ドラゴンの偉躯に慰むやどりかな　＜中心極限定理＞

2020 年 10 月下旬　白田由香利

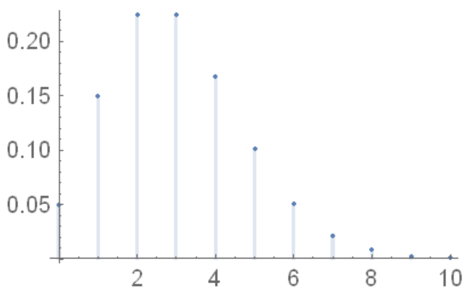
インドネシアの国立動物園等で，コモドドラゴンを見ることができる．大型のトカゲである．インドネシアの人はコモドドラゴンがインドネシアだけに生息することを，とても誇りに感じている．タマン・ミニ・インドネシアという巨大パークで，２m もある偉丈夫のコモドドラゴンが眠っているところを飽きずに眺めていた．普通の日の昼下がりで観客が私しかいない．熱中症にならないように，頻繁に水分を補給する．飼育係の人が私に何か言っている．よく分からないが身振りからして「ここに降りてきて触ってもいいぞ」と言っているらしい．本気か冗談か分からないが，おそらく「好意として触らしてあげる」ということだと思う．もちろん，丁重にお断りした．良い子の皆さんは，お腹がいっぱいの眠っている状態でもコモドドラゴンには絶対に触らないようにしましょう．毒ももっている．

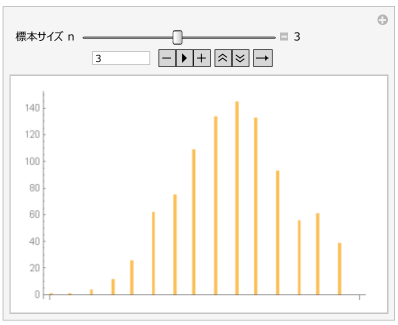
タマン・サファリ国立動物公園のコモドドラゴンも大きい．国家の威信をかけてりっぱな偉躯の個体を選んでいるのだろう．リリー教授の一家と一緒に行ったとき，「イスラムのお祈りをしてくるから 1 時間後にコモドドラゴンのおりの前で集合ね」と言われ，彼らのお祈りのすむのを，コモドドラゴンを見ながら息子と待っていたことがある．見ている分には楽しいが，ジャングルで絶対に遭遇したくない生物だ，と思った．

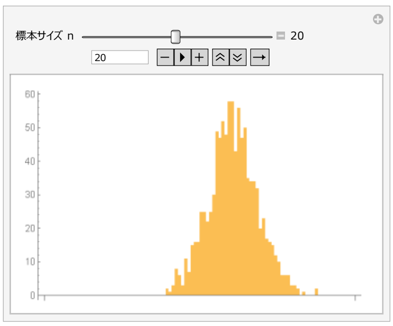
- **中心極限定理**：X_1，X_2，…が平均μ，分散σ^2（有限)の同じ分布に従う独立な変数ならば，標本サイズ $n \to \infty$　に対して標本平均は漸近的に$N(\mu, \frac{\sigma^2}{n})$　に従う．

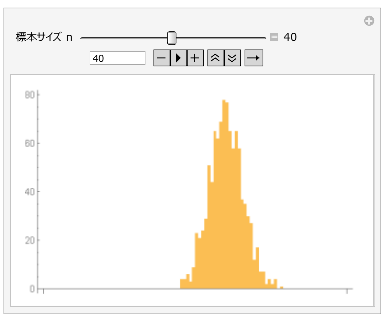
中心極限定理は標本平均の分布に関する定理であった．分布形状の異なる母集団を使って，標本平均の分布が正規分布に近づくようすを見る．

