第 2 章 確率の復習

James H. Stock 著・Mark W. Watson 著・宮尾 龍蔵 訳

『入門計量経済学』（共立出版、2016）
https://www.kyoritsu-pub.co.jp/book/b10003746.html

2025-05-08

はじめに

- 本章の目的は、**回帰分析と計量経済学の理解に不可欠な確率の基本概念**を復習することである。
- 入門レベルの確率論・統計学コースを履修済みという前提で説明を進める。
- 確率理論は、**確率的な現象を数量化して表現するための数学的な分析道具**である。

2.1 確率変数と確率分布

確率変数と確率分布

- **事象 (outcomes)**：起こりうる互いに重複しない結果。
- **標本空間 (sample space)**：起こりうるすべての事象の集合。
- **イベント (event)**：標本空間の中の部分集合（1 つあるいは複数の事象から構成される）。
- **確率 (probability)**：ある事象が発生する時間的な割合（長い時間をかけてみたときの）。
- **確率変数 (random variable)**：確率的な事象を数値を使って表現したもの。
  - 例：パソコンの故障回数、次に会う人の性別（数値化して）。
- 確率変数には、値が離れた**離散的な確率変数 (discrete random variable)** と、連続した値をとる**連続的な確率変数 (continuous random variable)** がある。

離散的な確率変数の確率分布

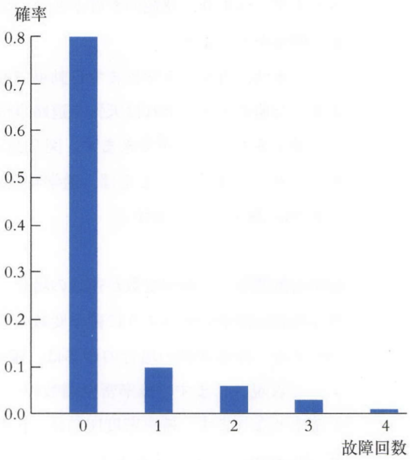
- **確率分布 (probability distribution)**：離散的な確率変数が取り得るすべての値のリストと、それぞれの値が起こる確率を表したもの。それぞれの確率の合計は 1 となる。
  - 例：パソコンが M 回故障する確率分布。
- イベントの確率：対応する事象（値）の確率を合計することで求められる。
  - 例：「1 回もしくは 2 回故障する」確率 $P(M = 1 \text{ または } M = 2) = P(M = 1) + P(M = 2) = 0.10 + 0.06 = 0.16$。
- **累積確率分布 (cumulative probability distribution)** または **累積分布関数 (cumulative distribution function, c.d.f.)**：確率変数がある値以下となる確率。
  - 例：最大で 1 回の故障が起こる確率 $P(M \leq 1) = P(M = 0) + P(M = 1) = 0.80 + 0.10 = 0.90$。
- **ベルヌーイ分布 (Bernoulli distribution)**：確率変数が二者択一（0 か 1）の場合の重要な離散確率分布。

表 2.1　パソコンが M 回故障する確率

	事象（故障回数）				
	0	1	2	3	4
確率分布	0.80	0.10	0.06	0.03	0.01
累積確率分布	0.80	0.90	0.96	0.99	1.00

図 2.1　パソコン故障回数の確率分布

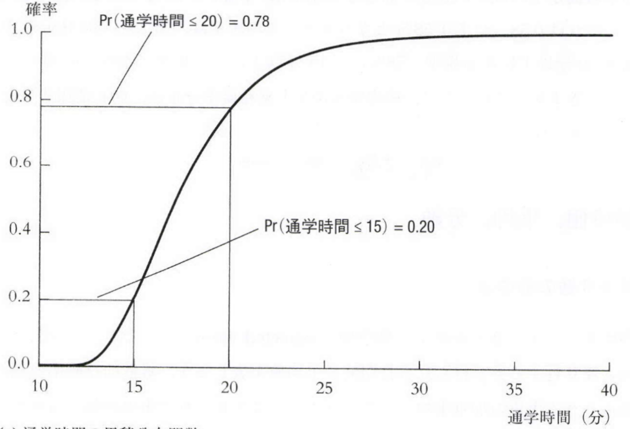
棒グラフのそれぞれの高さは，それぞれの回数だけパソコンが故障する確率を表す．棒グラフの最初の値は 0.80 なので，回数 0 の確率は 80%．棒グラフの次の値は 0.1，したがって回数 1 となる確率は 10%．



