具体的事象を取り入れた確率・統計の教材構築について

On a construction of teaching materials in probability and statistics introduced from concrete events

中　村　宗　敬*
NAKAMURA Munetaka

要約：具体的事象として
・ワールドカップやリーグ戦でのサッカーの試合の得点分布
・クラス内での誕生日一致ペア出現数
を取り上げ，確率論におけるさいころ投げとの関連を探る教材の構築について論じる。無関係に見えるこれらの分布が，ほぼ似通ったものになることを趣旨とした教材である。高校・大学での学習段階に応じた教材を提案する。それらの実践例として，筆者が行った２つの高校での出前講義の記録を付し，聴講者の反応を簡単に報告する。

キーワード：さいころ投げ，二項分布，ワールドカップの得点分布，誕生日一致ペアの分布，ポアソン分布

1　はじめに

本稿では具体的事象に基づきそれらを取り入れた確率・統計の教材構築を提案する。「具体的事象」は統計学一般に必然的に結びつくことではあるが，以下ではその具体的事象の例として，
・ワールドカップやリーグ戦でのサッカーの試合の得点分布
・クラス内での誕生日一致ペア出現数
を取り上げる。ワールドカップのような関心の高いイベントからデータを取得し分析をして，それらがまずどのような特性を持っているのかを探る。この部分はいわゆる記述統計の部分である。ここを含めて，統計データ取得 → 統計分析（記述統計）→ 確率モデル設定 → 統計的検証（推測統計），と進むのであるが，どこまで可能かは学習の段階による。

これらは確率モデルの学習にも良い影響があると考える。統計データからどのような現実を説明する確率モデルをどのように想定するか，すなわち仮説発見の論理（詳細は[3]を参照のこと）が重要なのであるが，ある種の飛躍的思考が必要になる。具体的事象の関心と合わせて，思考の面白さへの興味がそそられることだろう（これに類する報告は[8]を参照のこと）。

第２章に，長くなるが，実践例として筆者が行った２つの高校での出前講義の講義録を付しておく。これは次の第３章でも引用するためである。第３章では，提案教材の内容を第２章を引用しつつ詳しく見る。サッカーの得点分布，誕生日の（度数）分布がともにそれらの代表値に基づく二項分布，あるいはその極限としてのポアソン分布により説明できることを見る。

2　出前講義録

筆者は２つのA 高校，B 高校－ともに山梨県内の学校である－で出前講義をする機会を得た。実施日付を記しておくとA 高校：2018年７月25日（水），B 高校：2018年８月29日（水）である。時間

* 科学文化教育講座

は両校とも普段に合わせた１時間分の50分間であった。

内容的には，A 高校の方は誕生日一致ペア数の分布とさいころ投げ，B 高校の方はサッカーの得点分布とさいころ投げについて論じたものである。２つの高校ともに題目は「確率の目で見る身のまわりのできごと」としたが,「サッカーの得点とさいころ投げ」という副題を付した。先に行った A 高校では上記の一般的な題目のみにしたが，B 高校での講義に際してあらためて見直して，内容を最初から明確にした方がよいと判断したためである。

50分という時間は以下の内容と比して時間的にはかなり短い。数学的な厳密性を犠牲にしても，それらの類似性を提示して興味関心を惹かせるよう心がけた。

後の引用のため若干長くなるが，講義録を載せておく。以下の講義録内の節，小節名には本稿自身のそれらと区別するため A，B を冒頭に付しておく。

２．１　Ａ高校出前講義録

Ａ確率の目で見る身のまわりのできごと

Ａ１　はじめに

数学的知識・技能は学習を通して各発達段階で新たにされ，それに伴って理解力も深化されていく。一方で，現実世界はさまざまな偶然に起こる不思議なことであふれている。なぜこんなことが起こるのだろう，と感じるのは誰でも経験があるだろう。単に偶然に起こったことだとかたづけてしまえば簡単かもしれない。このような確率現象の場合，それまでに培ってきた知識を適用するのには困難が伴う。しかし，よくよく考え調べてみると，偶然がほとんど必然であることがわかることも実は少なからずある。すなわち，上述のすでに獲得した知識を有機的に組み合わせて説明できるということである。ただし，ここでいう「有機的」の語は，単に機械的な適用にとどまらない，一見見当はずれのような多方面からのアプローチをうまく結びつけることを意味している。それだけにやはり難しい面があることは確かではある。困難な面があるが，上に記したような不思議な思いののちに納得できるような話題提供をこの機会にしたいと考えている。話を具体的にするために，現実のデータに基づいて次の２つのことを調べてみたい。

１．誕生日の一致

２．サッカーの得点分布

確率的に見てみると，実はこれらが非常によく似た現象であるのだが，皆さんはそれが想像できるだろうか。

Ａ２　誕生日の一致

Ａ２．１　実験

次の実験をしてみよう。結果は果たしてどうなるだろうか。

実験１．この講座の参加者の一部44人（前後）の中に誕生日が同じ人たちがはたしているだろうか。

さて，結果についてこれを書いている時点では確定的なことは言えるべくもないが，実際に実験（調査）をしてみればおそらく「いる」という結果が出ている可能性が非常に高いだろう。比較対象として，昨年度に私が担当した授業出席者の一部41人（教室の座席１列）の誕生日を調べた結果を記しておく：

表２．１　41人の誕生日調査の結果

8/19	12/8	3/11	7/14	11/20	1/23	12/16	6/25	1/27	6/11
4/14	5/13	4/29	8/6	12/10	6/13	4/18	7/3	9/11	3/16
3/23	5/24	4/15	6/12	3/12	2/1	1/4	11/25	8/26	10/13
3/1	4/2	11/12	2/15	7/31	10/3	10/4	4/30	10/31	3/23
6/20									

