

第9回 続高橋セミナー  
最尤法によるポアソン回帰分析入門  
2020年6月7日

第6章 過分散・ゼロ過剰への対処

稀な現象ではないカウント・データは、過分散となりがちであることが知られている。このように過分散によりポアソン分布があてはめに難点があるカウント・データに対し、分散の大きさを調整できるガンマ・ポアソン分布（負の2項分布）のあてはめが望まれる。そこで、ガンマ・ポアソン分布の特徴について示し、さらに、カウントがゼロの反応を抑制するZero-Inflated Poisson 分布（ゼロ過剰ポアソン分布）、および、ゼロ過剰ガンマ・ポアソン分布についても検討を加える。さらに、これらの分布を誤差分布としたガンマ・ポアソン回帰、および、ゼロ過剰ガンマ・ポアソン回帰を適用した結果についても示す。

第6章 目次

6. 過分散・ゼロ過剰への対処	207
6.1. 負の二項分布	207
成功数を固定, 交通事故の件数, 負の2項分布 vs. ポアソン分布, 負の2項分布のパラメータ推定	
6.2. ガンマ・ポアソン分布	213
位置および形状パラメータに変換, ガンマ・ポアソン分布のパラメータ推定, 過分散パラメータを変化させた場合の形状	
6.3. 過分散の事例	218
6.4. ゼロ過剰ポアソン分布のあてはめ	221
6.5. ゼロ過剰ガンマ・ポアソン分布のあてはめ	225

6.6. ガンマ・ポアソン回帰 -----	228
ポアソン回帰 vs. ガンマ・ポアソン回帰, 甲羅の幅 $x$ に対するガンマ・ポアソン分布のあてはめ	
6.7. ゼロ過剰ガンマ・ポアソン回帰 -----	233
ガンマ・ポアソン回帰 vs. ゼロ過剰ガンマ・ポアソン回帰, 甲羅の幅 $x$ に対するゼロ過剰ガンマ・ポアソン分布のあてはめ	
文献索引, 索引, 解析用ファイル一覧 -----	237

## 第 9 回 続高橋セミナー 最尤法によるポアソン回帰分析入門

第 9 回 続高橋セミナー「最尤法によるポアソン回帰分析入門」は、ページ数が多いので章ごとに公開する。全体の章立てを次に示す。

### 目 次

はじめに -----	1
1. ポアソン分布に従う各種のカウント・データ-----	7
2. ニュートン・ラフソン法によるポアソン回帰 -----	63
3. 尤度比検定のためのデザイン行列-----	95
4. デザイン行列を用いた回帰分析入門-----	135
5. 反復重み付き最尤法によるポアソン回帰 -----	175
<b>6. 過分散・ゼロ過剰への対応-----</b>	<b>207</b>
7. 過分散がある場合の探索的ポアソン回帰 -----	237
8. 2 本の回帰直線の比較-----	269
9. 花数を共変量とした種子数の分析 -----	293
10. オフセットを含む探索的ポアソン回帰-----	323
11. デビアンス・逸脱度・残差・テコ比・4 種の残差 -----	359
12. パラメータの共分散行列の活用 -----	383
13. 最小 2 乗平均の謎を予測プロファイルで解く -----	421
文献, 文献索引, 索引, (解析用ファイル) 一覧 -----	461

## 6. 過分散・ゼロ過剰への対処

第 1.13 節で取り上げた、雌に連結する雄のサテライト数は、平均が 2.9191 匹、分散が 9.9120 匹であり、分散/平均の比が 3.40 と 1 よりもかなり大きな過分散であり、ポアソン分布のあてはめは否定的であった。そのために、過分散を考慮するガンマ・ポアソン分布（負の 2 項分布）のあてはめを JMP で行ったが、ゼロとなるサテライト数が多いこともあり、ゼロ過剰（Zero-Inflated）モデルの適用の検討も示唆した。また、第 1.12 節で取り上げた「医院への通院回数」のように、稀な現象ではないカウント・データは、過分散となりがちであることが知られている。このように過分散によりポアソン分布があてはめに難点があるカウント・データに対し、分散の大きさを調整できるガンマ・ポアソン分布（負の 2 項分布）のあてはめが望まれる。そこで、ガンマ・ポアソン分布の特徴について示し、ゼロ過剰ポアソン分布、および、ゼロ過剰ガンマ・ポアソン分布についても検討を加える。さらに、これらの分布を誤差分布としたガンマ・ポアソン回帰、および、ゼロ過剰ガンマ・ポアソン回帰を適用した結果についても示す。

### 6.1. 負の 2 項分布

#### 成功数を固定

負の 2 項分布は、成功の確率を  $\pi$  としたときに、一定の成功数  $k$  が得られるまでの試行を行なった場合の失敗数  $Y$  の分布であり、次式で与えられる。

$$\text{NegBinomial}(Y; k, \pi) = \left[ \binom{Y+k-1}{k-1} \pi^{k-1} (1-\pi)^Y \right] \pi \quad (6.1)$$

簡単な事例で説明しよう。成功の確率を  $\pi = 0.65$ ，成功数を  $k = 3$  と固定する。成功する場合を○，失敗する場合を×とした時に，3 回連続して成功する場合は，

1 2 3 回目  
○ ○ ○

となり，失敗の数は  $Y = 0$  なので，その確率は，

$$\text{NegBinomial}(0; k, \pi) = \left[ \binom{0+3-1}{3-1} \pi^{3-1} (1-\pi)^0 \right] \pi = \binom{2}{2} 0.65^3 = 0.2746$$

である。最後の第 3 回目の試行は，常に成功○なので，第 2 回目までの試行に対する確率が主体となる。

次に4回目の施行が成功○となる場合を考えよう．第1回目から第3回目まで1回は失敗，2回は成功の場合は，4回目で成功○なので成功数は $k=3$ となり，次の3通りの場合である．

1	2	3	4 回目
×	○	○	○
○	×	○	○
○	○	×	○

失敗数は $Y=1$ ，成功数は $k=3$ なので，この確率は，

$$NegBinomial(1; k, \pi) = \left[ \binom{1+3-1}{3-1} \pi^{3-1} (1-\pi)^1 \right] \pi = \binom{3}{2} 0.65^3 \times 0.35^1 = 3 \times 0.0961 = 0.2884$$

である，引き続き5回目成功○となる場合を考えよう．第1回目から第4回目までに2回失敗し，5回目で成功する場合は，次の6通りであり，

1	2	3	4	5 回目
×	×	○	○	○
×	○	×	○	○
×	○	○	×	○
○	×	×	○	○
○	×	○	×	○
○	○	×	×	○

その確率は，

$$NegBinomial(2; k, \pi) = \left[ \binom{2+3-1}{3-1} \pi^{3-1} (1-\pi)^2 \right] \pi = \binom{4}{2} 0.65^3 \times 0.35^2 = 6 \times 0.0336 = 0.2018$$

である．更に6回目で成功，7回目で成功と全体の成功の数を $k=3$ と固定し，順次計算して確率が0.0000となるまで計算した結果を表6.1に示す．2項分布と負の2項分布を対比すると，2項分布は， $N$ 回目の試行中  $Y=0,1,\dots,N$ 回の成功する確率であり，負の2項分布は，最後の試行が成功するとした場合に，試行回数は制限せずに $k-1$ 回が成功する確率に，最後の試行が成功する確率を掛ける．

分布	試行回数	成功の確率
2項分布	$N$ 回の試行中	$Y$ 回の成功する確率
負の2項分布	$Y+k-1$ 回の試行中	$k-1$ 回の成功する確率 $k$ 回目は成功

負の2項分布は，成功数 $k$ を固定し失敗数 $Y$ を変化させることに対応している．組み合わせ数は，Excelの  $\text{Combin}(Y+k-1, k-1)$  関数で計算し，(成功の確率の $k$ 乗) $\times$ (失敗の確率の

$Y$  乗) を求め、その積から負の 2 項分布の確率を計算している。なお、Excel には、負の 2 項分布の関数 `NegBinom.dist()`

$$\text{NegBinomial}(Y; k, \pi) = \text{NegBinom.dist}(Y, k, \pi, \text{false})$$

を使って直接計算することもできる。

## 交通事故の件数

ある地域で交通事故の発生の（あり，なし）の日を想定する。失敗を事故が起きた日とする。事故が起きなかった日を成功とし、観測を始めて事故が起きなかった日が  $k=3$  日となるまでに事故が起きた日の数を失敗数  $Y$  とする。事故が起きなかった日が 3 日となった翌日から再び観測を始める。事故が起きない日の確率を  $\pi=0.65$  とした場合の失敗数  $Y$ （事故が起きた日数）は、負の 2 項分布に従うと想定される。

表 6.1 に、負の 2 項分布に従うとしとしたときの失敗数  $Y$  の確率の計算結果を示す。試行数を  $(Y+k-1)$  としたときの  $(k-1)$  となる組み合わせ数を Excel の `Combin()` 関数で計算し、 $k$  日の成功確率、事故が起きた日数  $Y$  の確率を計算し、それらの積を計算している。組合せ数と確率の積から失敗数  $Y$  に対する負の 2 項分布の確率が計算されている。もちろん、`NegBinom.dist()` 関数で計算した結果と一致する。このような観察を 1,000 回間続けたとした場合の失敗  $Y$  の件数  $n_i$  を計算している。事故が 3 日間連続して起きなかった ( $Y_1=0$ ) のは 275 件、1 日だけ事故が起きた ( $Y_2=1$ ) のは 289 件で、最も事故が多発した ( $Y_{10}=9$ ) 場合には 1 件である。

表 6.1 負の 2 項分布 ( $k=3, \pi=0.65$ ) に対する事故件数の例

$i$	試行数	成功数	失敗数	負の二項分布			1,000回中
	$Y+k-1$	$k-1$	$Y$	組合せ	$\pi^{k-1}(1-\pi)^Y\pi$	確率 $P$	件数 $n$
1	2	2	0	1	0.2746	0.2746	275
2	3	2	1	3	0.0961	0.2884	289
3	4	2	2	6	0.0336	0.2018	202
4	5	2	3	10	0.0118	0.1177	118
5	6	2	4	15	0.0041	0.0618	62
6	7	2	5	21	0.0014	0.0303	30
7	8	2	6	28	0.0005	0.0141	14
8	9	2	7	36	0.0002	0.0064	6
9	10	2	8	45	0.0001	0.0028	3
10	11	2	9	55	0.0000	0.0012	1
11	12	2	10	66	0.0000	0.0005	0
12	13	2	11	78	0.0000	0.0002	0
13	14	2	12	91	0.0000	0.0001	0
14	15	2	13	105	0.0000	0.0000	0
		<b><math>k=3</math></b>		<b><math>\pi=0.65</math></b>		1.0000	1,000

## 負の 2 項分布 vs. ポアソン分布

表 6.2 に示すように、ある地域での事故がない日が 3 日となるまでの事故の発生件数の統計が得られたとして、この分布はポアソン分布に従うと見なすことができるのだろうか。事故件数の平均  $\bar{Y}$  は、

$$\bar{Y} = \frac{\sum_{i=1}^{10} Y_i n_i}{N} = \frac{1,604}{1,000} = 1.6040$$

となり、分散は

$$\bar{Y} = \frac{\sum_{i=1}^{10} (Y_i - \bar{Y})^2 n_i}{N - 1} = \frac{2399.1840}{999} = 2.4016$$

と、過分散（分散/平均=1.4972）である。ポアソン分布を仮定した場合の

$$P_i = \text{Poissonn.dist}(Y_i, \mu = 1.6040, \text{false})$$

の確率に 1,000 回を掛けた事故件数との差は、大きなプラスマイナスがあり、あてはまっているとは言い難い。事故件数の分布がポアソン分布と見なせない場合には、ある地域で特異的に事故が発生しやすい交差点がいくつかあるような場合、複数のポアソン分布が混合している場合にも過分散が起きやすい。

表 6.2 事故件数の人工データに対するポアソン分布のあてはめ

$i$	事故日数 $Y$	発生件数 $n$	ポアソン分布		件数 差
			確率 $P$	件数	
1	0	275	0.2011	201.1	-73.9
2	1	289	0.3225	322.5	33.5
3	2	202	0.2587	258.7	56.7
4	3	118	0.1383	138.3	20.3
5	4	62	0.0555	55.5	-6.5
6	5	30	0.0178	17.8	-12.2
7	6	14	0.0048	4.8	-9.2
8	7	6	0.0011	1.1	-4.9
9	8	3	0.0002	0.2	-2.8
10	9	1	0.0000	0.0	-1.0
11	10	0	0.0000	0.0	0.0
全体		1,000	1.6040	2.4016	1.4972
		$N$	平均	分散	比

Excel には、平均と分散の計算に `Avarege()`関数と `Ver.s()`関数があるが、度数  $n_i$  を含む場合の関数はないので、平均=SumProduct( $Y$  の範囲,  $n$  の範囲)/ $N$ 、分散=SumProduct( ( $Y$  の範囲 - 平均)<sup>2</sup>,  $n$  の範囲 ) / ( $N-1$ ) で直接計算することができる。

図 6.1 に負の 2 項分布を仮定して生成した事故件数データに対し、ポアソン分布をあてはめた場合を示す。負の 2 項分布から生成されたデータから、平均が 1.6040、分散が 2.4061 であり、ポアソン分布は、平均が 1.6040、分散が同じく 1.6040 なので、ポアソン分布は、負の 2 項分布に対して、相対的に尖った形状となっている。