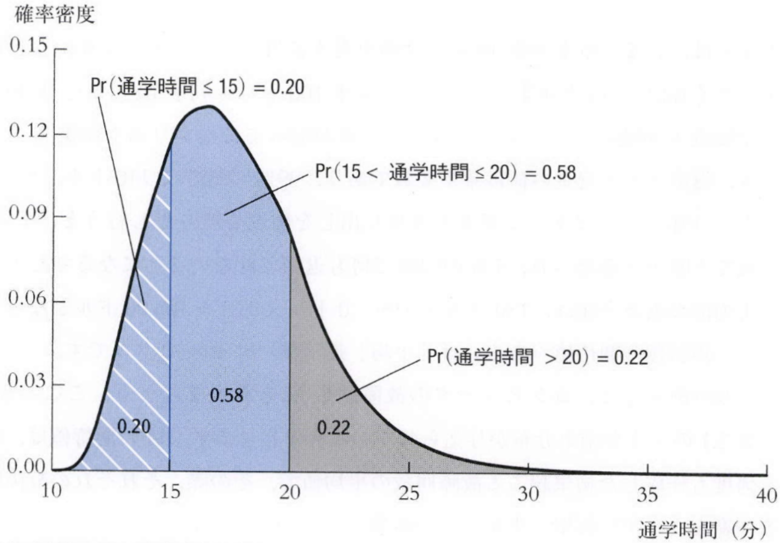
連続的な確率変数の確率分布

- **累積確率分布 (cumulative probability distribution)**：連続的な確率変数の値が特定の値以下である確率。定義は離散の場合と同様。
  - 例：通学時間が 15 分以内となる確率。
- **確率密度関数 (probability density function, p.d.f.)**：連続変数の確率を表現するために用いられる。
  - 確率密度関数の 2 点の間の面積は、確率変数がその 2 点の間に入る確率となる。
  - 例：通学時間の p.d.f.。15 分から 20 分の間となる確率は、p.d.f. のその区間の下の面積に相当する。
- 確率密度関数と累積確率分布は、同じ情報を異なった表現で表すものである。

図 2.2　通学時間の累積分布関数と確率密度関数



(a) 通学時間の累積分布関数



(b) 通学時間の確率密度関数

2.2 期待値, 平均, 分散

期待値、平均、分散

- **期待値 (expected value, $E(Y)$)**：試行を何回も繰り返すことで得られる長期的な平均値。
  - 離散的な確率変数の場合、起こりうる事象の加重平均（ウエイトはその事象が起こる確率)。
  - Y の期待値は **平均 (mean)** または**期待 (expectation)** とも呼ばれ、μ_Y と表記される。
  - **公式**: $E(Y) = \sum_{i=1}^{k} y_i P(Y = y_i)$
  - 例：パソコン故障回数 M の期待値 $E(M) = 0 \times 0.80 + 1 \times 0.10 + 2 \times 0.06 + 3 \times 0.03 + 4 \times 0.01 = 0.35$。
  - ベルヌーイ確率変数 G (成功確率 p) の期待値: $E(G) = 1 \times p + 0 \times (1 - p) = p$。

分散と標準偏差

- **分散 (variance, $var(Y)$ または σ_Y^2)**：確率分布の散らばり度合い、またはその「幅」を測るもの。
  - Y とその平均 μ_Y との差の二乗に関する期待値として定義される。
  - **定義**: $var(Y) = E[(Y - \mu_Y)^2]$
  - **離散的な確率変数の公式**: $var(Y) = \sum_{i=1}^{k} (y_i - \mu_Y)^2 P(Y = y_i)$
- **標準偏差 (standard deviation, σ_Y)**：分散に平方根を取った値。散らばりを測る指標としてよく利用される。標準偏差の単位は Y と同じである。
- 例：パソコン故障回数 M の分散と標準偏差。
- ベルヌーイ確率変数 G (成功確率 p) の分散: $var(G) = p(1-p)$。標準偏差: $\sigma_G = \sqrt{p(1-p)}$。