よくよく注意して調べてみると一致ペアが現れているが，どれとどれかが一致しているかおわかりだろうか。さて，それでは次の結果はどうなるだろう。

実験２. もっと大きな講座の参加者100人（前後）の中に誕生日が同じ人たちはどれくらいいるのだろうか。

前述の授業出席者全体 106 人の調査結果は次の通りであるが，事前予想は３～４組が多かった：

表２．２　106 人の誕生日調査の結果：表２．１に65人分のデータを追加した
（２重線は座席列の区切りにすぎない）

8/19	12/8	3/11	7/14	11/20	1/23	12/16	6/25	1/27	6/11
4/14	5/13	4/29	8/6	12/10	6/13	4/18	7/3	9/11	3/16
3/23	5/24	4/15	6/12	3/12	2/1	1/4	11/25	8/26	10/13
3/1	4/2	11/12	2/15	7/31	10/3	10/4	4/30	10/31	3/23
6/20									
10/17	7/25	11/22	11/25	4/12	11/11	2/17	1/22	4/13	11/27
6/11	5/20	12/30	11/10	6/29	1/5	7/29	8/18	7/7	9/16
10/27	12/11	2/28	4/12	4/10	1/8	6/5	3/11	10/13	4/30
12/14	2/17	5/19	8/19	12/24	7/16	9/25	7/27	12/24	8/31
12/7	9/23	8/21	9/25	7/19	1/9	9/8	5/2	7/13	8/21
10/1	6/8	11/27	6/22	3/11	10/25	1/27	10/21	10/22	12/3
7/27	12/2	10/22	11/21	10/12					

前より骨折りの作業になるが，それは別として，多くの人は調べているうちに信じられないような気持ちになるだろう。なぜこのようなことが起こるのだろう。率直に言ってしまえば，学生を含む我々の予想，考えが間違っていたといってもよい。それは何に起因するのだろう。

ところで，皆さんは前の表からどのように一致を探し出していただろうか。原理的には以下のように分析できるのではないだろうか。

- 最初の日付を見て，「ふーん。」=⇒ 一致比較０回
- ２番目と最初と見比べて，「違う。」=⇒ 一致比較１回
- ３番目と最初，２番目とをそれぞれ見比べて，「同じ人はいない。」=⇒ 一致比較２回
- ４番目と最初，２番目，３番目とをそれぞれ見比べて，「同じ人はいない。」=⇒ 一致比較３回

…

- 41番目と，最初，２番目, . . . , 40番目とをそれぞれ見比べて「. . . . . . 。」=⇒ 一致比較40回

要するに41人中のペア（二人組）をすべて取り上げて誕生日を比較していることになるが，その41人中のペア総数は

$$1 + 2 + 3 + \cdots + 38 + 39 + 40 = 820(\text{組})$$

になる。

一方で各ペアごとに見ると二人の誕生日一致の確率は

$$\frac{\text{二人が同じ誕生日になる総数}}{\text{二人の誕生日の割り振り方の総数}} = \frac{365}{365^2}\ \frac{1}{365}$$

と非常に小さいのだが，これだけ多くの比較機会があるとわかれば一致ペアが生じるのも不思議ではないだろう。ただし，簡単のため1年を365日として誕生日の確率はどの日も等確率 $\frac{1}{365}$ としているが，今後もこの設定の下で考える。

以上を踏まえると，106人の場合の原理的比較回数はそれぞれ単純なたし算の結果，

$$1+2+3+\cdots+103+104+105=5565$$

となる。1回の比較での一致確率は $\frac{1}{365}$ であったが，この膨大な回数だと私の授業での狐につままれたような結果も不思議ではないような気もする。次節でもう少し掘り下げて考えてみよう。

ところで上のたし算は単純な，しかし繰り返し回数が人間には苦痛になるほど多い。もっと簡単に答えを出せないだろうか。41人中のペア数は ${}_{30}\mathrm{C}_2$ という記号で表すが，実は ${}_{41}\mathrm{C}_2=\dfrac{41\times(41-1)}{2}$ $=820$ と簡単に計算される。一般に n 人中のペア数は ${}_n\mathrm{C}_2=\dfrac{n(n-1)}{2}$である。もちろん，${}_{71}\mathrm{C}_2=$ $\dfrac{71\times(71-1)}{2}=2485$，${}_{106}\mathrm{C}_2=\dfrac{106\times(106-1)}{2}=5565$ と上の計算結果と一致する。

Ａ２．２　コイン投げからの類推

もう少しくわしく考えてみよう。一般に n 人では，簡単のために前節の最後の部分を引用すると

$$\begin{cases} Ⓐ\ \text{各ペアについて誕生日一致の確率は } \dfrac{1}{365} \text{ である} \\ Ⓑ\ \text{ペア数，すなわち比較回数は全部で } \dfrac{n(n-1)}{2} \text{ である} \end{cases}$$

ここでコイン投げと比較して考えてみよう。10000回のコイン投げに言い換えると次のようになる。

$$\begin{cases} Ⓐ'\ \text{各回で 1 の目が出る確率は } \frac{1}{2} \text{ である} \\ Ⓑ'\ \text{投げる機会は全部で 10000 回ある} \end{cases}$$

10000回のコイン投げ投げで1の目は何回くらい出るだろう。確率論で平均あるいは期待値と呼ばれるものであるが、厳密な数学的計算によらなくても

$$\frac{1}{2}\times 10000 = 5000\ (\text{回})$$

くらいとはすぐに察しがつく。100回投げたら15回，500回投げたら250回，1000回投げたら500回くらい出るはずであるとは経験的になんとなく把握している。

このコイン投げの見積もりをⒶ，Ⓑにあてはめると，平均的に期待される一致ペア数は

$$\frac{1}{365}\times\frac{n(n-1)}{2}=\frac{n(n-1)}{2\times 365}$$

程度となる。先に述べた私の授業での例をこれにあてはめてみると，

$$41\text{ 人平均 (期待) 一致ペア数：}\frac{41\times(41-1)}{2\times 365}=2.24$$

となり，41人クラスでは平均すれば2.24組の一致ペアが現れるということなので，最初の結果はむしろ少ないとも見えるが，たまたま1組くらい少ないこともありうるだろう。後の106人はどうだろうか。計算してみると，

$$106\text{ 人平均 (期待) 一致ペア数：}\frac{106\times(106-1)}{2\times 365}=15.25.$$

一致ペア15組(+一致トリオ１組)をうまく言い当てている。表２．２の上つの区切り71人ではどうだろう。一致ペアをこれもまた辛抱強く数えてみると，７組いることがわかる。これについてもうまく説明できている。

Ａ３　一致ペア数の確率分布

Ａ３．１　疑問

期待される平均に近い結果が出てくれればよいが，残念なことにさる高校の出前講義のクラス35人中には誕生日一致ペアはいなかった。理論平均はこの現れなかったというのが珍しい方なのかどうか。一方で，私の授業クラスで最後に付け加えた表２．２の下側35人中には一致ペアがなんと５組いる。これも果たして珍事であろうか。

これに答えるには，単に期待値ではなく，

一致ペア０組の確率，一致ペア１組の確率，一致ペア２組の確率，...