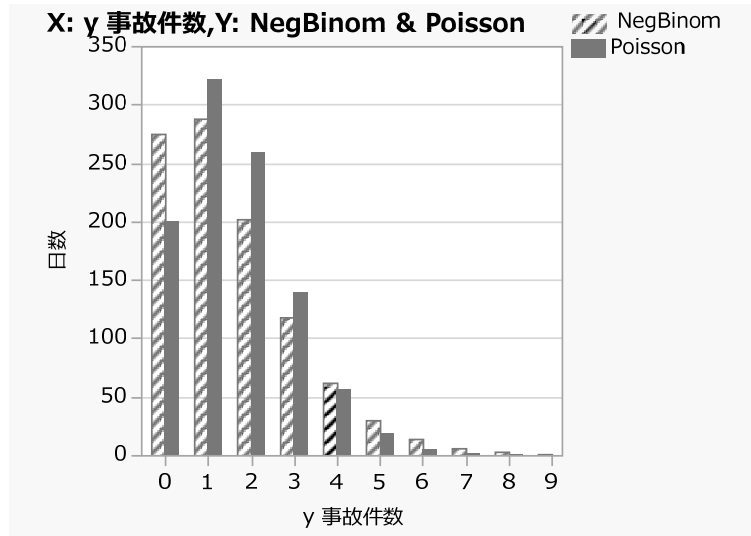


図 6.1 負の 2 項分布に従う事故件数データへのポアソン分布のあてはめ

### 負の 2 項分布のパラメータ推定

実際に観察されたデータで負の 2 項分布をあてはめてパラメータを推定したいが、どうしたら良いのだろうか。ポアソン分布の位置パラメータ  $\mu$  は、算術平均と等しいが、負の 2 項分布のパラメータである出現確率  $\pi$ 、成功数  $k$  を与えられた度数分布から推定したい。ある初期値を与えて負の 2 項分布の確率  $P_i$  を計算し、対数尤度

$$\ln L = \sum_{i=1}^{10} n_i (\ln \hat{P}_i) = \sum_{i=0}^9 n_i (\ln(\text{NegBinom.dist}(Y_i, \hat{k}, \hat{\pi}, \text{false})))$$

を求めて、Excel のソルバーで、 $\ln L$  を最大にするように出現確率  $\hat{\pi}$ 、成功数  $\hat{k}$  を変化させれば、最尤解が求められると思うかもしれない。残念ながら、成功数  $\hat{k}$  は整数なので、NegBinom.dist() 関数を使った場合には、ソルバーで変化させ最尤解を得ることができない。

負の 2 項分布は、階乗の計算を含むので、これをガンマ関数で置き換えて

$$\begin{aligned} \left[ \binom{Y+k-1}{k-1} \pi^{k-1} (1-\pi)^Y \right] \pi &= \frac{(Y+k-1)!}{[(Y+k-1)-(k-1)]! (k-1)!} \pi^k (1-\pi)^Y \\ &= \frac{\Gamma(Y+k)}{\Gamma(Y+1) \Gamma(k)} \pi^k (1-\pi)^Y \end{aligned} \quad (6.2)$$

のように、実数でも計算できるようにする。成功数  $k$  が整数であれば、式 (6.2) は、負の 2 項分布に一致するが、成功数  $k$  が実数の場合に離散分布の“負の 2 項分布”と言って良いのだろうか。失敗数  $Y$  も実数で与えることも可能であり、疑問である。

事故件数のデータにガンマ関数を用いた“負の 2 項分布”のパラメータを推定するための Excel シートを表 6.3 に示す。“成功数”  $\hat{k} = 5.0$  回、成功の確率  $\hat{\pi} = 0.50$  を初期値にし、ソルバーで対数尤度  $\ln L$  を最大化した結果を示す。最尤解として  $\hat{k} = 3.1516$ 、 $\hat{\pi} = 0.6627$  が得られ

ている。人工的に作成した事故件数のデータのパラメータは $k=3$ ， $\pi=0.65$ であったので，完全には一致しないが，“負の 2 項分布”のパラメータ推定を，Excel のソルバーを用いた最尤法により推定できる。なお，件数 $N$ を増やすことにより，パラメータの推定値は， $k=3$ ， $\pi=0.65$ に漸近する。

表 6.3 ガンマ関数を用いた“負の 2 項分布”のパラメータ推定

$i$	事故日数 $Y$	件数 $n$	試行数 $Y+k-1$	成功数 $k-1$	ガンマ関数・負の二項分布			対数尤度 $\ln L_i$
					組合せ	$\pi^k (1-\pi)^Y$	確率 $P_i^{\wedge}$	
1	0	275	2.1516	2.1516	1.0000	0.273457	0.2735	-356.57
2	1	289	3.1516	2.1516	3.1516	0.092233	0.2907	-357.07
3	2	202	4.1516	2.1516	6.5421	0.031109	0.2035	-321.58
4	3	118	5.1516	2.1516	11.2342	0.010493	0.1179	-252.30
5	4	62	6.1516	2.1516	17.2771	0.003539	0.0611	-173.26
6	5	30	7.1516	2.1516	24.7118	0.001194	0.0295	-105.70
7	6	14	8.1516	2.1516	33.5735	0.000403	0.0135	-60.25
8	7	6	9.1516	2.1516	43.8931	0.000136	0.0060	-30.74
9	8	3	10.1516	2.1516	55.6982	0.000046	0.0026	-17.91
10	9	1	11.1516	2.1516	69.0138	0.000015	0.0011	-6.84
11	10	0	12.1516	2.1516	83.8629	0.000005	0.0004	0.00
		最尤解	$k^{\wedge}=$	3.1516	$\pi^{\wedge}=$	0.6627	$\ln L=$	-1682.23
		初期値	$k^{\wedge}=$	5.0000	$\pi^{\wedge}=$	0.5000	$\ln L=$	-2602.52

注) なぜ，一定の成功数 $k$ が得られるまでの試行を行なった場合の失敗数 $Y$ の分布を負の 2 項分布と言うのだろうか。箕谷 (2010)，統計分布ハンドブック 増補版，p608-10 に「負の 2 項展開に現れる 2 項係数が，負の 2 項分布と同じであることから，この分布は負の 2 項分布とよばれる。」ことが示されている。Excel での組合せ数の計算のガンマ関数は，Gamma()を用いている。対数尤度 $\ln L_i = n_i \ln(P_i)$ で求め，それらを足し合わせて全体の対数尤度 $\ln L$ を計算している。

計算シートの $i=1$ の行は，

$$Y + \hat{k} - 1 = 0 + 3.1516 - 1 = 2.1516$$

$$\text{組合せ} = \frac{\text{Gamma}(Y_1 + \hat{k})}{\text{Gamma}(Y_1 + 1) \times \text{Gamma}(\hat{k})} = \frac{2.3106}{1 \times 2.3106} = 1.0000$$

$$\hat{\pi}^{\hat{k}} (1 - \hat{\pi})^{Y_1} = 0.6627^{3.1516} \times (1 - 0.6627)^0 = 0.273457$$

$$\hat{P}_1 = 1.0000 \times 0.273457 = 0.27351$$

$$\ln L_1 = n_1 \ln(\hat{P}_1) = n_1 \ln(0.2735) = 275 \times (-1.2966) = -356.571$$

として計算されている。対数尤度 $\ln L$ は，

$$\ln L = \sum_{i=1}^{10} \ln L_i = -356.57 - 357.07 - \dots - 6.84 = -1682.23$$

である。なお，負の 2 項分布のパラメータ推定ができたとしても，成功数 $\hat{k}=3.1516$ 回，成功の確率 $\hat{\pi}=0.6627$ と解釈はできても，ポアソン分布のパラメータである平均との関連付けができないので，このままでは使い物にならない。

## 6.2. ガンマ・ポアソン分布

### 位置および形状パラメータに変換

Agresti (2013), *Categorical Data Analysis 3rd.* の section 14.4 の Negative Binomial Regression に、ガンマ分布とポアソン分布を融合した場合に、“負の 2 項分布”となることが示されている。“負の 2 項分布”は、式 (6.2) で示したように、

$$\text{NegBinom}(Y; k, \pi) = \frac{\Gamma(Y+k)}{\Gamma(Y+1)\Gamma(k)} \pi^k (1-\pi)^Y \quad (6.3)$$

となる。負の 2 項分布のパラメータは、 $\pi$  と  $k$  であり、ポアソン分布のパラメータと全く異なるので、パラメータが平均  $\mu$  (位置パラメータ) と過分散  $\sigma$  (形状パラメータ) となるように変換したい。成功の確率  $\pi$  を平均  $\mu$  と  $k$  で

$$\pi = \frac{k}{\mu + k} \quad (6.4)$$

のように置き換えると

$$\text{GammaPoisson}(Y; k, \mu) = \frac{\Gamma(Y+k)}{\Gamma(Y+1)\Gamma(k)} \left( \frac{k}{\mu+k} \right)^k \left( 1 - \frac{k}{\mu+k} \right)^Y \quad (6.5)$$

となる。さらに、 $k$  を  $1/\sigma$  で置き換え、整理すると

$$\begin{aligned} \text{GammaPoisson}(Y; \mu, \sigma) &= \frac{\Gamma(Y+1/\sigma)}{\Gamma(Y+1)\Gamma(1/\sigma)} \left( \frac{1/\sigma}{\mu+1/\sigma} \right)^{1/\sigma} \left( 1 - \frac{1/\sigma}{\mu+1/\sigma} \right)^Y \\ &= \frac{\Gamma(Y+1/\sigma)}{\Gamma(Y+1)\Gamma(1/\sigma)} \left( \frac{1}{1+\sigma\mu} \right)^{1/\sigma} \left( 1 - \frac{1}{1+\sigma\mu} \right)^Y \\ &= \frac{\Gamma(Y+1/\sigma)}{\Gamma(Y+1)\Gamma(1/\sigma)} \cdot \frac{(\mu\sigma)^Y}{(1+\mu\sigma)^{Y+1/\sigma}} \end{aligned} \quad (6.6)$$

が得られる。前節の表 6.1 で、成功の数を  $k=3$ 、成功の確率  $\pi=0.65$  とする負の 2 項分布の確率を計算し、1,000 回の観察で事故件数の分布を例示した。ガンマ・ポアソン分布のパラメータは、式 (6.4) を平均  $\mu$  (位置パラメータ) について解くと、

$$\begin{aligned} \mu &= \frac{(1-\pi)k}{\pi} \\ &= \frac{(1-0.65) \times 3}{0.65} = 1.6154 \end{aligned} \quad (6.7)$$

と負の 2 項分布の期待値になり、過分散  $\sigma$  (形状パラメータ) は、

$$\sigma = \frac{1}{k} = \frac{1}{3} = 0.3333 \quad (6.8)$$

となる。

負の二項分布の分散は,

$$Var(Y) = \frac{(1-\pi)k}{\pi^2} \quad (6.9)$$

なので, 式 (6.4) の  $\pi = k / (\mu + k)$  を代入し, 整理し, さらに  $k$  を  $k = 1 / \pi$  置き換えると

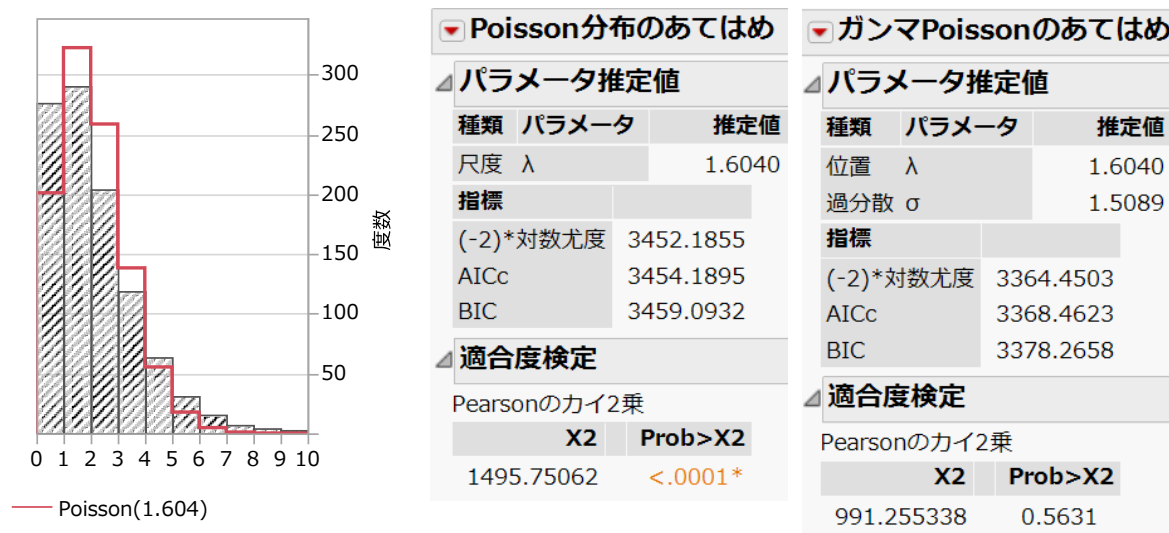
$$\begin{aligned} Var(Y) &= \frac{(1-\pi)k}{\pi^2} \\ &= \mu \cdot \frac{\mu + k}{k} \\ &= \mu(1 + \mu\sigma) \end{aligned} \quad (6.10)$$

が得られる. ガンマ・ポアソン分布の分散は, 期待値  $\mu$  を含んでいる. したがって, 過分散の形状パラメータ  $\sigma$  は, いわゆる分散の平方根に相当しているわけではないが, 期待値  $\mu$  が推定されれば, 分散も推定できるような「形状パラメータ」と理解される.

