

年賀状の番号の分布

2020.1.19
渡邊 俊夫

年賀状のお年玉くじ

年賀状にはお年玉くじが付いていて、6桁の数字が当選番号に合致すると賞品をもらうことができる。例年、1等の当選番号は6桁で1つ（100万本に1本）、2等は下4桁で1つ（1万本に1本）、3等の当選番号は下2桁で2つ（100本に2本）である。

※2019年以降、3等の当選番号が下2桁で3つ（100本に3本）になった。

毎年100枚くらいの年賀状が届く場合、当選本数から考えて、1等や2等が当たる可能性は低く、当たる可能性が高いのは3等（例年、賞品は年賀切手シート）である。

3等が100本に2本であれば、届いた100枚の年賀状の中には、2つの当選番号がそれぞれ1枚ずつありそうだが、それは正しいだろうか？

年賀状の当選枚数(実測)

2008年以降(2012年と2014年は喪中のため除く)の当選枚数を下表に示す。

年		干支	届いた枚数	当選番号		当選枚数		当選計
2008年	平成20年	子	95	37	64	2	0	2
2009年	平成21年	丑	104	46	94	1	1	2
2010年	平成22年	寅	95	00	52	0	2	2
2011年	平成23年	卯	104	02	69	0	2	2
2013年	平成25年	巳	102	29	70	0	1	1
2015年	平成27年	未	98	27	30	0	1	1
2016年	平成28年	申	96	69	90	0	1	1
2017年	平成29年	酉	95	45	51	0	1	1
2018年	平成30年	戌	93	27	68	1	1	2

年賀状の当選枚数(実測)

2019年以降の当選枚数を下表に示す。

年		干支	届いた枚数	当選番号			当選枚数			当選計
2019年	平成31年	亥	92	02	42	78	0	1	0	1
2020年	令和2年	子	86	16	37	67	0	1	0	1

年賀状の当選枚数(実測)

前掲の表中、当選本数の割合が100本に2本の9年(2008~2018年、喪中を除く)のうち、年2枚当たったのが5回で、年1枚当たったのが4回だった。2枚当たった年も、2つの当選番号が1枚ずつ当たったのは2回だけで、それ以外の3回は一方の当選番号が2枚、他の当選番号は0枚だった。

また、当選本数の割合が100本に3本になった2019年以降は2年とも、3つの当選番号のうち1つの番号が1枚で、他の2つの番号は0枚だった。

つまり、11年で24通りの当選番号のうち、3枚以上当たったものはなく、2枚当たったのは3通り、1枚当たったのは10通り、0枚だったのは11通りだった。

年賀状の当選枚数(考察)

前掲の結果(11年で24通りの当選番号のうち、11通りは0枚)は、一見、0枚だった当選番号が多すぎるようにも思えるが、これは年賀状の番号がランダムに届くとすると、定性的には妥当である。

なぜなら、ランダムならば、ある1枚の年賀状の番号は他の年賀状の番号とは無関係に決まるからである。100枚の年賀状が届くとき、00から99までの100通りの番号が均等に1枚ずつ届くのは、むしろランダムではない。ある1つの番号の枚数の期待値は1枚であるが、0枚の番号も2枚以上の番号もあり、全ての番号についての平均値が1枚になるというだけである。したがって、ある1つの当選番号が0枚でも不思議ではない。

それでは、年賀状の番号がランダムに届くとき、その番号の分布はどのようになるのだろうか？

二項分布

0 から $n - 1$ までの番号 (下2桁の場合は $n = 100$) のうちの1つが書かれた年賀状をランダムに m 枚受け取ったとき、同じ番号の年賀状が何枚あるかを考える。

ある番号の年賀状を k 枚 ($0 \leq k \leq m$) 受け取る確率 $P(k)$ は、 m 枚のうち k 枚がその番号で、残りの $m - k$ 枚がそれ以外の番号になることから

$$\begin{aligned} P(k) &= {}^m C_k \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} = \frac{m!}{(m-k)! k! n^k} \frac{(n-1)^{m-k}}{n^{m-k}} \\ &= \frac{m!}{(m-k)! k!} \frac{(n-1)^{m-k}}{n^m} \end{aligned}$$

であり、これは二項分布である。

ポアソン分布

二項分布

$$P(k) = \frac{m!}{(m-k)!k!} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k} = \frac{m!}{(m-k)!k!} \frac{(n-1)^{m-k}}{n^m}$$

は、 $n \gg 1$ のとき、 $m \gg k$ となるから、 $\mu = m/n$ とおくと

$$\begin{aligned} P(k) &= \frac{m(m-1)\cdots(m-k+1)}{n^k} \frac{1}{k!} \left(1 - \frac{1}{n}\right)^{m-k} \\ &= \frac{m}{n} \frac{m-1}{n} \cdots \frac{m-k+1}{n} \frac{1}{k!} \left(1 - \frac{1}{n}\right)^{m-k} \cong \left(\frac{m}{n}\right)^k \frac{1}{k!} \left(1 - \frac{1}{n}\right)^m \\ &= \frac{\mu^k}{k!} \left(1 - \frac{\mu}{m}\right)^m \cong \frac{\mu^k}{k!} e^{-\mu} \end{aligned}$$

と近似できる。これは平均値 μ のポアソン分布である。

二項分布とポアソン分布

二項分布

$$P(k) = \frac{m!}{(m-k)!k!} \frac{(n-1)^{m-k}}{n^m}$$

および、ポアソン分布 ($\mu = m/n$)

$$P(k) = \frac{\mu^k}{k!} e^{-\mu}$$

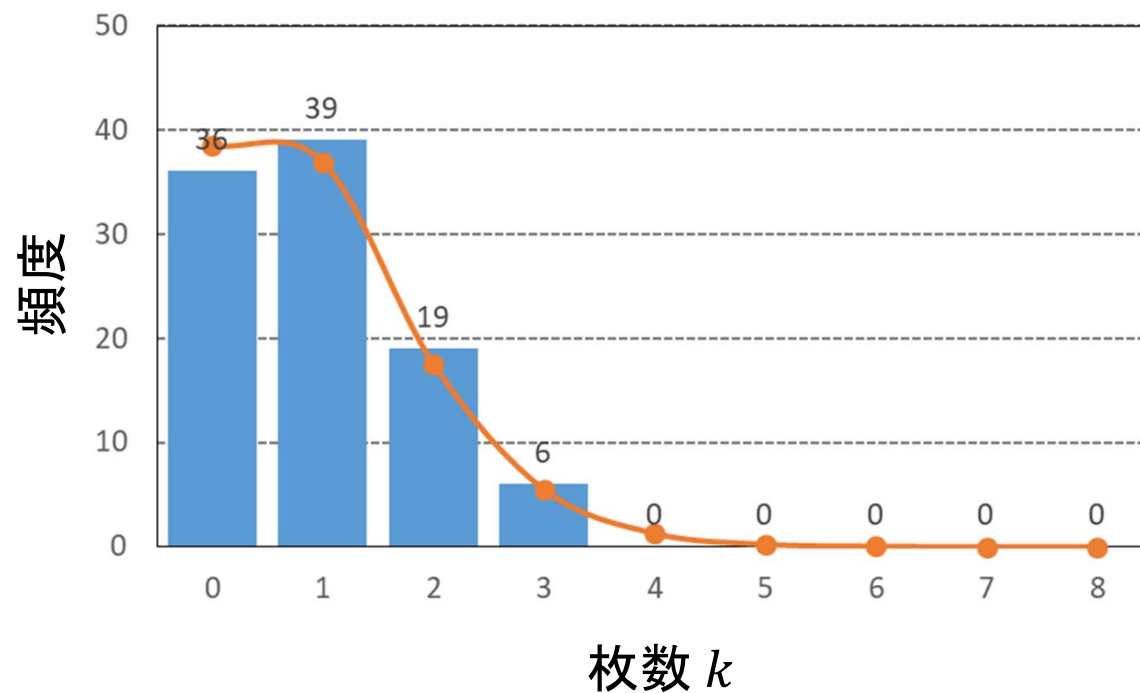
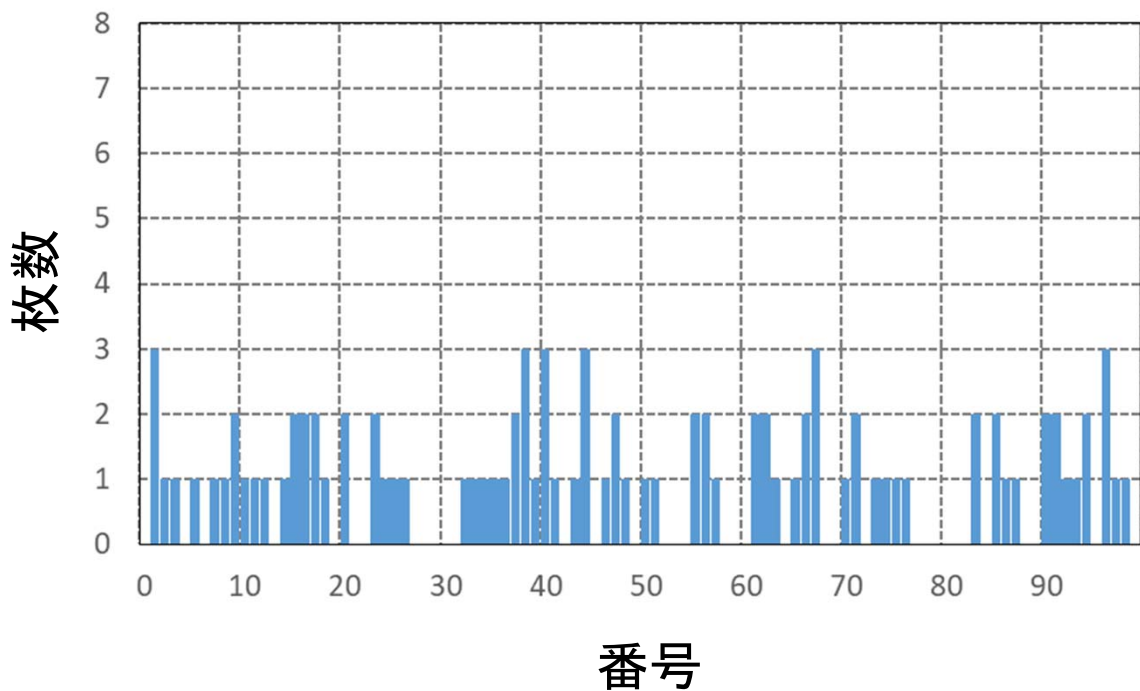
において、 $n = 100, m = 100$ のときの $P(k)$ の値を右表に示す。両者は、ほぼ一致している。

