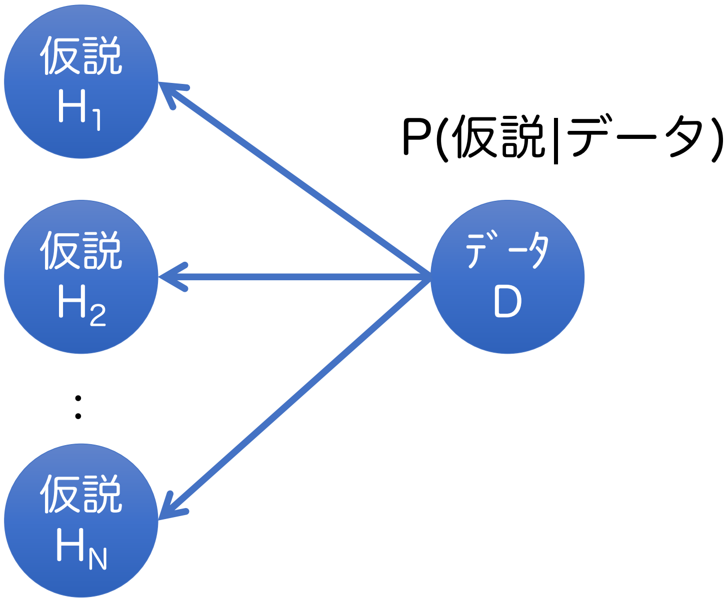
ベイズ統計

古谷知之

講義概要

- ベイズ統計の基礎
- ベイズの定理の確率分布への拡張
- 事前分布・尤度関数・事後分布
- 二項分布を用いたベイズ推論
- ポアソン分布を用いたベイズ推論
- 自然共役事前分布

データから仮説（モデル）を推定する



データから仮説（モデル）を推定する

- データDから仮説H_1が成立する確率は以下のようになる

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{\sum_{i=1}^{n} P(D \cap H_i)}$$

$$= \frac{P(D|H_1)P(H_1)}{\sum_{i=1}^{n} P(D|H_i)P(H_i)}$$

$$= \frac{P(D|H_1)P(H_1)}{P(D)}$$

事前確率・事後確率・尤度

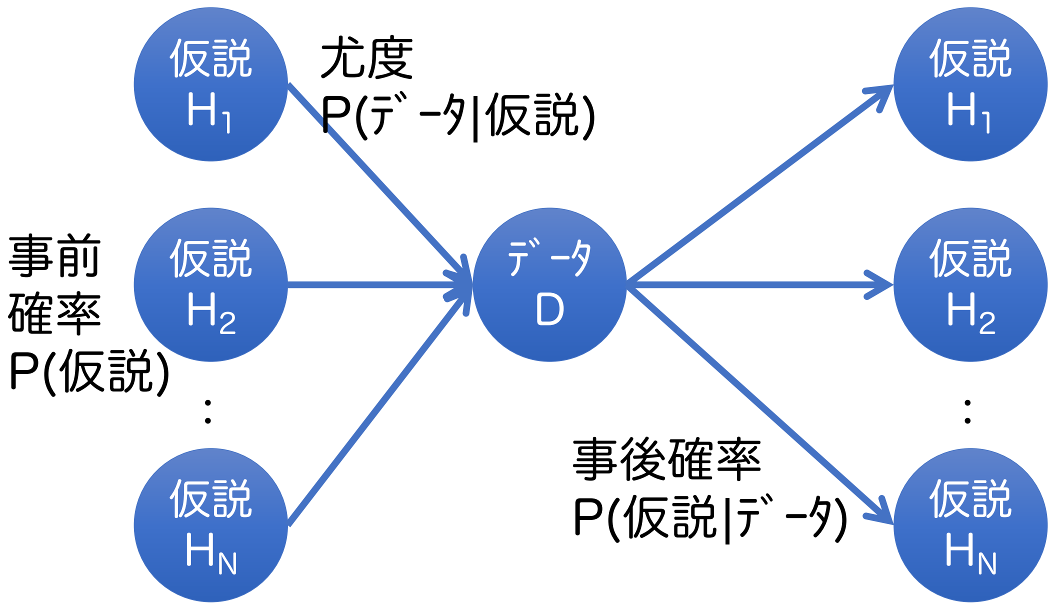
- ベイズの定理にもとづいて、$P(H)$を事前確率、$P(D|H)$を尤度、$P(H|D)$を事後確率という