### ガンマ・ポアソン分布のパラメータ推定

JMP の「一変量の分布」を用いて, 事故件数データにポアソン分布およびガンマ・ポアソン分布の当てはめ結果を表 6.4 に示す. ポアソン分布の当てはめは, Pearson の適合度検定の結果からも,  $p < 0.0001$  と全く支持されない. 当然ながらガンマ・ポアソン回帰は良く当てはまっていて  $p = 0.5631$  と棄却されない.

表 6.4 事故の人工データに対するポアソン分布の当てはめ



推定されたパラメータは,  $\hat{\lambda} = \hat{\mu} = 1.6040$  と期待値 1.6164 とほぼ同様の推定値が得られている. 推定された過分散パラメータは,  $\hat{\sigma} = 1.5089$  と式 (6.6) での  $\sigma = 1/k = 0.3333$  と全く異なる. これは, JMP (Ver. 14) の「一変量のあてはめ」でガンマ・ポアソン分布の当てはめた場合は, 過分散  $\sigma'$  として次式が

$$\sigma' = 1 + \mu\sigma = 1 + 1.6154 \times 0.3333 = 1.5385$$

使われている．この $\sigma'$ は，過分散がないポアソン分布の場合に， $\sigma'=1$ となり，さらに $\mu\sigma$ を加えて平均値と過分散のスケールを合わせるための対応と解される．表 6.3 で推定された負の 2 項分布のパラメータ $\hat{\pi}=0.6627$ および $\hat{k}=3.1513$ を使い，

$$\hat{\mu} = \frac{(1-\hat{\pi})\hat{k}}{\hat{\pi}} = \frac{(1-0.6627) \times 3.1516}{0.6627} = 1.6040$$

$$\hat{\sigma} = \frac{1}{\hat{k}} = \frac{1}{3.1516} = 0.3173$$

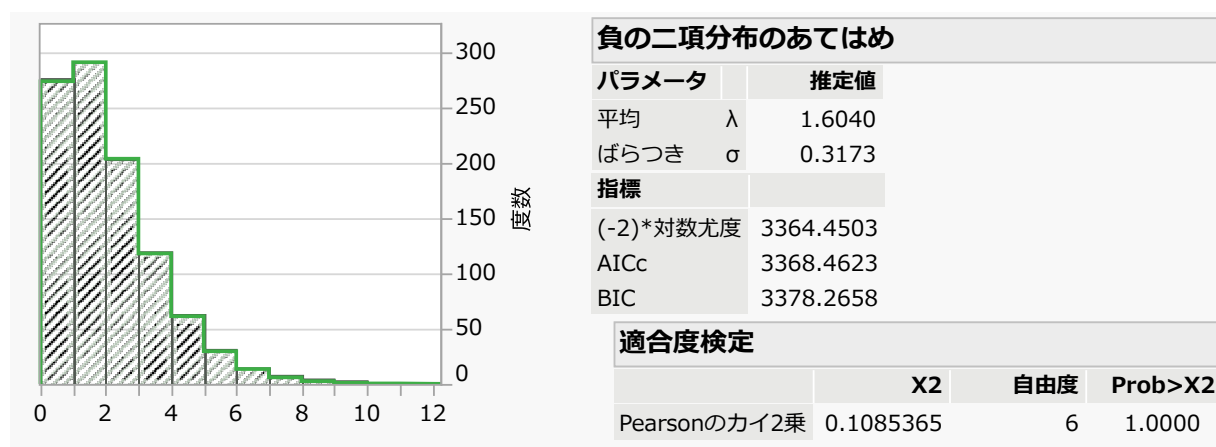
が得られ，表 6.4 で示し $\hat{\mu}$ に一致し， $\hat{\mu}$ と $\hat{\sigma}$ から，

$$\hat{\sigma}' = 1 + \hat{\mu}\hat{\sigma} = 1 + 1.6040 \times 0.3173 = 1.5089$$

が推定され，表 6.4 で示した「過分散 $\sigma$ 」に一致する．「指標」の欄の $(-2) \times \text{対数尤度} = 3364.4503$ は，表 6.3 で求めた対数尤度 $\ln L = -1682.23$ をマイナス 2 倍した結果に一致する．

2019 年にリリースされた JMP の Ver. 15 では，離散分布の選択画面から「ガンマ Poisson 分布のあてはめ」が消え，「負の 2 項分布のあてはめ」に置き換わっている．推定されるパラメータは，「平均 $\lambda$ 」は同じであるが，「過分散 $\sigma$ 」に代え「ばらつき $\sigma$ 」となり， $\hat{\sigma}=0.3173$ が推定されている．したがって，「負の 2 項分布」の通常のパラメータ $(k, \pi)$ が推定されているわけではない．

表 6.5 事故の人工データに対する負の 2 項分布のあてはめ



JMP (Ver 15.1) での負の 2 項分布のあてはめ。  
適合度検定で自由度が 6 となっているが，理由は調査中．

## 過分散パラメータを変化させた場合の形状

ガンマ・ポアソン分布の過分散パラメータ $\sigma$ が 0 に漸近した場合に，ポアソン分布となることが知られているので，実際に確認する．負の 2 項分布の場合は，成功数 $k$ は，1 以上の整数であるが，ガンマ・ポアソン分布の場合は実数となるので，組合せ数の計算をガンマ関数

に置き換えて、負の 2 項分布をガンマ・ポアソン分布として成功数  $k$  を実数化して、いくつかの  $k$  について分布の形状について検討する。

ガンマ・ポアソン分布の計算で、 $\sigma$  が小さくなると  $\Gamma(1/\sigma)$  が増大するために計算不能となるので、Excel の対数ガンマ関数  $\text{Gammaln}()$  を使って、計算する必要がある。

$$\begin{aligned} \text{GammaPoisson}(Y; \sigma, \mu) &= \frac{\Gamma(Y+1/\sigma)}{\Gamma(Y+1)\Gamma(1/\sigma)} \cdot \frac{(\mu\sigma)^Y}{(1+\mu\sigma)^{Y+1/\sigma}} \\ &= \exp\left[\ln\Gamma(Y+1/\sigma) - \ln\Gamma(Y+1) - \ln\Gamma(1/\sigma) + Y\ln(\mu\sigma) - (Y+1/\sigma)\ln(1+\mu\sigma)\right] \end{aligned} \quad (6.11)$$

表 6.6 および図 6.2 に示すように、 $\sigma=5$  の場合は、 $P_{y=0}=0.6433$  のように大きな確率で、 $\mu=1.6154$  を確保するために、 $y$  の大きい方に長く裾を引いている。小さな  $\sigma=0.0002$  に対しては、ポアソン分布との差はなくなっている。

表 6.6 ガンマ・ポアソン分布の形状の比較

Y	ガンマ・ポアソン 過分散 $\sigma$						ポアソン	$\sigma=0.0002$
	5.0	2.0	1.0	0.3333	0.1000	0.0002	P	との差
0	0.6433	0.4862	0.3824	0.2746	0.2237	0.1989	0.1988	0.0001
1	0.1145	0.1856	0.2362	0.2884	0.3111	0.3211	0.3212	0.0000
2	0.0611	0.1063	0.1459	0.2018	0.2380	0.2594	0.2594	0.0000
3	0.0399	0.0677	0.0901	0.1177	0.1324	0.1397	0.1397	0.0000
4	0.0284	0.0452	0.0556	0.0618	0.0598	0.0564	0.0564	0.0000
5	0.0212	0.0311	0.0344	0.0303	0.0233	0.0182	0.0182	0.0000
6	0.0164	0.0217	0.0212	0.0141	0.0081	0.0049	0.0049	0.0000
7	0.0129	0.0154	0.0131	0.0064	0.0026	0.0011	0.0011	0.0000
8	0.0103	0.0110	0.0081	0.0028	0.0008	0.0002	0.0002	0.0000
9	0.0084	0.0080	0.0050	0.0012	0.0002	0.0000	0.0000	0.0000
10	0.0069	0.0058	0.0031	0.0005	0.0001	0.0000	0.0000	0.0000
11	0.0057	0.0042	0.0019	0.0002	0.0000	0.0000	0.0000	0.0000
12	0.0047	0.0031	0.0012	0.0001	0.0000	0.0000	0.0000	0.0000
13	0.0039	0.0023	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
平均 $\mu=$	1.6154	1.6154	1.6154	1.6154	1.6154	1.6154	1.6154	
成功確率 $\pi=$	0.7558	0.5532	0.3824	0.1711	0.0583	0.0001		
成功数 $k=$	0.20	0.50	1.00	3.00	10.00	5000.00	—	
JMP: $1+\mu\sigma=$	9.0770	4.2308	2.6154	1.5385	1.1615	1.0003	1.0000	

$\text{GammaPoisson}(y=0; \sigma=5, \mu=1.6154)=0.6433$  に対する計算は、  
 $=\text{Exp}(\text{Gammaln}(0+1/5) - \text{Gammaln}(0+1) - \text{Gammaln}(1/5) + 0 \times \ln(1.6154 \times 5) - (0+1/5) \times \ln((1+1.6154 \times 5)))$   
 $=\text{Exp}(1.5241+0.0000 - 1.5241+0.0000 - 0.4411) = \text{Exp}(-0.4411) = 0.6433$  で求められている。

図 6.2 に示すように●印が、ほぼポアソン分布に一致し、平均  $\mu=1.6154$  に対して分散がほぼ等しい場合である。過分散パラメータ  $\sigma=0.10$  の場合（＊印）は、ほぼポアソン分布に重なっているが、 $\sigma=0.33$ （□印）になるとポアソン分布に比べ  $Y=0$  の確率が上方に乖離し、 $Y=1, 2, 3$  の場合は、下方に乖離し、 $Y \geq 4$  の場合は、図では識別しづらいが、●印から上方に

乖離していることがわかる．このような上下への乖離は，平均を  $\mu=1.6154$  と同じとすることで，反応を 0 とする割合が極端に大きくなり，右に長く裾を引く分布となる．

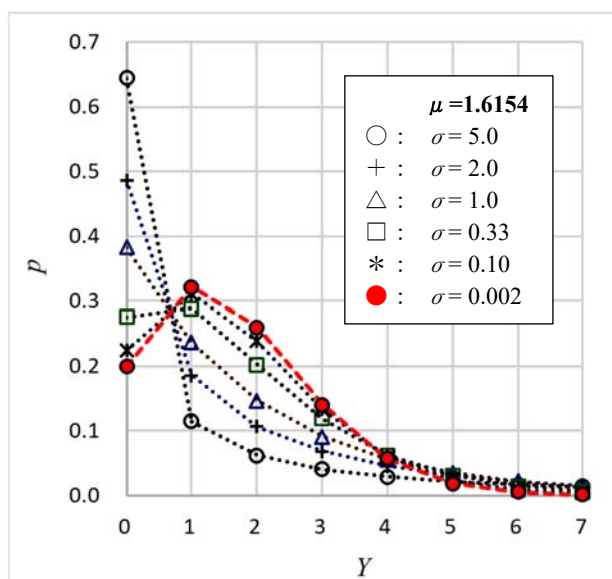


図 6.2 平均  $\mu=1.6154$  に対するガンマ・ポアソン分布の形状の比較

図 6.3 に平均を  $\mu=4.0$  とした場合のガンマ・ポアソン分布の形状を示す．形状パラメータ  $\sigma=0.33$  までは，ポアソン分布と同様に一山型であるが， $\sigma \geq 1.0$  以上では，指数分布的な片流れ型となっている．

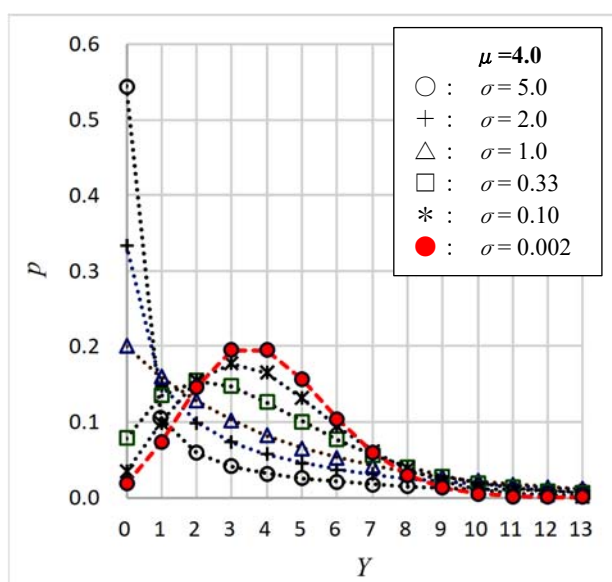


図 6.3 平均  $\mu=4.0$  に対するガンマ・ポアソン分布の形状の比較

なお，負の 2 項分布とガンマ・ポアソン分布の数学的な解説については，岩崎 (2010)，「カウントデータの統計解析」の「第 6 章 負の二項分布」に詳しく示されている．

### 6.3. 過分散の事例

過分散の事例とし、第 1.12 節の表 1.43 で「医院への通院回数」について JMP の「一変量の分布」でガンマ・ポアソン分布をあてはめた結果を示した [Cameron and Trivedi (1998)]. ここでは、Excel のソルバーを用いた最尤法によるあてはめを示す. 表 6.7 で示すように通院回数の合計は、5,190 回であり、平均は、SumProduct()関数を用いて

$$\text{平均} = \frac{\text{SumProduct}(n \text{ の範囲}, Y \text{ の範囲})}{\text{Sum}(n \text{ の範囲})} = \frac{1,566}{5,190} = 0.3017$$

として求められる. 分散は,

$$\text{分散} = \frac{\text{SumProduct}[n \text{ の範囲}, (Y \text{ の範囲} - \text{平均})^2]}{\text{Sum}(n \text{ の範囲}) - 1} = \frac{3,305.48}{5,190 - 1} = 0.6370$$

として求められる. その比は、 $0.6370 / 0.3017 = 2.11$  と過分散となっている.