のような，（ペア数の） 値とその確率の組にした確率分布がどうなるかがわかればよい。これがわかれば上記２例の珍しさ，あるいは平凡さも判明する。はたしてこれはどうなるか。

Ａ３．２　コイン投げの確率分布

確率分布を理解するために，前節でたとえとして引用したコイン投げを再び考えてみよう。１回あたり表が出る確率 $\frac{1}{6}$ の偏りのあるコインを８回投げることにしよう。さいころを８回投げて１の目が出るかどうかを見るのと同じである。８回投げを１セットと呼ぶことにし，これを1000セット繰り返すとどうなるだろう。下図のうちヒストグラムは，コンピューターシミュレーションの結果，８回中 x 回表が出たセット数の相対度数である。

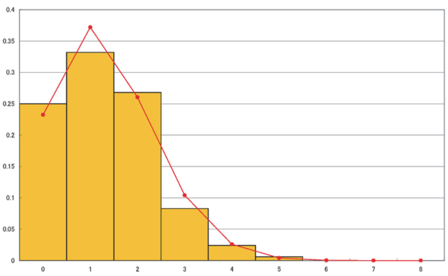

図３．１　シミュレーションによる x 回表のセット数の相対度数 (ヒストグラム) と理論確率 $B(x)$(折れ線)

折れ線でプロットしたのは x 回表が出る確率 $B(x)$ である。これは正確に計算できて（高校１年次に習う），

$$B(x) = {}_8\mathrm{C}_x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{n-x} \qquad (x = 0, 1, 2, \ldots, n)$$

となる。${}_n\mathrm{C}_x$ は n 回のうちどの x 回が表の回かという出方の総数を表している。例えば，

$$B(0) = (\frac{5}{6})^8 = 0.233, \quad B(1) = 8 \cdot \frac{1}{6} \cdot (\frac{5}{6})^7 = 0.372, \quad B(2) = 28 \cdot (\frac{1}{6})^2 \cdot (\frac{5}{6})^6 = 0.260, \quad \ldots\ldots$$

実際に起こった頻度（相対度数） が理論確率分布 $B(x)$ の折れ線にほぼ一致していることがわかる。この点に抽象的な確率計算の意味や価値があり，例えば８回中表が出るのはほぼ２回以下ということがわかる。

Ａ３．３　誕生日一致のシミュレーション

さて誕生日であるが，ここもシミュレーションを行ってみよう。比較対象のために31人のクラスと上述の35人のクラスで誕生日をランダムに発生させ各1000クラス調べ，一致ペア数の相対度数ヒストグラムである。

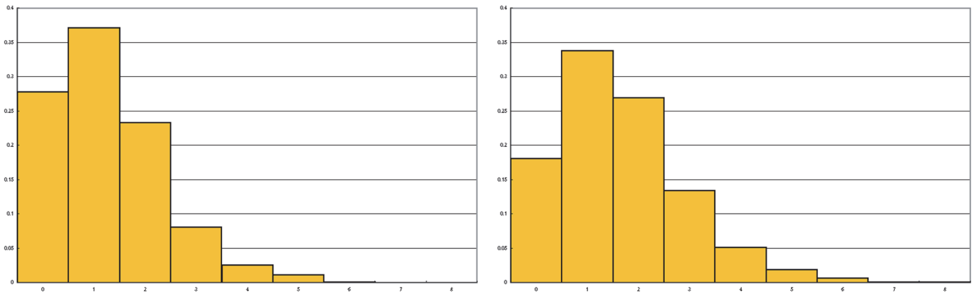

図３．２　左から 31 人 /35 人，各 1000 クラスの誕生日一致ペア数の相対度数のヒストグラム

このシミュレーションでは，35人で０ペアの相対度数＝0.181，５ペア以上の相対度数 ＝ 0.027なのであるが，これらのヒストグラムはそれぞれの理論的な確率分布にほぼ等しいはずである。したがって（確率もこれに近く，）前者は平凡な出来事で，後者はまれにしか起こらないことがわかる。これで一応冒頭の疑問には答えたことになる。

さて，根拠のない印象にすぎないが，この２つのグラフ（特に左側の31のそれ）はなんとなくコイン投げのそれ（図３．１）を想起させる。試しに

31人：確率 $\frac{1}{365}$ のコインを $_{31}\mathrm{C}_2 = 465$ 回投げる（平均 $= \frac{1}{365} \times 351 = 1.274$），

35人：確率 $\frac{1}{365}$ のコインを $_{35}\mathrm{C}_2 = 600$ 回投げる（平均 $= \frac{1}{365} \times 600 = 1.630$）

として，それぞれ表の回数の確率分布を上のヒストグラムに重ねてみるとどうなるだろう。

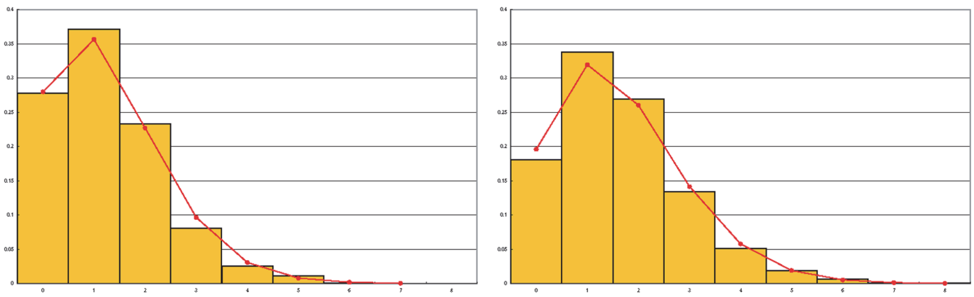

図３．３　図３．２にそれぞれに相当するコイン投げの確率分布（折れ線）を重ねた

結果は見ての通りほぼ一致する。確率分布が等しいということは何を意味するか。第２節で見たような単にその期待値が一致するという類似性だけではなく，誕生日の一致ペア数比較はあたかもコイン投げそのものと同じに見ることができるのである。突拍子もないことを言っているように聞こえるが，結果はそれを示唆しているのである。何か理由があるのだろうか。

