

日本赤十字九州国際看護大学/Japanese Red
Cross Kyushu International College of
Nursing

確率分布

メタデータ	言語: Japanese 出版者: 公開日: 2020-08-04 キーワード (Ja): キーワード (En): Statistical thinking, random variable, probability distribution, normal distribution, binomial distribution, Poisson distribution 作成者: 守山, 正樹 メールアドレス: 所属:
URL	https://jrckicn.repo.nii.ac.jp/records/714

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



第1章 統計と確率分布

皆さん、こんにちは。私が始めて皆さんにお会いしたのは昨年（2019）6月です。覚えていますか。基礎力総合ゼミの時間にWifiそしてプチプチからの問題提起をしました。さて今日からは、統計学の勉強を始めます。統計学は何かを測定・観察し、結果を数値で表し、集団や社会について考えていく科学です。

1 考え方の特徴

統計学に特徴的な考え方とは何でしょうか。統計学はこの世界の様々な事象を数値で表した上で、事象が「ランダムに・偶然に・確率的に」起きると考え、その起き方の法則を追究します。「全ての事象が偶然性・確率に支配されている」と聞くと「本当？」と疑う人もいるかもしれません。「私自身や私が住む世界は今ここに現存する；私は確率的な存在じゃない！；私も世界も偶然ではなく必然です」と思う人もいるでしょう。他方、今の新型コロナウイルス（COVID-19）の突然の流行を振り返ると「人類とウイルスの偶然の出会い」やその後の「ウイルスの確率的な変異」が世界を揺り動かしていることも事実です。ですから「偶然性・確率を基礎とする統計学の考え方」は今の激動する世界を生きる上でとても大切です。その統計学の中心になるのが、事象を数値（離散量または連続量）で捉え、それらの起こる確率の広がりを数学的に把握する確率分布の考え方です。

2 確率変数

統計学ではこの世界の「確率的に起こる事象E」は「①それが取り得る様々な数値（変数；x）を用いて E_x と表せる、②そうした数値には、その数値が現れる確率 $P(E_x)$ が対応する」と考え、その変数（x）を「確率変数」といいます。確率変数は、試験合否やコイン裏表のように0,1,2など自然数で表わせる離散型確率変数（離散量）と、身長・体重のように小数点以下何桁までも連続する値で表わせる連続型確率変数（連続量）に分かれます。

この世界の事象→ E_x x →離散量（離散型確率変数）、連続量（連続型確率変数）

1) 離散型確率変数（離散量）

その値が1と2だけとか、とびとびの値のみを取り、間の値をとることがない変数が離散量です。このコインには裏と表しかありません。コインを投げると、真ん中で止まることはなく、必ず裏か表が出ます。今度はサイコロです。サイコロは6面があり、投げると1から6のどれかが上になります。1.5といった値はとりません。これが離散量です。

事象（出来事）がコインの裏表のように、互いに排反する2項目しかない離散量は、私たちの毎日でも、看護や医療の世界でも広く出て来ます。試験の合格／不合格、ヒトの生死、疾患の有無は離散量です。皆さんの健康チェックでも離散量が活躍します；喉の痛みの有無、嗅覚の異常の有無、37.5度C以上の発熱の有無、これらも離散量です。以前は性別も、男性か女性か二つの値の離散量でした。現在は二つ以上の値を持つ離散量と位置付けられます。

2) 連続型確率変数（連続量）

サイコロの目のような、とびとびの値しかとらない離散量に対して、小数点以下何桁までも表すことができるデータが連続量です。

例えば私の手の人差し指の長さを計ってみると 7.1cm あります。7.1cm は連続量ですから 7 cm と 8 cm の間にあり、小数点以下をもっと精密に測定しようとする、理論的には無限に細かくすることが可能です。皆さんの場合はどうでしょうか。自分の指の長さを測り、連続量として表してみてください。自分の体だけでも様々な連続量を見出すことができます。挙げてみてください。

3 確率分布

1) 事例から確率分布へ

次は確率分布についてお話しします。分布とは「複数の事象が、ある広がりを持って存在するとき、その広がり」を示します。何か1回・1例だけ存在する事例の場合、分布という考え方は使いません。まず単独事例と分布について、考え方の違いを説明します。

「コインを投げて裏がでた、電車の最初の乗客が女性だった、ある学生の身長を測ったら 160.0 cm だった」などは単独の事例です。こうした事例を出発点として、詳しく聴き取り、記述を大切に進める事例研究は、看護でもよく用いられる研究方法です。一方、統計学で注目するのは、母集団における平均的事象／平均的個体です。1例だけで「コインは裏が出やすい、電車の乗客は女性が多い、学生の身長は 160.0cm！」とは結論しません。2回・3回・4回～n回と投げる試み（試行）、観察の試行を繰り返し重ねることで、初めて「コインは裏と表が同じ確率 0.5 が出る」「電車の乗客は 60%が女性だ」「学生の身長は平均 162.0cm」などと結論できます。

サイコロ（離散量）であれば投げる試み（試行）を繰り返し、身長や体重（連続量）であれば、一人二人と測る試みを増やすことで、いくつもの値が得られ、全体の広がり・分布が見えてきます。それが確率分布です。

2) 離散量に対応する確率分布

離散量は事象（出来事）の起こり方から得られますが、起こり方は一種類ではありません。起こり方に対応して、ここでは二つの確率分布を示します。二項分布とポアソン分布です。