表 6.7 医院への通院回数に対するポアソン分布およびガンマ・ポアソン分布のあてはめ

$i$	$Y$	度数	構成比	ポアソン分布		ガンマ・ポアソン分布		構成比との差
		$n$	$p$	$P$	$\ln L_i$	$P$	$\ln L_i$	
1	0	4,141	0.7979	0.7395	-1249.5	0.8011	-918.3	0.0032
2	1	782	0.1507	0.2231	-1173.0	0.1344	-1569.7	-0.0163
3	2	174	0.0335	0.0337	-590.1	0.0411	-555.3	0.0076
4	3	30	0.0058	0.0034	-170.6	0.0145	-127.1	0.0087
5	4	24	0.0046	0.0003	-198.5	0.0054	-125.2	0.0008
6	5	9	0.0017	0.0000	-99.7	0.0021	-55.4	0.0004
7	6	12	0.0023	0.0000	-168.8	0.0008	-85.0	-0.0015
8	7	12	0.0023	0.0000	-206.6	0.0003	-95.8	-0.0020
9	8	5	0.0010	0.0000	-102.5	0.0001	-44.4	-0.0008
10	9	1	0.0002	0.0000	-23.9	0.0001	-9.8	-0.0001
合計		5,190		$\ln L = -3983.2$		$\ln L = -3586.0$		
平均=		0.3017	$\mu^{\wedge} =$	0.3017	$\mu^{\wedge} =$	0.3017		
分散=		0.6370			$\sigma^{\wedge} =$	2.6487		
分散/平均=		2.1112			$\sigma^{\wedge} =$	1.7992	$\hat{\sigma}' = 1 + \hat{\mu}\hat{\sigma}$	

表 6.7 のポアソン分布のあてはめは、適当な  $\hat{\mu}$  を設定し、対数尤度  $\ln L$  を Excel のソルバーで最大化した結果を示す. もちろん  $\hat{\mu} = 0.3017$  は、算術平均に一致している. ガンマ・ポアソン分布は、式 (6.7) を用いて適当な  $\hat{\mu}$  と  $\hat{\sigma}$  を初期値を設定し、分布の確率  $P_i$  を計算し、対数尤度  $\ln L_i = n_i \ln P_i$  の和としての対数尤度  $\ln L$  を Excel のソルバーで最大化した結果を示す.

ガンマ・ポアソン分布の確率は、 $Y=0$  の場合

$$\begin{aligned} P_1 &= \frac{\Gamma(Y+1/\sigma)}{\Gamma(Y+1)\Gamma(1/\sigma)} \cdot \frac{(\mu\sigma)^Y}{(1+\mu\sigma)^{Y+1/\sigma}} \\ &= \frac{\Gamma(0+1/2.6487)}{\Gamma(0+1)\Gamma(1/2.6487)} \cdot \frac{(2.6487)^0}{(1+0.3017 \times 2.6487)^{0+1/2.6487}} \\ &= 1 \times \frac{1}{1.2483} = 0.8011 \end{aligned}$$

として計算される。その対数尤度  $\ln L_1$  は、

$$\begin{aligned} n L_1 &= n_1 \ln P_1 \\ &= 4,141 \times \ln(0.8011) \\ &= -918.3 \end{aligned}$$

である。全体の対数尤度  $\ln L_{GP}$  は、

$$\begin{aligned} \ln L_{GP} &= \sum_{i=1}^{10} \ln L_i \\ &= -918.3 - 1569.7 - \dots - 9.8 \\ &= -3586.0 \end{aligned}$$

となっている。ポアソン分布の場合の対数尤度  $\ln L_P$  との差  $\ln L_{diff}$  は、

$$\begin{aligned} \ln L_{diff} &= \ln L_{GP} - \ln L_P \\ &= -3586.0 - (-3983.2) = 397.2 \end{aligned}$$

と桁違いに大きい。差の2倍がおおよそ自由度1のカイに乗分布に従うことから、分布のあてはめの比較の目安として使うとすれば、明らかにガンマ・ポアソン分布のあてはめが良いと判断される。JMPの「一変量の分布」の結果は省略するが、表6.4に示したと同様に「(-2)\*対数尤度の出力があり、一致する。

ポアソン分布があてはまるかの目安に分散と平均の比をこれまで用いてきた。医院への通院回数のデータでは、分散/平均=2.112であり、ガンマ・ポアソン分布の過分散パラメータでは、 $\hat{\sigma}=2.6487$ とやや大きい。JMPでの過分散  $\hat{\sigma}'=1+\hat{\mu}\hat{\sigma}=1.7992$ であり、過分散が $\hat{\sigma}=0.0$ の場合に $\hat{\sigma}'=1+\hat{\mu}\hat{\sigma}=1.0$ となることから、分散と平均の比と比較する場合に相性が良い。

図6.4に示すように、ポアソン分布の場合には、通院回数が0回の場合に低めな推定となり、1回の場合に高めの推定となっているのに対し、ガンマ・ポアソン分布の場合には、良くあてはまっていることが読み取れる。

「医院への通院回数」は、カウント・データではあるが、ポアソン分布を仮定する稀な現象とは言い難くなっている。「医院へ通院」する回数は、病気の種類あるいは治療の内容によって一定の期間あたりの通院回数が異なる幾つかの集団から構成されていると想定すれば、必然的に過分散となりがちである。単純に分布のあてはめとするならば、過分散パラメータを持つガンマ・ポアソン分布があてはめられると言っていいのかも知れない。

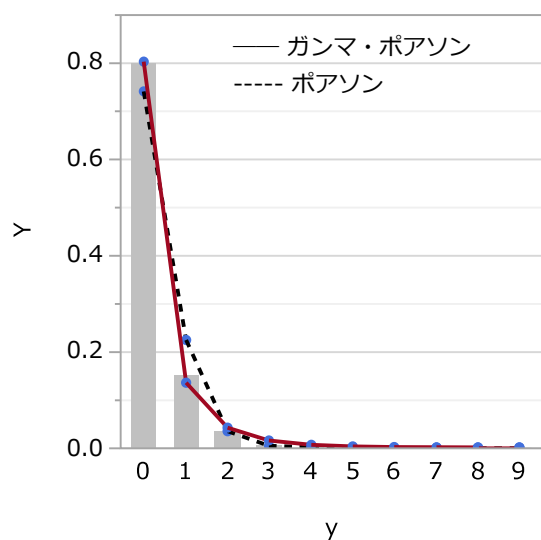


図 6.4 ガンマ・ポアソン分布とポアソン分布のあてはめ

このような得られたデータから、過分散であることが判明した場合には、数院回数に影響を与えるような未知の因子が存在している判断し、探索的な解析が必要と思うべきである。これについては、[第7章](#)を参照してもらいたい。

## 6.4. ゼロ過剰ポアソン分布のあてはめ

ゼロの反応が過大となっている場合に、ゼロとなる件数を割り引いた残りでポアソン分布をあてはめるゼロ過剰 (Zero-Inflated) モデルが、Cameron and Trivedi (1998) に示されている。SAS/GENMOD プロシジャでは、ゼロ過剰ポアソン回帰が使えるようになっているが、どのような性質なのか Excel を用いて検討する。なお、JMP の Ver. 15.1 からゼロ過剰ポアソン分布のあてはめがサポートされるようになっている。

反応が 0 の場合にポアソン分布のあてはめから除外する確率を  $\omega$  とする。反応 0 を含む残りのデータに対して、ポアソン分布の確率の和が 1 ではなく  $(1-\omega)$  となるようにパラメータ  $\omega$  および  $\mu$  を推定する。ゼロ過剰ポアソン分布の確率関数は、次のように定義される。

$$\begin{aligned} f(Y; \omega, \mu) &= \omega + (1-\omega) \text{Poisson}(Y; \mu), & Y = 0 \\ f(Y; \omega, \mu) &= (1-\omega) \text{Poisson}(Y; \mu), & Y = 1, 2, \dots \end{aligned} \quad (6.12)$$

このモデルは、 $Y$  が 0 となる部分集団、 $Y$  が 1 以上となる部分集団が混在しているような混合分布を想定することになる。

過分散データの事例として、第 1.13 節の「雌のカブトガニに結合する雄のサテライト数」のデータを引き続き用いる [アグレスティ (2003)]。表 6.8 に式 (6.12) を用いたゼロ過剰ポアソン分布、式 (6.6) を用いたガンマ・ポアソン分布のあてはめ結果を示す。表 6.8 に示すように、ゼロ過剰ポアソン分布のあてはめは、適当な初期値 ( $\hat{\mu} = 3.0000$ ,  $\hat{\omega} = 0.3000$ ) をセットし、ソルバーで対数尤度を最大化するような  $\hat{\mu}$  と  $\hat{\omega}$  を変化させ、 $\hat{\mu} = 4.4990$ ,  $\hat{\omega} = 0.3512$  を得る。

「確率  $\omega$ 」の欄は、 $y=0$  に対する  $\hat{\omega}$  の推定結果であり、「確率  $P$ 」の欄は、平均が 4.4990 のポアソン分布に  $(1-\hat{\omega})$  を掛けた結果である。従って、「確率  $P$ 」の合計は、 $(1-\hat{\omega}) = 0.6488$  となる。それぞれの  $Y_i$  に対する確率  $P_i$  と尤度  $\ln L_i$  は、

$$\begin{aligned} Y_1 = 0 : & \left[ \begin{aligned} P_1 &= \hat{\omega} + (1-\hat{\omega}) \times \text{Poisson.dist}(Y_1 = 0; \hat{\mu} = 4.4990, \text{false}) \\ &= 0.3512 + (1-0.3512) \times 0.0111 \\ &= 0.3512 + 0.0072 \\ \ln L_1 &= 62 \times \ln(0.3512 + 0.0072) \\ &= -63.6217 \end{aligned} \right. \\ Y_2 = 1 : & \left[ \begin{aligned} p_{y=1} &= (1-\hat{\omega}) \times \text{Poisson.dist}(Y = 1; \hat{\mu} = 4.4990, \text{false}) \\ &= (1-0.3512) \times 0.0500 = 0.0325 \end{aligned} \right. \end{aligned}$$

$$\left\{ \begin{array}{l} \ln L_2 = 16 \times \ln(0.0325) \\ = -54.8430 \end{array} \right.$$

:

全体の対数尤度  $\ln L_{ZP}$  は,

$$\begin{aligned} \ln L_{ZP} &= \sum_i \ln L_i \\ &= -63.6217 - 54.8430 - \dots - 10.2731 \\ &= -381.6146 \end{aligned}$$

として計算されている.

表 6.8 ゼロ過剰ポアソン分布およびガンマ・ポアソン分布のあてはめ

	初期値	$\mu^{\wedge}=$	3.0000	$\ln L_{ZP}$	$\mu^{\wedge}=$	3.0000	$\ln L_{GP}$		
		$\omega^{\wedge}=$	0.3000	-415.6585	$\sigma^{\wedge}=$	1.0000	-385.1084		
	推定結果	$\mu^{\wedge}=$	4.4990	$\ln L_{ZP}$	$\mu^{\wedge}=$	2.9191	$\ln L_{GP}$		
		$\omega^{\wedge}=$	0.3512	-381.6146	$\sigma^{\wedge}=$	1.3197	-383.7046		
	サテライト	ゼロ過剰 ポアソン				ガンマ・ポアソン			
$i$	$Y$	$n$	確率 $\omega^{\wedge}$	確率 $P^{\wedge}$	$\ln L_i$	$n^{\wedge}$	確率 $P^{\wedge}$	$\ln L_i$	$n^{\wedge}$
1	0	62	0.3512	0.0072	-63.6217	62.0	0.3021	-74.2038	52.3
2	1	16		0.0325	-54.8430	5.6	0.1818	-27.2802	31.4
3	2	9		0.0730	-23.5529	12.6	0.1268	-18.5841	21.9
4	3	19		0.1095	-42.0234	18.9	0.0926	-45.2178	16.0
5	4	19		0.1232	-39.7899	21.3	0.0690	-50.7897	11.9
6	5	15		0.1108	-32.9970	19.2	0.0522	-44.3038	9.0
7	6	13		0.0831	-32.3403	14.4	0.0397	-41.9326	6.9
8	7	4		0.0534	-11.7191	9.2	0.0305	-13.9664	5.3
9	8	6		0.0300	-21.0323	5.2	0.0234	-22.5187	4.1
10	9	3		0.0150	-12.5963	2.6	0.0181	-12.0336	3.1
11	10	3		0.0068	-14.9925	1.2	0.0140	-12.7995	2.4
12	11	1		0.0028	-5.8915	0.5	0.0109	-4.5196	1.9
13	12	1		0.0010	-6.8726	0.2	0.0085	-4.7707	1.5
14	13	0		0.0004	0.0000	0.1	0.0066	0.0000	1.1
15	14	1		0.0001	-9.0689	0.0	0.0052	-5.2686	0.9
16	15	1		0.0000	-10.2731	0.0	0.0040	-5.5156	0.7
	計	173	0.3512	0.6488		173.0	0.9854		170.5
			1.0000						

