なら：$\sigma_{W1}^2 = \sigma_{W2}^2, \mu_{W1} > \mu_{W2}$　→W1を採用。
2. **低リスク選好**なら：$\mu_{W1} = \mu_{W2}, \sigma_{W1}^2 < \sigma_{W2}^2$　→W1を採用。

ここで質問：もしポートフォリオリスクを最小化にしたいなら、比率rをいくらにするか？(Hint:このアプローチは**最小分散ポートフォリオ**（minimum variance portfolio）と呼ぶ。)

相関係数 共分散 $Cov(X,Y)$ は，X,Yの関係の方向を表すが，その強さの程度を判断する基準ではない．そこで，この値を標準偏差で割って調整し，確率変数X,Yの**相関係数**correlation coefficientを

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)}\cdot\sqrt{V(Y)}} \tag{18}$$

で定義される。

ρの定義から，$\rho>0$ ならX,Yは同じ大小の向きに変化する傾向があり，$\rho<0$なら逆である．もっとも極端な場合は，$\rho=\pm1$であり，このときはX,Yの間には厳密に次の1次式の関係が成り立つ：

$$Y = aX + b \qquad \text{ただし}\rho = 1\text{なら、}a > 0\text{、}\rho = -1\text{なら、}a < 0 \tag{19}$$

逆に$\rho=0$ (つまり$Cov(X,Y=0$の場合はX,Yはどちらの関係をもつともいえない．この場合，X,Yは無相関 uncorrelated であるという．無相関とは,「関連がない」ということの一つの表現である.

1.3　条件付確率分布と独立な確率変数

条件付確率関数 Y が $Y=y_j$ と与えられたときの $X=X_i$ の条件付確率関数

$$\begin{aligned} f(x_i|y_j) &= P(X=x_i|Y=y_j) \\ &= \frac{P(X=x_i,Y=y_j)}{P(Y=y_j)} \\ &= \frac{f(x_i,y_j)}{f_y(y_j)} \end{aligned} \tag{20}$$

と定義される。同じで、X が $X=x_i$ と与えられたときの $Y=y_j$ の条件付確率密度関数$f(y_j|x_i)=\frac{f(x_i,y_j)}{f_x(x_i)}$と定義される。

$|$の後が条件，前が変数である．即ち、$f(x_i|y_j)$はxの関数で，他方yは指定された値に固定されている. xで和をとると，$\sum_x f(x_i|y_j) = \frac{\sum_x f(x_i,y_j)}{f_y(y_j)} = \frac{f_y(y_j)}{f_y(y_j)} = 1$が成立し， $f(x_i|y_j)$は x の関数として確かに確率分布の条件を満たしている．このように，条件付確率分布もひとつの確率分布であるから，その条件付期待値 conditional expectationは$E(X|y)=\sum_x xf(x_i|y_j)=\mu_{x|y}$, 条件付分散 conditional variance は$V(X|y)=\sum_x(x-\mu_{x|y})^2f(x_i|y_j)$で計算できる。$f(y_j|x_i)$も、連続型も同じの考え方で計算できる。

独立な確率変数 Xがx_iという値をとるという事象とYがy_jという値をとるという事象が独立であるということは

$$P(X=x_i,Y=y_j) = P(X=x_i)(P(Y=y_j) \tag{21}$$

となる.

これは，同時確率関数$f(x_i,y_j)$と周辺確率関数$f_x(x_i), f_y(y_j)$を用いれば，条件

$$f(x_i,y_j) = f_x(x_i)f_y(y_j) \tag{22}$$

が成り立つならば，X,Yはたがいに**独立** independent であるという．n 個の確率変数$X_1,X_2,..,X_n$に拡張したら、その同時確率分布 f が周辺確率分布$f_1(x_1),f_2(x_2),...,f_n(x_n)$の積$f(x_1,x_2,..,x_n)=f_1(x_1)f_2(x_2)\cdots f_n(x_n)$のように分解されるなら，$X_1,X_2,..,X_n$は独立であるという.

独立のとき条件付き確率関数 式20と式22より、X, Yはたがいに独立の時、

$$f(x_i|y_j) = f_x(x_i), \qquad f(y_j|x_i) = f_y(y_j) \tag{23}$$

が成り立つ。

上記は離散型確率変数についてであるが，確率関数を確率密度関数に置き換える事で，同様の議論が連続型確率変数に関しても成り立つ．

独立の時の性質:

- X, Yが独立ならば

$$E(XY) = E(X)E(Y) \tag{24}$$

が成立する．

離散型の場合の証明は：

$$\begin{aligned} E(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j f(x_i, y_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j f_x(x_i) f_y(y_j) \\ &= \sum_{i=1}^{n} x_i f_x(x_i) \sum_{j=1}^{m} y_j f_y(y_j) \\ &= E(X)E(Y) \end{aligned}$$

