9　統計量から誘導される分布

9　代表的な統計量の分布

ここでは，推測統計学の理解に必要な各種統計量の分布について述べる．

T1　χ^2 分布　(Chi-square distribution)

T2　t 分布　(t distribution)

T3　F 分布　(F distribution)

9.1　χ^2 分布 $[\chi_n^2]$

確率変数 X_i, $(i = 1, 2, \cdots, n)$ が独立で同一な標準正規分布 $N(0,1)$ で定義されているとし，これら n 個の自乗和

$$S_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

の従う分布．n は自由度 (degree of freedom) といわれる．

確率密度関数

$$p_X(x) = \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \qquad (0 < x < \infty)$$

証明　まず，χ^2 分布の再生性を示す．確率変数 X が χ^2 分布に従うなら，積率母関数は

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx \\ &= \frac{1}{2\Gamma(\frac{n}{2})} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-(\frac{1}{2}-t)x} dx \end{aligned}$$

となる．ここで，変数変換

$$(\frac{1}{2}-t)x = z, \qquad (t < \frac{1}{2})$$

を施すと

$$\begin{aligned} M_X(t) &= \frac{1}{2\Gamma(\frac{n}{2})}\int_0^\infty \left(\frac{z}{1-2t}\right)^{\frac{n}{2}-1} e^{-z}\frac{2}{1-2t}dz \\ &= (1-2t)^{-\frac{n}{2}}\frac{1}{\Gamma(\frac{n}{2})}\int_0^\infty z^{\frac{n}{2}-1}e^{z}dz = (1-2t)^{-\frac{n}{2}} \end{aligned}$$

が得られる．よって，χ^2 分布に従う確率変数 X_1 と X_2 の積率母関数はそれぞれ，$M_{X_1}(t) = (1-2t)^{-\frac{n_1}{2}}$ と $M_{X_2}(t) = (1-2t)^{-\frac{n_2}{2}}$ となる．このとき，$S_2 = X_1 + X_2$ の積率母関数は

$$M_{S_2}(t) = M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = (1-2t)^{-\frac{(n_1+n_2)}{2}}$$

となる．このことから，$X_1 + X_2$ の分布が χ^2 分布 $[\chi^2_{n_1+n_2}]$ であることがわかる．

つぎに，確率変数 X が標準正規分布 $N(0,1)$ に従うとき，X^2 の積率母関数は

$$M_{X^2}(t) = \int_0^\infty e^{tx^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = \int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{(1-2t)x^2}{2}}dx$$

となる．ここで，変数変換

$$\left(\frac{1}{2}-t\right)^{\frac{1}{2}} x = z$$

を施すと

$$M_{X^2}(t) = \left(\frac{1}{2}-t\right)^{\frac{1}{2}}\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz = (1-2t)^{\frac{1}{2}}$$

が得られる．よって，$X^2 \sim \chi^2_1$ である．

確率変数 X_i, $(i = 1, 2, \cdots, n)$ が互いに独立であるなら，これらを異なる要素をもつ任意の k $(\leq n)$ 組に分けるとき，各組ごとに得られる k 個の統計量は互いに独立であることから χ^2 分布 $[\chi^2_n]$ が求められる．

平均・分散・特性関数

$$\mathrm{E}[X] = n, \qquad \mathrm{V}[X] = 2n, \qquad \phi_X(t) = (1-2it)^{-\frac{n}{2}}.$$

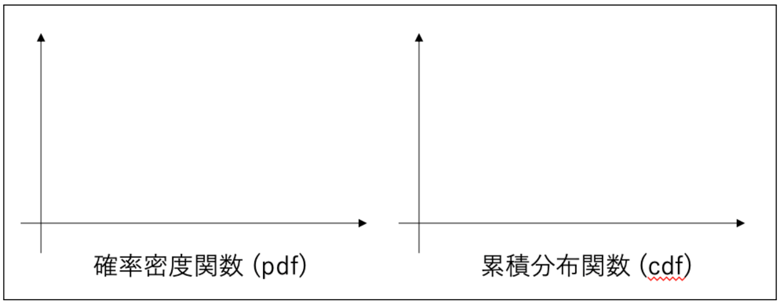

図 1: χ_n^2 の pdf と cdf

9.2 t 分布 $[t_n]$

確率変数 X と Y は独立な確率変数で，X は標準正規分布 $N(0,1)$ に従い，Y は自由度 n のカイ自乗分布 χ_n^2 に従う．これらから導かれる

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

が従う分布

確率密度関数

$$p_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\ \Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad (-\infty < x < \infty)$$

