確率数理工学6

特性関数の性質

Thm (分布関数と特性関数の関係) 重要

2つの分布関数 F_1, F_2 に対して、
その特性関数を ϕ_1, ϕ_2 とする.

F_1 と F_2 が等しい ($F_1(x) = F_2(x)$ $({}^\forall x \in \mathbb{R})$)

$\iff$ ϕ_1 と ϕ_2 が等しい ($\phi_1(t) = \phi_2(t)$ $({}^\forall t \in \mathbb{R})$) //

($\Rightarrow$ は自明, $\Leftarrow$ は非自明)

☆ 分布の等しさを、ある関数クラスの期待値の等しさで測る.

$$\int e^{itx} dF_1(x) = \int e^{itx} dF_2(x) \quad ({}^\forall t \in \mathbb{R})$$

これより、全てのモーメントが等しければ分布が等しいこともわかる.
(c.f. ベクトル x と x' が等しい $\iff$ $\langle x, y \rangle = \langle x', y \rangle$ $({}^\forall y \in \mathbb{R}^d)$)

より強く以下のことが言える.

Thm (Lévyの反転公式)

$$P(X \in (a,b)) + \frac{1}{2}\left(P(X=a) + P(X=b)\right)$$
$$= \lim_{R\to\infty} \frac{1}{2\pi} \int_{-R}^{R} \frac{e^{-i\xi b} - e^{-i\xi a}}{-i\xi} \phi(\xi)\, d\xi \quad //$$

(c.f. フーリエ逆変換. $P(X \in (a,b))$ が b で微分可能なら、両辺 b で微分してみる)

先の定理をLévyの反転公式を用いて示す.

$\Leftarrow$ を示す.

a, b が F の連続点なら. $P(X=a)=P(X=b)=0$ で

$$F(b)-F(a)=P(X\in(a,b))$$

となる. よって, $\phi_1=\phi_2$ なら. F_1 と F_2 が同時に連続である点 $x, x' \in \mathbb{R}$ において

$$F_1(x)-F_1(x')=F_2(x)-F_2(x')$$

である. さらに, 不連続点は高々可算個なので, x' はいくらでも大きくできて. $x'\to-\infty$ とすれば分布関数の性質より

$$F_1(x)=F_2(x) \qquad (x \text{ は } F_1, F_2 \text{ の連続点})$$

を得る.

次に任意の $x\in\mathbb{R}$ (不連続点でも良い) を固定すると,

$\forall n>0$ で $|x-x_n|\leqq\frac{1}{n}$ かつ $x\leqq x_n$ であり, x_n で F_1 と F_2 が連続であるような点 x_n が存在する. つまり

$$F_1(x_n)=F_2(x_n) \quad \text{とできる.}$$

すると, F_1 と F_2 の右連続性より.

$$F_1(x)\underset{\text{右連続性}}{=}\lim_{n\to\infty}F_1(x_n)=\lim_{n\to\infty}F_2(x_n)\underset{\text{右連続性}}{=}F_2(x)$$

を得る. //

Lévyの反転公式の証明

Rを止めた時の右辺の積分

$$= \frac{1}{2\pi}\int_{-R}^{R}\left(\frac{e^{-i\xi b}-e^{-i\xi a}}{-i\xi}\right)\left(\int e^{i\xi x}dF(x)\right)d\xi$$

$$= \frac{1}{2\pi}\int dF(x)\left(\int_{-R}^{R}\frac{e^{-i\xi(b-x)}-e^{-i\xi(a-x)}}{-i\xi}d\xi\right) \qquad (\text{Fubini})$$

$$= \frac{1}{2\pi}\int dF(x)\int_{-R}^{R}\left\{\frac{\cos(\xi(x-b))-\cos(\xi(x-a))}{-i\xi} - \frac{\sin(\xi(x-b))-\sin(\xi(x-a))}{\xi}\right\}d\xi$$

(注: $\frac{\cos(\xi(x-b))-\cos(\xi(x-a))}{\xi}$, $\frac{\sin(\xi(x-b))-\sin(\xi(x-a))}{\xi}$ はともに有界

$\xi\to 0$ においても ⟶ これは $(x-b)\sin(\xi(x-b))-(x-a)\sin(\xi(x-a))$ に収ま)

