母/標本 平均値と分散

樋口さぶろお

龍谷大学理工学部数理情報学科

計算科学☆演習 II L03(2013-04-24 Wed)

今日の目標

1. 母平均値・母分散・母標準偏差の意味を説明できる
2. 母分布から母平均値・母分散・母標準偏差を計算できる
3. 標本から標本平均値・標本分散・標本標準偏差を計算できる

http://hig3.net

ここまで来たよ

Quiz 解答:期待値

1. $\frac{3}{11} \times (-2)^2 + \frac{1}{11} \times (-1)^2 + \frac{2}{11} \times (0)^2 + \frac{5}{11} \times (+1)^2 = \frac{18}{11}$.
2. $\frac{3}{11} \times (-2) + \frac{1}{11} \times (-1) + \frac{2}{11} \times (0) + \frac{5}{11} \times (+1) = -\frac{2}{11}$.
3. $R = 0$ と $R = 1$ の和事象なので，$\frac{2}{11} + \frac{5}{11} = \frac{7}{11}$.

Quiz 解答:標本期待値

1.

階級	度数	相対度数
1 以上 2 未満	4	0.4
2 以上 3 未満	3	0.3
3 以上 4 未満	1	0.1
4 以上 5 未満	1	0.1
5 以上 6 未満	0	0.0
6 以上 7 未満	1	0.1

2. 略
3. 標本期待値で推定できる.
$\frac{1}{10}(\frac{1}{1}+\frac{1}{6}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{3}+\frac{1}{4}+\frac{1}{2}+\frac{1}{2}+\frac{1}{1})=\frac{75}{12}=\frac{5}{8}$.
4. 標本期待値で推定できる.
$\frac{1}{10}(1+6+1+2+1+3+4+2+2+1)=\frac{23}{10}$.

罪の重い誤り:
この問で, $E(\frac{1}{R}) = \cdots$ とか, $E(R) = \cdots$ とか答えてはいけない. サンプルから, 母期待値 $\mathrm{E}(\cdots)$ を正確に ('=') 求めることは不可能. この問では, 標本期待値を求めろといわれていて, 母期待値 $\mathrm{E}(\cdots)$ は標本期待値で推定される. やや大胆な例えですが, 標本期待値は母期待値の近似値みたいなもの. 近似値について $=$ を使って, $3.1 = \pi = 3$ とか書くのはおかしいでしょ.

ここまで来たよ

1. 母分布と標本分布, 期待値の計算
   - Quiz 解説
2. 母/標本 平均値と分散
   - 復習:母期待値と標本期待値
   - 平均値
   - 分散
   - 標準偏差
   - ランダムウォーカーの座標の場合

復習:母期待値と標本期待値

R:確率変数

$f(R)$ の母期待値

$$確率の表 \rightsquigarrow \mathrm{E}(f(R)) = \sum_{j=1}^{m} p_j \times f(r_j)$$

.

標本期待値

$$サンプル \rightsquigarrow \bar{R} = \frac{1}{N}\sum_{i=1}^{N} f(R^{(i)})$$

N:サンプルサイズ. (i): サンプル内の通し番号. $R^{(i)}$ は値に重複あり.

N が大きくなるほど, 標本期待値は母期待値に ‘近くなる’
つまり, 標本期待値を使って, 母期待値を推定できる.

ここまで来たよ

1. 母分布と標本分布, 期待値の計算
   - Quiz 解説
2. 母/標本 平均値と分散
   - 復習:母期待値と標本期待値
   - 平均値
   - 分散
   - 標準偏差
   - ランダムウォーカーの座標の場合

平均値 mean

母平均値の定義

$$\mu = \mathrm{E}(R) \quad \text{つまり } f(R) = R \text{ の母期待値}$$

母平均値の性質

R: 確率変数, a, b:定数 のとき,

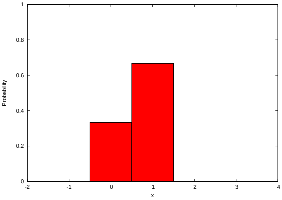
$$\mathrm{E}(aR + b) = \sum_{j=1}^{m} p_j \times (ar_j + b) = a\mathrm{E}(R) + b.$$

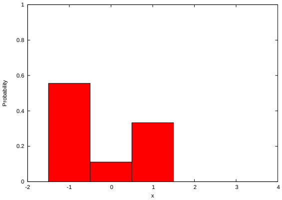
R_1, R_2: (異なる) 確率変数のとき,

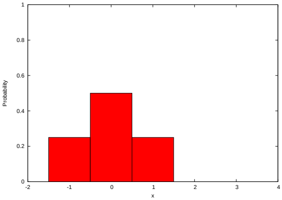
$$\mathrm{E}(R_1 + R_2) = \mathrm{E}(R_1) + \mathrm{E}(R_2).$$

数学 C, 確率統計演習 I

平均値の直観的意味







母平均値をサンプルから推定するには?