証明 確率変数 X と Y の同時確率密度関数は

$$p_X(x)p_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \cdot \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}$$

となる．ここで，変数変換

$$T = \frac{X}{\sqrt{Y}n}, \qquad Y' = Y$$

を施すとヤコビアンは

$$J = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} \\ \frac{\partial x}{\partial y'} & \frac{\partial y}{\partial y'} \end{vmatrix} = \begin{vmatrix} \sqrt{y'}n & \frac{t}{2\sqrt{ny'}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{y'}{n}}$$

となる．よって，新たな確率変数 T と Y' の同時確率密度関数は

$$p_{TY'}(t,y') = \frac{1}{2^{\frac{n+1}{2}}\sqrt{n\pi}\ \Gamma(\frac{n}{2})}y'^{\frac{n-1}{2}}e^{-\frac{1}{2}y'(1+\frac{t^2}{n})}$$

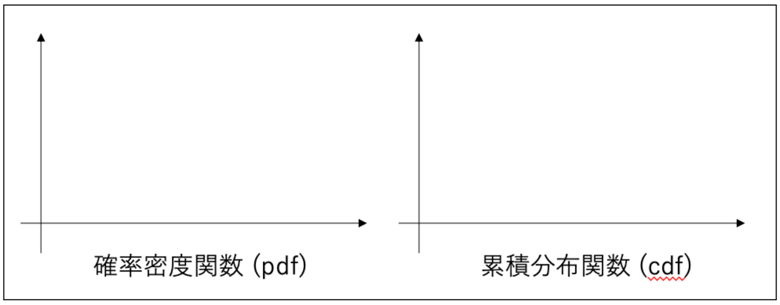

図 2: t_n の pdf と cdf

とできる．したがって T の確率密度関数は

$$p_T(t) = \int_0^\infty p_{TY'}(t, y')dy'$$

で与えられる．ここで，変数変換

$$\frac{1}{2}y'\left(1 + \frac{t^2}{n}\right) = z$$

を施すと t 分布が得られる．

$$p_T(t) = \frac{1}{2^{\frac{n+1}{2}}\sqrt{n\pi}\ \Gamma(\frac{n}{2})}\int_0^\infty \left(\frac{2z}{1+\frac{t^2}{n}}\right)^{\frac{n-1}{2}} e^{-z}\frac{2}{1+\frac{t^2}{n}}dz$$

$$= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\ \Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \int_0^\infty \frac{1}{\Gamma(2)} z e^{-z} dz$$

ここで，被積分関数が Ga(2,1) であるので

$$与式 = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\ \Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

平均・分散

$$\mathrm{E}[X] = 0 \quad (n > 1), \qquad \mathrm{V}[X] = \frac{n}{n-2} \quad (n > 2).$$

9.3 F 分布 $[F_n^m]$

確率変数 X と Y は独立な確率変数で，X は自由度 m のカイ自乗分布 χ_m^2 に従い，Y は自由度 n のカイ自乗分布 χ_n^2 に従う．これらから導かれる

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}$$

が従う分布

確率密度関数

$$p_X(x) = \frac{(\frac{m}{n})^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\beta(\frac{m}{2}, \frac{n}{2})} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, \qquad (0 < x < \infty)$$

証明 確率変数 X と Y の同時確率密度関数は

$$p_X(x)p_Y(y) = \frac{1}{2\Gamma(\frac{m}{2})} \left(\frac{x}{2}\right)^{\frac{m}{2}-1} e^{-\frac{x}{2}} \cdot \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{y}{2}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2}}$$

