二項分布と幾何分布

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 II L12(2016-07-14 Thu)

最終更新: Time-stamp: "2016-07-14 Thu 07:34 JST hig"

今日の目標

- 二項分布, 幾何分布のモーメント母関数, 母平均値, 母分散が求められる
- 現象を二項分布, 幾何分布でモデルできる

http://hig3.net

L11-Q1
Quiz 解答:2 次元正規分布

1. 指数関数の引数を平方完成する.

$$
\begin{aligned}
&-x^2-4x-2y^2+12y-5\\
&=-\frac{1}{2}(2x^2+8x+4y^2-24y+10)\\
&=-\frac{1}{2}\left(\frac{(x+2)^2}{(1/\sqrt{2})^2}-8+\frac{(y-3)^2}{(1/2)^2}-36+10\right)
\end{aligned}
$$

より, $\mathrm{E}[X]=-2, \mathrm{E}[Y]=3, \mathrm{V}[X]=1/2, \mathrm{V}[Y]=1/4$. $f(x,y)$ は x 部分と y 部分の積だから X, Y は独立であり, $\mathrm{Cov}[X,Y]=0$.

2. $f(x,y)$ は次の 2 つの形に書ける.

$$C\cdot \mathrm{e}^{-\frac{(x+2)^2}{2(1/\sqrt{2})^2}-\frac{(y-3)^2}{2(1/2)^2}}\mathrm{e}^{-\frac{1}{2}(-34)} = \frac{1}{\sqrt{2\pi(1/2)}}\frac{1}{\sqrt{2\pi(1/4)}}\mathrm{e}^{-\frac{(x+2)^2}{2(1/\sqrt{2})^2}-\frac{(y-3)^2}{2(1/2)^2}}$$

よって, $C = \frac{1}{\sqrt{2\pi(1/2)}}\frac{1}{\sqrt{2\pi(1/4)}}\mathrm{e}^{-17}$.

L11-Q2
Quiz 解答:2 次元正規分布

1. 指数関数の引数を平方完成する.

$$
\begin{aligned}
&-4x^2-\tfrac{1}{6}y^2+2y\\
&=-\frac{1}{2}(8x^2+\tfrac{1}{3}y^2-4y)\\
&=-\frac{1}{2}\left(\frac{(x-0)^2}{(1/2\sqrt{2})^2}+\frac{(y-6)^2}{(\sqrt{3})^2}-12\right)
\end{aligned}
$$

より, $\mathrm{E}[X]=0, \mathrm{E}[Y]=6, \mathrm{V}[X]=\frac{1}{8}, \mathrm{V}[Y]=3, \mathrm{Cov}[X,Y]=0$.

2. $f(x,y)$ は次の 2 つの形に書ける.

$$
C\cdot \mathrm{e}^{-\frac{(x-0)^2}{2(1/2\sqrt{2})^2}-\frac{(y-6)^2}{2(\sqrt{3})^2}}\mathrm{e}^{-\frac{1}{2}(-12)}=\frac{1}{\sqrt{2\pi(1/8)}}\frac{1}{\sqrt{2\pi\cdot 3}}\mathrm{e}^{-\frac{(x-0)^2}{2(1/2\sqrt{2})^2}-\frac{(y-6)^2}{2(\sqrt{3})^2}}
$$

であるべきだから, $C=\frac{1}{\sqrt{2\pi(1/8)}}\frac{1}{\sqrt{2\pi\cdot 3}}\mathrm{e}^{-6}$.

L11-Q3
Quiz 解答:2 次元正規分布

1. 指数関数の引数の 2 次式を対称行列を用いて書く. x, y の 1 次の項はないので, X, Y の母平均値は 0 であり,

$$-x^2 + 4xy - 7y^2 = -\frac{1}{2}\begin{pmatrix} x-0 & y-0 \end{pmatrix}\begin{pmatrix} 2 & -4 \\ -4 & 14 \end{pmatrix}\begin{pmatrix} x-0 \\ y-0 \end{pmatrix}$$

よって,

$$V^{-1} = \begin{pmatrix} 2 & -4 \\ -4 & 14 \end{pmatrix}, \quad V = \frac{1}{2 \cdot 14 - (-4)(-4)}\begin{pmatrix} 14 & 4 \\ 4 & 2 \end{pmatrix}.$$

2. $f(x, y) = C\mathrm{e}^{-x^2+4xy-7y^2} = \frac{1}{2\pi\sqrt{\det V}}\mathrm{e}^{-\frac{1}{2}{}^t\boldsymbol{x}V^{-1}\boldsymbol{x}}$. よって $C = \frac{1}{2\pi\sqrt{1/12}}$.