ロピタル

ここで、$\frac{\cos(\xi(x-b))-\cos(\xi(x-a))}{\xi}$ は奇関数なので積分は0

よって、$f_R(x)=\int_0^R\frac{\sin(\xi x)}{\xi}d\xi \quad (x\in\mathbb{R})$ とすると、

$$右辺 = \frac{1}{2\pi}\cdot 2\int dF(x)\left(f_R(x-a)-f_R(x-b)\right)$$

ここで、
$$\lim_{R\to\infty}f_R(x)=\begin{cases}\int_0^\infty\frac{\sin(\xi)}{\xi}d\xi=\frac{\pi}{2} & (x>0)\\ -\frac{\pi}{2} & (x<0)\\ 0 & (x=0)\end{cases}$$

より、
$$\frac{1}{2\pi}\cdot 2\cdot\lim_{R\to\infty}\int f_R(x-a)-f_R(x-b)\,dF(x)$$

$$=\frac{1}{\pi}\int\lim_{R\to\infty}\left(f_R(x-a)-f_R(x-b)\right)dF(x) \qquad (\text{優収束定理})$$

積分の中身は有界

$$=\frac{1}{\pi}\cdot\frac{\pi}{2}\int\left(\mathbb{1}\{x>a\}-\mathbb{1}\{x<a\}\right)-\left(\mathbb{1}\{x>b\}-\mathbb{1}\{x<b\}\right)dF(x)$$

$$=\frac{1}{2}\int\left(2\,\mathbb{1}\{a<x<b\}+\mathbb{1}\{x=a\}+\mathbb{1}\{x=b\}\right)dF(x)$$

$$= \int \mathbb{1}\{a<x<b\}\, dF(x) + \frac{1}{2}\int \mathbb{1}\{x=a\}\, dF(x) + \frac{1}{2}\int \mathbb{1}\{x=b\}\, dF(x)$$

$$= P(a<X<b) + \frac{1}{2}\left(P(X=a) + P(X=b)\right)$$

↑ 定義関数の期待値.

※分布の性質を調べたければ特性関数の性質を調べれば良い.

Thm (独立な確率変数の和の特性関数) ㊥重要!

$X_1, \dots, X_n$: 互いに独立な r.v.

$\phi_j(t)$: X_j の特性関数 ($j=1,2,\dots,n$)

$S_n = X_1 + \cdots + X_n$

S_n の特性関数は

$$\phi(t) = \phi_1(t)\cdot\phi_2(t)\cdot\cdots\cdot\phi_n(t)$$

で与えられる.

(和の特性関数は、特性関数の積)

Proof

$$\phi(t) = E\left[e^{it(X_1+\cdots+X_n)}\right] = E\left[e^{itX_1}\cdots\cdots e^{itX_n}\right]$$

$$= \prod_{j=1}^{n} E\left[e^{itX_j}\right] = \phi_1(t)\times\cdots\times\phi_n(t)$$

↑ 独立性

Thm (独立性と特性関数) ← 特性関数と分布の一対一対応

$X_1, \dots, X_n$: r.v.

同時分布の特性関数 : $\phi(t_1,\dots,t_n) \overset{\text{def}}{=} E\left[e^{i(t_1X_1+\cdots+t_nX_n)}\right]$

ϕ_j : X_j の c.f. (特性関数)

$X_1,\dots,X_n$ が独立 $\iff \phi(t_1,\dots,t_n) = \phi_1(t_1)\cdots\phi_n(t_n)$

例

Ex.（二項分布）

$X_i \sim \mathrm{Bernoulli}(\theta)$ $\left(P(X_i=1)=\theta,\ P(X_i=0)=1-\theta \right)$
(i.i.d.)