となる．ここで，変数変換

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}, \qquad Y' = Y$$

を施すとヤコビアンは

$$J = \begin{vmatrix} \frac{\partial x}{\partial f} & \frac{\partial y}{\partial f} \\ \frac{\partial x}{\partial y'} & \frac{\partial y}{\partial y'} \end{vmatrix} = \begin{vmatrix} \sqrt{m} n y' & \frac{m}{n} f \\ 0 & 1 \end{vmatrix} = \frac{m}{n} y'$$

となる．よって，新たな確率変数 F と Y' の同時確率密度関数は

$$p_{FY'}(f, y') = \frac{2^{-\frac{(m+n)}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n} f y'\right)^{\frac{m}{2}-1} y'^{\frac{n}{2}-1} e^{-\frac{1}{2}(\frac{m}{n}f+1)y'} \frac{m}{n} y'$$

とできる．したがって F の確率密度関数は

$$p_F(f) = \int_0^\infty p_{TY'}(t, y') dy'$$

で与えられる．ここで，変数変換

$$\frac{1}{2}\left(1 + \frac{m}{n} f\right) y' = z$$

を施すと F 分布が得られる．

$$p_F(f) = \frac{2^{-\frac{(m+n)}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} f^{\frac{m}{2}-1} \int_0^\infty y'^{\frac{m+n}{2}-1} e^{-\frac{1}{2}(\frac{m}{n}f+1)y'} dy'$$

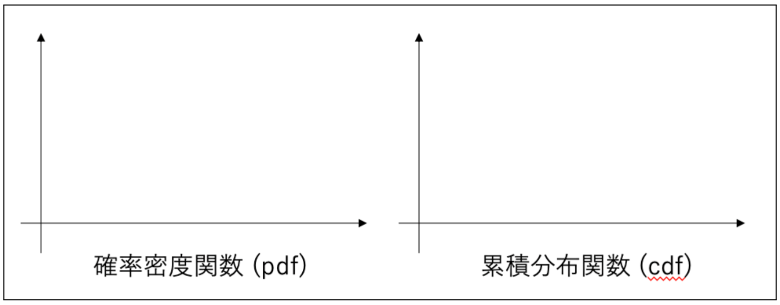

図 3: F_n^m の pdf と cdf

$$
\begin{aligned}
&= \frac{2^{-\frac{(m+n)}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}} f^{\frac{m}{2}-1}\left\{\frac{1}{2}\left(\frac{m}{n}f+1\right)\right\}^{-\frac{m+n}{2}}\int_0^\infty z^{\frac{m+n}{2}-1}e^{-z}dz \\
&= \frac{\Gamma(\frac{m+n}{2})2^{-\frac{(m+n)}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}} f^{\frac{m}{2}-1}\left\{\frac{1}{2}\left(\frac{m}{n}f+1\right)\right\}^{-\frac{m+n}{2}} \\
&= \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} f^{\frac{m}{2}-1}}{\beta(\frac{m}{2},\frac{n}{2})}\left(1+\frac{m}{n}f\right)^{-\frac{m+n}{2}}
\end{aligned}
$$

平均・分散

$$\mathrm{E}[X] = \frac{n}{n-2} \quad (n>2), \qquad \mathrm{V}[X] = \frac{2n^2(m+n-2)}{n(n-2)^2(n-4)} \quad (n>4).$$

その他 確率変数 F が F_n^m に従うとき，確率変数 $\frac{1}{F}$ は F_m^n に従う．

10 標本と統計量

母平均 (population mean)$\mathrm{E}[X]=m$，母分散 (population variance)$\mathrm{V}[X]=\sigma^2<\infty$ をもつ母集団から抽出した大きさ n の確率変数 X_i, $(i=1,2,\cdots,n)$ から導かれる標本平均 $\bar{X}_n$ の平均と分散は

$$\mathrm{E}\left[\bar{X}_n\right] = m, \qquad \mathrm{V}\left[\bar{X}_n\right] = \frac{\sigma^2}{n}$$

