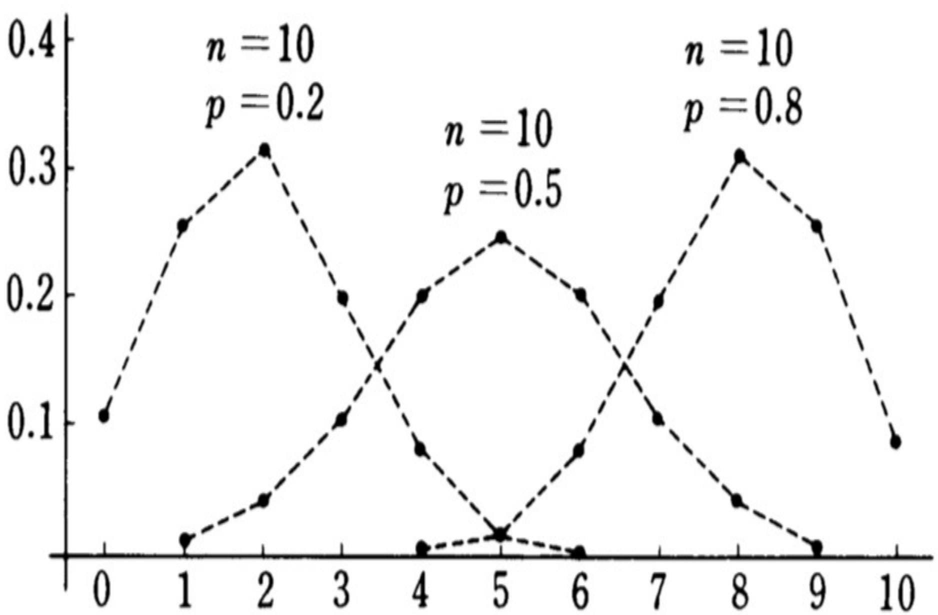
確率分布間の関係

二項分布

$$P(X=k)=C_n^k p^k(1-p)^{n-k}=\frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k}$$