$P(0) \cong 0.37$ であるから、100通りの番号のうち約37通りは0枚である。

k	二項分布 ($n = m = 100$)	ポアソン分布 ($\mu = 1$)
0	0.3660	0.3679
1	0.3697	0.3679
2	0.1849	0.1839
3	0.0610	0.0613
4	0.0149	0.0153
5	0.0029	0.0031
6	4.6×10^{-4}	5.1×10^{-4}
7	6.3×10^{-5}	7.3×10^{-5}
8	7.4×10^{-6}	9.1×10^{-6}

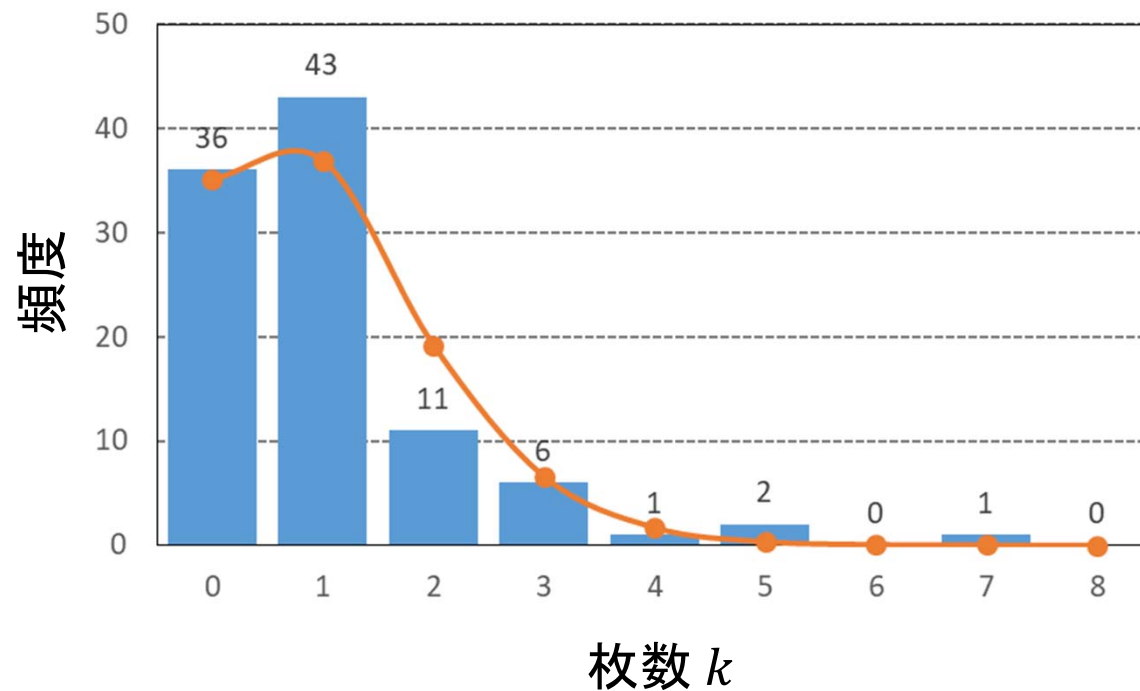
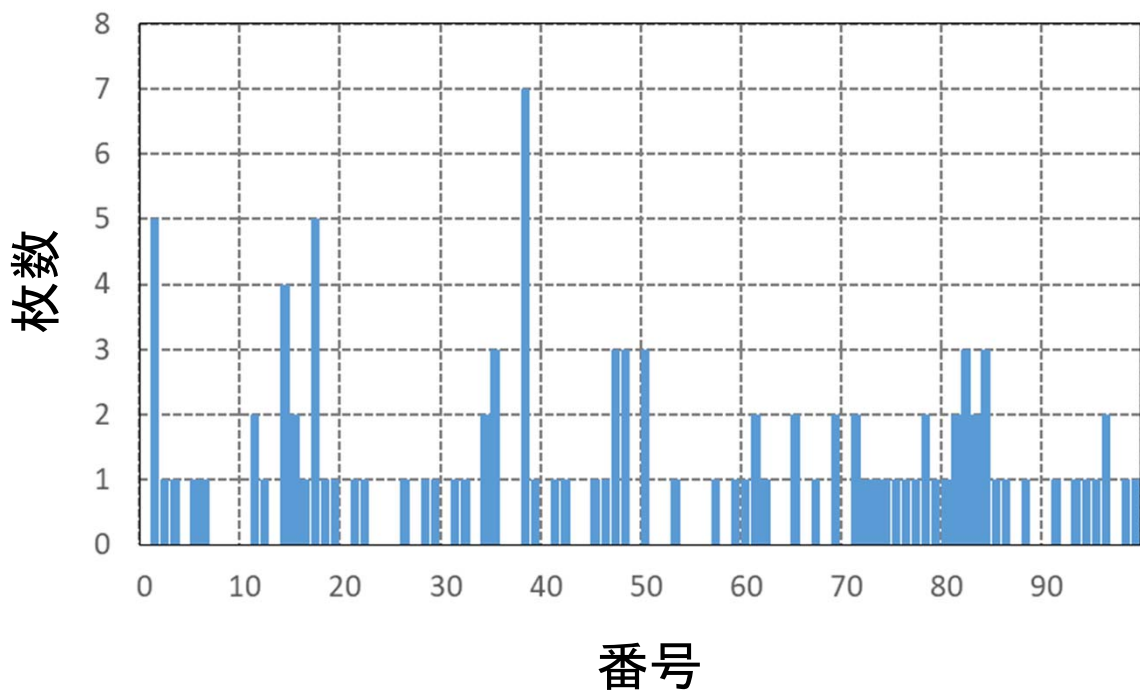
年賀状の番号と枚数(実測):2008年

2008年に届いた年賀状($m = 95$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)とほぼ一致している。