尤度　事前確率

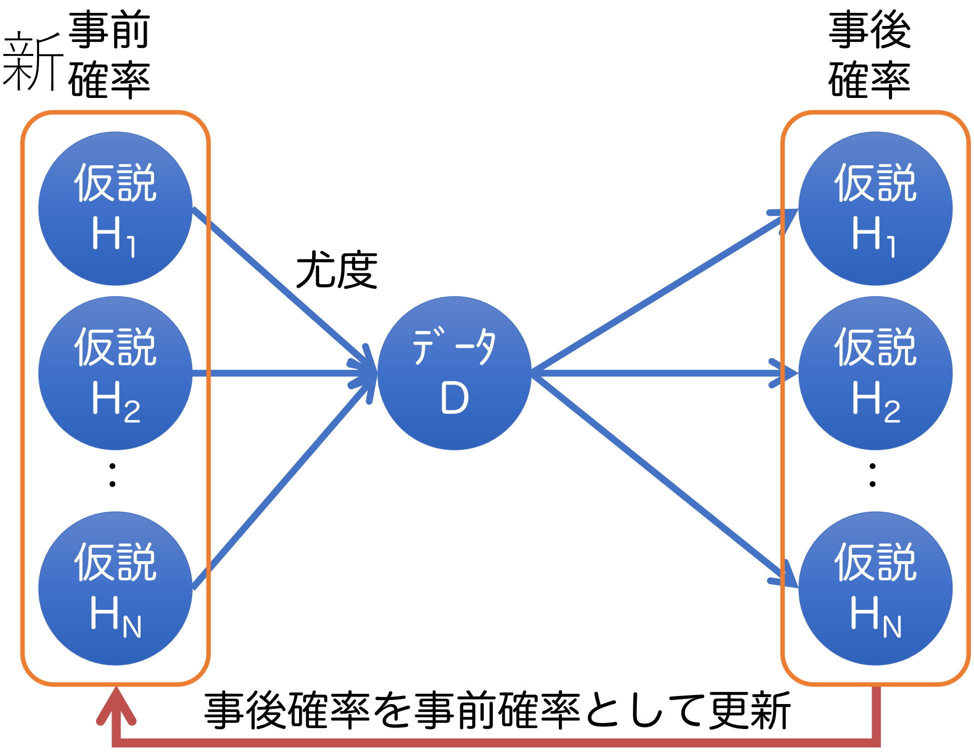
事後確率

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

事前確率・尤度・事後確率



ベイズ更新



ベイズの定理の拡張

- ベイズの定理の分母は定数であるため、ベイズの定理を次式のように置き換えることができる

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \propto P(D|H)P(H)$$

- 尤度関数と事前確率に確率分布を与えることで、事後確率の確率分布を計算することができる

確率分布の種類

- 離散的な確率分布
  - 一様分布
  - 二項分布
  - ポアソン分布
- 連続的な確率分布
  - 正規分布
  - ベータ分布
  - ガンマ分布

試合の勝敗とベイズ推定

- あなたはあるスポーツチームAのアナリスト兼ストラテジストであるとする。
- あなたのチームAは、ある別のチームBと初めて戦うことになった。
- チームBと実力が似通っている別のチームとは、既に10戦して6勝している。
- このとき、チームBに勝つことができると期待できるだろうか？

試合の勝敗とベイズ推定

- 試合$= G$、勝ち$= V$とすると、
- 求めたい事後確率は　$P(V|G)$
- 得られている情報（尤度）$P(G|V)$は、「これまでに10試合して6勝している」
- 事前確率$P(V)$と事後確率$P(V|G)$はどのようなものになるのだろうか？

- 試合の勝敗は二項分布に従う

ベルヌーイ試行と二項分布

- 0か1かしかない試行において、n回の試行でs回成功し、その確率pがわかっているとき、実験が成功する期待値は以下のベルヌーイ試行に従う
- ベルヌーイ試行の確率分布を二項分布といい、その分布は次式の確率密度関数に従う

$$Binom(n, p) = {}_nC_s \cdot p^s \cdot (1-p)^{n-s} \approx p^s \cdot (1-p)^{n-s}$$

ベルヌーイ試行と二項分布

- ベルヌーイ試行の確率変数と確率値を下表のように与えた時、次式が成立する
- 平均：$\mu = np$
- 分散：$\sigma^2 = np(1-p)$

確率変数X	0	1
確率	$1-p$	p

試合の勝敗とベイズ推定

- 尤度$P(G|V)$の二項分布は、

$$Binom(10, 0.6) = {}_{10}C_6 \cdot p^6 \cdot (1-p)^{10-6} \approx p^6 \cdot (1-p)^4$$

試合の勝敗とベイズ推定

- 二項分布の尤度から事後確率を求めたい場合、どのような事前確率が適しているだろうか？

- $p^s \cdot (1-p)^{n-s}$の形をした確率分布
  ⇒ベータ分布$B(\alpha,\beta)$

ベータ分布

- 二項分布を式変形すると次式のベータ関数に従うベータ分布となる

$$s = \alpha - 1, n - s = \beta - 1$$

$$Binom(n,p) \approx p^s \cdot (1-p)^{n-s} = \frac{(\alpha-1)!\,(\beta-1)!}{(\alpha+\beta-1)!}$$

$$B(\alpha,\beta) \approx k \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

$$0 < p < 1, 0 < \alpha, \beta, k = \text{定数}$$

試合の勝敗とベイズ推定

- チームBとは初顔合わせであるため、「理由不十分の原則」から勝敗の期待値（平均）が$1/2$となるようなベータ分布を事前確率として採用するのが良い。
- $\alpha - 1 > 0, \beta - 1 > 0, \alpha = \beta$となる最小の整数$\alpha = \beta = 2$

