母期待値と母比率の区間推定・中心極限定理

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 L11(2015-06-26 Fri)

最終更新: Time-stamp: "2015-06-26 Fri 16:07 JST hig"

今日の目標

- 母期待値の t 分布を用いた区間推定ができる
- ランダムウォークの座標 $X(t)$ の母分布が中心極限定理から近似的に求められる

http://hig3.net

L11-Q1

Quiz 解答:ラグランジュ表現とオイラー表現

1. 6 羽なのでサイズは 6.
   各要素は, `x[]={1,1,3,3,3,8};` (順序はこうである必要はない)
2. 座標が $x = 0, 1, 2, \ldots, 9$ の計 10 か所なので, サイズは 10.
   各要素は `u[]={0,2,0,3,0,0,0,0,1,0};` (順序はこうである必要がある)

L11-Q6

Quiz 解答:逆関数法

r の累積分布関数は
$F(r) = \int_0^r \frac{3\sqrt{2}}{8}\sqrt{r'}\,\mathrm{d}r' = 2^{-3/2}r^{3/2}$. $(0 \le r < 2)$.
$y = 2^{-3/2}r^{3/2}$ を解くと, $0 \le r$ より, $r = g(y) = 2y^{2/3}$.

ここまで来たよ

1. 略解:ラグランジュ表現とオイラー表現・逆関数法

2. 母期待値と母比率の区間推定・中心極限定理
   - 空間連続のランダムウォーク
   - 母期待値の区間推定
   - 中心極限定理

空間連続のランダムウォーク

$t \in \mathbb{Z}$: 時刻 時間離散

$$X(t+1) = X(t) + R(t+1)$$

これまで 空間離散 (離散座標), 時間離散ランダムウォーク
$R(t)$ 整数値をとる離散型確率変数 $\rightsquigarrow$ $X(t)$ 整数値をとる離散型確率変数
離散型確率変数 $R(t)$ は確率関数 $P(R=k)=p_k$ で指定される.

これから 空間連続 (連続座標), 時間離散ランダムウォーク
$R(t)$ 連続型確率変数 $\rightsquigarrow$ $X(t)$ 連続型確率変数
連続型確率変数 $R(t)$ は確率密度関数 $f_R(r)$ で指定される.

L11-Q1

Quiz(逆変換法による擬似乱数生成)

次の確率密度関数 $p(r)$ に従う連続値擬似乱数 R を, 逆変換法で生成したい.

$$p(r) = \begin{cases} -\frac{3}{2}r + \frac{1}{4} & (-1 \leq r < 0) \\ 0 & (\text{他}) \end{cases}$$

関数 `double getrandom(double y)` を定めて ($g(y)$ を定めて), `getrandom(getuniform())` の返り値 ($r = g(y)$) が $p(r)$ に従うようにしよう. ただし, `double getuniform()` は $[0, 1)$ 一様乱数 Y を返す関数.

ここまで来たよ

1. 略解:ラグランジュ表現とオイラー表現・逆関数法

2. 母期待値と母比率の区間推定・中心極限定理
   - 空間連続のランダムウォーク
   - 母期待値の区間推定
   - 中心極限定理

シミュレーションによる母期待値の (点) 推定

座標の標本 $X^{(1)}(T), \cdots, X^{(N)}(T)$. 標本サイズ (例)$N = 1000$
標本からの推定

- 座標 $X(T)$ の母平均値 $\mathrm{E}[X(T)]$ の推定値

$$\text{標本平均値 average} \quad \overline{X} = \frac{1}{N}[X^{(1)}(T) + \cdots + X^{(N)}(T)]$$

.

- 座標の関数 $Y = \phi(X(T))$ の母期待値 $\mathrm{E}[Y]$ の推定値

$$\text{標本期待値 average} \quad \overline{\phi(X(T))} = \overline{Y} = \frac{1}{N}[Y^{(1)} + \cdots + Y^{(N)}], \quad Y^{(i)} = \phi(X^{(i)}(T))$$

- Y の母分散 $\mathrm{V}[Y]$ の の推定値

$$\text{(不偏) 標本分散 var} \quad S_Y^2 = \frac{1}{N-1}[(Y^{(i)} - \overline{Y})^2 + \cdots (Y^{(N)} - \overline{Y})^2] = \frac{N}{N-1}[\overline{Y^2} - (\overline{Y})^2]$$

- (例)$X(T) > a$ の母比率 p の推定値

$$\text{標本比率}\hat{p} = \frac{1}{N}[0 + 1 + 1 + 0 + \cdots + 0] = \frac{\text{条件をみたすデータの個数}}{\text{サンプルサイズ}}$$

(点) 推定っていうけど, 標本ナントカって, 母ナントカにどのくらい近いの?

(母分散未知の場合の) 区間推定

区間推定

標本平均値 $\overline{Y}$ として m, 不偏標本分散 $\overline{Y}$ として S^2 が得られたとき, 母平均値 $\mu = \mathrm{E}[Y]$ の,
信頼係数 $1-\alpha$ の信頼区間は

$$m - t_{\alpha/2}(N-1) \times \sqrt{S^2/N} < \mu < m + t_{\alpha/2}(N-1) \times \sqrt{S^2/N}$$

異なる シード で $\boxed{}$ とき, μ がこの不等式を満たす (=信頼区間に含まれる) 確率は $1-\alpha$.