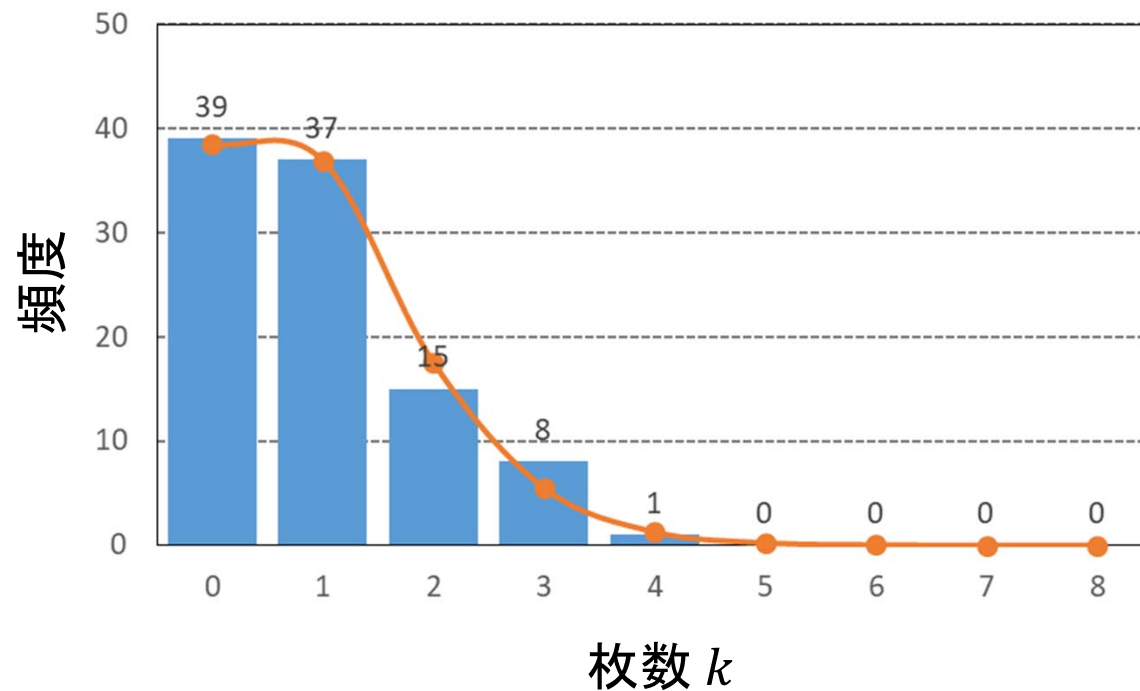
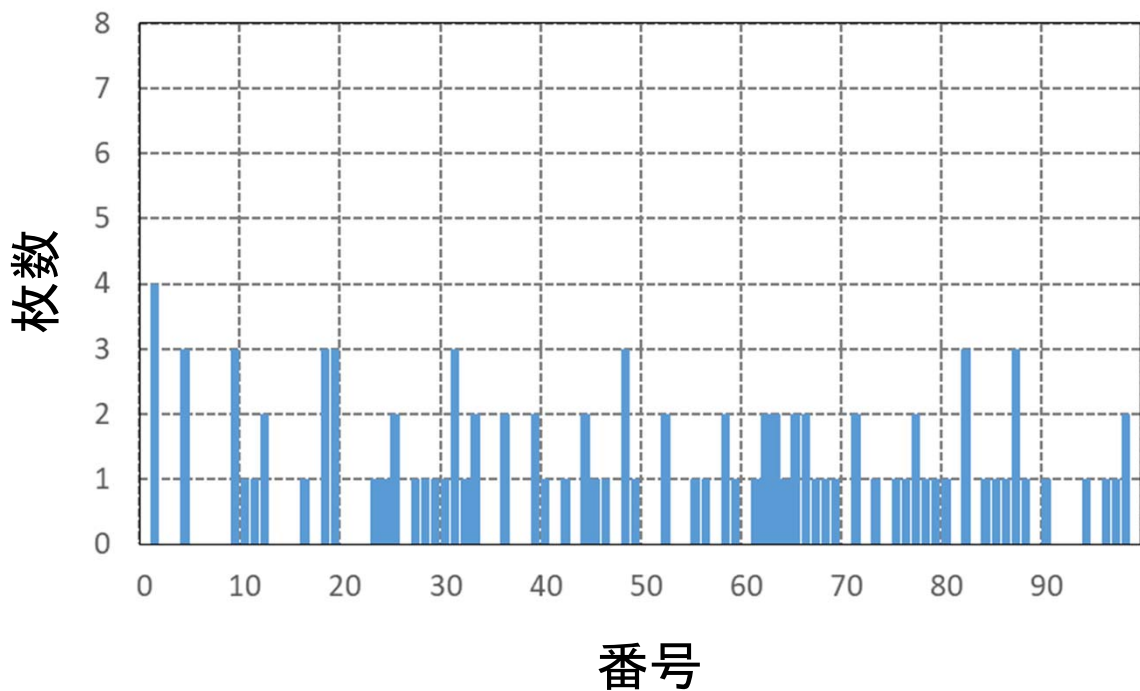
年賀状の番号と枚数(実測):2009年

2009年に届いた年賀状($m = 104$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。二項分布による計算結果(赤実線)と比べると、 $k = 1, 5, 7$ の頻度が高くて $k = 2$ の頻度が低く、少し外れている。



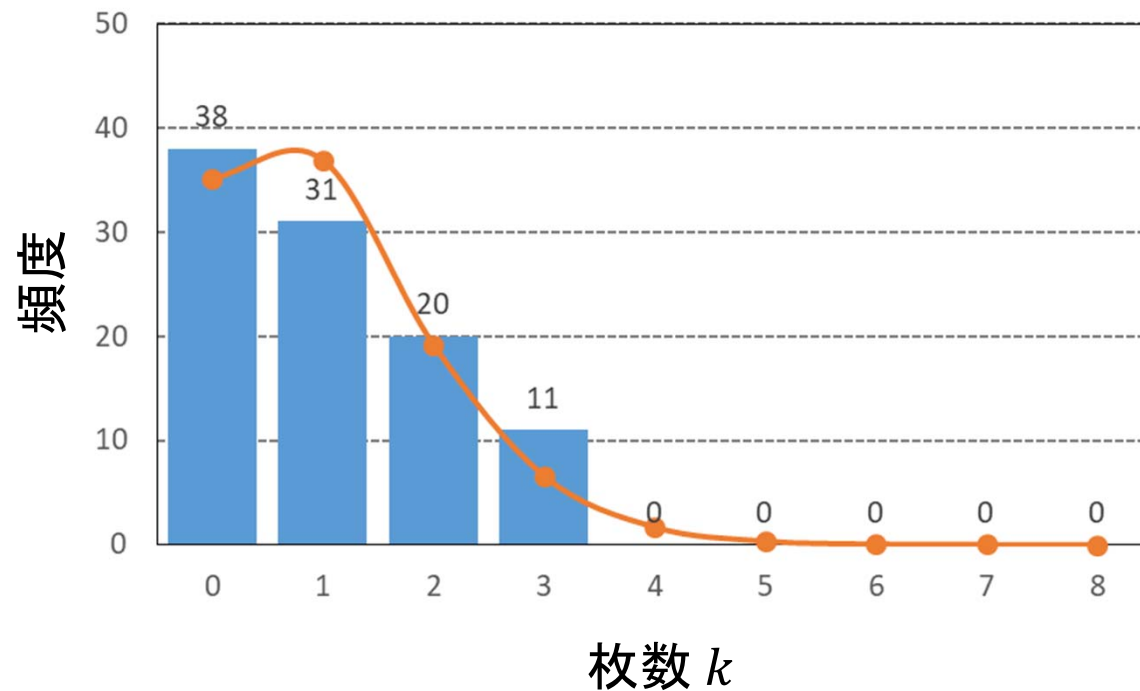
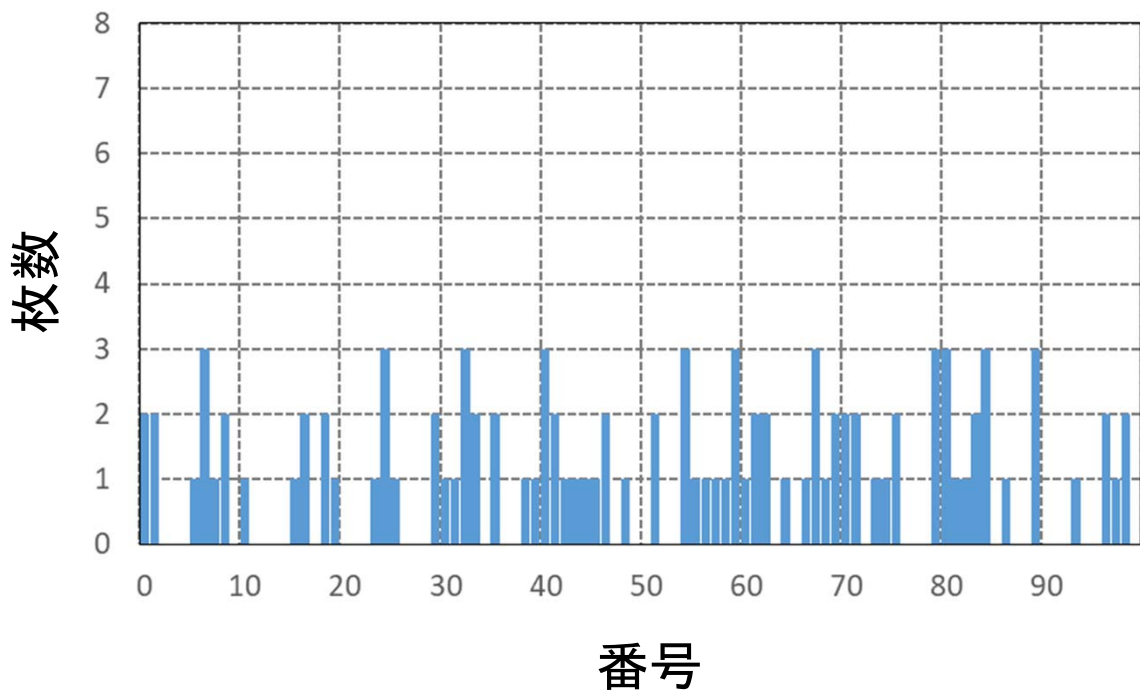
年賀状の番号と枚数(実測):2010年

2010年に届いた年賀状($m = 95$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)に近い分布となっている。



