正規分布と中心極限定理

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 L12(2013-07-03 Wed)

今日の目標

1. 正規分布の確率密度, 確率を計算できる
2. 中心極限定理を使って, 独立同分布に従う確率変数の和を, 正規分布で近似できる

http://hig3.net

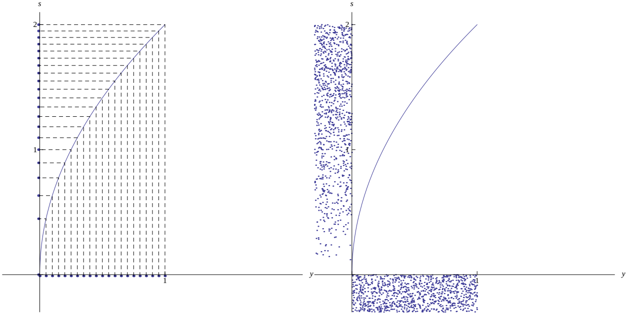
ここまで来たよ

Quiz 解答:逆変換法

r の累積密度関数は

$F(r) = \int_{-\infty}^{r} p(r_1)\,\mathrm{d}r = \int_0^r \frac{1}{2} r_1\,\mathrm{d}r_1 = \frac{1}{4}r^2.\ (0 \le r < 2).$

$y = \frac{1}{4}r^2$ を解くと, $0 \le r < 2$ より, $r = g(y) = 2\sqrt{y}$.



Quiz 解答:逆変換法

r の累積分布関数は

$F(r) = \int_{-\infty}^{r} \frac{3\sqrt{2}}{8}\sqrt{r_1}\,\mathrm{d}r_1 = 2^{-3/2} r^{3/2}.\ (0 \le r < 2).$

$y = 2^{-3/2} r^{3/2}$ を解くと, $0 \le r$ より, $r = g(y) = 2y^{2/3}$.

講評:積分変数

累積分布関数を求める積分で $\int \mathrm{d}r_1$ と書いたのは, 定積分の上限 r と別の文字を使いたかったから. 別の文字例えば z とかでもよかった.
p_2 の 2 とあわせる必要もないし, 項ごとに $r_2, r_3, \cdots$ などとリネームしていく必要もない.

Quiz 解答:逆変換法

1.

$$F(r) = \int_{-\infty}^{r} p(r_1)\,\mathrm{d}r_1 = \begin{cases} \displaystyle\int_{-\infty}^{r} 0\,\mathrm{d}r = 0 & (r < 0) \\ \displaystyle 0 + \int_{0}^{r} 3\,\mathrm{d}r_1 = 3r & (0 \leq r < \frac{1}{4}) \\ \displaystyle 0 + \int_{0}^{\frac{1}{4}} 3\,\mathrm{d}r_1 = \frac{3}{4} & (\frac{1}{4} \leq r < 2) \\ \displaystyle 0 + \int_{0}^{\frac{1}{4}} 3\,\mathrm{d}r_1 + \int_{2}^{r} \frac{1}{8}\,\mathrm{d}r_1 & \\ = \frac{3}{4} + \frac{1}{8}(r-2) & (2 \leq r < 4) \\ 1 & (4 \leq r) \end{cases}$$

2. 区間 $0 \le r < \frac{1}{4}$ で, 値域は $0 \le y < \frac{3}{4}$. $y = 3r$ を解いて, $r = \frac{1}{3}y$. すなわち,
$$F^{-1}(y) = \frac{1}{3}y \quad (0 \le y < \tfrac{3}{4})$$
区間 $2 \le r < 4$ で, 値域は $2 \le y < 4$. $y = \frac{3}{4} + \frac{1}{8}(r-2)$ を解いて, $r = 8(y - \frac{3}{4}) + 2 = 8y - 4$. すなわち,
$$F^{-1}(y) = 8y - 4 \quad (\tfrac{3}{4} \le y < 1)$$

3. $$g(y) = \begin{cases} \frac{1}{3}y & (0 \le y < \frac{3}{4}) \\ 8y - 4 & (\frac{3}{4} \le y < 1) \end{cases}$$

講評:場合分け

$g(y)$ の y は $[0,1)$ 一様乱数だから, g の定義域は $[0,1)$. その外を 'その他' などとして考慮する必要はない.

$$g(y) = \begin{cases} & \\ 0 & (\text{他}) \end{cases}$$

って書いてる人もいたけど, なんで $r=0$ を持ってくるの? $r=0$ が特別だってことないでしょ. これ確率や確率密度じゃないんだから.

ここまで来たよ

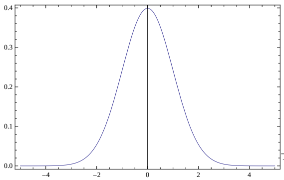
標準正規分布 (ガウス分布)

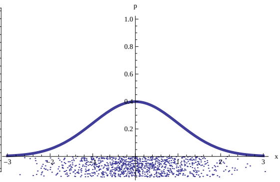
標準正規分布 $\mathrm{N}(0,1)$

$$\text{確率密度関数 } p(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} = \frac{1}{\sqrt{2\pi \cdot 1^2}} \mathrm{e}^{-\frac{(z-0)^2}{2\cdot 1^2}}$$

$$\text{累積分布関数 } \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z'^2}{2}}\, \mathrm{d}z'$$

母平均値 $\mathrm{E}(Z) = 0$, 母分散 $\mathrm{V}(Z) = 1$.