Ａ３．４　誕生日一致比較とコイン投げとの同一視

そもそも「コイン投げ」とはとは何であったかを思い起こしてみよう。すなわち，表の確率 p のコイン投げを簡単に図式化すると次のようになる。

$\boxed{\text{コイン投げ}} \Longleftarrow$

$$\text{繰り返し}\begin{cases}\text{① 1 回投げるとき表の出る確率が } p \text{ と一定である (確率一定条件)}\\ \text{② 各回の投げの結果は他の回の結果に影響しない (独立性条件)}\end{cases}$$

図３．４　コイン投げの原理

①，②が数学モデルを構築する決定的な要素となっている。これを流用して次の図式が成り立たないだろか。

$\boxed{\text{誕生日一致とコイン投げの同一視}} \Longleftarrow$

$$\text{繰り返し}\begin{cases}\text{①}' \text{ 1 回の比較あたりの一致確率が} \frac{1}{365} \text{と一定である (確率一定条件)}\\ \text{②}' \text{ 各比較は互いに影響・関係がない (独立性条件)}\end{cases}$$

図３．５　誕生日一致比較とコイン投げとの同一視の原理

①′ に相当する確率一定条件については，ペアごとの誕生日一致確率は $\frac{1}{365}$ であるから満たされる。
②′ の独立性条件については厳密には不可に思える。

ある３人 A，B，C についてペア（A, B），（B, C）の誕生日が一致すれば，ペア（A, C）も必然的に一致しなければならないからである。ペアごとに無関係に一致が起こるというわけではない。

しかし，クラスの人数がそれほど多くなければ，このようなことはまず起こらない。したがって，②′ も成り立つと見てよいだろう。きわめて素朴な説明であるが，同じ考え方により，これは数学的に正当化されている。「近似的に成り立つ」という意味で，誕生日一致比較とコイン投げとの同一視できるのである。

Ａ３．５　ワールドカップを振り返る

ワールドカップもフランスの優勝とともについに終わった。リーグ戦ではなく試合数もそれほど多くないが，コイン投げ化があてはまるのだろうか。早速集計して調べてみた。

表３．１

得点 x	0	1	2	3	4	5	6	7	8
チーム数 y	33	45	35	10	2	2	1	0	0
相対度数 z	0.258	0.352	0.273	0.078	0.016	0.016	0.008	0	0
$B(x)$	0.264	0.354	0.235	0.103	0.033	0.009	0.002	0.000	0.000

平均は $0\times\dfrac{33}{128}+1\times\dfrac{45}{128}+\cdots+6\times\dfrac{1}{128}=1.320$ であり，比較すべき理論的な確率分布は

$$B(x)={}_{90}\mathrm{C}_x\left(\frac{1.320}{90}\right)^x\left(1-\frac{1.320}{90}\right)^{90-x},\quad x=0,1,2,\ldots,90\ .$$

（１分あたり1.320/90の得点があり，これを90回繰り返すと見ている）

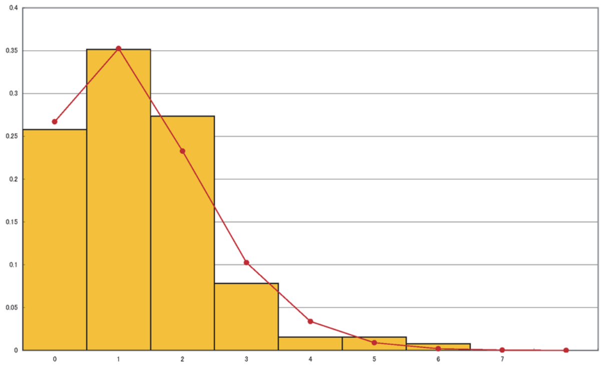

図３．６　2018 FIFA ワールドカップ ロシア の得点分布

２．２　Ｂ高校出前講義録

Ｂ確率の目で見る身のまわりのできごと
～サッカーの得点とさいころ投げ～

概要　現実世界はさまざまな偶然に起こる不思議なことであふれている。しかし，よくよく考え調べてみると，偶然が実はほとんど必然であることも少なくない。今回は皆さんにこのようなお話をしたいと考えている。具体的に現実のデータに基づいて次の２つのことを調べてみたい。

- さいころ投げ
- サッカーの得点分布

確率的に見てみるとこれらが非常によく似た現象であるのだが，皆さんはそれが想像できるだろうか。

Ａ１　奇妙な一致

Ａ１．１　さいころを投げる

まずはさいころ投げから考えてみよう。各目が出る確率が $\frac{1}{6}$ の標準的なさいころである。さて，このさいころを３回投げるとき，１の目（以下では簡単のため $\boxed{1}$ と略記する）が出る回数が０の確率，１の確率，２の確率，３の確率はそれぞれどうなるだろう。実はこれは次の樹形図から計算することができる。

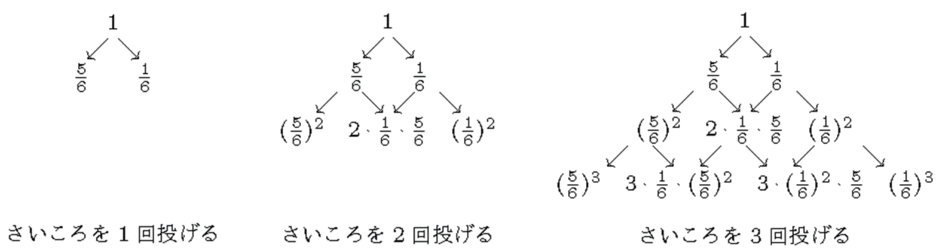

図１．１　$\boxed{1}$ の回数の確率を計算する樹形図：０回目の（確率）１から始めて，↙は $\boxed{1}$ 以外が出たことを表し $\frac{5}{6}$ をかける。↘は $\boxed{1}$ が出たことを表し $\frac{1}{6}$ をかける。さらに上段から来る↘と↙の和を記入していく。