2 項分布

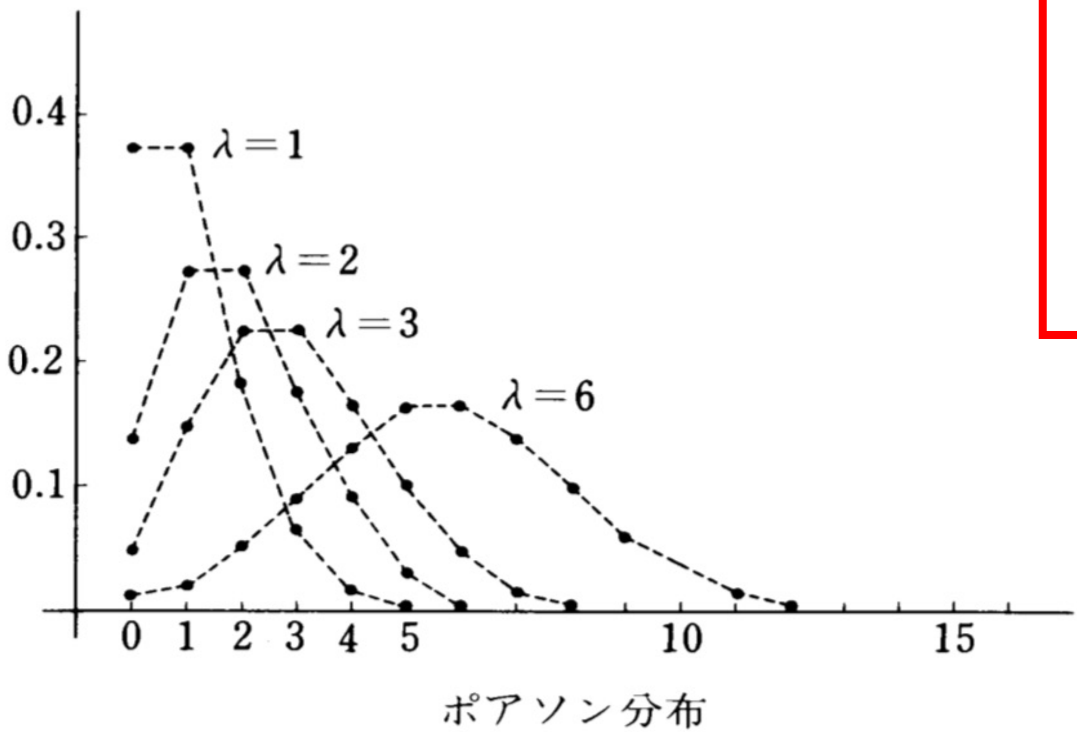
2 項分布の平均と分散

$$\mu = np$$

$$\sigma^2 = npq$$

ポアソン分布

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



ポアソン分布の平均と分散

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$

ポアソン分布

例題 7 （ポアソン分布の応用）

りんごを250個ずつ箱に詰める．箱詰めされたりんごは平均して0.8％が腐るという．箱詰めりんご1箱をあけたとき，腐ったりんごが3個以上見出される確率を求めよ．

ポアソン分布

例題 7 （ポアソン分布の応用）

りんごを250個ずつ箱に詰める．箱詰めされたりんごは平均して0.8％が腐るという．箱詰めりんご1箱をあけたとき，腐ったりんごが3個以上見出される確率を求めよ．

ポイント

平均を求める！

ポアソン分布

解　1箱中の腐ったりんごの数を X とすると，X は $n=250$，$p=0.008$ の2項分布に従う．この2項分布は n が50より大きく，p は非常に小さく，$\lambda=np=250\times0.008=2\leqq5$ であるから，$\lambda=2$ のポアソン分布

$$P(X=r)=\frac{2^r e^{-2}}{r!}, \quad (r=0, 1, 2, \cdots)$$

で近似できる．よって，求める確率は

$$\begin{aligned} P(X\geqq3)&=1-P(0)-P(1)-P(2) \\ &=1-e^{-2}-2e^{-2}-2e^{-2} \\ &=1-5e^{-2}=\mathbf{0.32} \end{aligned}$$

ポアソン分布

例題 8 （ポアソン分布の応用）

A 駅の売店でのある月刊雑誌の販売数は平均 2 冊のポアソン分布に従う．売店では毎月この雑誌を 3 冊仕入れている．

（a） この店がある月，客の需要を満たせなくなる確率を求めよ．

（b） この店でこの雑誌が 1 月当たりに売れる平均販売数を求めよ．

（c） この店が毎月の雑誌の需要を満たす確率を少なくとも 0.95 にするには毎月最低何冊を仕入れねばならないか．

解　この店の毎月の雑誌販売数を X とすると

$$P(X=x)=\frac{2^x e^{-2}}{x!} \qquad (x=0,1,2,\cdots)$$

(a)　客の需要が満たせないのは，$X \geqq 4$ のときだから

$$\begin{aligned}P(X\geqq 4)&=1-P(X\leqq 3)\\&=1-P(0)-P(1)-P(2)-P(3)\\&=1-e^{-2}-2e^{-2}-2e^{-2}-\frac{4}{3}e^{-2}\\&=1-\frac{19}{3}e^{-2}=\mathbf{0.14}\end{aligned}$$

(b)　このポアソン分布の平均は 2 だから，**2 冊**.

(c)　求める冊数を n とすると

$$P(X\leqq n)\geqq 0.95$$

$$e^{-2}\sum_{x=0}^{n}\frac{2^x}{x!}\geqq 0.95$$

$$\sum_{x=0}^{n}\frac{2^x}{x!}\geqq 7.01$$

ここで，$f_n=\sum_{x=0}^{n}\frac{2^x}{x!}$ とおくと，

$$f_3=6.3, \qquad f_4=7, \qquad f_5=7.2$$

よって，$n\geqq 5$. 最低 **5 冊**仕入れねばならない.