表 6.8 の右側のガンマ・ポアソン分布のあてはめは、式 (6.6) により確率を計算し、それぞれの対数尤度から、全体の対数尤度が計算されている。もちろん、初期値 ( $\hat{\mu}=3.0000$ ,  $\hat{\sigma}=1.0000$ ) からソルバーで、全体の対数尤度  $\ln L$  を最大化した結果である。

$$\text{確率} \quad \hat{p}_i = \frac{\Gamma(Y_i + 1 / \hat{\sigma})}{\Gamma(Y_i + 1) \Gamma(1 / \hat{\sigma})} \cdot \frac{(\hat{\mu}_i \hat{\sigma})^{Y_i}}{(1 + \hat{\mu}_i \hat{\sigma})^{Y_i + 1 / \hat{\sigma}}}$$

$$\text{対数尤度} \quad \ln L_i = n_i \ln(\hat{p}_i)$$

$$\begin{aligned}
\ln L_{GP} &= \sum_{i=1}^{16} \ln L_i \\
&= -74.2038 - 27.2802 - \dots - 5.5156 \\
&= -383.7046
\end{aligned}$$

図 6.5 にデータ数  $N=173$  のそれぞれの分布の確率  $Y$  を掛けて推定件数  $\hat{n}_i = N\hat{P}_i$  を算出し、元のツガイ数  $n_i$  のヒストグラムの上に上書きした結果を示す。ゼロ過剰ポアソン分布のあてはまりが、ガンマ・ポアソン分布よりもサテライト数が 1 および 2 に対して推定値が追従している。

対数尤度の比較でも、ゼロ過剰ポアソン分布の  $\ln L_{ZP}$  とガンマ・ポアソン分布の  $\ln L_{GP}$  の比較では、

$$\begin{aligned}
\ln L_{diff} &= \ln L_{ZP} - \ln L_{GP} \\
&= -381.6146 - (-383.7046) \\
&= 2.0900
\end{aligned}$$

と、わずかではあるが、ゼロ過剰ポアソン分布のあてはめがほんの少し良い。

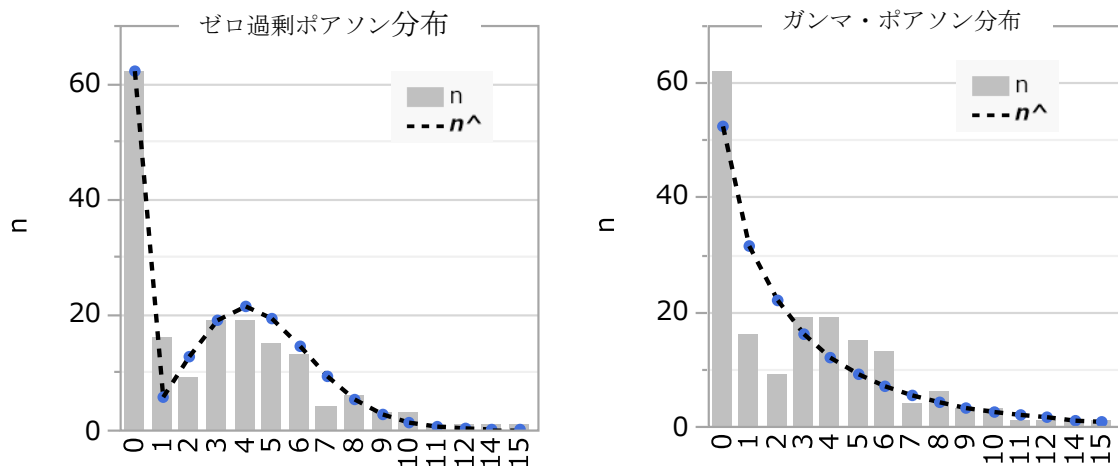


図 6.5 Zero-Inflated ポアソン分布およびガンマ・ポアソン分布のあてはめ

JMP (Ver 15.1) から、ゼロ過剰ポアソン分布のあてはめができるようになったので、表 6.9 に結果を示す。推定値は、 $(\lambda: \hat{\mu} = 4.4990, \pi: \hat{\omega} = 0.3512)$  と、表 6.8 の結果に一致することが確認される。

表 6.9 JMP によるゼロ過剰ポアソン分布のあてはめ

ゼロ強調 Poisson分布のあてはめ					
パラメータ		推定値	標準誤差	下側95%	上側95%
平均	$\lambda$	4.4990	0.2055	4.1074	4.9129
ゼロ強調	$\pi$	0.3512	0.0369	0.2813	0.4252
指標					
(-2)*対数尤度		763.2292			
AICc		767.2998			
BIC		773.5357			

雄の結合がない雌の数に対して、ポアソン分布のあてはめから除外する雌の数は  $\hat{\omega} = 0.3512$  と推定されているので、

$$N\hat{\omega} = 173 \times 0.3512 = 60.75$$

と、62 匹中の大部分となる。このことから、雄が雌に結合しやすい場合と、結合しにくい何らかの理由があると推論される。図 6.6 に示すように、サテライト数がゼロを除いた場合に、平均=4.5494、分散=8.0134 と分散が平均の 2 倍弱であり、過分散となっており、適合度の検定でもポアソン分布のあてはめは支持されない。



図 6.6 ゼロを除いた場合のポアソン分布のあてはめ

## 6.5. ゼロ過剰ガンマ・ポアソン分布のあてはめ

ゼロの反応が過大となっている場合に、ゼロとなる件数を割り引いた残りにガンマ・ポアソン分布をあてはめる。全体の数を  $N$  としたときに、反応が 0 の場合にガンマ・ポアソン分布のあてはめをしない割合を  $\omega$  とする。残りの  $(1-\omega)$  にガンマ・ポアソン分布をあてはめる。反応が 0 の場合の確率は、 $\omega$  にガンマ・ポアソン分布の確率に  $(1-\omega)$  を掛けたものを加える。

$$\left. \begin{aligned} f(Y; \omega, \mu, \sigma) &= \omega + (1-\omega) \text{GammaPoisson}(Y; \mu, \sigma), & y = 0 \\ f(Y; \omega, \mu, \sigma) &= (1-\omega) \text{GammaPoisson}(Y; \mu, \sigma), & y = 1, 2, \dots \end{aligned} \right] \quad (6.13)$$

$$\text{GammaPoisson}(Y; \mu, \sigma) = \frac{\Gamma(Y+1/\sigma)}{\Gamma(Y+1)\Gamma(1/\sigma)} \frac{(\mu\sigma)^Y}{(1+\mu\sigma)^{Y+1/\sigma}}$$

このモデルは、ゼロ過剰ポアソン分布と同様に、 $Y$  が 0 となる部分集団、 $Y$  が 1 以上となる部分集団が混在しているような混合分布を想定することになる。表 6.10 にゼロ過剰ガンマ・ポアソン分布をあてはめた結果を示す。

表 6.10 最尤法によるゼロ過剰ガンマ・ポアソン分布のあてはめ

	初期値	4.0000	$\mu^{\wedge} =$	4.3287		
		1.0000	$\sigma^{\wedge} =$	0.2242	$\ln L$	
		0.4000	$\omega^{\wedge} =$	0.3256	-369.3516	
	サテライト数	ツガイ数	ゼロ過剰ガンマ・ポアソン			
$i$	$Y$	$n$	確率 $\omega^{\wedge}$	確率 $P^{\wedge}$	$\ln L_i$	$n^{\wedge}$
1	0	62	0.3256	0.0327	-63.6218	62.0
2	1	16		0.0719	-42.1183	12.4
3	2	9		0.0967	-21.0263	16.7
4	3	19		0.1025	-43.2710	17.7
5	4	19		0.0942	-44.8846	16.3
6	5	15		0.0785	-38.1695	13.6
7	6	13		0.0610	-36.3678	10.5
8	7	4		0.0449	-12.4163	7.8
9	8	6		0.0317	-20.7172	5.5
10	9	3		0.0216	-11.5073	3.7
11	10	3		0.0143	-12.7406	2.5
12	11	1		0.0093	-4.6816	1.6
13	12	1		0.0059	-5.1365	1.0
14	13	0		0.0037	0.0000	0.6
15	14	1		0.0023	-6.0961	0.4
16	15	1		0.0014	-6.5967	0.2
	計	173	0.3256	0.6724		172.7
			0.9980			

ゼロ過剰ポアソン分布をあてはめた場合に比べ、 $\hat{\omega}=0.3256$ とやや小さ目になり、 $Y_1=0$ に対する推定値 $\hat{n}_1$ は、

$$\hat{n}_1 = N\hat{\omega} = 173 \times 0.3256 = 56.34$$

となる。残りの  $173 \times (1 - 0.3256) = 116.66$  ツガイ数に対するガンマ・ポアソン分布のあてはめは、平均が  $\mu = 4.3287$ 、過分散が  $\sigma = 0.2243$  となり、全体的にあてはまりがよくなっている。

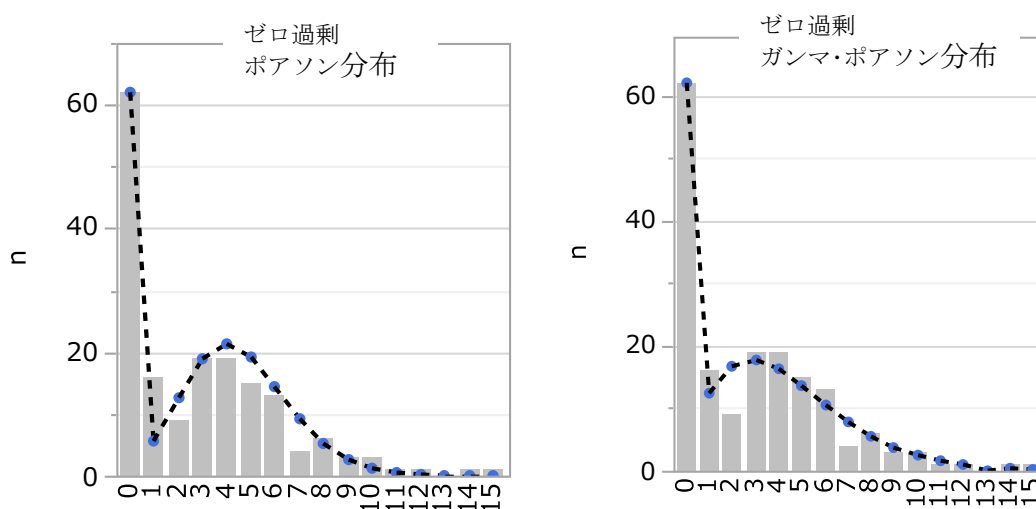


図 6.7 ゼロ過剰(ポアソン分布およびガンマ・ポアソン分布)のあてはめ

これまでに検討した結果を、対数尤度の大きさによってモデルのあてはめについて評価する。基本は、ポアソン分布のあてはめで、表 1.46 に示すように、JMP の「一変量の分布」によりサテライト数の平均は 2.9191、分散は 9.2129 であり、その比は 3.40 と過分散になっている。ポアソン分布をあてはめ、棒グラフ上に上書きした結果を見ても、誤差分布にポアソン分布を仮定することは絶望的である。もちろん、適合度の検定でもポアソン分布であるとは言えない。

過分散となる場合は、何らかの条件によりサテライト数の平均が大きく異なる部分集団の集まりが複数存在すると推測される。甲羅の色によってサテライト数の平均が大きく異なり、それに伴って過分散が解消されるのであろうか。あるいは、甲羅の幅、あるいは、体重を幾つかに区分したときにサテライト数の平均が大きく異なり、過分散が解消されるのだろうか。検討すべき事項である。表 1.46 に JMP の出力でマイナス 2 倍の対数尤度が 988.0893 である。次に、ガンマ・ポアソン分布で、表 6.8 から、対数尤度が -383.7046 である。ゼロ過剰ポアソン分布のあてはめは、対数尤度が -381.6146 であり、ゼロ過剰ガンマ・ポアソン分布のあてはめは、表 6.10 に対数尤度が -369.3516 とある。これらを表 6.11 にまとめ直す。

表 6.11 4 種のモデルのマイナス 2 倍の対数尤度の比較

		$\ln L$	$(-2)\ln L$	ポアソン vs. ガンマ・ポアソン	ゼロ過剰 なし・あり
ゼロ過剰	分布タイプ				
なし	ポアソン分布	-494.0447	<b>988.0893</b>	基準	基準
	ガンマ・ポアソン分布	-383.7046	767.4092	220.6801	
あり	ゼロ過剰 ポアソン分布	-381.6146	<b>763.2292</b>	基準	224.8602
	ゼロ過剰 ガンマ・ポアソン分布	-369.3516	738.7032	24.5260	28.7061

マイナス 2 倍の対数尤度の差について、自由度 1 のカイ 2 乗分布の 5 パーセント点 3.84 をおおよその目安として用いる。ポアソン分布 vs. ガンマ・ポアソン分布のマイナス 2 倍の対数尤度の差は、220.6801 であり、圧倒的にガンマ・ポアソン分布のあてはめが支持される。