図１．１の１番右を見て表にまとめると次のようになる。

表１．１　さいころを３回投げたときの $\boxed{1}$ の回数（上段）とその確率（下段）

回数	0	1	2	3
確率（式）	$(\frac{5}{6})^3$	$3\cdot\frac{1}{6}\cdot(\frac{5}{6})^2$	$3\cdot(\frac{1}{6})^2\cdot\frac{5}{6}$	$(\frac{1}{6})^3$

注１．数学Aで２，３年生の人は反復試行の確率を習ったと思う。これを思い出して，図なしで上記確率を計算してもよい。たとえば，③は

$$ {}_3\mathrm{C}_2\cdot(\frac{1}{6})^2\cdot\frac{5}{6}=\frac{3\cdot 2}{2\cdot 1}\cdot(\frac{1}{6})^2\cdot\frac{5}{6}=3\cdot(\frac{1}{6})^2\cdot\frac{5}{6} $$

である。さらに述べると，x 回 $\boxed{1}$ が出るの確率は，

$$ {}_3\mathrm{C}_x\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{3-x},\quad x=0,1,2,3\ . \qquad \cdots\cdots Ⓐ $$

である。樹形図といっても実は図１．１は少し細工を施したもので，二項係数に関係するパスカルの三角形というアルゴリズムの確率版と見ることもできる。今後も Ⓐ に類する式を何回か出すが，その意味がわからなくても，とにかく図なしで計算できると思ってもらえばよい。なお，分数はもっと簡単な形に整理することができるが，それが目的ではないのでここではこの形にとどめておく。

Ａ１．２　さいころを８回投げる

それでは，さいころを８回投げたとき，$\boxed{1}$ のそれぞれの回数の確率はどうなるだろう。３回投げのときのように樹形図を辛抱強く書いて求める（あるいは注１のように反復試行の確率にあてはめる）と，次のような結果になる。なお，８回投げのときの①の回数が x である確率 $B(x)$ は，一般的に次のように書ける：

$$ B(x)={}_8\mathrm{C}_x\left(\frac{1}{6}\right)\ \left(\frac{5}{6}\right)\quad,\quad x=0,1,2,\ldots,8\ . \qquad \cdots\cdots Ⓑ $$

表１．２さいころを８回投げたときの $\boxed{1}$ の回数（上段）とその確率（下段）

回数	0	1	2	3	4
確率（式）	$(\frac{5}{6})^8$	$8\cdot\frac{1}{6}\cdot(\frac{5}{6})^7$	$28\cdot(\frac{1}{6})^2(\frac{5}{6})^6$	$56\cdot(\frac{1}{6})^3(\frac{5}{6})^5$	$70\cdot(\frac{1}{6})^4(\frac{5}{6})^4$
回数	5	6	7	8	
確率（式）	$56\cdot(\frac{1}{6})^5(\frac{5}{6})^3$	$28\cdot(\frac{1}{6})^6(\frac{5}{6})^2$	$8\cdot(\frac{1}{6})^7\cdot\frac{5}{6}$	$(\frac{1}{6})^8$	

苦労して計算した割には，表を見てもいまいちピンと来ないので数値計算して小数表示にしてみる。

表１．２′　さいころを８回投げたときの $\boxed{1}$ の回数の確率（数値）

回数	0	1	2	3	4	5	6	7	8
確率（数値）	0.233	0.372	0.260	0.104	0.026	0.004	0.000	0.000	0.000

これらからは日常的な間隔で納得できる結果が読み取れるのだが，それは皆さん自身で考えてもらうことにして先を急ごう。

Ａ１．３　ワールドカップを振り返る

さて，話は突然変わって，別表を参照して先頃のサッカー・ワールドカップの結果を振り返ってみよう。別表の中で，０点のチームは何チームだろうか？１点のチームは？ ... と計64試合，延べ128チームの結果を順次数えてまとめると下の表のようになる。

表１．３　FIFA WORLD CUP 2018の得点ごとのチーム数とその相対度数

得点数	0	1	2	3	4	5	6	7	8
チーム数	33	45	35	10	2	2	1	0	0
相対度数	0.258	0.352	0.273	0.078	0.016	0.016	0.008	0	0

表１．２と表１．３を見比べてみて何か感じたり，思いつくことはないだろうか。表１．２の「確率（数値）」と表１．３の「相対度数」が似ているように見えるが，単なる気のせいだろうか。並べて比較してみよう。

表１．４　表１．２「確率（数値）」と表１．３「相対度数」の比較

回数／得点数	0	1	2	3	4	5	6	7	8
確率（数値）	0.233	0.372	0.260	0.104	0.026	0.004	0.000	0.000	0.000
相対度数	0.258	0.352	0.273	0.078	0.016	0.016	0.008	0	0

さらに，グラフを重ねて可視化してみる。さいころ投げ８回の $\boxed{1}$ の確率が折れ線，ヒストグラムがサッカーの各得点の相対度数である。

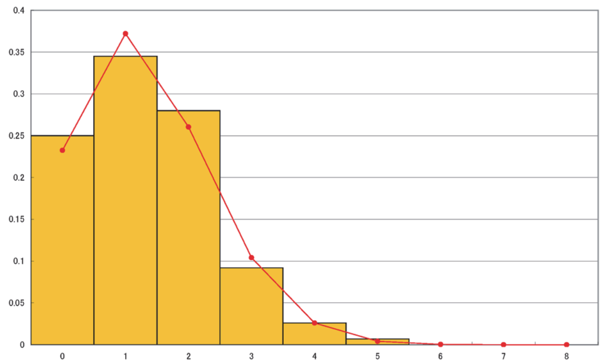

図１．２　表１．４の「確率（数値）」（折れ線）と「相対度数」（ヒストグラム）の重ね合せ

目で判断してもやはり似ている。奇妙なことだが，何かあるのだろうか ... と考える前に，準備として得点の平均値を計算しておこう：

$$
\begin{aligned}
（１試合１チームの得点数の）平均値 &= \frac{総得点}{延べチーム数} \quad （\leftarrow これが定義式）\\
&= \frac{1}{128}\,(0\cdot 33+1\cdot 45+2\cdot 35+3\cdot 10+4\cdot 2+5\cdot 2+6\cdot 1)\\
&= \frac{169}{128} = 1.320
\end{aligned}
$$