＜＜ポアソン分布＞＞



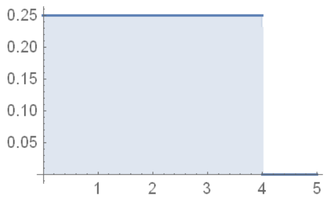


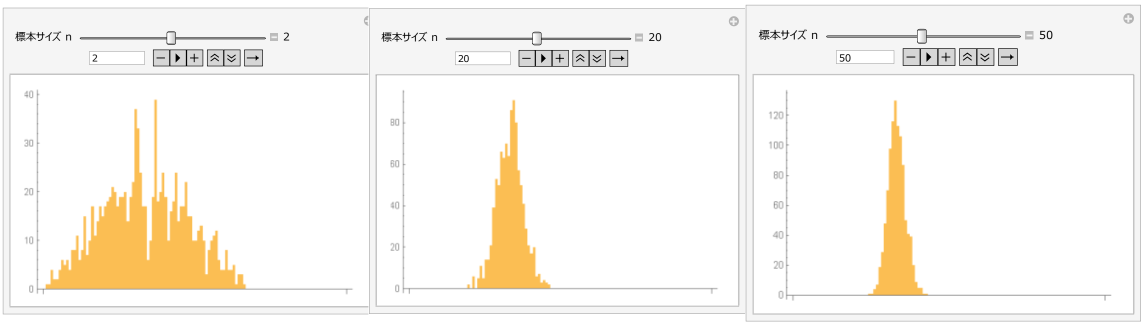




平均３のポワソン分布（上図）と，それを母集団として標本を取ったときの標本平均の分布である．一回の試行で値２と値 3 を取る確率が最も大きくそれぞれ 0.2 を超えている．この母集団から無作為に標本を抽出する．下の列，左から標本サイズが３，20，40 のときの標本平均のヒストグラムである．シミュレーションであるので，毎回分布は若干変わるが，サイズを大きくするに従い，正規分布に近づくことが分かる．そして，その分散も小さくなっていく．

＜一様分布＋ α ＞





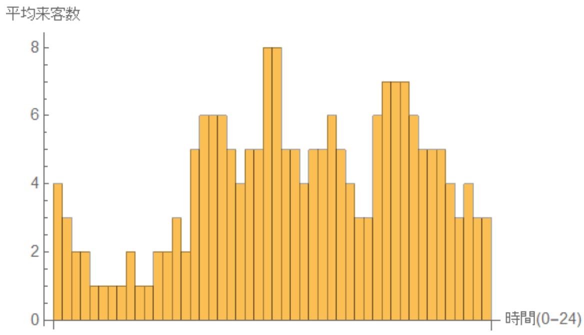
次の母集団として，上図にあるような，領域０から４までの一様分布を考える．一回の試行で０から４までの実数が等確率で得られる．3.6，0.8，2.6．．．というようにである．これを母集団として標本を取ったときの標本平均の分布をみる．左から標本サイズが 2，20，50．標本サイズを大きくすると，正規分布に近づき，その分散は小さくなる．

次に一般化された中心極限定理を見る．一応，定理を説明するが，分からなくても気にしないで，後述のグラフィクスの説明を見てほしい．

- **一般化された(Liapounoff リアプノフの)中心極限定理**：X_i が平均 μ_i ，分散 σ_i^2 の独立な確率変数で，(1) $\rho_i^3 = E(|X_i - \mu_i|^3) < \infty$ (i=1,2,⋯)， (2) $\lim_{n\to\infty}\left(\frac{\rho}{\sigma}\right) = 0$， が成り立つときは，$n \to \infty$ に対して標本平均$\bar{X}$は漸近的に$N(\frac{\mu}{n}, \frac{\sigma^2}{n^2})$ に従う．ただし $\mu = \sum_{i=1}^{n} \mu_i$ ．ただし$\rho^3 = \sum_{i=1}^{n} \rho_i^3$，$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

