統計学　Lecture 2

確率変数，確率分布，期待値，分散

北門　利英（海洋生物資源学科）

Attention: 授業に関してわからないことがあれば，メールかリアルタイム接続時（水曜13:00-14:00，質問が続けば14:30まで延長します）に遠慮なく質問してください．

POINT:

- 次の用語と意味を理解すること
  - 確率変数とは，どのような値をとるかが確率的にしか分からない変数
  - 確率分布とは，確率変数の値と確率の関係を規定するメカニズム
  - 期待値とは，確率変数の確率性を考慮した平均値
  - 分散とは，散らばりの尺度で，確率変数の(期待値周りでの)散らばりの期待値
- 簡単な例を通して，上記の定義を確認すること
- 平均と分散という言葉について，観測値の「標本平均と標本分散」と確率変数の「理論平均（期待値）と理論分散」の違いを頭に入れておくこと

T. Kitakado　Statistics　Lecture 2　May 20, 2020

生物資源学科の北門です．このビデオは統計学のLecture2の要約版です．最初にこのビデオを見てから，ハンドアウトをご覧いただき，最後に確認のために再度ビデオをご覧いただくと頭の整理になると思います．今回は，確率変数，確率分布，期待値，分散という言葉を簡単な例を通して理解してください．それでは始めます．

簡単な例１：サイコロ

（設定）サイコロを1回振ったときの出る目の数に興味があるとする

- 確率変数　出る目の数を表す確率変数 Y とおく
- 標本空間　{1, 2, ..., 6}
- 確率分布　$P(Y=y)=\frac{1}{6}\qquad (y=1,2,\dots,6)$

サイコロの目の数	1	2	3	4	5	6
確率	1/6	1/6	1/6	1/6	1/6	1/6

- 期待値

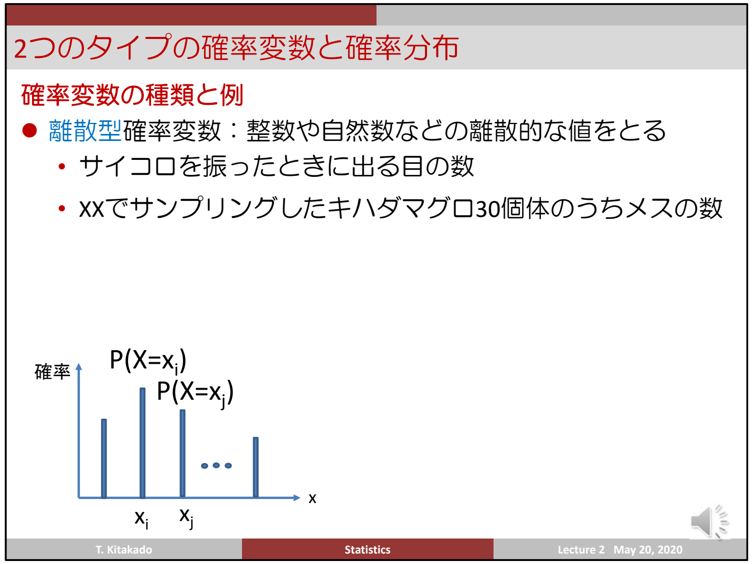
$$E[Y]=\sum_{y=1}^{6} y\cdot P(Y=y)=\sum_{y=1}^{6} y\cdot\frac{1}{6}=\frac{7}{2}$$

- 分散

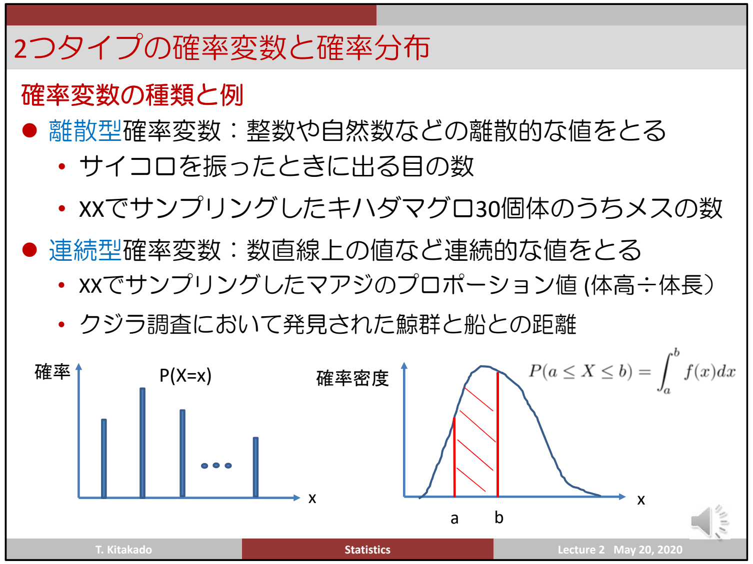
$$V[Y]=\sum_{y=1}^{6}(y-E[Y])^2\cdot P(Y=y)=\sum_{y=1}^{6}\left(y-\frac{7}{2}\right)^2\cdot\frac{1}{6}=\frac{35}{2}$$