確率変数の線形関数に関する平均と分散

- 確率変数 X の平均を μ_X, 分散を σ_X^2 とする。
- Y が X の線形関数 $Y = \alpha + \beta X$ で表されるとき (α, β は定数)、
  - Y の平均は: $E(Y) = \alpha + \beta E(X) = \alpha + \beta\mu_X$
  - Y の分散は: $var(Y) = \beta^2 var(X) = \beta^2\sigma_X^2$
  - Y の標準偏差は: $\sigma_Y = |\beta|\sigma_X$
- 例：税引き後所得 $Y = 2000 + 0.8X$ の平均と標準偏差。

確率分布の形状に関する他の指標

- 平均と標準偏差は分布の中心と散らばりを測る。
- **歪度 (skewness)**：分布の歪み・左右非対称の度合いを測る指標。
  - 対称な分布の歪度はゼロである。
  - 右側のすそが長いと歪度は正、左側のすそが長いと歪度は負となる。
  - 歪度は単位のない値である。
- **尖度 (kurtosis)**：分布のすその厚み（重さ）を測る指標。異常値 (outlier) の起こる可能性を表す。
  - 正規分布に従う確率変数の尖度は **3** である。
  - 3 を超える尖度を持つ分布は、正規分布よりもすそが厚く、急尖的 (leptokurtic) または「厚いすそ (heavy tail)」と呼ばれる。
  - 尖度も単位のない値である。

2.3 2つの確率変数

2 つの確率変数：結合・限界・条件付分布

- 経済問題は複数の変数に関わることがほとんどである。
- **結合確率分布 (joint probability distribution)**：2 つの離散的な確率変数 X, Y が同時に特定の値 (x, y) を取る確率を示す。$P(X = x, Y = y)$ で表現される。
  - 取りうるすべての (x, y) の確率を合計すれば 1 になる。
  - 例：天気と通学時間の結合分布。
- **限界確率分布 (marginal probability distribution)**：確率変数 Y だけの確率分布のことである。
  - 結合分布から計算できる。$P(Y = y) = \sum_i P(X = x_i, Y = y)$。
  - 例：通学時間が長くなる限界確率 $P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = 0.15 + 0.07 = 0.22$。

表 2.2　天気と通学時間の結合分布

	雨が降る（$X = 0$）	雨は降らない（$X = 1$）	合計
通学時間が長い（$Y = 0$）	0.15	0.07	0.22
通学時間が短い（$Y = 1$）	0.15	0.63	0.78
合計	0.30	0.70	1.00

表 2.3 パソコンの故障回数（M）と年式（A）の結合分布と条件付分布

A. 結合分布

	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	合計
古いパソコン（$A=0$）	0.35	0.065	0.05	0.025	0.01	0.50
新しいパソコン（$A=1$）	0.45	0.035	0.01	0.005	0.00	0.50
合計	0.8	0.1	0.06	0.03	0.01	1.00

B. A が与えられた下での M の条件付分布

	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	合計
$\Pr(M\|A=0)$	0.70	0.13	0.10	0.05	0.02	1.00
$\Pr(M\|A=1)$	0.90	0.07	0.02	0.01	0.00	1.00

条件付分布

- **条件付分布 (conditional distribution)**：別の確率変数 X が特定の値を取るという条件の下で導出される確率変数 Y の分布。
  - X が x という値を取るとき、Y が y という値を取る条件付確率は $P(Y=y|X=x)$ と表記される。
  - **定義**: $P(Y=y|X=x)=\frac{P(X=x,Y=y)}{P(X=x)}$
  - 例：雨が降る $(X=0)$ という条件下で通学時間が長くなる $(Y=0)$ 確率 $P(Y=0|X=0)=\frac{P(X=0,Y=0)}{P(X=0)}=\frac{0.15}{0.30}=0.50$。