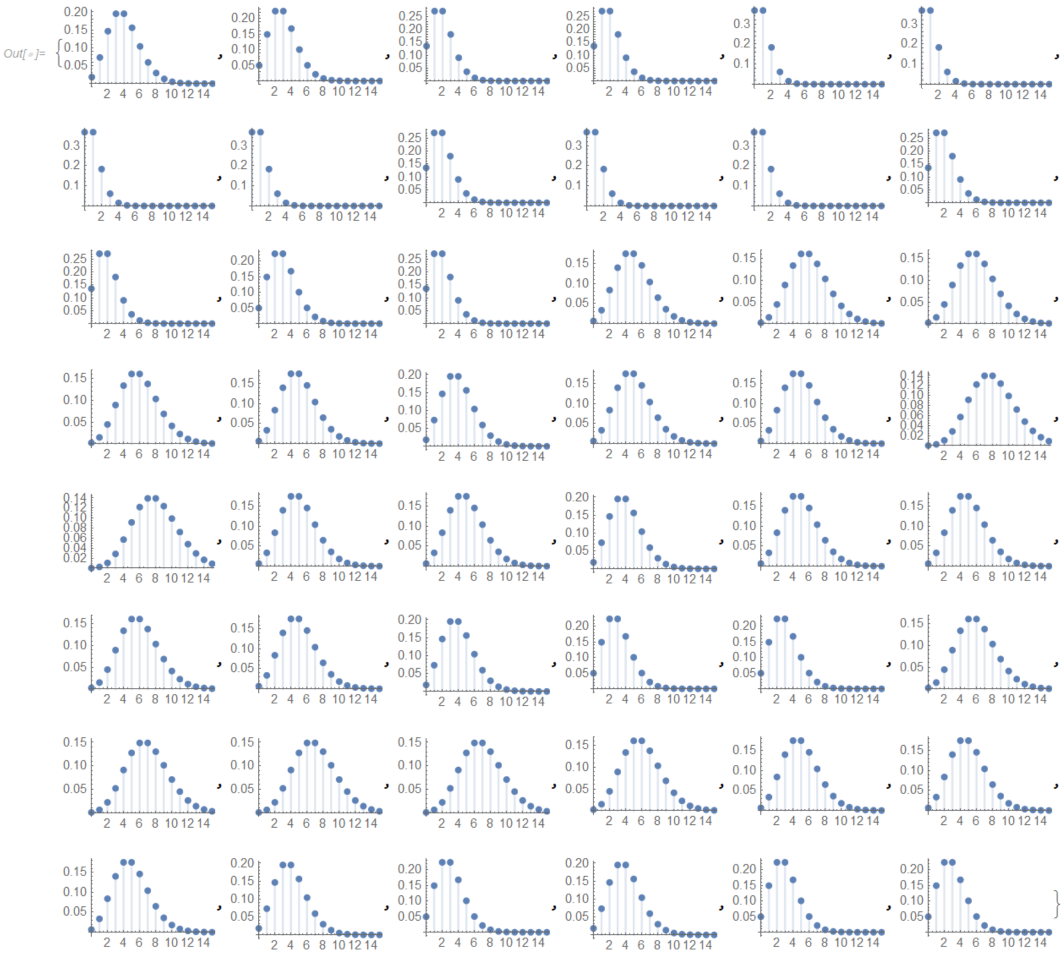
普通の中心極限定理では，標本は同じ母集団から抽出した．一般化中心極限定理では、標本を抽出する際，「母集団の分布は同じ」という制約条件がなくなる。標本１個ごとに，その母集団を変えても，公式に示した２個の条件を満たしていれば，標本平均の分布は，正規分布に近づく，と言っている．母集団の分布の形が違う場合とは、どのように考えればよいのであろうか？

例えば、店への来店者の数を確率変数とすると、時間帯ごとに母集団の分布が変わる。24 時間オープンの店を仮定して、夜中の 0 時から 30 分ごとの平均が

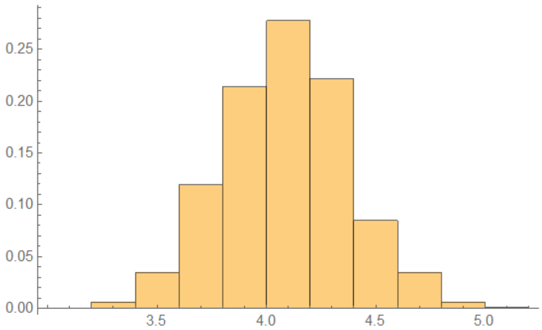


{4,3,2,2,1,1,1,1,2,1,1,2,2,3,2,5,6,6,6,5,4,5,5,8,8,5,5,4,5,5,6,5,4,3,3,6,7,7,7,6,5,5,5,4,3,4,3,4}であったとしよう。