$S = X_1 + \cdots + X_n$ とおく.

$$\begin{cases} X_i \text{の特性関数} : \theta e^{it} + (1-\theta) \\ S \text{の} \quad 〃 \quad : \left(\theta e^{it} + (1-\theta)\right)^n \end{cases}$$

前回の講義

⇝ 二項分布の特性関数

$\Longrightarrow S \sim B(n, \theta)$ （二項分布）

Ex.（正規分布の和の分布）

$X_1, \ldots, X_n$: r.v.，独立

$X_j \sim N(\mu_j, \sigma_j^2)$

$S_n = X_1 + \cdots + X_n$

S_n の c.f.　　X_j の c.f.

$$\begin{aligned} \phi(t) &= \prod_{j=1}^{n} \phi_j(t) \\ &= \prod_{j=1}^{n} \exp\left(it\mu_j - \frac{1}{2}\sigma_j^2 t^2\right) \\ &= \exp\left(it\bar{\mu} - \frac{1}{2}\bar{\sigma}^2 t^2\right) \end{aligned}$$

ただし，$\bar{\mu} = \mu_1 + \cdots + \mu_n$，$\bar{\sigma}^2 = \sigma_1^2 + \cdots + \sigma_n^2$

よって，$\underline{S_n \sim N(\bar{\mu}, \bar{\sigma}^2)}$ ：正規分布の再生性

⟶ 和の分布がすぐに求まる．

様々な分布

- 連続な分布
  - 指数分布 ($Ex(\lambda)$)

p.d.f. $f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$

($\lambda > 0$: パラメータ)

$$F(x) = \int_0^x \lambda e^{-\lambda s} ds = 1 - e^{-\lambda x}$$

〜 単位時間にある事象が平均λ回発生するとして、その事象の発生間隔の分布
(ポアソン分布と関係)

特性関数 $\phi(t) = \int_0^\infty \lambda e^{-\lambda x} e^{ixt} dx$

$$= \frac{\lambda}{\lambda - it} = \left(1 - \frac{it}{\lambda}\right)^{-1}$$

キュムラント母関数

$$\psi(t) = \ln\left(\frac{\lambda}{\lambda - it}\right) = -\ln\left(1 - \frac{it}{\lambda}\right)$$

$$\psi'(t) = \left(\frac{i}{\lambda}\right)\left(1 - \frac{it}{\lambda}\right)^{-1}$$

$$\psi''(t) = \left(\frac{i}{\lambda}\right)^2\left(1 - \frac{it}{\lambda}\right)^{-2}$$

$$\begin{cases} \psi'(0) = \frac{i}{\lambda} & \longrightarrow & \kappa_1 = \frac{1}{\lambda} = \mu & (\text{平均}) \\ \psi''(0) = \left(\frac{i}{\lambda}\right)^2 & \longrightarrow & \kappa_2 = \frac{1}{\lambda^2} & (\text{分散}) \\ \psi^{(3)}(0) = 2\left(\frac{i}{\lambda}\right)^3 & \longrightarrow & \kappa_3 = \frac{2}{\lambda^3} & \\ \psi^{(4)}(0) = 3!\left(\frac{i}{\lambda}\right)^4 & \longrightarrow & \kappa_4 = \frac{6}{\lambda^4} & \end{cases}$$

$\Longrightarrow$ $\beta_1 = 2$, $\beta_2 = 6$

(λ によらない)

ーガンマ分布 ($G(\lambda, k)$)

p.d.f. $$f(x) = \begin{cases} \dfrac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

Note $\Gamma(s) \overset{\text{def}}{=} \int_0^\infty e^{-x} x^{s-1} dx$: ガンマ関数

$$\Gamma(s+1) = s\,\Gamma(s)$$

nが自然数のとき、$\Gamma(n) = (n-1)!$, $\Gamma(1) = 1$

$$\Gamma\left(n+\tfrac{1}{2}\right) = \left(n-\tfrac{1}{2}\right)\left(n-\tfrac{3}{2}\right)\cdots\tfrac{1}{2}\sqrt{\pi}$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

特性関数

$$\phi(t) = \int_0^\infty \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} e^{ixt} dx$$

$$= \frac{\lambda^k}{(\lambda - it)^k} \cdot \frac{1}{\Gamma(k)} \int_0^\infty ((\lambda - it)x)^{k-1} e^{-(\lambda - it)x} (\lambda - it) dx$$

$$= \frac{\lambda^k}{(\lambda - it)^k} = \left(1 - \frac{it}{\lambda}\right)^{-k}$$

$X_j \sim Ex(\lambda)$ $(j = 1, \cdots, k)$ 互いに独立、とおくと

$S_k = X_1 + \cdots + X_k$ の c.f. は

$\phi(t) = \left(1 - \frac{it}{\lambda}\right)^{-k}$: ガンマ分布 $G(\lambda, k)$ の特性関数