連続型の場合も同じで証明できる。

- X, Yが独立ならば

$$Cov(X, Y) = 0 \tag{25}$$

が成り立つ。即ち、独立なら無相関であるが、逆に、無相関でも独立とは限らない．

1.4　独立な確率変数の和

統計学では，ランダムな誤差を含んだ観測値，測定値，データなどの集計を行うが，そのなかでも和をとることは基本的な操作である．ところが $2+3=5$のように和を作ることは簡単だが，確率変数の和 $X+Y$ については，その確率分布は意外と求めにくい．独立性があるときは，事はやや扱いやすくなる．

分散の加法性 2 個の確率変数 X, Yに対して，$E(X+Y) = E(X) + E(Y)$という期待値の加法性は常に成り立つ。さらに、X, Yが独立であるときには，分散の加法性など

$$V(X \pm Y) = V(X) + V(Y) \tag{26}$$

が成立する。

n個の場合 n個の確率変数$X_1, X_2, .., X_n$に対しても同じく、$E(X_1 + X_2 + .. + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$であるが、さらに、$X_1, X_2, .., X_n$が独立のときには、

$$V(X_1 + X_2 + .. + X_n) = V(X_1) + V(X_2) + ... + V(X_n) \tag{27}$$

も成立する。

同一分布 $X_1, X_2, .., X_n$が同一の（共通の）確率分布に従うとし，これらの期待値，分散をμ, σ^2とすれば， $E(X_1) = E(X_2) = ... = E(X_n) = \mu, V(X_1) = V(X_2) = ... = V(X_n) = \sigma^2$であるから，

$$E(X_1 + X_2 + .. + X_n) = n\mu, \qquad V(X_1 + X_2 + .. + X_n) = n\sigma^2 \tag{28}$$

が成立する。したがつて，標準偏差は，$D(X_1 + X_2 + .. + X_n) = \sqrt{n}\sigma$となる.

相加平均 $X_1, X_2, .., X_n$をnで割った相加平均を$\bar{X} = \frac{(X_1+X_2+..+X_n)}{n}$とおくと、

$$E(\bar{X}) = \mu, \qquad V(\bar{X}) = \frac{\sigma^2}{n} \tag{29}$$

を得る．この結果は重要である．相加平均$\bar{X}$は，期待値はnに無関係に常にμに一致するが，分散はnに反比例して減少し0に収束する．この安定化の傾向を定理の形に発展させたのが**大数の法則**である。

和の確率分布 離散型の確率変数X, Yが独立であるとし，その確率分布を$g(x), h(y)$としよう．和 $X + Y$の確率分布$k(z)$は確率 $P(X + Y = z)$ を考えれば得られる.

関数$k(z)$は

$$\begin{aligned}
k(z) &= P(X + Y = z) \\
&= \sum_x P(X = x, Y = z - x) \\
&= \sum_x P(X = x)P(Y = z - x|X = x) \qquad \text{条件付き確率関数定義の変形より} \\
&= \sum_x P(X = x)P(Y = z - x) \\
&= \sum_x g(x)h(z - x)
\end{aligned} \tag{30}$$

関数 g, hからkを作る数学操作をg, hのたたみこみ convolution といい， $k = g * h$ と書く.g, hが密度関数のときも同様で，たたみこみは積分

$$k(z) = \int_{-\infty}^{\infty} g(x)h(z - x)dx \tag{31}$$

となる．このたたみこみを用いて，いくつかのよく知られた確率分布に対しては，次の結果が得られる。

再生的 (a) 二項分布　$Bi(n,p)*Bi(m,p) = Bi(n+m,p)$ X, Yが独立で，それぞれ$Bi(n,p), Bi(m,p)$に従っているとき，$X + Y$は$Bi(n + m, p)$に従う.

(b) ポアソン分布　$Po(\lambda1)*Po(\lambda2) = Po(\lambda1+\lambda2)$ X, Yが独立で，それぞれ$Po(\lambda1), Po(\lambda2)$に従っているとき，$X + Y$は$Po(\lambda1 + \lambda2)$に従う.

(c) 正規分布　$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ X, Yが独立で，それぞれ$N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$に従っているとき，$X + Y$は$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$に従う.

このように，たたみこみの結果として，全く別の分布でなくふたたび同一種類の確率分布（確率分布族）が得られるならば，取扱いが便利である．このとき，この確率分布族は**再生的** reproductive であるという.

正規分布の再生性 正規分布については，再生性は大きなメリットがあり，統計的推測によく用いられているのであらかじめ知っておくとよい.