(1) 二項分布

コインの裏表、生死、疾患の有無など、取り得る場合が2項目しかない離散量に対応する確率分布が二項分布です。

コイン投げを例にとります。2回・3回・4回～n回とコインを投げる試み（試行）を増やし、N回振って、何回表が出たかを横軸に、またその確率を縦軸にとってヒストグラム（棒グラフの一種）を描くと、山の形の確率分布が現れます。これを二項分布といいます。

二項分布の例としては、コイン投げの他に、視聴行動（ある番組を見ない0、見た1）、投票行動（ある候補に投票しない0、投票する1）、治療効果（ある治療が効かない0、効く1）などがあり、どれも選択（どちらかを選ぶ行為）に関連しています。

二項分布を描いてみましょう。以下はアメリカのアイオワ大学によるウェブサイトです。n = 生起確率 p と試行回数 n を入力すると、対応する二項分布のグラフを描いてくれます。たとえば、

ある治療が効く確率 p を 0.6、その治療を試した患者さんの数 n を 20 などと入力して、その条件に合わせた二項分布を描いてみましょう。

<https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html>

(2) ポアソン分布

19 世紀に活躍したフランス人の数学者、シメオン・デニス・ポアソンは、ラバ蹴られて亡くなったフランス軍の死亡者の発生の確率分布を研究し、1837 年にポアソン分布を発表しました。事故で亡くなる人の発生は、「どちらかを選ぶ際の離散量」ではなく「自然現象が発生する際の離散量」です。「ある時間内やある領域内で、ときどき発生する自然現象の回数」から求められるのがポアソン分布です。ポアソン分布の例としては「時間内の来客数・来院者数」「時間内の電話相談件数」「空気の体積当たりの特定のウイルス数」などが考えられます。

ポアソン分布もアイオワ大学のウェブサイトで描けます。ポアソン分布は試行回数 n が十分に大きく、また生起確率 p が非常に小さいときに導かれる二項分布の極限と考えられます。ポアソン分布を計算するときは、 n と p とを掛け算した値 ($\lambda ; n \times p$) が大切です。この λ の値を指定すると、アイオワ大学のウェブサイトから、ポアソン分布のグラフを描けます。

<https://homepage.divms.uiowa.edu/~mbognar/applets/pois.html>

3) 連続量の確率分布＝正規分布

さて連続量の場合は、観察を重ねると、どのような形のグラフになるでしょうか。

教科書 65 頁の最後を見ると「連続型データの確率変数 x は（離散型データの場合のような 1、2 などではなく）どのような値でもとりうるため、確率の計算は簡単にはできない」と書いてあります。しかし理論的な計算は難しくても、その実例は至るところにあります。

連続型確率変数の確率分布がどのような形になるか、実は皆さんは経験的に知っているはずで、健康診断で測定した皆さんの身長や体重、試験の点数など様々な連続量を、たとえば学年単位でヒストグラムに描いてみてください。釣鐘型／ベル型の分布になるはずで、これを正規分布といいます。

アイオワ大学が提供しているサイトで、正規分布の曲線も描けます。試してみてください。

<https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html>

4 終わりに

さて今日は世界の様々な出来事を統計的に見る考え方の導入として、出来事の起こり方が離散型確率変数または連続型確率変数で表せることをお話ししました。またそれらの値が存在する範囲を目に見える形で示す分布の話をしました。病気の起こり方から身長や体重に至るまで統計的に考える時は、一例の一つの値で「ここだ！」と決めつけるのではなく、試行や観察を繰り返す中で「中心はこのあたり、全体はだいたいこの範囲に分布する」との捉え方をします。

分布としては、代表的な三つ、二項分布、ポアソン分布、正規分布についてお話ししました。中でも最もよく使うのは正規分布です。

さて、世界には様々な出来事・事象があり、それらの分布を全て数式で表わすと、実は代表的な三つでは足りず、多くの数式・分布が必要になります。どのような分布があるか、その全てを示したのが最後の図です。分布はたくさんありますが、心配しないでください。5 回目までの授業で実際に用

いるのは、正規分布だけです。6回目以降の授業では正規分布の他にカイ二乗分布、t分布、F分布の名前が出て来ます。これらは事象に対応する確率分布ではなく、基本的な統計計算で得られた統計量の存在範囲を示す分布です。名前だけ頭に入れておいてください。では今日はこれで終わります。

演習問題

1. 離散量（離散型確率変数）とはどのような量ですか。あなたの生活に関連して、具体例を挙げてください。
2. あなたの生活に関連して、連続量（連続型確率変数）の具体例を挙げてください。
3. 二項分布とは何ですか？あなたの生活に関連する例を挙げてください。
4. ポアソン分布について、あなたの生活に関連する例を挙げてください。
5. 正規分布について、あなたの生活に関連する例を挙げてください。
6. 二項分布など確率分布曲線を、実際にコイン投げなどを行って描くのは大変です。しかしネットを介し、コンピューターで電子的に曲線を描くのは難しくありません。以下は二項分布を描くサイトです。米国のアイオワ大学が運営しています。説明は英語です。チャレンジしてみてください。
<https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html>
7. 毎日同じような平和な生活が続くと、私たちは世界の事象が偶然性・確率に支配されているなんて、あまり考えません。しかし新型コロナウイルス COVID-19 の流行で明日が見通せない、今の時代には、確率的に考えることは大切です。今日の講義への感想、質問など、何でも構いませんので、100文字以内で書いてください。