- ベータ分布$B(2, 2)$ を事前確率$P(V)$の確率分布に採用するとどうなるか？

ベータ分布

- ベータ分布

$$B(\alpha, \beta) \approx k \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1}$$
$$0 < p < 1, 0 < \alpha, \beta, k = \text{定数}$$

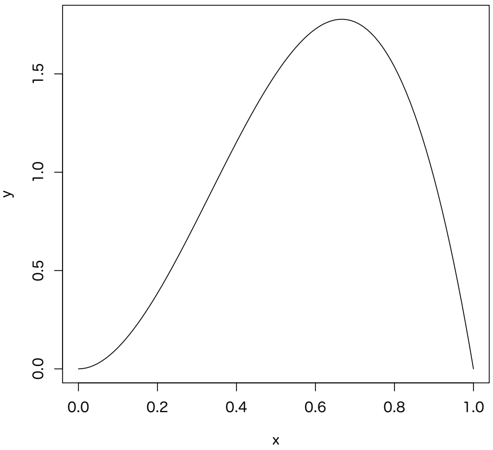
については、次式が成立する

$$\mu = \frac{\alpha}{\alpha + \beta}$$

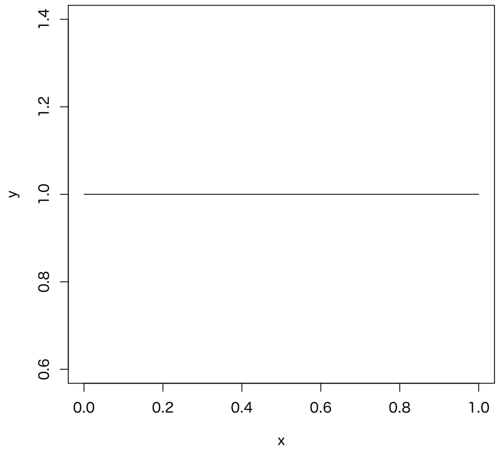
$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