T. Kitakado　Statistics　Lecture 2　May 20, 2020

まず確率変数ですが，サイコロを振ったときの出る目は1～6のどの目がでるか振るまで分からないですが，そのような確率的にしかメカニズムが分からない変数を確率変数と言います．また確率変数の確率的メカニズムのことを確率分布と呼びます．このサイコロの例では，出る目の数を確率変数Yで表すと，確率分布はこの式のように表現できます．これがこの場合の確率分布です．また，期待値は，値とそれが生じる確率をすべて足し合わせたもの，また分散は，値が期待値からどの程度離れているかを平均したものと言えます．



Lecture3以降で詳しく述べますが，確率変数には2種類あって，一つは整数や自然数などの離散的な値をとる離散型確率変数です．この図のように，確率変数のすべての値に対して確率を定義できます．これが離散型確率変数とその確率分布の特徴です．

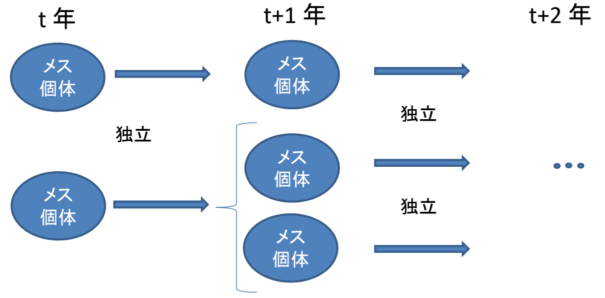


もう一つが連続型確率です．これも後の授業で改めて詳しく説明しますが，数直線上の値をとるため，1点1点には確率を与えられず，この場合には積分の面積によって確率を計算します．今回と続く2回は離散型を中心に話します．

例題：メスの個体数の確率的変化

例題

ある生物種のメスは寿命が1年で，子供を産むとすぐに死亡する．次の世代にメス1個体を産む確率と2個体を産む確率はともに50%と仮定する．このとき，この生物の増加の様子は確率的にしか表現できない．実験開始時のメスの個体数を1とするとき，メスの数は年々どのように変化するであろうか．



サイコロの例は直感的にわかりやすいのですが，少し現実味に欠けますので，生物の個体数の確率的変化を考えてみましょう．毎年メス個体はどの個体も独立に，50%の確率で1個体の雌を，残りの50%の確率で2個体の雌を産み，産んだ後はすぐに死ぬとします．この図の場合ですと，上の個体はたまたま1個体を産み，下の個体はたまたま2個体を産むという例です．したがって時間とともに確率的に個体数は変化します．

1年後のメスの個体数

$Y_1 \in \{1,2\}$

$$\begin{cases} P(Y_1 = 1) = \dfrac{1}{2} \\ P(Y_1 = 2) = \dfrac{1}{2} \end{cases}$$

$$E[Y_1] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}$$

$$V[Y_1] = \left(1 - \frac{3}{2}\right)^2 \cdot \frac{1}{2} + \left(2 - \frac{3}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}$$

その様子を確率変数と確率分布を通して表すと，Y1を1年後のメス個体数を表す確率変数とすると，確率の値ははいずれの場合とも1/2となります．期待値と分散も先ほどのサイコロの時と同じように計算されます．

2年後のメスの数

$Y_0 = 1, Y_1 \in \{1,2\}, Y_2 \in \{1,2,3,4\}$

$$\begin{cases} P(Y_2 = 1 | Y_1 = 1) = 1/2 \\ P(Y_2 = 2 | Y_1 = 1) = 1/2 \\ P(Y_2 = 3 | Y_1 = 1) = 0 \\ P(Y_2 = 4 | Y_1 = 1) = 0 \end{cases} \qquad \begin{cases} P(Y_2 = 1 | Y_1 = 2) = 0 \\ P(Y_2 = 2 | Y_1 = 2) = 1/4 \\ P(Y_2 = 3 | Y_1 = 2) = 2/4 \\ P(Y_2 = 4 | Y_1 = 2) = 1/4 \end{cases}$$

$$P(Y_2 = \mathbf{1}) = P(Y_1 = 1)P(Y_2 = \mathbf{1} | Y_1 = 1) + P(Y_1 = 2)P(Y_2 = \mathbf{1} | Y_1 = 2) = \frac{1}{4}$$

$$P(Y_2 = \mathbf{2}) = P(Y_1 = 1)P(Y_2 = \mathbf{2} | Y_1 = 1) + P(Y_1 = 2)P(Y_2 = \mathbf{2} | Y_1 = 2) = \frac{3}{8}$$

$$P(Y_2 = \mathbf{3}) = \frac{2}{8}$$

$$P(Y_2 = \mathbf{4}) = \frac{1}{8}$$

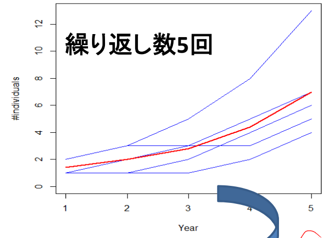
$$E[Y_2] = 1 \cdot \frac{2}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{2}{8} + 4 \cdot \frac{1}{8} = \frac{9}{4} = \left(\frac{3}{2}\right)^2$$