→ 単位時間あたり平均λ回発生する事象が k回起きるまでの時間 (指数分布の和はガンマ分布)

キュムラント母関数 $\psi(t) = -k \ln\left(1 - \frac{it}{\lambda}\right)$

$$\left.\begin{aligned} \kappa_1 &= \frac{k}{\lambda} & \kappa_3 &= \frac{2k}{\lambda^3} \\ \kappa_2 &= \frac{k}{\lambda^2} & \kappa_4 &= \frac{6k}{\lambda^4} \end{aligned}\right\} \longrightarrow \quad \beta_1 = \frac{2}{\sqrt{k}}, \quad \beta_2 = \frac{6}{k}$$

$k \to \infty$で $\beta_1 \to 0$, $\beta_2 \to 0$ (正規分布に近づく)

$$\frac{X - \frac{k}{\lambda}}{\sqrt{k/\lambda^2}} \xrightarrow{w} N(0,1)$$

ー χ^2分布（カイ2乗分布）

$G(\frac{1}{2}, \frac{n}{2})$ を自由度 n の χ^2-分布と言う。

統計でよく現れる。

使い所：（正規分布と χ^2分布の関係）

$X \sim N(0,1)$：標準正規分布

$Y = X^2$ の分布は $G(\frac{1}{2}, \frac{1}{2})$ に従う

--- 自由度1の χ^2分布

$X_j \sim N(0,1)$ $(j=1, \dots, k)$ (i.i.d.) なら

$Y = X_1^2 + \cdots + X_k^2$ は $G(\frac{1}{2}, \frac{k}{2})$ に従う。

--- 自由度 k の χ^2分布

(Proof) ・ $k=1$ のとき。$X \sim N(0,1)$ なら

標準正規分布のp.d.f.

$$F(y) = P(X^2 \leqq y) = \int_{-\sqrt{y}}^{\sqrt{y}} p(x)\,dx = 2\int_0^{\sqrt{y}} p(x)\,dx$$

$$f(y) = \frac{d}{dy}F(y) = \frac{d}{dy}\left(2\int_0^{\sqrt{y}} p(x)\,dx\right)$$

$$= \frac{d\sqrt{y}}{dy} \cdot 2p(\sqrt{y}) = \frac{1}{\sqrt{y}} \underset{\frac{\sqrt{\frac{1}{2}}}{\Gamma(\frac{1}{2})}}{\underbrace{\frac{1}{\sqrt{2\pi}}}} \exp\left(-\frac{\sqrt{y}^2}{2}\right)$$

$$= \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \cdot y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}$$

$\longrightarrow G(\frac{1}{2}, \frac{1}{2})$ の p.d.f.

・ $k > 1$ のとき。

$Y = X_1^2 + \cdots + X_k^2$ の特性関数：

$$\phi(t) = \left[(1-2it)^{-\frac{1}{2}}\right]^k = (1-2it)^{-\frac{k}{2}}$$

$\Longrightarrow G(\frac{1}{2}, \frac{k}{2})$ の特性関数：自由度 k の χ^2分布

- Cauchy分布（コーシー分布）

p.d.f. $f(x) = \dfrac{1}{\pi} \dfrac{a}{a^2+x^2}$

$$E[|x|] = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{|x|}{a^2+x^2}dx$$
$$= \frac{2a}{2\pi}\left[\ln(a^2+x^2)\right]_0^{\infty}$$

$\longrightarrow$ 有限に確定しない
E[X]は存在しない（非可積分）

c.f. : $\phi(t) = \exp(-a|t|)$

- ベータ分布（Be(a,b)）

p.d.f. $f(x) = \begin{cases} \dfrac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & (0<x<1) \\ 0 & (\text{その他}) \end{cases}$

a,bはパラメータ（形状母数）

Note $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$: ベータ関数

平均 : $\mu = \dfrac{a}{a+b}$

分散 : $\sigma^2 = \dfrac{a(a+1)}{(a+b)(a+b+1)}$

実は $X \sim G(a,1),\ Y \sim G(b,1)$ （独立） $\Longrightarrow \dfrac{X}{X+Y} \sim Be(a,b)$