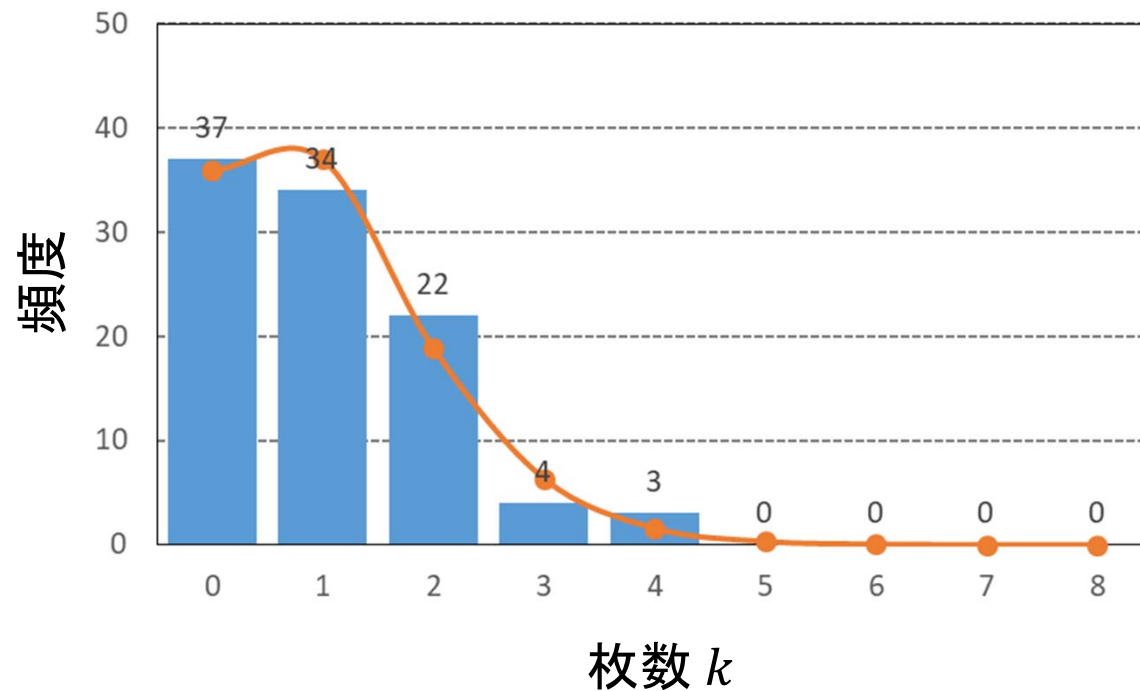
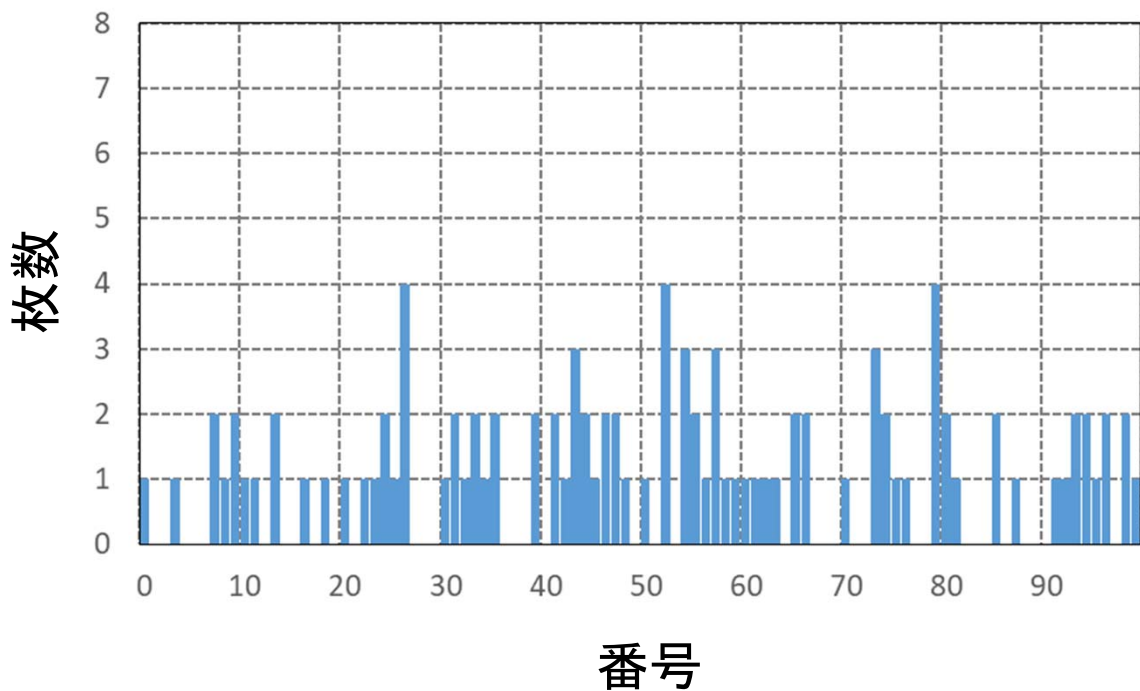
年賀状の番号と枚数(実測):2011年

2011年に届いた年賀状($m = 104$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。二項分布による計算結果(赤実線)と比べると、 $k = 1$ の頻度が低くて $k = 3$ の頻度が高く、少し外れている。



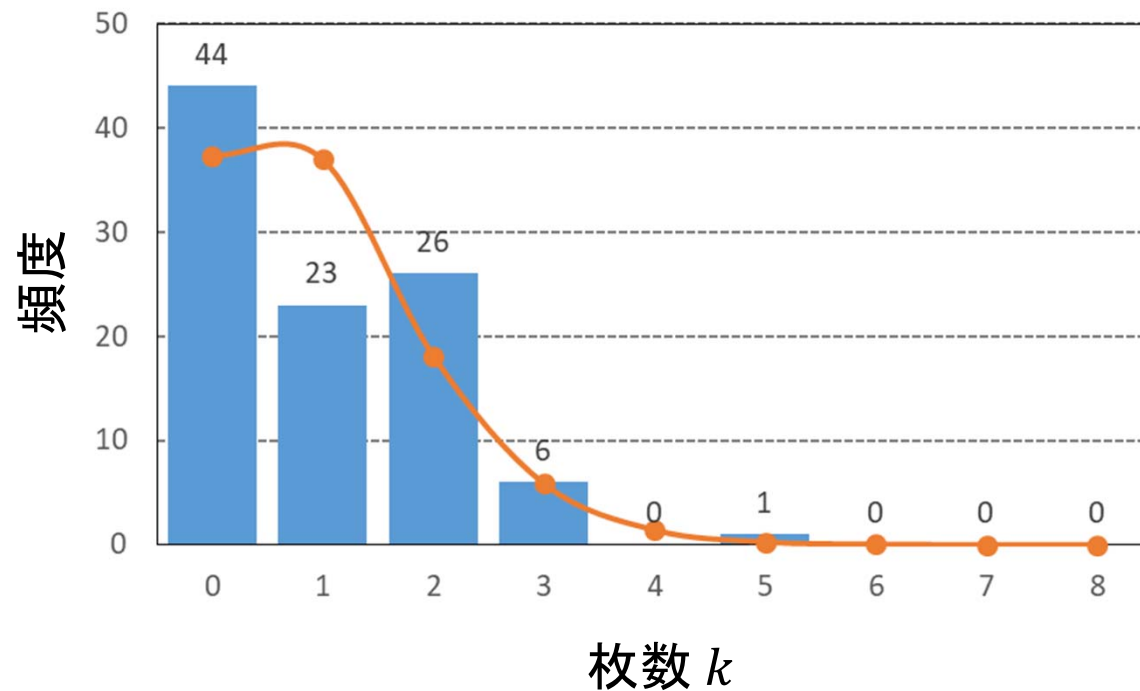
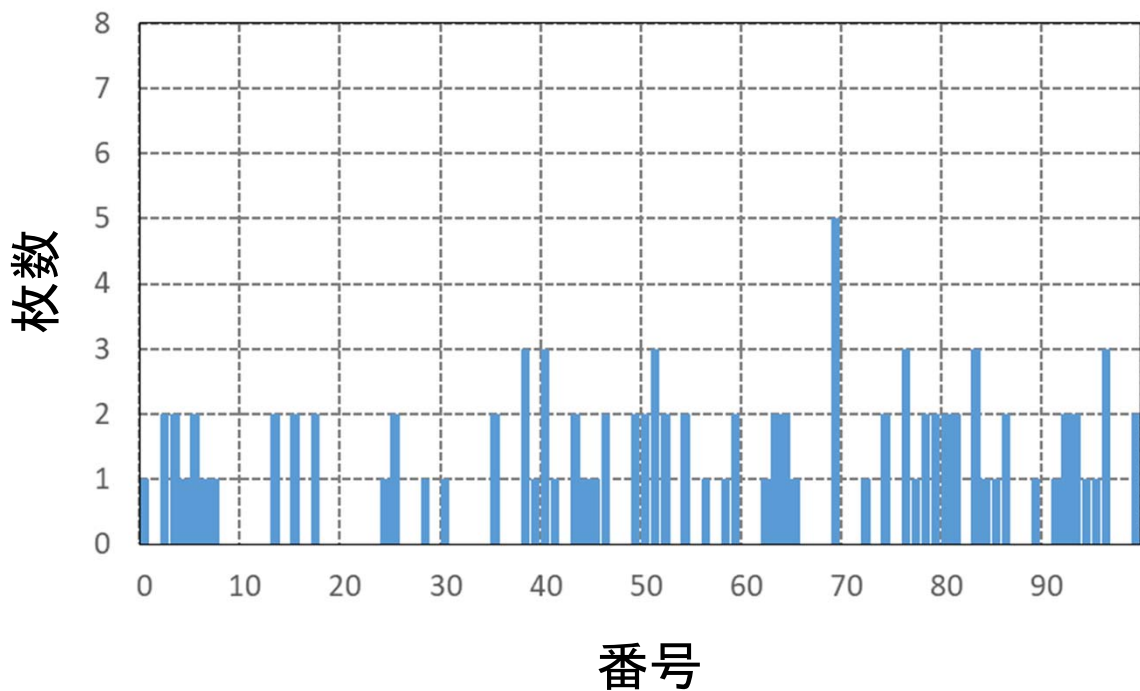
年賀状の番号と枚数(実測):2013年

2013年に届いた年賀状($m = 102$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)に近い分布となっている。