昼時が一番来店者が多く，夜中の丑三つ時は少ない．来店者数の分布はポワソン分布を仮定する。30 分ごとにポワソン分布を仮定し，48 個の確率分布を作った．それらの分布は以下に示すように同じではない．



この 48 個の母集団から順番に 1 つずつ標本を無作為に抽出し，48 個の標本の値を足し合わせれば，ある日の来客数となる．それを 48 で割れば，時間帯によらない 30 分当たりの平均来客数となる．繰り返すが，上記の 48 個の母集団のそれぞれから、標本 1 個を取り出して、合計して割って，その平均を取った値が、30 分あたりの標本平均の値である．シミュレーションによって，この標本平均分布を作った図が以下である．定理の言うように，正規分布になっている．



図；２４時間オープンの店の３０分間の平均来店者数の分布（シミュレーション 5000 回）.

標本平均分布が正規分布に近い形をしている．これが一般化された中心極限定理の意味である．

定理から，この標本平均は正規分布　$N(\frac{\mu}{n},\frac{\sigma^2}{n^2})$に従う．

$N(\frac{\mu}{n},\frac{\sigma^2}{n^2})$の２つのパラメータ値　$\frac{\mu}{n},\frac{\sigma^2}{n^2}$をこの例で計算してみよう．サイズ n は 48 である．ポワソン分布では，理論的に，平均＝分散　であることが分かっている．

μ ＝Total{4,3,2,2,1,1,1,1,2,1,1,2,2,3,2,5,6,6,6,5,4,5,5,8,8,5,5,4,5,5,6,5,4,3,3,6,7,7,7,6,5,5,5,4,3,4,3,4}= 196

よって，$\frac{\mu}{n} = 4.08$，$\frac{\sigma^2}{n^2} = 0.085$ となる．標本平均の分布をシミュレーションした結果の分布の図からその平均と分散が 4.08，0.081 となった．シミュレーションによる分布の平均と分散は，理論の示す値に非常に近く，また，分布の形も正規分布に近いことが分かる．

このような各種のプロセスで異なる確率分布を取ったとき，その標本平均は正規分布に近づくということは，現実的にどのような意味をもつのか．10000 鉢のヒマワリの高さ測定をすると，ほぼ正規分布に従う．ヒマワリの成長に関与する，水の量，肥料の量，日光の量，の３つの要素が，高さに独立に貢献すると仮定する．水の量による高さへの貢献＋肥料の量による高さへの貢献＋日光の量による高さへの貢献で最終的な高さが決定されるとする．相乗効果や相殺効果はどうなのだ，ということは話を単純化するため，今は考えないとする．これは大筋理解のためのラフな議論である．そうすると，３つの要素の高さへの貢献の和が高さとなり，各，確率分布が異なっていても，標本平均の分布は正規分布になる．

世の中にこれだけ多く，正規分布に従う確率変数が存在するということは，このように各種の要因からの貢献が和となって働いているからだと思う．それでも，正規分布となるということは非常に不思議である．毎年，学生から「今までそんな定理を知らずに生きてきました．誰も教えてくれませんでした．世の中がそうなっていたとは驚きです」と感想を頂く．

コモドドラゴンも 10000 匹の体重測定をすれば，正規分布になる．では，コモドドラゴンの体重に寄与する要因とは，何であろう．やはり食事量が圧倒的なのだろうか？　日光によく当たっていると大きくなれる，のかもしれない．この件に関しても情報力不足のため考察不能.

スーパーマーケットの 1 日の売上金額が正規分布になったとする．この売上に関与する要素としては，品物の価格に対する品質の良さ，接客態度の良さ，天候，隣のスーパーの安売り日等のイベント，等々が考えられる．腐りかけの食材を，態度のわるい店員さんが売っているようになったら，売り上げは落ち

る．その日によって，これらの要素の確率分布は異なり，その日の値は異なる．それらの総合された結果としての標本平均の売上が正規分布になる，と解釈できる．個々の要素の売上への貢献の確率分布をどう表現していいか難しいが，考え方としてはご理解頂けると思う．たくさんのプロセス要因の影響の標本平均が正規分布なのである．

終わり

引用元：行く駒の麦に慰むやどりかな　　芭蕉