ベータ分布

$B(3,2)$

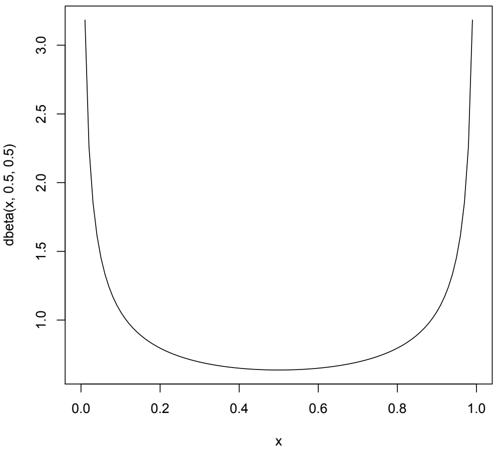


$B(3,2)$=一様分布

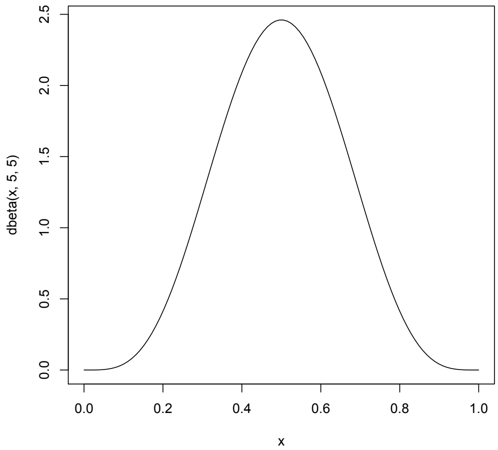


ベータ分布

$B(0.5,0.5)$



$B(5,5)$



ベータ分布は万能な確率分布

- ベータ分布は以下の分布に変化できる
  - 一様分布、線形分布
  - 単調増加・単調減少分布
  - 単峰分布
  - 左右対称分布

試合の勝敗とベイズ推定

- 尤度$P(G|V) = p^6(1-p)^4$
- 事前確率$P(V) = B(2,2) = p(1-p)$
- 事後確率

$$P(V|G) \approx p^6(1-p)^4 \times p(1-p) = p^7(1-p)^5 = B(8,6)$$

試合の勝敗とベイズ推定

- チームBとの試合の勝敗に関する期待値（平均）は、ベータ関数の平均から、

$$E\big(P(V|G)\big) \approx E\big(B(8,6)\big) = \frac{8}{8+6} \approx 0.571$$

- $0.571 > 0.5$なので、勝利への期待が上回る？

事前分布・尤度関数・事後分布

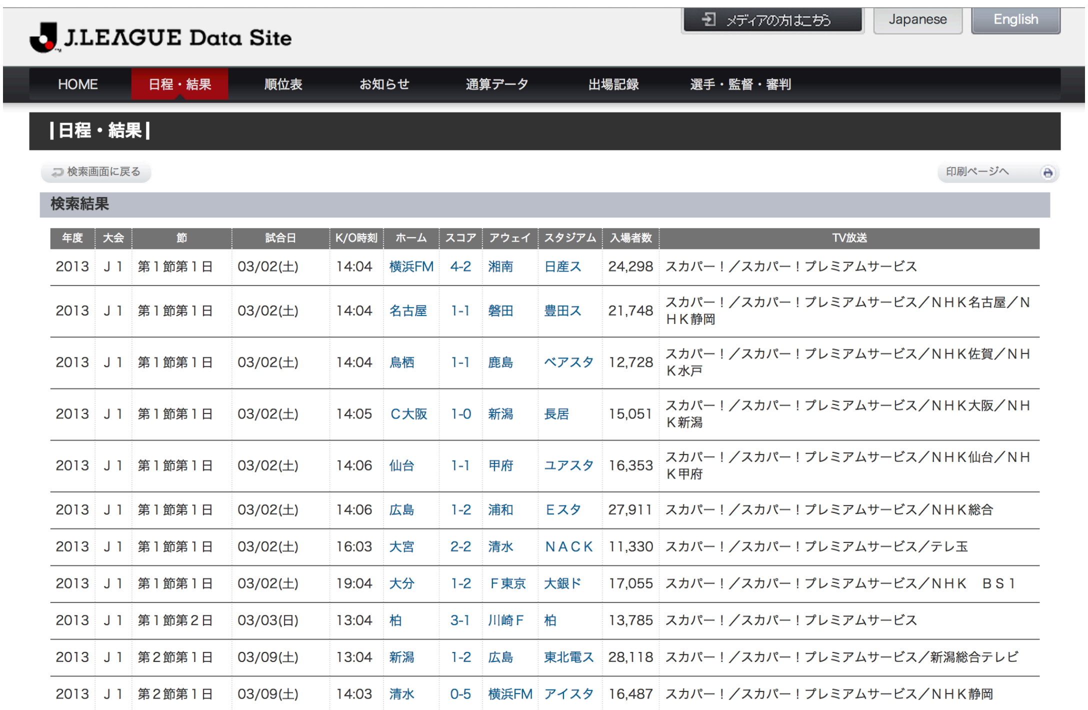
- 事前確率・尤度・事後確率は、ある確率値ではなく、確率分布を取ることがある。このとき、
- 事前確率⇒事前分布
- 尤度⇒尤度関数
- 事後確率⇒事後分布

  という。

試合の得点とベイズ推定

- あなたが関わっている競技は、野球やサッカー、アメフトなどのように少数得点で勝敗がきまります。
- 例えば、サッカーJリーグ(J1)の2013年シーズンの全306試合における各チームの得点分布（1試合での各チームの得点）は、下表のようになっています。(612チーム分、平均=1.44)