エクセルを使って解く

例題 11 （ポアソン分布の当てはめ）

次の表は，高速道路のある地点で観測した車の交通量の度数分布である．

車の数（10秒間ごとの）	0	1	2	3	4	計
観測度数	68	81	38	9	4	200

データから平均と分散を求めよ．この分布にポアソン分布を当てはめたときの理論度数を求めよ．

解　度数分布から，車の数の平均と分散を求めると

$$\bar{x}=\frac{0\times68+1\times81+2\times38+3\times9+4\times4}{200}=\mathbf{1}$$

$$s^2=\frac{0^2\times68+1^2\times81+2^2\times83+3^2\times9+4^2\times4}{200}-1^2=\mathbf{0.89}$$

平均と分散はほぼ等しいので，車の交通量の分布は近似的にポアソン分布に従うことが示唆される．

平均 $\bar{x}$ は λ の推定値を与えるから，このデータに当てはめるポアソン分布は，

$$P(X=x)=\frac{e^{-1}}{x!}\qquad(x=0,1,2,\cdots)$$

理論度数はこれら確率に 200 をかけて求める．

$$200\times\frac{e^{-1}}{0!}\fallingdotseq74,\quad 200\times\frac{e^{-1}}{1!}\fallingdotseq74,\quad 200\times\frac{e^{-1}}{2!}\fallingdotseq37,$$

$$200\times\frac{e^{-1}}{3!}\fallingdotseq12,\quad 200\times\frac{e^{-1}}{4!}\fallingdotseq3$$

よって，

車の数	0	1	2	3	4	計
観測度数	68	81	38	9	4	200
理論度数	74	74	37	12	3	200

［注］　ポアソン分布のデータへの当てはまりが良いか否かは，通常カイ２乗検定によってなされる（9章カイ２乗検定を参照）．

二項分布とポアソン分布と正規分布の関係

2項分布のポアソン分布による近似　2項分布 $B(n, p)$ がポアソン分布で近似できるための一般的条件は

$$n > 50, \quad np \leqq 5$$

二項分布とポアソン分布と正規分布の関係

	A	B	C	D
1	**X**	二項分布	ポアソン分布	正規分布
2	0			
3	1			
4	2			
5	3			
6	4			
99	97			
100	98			
101	99			
102	100			
103				

二項分布とポアソン分布と正規分布の関係

	A	B	C	D
1	X	二項分布	ポアソン分布	正規分布
2	0			
3	1			
4	2			
5	3			
6	4			

=BINOM.DIST(A2,100,0.5,0)

99	97			
100	98			
101	99			
102	100			
103				

二項分布とポアソン分布と正規分布の関係

	A	B	C	D
1	X	二項分布	ポアソン分布	正規分布
2	0			
3	1			
4	2			
5	3			
6	4			

=POISSON.DIST(A2,50,0)