ゼロ過剰の場合のポアソン分布 vs. ガンマ・ポアソン分布の差は、24.5360 あり、ガンマ・ポアソン分布のあてはめがかなり改善している。ポアソン分布でゼロ過剰にした場合には、マイナス 2 倍の対数尤度の差は、224.8602 とゼロ過剰にした場合が圧倒的に支持される。ガンマ・ポアソン分布でゼロ過剰にした場合には、マイナス 2 倍の対数尤度の差は、28.7061 とゼロ過剰にした場合にあてはめがかなり改善される。

この結果から、ゼロ過剰ガンマ・ポアソン分布が、最もあてはまりがいいとの結論である。ただし、この分布を使った回帰分析を実施することは、可能ではあるが、ゼロの割合が、説明変数の値により変化するような場合に、ゼロがないにも関わらず確率的には存在するというような矛盾が内在し、結果の解釈に困難さを感じる。

JMP (Ver 15.1) から、ゼロ強調負の 2 項分布（ゼロ過剰ガンマ・ポアソン分布）のあてはめができるようになったので、表 6.12 に結果を示す。推定値は、( $\lambda: \hat{\mu}=4.3287$ ,  $\hat{\sigma}=0.2242$   $\pi: \hat{\omega}=0.3257$ ) と、表 6.10 に一致することが確認される。注)  $\pi: \hat{\omega}$  は少数点以下 4 桁で 0.0001 の食い違いがあるが、正確には、 $\pi=0.325650$ ,  $\hat{\omega}=0.325649$  あり、少数点以下 6 桁での 0.000001 の違いである。

表 6.12 JMP によるゼロ強調・負の 2 項分布のあてはめ

ゼロ強調 負の二項分布のあてはめ					
パラメータ		推定値	標準誤差	下側95%	上側95%
平均	$\lambda$	4.3287	0.2958	3.7497	4.9289
ばらつき	$\sigma$	0.2242	0.0783	0.1039	0.4317
ゼロ強調	$\pi$	0.3257	0.0402	0.2464	0.4050
指標					
(-2)*対数尤度		738.7032			
AICc		744.8452			
BIC		754.1630			

## 6.6. ガンマ・ポアソン回帰

第 1.13 節では、雌のカブトカニに連結する雄のサテライト数のデータにポアソン分布をあてはめ、過分散データの場合にどのような不具合が起きるかを示した[アグレスティ (2003)]. サテライト数を反応変数、甲羅の幅を説明変数とする対数リンクによるポアソン回帰を行い、過分散があるデータの場合には、個別データの 95%信頼区間が極端に過小評価になることを示した.

### ポアソン回帰 vs. ガンマ・ポアソン回帰

第 1.13 節の表 1.50 で、ガンマ・ポアソン分布のあてはめを行なった結果、見た目にはあてはまりが悪いと思われた. ただし、適合度検定ではあてはめは否定されなかった. この様に過分散データに対して、ガンマ・ポアソン分布を誤差分布とする回帰分析を行った場合、どのような結果になるのでしょうか. あるいは、どのような注意点があるのだろうか. そこで、Excel による対数リンクによるガンマ・ポアソン回帰を行い、ポアソン回帰の場合と比較検討することにした. 表 6.13 に結果を示す.

表 6.13 甲羅の幅を説明変数とするガンマ・ポアソン回帰

			$\beta_0^{\wedge} =$	<b>-3.3048</b>		$\beta_0^{\wedge} =$	<b>-4.0523</b>	
			$\beta_1^{\wedge} =$	<b>0.1640</b>	$\ln L_P$	$\beta_1^{\wedge} =$	<b>0.1921</b>	$\ln L_{GP}$
					<b>-461.59</b>	過分散 $\sigma^{\wedge} =$	<b>1.1055</b>	<b>-375.65</b>
			ポアソン回帰			ガンマ・ポアソン回帰		
	幅	サテ	対数	確率	対数尤度	対数	確率	対数尤度
No	$x$	ライト $y$	リンク $y^{\wedge}$	$P^{\wedge}$	$\ln L_i$	リンク $y^{\wedge}$	$P^{\wedge}$	$\ln L_i$
1	28.3	8	3.8103	0.0244	-3.7132	3.9874	0.0324	-3.4291
2	22.5	0	1.4715	0.2296	-1.4715	1.3089	0.4451	-0.8094
3	26.0	9	2.6128	0.0011	-6.7709	2.5636	0.0148	-4.2130
4	24.8	0	2.1459	0.1170	-2.1459	2.0358	0.3443	-1.0663
5	26.0	4	2.6128	0.1424	-1.9492	2.5636	0.0721	-2.6302
:								
171	28.0	0	3.6274	0.0266	-3.6274	3.7642	0.2266	-1.4846
172	27.0	0	3.0786	0.0460	-3.0786	3.1064	0.2600	-1.3472
173	24.5	0	2.0429	0.1297	-2.0429	1.9218	0.3568	-1.0306

ポアソン回帰の場合は、

$$\begin{aligned}
 \text{対数リンク} & \quad \hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i) \\
 \text{確率} & \quad \hat{P}_i = \text{Poisson.dist}(Y_i, \hat{y}_i, \text{false}) \\
 \text{対数尤度} & \quad \ln L_i = \ln(\hat{P}_i) \\
 \ln L & \quad \ln L_P = \sum_{i=1}^{173} \ln L_i
 \end{aligned}$$

とし、 $\hat{\beta}_0$  と  $\hat{\beta}_1$  に  $\ln L_i$  が計算されるような初期値を設定し、ソルバーで、 $\ln L_i$  の和である  $\ln L_p$  が最大になるように  $\hat{\beta}_0$  と  $\hat{\beta}_1$  を変化させた結果である。推定された  $\hat{\beta}_0 = -3.3048$  および  $\hat{\beta}_1 = 0.1640$  が最尤解として求まっている。これは、表 1.47 に示した JMP による結果と一致している。

次に、ガンマ・ポアソン回帰の場合は、式 (6.6) のガンマ・ポアソン分布関数を用いて、

$$\text{対数リンク} \quad \hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$\text{確率} \quad \hat{P}_i = \frac{\Gamma(y_i + 1/\hat{\sigma})}{\Gamma(y_i + 1)\Gamma(1/\hat{\sigma})} \cdot \frac{(\hat{y}_i \hat{\sigma})^{y_i}}{(1 + \hat{y}_i \hat{\sigma})^{y_i + 1/\hat{\sigma}}}$$

$$\text{対数尤度} \quad \ln L_i = \ln(\hat{P}_i)$$

$$\ln L \quad \ln L_{GP} = \sum_{i=1}^{173} \ln L_i$$

とし、 $\hat{\beta}_0$  と  $\hat{\beta}_1$  に  $\ln L_i$  が計算されるような初期値を入れ、ソルバーで  $\ln L_{GP}$  が最大になるように  $\hat{\beta}_0$ 、 $\hat{\beta}_1$  および  $\hat{\sigma}$  を変化させた結果である。推定された  $\hat{\beta}_0 = -4.0523$ 、 $\hat{\beta}_1 = 0.1921$  および  $\hat{\sigma} = 1.1055$  が最尤解として求まっている。

ポアソン回帰の対数尤度  $\ln L_p = -461.58$  に比べ、ガンマ・ポアソン回帰の対数尤度は、 $\ln L_{GP} = -375.65$  と大きくなり、対数尤度の差は、

$$\begin{aligned} \ln L_{diff} &= \ln L_{GP} - \ln L_p \\ &= -375.65 - (-461.58) = 85.94 \end{aligned}$$

と大きく、ガンマ・ポアソン回帰のあてはめが支持される。また、ガンマ・ポアソン回帰場合に、切片  $\hat{\beta}_0$  は小さくなり、傾きは大きくなっている。図 6.8 の甲羅の幅とサテライト数の散布図上に、対数リンクの回帰直線（指数曲線）を上書きした結果を示す。

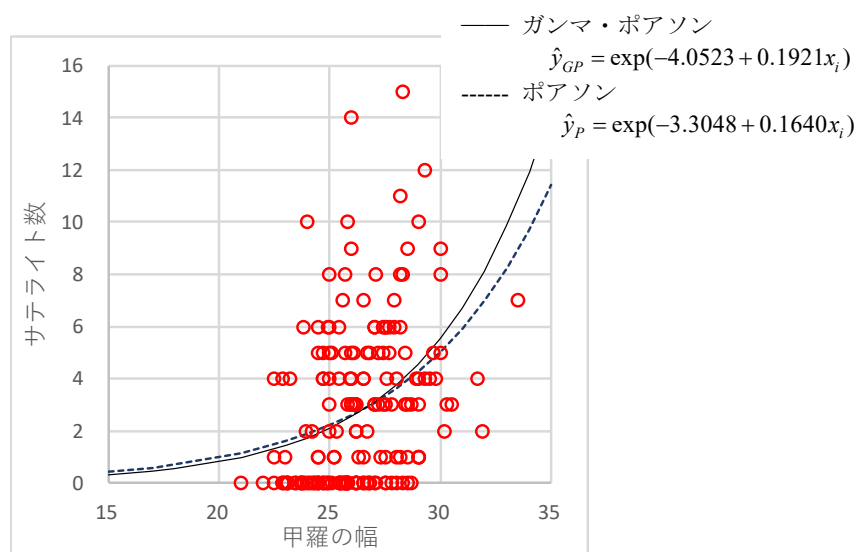


図 6.8 対数リンクでのガンマ・ポアソン回帰

甲羅の幅に対するポアソン回帰およびガンマ・ポアソン回帰の指数曲線の推定値を表 6.14 に示す。ポアソン回帰の切片は,

$$\hat{y}_{GP(x=0)} = \exp(-3.3047 + 0.1640 \times 0) = 0.0367$$

であり, ガンマ・ポアソン回帰の切片は,

$$\hat{y}_{P(x=0)} = \exp(-4.0524 + 0.1921 \times 0) = 0.0174$$

と低めに推定されている。

表 6.14 ポアソン回帰およびガンマ・ポアソン回帰の推定値

$x$	ポアソン	ガンマ・ポアソン	
	$y_P^{\wedge}$	$y_{GP}^{\wedge}$	$1 + y_{GP}^{\wedge} \sigma^{\wedge}$
0	0.0367	0.0174	1.0192
10	0.1893	0.1186	1.1311
20	0.9764	0.8098	1.8952
25	2.2175	2.1156	3.3388
30	5.0359	5.5271	7.1102
35	11.4366	14.4401	16.9636

甲羅の幅  $x = 25$  の場合には, ポアソン回帰の場合は,

$$\hat{y}_{P(x=25)} = \exp(-3.3047 + 0.1640 \times 25) = 2.2175$$

であり, ガンマ・ポアソン回帰の場合は,

$$\hat{y}_{GP(x=25)} = \exp(-4.0524 + 0.1921 \times 25) = 2.1156$$

と低めであるが,  $x = 30$  の場合には, ガンマ・ポアソン回帰の方が 14.4401 と高めとなっている。

ポアソン回帰の分散は, 回帰の推定値の伸びと同値で大きくなる。ガンマ・ポアソン回帰の場合に過分散  $\sigma$  は, 一定で変化しないので, JMP の過分散の定義式

$$\hat{\sigma}' = 1 + \hat{y}_{GP} \hat{\sigma}$$

を用いて計算した結果を示す。

#### 甲羅の幅 $x$ に対するガンマ・ポアソン分布のあてはめ

ガンマ・ポアソン回帰は, 甲羅の幅  $x$  の変化に対し, どのような分布をあてはめているのだろうか。表 6.15 に  $x = (20, 25, 30)$  の場合の推定値  $\hat{y}_j = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_j)$  に対するポアソン分布の確率

$$\hat{P}_i = \text{Poisson.dist}(Y_i, \hat{y}_j, \text{false})$$

を計算した結果を示す。同様にガンマ・ポアソン回帰の推定値  $\hat{y}_j = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_j)$  に対するガ

ンマ・ポアソン分布の確率

$$\hat{P}_i = \frac{\Gamma(Y_i + 1 / \hat{\sigma})}{\Gamma(Y_i + 1) \Gamma(1 / \hat{\sigma})} \cdot \frac{(\hat{y}_j \hat{\sigma})^{Y_i}}{(1 + \hat{y}_j \hat{\sigma})^{Y_i + 1 / \hat{\sigma}}}$$

を計算した結果も示す. 甲羅の幅  $x=25$  の場合では, ポアソン分布の  $y=0$  場合は確率が 0.1089 であるのに対し, ガンマ・ポアソン分布の確率は 0.3360 とかなり上回っている.