得点	0	1	2	3	4	5	6	7
チーム数	132	227	154	66	23	6	4	0


J.LEAGUE Data Site

メディアの方はこちら　Japanese　English

HOME　日程・結果　順位表　お知らせ　通算データ　出場記録　選手・監督・審判

|日程・結果|

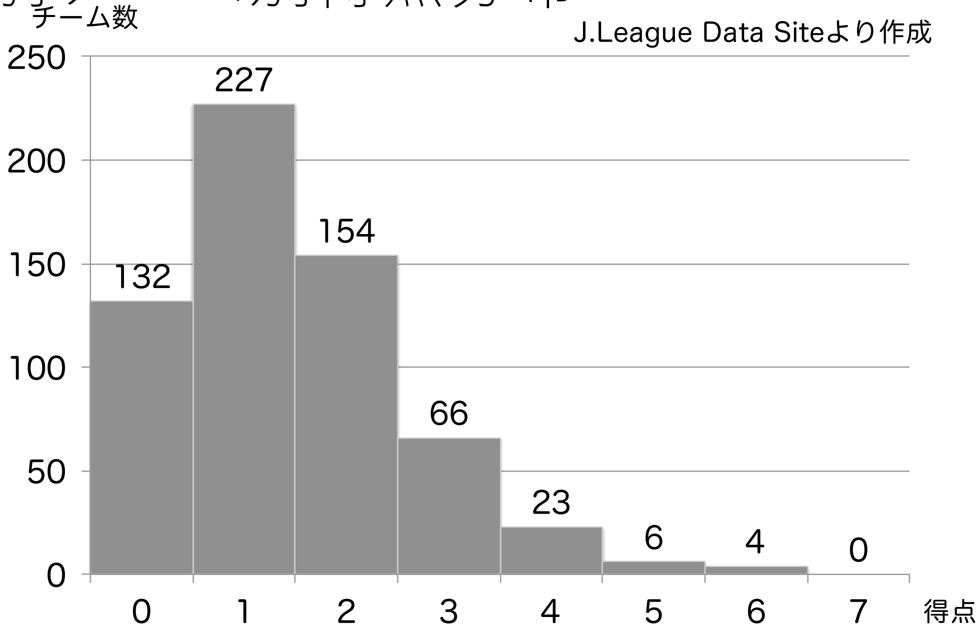
検索画面に戻る　印刷ページへ

検索結果

年度	大会	節	試合日	K/O時刻	ホーム	スコア	アウェイ	スタジアム	入場者数	TV放送
2013	Ｊ１	第１節第１日	03/02(土)	14:04	横浜FM	4-2	湘南	日産ス	24,298	スカパー！／スカパー！プレミアムサービス
2013	Ｊ１	第１節第１日	03/02(土)	14:04	名古屋	1-1	磐田	豊田ス	21,748	スカパー！／スカパー！プレミアムサービス／ＮＨＫ名古屋／ＮＨＫ静岡
2013	Ｊ１	第１節第１日	03/02(土)	14:04	鳥栖	1-1	鹿島	ベアスタ	12,728	スカパー！／スカパー！プレミアムサービス／ＮＨＫ佐賀／ＮＨＫ水戸
2013	Ｊ１	第１節第１日	03/02(土)	14:05	Ｃ大阪	1-0	新潟	長居	15,051	スカパー！／スカパー！プレミアムサービス／ＮＨＫ大阪／ＮＨＫ新潟
2013	Ｊ１	第１節第１日	03/02(土)	14:06	仙台	1-1	甲府	ユアスタ	16,353	スカパー！／スカパー！プレミアムサービス／ＮＨＫ仙台／ＮＨＫ甲府
2013	Ｊ１	第１節第１日	03/02(土)	14:06	広島	1-2	浦和	Ｅスタ	27,911	スカパー！／スカパー！プレミアムサービス／ＮＨＫ総合
2013	Ｊ１	第１節第１日	03/02(土)	16:03	大宮	2-2	清水	ＮＡＣＫ	11,330	スカパー！／スカパー！プレミアムサービス／テレ玉
2013	Ｊ１	第１節第１日	03/02(土)	19:04	大分	1-2	Ｆ東京	大銀ド	17,055	スカパー！／スカパー！プレミアムサービス／ＮＨＫ　ＢＳ１
2013	Ｊ１	第１節第２日	03/03(日)	13:04	柏	3-1	川崎Ｆ	柏	13,785	スカパー！／スカパー！プレミアムサービス
2013	Ｊ１	第２節第１日	03/09(土)	13:04	新潟	1-2	広島	東北電ス	28,118	スカパー！／スカパー！プレミアムサービス／新潟総合テレビ
2013	Ｊ１	第２節第１日	03/09(土)	14:03	清水	0-5	横浜FM	アイスタ	16,487	スカパー！／スカパー！プレミアムサービス／ＮＨＫ静岡

https://data.j-league.or.jp/SFMS01/search?competition_years=2013&competition_frame_ids=1&competition_ids=347&tv_relay_station_name=

2013年サッカーJリーグ(J1)
試合別チーム別得点分布



試合の得点のベイズ推定

- いまあなたは、チームの得点力がリーグの平均と比較して弱く、オフシーズンに攻撃力を強化すべきか検討しています。
- 昨シーズンはリーグ全体で1試合1チームあたりの平均得点が1.44点でしたが、あなたのチームは平均得点が1.38点でした。

得点	0	1	2	3	4	5	6	7
リーグ全体	132	227	154	66	23	6	4	0
あなたのチーム	5	16	8	5	0	0	0	0

試合の得点のベイズ推定

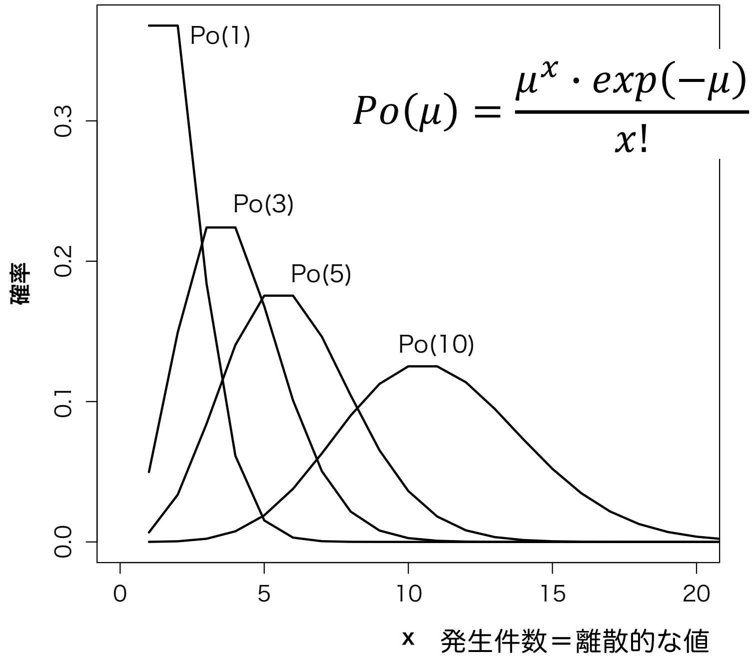
- あなたのチームがJ1リーグ平均並みまで得点力を強化するとすれば、チームの得点力はどの程度になるだろうか？
- 試合別チーム別得点を何らかの確率分布で表すことができればよいのではないか？
- このとき、
  - 事前分布＝昨季のあなたのチームの得点分布
  - 尤度関数＝昨季のリーグ全試合得点分布
  - 事後分布＝期待されるチーム得点分布