99	97			
100	98			
101	99			
102	100			
103				

二項分布とポアソン分布と正規分布の関係

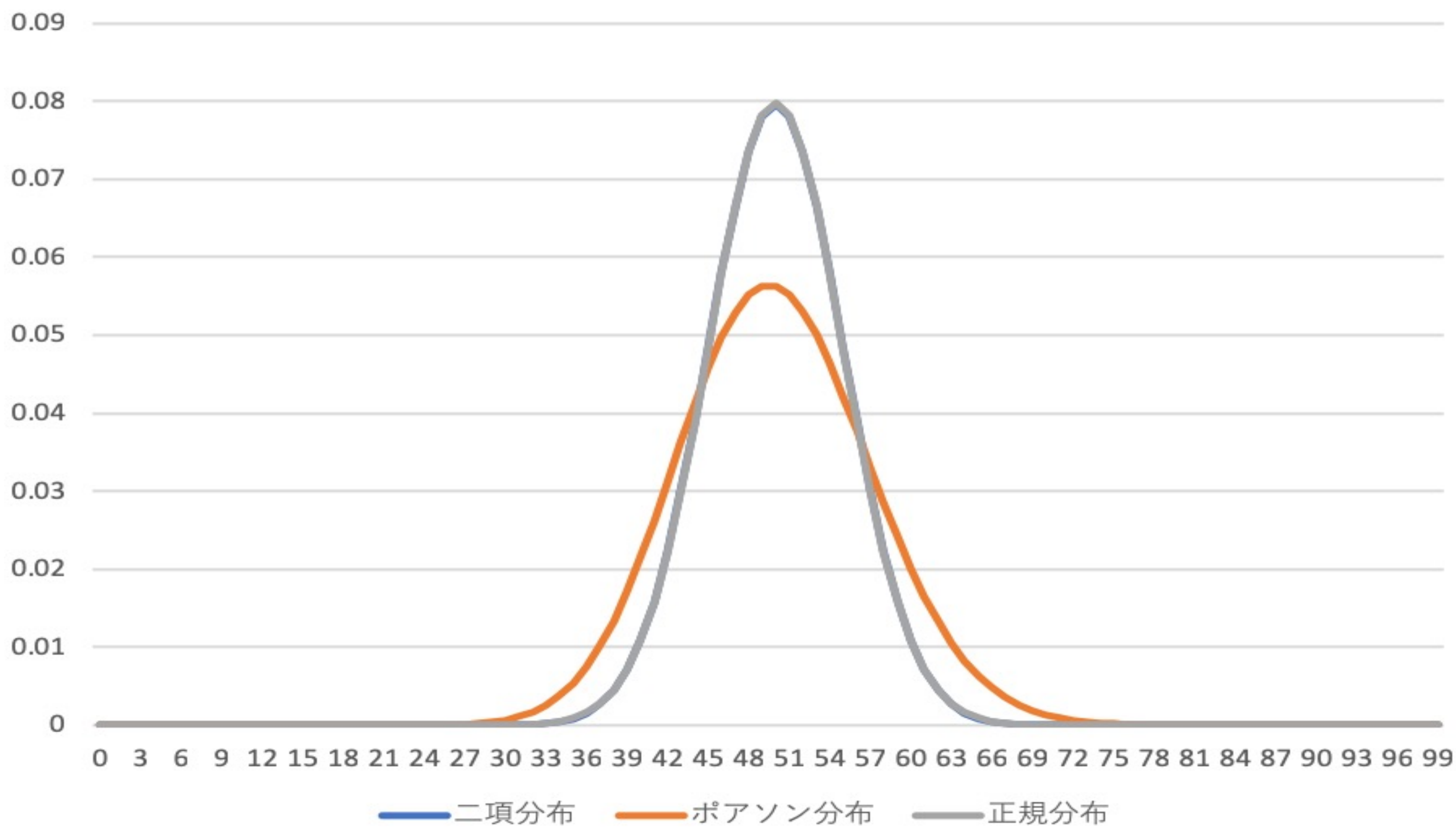
	A	B	C	D
1	X	二項分布	ポアソン分布	正規分布
2	0			
3	1			
4	2			
5	3			
6	4			
99	97			
100	98			
101	99			
102	100			
103				

=NORM.DIST(A2,50,5,0)

二項分布とポアソン分布と正規分布の関係

	A	B	C	D
1	X	二項分布	ポアソン分布	正規分布
2	0	7.88861E-31	1.92875E-22	1.53892E-23
3	1	7.88861E-29	9.64375E-21	1.1146E-22
4	2	3.90486E-27	2.41094E-19	7.75622E-22
5	3	1.27559E-25	4.01823E-18	5.18573E-21
6	4	3.0933E-24	5.02279E-17	3.33118E-20

二項分布とポアソン分布と正規分布の関係



二項分布とポアソン分布と正規分布の関係

2 項分布のポアソン分布による近似　2 項分布 $B(n, p)$ がポアソン分布で近似できるための一般的条件は

$$n > 50, \quad \boxed{np \leqq 5}$$

2項分布のポアソン分布近似

例題 10 (2項分布のポアソン近似)

大都市では平均80人に1人が α 型の血液をもつという.

(a) 無作為に200人の血液提供者を選ぶとき, その中に α 型の血液の人が少なくとも4人含まれる確率を求めよ.