Ａ２　サッカーはさいころ投げか

Ａ２．１　突拍子もない仮説

前節で見た思いがけない一致は，突拍子もないことを言ってしまえば，

　サッカーは確率論的にはさいころ投げと変わらない

ということではないだろうか。そう信じて，サッカーとさいころ投げの類似性を探して，次のような仮説（モデル）を作ってみる。

[さいころ投げ仮説]

(1)　サッカーの１試合90分のうち，１チームあたり８回の攻撃機会がある。

(2)　異なる機会の攻撃はほとんど無関係である。一方で，さいころも同じく各回の結果は互いに関係がない。

(3)　平均的に考えて，１機会で得点できる確率を $\frac{1}{6}$ とする。別の表現をすると，６機会をあわせると平均的に１点とれるとする。この意味は，１回の攻撃機会で得点できるか否かは，さいころを１回投げ１の目が出るか否かと同一視できると見なせる，ということである。

考察：なぜ $\frac{1}{6}$ か

大会での攻撃機会総数は 8×128 (回) であるから，攻撃機会１回あたりの得点確率を p として，まずどのような計算をするのかという意味を確認すると，

$$p = \frac{\text{総得点}}{\text{攻撃機会総数}} = \frac{\text{総得点}}{(1\text{試合の}1\text{チームの})\text{攻撃機会数} \times \text{延べチーム数}}$$

$$= \frac{\frac{\text{総得点}}{\text{延べチーム数}}}{\text{攻撃機会数}} = \frac{\text{平均得点}}{\text{攻撃機会数}} \quad (\longleftarrow \text{これは後でも使う})$$

とすればよい。平均得点 $= 1.320$ (前節最後に計算した)，攻撃機会数 $= 8$ により具体的に数値計算すると，

$$p = \frac{1.32}{8} = 0.165 \fallingdotseq \frac{1}{6} .$$

(4)　上述の (1)，(2)，(3) から１試合の得点は，さいころ８回投げの $\boxed{1}$ の回数と同じである。

1		2		3		4	
	5		6		7		8

図２．１　さいころ投げ仮説：１～８の攻撃機会それぞれで，確率 $\frac{1}{6}$ で１点のみ得点でき，８機会の合計が１試合の得点である。これはさいころを８回投げて１が何回出るかを見るのと同じである。番号のついた箇所は攻撃機会，空白箇所は守備機会を表す

Ａ２．２　もっと細かく

しかし，そもそも１試合中の攻撃機会が８回というのはいくらなんでも少なすぎる。そこで次のような修正仮説を立ててみる。

注２．さいころ投げ仮説には次の問題点も思い浮かぶ。すなわち，強いチーム，弱いチームが混在する中で，１攻撃機回あたりの得点確率を一定としてよいのだろうか。実はこれは難しくなるのでここでは不問にしておく。

[特異さいころ投げ仮説]

(1)　サッカーの試合時間90分のうち，90回の攻撃機会がある。

(2)　異なる機会の攻撃はほとんど無関係である。一方でさいころも同じく各回の結果は互いに関係がない。

(3)　平均的に考えて，１区間で得点できる確率 p を $p=\dfrac{\text{平均得点}}{\text{攻撃機会数}}=\dfrac{1.320}{90}=0.0147$ とする([さいころ投げ仮説](3)の「考察」に出てくる“１攻撃機会あたりの得点確率 $p=\dfrac{\text{平均得点}}{\text{攻撃機会数}}$”を思い出そう)。すなわち，１回の攻撃機会で得点できるか否かは，$\boxed{1}$ の出る確率が0.0147と非常に特異なさいころを１回投げ，$\boxed{1}$ が出るか否かと同一視できると見なす。

(4)　上述の(1)，(2)，(3)から１試合の得点は，この特異さいころ90回投げの $\boxed{1}$ の回数と同じである。

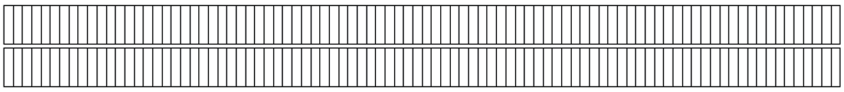

図２．２　特異さいころ投げ仮説：１攻撃機会それぞれで，確率 $p=\dfrac{1.320}{90}=\dfrac{\text{平均得点}}{\text{攻撃回数}}$ で１点のみ得点でき，90回の機会の合計が１試合の得点である。これは $\boxed{1}$ の確率が p の特異さいころを90回投げて $\boxed{1}$ が何回出るかを見るのと同じである。

このように修正して，特異さいころ90回投げのうち，$\boxed{1}$ の回数が x である確率を

$$B_*(x)={}_{90}\mathrm{C}_x\left(\frac{1.320}{90}\right)^x\left(1-\frac{1.320}{90}\right)^{90-x},\quad x=0,1,2,\ldots,90\ . \qquad \cdots\cdots\text{©}$$

に置き換えて，前のグラフに重ねてみる。ただし，$B_*(x)$ $(x \geqq 9)$ は描いていない（ほぼ０である)。

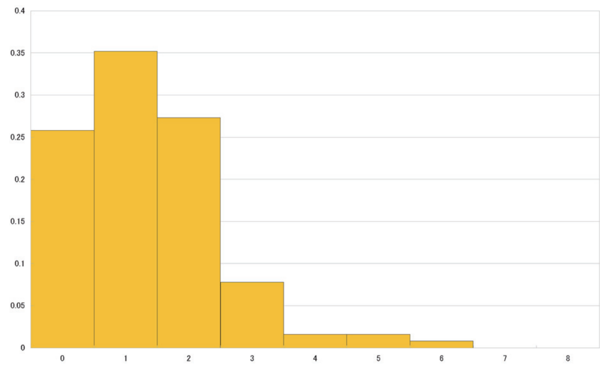

図２．３　図１．２にさらに©の特異さいころ投げ確率 $B_*(x)$ (三角マーカーの太折れ線) を追加した

思惑通り，確かに細い線の通常さいころよりも当てはまりは良くなったように見える。

２．３　Ｊ１リーグでは

粗を探すと，グラフでヒストグラムと折れ線との差異が若干気になる。これはデータ数が128と少ないせいかもしれない。もっと期間が長くデータ数の多いリーグ戦ではどうだろうか。データ数が612の2017年のＪ１リーグのデータを調べてみる。