ポアソン分布

- 非常に多くの観測回数が繰り返されるものの、観測ケースの発生頻度が非常に低い場合に用いられる確率分布
- 一定範囲内（時間、回数、空間）である事象が発生する平均値μをもちいて計算される
- 試行回数n、発生頻度pとすると、$\mu = np$

$$Po(\mu) = \frac{\mu^x \cdot exp(-\mu)}{x!}$$

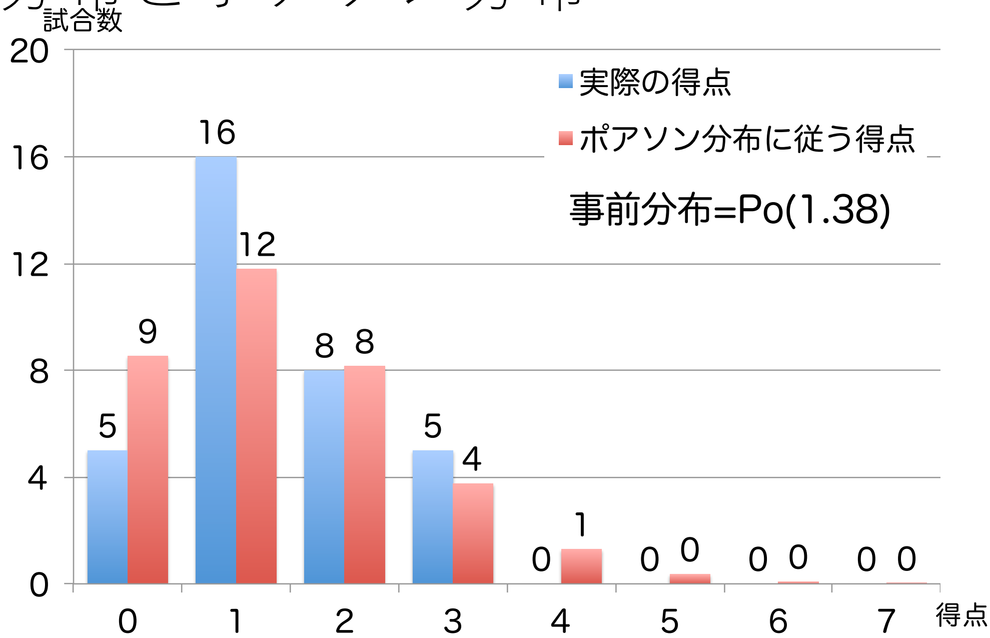
ポアソン分布



あなたのチームの得点のポアソン分布は？

得点	全チーム
0	$34*1.38^0 \times exp(-1.38)/0! \fallingdotseq$
1	$34*1.38^1*exp(-1.38)/1! \fallingdotseq$
2	$34*1.38^2*exp(-1.38)/2! \fallingdotseq$
3	$34*1.38^3*exp(-1.38)/3! \fallingdotseq$
4	$34*1.38^4*exp(-1.38)/4! \fallingdotseq$
5	$34*1.38^5*exp(-1.38)/5! \fallingdotseq$
6	$34*1.38^6*exp(-1.38)/6! \fallingdotseq$
7	$34*1.38^7*exp(-1.38)/7! \fallingdotseq$