条件付期待値

- **条件付期待値 (conditional expectation of Y given X, $E(Y|X=x)$)** または**条件付平均 (conditional mean of Y given X)**：X が与えられた下での Y の条件付分布の平均。
  - **離散的な場合の公式**:
    $E(Y|X=x) = \sum_{i=1}^{k} y_i P(Y = y_i | X = x)$
  - 例：古いパソコン ($A=0$) の下で予想される故障回数
    $E(M|A=0) = 0 \times 0.70 + 1 \times 0.13 + \cdots + 4 \times 0.02 = 0.56$。

繰り返し期待値の法則

- **繰り返し期待値の法則 (law of iterated expectations)**：Y の平均は、X が与えられた下での Y の条件付期待値の加重平均（ウエイトは X の確率分布）である。
  - **法則**: $E(Y) = E[E(Y|X)]$
  - 数式表現（X が m 個の値を取る場合）：$E(Y) = \sum_{j=1}^{m} E(Y|X = \chi_j)P(X = \chi_j)$
  - 例：パソコン故障回数 M の平均 $E(M) = E(M|A = 0)P(A = 0) + E(M|A = 1)P(A = 1) = 0.56 \times 0.50 + 0.14 \times 0.50 = 0.35$。

条件付分散

- **条件付分散 (variance of Y conditional on X, $var(Y|X=x)$)**：X が与えられた下での Y の条件付分布の分散。
  - **離散的な場合の公式**: $var(Y|X=x) = \sum_{i=1}^{k}[y_i - E(Y|X=x)]^2 P(Y=y_i|X=x)$
  - 例：古いパソコン ($A=0$) の下での故障回数の条件付分散は 0.99。

確率変数の独立、共分散、相関

- **独立 (independence)**：2 つの確率変数 X, Y があり、一方の値を知ることでもう一方の情報が何も得られないとき、両者は独立である。
  - X が与えられた下での Y の条件付分布が Y の限界分布と等しいとき、独立となる。つまり、$P(Y = y|X = x) = P(Y = y)$ がすべての x, y について成立する場合。
  - X, Y が独立であれば、結合分布は限界分布の積で表される: $P(X = x, Y = y) = P(X = x)P(Y = y)$。
- **共分散 (covariance, $cov(X, Y)$ または σ_{XY})**：2 つの確率変数が共に動く程度を測る尺度。
  - **定義**: $cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
  - X, Y が独立なら、共分散はゼロとなる。

相関

- **相関 (correlation, $corr(X,Y)$)**：共分散をそれぞれの標準偏差で割ったもの。共分散の単位問題を解決する指標。
  - **定義**: $corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$
  - 相関は常に -1 から 1 の間の値をとる。$-1 \leq corr(X,Y) \leq 1$。
  - **無相関 (uncorrelated)**：$corr(X,Y) = 0$ のとき。
  - Y の条件付平均が X に依存しない場合、Y と X は無相関となる ($E(Y|X) = \mu_Y$ のとき $cov(X,Y) = 0$)。ただし逆は必ずしも成立しない。

確率変数の和に関する平均と分散

- 2 つの確率変数 X, Y の和の平均は、それぞれの平均を合計したものになる。
  - $E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$
- X, Y の和の分散は、それぞれの分散を合計したものに両者の共分散を加えたものになる。
  - $var(X+Y) = var(X) + var(Y) + 2cov(X,Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$
- もし X と Y が独立ならば共分散はゼロなので、和の分散はそれぞれの分散の和になる。
  - $var(X+Y) = var(X) + var(Y) = \sigma_X^2 + \sigma_Y^2$ (X, Y が独立の場合)
- ウエイト付きの和に関する平均、分散、共分散の公式も存在する。