$\Longrightarrow$ 2項分布の事前分布に使われたりする。

○ 離散分布

- 二項分布（省略）

- Poisson分布 $(P_o(\lambda))$

p.m.f. $f(x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$ $(x=0,1,2,\dots)$

$\displaystyle\sum_{x=0}^{\infty} f(x) = 1$ をチェックせよ

特性関数

$$\phi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda}$$
$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!}$$
$$= e^{-\lambda} e^{\lambda e^{it}} = \exp\left(\lambda(e^{it}-1)\right)$$

キュムラント母関数

$$\psi(t) = \ln \phi(t) = \lambda(e^{it}-1)$$

$$\psi'(t) = i\lambda e^{it}$$
$$\psi''(t) = (i)^2 \lambda e^{it}$$
$$\psi^{(k)}(t) = (i)^k \lambda e^{it}$$

$\longrightarrow \begin{cases} \kappa_1 = \lambda \\ \kappa_2 = \lambda \\ \vdots \\ \kappa_k = \lambda \quad (^\forall k) \end{cases}$

$\longrightarrow \beta_1 = \dfrac{1}{\sqrt{\lambda}},\ \beta_2 = \dfrac{1}{\lambda}$

事象の発生間隔が指数分布に従うとき、

単位時間あたりに発生する事象の数の分布

$X_1, X_2, \dots$：独立な指数分布 $(Ex(\lambda))$、客の到着間隔

客が k人以上来た：$X_1 + \dots + X_k \leqq 1$

つまり、$Y \sim G(\lambda, k)$ として $Y \leqq 1$ である確率

事象の発生間隔が指数分布に従うとき、

単位時間あたりに発生する事象の数の分布

$X_1, X_2, \ldots$：独立な指数分布 $(Ex(\lambda))$、客の到着間隔

客が k人以上来た：$X_1 + \cdots + X_k \leqq 1$

つまり、$Y \sim G(\lambda, k)$ として $Y \leqq 1$ である確率

$$\int_0^1 \frac{\lambda^k}{\Gamma(k)} y^{k-1} e^{-\lambda y} dy$$

$$= \int_0^1 -\frac{\lambda^{k-1}}{\Gamma(k)} y^{k-1} \frac{d}{dy}\left(e^{-\lambda y}\right) dy$$

$$= \underbrace{\left[-\frac{\lambda^{k-1}}{\Gamma(k)} y^{k-1} e^{-\lambda y}\right]_0^1}_{= -\frac{\lambda^{k-1}}{\Gamma(k)} e^{-\lambda}} + \int_0^1 \frac{\lambda^{k-1}}{\Gamma(k)} e^{-\lambda y} y^{k-2} dy$$

$\vdots$

$$= 1 - \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} e^{-\lambda} = 1 - \underbrace{F(k-1)}_{\text{Poisson分布の分布関数}}$$

つまり $P(k\text{人以上}) = 1 - F(k-1)$

$\Longrightarrow P(k-1\text{人以下}) = F(k-1)$ ：Poisson分布

負の二項分布（$NB(k,\theta)$）

コイン投げをして、k回表が出るまでに裏が出る回数(x)の分布
（ただし、その1回前までは裏x回、表$k-1$回とする）

$k=1$のとき、$NB(1,\theta)=G(\theta)$：幾何分布と言う

一般のとき、　p.m.f

$$f(x)=\binom{k+x-1}{k-1}\theta^{k-1}(1-\theta)^x\underbrace{\theta}_{\text{最後に成功}}$$

$$=\binom{k+x-1}{k-1}\theta^k(1-\theta)^x$$

特性関数

$$\phi(t)=\sum_{x=0}^{\infty}e^{itx}\binom{k+x-1}{k-1}\theta^k(1-\theta)^x$$

$$=\sum_{x=0}^{\infty}\binom{k+x-1}{k-1}\theta^k\left(e^{it}(1-\theta)\right)^x$$