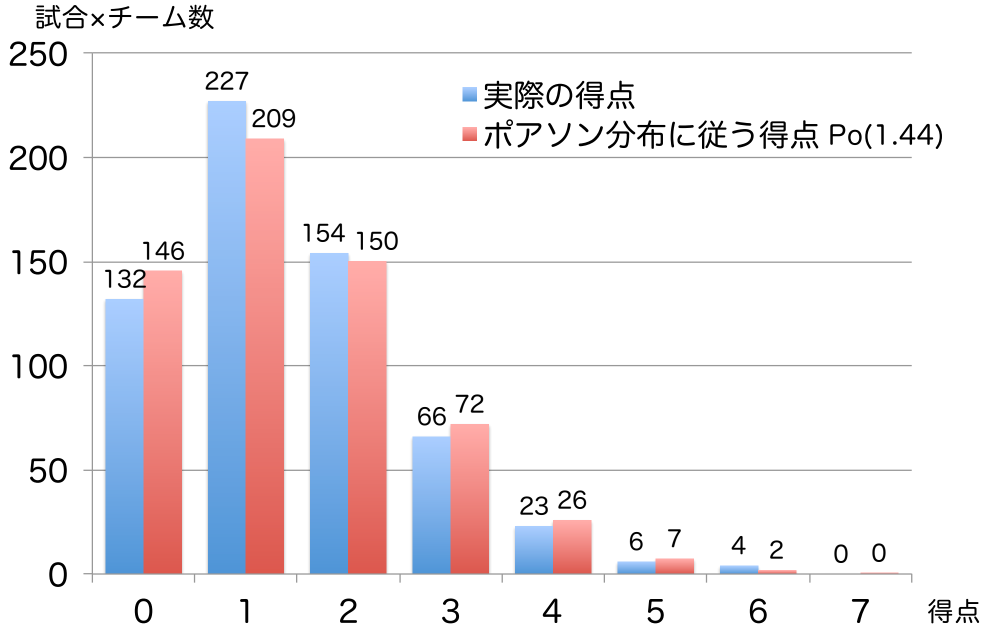
あなたのチームの全試合の得点分布とポアソン分布



全試合得点のポアソン分布の計算

得点	全チーム
0	612*1.44^{0}×exp(-1.44)/0! ≒146
1	612*1.44^{1}*exp(-1.44)/1! ≒209
2	612*1.44^{2}*exp(-1.44)/2! ≒150
3	612*1.44^{3}*exp(-1.44)/3! ≒72
4	612*1.44^{4}*exp(-1.44)/4! ≒26
5	612*1.44^{5}*exp(-1.44)/5! ≒7
6	612*1.44^{6}*exp(-1.44)/6! ≒2
7	612*1.44^{6}*exp(-1.44)/6! ≒0

全試合の得点分布とポアソン分布



ガンマ関数

- ポアソン分布の形と似た形の確率密度関数に、ガンマ分布がある
- ガンマ関数$\Gamma(\alpha)$は次式で表される

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$$
$$\alpha > 0$$

- ガンマ関数$\Gamma(\alpha)$は以下のような性質を持つ

$$\Gamma(\alpha) = (\alpha - 1)!$$
$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$
$$\Gamma(1/2) = \sqrt{\pi}$$

ガンマ分布

- ガンマ分布の確率密度関数$f(x)$は次式のようになる

$$f(x) = Ga(\alpha, \lambda) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x \geq 0$$

- ガンマ分布の平均$E(X)$と分散$V(X)$は以下の通り

$$E(X) = \frac{\alpha}{\lambda}$$

$$E(X) = \frac{\alpha}{\lambda^2}$$

ベータ分布とガンマ分布

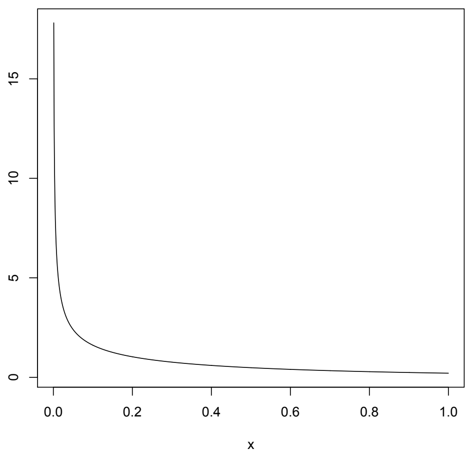
- ベータ関数の確率密度関数は以下のように式変形できる

$$Binom(n, p) \approx p^{\alpha} \cdot (1-p)^{\beta-1}$$
$$= \frac{(\alpha-1)!\,(\beta-1)!}{(\alpha+\beta-1)!}$$
$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