表２．１　2017Ｊ１リーグの得点ごとのチーム数とその相対度数

得点	0	1	2	3	4	5	6	7	8
チーム数	171	215	136	66	16	5	2	1	0
相対度数	0.279	0.351	0.222	0.108	0.026	0.008	0.002	0.001	0.000

得点の平均値は，

$$\frac{1}{612}(0\times171+1\times215+2\cdot136+3\cdot66+4\cdot16+5\cdot5+6\cdot2+7\cdot1)=1.296$$

となるので，特異さいころ投げ仮説で③の p はここでは $p=\dfrac{1.296}{90}$ であり，相対度数と比較すべき確率 $B_*(x)$ は

$$B_*(x)={}_{90}\mathrm{C}_x\left(\frac{1.296}{90}\right)^x\left(1-\frac{1.296}{90}\right)^{90-x},\quad x=0,1,2,\dots,90\ . \qquad \cdots\cdots Ⓓ$$

となるが，前のように重ねあわせてグラフにしてみる。

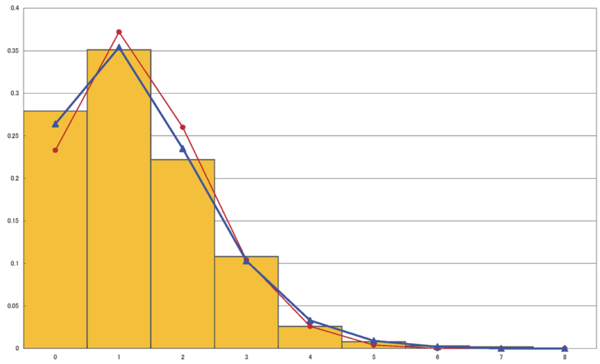

図２．４　2017 Ｊ１リーグの得点分布（相対度数）のヒストグラムに，Ⓑの $B(x)$ (細折れ線）とⒹの $B_*(x)$ (太折れ線）を追加した

特異さいころ投げ仮説が得点状況をよく説明できていることがわかる。なんとまあ，世界は不思議に満ちていることか。

3　さいころ投げを基軸として

3．1　さいころ投げモデル

第２章を見てわかる通り筆者が提案するのは，

・リーグ戦等を通じた一定試合数のサッカーの得点分布，

・一定数の集団内での誕生日の一致ペア数の分布，

のデータを基にしたこれらのさいころ投げへの関連付けである。前記のような現実の分布がさいころ投げで１の目の確率分布 $B(n,p)$，$B_n(x):={}_n\mathrm{C}_x p^x(1-p)^{n-x}$，$x=0,1,2,\dots,n$，$n$ は投げる回数，p は１の目が出る確率（高校では反復試行の確率と呼ばれるが，大学では二項分布という名称が付される）に従うのは多くの学習者にとっては興味が惹かれることと思う。筆者の大学における授業実践でもそのことは明らかにされている（これについては長くなるので別の機会に報告したい）。

このように現実の具体的データと確率とを結びつけることは，推測統計の段階では必然的なことである。しかし，推測統計の基礎として確率論導入の段階ではほとんど考慮されていない。このことは次節以降でまた述べるとして，学習者にとっては何のために確率を学ぶのかという動機づけが難しいところである。確率統計の早い段階から興味・関心を持たせ，現実と向き合う姿勢を持続させることに，このさいころ投げモデルの教材の意義がある。

３．２　高校段階での教材

高校では現行の学習指導要領では，数学Ⅰに「データの分析」が新たに導入され，標準的な１年次段階では従来より続いている数学Ａの「場合の数と確率」とあわせて，確率・統計分野で並び立つ形となった。その後の数学Ｂ「確率分布と統計的な推測」も設置されているが，少数しか選択されていない（平成30年公示学習指導要領では数学Ｂに多数生徒が選択可能な形で設置されている）ので，高校段階では１年次の上記２単元が実質確率・統計に該当する分野といえる。「場合の数と確率」を組み合せ・離散数学として捉える見方もできるが，ここではそれは措く。

しかしながら，教科書の記述を見ると，両者の連携はほとんど考慮されていない。どちらを先行して学習するにしても，筆者が前節講義録で提示したようなデータと確率分布の連携を考えてもよいのではないか。厳密に言えば，確率分布は数学Ｂの学習事項である。数学Ａでは個々の事象の確率を計算するに留めるということである。しかしながら，８回さいころを投げたとき，何回くらい１の目が出るのだろう，とか最も出やすい回数はどれか，と１の目が出る回数のそれぞれの確率を比較するのはごく素朴に疑問に思うことであり，また自然なことである。また，単に並べてグラフ化するだけであるから，数学Ⅰでも学習可能である。さらにいうと，確率分布（的）の見方をもってデータの度数分布表やヒストグラムを見ると，多くの場合，その背後に理想形の数学モデルとしての確率分布が潜んでいるという考え方を養うことができる。

さて，前節で誕生日一致とサッカー得点分布という２つの現実問題を提示したのであるが，高校段階でより適しているのは後者の方であると判断する。クラスの人の誕生日を書き出して同じ日を見つけるという作業を通じて問題を発見するというのは魅力的ではあるが，以下の困難点が見出される：

①　誕生日の一致が起きるかどうか，また一致ペアが何組生じるかはその場の一回限りのことで，他の集団の状況を想定しにくい。

②　各ペアでの誕生日一致出現が「ほぼ独立」ということが理解しにくい。

一方で，サッカーの方は上の問題点に即して述べると，

③　多数のサッカーの試合結果のデータを得ることができる。またテレビ等で実際にいくつもの試合を見ることができる。

④　攻撃が終わったら再度１からやり直す。すなわち，各攻撃機会はほぼ独立である。

ということになる。加えていうと，誕生日一致があったからといってその現実的意味付けは難しいが，サッカーの結果自体は多くの若い学習者にとっては興味があり，ひいきのチームがあ ればより関心を持てる。実は前節でＡ高の内容をＢ高で変更したのは，上述の反省による。

以上の観点から高校段階ではサッカーの方を提案したい。１時間内で完結するのは難しいだろうが，事前調査を取り入れたり，授業外学習と連携させれば可能であろう。場合によっては，課題学習として取り組むのケースもありうる。いずれにしても，生徒の主体的な学習を促す学習支援計画が必要である