ここで、

$$\left(\frac{1}{1-z}\right)^k=\sum_{x=0}^{\infty}\frac{k(k+1)\cdots(k+x-1)}{x!}z^x$$

$$=\sum_{x=0}^{\infty}\binom{k+x-1}{k-1}z^x$$

より、

$$\phi(t)=\theta^k\left(1-e^{it}(1-\theta)\right)^{-k}$$ である。

$$\psi(t)=\ln(\phi(t))=k\ln(\theta)-k\ln\left(1-e^{it}(1-\theta)\right)$$

$$\psi'(t)=-k\frac{-i(1-\theta)e^{it}}{1-e^{it}(1-\theta)}\xrightarrow[t=0]{}\mu=k\frac{1-\theta}{\theta}$$

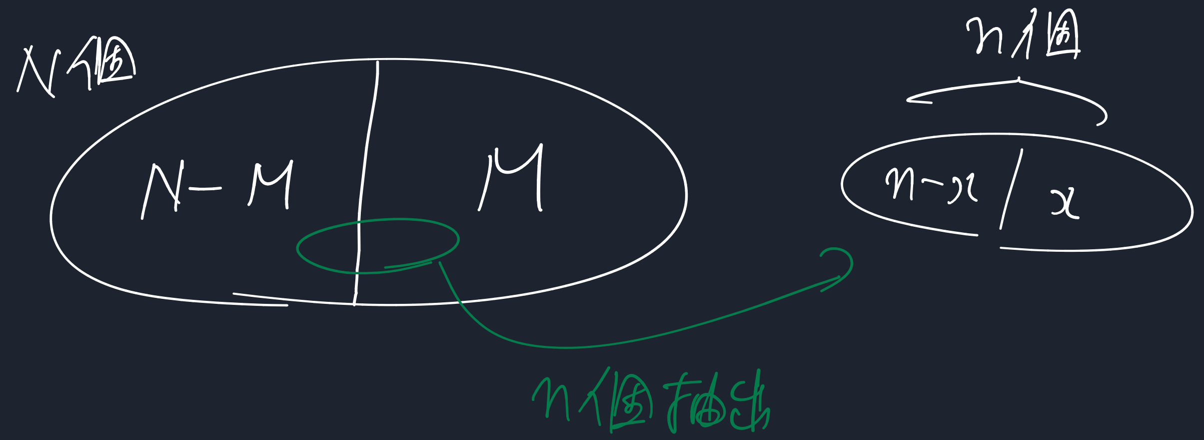
$$\psi''(t)=k\frac{i^2(1-\theta)e^{it}}{(1-e^{it}(1-\theta))^2}\xrightarrow[t=0]{}\sigma^2=k\frac{1-\theta}{\theta^2}$$