L11-Q4
Quiz 解答:2 次元正規分布

1. 指数関数の引数の 2 次式を対称行列を用いて書く. x, y の 1 次の項はないので, X, Y の母平均値は 0 であり,

$$-2x^2 - xy - \tfrac{1}{2}y^2 = -\frac{1}{2}\begin{pmatrix} x-0 & y-0 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x-0 \\ y-0 \end{pmatrix}$$

よって,

$$V^{-1} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}, \quad V = \frac{1}{4\cdot 1 - 1\cdot 1}\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}.$$

2. $f(x, y) = C\mathrm{e}^{-2x^2 - xy - \frac{1}{2}y^2} = \frac{1}{2\pi\sqrt{\det V}}\mathrm{e}^{-\frac{1}{2}{}^t\boldsymbol{x}V^{-1}\boldsymbol{x}}$. よって $C = \frac{1}{2\pi\sqrt{1/3}}$.

復習:定義と公式

離散型確率変数のモーメント母関数

$$M_{\mathrm{X}}(t) = \mathrm{E}[\mathrm{e}^{Xt}] = \sum_{x=-\infty}^{+\infty} P(X = x)\mathrm{e}^{xt}$$

確率統計☆演習 II(2016)L05

等比級数の和

$$\sum_{k=0}^{n} ar^k = \frac{a(1 - r^{n+1})}{1 - r}$$

数学 (中学)

ここまで来たよ

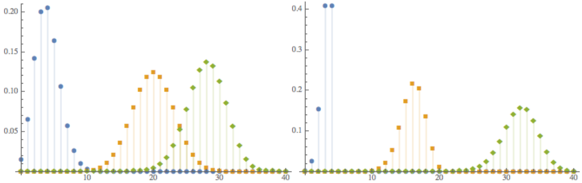
二項分布

二項分布

離散型確率変数 X が次の確率分布を持つとき，X は二項分布 $\mathrm{B}(n,p)$ に従うという．

$$P(X=k)=\begin{cases} {}_n\mathrm{C}_k\, p^k(1-p)^{n-k} & (k=0,1,2,3,\dots,n) \\ 0 & (\text{他}) \end{cases}$$

意味: 確率 p で表の出るコインを n 回投げたとき，k 回表が出る確率．



$\mathrm{B}(40,0.1),\mathrm{B}(40,0.5),\mathrm{B}(40,0.7)$, $\mathrm{B}(4,0.8),\mathrm{B}(20,0.8),\mathrm{B}(40,0.8)$

二項分布のモーメント母関数と母期待値

$$M_X(t) = (pe^t + (1-p))^n$$

$$\mathrm{E}[X] = \boxed{}, \mathrm{V}[X] = \boxed{}$$

二項分布の再生性

分布 $\mathrm{B}(n,p)$ は再生的.
すなわち, $X_1 \sim \mathrm{B}(n_1,p)$, $X_2 \sim \mathrm{B}(n_2,p)$ ならば
$Y = X_1 + X_2 \sim \mathrm{B}(n_1+n_2,p)$.

なぜなら

意味: X_i は n_i 枚のコインのうち表が出た枚数. Y は n_1+n_2 枚のコインのうち表が出た枚数.

ここまで来たよ

3 2 次元正規分布

4 二項分布と幾何分布
- 二項分布
- ベルヌーイ分布
- 幾何分布

ベルヌーイ分布

ベルヌーイ分布

$n=1$ の二項分布 $\mathrm{B}(1,p)$ のこと

$$P(X=k)=\begin{cases} 1-p & (k=0) \\ p & (k=1) \\ 0 & (\text{他}) \end{cases}$$

意味:ベルヌーイ試行=(不公平な) コイン投げ．表がでる確率 p.

ベルヌーイ分布のモーメント母関数と期待値

$$M_X(t)=p\mathrm{e}^t+(1-p)\mathrm{e}^0$$

$$\mathrm{E}[X]=\boxed{}, \qquad \mathrm{V}[X]=\boxed{}$$

ベルヌーイ分布と二項分布の関係

$X_i \sim \mathrm{B}(1,p)$ $(i=1,2,\ldots,n)$ を独立同分布にしたがう確率変数とするとき, $Y = X_1 + \cdots + X_n$ は二項分布 $\mathrm{B}(n,p)$ にしたがう.

数学 B

なぜなら $\boxed{}$

カウントデータ

独立同分布にしたがう確率変数 $X_1, X_2, \ldots, X_n$ と, X に対する条件 A あるとき
集合の位数 $Y = |\{i | X_i$ が A を満たす $\}|$ は整数値確率変数になる.
この確率変数 (の標本) を**カウントデータ**という.