サイズ N のサンプル $R^{(1)}, R^{(2)}, \ldots, R^{(N)}$ が与えられたとき,
標本平均値 $\bar{R} = \frac{1}{N}(R^{(1)} + R^{(2)} + \cdots + R^{(N)})$.
Excel では標本平均値は `average`

ここまで来たよ

1. 母分布と標本分布, 期待値の計算
   - Quiz 解説
2. 母/標本 平均値と分散
   - 復習:母期待値と標本期待値
   - 平均値
   - 分散
   - 標準偏差
   - ランダムウォーカーの座標の場合

分散 variance

母分散の定義

$$\sigma^2 = \mathrm{V}(R) = \mathrm{E}((R-\mu)^2) \quad \text{つまり } f(R) = (R-\mu)^2.$$

同じ結果を与える別の計算手順

$$\mathrm{V}(R) = \cdots = \mathrm{E}(R^2) - (\mathrm{E}(R))^2.$$

数学 C, 確率統計演習 I

母分散の性質

R_1, R_2: 独立な確率変数のとき,

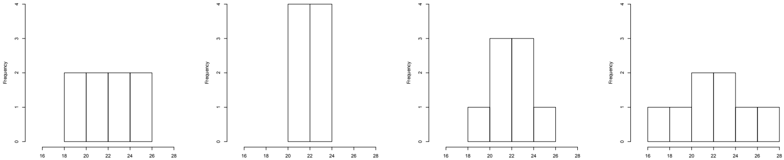
$$\mathrm{V}(R_1 + R_2) = \mathrm{V}(R_1) + \mathrm{V}(R_2).$$

当たり前ではない. ‘確率関係のものはみんな線形’ なわけではない.

数学 C, 確率統計演習 I

分散の直観的意味

平均値は同じでも母分布はいろいろ．すみません縦軸は信じないで．



母分散をサンプルから推定するには?

サイズ N のサンプル $R^{(1)}, R^{(2)}, \ldots, R^{(N)}$ が与えられたとき,

$$
\begin{aligned}
\text{標本分散 } s^2 =& \frac{1}{N}[(R^{(1)}-\mu)^2 + (R^{(2)}-\mu)^2 + \cdots + (R^{(N)}-\mu)^2] \\
& \text{でも } \mu \text{ はわからないので } \bar{R} \text{ で代用したい…} \\
=& \frac{1}{\color{red}{N-1}}[(R^{(1)}-\bar{R})^2 + (R^{(2)}-\bar{R})^2 + \cdots + (R^{(N)}-\bar{R}]^2) \\
=& \frac{N}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N}(R^{(i)})^2 - \bar{R}^2\right].
\end{aligned}
$$

Excel では `var` (`varp` ではない)

s^2 は不偏分散, 標本不偏分散などといわれることもある.

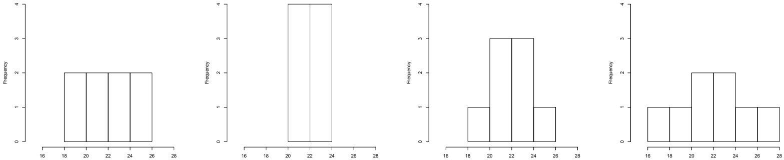
ここまで来たよ

標準偏差 standard deviation

母標準偏差の定義

$$\sigma = \sqrt{\mathrm{V}(R)}$$

標準偏差の直観的意味



母標準偏差をサンプルから推定するには?

標本分散を求めることで母分散を推定する. そして, その平方根をとる.
つまり標本標準偏差 $s = \sqrt{s^2}$.
Excel では `stdev` (`stdevp` ではない)

Quiz(平均値の意味)

次の操作のうち, やって意味あるのはどれ (数学としてじゃなく自然科学として)?