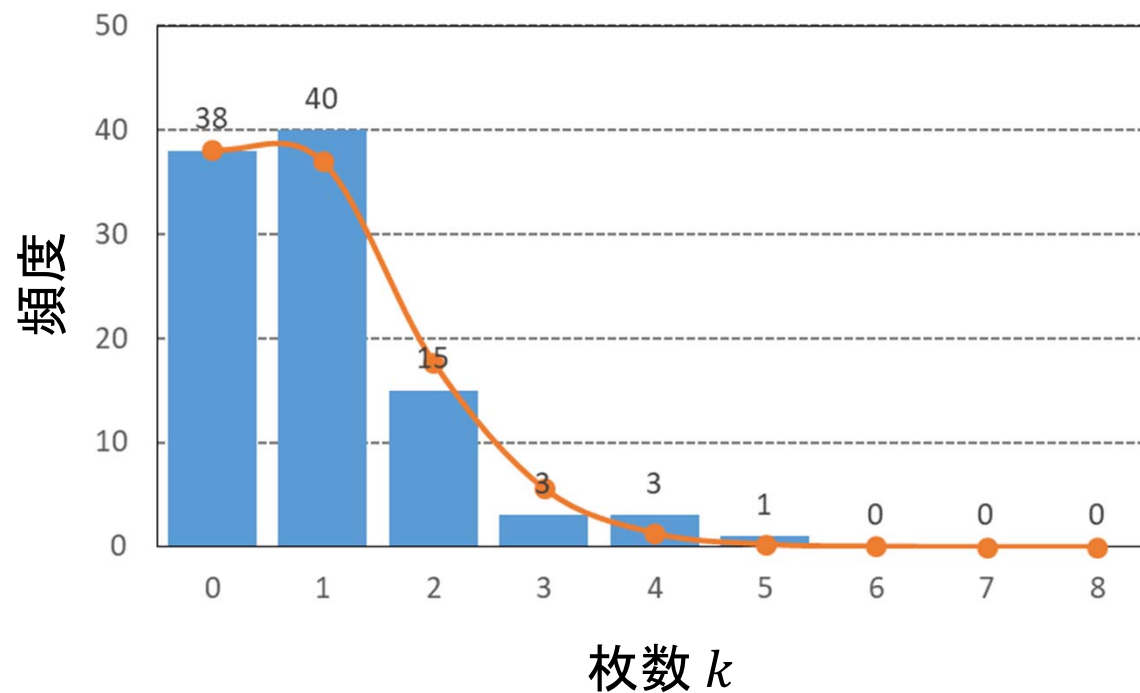
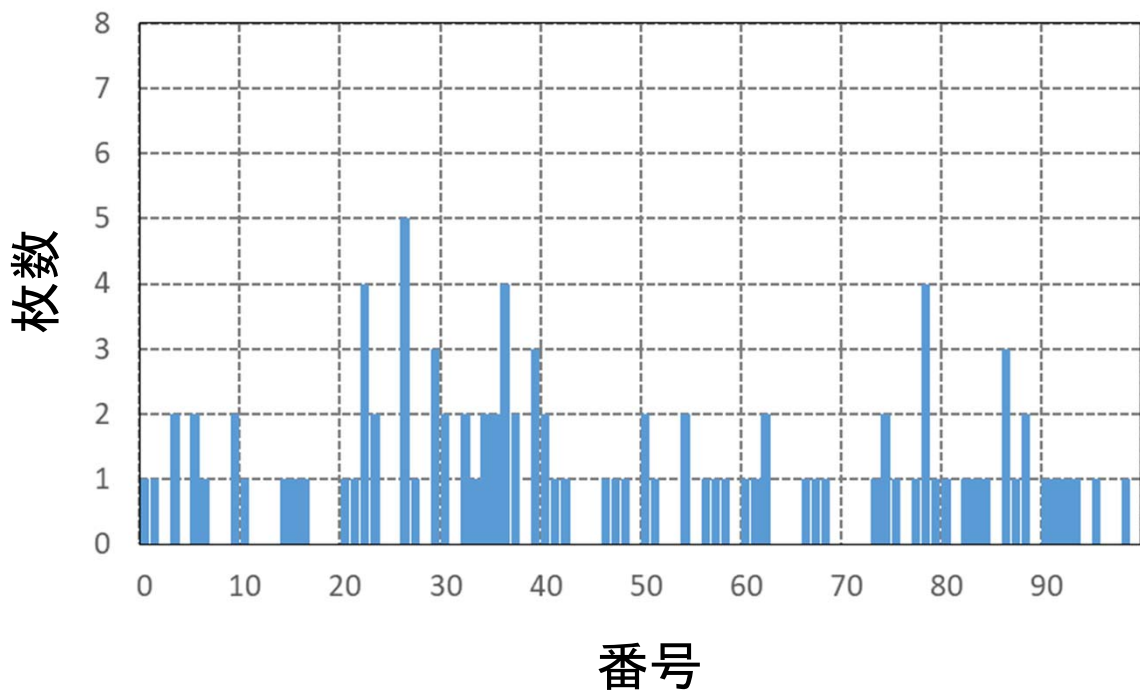
年賀状の番号と枚数(実測):2015年

2015年に届いた年賀状($m = 98$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。二項分布による計算結果(赤実線)と比べると、 $k = 0, 2$ の頻度が高くて $k = 1$ の頻度が低く、乖離している。



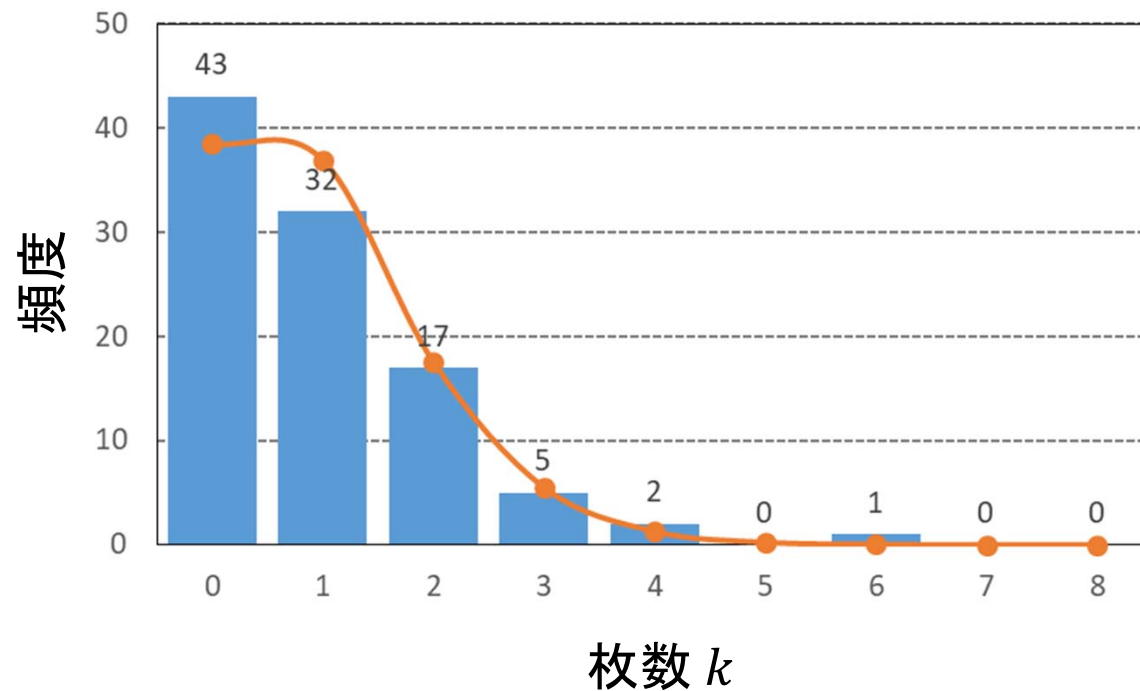
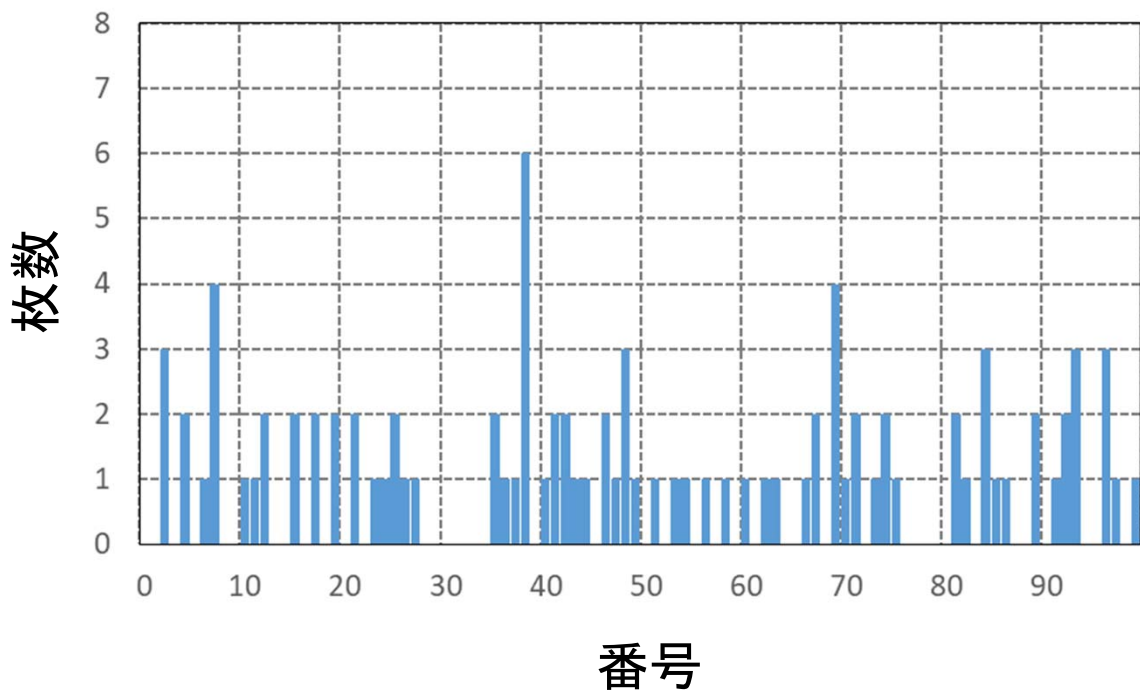
年賀状の番号と枚数(実測):2016年