カウントデータとみなせる確率変数

- 二項分布にしたがう確率変数
- 度数分布表の各階級の度数
- 2 元分割表・クロス集計表の各欄の度数
- (ちょっと拡大解釈) ポアソン分布にしたがう確率変数

L12-Q1

Quiz(二項分布)

確率 $p = \frac{2}{3}$ で表のでるいかさまコインがある．100 回投げる．

1. 表が 50 回でる確率を求めよう．
2. 表がでる回数の母平均値を求めよう．
3. 表がでる回数の母分散を求めよう．

L12-Q2

Quiz(二項分布)

ある宝くじは, あたりと残念賞の 2 種類の結果だけがある．あたりの確率は 0.05 である．あたりの賞金は 1050 円, 残念賞の賞金は 50 円である．このくじを 10 回ひいたときの賞金の合計額を確率変数 Y とする．

1. Y と, あたりの回数 X との関係を書こう．X はどのような分布にしたがうか．
2. 確率 $P(Y = 2500)$ を求めよう (小数の積に書けば, それ以上整理しなくてよい).
3. Y の母平均値と母分散を求めよう．

2,3 では過程が必要だが, 二項分布の確率や母平均値や母分散やモーメント母関数をおぼえていれば記してそれを使ってもよい．二項分布と無関係に解いてもよい．モーメント母関数を自分で求めて使ってもよい．

二項分布と正規分布

$X_i \sim \mathrm{B}(1,p)$ のとき, $T_n = X_1 + \cdots + X_n \sim \mathrm{B}(n,p)$
よって, $\mathrm{B}(n,p)$ は $n \to \infty$ で正規分布 $\mathrm{N}(np, np(1-p))$ に「似る」

数学 B

なぜなら

確率統計☆演習 II(2016)L06

L12-Q3

Quiz(二項分布と正規分布と中心極限定理)

表が $\frac{3}{5}$, 裏が $\frac{2}{5}$ でてるコインを, 100 回投げる. 表が 64 回より多くでる確率を, 正規分布を表を用いて近似的に求めよう.

ここまで来たよ

幾何分布 I

Quiz(初めて表, の確率)

ある 1 日に死者 10 名以上の交通事故が起きる確率を $1/10000 = 0.0001$ とする.
今日そのような事故が起きた. 次にそのような交通事故が起きるまでの間隔として確率が一番大きい間隔は?

1. 1 日 (=次の日)
2. 100 日
3. 5000 日
4. 10000 日
5. 20000 日

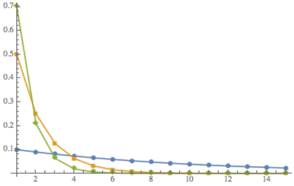
幾何分布の例

幾何分布

離散型確率変数 X が次の確率分布を持つとき, X はパラメタ p の幾何分布に従うという.

$$P(X=k)=\begin{cases} p(1-p)^{k-1} & (k=1,2,3,\dots) \\ 0 & \text{他} \end{cases}$$

意味: ベルヌーイ試行を繰りかえしたとき, k 回目に初めてコインの表を得る確率.
これは統計学の流儀. 確率論では, $p(1-p)^k$ $(k=0,1,2,\dots)$ とひとつずらして定義することが多い. 要確認.



$p=0.1,0.5,0.7$

幾何分布のモーメント母関数と母期待値

$$M_X(t) = \frac{p}{\mathrm{e}^{-t} - (1-p)},$$

$$\mathrm{E}[X] = \boxed{}, \qquad \mathrm{V}[X] = \boxed{}$$

L12-Q4

Quiz(幾何分布)

ある野球の打者は, 100 打席に 1 回の割合でホームランを打つ. 各打席は独立な試行であると考える.

1. あるホームランのあと, x 打席目 $(x = 1, 2, \ldots,)$ に次のホームランが出る確率 $f(x)$ を求めよう (連続ホームランなら「1 打席目」). この X は何分布にしたがうか (シーズン終了とか引退のことは考えず, 無限に打席は続くと考えてよい).
2. 確率分布 $f(x)$ の最大値とそのときの x を求めよう.
3. 確率分布 $f(x)$ のもとで X の母平均値を求めよう.
4. ホームランを打った後, 100 打席目までに次のホームランが出る確率を求めよう.

瀬田龍大生調査プロジェクト

何回かの授業にまたがって，チーム別で，問題 (RQ=Research Question) をたて，調査し，検定して答をだします．
来週のタスク 2016-07-21 木 2 は 1-542 に臨時教室変更
チームでスライドを提出
授業前に可能

1. 教員の整理した Web アンケート 1 に答えておく ← お知らせします
2. 教員の整理した Web アンケート 2 に答えておく
3. まとめの PowerPoint スライドに 1-4 を整理して書く
4. 担当の検定の手順を復習しておく

授業中に

1. Web アンケートの結果をダウンロードする
2. Excel でグラフを描いて PowerPoint スライドに貼る
3. Excel で計算して，**手順に従って検定を行って** PowerPoint スライドに書く
4. PowerPoint スライドに結論を書く

標準正規確率表 (上側確率$=Q(z) = 1 - F(z)$)

$Z \sim \mathrm{N}(0, 1^2)$. $P(Z > z) = 1 - F(z) = Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^{+\infty} \mathrm{e}^{-\frac{z'^2}{2}}\, \mathrm{d}z'$ の表.

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