（θが小さいと大きい）

例：クーポンコレクター問題、コンプガチャ

一 超幾何分布 $(H(n, N, M))$

N個のアイテムにM個の不良品が入っているとき、
n個を非復元抽出した時、そのn個にx個の
不良品が入っている確率



p.m.f. $$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

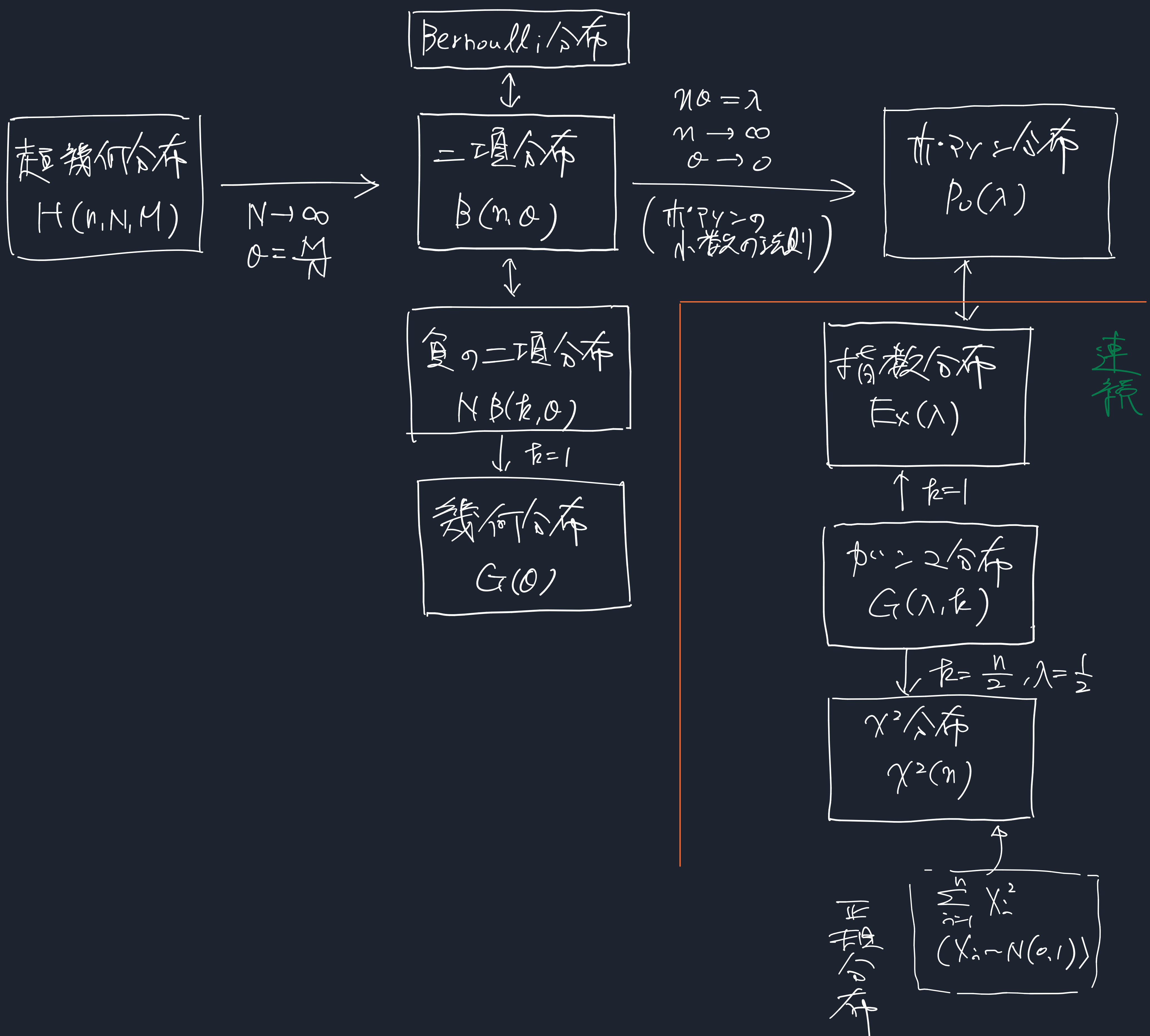
$\theta = \frac{M}{N}$ として、$N \to \infty$ とすると。(θは固定)

$$\longrightarrow \binom{n}{x}\theta^x(1-\theta)^{n-x} \quad (\text{二項分布})$$

$$\left(f(x) = \frac{\dfrac{M(M-1)\cdots(M-x+1)}{x!} \cdot \dfrac{(N-M)(N-M-1)\cdots(N-M-n+x+1)}{(n-x)!}}{\dfrac{N\cdots(N-n+1)}{n!}} \right.$$

$$= \binom{n}{x}\frac{\frac{M}{N}\cdot\left(\frac{M}{N}-\frac{1}{N}\right)\cdots\cdots\left(\frac{M}{N}-\frac{x-1}{N}\right)\left(1-\frac{M}{N}\right)\left(1-\frac{M+1}{N}\right)\cdots\left(1-\frac{M+n-x-1}{N}\right)}{1\times\left(1-\frac{1}{N}\right)\times\cdots\times\left(1-\frac{n-1}{N}\right)}$$

$$\left. \xrightarrow[\substack{N\to\infty \\ M\to\infty \\ (\frac{M}{N}=\theta)}]{} \binom{n}{x}\theta^x(1-\theta)^{n-x} \right)$$


Bernoulli分布
超幾何分布
H(n,N,M)
N→∞
θ=M/N
二項分布
B(n,θ)
nθ=λ
n→∞
θ→0
(ポアソンの小数の法則)
ポアソン分布
Po(λ)
負の二項分布
NB(k,θ)
k=1
幾何分布
G(θ)
連続
指数分布
Ex(λ)
k=1
ガンマ分布
G(λ,k)
k=n/2, λ=1/2
χ²分布
χ²(n)
Σ_{i=1}^{n} X_i²
(X_i〜N(0,1))
正規分布