(a) $X_1, X_2, .., X_n$が独立で，それぞれ正規分布$N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2), .., N(\mu_n, \sigma_n^2)$従っているならば，

(i) $X_1+X_2+..+X_n$は$N(\mu_1+\mu_2+...+\mu_n,\sigma_1^2+\sigma_2^2+...+\sigma_n^2)$従い，

(ii) $c_1X_1+c_2X_2+..+c_nX_n$は$N(c_1\mu_1+c_2\mu_2+...+c_n\mu_n,c_1^2\sigma_1^2+c_2^2\sigma_2^2+...+c_n^2\sigma_n^2)$従う。

(b) とくに$X_1,X_2,..,X_n$の確率分布がすべて正規分布$N(\mu,\sigma^2)$なら，

(i) $X_1+X_2+..+X_n$は$N(n\mu,n\sigma^2)$に従い，

(ii) $\bar{X}=\frac{(X_1+X_2+..+X_n)}{n}$は$N(\mu,\frac{\sigma^2}{n})$に従う。

補足 1.2 モーメントと積率母関数(モーメント母関数)

モーメント 以前勉強したように、平均値と分散によって分布の形は決めることを思い出しよう。確率分布の場合も同じで、$E(X-\mu)^1$は期待値、$E(X-\mu)^2$は分散、このように、$E(X-\mu)^k$なる量で確率分布の形は決まつてくることがわかる．あるいは，$E(X^k)$を考えてもよい．一般に

$$\mu_k=E(X^k) \tag{32}$$

を，X の（原点のまわりの）k 次のモーメント moment,または積率といい，

$$\mu_k'=E(X-\mu)^k \qquad ((\text{ただし，}\mu=E(X)) \tag{33}$$

を，X の期待値（平均）のまわりの k 次のモーメントという．用語「モーメント」は，力学のモーメント（積率，能率）と数学的に似ていることによる．

期待値，分散はモーメントの基礎的なものであって

$$\mu_1=E(X), \qquad \mu_2'=V(X) \tag{34}$$

である．高次の標準化モーメントとしては，歪度と尖度を定義することもできる。

即ち、すべての次数のモーメントを指定すれば，それにより一つの確率分布が決定されるはずである．従って、すべての次数のモーメントを生成するモーメント母関数 moment generating functionという。

積率母関数 確率変数 X の積率母関数(moment generating function) は

$$M_X(t)=E(e^{tX}) \tag{35}$$

で定義される．ただし e は自然対数の底である．

e^{tX}をマクローリン展開を用いて展開すると

$$\begin{aligned} e^{tX} &= 1+tx+\frac{(tx)^2}{2!}+\cdots+\frac{(tx)^k}{k!}+\cdots \\ &= \sum_{k=0}^{\infty}\frac{(tx)^k}{k!} \end{aligned} \tag{36}$$

である．このことを利用すると，式35積率母関数は次のように書き換える．

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tX} f(x)dx \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} f(x)dx \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \int_{-\infty}^{\infty} x^k f(x)dx \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(x^k) \end{aligned} \tag{37}$$

$M_X(t)$を t で微分すると

$$\frac{M_X(t)}{dt} = E(X) + \frac{2t}{2!}E(X^2) + \frac{3t^2}{3!}E(X^3) + \cdots \tag{38}$$

となる．この式を $t=0$ で評価すると

$$\frac{dM_X(t)}{dt}|_{t=0} = E(X) \tag{39}$$

となり，X の 1 次の積率が求まる．$\frac{dM_X(t)}{dt}$を t でさらに微分すると

$$\frac{d^2M_X(t)}{dt^2} = \frac{2}{2!}E(X^2) + \frac{3 \cdot 2t}{3!}E(X^3) + \cdots \tag{40}$$

この式を $t=0$ で評価すると

$$\frac{d^2M_X(t)}{dt^2}|_{t=0} = E(X^2) \tag{41}$$

となり，X の 2 次の積率が求まる．このように積率母関数を k 回微分して 0 で評価すると，一般に

$$\frac{d^kM_X(t)}{dt^k}|_{t=0} = E(X^k) \tag{42}$$

となり，k 次の積率が得られることになる．このように，積率母関数は確率変数の積率を求める上で非常に便利である．

2 つの確率変数 X, Y の確率密度関数 (または確率関数) を$f_x(x), f_y(y)$とし，対応する積率母関数$M_X(t), M_Y(t)$が存在するとする．このとき，X, Yが独立ならば

$$M_{X+Y}(t) = M_X(t)M_Y(t) \tag{43}$$

が成立する．

証明：

$$
\begin{aligned}
M_{X+Y}(t) &= E(e^{t(X+Y)}) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(X+Y)} f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(X+Y)} f_x(x)f_y(y)dxdy \\
&= \int_{-\infty}^{\infty} e^{tX} f_x(x)dx \int_{-\infty}^{\infty} e^{tY} f_y(y)dy \\
&= E(e^{tX})E(e^{tY}) \\
&= M_X(t)M_Y(t)
\end{aligned} \tag{44}
$$