表 6.15 ポアソン回帰およびガンマ・ポアソン回帰の誤差の確率分布

			$\beta^{\wedge}_0=$	<b>-3.3047</b>		$\beta^{\wedge}_0=$	<b>-4.0524</b>	
			$\beta^{\wedge}_1=$	<b>0.1640</b>		$\beta^{\wedge}_1=$	<b>0.1921</b>	
						過分散 $\sigma^{\wedge}=$	<b>1.1055</b>	
			ポアソン回帰			ガンマ・ポアソン回帰		
		$x=$	20	25	30	20	25	30
$j$	$y^{\wedge}=$		0.9764	2.2175	5.0359	0.8098	2.1156	5.5271
1	0		0.3767	0.1089	0.0065	0.5608	0.3360	0.1696
2	1		0.3678	0.2414	0.0327	0.2396	0.2129	0.1318
3	2		0.1796	0.2677	0.0824	0.1078	0.1420	0.1079
4	3		0.0584	0.1979	0.1384	0.0493	0.0963	0.0898
5	4		0.0143	0.1097	0.1742	0.0227	0.0659	0.0753
6	5		0.0028	0.0486	0.1754	0.0105	0.0453	0.0635
7	6		0.0005	0.0180	0.1473	0.0049	0.0312	0.0537
8	7		0.0001	0.0057	0.1059	0.0023	0.0216	0.0455
9	8		0.0000	0.0016	0.0667	0.0011	0.0149	0.0386
10	9		0.0000	0.0004	0.0373	0.0005	0.0103	0.0328
11	10		0.0000	0.0001	0.0188	0.0002	0.0072	0.0280
12	11		0.0000	0.0000	0.0086	0.0001	0.0050	0.0238
13	12		0.0000	0.0000	0.0036	0.0001	0.0035	0.0203
14	13		0.0000	0.0000	0.0014	0.0000	0.0024	0.0173
15	14		0.0000	0.0000	0.0005	0.0000	0.0017	0.0148
16	15		0.0000	0.0000	0.0002	0.0000	0.0012	0.0126
17	16		0.0000	0.0000	0.0001	0.0000	0.0008	0.0108

表 6.15 に示したポアソン回帰およびガンマ・ポアソン回帰の甲羅の幅  $x=(20, 25, 30)$  に対するサテライト数の分布の確率を図 6.9 に示す. 甲羅の幅が 20 cm の場合は, 片流れ的であ

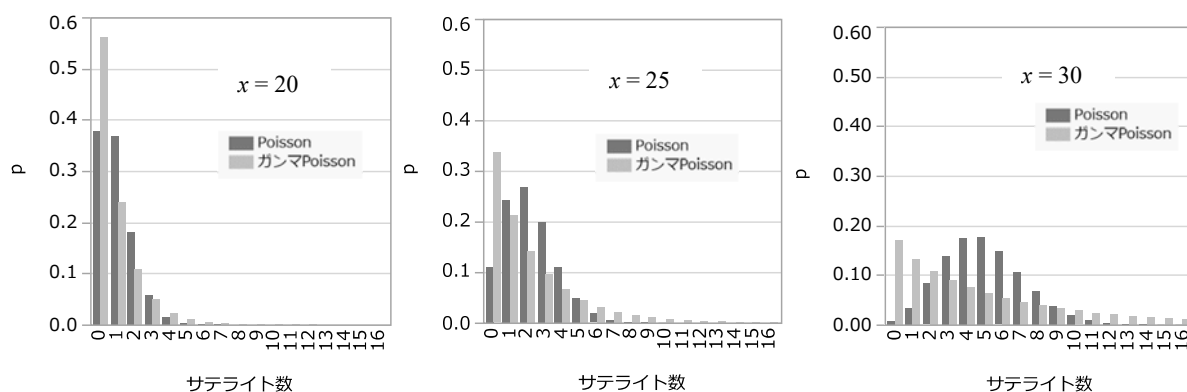


図 6.9 対数リンクでの確率分布

るのが，甲羅の幅が大きくなるにつれて，ポアソン回帰の場合は，一山型であるが，ガンマ・ポアソン回帰の場合は，片流れが続いている．

対数リンクによるガンマ・ポアソン回帰の指数曲線は，図 6.8 に示したのであるが， $x=(20, 25, 30)$  に対して表 6.15 で計算したガンマ・ポアソン分布の確率を重ね書きした結果を図 6.10 に示す．

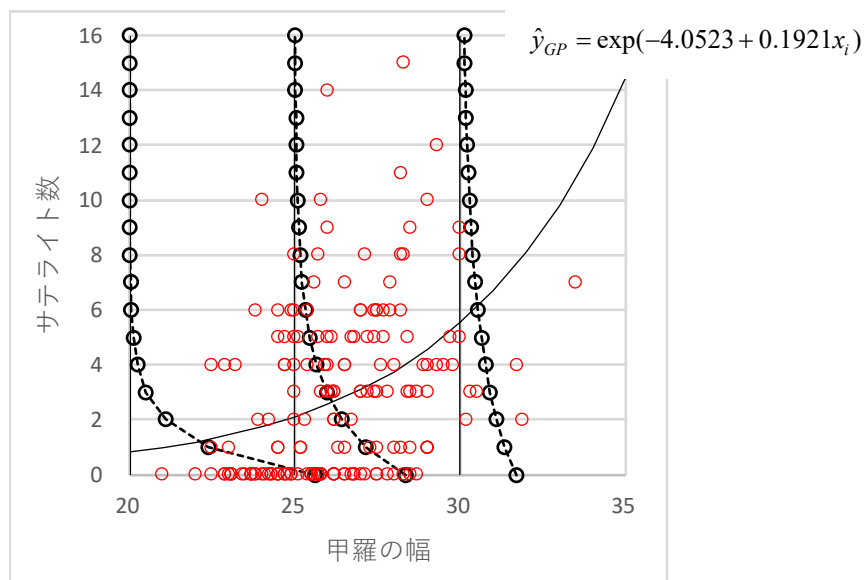


図 6.10 対数リンクでのガンマ・ポアソン回帰における確率分布

甲羅の幅  $x=25$  の場合にサテライト数  $y$  に対する出現確率は，○印の分布を適切に反映していると判断される．ただし，甲羅の幅  $x=30$  では，サテライト数  $y=0$  が観察されないにもかかわらず出現確率が高止まりしている．このことは，甲羅の幅に関わらずサテライト数の分布が，一律にガンマ・ポアソン分布に従っていると仮定することが妥当ではないとの判断になる．

## 6.7. ゼロ過剰ガンマ・ポアソン回帰

第 6.6 節で示したようにガンマ・ポアソン回帰に引き続き、ゼロ過剰ガンマ・ポアソン回帰の振る舞いについて示す。第 6.3 節で示したように稀な現象ではないカウント・データでは、ゼロ過剰ガンマ・ポアソンのあてはめが支持されていた。雌のカブトカニに連結する雄のサテライト数のデータに対し、甲羅の幅を説明変数とする対数リンクによる、ゼロ過剰ガンマ・ポアソン回帰を適用し、その振る舞いについて検討する。

### ガンマ・ポアソン回帰 vs. ゼロ過剰ガンマ・ポアソン回帰

表 6.16 にゼロ過剰ガンマ・ポアソン回帰の結果を示す。対数リンクは、

$$\text{対数リンク} \quad \hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

で計算し、サテライト数  $y_i$  がゼロの場合は、

$$\begin{aligned} \text{ゼロ対応} \quad Y_i = 0 : \quad & \hat{P}_{0i} = \hat{\omega} \\ Y_i \neq 0 : \quad & \hat{P}_{0i} = 0 \end{aligned}$$

とし、全てのサテライト数  $y_i$  について

$$\text{確率} \quad \hat{P}_i = (1 - \hat{\omega}) \cdot \frac{\Gamma(Y_i + 1/\hat{\sigma})}{\Gamma(Y_i + 1)\Gamma(1/\hat{\sigma})} \cdot \frac{(\hat{\mu}_i \hat{\sigma})^{Y_i}}{(1 + \hat{\mu}_i \hat{\sigma})^{Y_i + 1/\hat{\sigma}}}$$

$$\text{対数尤度} \quad \ln L_i = \ln(\hat{P}_{0i} + \hat{P}_i)$$

$$\ln L \quad \ln L_{ZGP} = \sum_{i=1}^{173} \ln L_i$$

表 6.16 甲羅の幅を説明変数とするゼロ過剰ガンマ・ポアソン回帰

							$\omega =$	<b>0.3099</b>	
			$\beta^{\wedge}_0 =$	<b>-4.0523</b>			$\beta^{\wedge}_0 =$	<b>-0.3638</b>	
			$\beta^{\wedge}_1 =$	<b>0.1921</b>	$\ln L_{GP}$		$\beta^{\wedge}_1 =$	<b>0.0671</b>	$\ln L_{ZGP}$
			過分散 $\sigma^{\wedge} =$	<b>1.1055</b>	<b>-375.65</b>		過分散 $\sigma^{\wedge} =$	<b>0.2442</b>	<b>-367.49</b>
			ガンマ・ポアソン回帰			ゼロ過剰ガンマ・ポアソン回帰			
	幅	サテラ	対数	確率	対数尤度	対数	ゼロ	確率	対数尤度
No	$x$	イト $Y$	リンク $y^{\wedge}$	$P$ 値	$\ln L_i$	リンク $y^{\wedge}$	$P_0$	$P$	$\ln L_i$
1	28.3	8	3.9874	0.0324	-3.4291	4.6389	0.0000	0.0363	-3.3168
2	22.5	0	1.3089	0.4451	-0.8094	3.1438	0.3099	0.0669	-0.9758
3	26.0	9	2.5636	0.0148	-4.2130	3.9757	0.0000	0.0182	-4.0083
4	24.8	0	2.0358	0.3443	-1.0663	3.6682	0.3099	0.0503	-1.0210
5	26.0	4	2.5636	0.0721	-2.6302	3.9757	0.0000	0.0949	-2.3547
:									
170	29.0	4	4.5612	0.0776	-2.5565	4.8619	0.0000	0.0913	-2.3931
171	28.0	0	3.7642	0.2266	-1.4846	4.5465	0.3099	0.0324	-1.0719
172	27.0	0	3.1064	0.2600	-1.3472	4.2515	0.3099	0.0374	-1.0575
173	24.5	0	1.9219	0.3568	-1.0306	3.5952	0.3099	0.0523	-1.0155

とし、 $(\hat{\omega}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$  に  $\ln L_i$  が計算されるような初期値を入れ、ソルバーで  $\ln L_{ZGP}$  が最大になるように  $(\hat{\omega}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$  を変化させた結果である。推定された  $\hat{\omega}=0.3099$   $\hat{\beta}_0=-0.3638$ ,  $\hat{\beta}_1=0.0671$  および  $\hat{\sigma}=0.2442$  が最尤解として求まっている。

ガンマ・ポアソン回帰の対数尤度  $\ln L_{GP}=-375.65$  に比べ、ゼロ過剰ガンマ・ポアソン回帰の対数尤度は、 $\ln L_{ZGP}=-367.49$  と大きくなり、対数尤度の差は、

$$\begin{aligned}\ln L_{diff} &= \ln L_{ZGP} - \ln L_{GP} \\ &= -367.49 - (-375.65) = 8.15\end{aligned}$$

と大きく、ゼロ過剰ガンマ・ポアソン回帰のあてはめが支持される。また、ガンマ・ポアソン回帰場合に、切片  $\hat{\beta}_0$  は大きくなり、傾き  $\hat{\beta}_1$  は小さく、過分散  $\hat{\sigma}=0.2442$  とかなり小さくなっている。図 6.11 に甲羅の幅とサテライト数の散布図上に、対数リンクの回帰直線（指数曲線）を上書きした結果を示す。

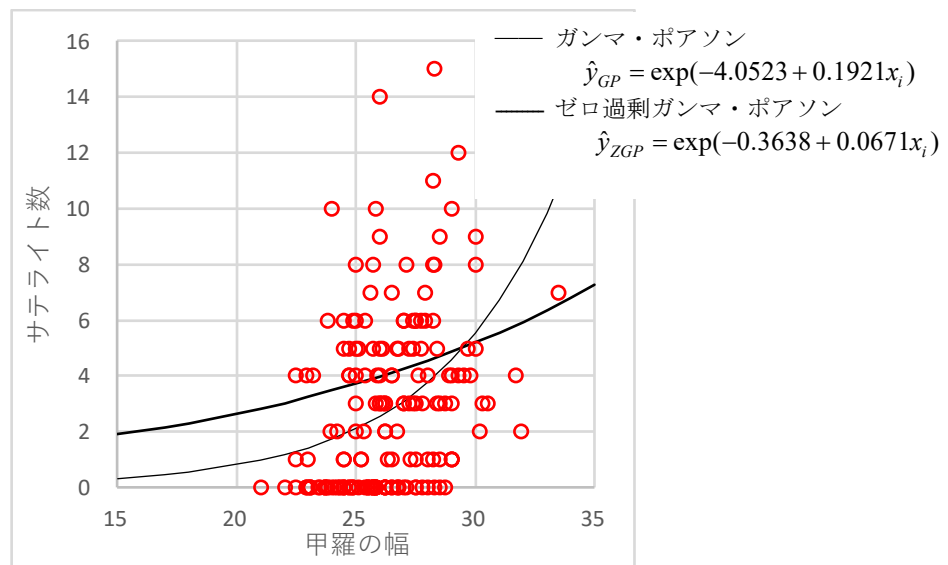


図 6.11 対数リンクでのゼロ過剰ガンマ・ポアソン回帰

甲羅の幅  $x$  に対するガンマ・ポアソン回帰およびゼロ過剰ガンマ・ポアソン回帰の指数曲線を次式で推定する。ガンマ・ポアソン回帰の切片は、

$$\hat{y}_{GP(x=0)} = \exp(-4.0524 + 0.1921 \times 0) = 0.0174$$

であり、ゼロ過剰ガンマ・ポアソン回帰の切片は、

$$\hat{y}_{ZGP(x=0)} = \exp(-0.3638 + 0.0671 \times 0) = 0.6950$$

と高めに推定されている。

甲羅の幅  $x=25$  の場合には、ガンマ・ポアソン回帰の場合は、

$$\hat{y}_{GP(x=25)} = \exp(-4.0524 + 0.1921 \times 25) = 2.1156$$

であり、ガンマ・ポアソン回帰の場合は、

$$\hat{y}_{ZGP(x=25)} = \exp(-0.3638 + 0.0671 \times 25) = 3.7178$$

と低めであるが、 $x=30$  の場合には、ゼロ過剰ガンマ・ポアソン回帰の方が低めと逆転している。

### 甲羅の幅 $x$ に対するゼロ過剰ガンマ・ポアソン分布のあてはめ

ゼロ過剰ガンマ・ポアソン回帰は、甲羅の幅  $x$  の変化に対し、どのような分布をあてはめているのか図 6.12 に  $x=(20, 25, 30)$  に対応するゼロ過剰ガンマ・ポアソン回帰の推定値  $\hat{\mu}_j = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_j)$  に対するゼロ過剰ガンマ・ポアソン分布の確率  $\hat{P}$  を  $Y_i, i=0,1,...,16$  について