1. ある確率的地震発生理論が正しいか確かめるため, 滋賀県内で 1 か月間に起きた地震のマグニチュードの標本平均値を求めた.
2. ある確率的恒星誕生理論が正しいか確かめるため, 空に見える星の絶対等級の標本平均値を求めた
3. 確率的に成長するペンギンの群れで卵を採集して, 産卵時の卵の質量の標本平均値を求めた.
4. スピードが確率的に変化するマラソンランナーの, 1km 区間ごとの時速を測定して, その標本平均値を求めた.

Quiz(ランダムウォークの到達点の座標の母平均・母分散)

確率変数 R_{t+1} は,

- 確率 $q = 1 - p$ で $R_{t+1} = 0$
- 確率 p で $R_{t+1} = 1$

の値をとる. ($t = 0, 1, 2, \ldots$, $t \neq t'$ のとき R_t と $R_{t'}$ は独立).
時刻 t におけるランダムウォーカーの座標を, 次の漸化式で定める.

$$X_{t+1} = X_t + R_{t+1}, \quad X_0 = 0.$$

1. R_{t+1} の母平均を求めよう.
2. R_{t+1} の母分散を求めよう.
3. R_{t+1} の母標準偏差を求めよう.
4. X_t の母平均を求めよう.
5. X_t の母分散を求めよう.
6. X_t の母標準偏差を求めよう.

サイコロをふる試行で，下の確率分布に従うサンプル生成

R	確率	サイコロの目
0	$q = 1 - p = \frac{1}{3}$	1,2
1	$p = \frac{2}{3}$	3,4,5,6

ベルヌーイ試行

$T = 1, 5, 9$ で.

ここまで来たよ

1. 母分布と標本分布, 期待値の計算
   - Quiz 解説
2. 母/標本 平均値と分散
   - 復習:母期待値と標本期待値
   - 平均値
   - 分散
   - 標準偏差
   - ランダムウォーカーの座標の場合

ランダムウォーク

$$X_{t+1} = X_t + R_{t+1}, \quad X_0 = 0$$

ということは

$$X_T = 0 + \sum_{t=1}^{T} R_t.$$

ここで, R_t $(t = 1, 2, \ldots, T)$ は独立同分布, $\mathrm{E}(R_t) = \mu$,$\mathrm{V}(R_t) = \sigma^2$ とする.

X_T の母平均値

$$\mathrm{E}(X_T) = \mathrm{E}\left(\sum_{t=1}^{T} R_t\right) = \sum_{t=1}^{T} \mathrm{E}(R_t) = T \times \mu.$$

直観的解釈:

X_T の母分散 R_t が互いに独立なので

$$\mathrm{V}(X_T) = \mathrm{V}\left(\sum_{t=1}^{T} R_t\right) = \sum_{t=1}^{T} \mathrm{V}(R_t) = T \times \sigma^2.$$

直観的解釈: $\boxed{}$

X_T の母標準偏差

$\sigma = \boxed{}$

Quiz(ランダムウォーカーの到達点の座標の母平均・母分散)

確率変数 R_{t+1} は,

- 確率 $5/9$ で $R_{t+1} = -1$,
- 確率 $1/9$ で $R_{t+1} = 0$,
- 確率 $3/9$ で $R_{t+1} = +1$,

の値をとる. ($t = 0, 1, 2, 3, \ldots$, $t \neq t'$ のとき R_t と $R_{t'}$ は独立).
時刻 t におけるランダムウォーカーの座標 X_t を, 次の漸化式で定める.

$$X_{t+1} = X_t + R_{t+1}, \quad X_0 = 0$$

1. R_{t+1} の母平均を求めよう.
2. R_{t+1} の母分散を求めよう.
3. R_{t+1} の母標準偏差を求めよう.
4. X_{20} の母平均を求めよう.
5. X_{20} の母分散を求めよう.

予習復習問題これからは毎週

- 金 11:05 締切の予習復習問題は RaMMoodle http://el.math.ryukoku.ac.jp/moodle → 計算科学 II(講義) でやってね.
- 水 13:30 締切の予習復習問題は C-learning http://asp.c-learning.jp/s/ でやってね.

プチテスト演習のプチテストやります! 2013-05-10

自宅で演習の課題をやろう Visual Studio には Express Edition という '無料版' があります. 数理情報学科の学生は DreamSpark 経由で Visual Studio 製品を自宅で使えます.

https://www.a.math.ryukoku.ac.jp/dreamspark/