標本サイズ N が大きくなると信頼区間は $\boxed{}$

信頼係数 $1-\alpha$ が大きくなると信頼区間は $\boxed{}$

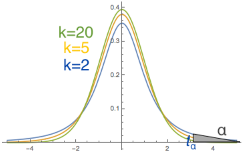
確率統計☆演習 I(2014)L10

確率統計☆演習 II(2015)L12?

t-分布表

$\alpha = P(T > t_\alpha(k))$ となる, $t_\alpha(k)$ の値の表.

$k\backslash\alpha$	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.00025
1	1.000	1.376	1.963	3.078	6.314	12.710	31.820	63.660	127.300	318.300	636.600
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.090	22.330	31.600
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.210	12.920
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
$+\infty$	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291



L11-Q2

Quiz(母期待値の区間推定 (母分散未知))

ランダムウォークの座標の標本を出力するプログラムで, シードそれなり, 標本サイズ $N=5$ で実行したところ, 次のような出力を得た.

```
# それなり
# 5
1,-15
2,-35
2,-35
3,-55
2,-35
```

コンマで区切られた二つの数値は, $X, Y=\phi(X)$.

1. 母平均値 $\mu=\mathrm{E}[Y]$ と母分散 $\mathrm{V}[Y]$ を (点) 推定しよう.
2. 母平均値 $\mu=\mathrm{E}[Y]$ を, 信頼係数 $\alpha=0.95$ で区間推定しよう.
3. 母平均値 $\mu=\mathrm{E}[Y]$ を, 信頼係数 $\alpha=0.99$ で区間推定しよう.

母比率の区間推定

母比率の区間推定

母比率の信頼係数 $1-\alpha$ の信頼区間は, 標本サイズ N が大きいとき,

$$\hat{p} - t_{\alpha/2}(\infty) \times \sqrt{\frac{1}{N}\hat{p}(1-\hat{p})} < p < \hat{p} + t_{\alpha/2}(\infty) \times \sqrt{\frac{1}{N}\hat{p}(1-\hat{p})}$$

注: $N-1=\infty$ としてるのは, 他のところでもすでに近似 $N=\infty$ を使ってるから.
注: 上の計算で比率の下限, 上限が 0,1 を越えるときは, 0,1 に直してしまっていい.

L11-Q3

Quiz(母比率の区間推定)

ランダムウォークの座標の標本を出力するプログラムを実行した

1. 標本サイズ $N = 1000$ で実行したところ, $(X(10))^2 > 20$ を満たすものが 1000 個中 200 個だった. $(X(10))^2 > 20$ の母比率 p を, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう.
2. 標本サイズ $N = 100000$ で実行したところ, $(X(10))^2 > 20$ を満たすものが 100000 個中 20000 個だった. $(X(10))^2 > 20$ の母比率 p を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

ここまで来たよ

1. 略解:ラグランジュ表現とオイラー表現・逆関数法

2. 母期待値と母比率の区間推定・中心極限定理
   - 空間連続のランダムウォーク
   - 母期待値の区間推定
   - 中心極限定理

中心極限定理

$X(0) = 0$ としよう.

$$X(T) = X(0) + \sum_{t=1}^{T} R(t) = \sum_{t=1}^{T} R(t)$$

中心極限定理 (いいかげんバージョン)

$R(1), \ldots, R(T)$ が, 母平均値 μ, 母分散 σ^2 の独立同分布に従うとする.
どんな分布でも可 T が大きいとき,

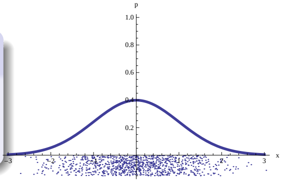
- $X(T) = R(1) + \cdots + R(T)$, の確率分布は,
  の $\boxed{}$ に似る

$T \to +\infty$ では分布の個性がなくなる!

標準正規分布 (ガウス分布) $\mathrm{N}(0, 1^2)$

標準正規分布 $\mathrm{N}(0, 1^2)$

確率密度関数 $f(x) = \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$



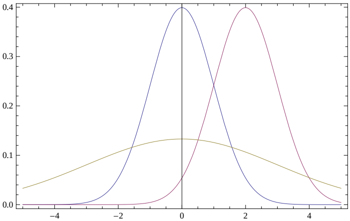
$\mathrm{N}(0, 1^2)$ の性質

連続型確率変数 Z が標準正規分布 $\mathrm{N}(0, 1^2)$ に従うとき

- 母平均値 $\mathrm{E}[Z] = 0.$
- 母分散 $\mathrm{V}[Z] = 1 - 0^2 = 1.$

一般の正規分布 $\mathrm{N}(\mu, \sigma^2)$

確率変数 $X = aZ + b$ を考える ($a > 0, b$ は定数).
確率密度関数は, 横に a 倍に拡大, b 平行移動, 縦に $\frac{1}{a}$ 倍.