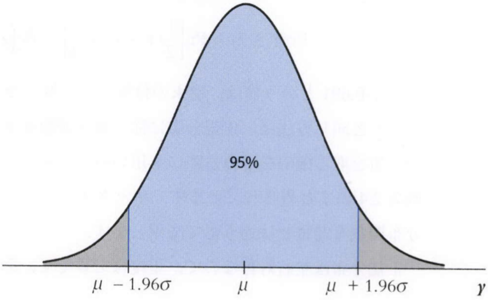
2.4 正規分布、カイニ乗分布、ステューデントの t 分布、F 分布

計量経済学で重要な確率分布 (1/3)

- **正規分布 (normal distribution)**：連続的な確率変数が従う分布で、よく知られる**釣鐘型の形状**をしている。$N(\mu, \sigma^2)$ と表記される。
  - 平均 μ, 分散 σ^2 を持つ。
  - **標準正規分布 (standard normal distribution)**：平均 $\mu = 0$, 分散 $\sigma^2 = 1$ を持つ正規分布 ($N(0, 1)$)。
  - 正規分布に従う変数は、平均を差し引き標準偏差で割ることで**標準化 (standardize)** すると、標準正規分布に従う。$Z = (Y - \mu)/\sigma$。
  - 確率の計算には標準正規分布の累積分布関数 (c.d.f., Φ) の表を利用する。例：$P(Y \leq y) = P(Z \leq (y - \mu)/\sigma) = \Phi((y - \mu)/\sigma)$。
  - 多変数正規分布：複数の確率変数の結合分布への拡張。

図 2.5 正規分布の確率密度関数

平均 μ, 分散 σ^2 を持つ正規分布の確率密度関数は, μ を中心とした釣鐘型の曲線を描く. 密度関数の $\mu - 1.96\sigma$ と $\mu + 1.96\sigma$ の間の面積は 0.95. この正規分布は, $N(\mu, \sigma^2)$ と表記される.



計量経済学で重要な確率分布 (2/3)

- **カイ二乗分布 (chi-squared distribution, χ^2_ν)**：標準正規分布に従う ν 個の独立した確率変数をそれぞれ二乗して合計した値が従う分布。「**二乗和**」の分布である。
  - 自由度 ν に依存する。
  - 例: Z_1, Z_2, Z_3 が独立な標準正規分布に従うとき、$Z_1^2 + Z_2^2 + Z_3^2$ は自由度 3 のカイ二乗分布 χ^2_3 に従う。

計量経済学で重要な確率分布 (3/3)

- **ステューデント t 分布 (Student’s t distribution, t_ν)**：標準正規変数と、それとは独立な自由度 ν のカイ二乗変数を ν で割った値の平方根との比率が従う分布。
  - 自由度 ν に依存する。
  - 釣鐘型だが、ν が小さいときは正規分布よりすそが厚い性質を持つ。ν が大きくなると標準正規分布に近似する。
- **F 分布 (F distribution, F_{ν_1,ν_2})**：自由度 ν_1 のカイ二乗変数を ν_1 で割ったものと、それとは独立な自由度 ν_2 のカイ二乗変数を ν_2 で割ったものとの比率が従う分布。
  - 分子の自由度 ν_1 と分母の自由度 ν_2 に依存する。
  - 分母の自由度 ν_2 が十分大きい場合、F 分布はカイ二乗分布を自由度で割ったものに近似する。

2.5 無作為抽出と標本平均の分布

無作為抽出と i.i.d.

- 計量手法のほとんどは、**母集団 (population)** から抽出された**標本 (サンプル)** データに基づいている。
- **単純な無作為抽出 (simple random sampling)**：母集団のどのメンバーも、標本として選ばれる確率が同じになるようにランダムに抽出する方法。
- 単純な無作為抽出の下では、抽出された　個の観測値 $Y_1, \dots, Y_n$ は、**独立かつ同一の分布 (independently and identically distributed, i.i.d.)** に従う。
  - **独立 (independently distributed)**：ある観測値の値を知ることが、他の観測値に関して何ら情報を持たないこと。
  - **同一の分布 (identically distributed)**：各観測値の限界分布（母集団における Y の分布）がすべて等しいこと。