2016年に届いた年賀状($m = 96$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)にほぼ一致している。



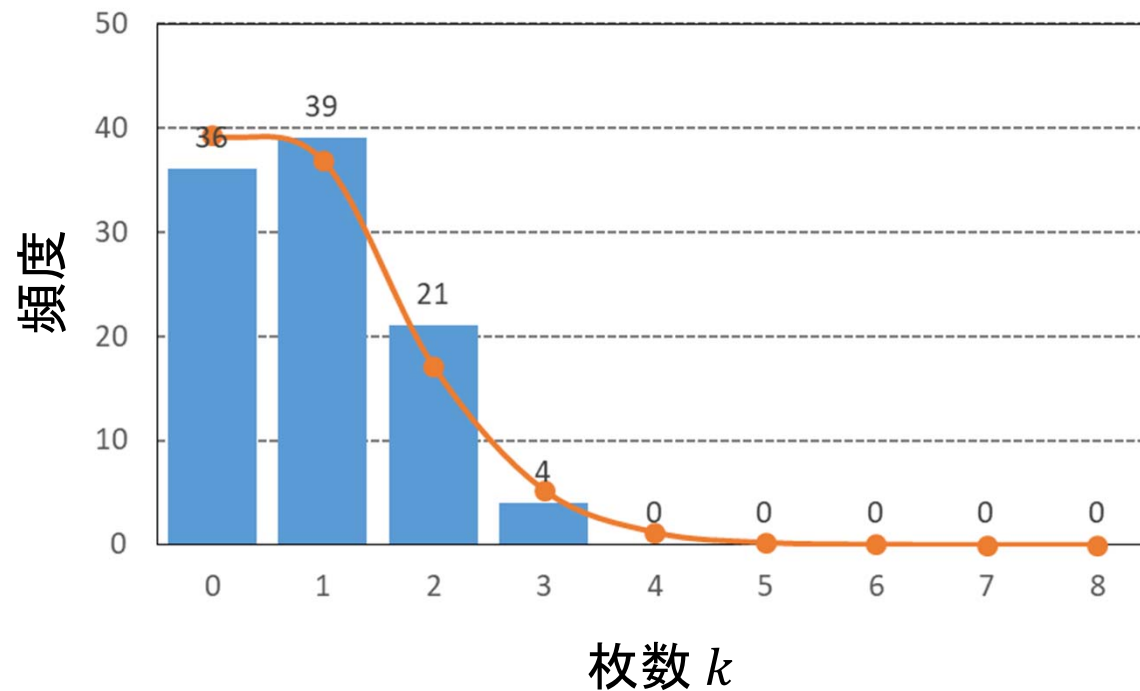
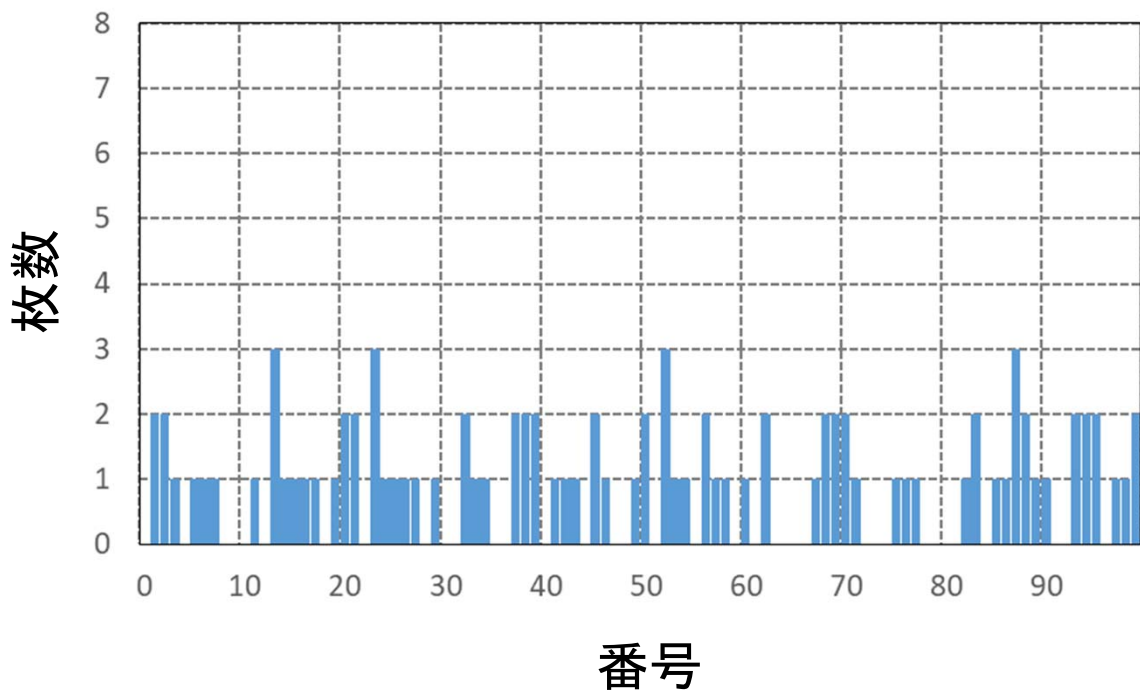
年賀状の番号と枚数(実測):2017年

2017年に届いた年賀状($m = 95$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)に近い分布となっている。



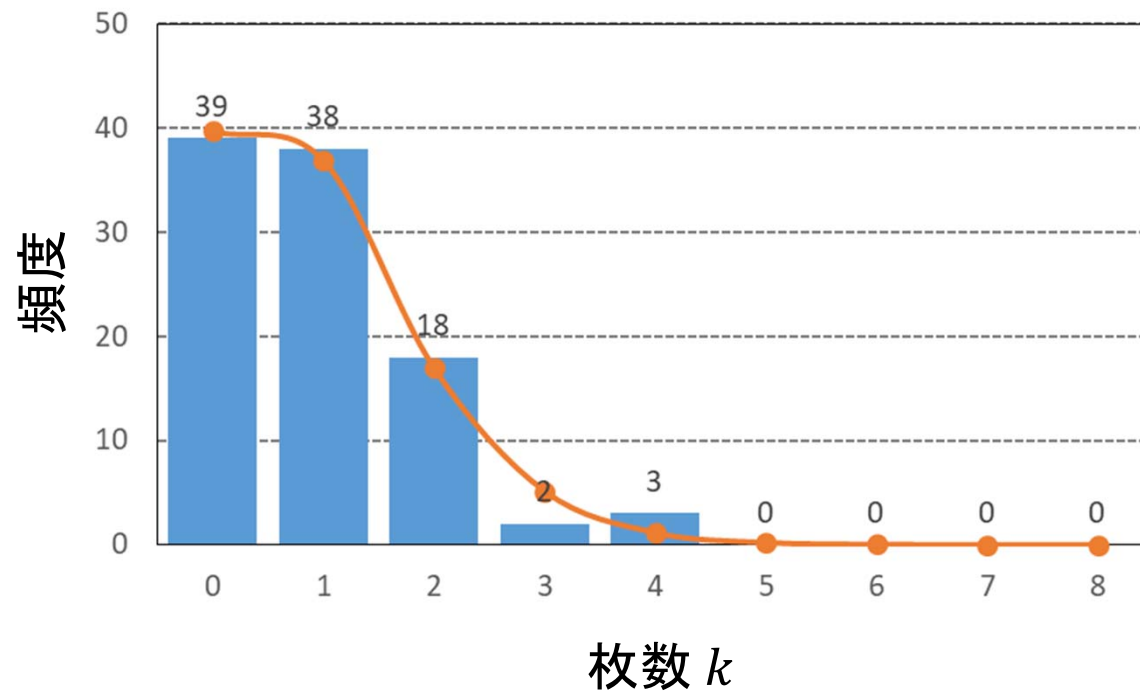
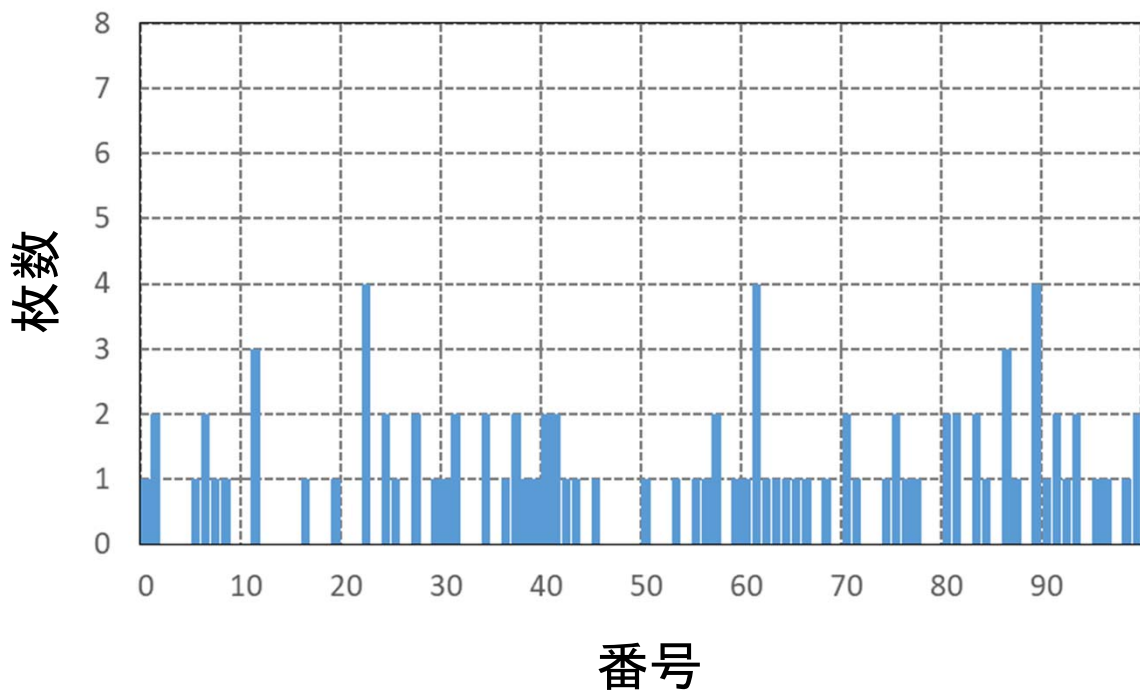
年賀状の番号と枚数(実測):2018年