$$f(x; b, a^2) = \frac{1}{\sqrt{2\pi a^2}} \mathrm{e}^{-\frac{(x-b)^2}{2a^2}}$$

- 母平均値 $\mu = \mathrm{E}[X] = \mathrm{E}[aZ + b] = \boxed{\qquad\qquad} = b.$
- 母分散 $\sigma^2 = \mathrm{V}[X] = \mathrm{V}[aZ + b] = \boxed{\qquad\qquad} = a^2.$

(一般の) 正規分布 $\mathrm{N}(\mu, \sigma^2)$

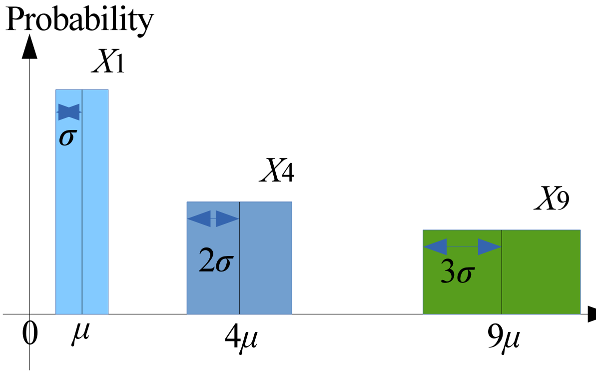
確率密度関数

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

母平均値 μ, 母分散 σ^2.

長方形じゃなくてガウス分布だった!

$X(T)$ の確率密度関数はこんな感じ?



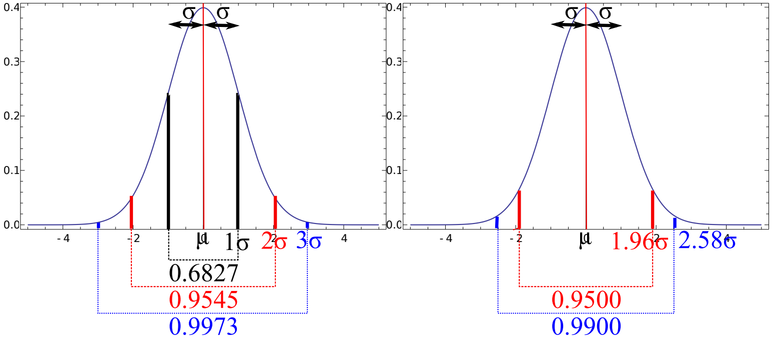
L11-Q4

Quiz(中心極限定理)

$X_t = R_1 + \cdots + R_t$. $R_1, R_2, \ldots$ は連続型確率変数で, 独立同分布に従う. X_t の確率密度関数 $f(x)$ のグラフは, t が増加するとともにどうなる?

1. 幅は広く, 高さは高くなっていく.
2. 幅は広くなっていく. 高さは変わらない.
3. 幅は広く, 高さは低くなっていく.
4. 幅は狭く, 高さは高くなっていく.
5. 幅は狭くなっていく. 高さは変わらない.
6. 幅は狭く, 高さは低くなっていく.

正規分布 (ガウス分布) のグラフに関係した面積



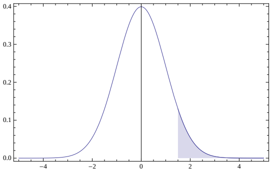
標準正規確率表 (上側確率 $Q(x)$)

$\mathrm{N}(0, 1^2)$ で, $P(X \geq x) = Q(x) = 1 - F(x) = \frac{1}{2}\mathrm{erfc}(x/\sqrt{2})$. F:累積分布関数.

紙と鉛筆では計算できない. 表またはソフトウェアに頼る.

標準正規確率表 (上側確率=$Q(x) = 1 - F(x)$)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010



L11-Q5

Quiz(ランダムウォークと中心極限定理)

$X(t+1) = X(t) + R(t+1), X(0) = 0$ で定まるランダムウォークの座標を考える. ただし, $R(1), R(2), \ldots$ は連続型確率変数で, 母平均値 $\mathrm{E}[R(t)] = -\frac{1}{4}$, 母分散 $\mathrm{V}[R(t)] = \frac{1}{5}$ の独立同分布に従う.

1. $X(20)$ の母平均値と母分散を求めよう.
2. $X(20) > 0$ となる確率を (近似的でよいので紙と鉛筆で) 求めよう.
3. $|X(20)| > 1$ となる確率を (近似的でよいので紙と鉛筆で) 求めよう.

お知らせ

Math ラウンジ=チューター月火水木昼.
スケジュール
2015-07-29 水 3 演習の真夏のプチテスト 35 ピーナッツ
2015-07-31 水 2 講義のファイナルトライアル (外部記憶ペーパー使用) 50 ピーナッツ
e ラーニング予習問題 次は 2015-06-30 火 23:55 締切

manaba 出席カード提出
`https://attend.ryukoku.ac.jp`