是非3年目も計算してみてください

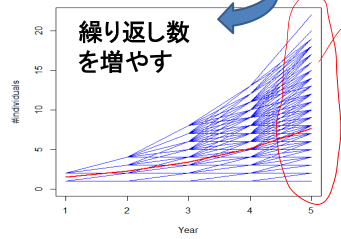
T. Kitakado Statistics Lecture 2 May 20, 2020

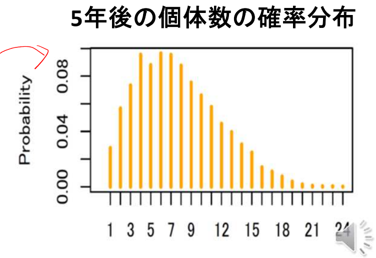
2年目の状態は1年目の状態に依存しますので，このような条件付き確率を考えることになりますが，場合分けすることで，Y2の確率分布が求まります．期待値もこの通りです．3年目以降も同様に計算可能ですが，いずれにしましてもこのように，確率的にしか変化の様子が分からないのですが，その様子を数学的に表現するための道具が確率変数と確率分布となります．

Rのプログラム例



- 2年～3年後くらいまでは簡単に計算できるが，長期にわたるような場合には手計算では大変
- そこで，コンピュータを用いてシミュレーションをすることが多い





実際には，２年から3年後くらいまでは簡単に計算できますが，長期にわたるような場合には手計算では大変なのでコンピュータを用いてシミュレーションをすることが多いです．ハンドアウトにはRのコードの例を書いています．後の授業で説明するチャンスもあるかと思いますので，こんなことができるんだ，という印象だけもって頂ければと思います．

観測値（データ）の平均と分散

観測値 $x_1, x_2, \ldots, x_n$

確率変数 X

標本平均(Sample mean)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

（理論）平均 (theoretical mean)
・期待値(Expectation)

$$E[X] = \begin{cases} \displaystyle\sum_{x=0}^{\infty} x f(x) & \text{(離散型確率分布のとき)} \\ \displaystyle\int_{-\infty}^{\infty} x f(x)dx & \text{(連続型確率分布のとき)} \end{cases}$$

標本(不偏)分散 (Sample variance)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

（理論）分散(theoretical variance)

$$V[X] = \begin{cases} \displaystyle\sum_{x=0}^{\infty}(x - E[X])^2 f(x) & \text{(離散型確率分布のとき)} \\ \displaystyle\int_{-\infty}^{\infty}(x - E[X])^2 f(x)dx & \text{(連続型確率分布のとき)} \end{cases}$$

T. Kitakado | Statistics | Lecture 2 May 20, 2020

今日の授業で，期待値と分散という言葉が出てきました．これはあくまで確率変数の理論的な値のことで，実際のデータを複数個観察した時の標本平均と標本分散とは無関係ではないですが，概念的には別のものですので，区別できるようにしてください．

まとめ

Attention:

- 授業 HP に授業に関する情報をアップデートしていきますので参照ください．(URL: https://toshihidekitakado.github.io/STAT2020/index.html)
- 授業に関してわからないことがあれば，メールかリアルタイム接続時（水曜 13:00-14:00，質問が多ければ 14:30 まで延長します）に遠慮なく質問してください．

Point:

- 次の用語と意味を理解すること
  - 確率変数とは，どのような値をとるかが確率的にしか分からない変数
  - 確率分布とは，確率変数の値と確率の関係を規定するメカニズム
  - 期待値は，確率変数の確率性を考慮した平均値
  - 分散は，散らばりの大きさの尺度で，確率変数の（期待値の周りでの）散らばりの期待値
- 簡単な例を通して，上記の定義を確認すること
- 平均と分散という言葉について，観測値の「標本平均と標本分散」と確率変数の「理論平均（期待値）と理論分散」が異なることを頭に入れておくこと

T. Kitakado | Statistics | Lecture 2 May 20, 2020

Lecture2では確率変数と確率分布から始め，期待値と分散について解説しています．詳しい内容はハンドアウトをご覧ください．水曜日13時から14時までリアルタイムでWebexにて質問を受け付けますので遠慮なく質問ください．これでLecture2のビデオを終わります．