標本平均と標本分布

- **標本平均 (sample mean, $\bar{Y}$)**： 個の観測値 $Y_1, \ldots, Y_n$ の平均。
  - $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$
- 無作為抽出により標本が異なるごとに観測値 Y_i は異なるため、**標本平均 $\bar{Y}$ 自体も確率変数**となる。
- **標本分布 (sampling distribution)**：標本平均 $\bar{Y}$ の確率分布のこと。異なる標本から得られるであろうさまざまな $\bar{Y}$ の値に関する確率分布である。

標本平均の平均と分散 (i.i.d. の場合)

- 観測値 $Y_1, \ldots, Y_n$ が i.i.d. で、母集団（各 Y_i）の平均を μ_Y, 分散を σ_Y^2 とする。
- 標本平均 $\bar{Y}$ の期待値（平均）は、母集団の平均に等しくなる。
  - $E(\bar{Y}) = \mu_Y$
- 標本平均 $\bar{Y}$ の分散は、母集団の分散を標本数 n で割ったものになる。
  - $var(\bar{Y}) = \sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$
- 標本平均 $\bar{Y}$ の標準偏差は、$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$。
- これらの結果は、母集団の分布がどのようなものであっても（正規分布でなくても）成り立つ。

2.6 大標本の場合の標本分布の近似

大標本の場合の標本分布の近似

- 標本分布を求めるアプローチには「正確な (exact)」アプローチと「近似的な (approximate)」アプローチがある。
- 「正確な」標本分布は、標本数 n がどんな値でも正確に成り立つ公式である。一般に母集団の分布に依存し複雑である。
- 「近似的な」標本分布（**漸近分布 (asymptotic distribution)**）は、**標本数 n が大きい場合**に適用できる近似である。
  - 漸近分布はシンプルで、特に標準化された標本平均の分布は母集団の分布に依存しない。
  - 計量分析で使う標本数は大きいことが多いので、この近似は信頼できるものとなる。
- 大標本近似に使われる 2 つの基本ツール: **大数の法則**と**中心極限定理**。

大数の法則 (Law of Large Numbers)

- **大数の法則 (law of large numbers)**：一般的な条件の下で、標本数 n が大きいとき、標本平均 $\bar{Y}$ は母集団の平均 μ_Y に非常に高い確率で近づくという性質である。
- これは**確率における収束 (convergence in probability)** と呼ばれ、$\bar{Y}$ は μ_Y に対して**一致性 (consistency)** を持つと言う。
  - 記号で $\bar{Y} \xrightarrow{p} \mu_Y$ と表記されることもある。
- 大数の法則に必要な主な条件は、観測値 Y_i が **i.i.d.** であることと、分散 σ_Y^2 が有限であることである。

中心極限定理 (Central Limit Theorem)

- **中心極限定理 (central limit theorem)**：ある一般的な条件の下、標本数 n が大きいとき、標本平均 $\bar{Y}$ の分布は正規分布でうまく近似できるというものである。
  - $\bar{Y}$ の平均は μ_Y, 分散は $\sigma^2_{\bar{Y}} = \sigma^2_Y/n$ である。
  - 定理によれば、n が大きいとき、$\bar{Y}$ の分布は近似的に $N(\mu_Y, \sigma^2_Y/n)$ に従う。
- さらに重要なのは、標準化された標本平均 $(\bar{Y} - \mu_Y)/\sigma_{\bar{Y}}$ の標本分布は、n が大きいとき近似的に標準正規分布 $N(0, 1)$ に従うということである。
- この近似は、たとえ母集団の分布が正規分布でなくとも成り立つ。

中心極限定理の重要性

- 標準化された標本平均が正規分布に近似するという性質は、**母集団の分布に依存しない**という点で非常に重要である。
- これにより、母集団の分布が未知であっても、大標本の下では標本平均に関する確率計算や統計的推論に正規分布を利用できるようになる。
- 中心極限定理による正規分布への近似は、現代の統計分析に不可欠な考え方であり、**本書の回帰分析理論の基礎**となる。
- $\bar{Y}$ は **漸近的に正規分布に従う (asymptotically normally distributed)** と呼ばれる。