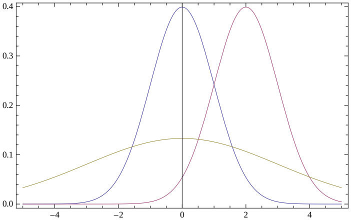




一般の正規分布 (ガウス分布)

Z が標準正規分布 $\mathrm{N}(0,1)$ に従うとき，確率変数 $X = aZ + c$ を考える．$p_X(x)\,\mathrm{d}x = p_Z(z)\,\mathrm{d}z$ より，

$$p_X(x) = \frac{1}{a} \cdot p_Z\left(\frac{x-c}{a}\right) = \frac{1}{a} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-c)^2}{2a^2}}$$



- 母平均値 $\mu = \mathrm{E}(X) = \mathrm{E}(aZ + c) = a\mathrm{E}(Z) + c = c$
- 母分散 $\sigma^2 = \mathrm{V}(X) = \mathrm{V}(aZ + c) = a^2\mathrm{V}(Z) = a^2$

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

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

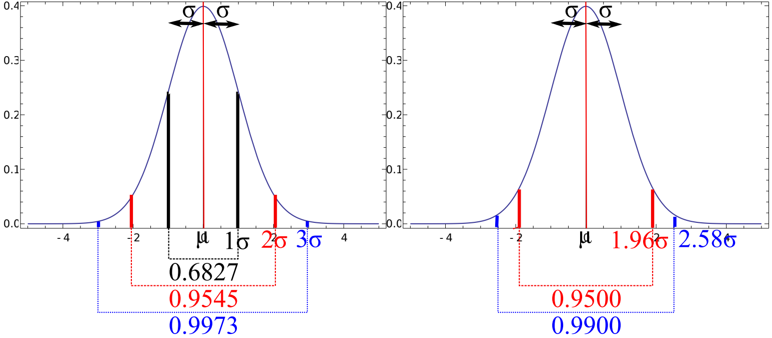
母平均値 μ, 母分散 σ^2.

Quiz(正規分布の確率密度関数の拡大縮小平行移動)

平均値 3, 分散 4 の正規分布のグラフの概形を描こう.

ここまで来たよ

正規分布 (ガウス分布) のグラフに関係した面積



上側確率

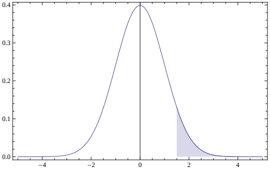
N(0, 1) で, $Z \geq z$ となる確率 $= 1 - \Phi(z)$. Φ:累積分布関数.
紙と鉛筆では計算できない. 表またはソフトウェアに頼る.

$2(1-\Phi(1)) = \boxed{}$, $2(1-\Phi(2)) = \boxed{}$

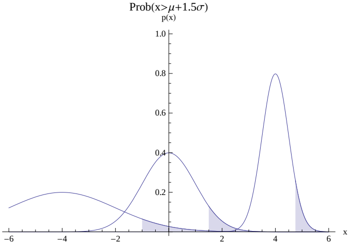
$2(1-\Phi(\boxed{})) = 1 - 0.9500$, $2(1-\Phi(\boxed{})) = 1 - 0.9900$.

標準正規確率表 (上側確率$=1-\Phi(z)$)

z	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



$\mathrm{N}(0,1)$ と $\mathrm{N}(\mu,\sigma^2)$ の面積の求め方ほとんど同じ



$p_X(x)\,\mathrm{d}x = p_Z(z)\,\mathrm{d}z$ なので, どんな μ,σ でも, $\mathrm{N}(\mu,\sigma)$ の, $\mu + a\sigma \le x \le \mu + b\sigma$ 部分の面積は同じ (a, b だけで決まる).

Quiz(正規分布の確率)

平均値 3, 分散 4 の正規分布で,

1. $x \geq 5$ となる確率を求めよう.
2. $+1 \leq x \leq 7$ となる確率を求めよう.

Quiz(正規分布の確率)

1. 平均 0, 分散 1^2 の正規分布で, $0.5 \le x \le 0.7$ となる確率を求めよう.
2. 平均 0, 分散 2^2 の正規分布で, $0.5 \le x \le 0.7$ となる確率を求めよう.
3. 平均 3, 分散 2^2 の正規分布で, $4.0 \le x \le 4.4$ となる確率を求めよう.

ここまで来たよ

1. 逆変換法による乱数生成 (2)
   - Quiz 解説
2. 正規分布と中心極限定理
   - 正規分布
   - 正規分布の確率
   - ランダムウォークの座標の平均値と分散の復習
   - 中心極限定理

ランダムウォークの座標の平均値と分散 (L04 の復習)

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

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.