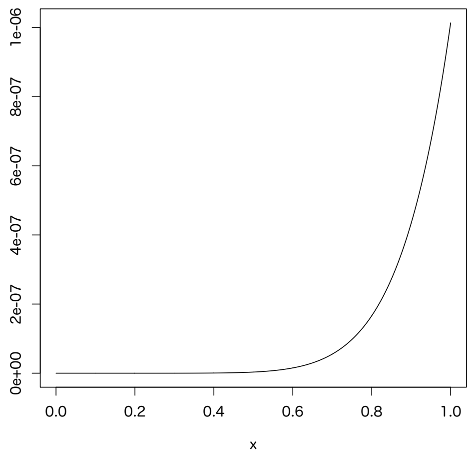
- ここで$\Gamma(\alpha)$はガンマ関数という

ガンマ分布

$\Gamma(0.5)$



$\Gamma(10)$



試合の得点とベイズ推定：事前分布

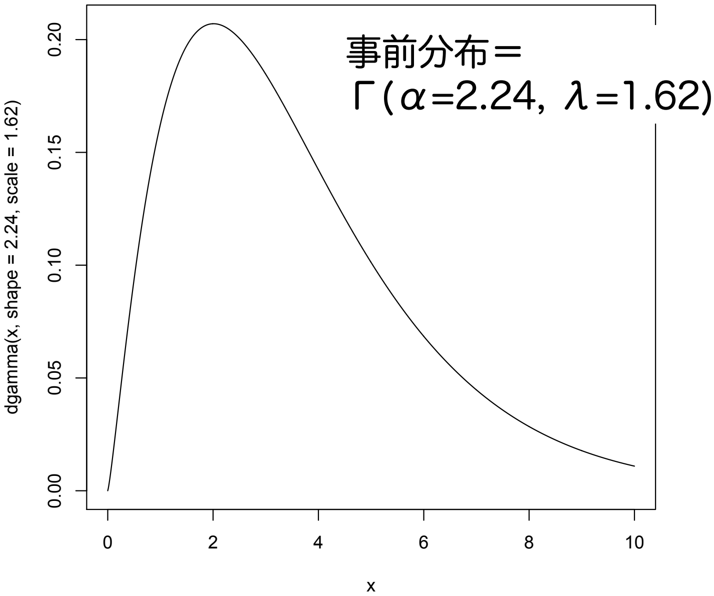
あなたのチームの1試合あたりの得点の平均と分散は以下のとおりである

- 平均1.38
- 分散0.85

これをガンマ分布で表すには、

- 平均$\alpha/\lambda = 1.38$
- 分散$\alpha/\lambda^2 = 0.85$ より、
- $\lambda \approx 1.62$、$\alpha \approx 2.24$を用いる

事前分布のガンマ分布



試合の得点とベイズ推定：尤度関数

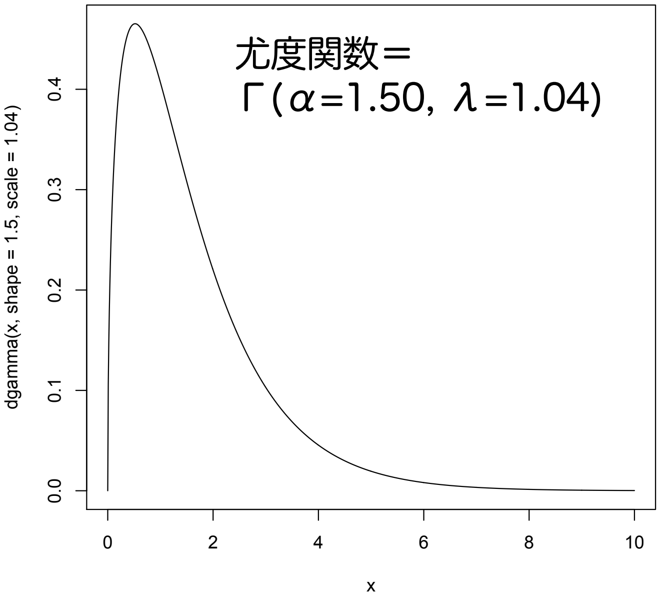
リーグ全体での1試合1チームあたりの得点の平均と分散は以下のとおりである

- 平均1.44
- 分散1.38

これをガンマ分布で表すには、

- 平均$\alpha/\lambda = 1.44$
- 分散$\alpha/\lambda^2 = 1.38$より、
- $\lambda \approx 1.04$、$\alpha \approx 1.50$を用いる

尤度関数のガンマ分布



試合の得点のベイズ推定

- 事前分布$\Gamma(2.24, 1.62)$
- 尤度関数$\Gamma(1.50, 1.04)$

- 事後分布
$$Ga(2.24, 1.62) \cdot Ga(1.50, 1.04) \approx Ga(2.27, 2.66)$$

試合の得点のベイズ推定

- 事後分布

$$
\begin{aligned}
&Ga(2.24, 1.62) \cdot Ga(1.50, 1.04) \\
&= \frac{1.62^{2.24} \cdot x^{(2.24-1)} \cdot exp(-1.62x)}{\Gamma(2.24)} \cdot \frac{1.04^{1.50} \cdot x^{(1.50-1)} \cdot exp(-1.04x)}{\Gamma(1.50)} \\
&\propto x^{(2.24-1)} \cdot x^{(1.50-1)} \cdot exp(-1.62x) \cdot exp(-1.04x) \\
&= x^{(2.74-1)} \cdot exp(-2.66x) \\
&\propto Ga(2.74, 2.66)
\end{aligned}
$$

ポアソン分布を事前分布とする場合の事後分布

- 事前分布がポアソン分布の場合、尤度関数にガンマ分布を採用すると、事後分布はガンマ分布となる。

ガンマ関数(尤度)×ポアソン分布(事前)
∝ガンマ関数(尤度)×ガンマ関数(事前)
∝ガンマ関数(事後)

自然共役事前分布

- 事前分布と事後分布が、互いに似たような確率分布を持つような場合に、その事前分布を自然共役事前分布という
- 主な自然共役事前分布は以下のとおり

事前分布	尤度関数	事後分布
二項分布	ベータ分布	ベータ分布
ポアソン分布	ガンマ分布	ガンマ分布
正規分布	正規分布	正規分布
正規分布	正規分布	逆ガンマ分布