2018年に届いた年賀状($m = 93$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)と比べると $k = 2$ の頻度が高く、少し外れている。



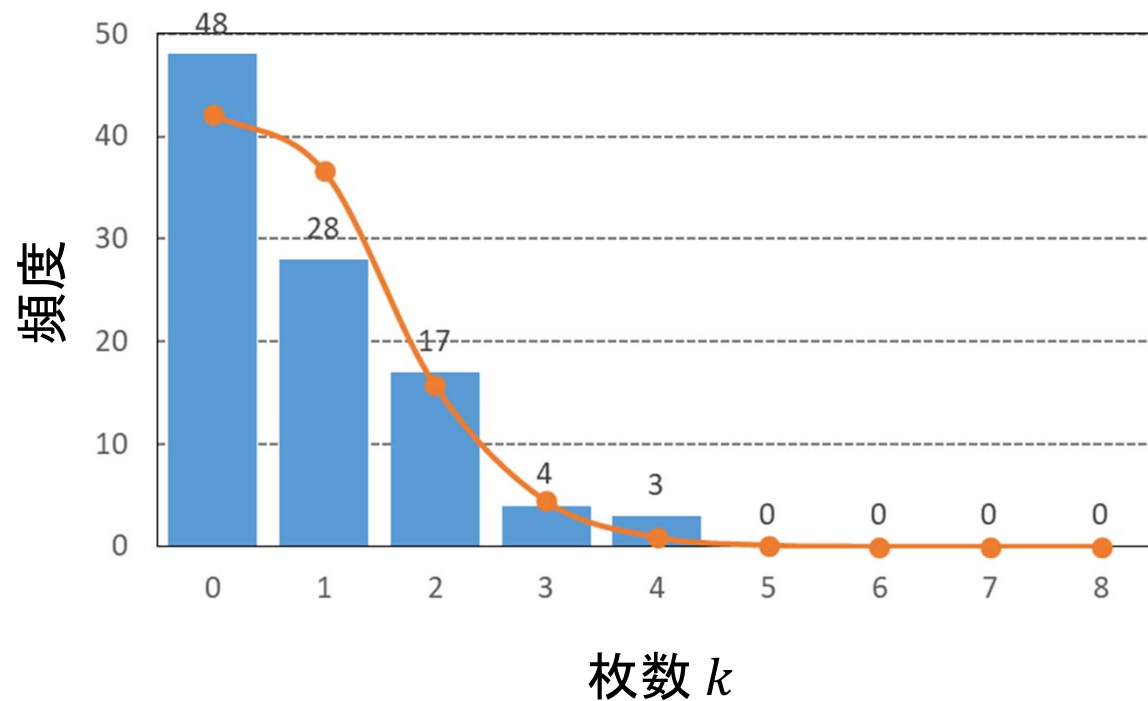
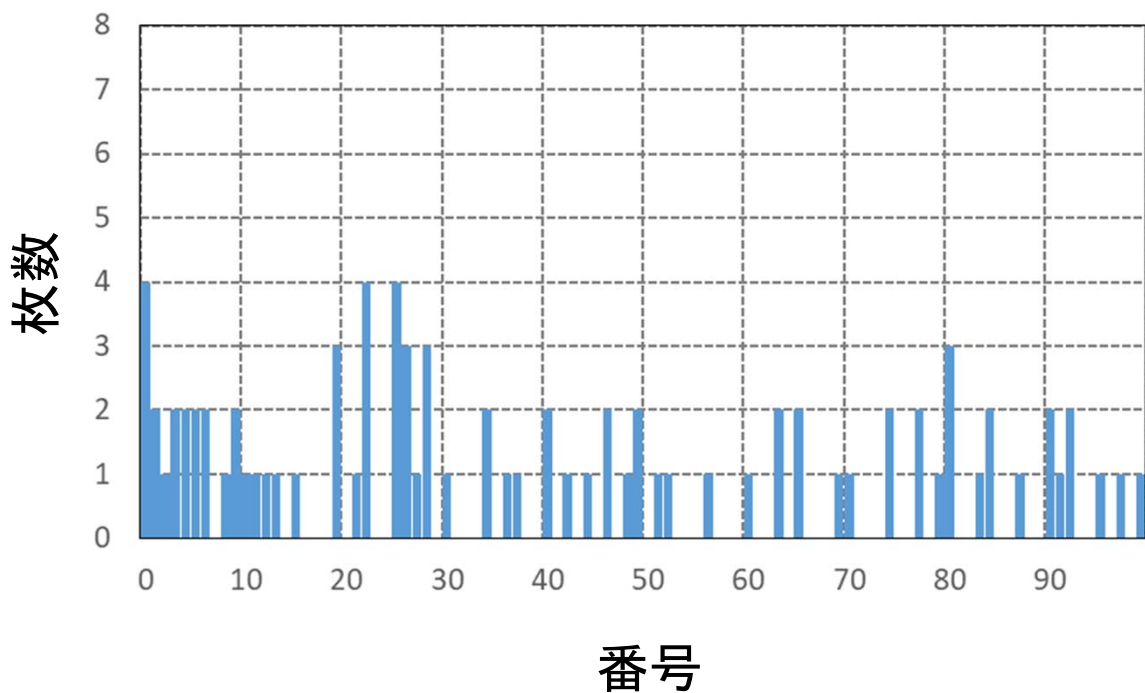
年賀状の番号と枚数(実測):2019年

2019年に届いた年賀状($m = 92$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)にほぼ一致している。



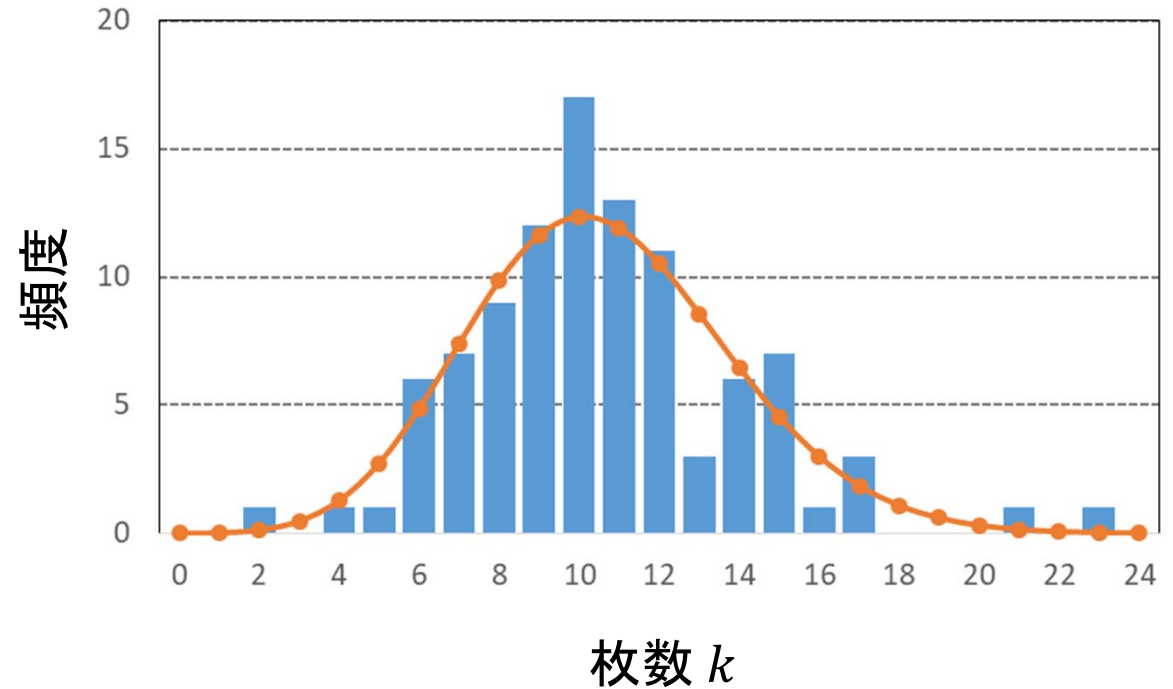
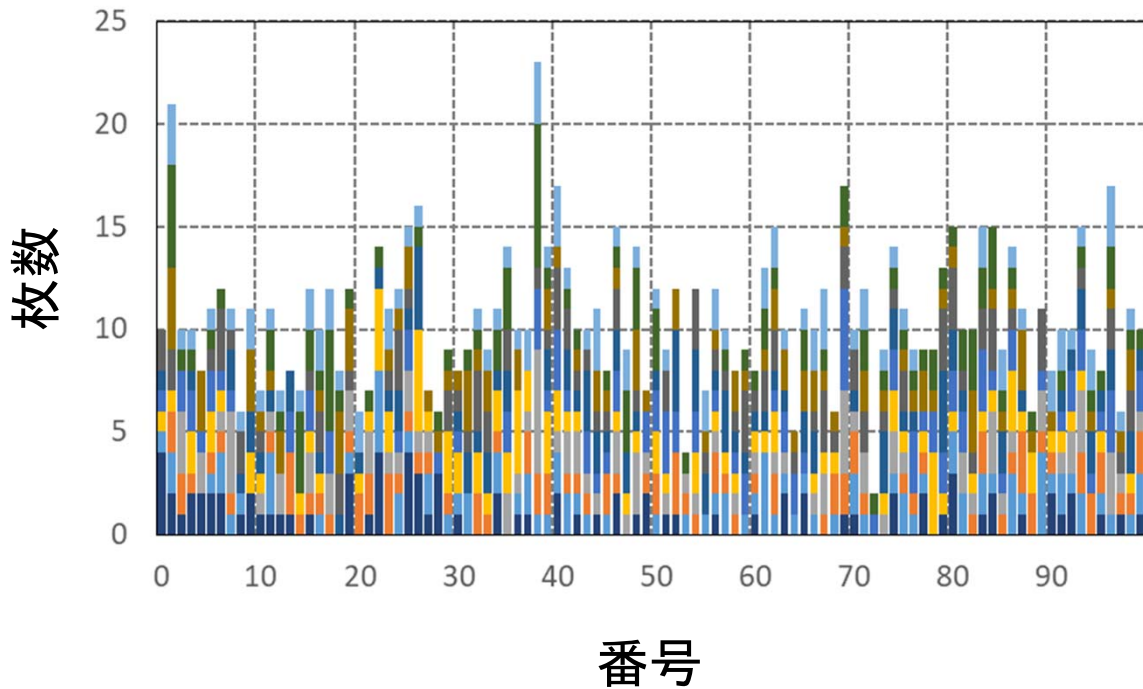
年賀状の番号と枚数(実測):2020年