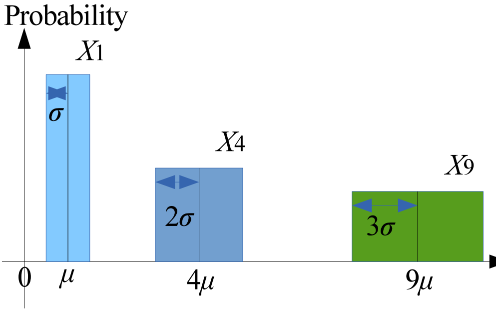
X_T の母平均値

$$\mathrm{E}(X_T) = T \times \mu_R.$$

X_T の母分散 R_t が互いに独立なので

$$\mathrm{V}(X_T) = T \times \sigma_R^2.$$

ってことは, ヒストグラムの時間変化はこんな感じ?



いつでもこんな長方形? 待て中心極限定理

ここまで来たよ

1. 逆変換法による乱数生成 (2)
   - Quiz 解説
2. 正規分布と中心極限定理
   - 正規分布
   - 正規分布の確率
   - ランダムウォークの座標の平均値と分散の復習
   - 中心極限定理

中心極限定理の実験

R: 連続型確率変数.

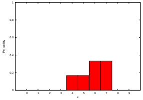
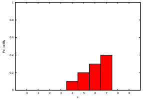
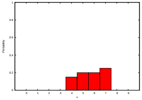
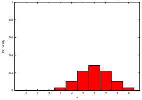
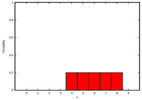
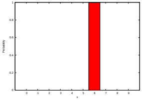
$$p(r) = \begin{cases} \frac{2}{3} & (0 \le r < \frac{1}{2}) \\ \frac{4}{3} & (\frac{1}{2} \le r < 1) \\ 0 & (\text{他}) \end{cases}$$

- $\mu_R = \cdots = \frac{7}{12}$
- $\sigma_R^2 = \cdots = \frac{11}{144}$.

$X_t = R_1 + R_2 + \cdots + R_t$.

$\mathrm{E}(X_9) = \boxed{}$

$\mathrm{V}(X_9) = \boxed{}$

中心極限定理 (いいかげんバージョン)

$R_1, \ldots, R_T$ が, 同じ確率分布に従い, 独立であるとする.
R_t の

- 母平均値は μ
- 母分散は σ^2

だとする (何でもいい. 正規分布でなくていい). これを $R_1, \ldots, R_T$ は$\boxed{}$, 独立同分布に従う, という
このとき, $X_T = R_1 + \cdots + R_T$ の確率分布は, $T \to +\infty$ で

- 母平均値 $\boxed{}$
- 母分散 $\boxed{}$

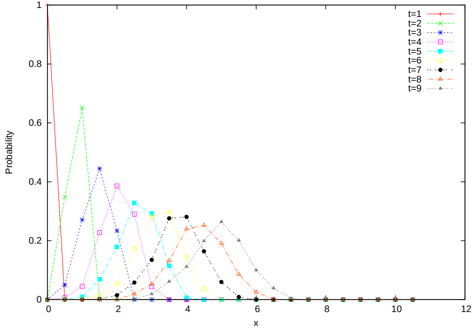
の$\boxed{}$に近づいていく.

ってことは?

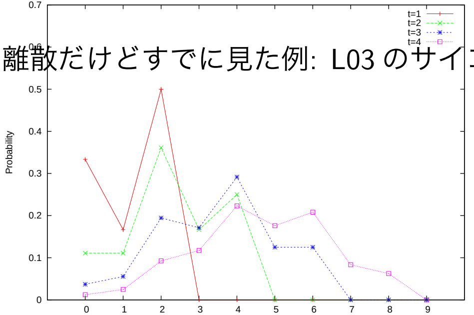
Quiz(中心極限定理)

$X_t = R_1 + \cdots + R_t$. $R_1, R_2, \ldots$ 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 $p_t(x)$ 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.



離散だけどすでに見た例: L03 のサイコロ実験, 課題 rw6, rw7



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, $R_0, R_1, \ldots$ は連続型確率変数で, 母平均値 $\mu = -\frac{1}{4}$, 母分散 $\sigma^2 = \frac{1}{5}$ 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.

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 13:35 締切の予習復習問題は RaMMoodle `http://el.math.ryukoku.ac.jp/moodle` → 計算科学演習 II でやってね.
- RaMMoodle にはスマートフォンからもアクセスできます. `http://hig3.net` > Links > RaMMoodle.

演習の休講/補講/プチテスト計画

- 2013-07-05 金 3(ここは休講する総合演習の裏) に補講. 金 2 はふつう
- 2013-07-12 金 2 ふつう
- 2013-07-19 金 2 演習の夏のプチテスト. 35 ピーナッツ.
- 2013-07-26 金 2 は予備日 (休講の有力候補),
- 2013-07-31 水 3 講義のファイナルトライアル