であった．ここで，標本分散 (sample variance) を

$$\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

で定義すると，標本分散 $\bar{\sigma}^2$ の平均は

$$\mathrm{E}[\bar{\sigma}^2] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}X_i^2\right] - \mathrm{E}\left[\bar{X}_n^2\right]$$

$$= (\sigma^2 + m^2) - \left(\frac{\sigma^2}{n} + m^2\right) = \frac{n-1}{n}\sigma^2$$

となる．これらの結果，標本平均 $\bar{X}_n$ の平均は母平均 m に一致することから偏りなく分布しているが，標本分散の平均は母分散とは一致せず偏った分布をしていることがわかる．このことは標本分散 $\bar{\sigma}^2$ が母分散 σ^2 を必ずしも反映していないことを示している．そこで，標本分散 $\bar{\sigma}^2$ を修正した

$$\tilde{\sigma}_n^2 = \frac{n}{n-1}\bar{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

を標本不偏分散として定義する．当然のことながら，標本不偏分散 $\tilde{\sigma}_n^2$ の平均は母分散 σ^2 と一致することとなる．

$$\mathrm{E}\left[\tilde{\sigma}_n^2\right] = \sigma^2$$

確率変数 $X_i,\ (i = 1, 2, \cdots, n)$ が正規分布 $N(m, \sigma^2)$ に従うものとする．このとき，標本平均 $\bar{X}_n$ と標本分散 $\bar{\sigma}^2$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

は独立であり，それぞれ以下の分布に従う．

$$\frac{\sqrt{n}\left(\bar{X}_n - m\right)}{\sigma} \sim N(0,1), \qquad \frac{n\bar{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

証明 確率変数 $X_i,\ (i = 1, 2, \cdots, n)$ を Z 変換 (Z-transform) により標準化した確率変数 Z_i を

$$Z_i = \frac{X_i - m}{\sigma}, \qquad (i = 1, 2, \cdots, n)$$

とすると，確率変数 $Z_i,\ (i = 1, 2, \cdots, n)$ は互いに独立で同一な分布 $N(0, 1)$ に従う．よって，確率ベクトル $\mathbf{Z} = [Z_1, Z_2, \cdots, Z_n]^{\mathrm{T}} \in \Re^n$ の同時確率密度関数は

$$p_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^{n}(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}z_i^2} = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{z}}$$

となる．ここで，以下のような確率ベクトル $\mathbf{Z}$ の直交変換 $\mathbf{Y} = \mathcal{D}\mathbf{Z}$ を考える．

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \vdots & \ddots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$$

もちろん，演算子 $\mathcal{D}$ は $\mathcal{D}^{\mathrm{T}}\mathcal{D} = \mathcal{D}\mathcal{D}^{\mathrm{T}} = \mathbf{I}$ を満足するものである．したがって，確率ベクトル $\mathbf{Y}$ の同時確率密度関数は変数変換 $\mathbf{Z} = \mathcal{D}^{\mathrm{T}}\mathbf{Y}$ のヤコビアンが

$$J = |\mathcal{D}^{\mathrm{T}}| = \pm 1$$

であるので，

$$p_{\mathbf{Y}}(\mathbf{y}) = p_{\mathbf{Z}}(\mathcal{D}^{\mathrm{T}}\mathbf{y}) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(\mathcal{D}^{\mathrm{T}}\mathbf{y})^{\mathrm{T}}(\mathcal{D}^{\mathrm{T}}\mathbf{y})} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{y}}$$

$$= \prod_{i=1}^{n} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y_i^2}$$

となる．よって，確率変数 $Y_i,\ (i = 1, 2, \cdots, n)$ は互いに独立で同一の確率分布 $N(0,1)$ に従うことがわかる．また，直交行列 $\mathcal{D}$ の第一行を考えると

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i = \frac{\sqrt{n}\left(\bar{X}_n - m\right)}{\sigma} \sim N(0,1)$$

であることがわかる．さらに，

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \frac{(X_i - m)}{\sigma^2} = \sum_{i=1}^{n} \frac{\left\{\left(X_i - \bar{X}_n\right) + \left(\bar{X}_n - m\right)\right\}}{\sigma^2}$$

$$= \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}_n\right)}{\sigma^2} + \frac{n\left(\bar{X}_n - m\right)}{\sigma^2} = \frac{n\bar{\sigma}^2}{\sigma^2} + Y_1^2$$