$$y_0 = 0 \quad \hat{P}_i = \hat{\omega} + (1 - \hat{\omega}) \cdot \frac{\Gamma(Y_i + 1/\hat{\sigma})}{\Gamma(Y_i + 1)\Gamma(1/\hat{\sigma})} \cdot \frac{(\hat{\mu}_j \hat{\sigma})^{Y_i}}{(1 + \hat{\mu}_j \hat{\sigma})^{Y_i + 1/\hat{\sigma}}}$$

$$y_0 \neq 0 \quad \hat{P}_i = (1 - \hat{\omega}) \cdot \frac{\Gamma(Y_i + 1/\hat{\sigma})}{\Gamma(Y_i + 1)\Gamma(1/\hat{\sigma})} \cdot \frac{(\hat{\mu}_j \hat{\sigma})^{Y_i}}{(1 + \hat{\mu}_j \hat{\sigma})^{Y_i + 1/\hat{\sigma}}}$$

計算し（表 6.17）、図 6.12 にプロットとした結果を示す。

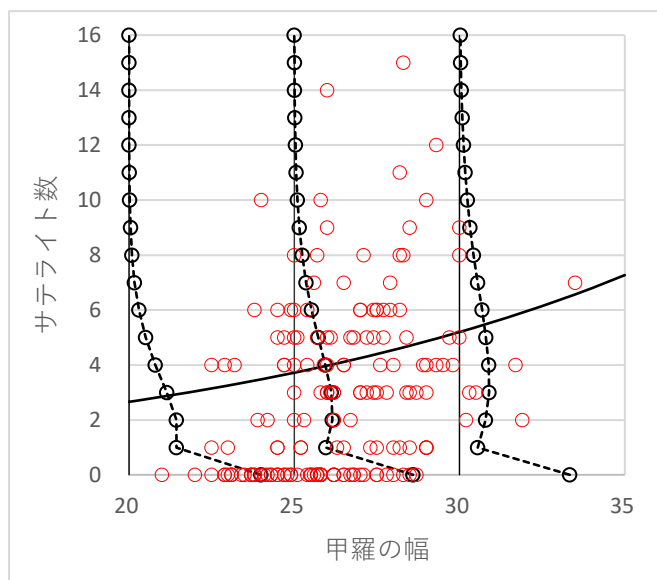


図 6.12 対数リンクでのゼロ過剰ガンマ・ポアソン回帰における確率分布

甲羅の幅  $x=25$  の場合にサテライト数  $y$  に対する出現確率は、○印の分布を適切に反映していると判断される。ただし、甲羅の幅  $x=30$  では、サテライト数  $Y=0$  が観察されないにもかかわらず出現確率が高止まりしている。このことは、甲羅の幅に関わらずサテライト数の分布が、一律にゼロ過剰ガンマ・ポアソン分布に従っていると仮定することが妥当ではないとの判断になる。

表 6.17 ガンマ・ポアソン回帰およびゼロ過剰ガンマ・ポアソン回帰の誤差の確率分布

						$\omega=$	0.3099		
			$\beta^{\wedge}_0=$	-4.0524		$\beta^{\wedge}_0=$	-0.3638		
			$\beta^{\wedge}_1=$	0.1921		$\beta^{\wedge}_1=$	0.0671		
			過分散 $\sigma^{\wedge}=$	1.1055		過分散 $\sigma^{\wedge}=$	0.2442		
			ガンマ・ポアソン回帰			ゼロ過剰 ガンマ・ポアソン回帰			
		$x=$	20	25	30		20	25	30
	$j$	$y^{\wedge}=$	0.8098	2.1156	5.5271		2.6585	3.7178	5.1992
1	$y=$	0	0.5608	0.3360	0.1696		0.3989	0.3589	0.3340
2		1	0.2396	0.2129	0.1318		0.1434	0.0954	0.0551
3		2	0.1078	0.1420	0.1079		0.1438	0.1157	0.0785
4		3	0.0493	0.0963	0.0898		0.1150	0.1119	0.0893
5		4	0.0227	0.0659	0.0753		0.0803	0.0944	0.0886
6		5	0.0105	0.0453	0.0635		0.0512	0.0727	0.0802
7		6	0.0049	0.0312	0.0537		0.0305	0.0525	0.0680
8		7	0.0023	0.0216	0.0455		0.0173	0.0360	0.0549
9		8	0.0011	0.0149	0.0386		0.0095	0.0238	0.0426
10		9	0.0005	0.0103	0.0328		0.0050	0.0152	0.0320
11		10	0.0002	0.0072	0.0280		0.0026	0.0095	0.0234
12		11	0.0001	0.0050	0.0238		0.0013	0.0058	0.0168
13		12	0.0001	0.0035	0.0203		0.0006	0.0035	0.0118
14		13	0.0000	0.0024	0.0173		0.0003	0.0020	0.0082
15		14	0.0000	0.0017	0.0148		0.0002	0.0012	0.0056
16		15	0.0000	0.0012	0.0126		0.0001	0.0007	0.0038
17		16	0.0000	0.0008	0.0108		0.0000	0.0004	0.0025

## 第6章 文献索引

Agresti (2013) - Categorical Data Analysis 3rd.	213
岩崎 (2010) - カウントデータの統計解析	217
蓑谷 (2010) - 統計分布ハンドブック 増補版	212

## 第6章 索引

あ Agresti (2013) - 負の2項分布	213	か 甲羅の幅 - ガンマ・ポアソン回帰	228
アグレスティ(2003) - カブトガニ	221, 228	- 説明変数	228
幾つかの集団 - 必然的に過分散	220	- ゼロ過剰ガンマ・ポアソン回帰	233
位置パラメータ - ガンマ・ポアソン分布	213	Combin() 関数 - Excel	208
- 負の2項分布	213	混合 - ポアソン分布	210
- 平均 $\mu$	213	混合分布 - 部分集団	221
岩崎 (2010) - 負の2項分布	217	さ SumProduct() 関数 - Excel	210, 219
Excel - Gamma() 関数	212, 216	- 度数 ni	210
- Combin() 関数	208	事故件数 - 人工データ	210
- SumProduct() 関数	210, 219	失敗数 - 負の2項分布	207
- 対数ガンマ関数	216	JMP - ガンマPoisson分布	214
- NegBinom.dist() 関数	209	- ゼロ過剰ポアソン分布	221, 224
か 階乗 - ガンマ関数	211	- ゼロ強調負の2項分布	227
確率関数 - ゼロ過剰ガンマ・ポアソン分布	225	- 負の2項分布	215
- ゼロ過剰ポアソン分布	221, 225	人工データ - 事故件数	210
確率分布 - ガンマ・ポアソン回帰	231	推定値 - ガンマ・ポアソン回帰	230
- ゼロ過剰ガンマ・ポアソン回帰	235	数学的な解説 - ガンマ・ポアソン分布	217
カブトガニ - アグレスティ(2003)	221, 228	- 負の2項分布	217
過分散 - 通院回数	218	成功数を実数 - 負の2項分布	211
- 必然的に過分散	220	成功数をと固定 - 負の2項分布	207
- 分散/平均	210	説明変数 - 甲羅の幅	228
- ポアソン分布	207	Zero-Inflated - ゼロ過剰	221
過分散パラメータ - 変化	216	ゼロ過剰 - Cameron and Trivedi (1998)	221
過分散 $\sigma$ - ガンマ・ポアソン分布	213	- Zero-Inflated	221
- 形状パラメータ	213	- ポアソン分布	207, 221
Cameron and Trivedi (1998) - ゼロ過剰	221	ゼロ過剰ガンマ・ポアソン回帰 - 確率分布	235
- 通院回数	219	- ガンマ・ポアソン回帰	233
Gamma() 関数 - Excel	212	- 甲羅の幅	233
GammaLn() 関数 - Excel	216	ゼロ過剰ガンマ・ポアソン分布 - 確率関数	225
ガンマ・ポアソン回帰 - 確率分布	231	ゼロ過剰ポアソン分布 - 確率関数	221, 225
- 甲羅の幅	228	- JMP	221, 224
- 推定値	230	- ソルバー	223
- 対数尤度	229	ゼロ強調負の2項分布 - JMP	227
- ポアソン回帰	228	ソルバー - ガンマ・ポアソン分布	219
ガンマ・ポアソン分布 - 位置パラメータ	213	- ゼロ過剰ポアソン分布	223
- 過分散 $\sigma$	213	- 負の2項分布	211
- 形状の比較	216	た 対数ガンマ関数 - Excel	216
- 数学的な解説	217	対数尤度 - ガンマ・ポアソン回帰	229
- ソルバー	219	- ゼロ過剰ガンマ・ポアソン回帰	233
- パラメータ推定	214	- ポアソン回帰	228
- 負の2項分布	213	対数尤度の比較 - 4種のモデル	227
ガンマPoisson分布 - JMP	214	通院回数 - Cameron and Trivedi (1998)	219
ガンマ関数 - 階乗	211	- 過分散	219
- パラメータ推定	212	度数 ni - SumProduct() 関数	210
- 負の2項分布	211	- 平均と分散	210
形状の比較 - ガンマ・ポアソン分布	216	な NegBinom.dist() 関数 - Excel	209
形状パラメータ - 過分散 $\sigma$	213	は パラメータ推定 - ガンマ・ポアソン分布	214
- 負の2項分布	213	- ガンマ関数	212
形状パラメータ $\sigma$ - 分散	214	- 負の2項分布	210
交通事故の件数 - 負の2項分布	209	必然的に過分散 - 幾つかの集団	220

は	複数 - ポアソン分布	210
	負の2項分布 - Agresti (2013)	213
	- 位置パラメータ	213
	- 岩崎 (2010)	217
	- ガンマ・ポアソン分布	213
	- ガンマ関数	211
	- 形状パラメータ	213
	- 交通事故の件数	209
	- 失敗数	207
	- JMP	215
	- 数学的な解説	217
	- 成功数が実数	211
	- 成功数をと固定	207
	- ソルバー	211
	- パラメータ推定	210
	- ポアソン分布	209
	- ポアソン分布のあてはめ	211
	- 蓑谷 (2010)	212
	負の二項分布 - 分散	214
	部分集団 - 混合分布	221
	分散 - 形状パラメータ $\sigma$	214
	- 負の二項分布	214
	分散/平均 - 過分散	210
	平均 $\mu$ - 位置パラメータ	213
	平均と分散 - 度数 $n_i$	210
	変化 - 過分散パラメータ	216
	ポアソン回帰 - ガンマ・ポアソン回帰	228
	- 対数尤度	228
	ポアソン分布 - 過分散	207
	- 混合	210
	- ゼロ過剰	207, 221
	- 複数	210
	ポアソン分布のあてはめ - 負の2項分布	211
ま	蓑谷 (2010) - 負の2項分布	212
や	4種のモデル - 対数尤度の比較	227

## 第 6 章 解析ファイル一覧

	24 KB	第6章01_負の二項分布	Microsoft Excel ワークシート
	2 KB	第6章01_負の二項分布_ポアソン	JMP Data Table
	28 KB	第6章02_ガンマポアソン	Microsoft Excel ワークシート
	3 KB	第6章02_事故_ポアソン	JMP Data Table
	4 KB	第6章03_通院回数	JMP Data Table
	38 KB	第6章03_通院回数	Microsoft Excel ワークシート
	2 KB	第6章03_通院回数グラフ	JMP Data Table
	6 KB	第6章04_ゼロ_ポアソン	JMP Data Table
	28 KB	第6章04_ゼロ_ポアソン	Microsoft Excel ワークシート
	5 KB	第6章05_ゼロ_ガンマ・ポアソン	JMP Data Table
	21 KB	第6章05_ゼロ_ガンマ・ポアソン	Microsoft Excel ワークシート
	74 KB	第6章06_カプトガニ_ガンマポアソン	Microsoft Excel ワークシート
	6 KB	第6章06_過大分散比較	JMP Data Table
	76 KB	第6章07_カプトガニ_zero過剰ガンマポアソン	Microsoft Excel ワークシート

非売品, 無断複製を禁ずる

第 9 回 続高橋セミナー

最尤法によるポアソン回帰分析入門 <<第 6 章>>

**第 6 章 過分散・ゼロ過剰への対処**

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3 の 1005

2020 年 6 月 7 日 高橋 行雄