３．３　大学段階での教材

既に記したように，筆者の授業実践例とともに稿をあらためて詳述したいので，前小節で問題として書いた２点の克服についてと若干の補足を述べるに留める。

まず，順序が前後するが，まず②の「各ペアでの誕生日一致出現がほぼ独立か」については n 人の集団内のメンバーを１，２，．．．，n と番号付けして，

$$X_{i,j} = \begin{cases} 1 & (i \text{ と } j \text{ の誕生日が一致}) \\ 0 & (i \text{ と } j \text{ の誕生日が不一致}) \end{cases}$$

のように確率変数をつくり，$X_{i,j}, X_{k,l}$ の相関係数を計算すれば二項間の相関について知ることができる。これと「ほぼ独立」とは差異があるが，概略を知るにはこれで十分であろう。これについては [1] を参照されたい。

次に②の一回性の克服についてであるが，これはコンピューターによるシミュレーションにより可能であろう。というのも，この段階では多くの学習者にとって，コンピューターは高校時代よりアクセスが容易になっているからである。30 人クラス，70 人クラス，100 人クラスそれぞれの 10000 回の繰り返しシミュレーション例を記しておく。これらと対応する二項分布あるいは後述するポアソン分布と比較すればよい。シミュレーションでは，簡単のため 1 年を 365 日とし，どの日に生まれる確率も $\frac{1}{365}$ である標準的なモデルを用いている。30 人，70 双方とも一致トリオが生じているクラスが出現しているが，一致トリオは一致ペア数にある方法で換算している。ここではその換算方法は本筋から逸れるのでここでは触れずにおく。

下図はいずれも横軸が一致ペア数，縦軸が該当する一致ペア数を生じたクラスの 10000 クラス中の相対度数である。

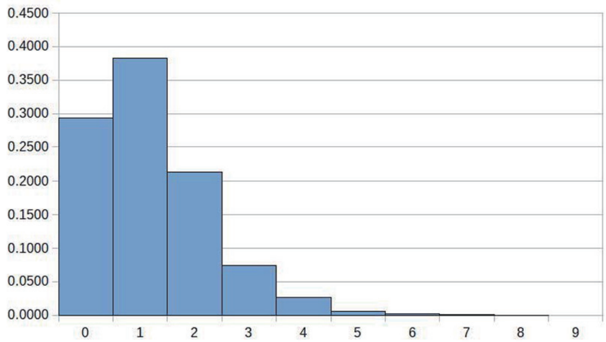

図４．１　30 人 (10000 クラス) の結果。
平均は 1.1914

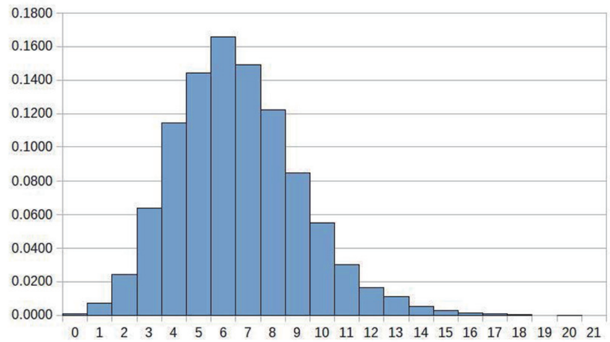

図４．２　相対度数：70 人 (1000 クラス) の結果。
平均は 6.5799

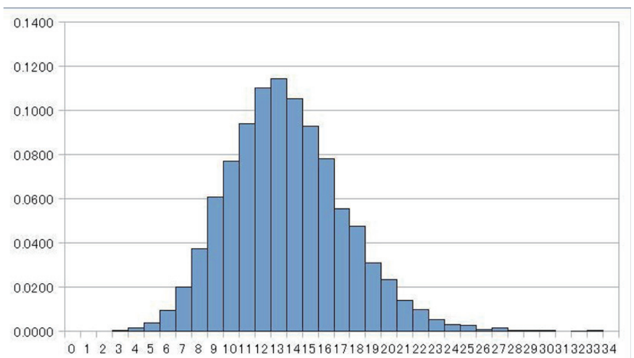

図４．３　100 人 (10000 クラス) の結果平均は 13.5126

補足として次の 2 点を期待することとして指摘しておく。まず，二項分布 $B(n,p)$ の平均（期待値）が np であることは，大学の統計で標準的に学習するので，ヒストグラムの峰の部分の意味付けができるだろう。

次に，$np \approx \lambda$ ならば

$$B_n(x) \approx e^{-\lambda}\frac{\lambda^x}{x!}$$

というポアソンの極限定理も，標準的に知ることになる。二項分布を 1 パラメーターでより簡便な

ポアソン分布で置き換えることの利便性が実感できるとともに，それにあわせてポアソン分布と希少事象に関する小数の法則の意味付けを理解が期待できる。

これらは既に筆者自身の授業実践で確認していることであるが，さらに詳しい調査を今後の課題としたい。

参考文献

［１］ 鈴木武，確率入門ーモデルで学ぶ，培風館，1997.
［２］ 森口繁一，応用数学夜話，筑摩書房（ちくま学芸文庫），2011.
［３］ 米盛裕二，アブダクション仮説と発見の論理，勁草書房，2007.
［４］ フェラー，W.，確率論とその応用，紀伊國屋書店，1960.
［５］ ブロム，G.，ホルスト，L.，サンデル，D.，確率論へようこそ，シュプリンガー・フェアラーク東京，2012.
［６］ Maher, M.J., Modellinfg association football scores, Statistica Neerlandica 36(1982), nr.3
［７］ Wild,C.J., Pfannkuch,M., Statistical Thinking in Empirical Enquiry, lntrnational Sta- tistical Review (1999). 67, 3, 223-265.
［８］ 中村宗敬，統計学習における学習者のPPDACの実践，山梨大学・教育実践学研究23，2018, pp. 115-122.

以下は本稿で用いたデータの入手サイトである。

［９］ 2018 FIFA WORLD CUP RUSSIA, https://www.fifa.com/worldcup/matches/#groupphase, https://www.fifa.com/worldcup/matches/#knockoutphase.
［10］ 2017 年の明治安田生命Ｊ１リーグ，http://www.jleague.ne.jp/special/match/2017/j1.htm