であるので

$$\frac{n\bar{\sigma}^2}{\sigma^2} = \sum_{i=2}^{n} Y_i \sim \chi^2_{n-1}$$

であることが示される．標本平均 $\bar{X}_n$ は Y_1 にのみ依存し，標本分散 $\bar{\sigma}^2$ は $Y_2, Y_3, \cdots, Y_n$ にのみ依存するので，これらは独立である．

確率変数 $X_i,\ (i = 1, 2, \cdots, n)$ が正規分布 $N(m, \sigma^2)$ に従うものとする．このとき，標本平均 $\bar{X}_n$ と標本不偏分散 $\tilde{\sigma}_n^2$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad \tilde{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

を用いて T 変換 (T-transform) した確率変数 T は自由度 $n-1$ の t 分布に従う．

$$T = \frac{\sqrt{n}\left(\bar{X}_n - m\right)}{\tilde{\sigma}_n} \sim t_{n-1}$$

11　問題 7

7.1　確率変数 X は一様分布 $[U(0,1)]$ に従うものとする．これから導かれる

$$Y = -2\log X$$

が従う分布は指数分布 $[E_x(\frac{1}{2})]$，すなわち，自由度 2 の χ^2 分布 $[\chi_2^2]$ となることを示せ．

7.2　確率変数 X と Y は独立な確率変数で，ともに標準正規分布 $[N(0,1)]$ に従うものとする．これから導かれる

$$C = \frac{X}{Y}$$

が従う分布はコーシー分布 $[Cy(0,1)]$ となることを示せ．

7.3　確率変数 X と Y は独立な確率変数で，それぞれガンマ分布 $[Ga(\alpha,\sigma)]$ とガンマ分布 $[Ga(\beta,\sigma)]$ に従うものとする．これから導かれる

$$Z = \frac{X}{X+Y}$$

が従う分布を求め，何分布となるかを述べよ．

7.4　(二次元正規分布)　確率変数 X_1 と X_2 が独立な確率変数で，ともに標準正規分布 $[N(0,1)]$ に従うものとする．これから導かれる

$$Y_1 = aX_1 + bX_2, \qquad Y_2 = cX_1 + dX_2$$

について以下の問いに答えよ．ただし，$a,\ b,\ c,\ d$ は定数であり，$ad-bc \neq 0$ である．また，既に前問で使用した記号は用いてよい．

(1)　$Y_1,\ Y_2$ の期待値 $m_1,\ m_2$，ならびに分散 $\sigma_1^2,\ \sigma_2^2$ を求めよ．

(2)　$Y_1,\ Y_2$ の共分散 σ_{12}^2，ならびに相関係数 ρ を求めよ．

(3)　$Y_1,\ Y_2$ の同時確率密度関数 $p_{Y_1Y_2}(y_1,y_2)$ を求めよ．

7.5 (ショックモデル) システム $S_1,\ S_2$ の寿命を確率変数 $X_1,\ X_2$ で表す．これらは互いに独立で指数分布 $[E_x(\lambda_1)]$ と指数分布 $[E_x(\lambda_2)]$ に従うものとする．さらに，システムはショックを受けると故障することがある．ショックの発生間隔 Y は指数分布 $[E_x(\tau)]$ に従うものとする．システムの寿命が尽きるか，それまでにショックで故障する場合の時間は

$$X'_1 = \min(X_1, Y), \qquad X'_2 = \min(X_2, Y)$$

で与えられる．以下の問いに答えよ．

(1) (X_1, X_2) の同時確率分布関数を求めよ．

(2) X_1 と X_2 の相関係数 $\rho(X_1, X_2)$ を求めよ．

(3) $S_1,\ S_2$ が直列に結合されているシステム全体の寿命を確率変数 Z とする．確率変数 Z が従う確率分布関数，ならびにその平均を求めよ．

(4) $S_1,\ S_2$ が並列に結合されている場合はどうなるか．