2020年に届いた年賀状($m = 86$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。頻度は二項分布による計算結果(赤実線)と比べると、 $k = 0$ の頻度が高くて $k = 1$ の頻度が低く、少し外れている。



年賀状の番号と枚数(実測): 2008~2020年

2008~2020年(2012年, 2014年を除く)の11年に届いた年賀状($m = 1060$)の下2桁の各番号が何枚あったかを下左図に示す。また、下右図の棒グラフは同じ番号が k 枚あった番号の頻度を示す。二項分布による計算結果(赤実線)と比べると、 $k = 10$ の頻度が高くて $k = 13$ の頻度が低い。



χ^2 検定: 2008年

2008年に届いた年賀状 ($m = 95$) について、 χ^2 検定を行う。同じ番号が k 枚あった番号の頻度の実測値を O_k 、二項分布による期待値を E_k として

$$\chi^2 = \sum_k \chi_k^2 = \sum_k \frac{(O_k - E_k)^2}{E_k}$$

を求めると、右表に示すように $\chi^2 = 0.55$ となる。ただし、 $E_k > 5$ となるように $k \geq 3$ はひとまとめとした。

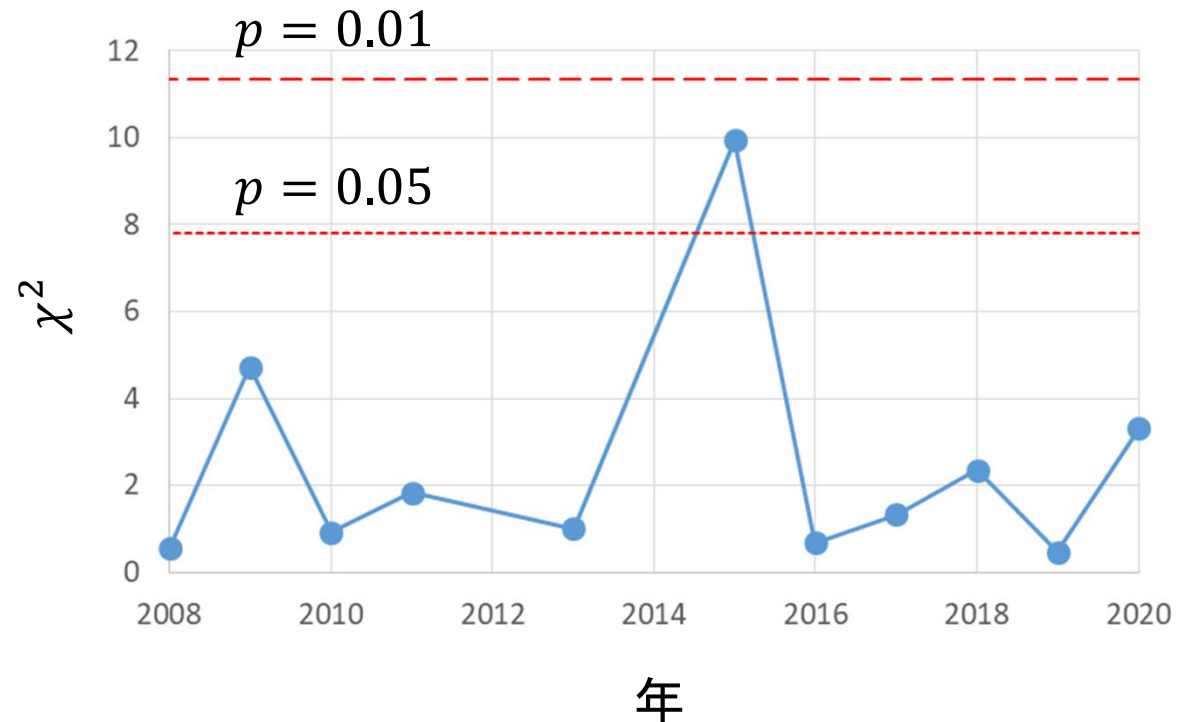
自由度 3 の χ^2 分布において、 $\chi^2 = 0.55$ となる差が見いだされる確率は 0.91 であり、統計学的有意水準 (0.05 または 0.01) より大きい。

k	実測	二項分布 ($m = 95$)	χ_k^2
0	36	38.49	0.16
1	39	36.93	0.12
2	19	17.53	0.12
≥ 3	6	7.04	0.15
計	100	100	0.55

χ^2 検定: 2008~2020年

2008~2020年(2012年, 2014年を除く)の11年について、二項分布に対する χ^2 を下図に示す。自由度3の χ^2 分布において、それより大きな差が見いだされる確率が $p = 0.05$ となるのは $\chi^2 = 7.81$ のとき、 $p = 0.01$ となるのは $\chi^2 = 11.34$ のときである。

2015年は $\chi^2 = 9.94$ で、 $p = 0.05$ となる $\chi^2 = 7.81$ を上回っているが、それ以外の年は $\chi^2 = 7.81$ を下回っていて $p > 0.05$ である。よって、年賀状の番号の分布が二項分布にいたがっているという仮説は棄却されないと言える。



まとめ

自分宛てに届く年賀状のお年玉くじの下2桁の番号は、00～99の100通りが均等には分布していない。

年賀状の番号がランダムに届くとすれば、その分布は二項分布(ポアソン分布)になる。

100枚の年賀状が届く場合、下2桁の番号100通りのうち約37通りは0枚であると見積もられる。

謝辞

いつも年賀状をいただき皆様に心より感謝申し上げます。

補足：当選枚数の確率

100通り ($n = 100$) の番号の中に当選番号が2通りあるとき、届いた m 枚のうち l 枚が当たる確率は

$$\sum_{i=0}^l p(i)p(l-i)$$

となる。 $n = 100, m = 100$ の二項分布についての計算結果を右表に示す。

100枚の年賀状が届く場合、1枚も当たらない確率は約13%あり、1枚だけ当たる確率は約27%、2枚以上が当たる確率は約60%である。

l	確率	累積確率
0	0.1340	0.1340
1	0.2707	0.4046
2	0.2720	0.6767
3	0.1814	0.8580
4	0.0902	0.9483
5	0.0357	0.9840
6	0.0117	0.9957
7	0.0033	0.9990
8	0.0008	0.9998
9	0.0002	1.0000
10	0.0000	1.0000

補足：当選枚数の確率

100通り ($n = 100$) の番号の中に当選番号が3通りあるとき、届いた m 枚のうち l 枚が当たる確率を $m = 100$ の二項分布について計算すると右表のようになる。

100枚の年賀状が届く場合、1枚も当たらない確率は約5%であり、1枚だけ当たる確率は約15%、2枚が当たる確率は約22%、3枚以上が当たる確率は約58%である。

l	確率	累積確率
0	0.0490	0.0490
1	0.1486	0.1976
2	0.2244	0.4221
3	0.2252	0.6472
4	0.1689	0.8161
5	0.1010	0.9171
6	0.0502	0.9672
7	0.0213	0.9885
8	0.0079	0.9964
9	0.0026	0.9990
10	0.0006	0.9996
11	0.0002	0.9998
12	0.0001	0.9999