(b) α 型の血液提供者をその中に少なくとも1人含む確率を0.9以上にするには何人を選ばねばならないか.

解　200人中血液が a 型の人の数を X とすると

$$X \sim B\left(200,\ \frac{1}{80}\right)$$

この2項分布は，$n=200$ が大きく，$p=\dfrac{1}{80}$ は十分小さく，$\lambda=np=200\times\dfrac{1}{80}=2.5$ は5より小さいから，$\lambda=2.5$ のポアソン分布

$$P(X=x)=\frac{(2.5)^x e^{-2.5}}{x!} \qquad (x=0,\ 1,\ 2,\ \cdots)$$

で近似できる．よって，

（a）　$P(X \geqq 4)=1-P(0)-P(1)-P(2)-P(3)$

$$=1-e^{-2.5}-2.5e^{-2.5}-\frac{(2.5)^2e^{-2.5}}{2!}-\frac{(2.5)^3e^{-2.5}}{3!}$$

$\fallingdotseq$ **0.24**

（b） $X \sim B\left(n, \frac{1}{80}\right)$. 題意より

$$P(X \geqq 1) \geqq 0.9$$

$$1 - P(X = 0) \geqq 0.9$$

$$\left(\frac{79}{80}\right)^n \leqq 0.1$$

$$n \geqq \frac{\log 0.1}{\log 0.9875} \fallingdotseq 183.1$$

よって，**184 人以上**.

二項分布の正規分布による近似

例題 9 （2 項分布の正規近似）

硬貨を 400 回投げるとき，おもてが 180 回から 210 回まで出る確率を求めよ．

二項分布の正規分布による近似

解　X をおもての出る数とすると，$X \sim B\left(400, \frac{1}{2}\right)$

$$n=400, \quad np=nq=400\times\frac{1}{2}=200\geqq 5, \quad npq=400\times\frac{1}{2}\times\frac{1}{2}=100$$

$n=400$ は十分大きいから，2 項分布 $B\left(400, \frac{1}{2}\right)$ は正規分布 $N(200, 10^2)$ で近似できる．

Y を平均が 200 で，分散が 10^2 の正規変量とすると，

$$Z=\frac{Y-200}{10}\sim N(0, 1)$$

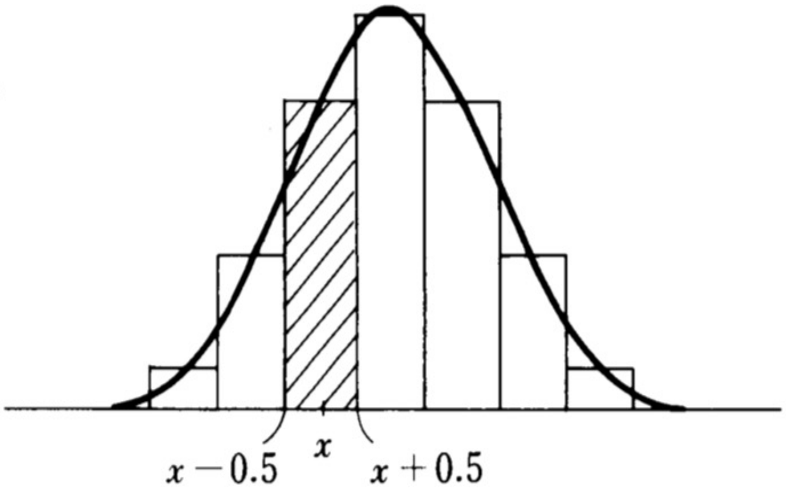
よって求める確率は，半整数補正により

$$\begin{aligned}
P(180\leqq X\leqq 210) &\fallingdotseq P(179.5<Y<210.5) \\
&=P\left(\frac{179.5-200}{10}<Z<\frac{210.5-200}{10}\right) \\
&=P(-2.05<Z\leqq 1.05) \\
&=\Phi(1.05)+\Phi(2.05)-1 \\
&=0.8531+0.9798-1=\mathbf{0.8329}
\end{aligned}$$

半整数補正

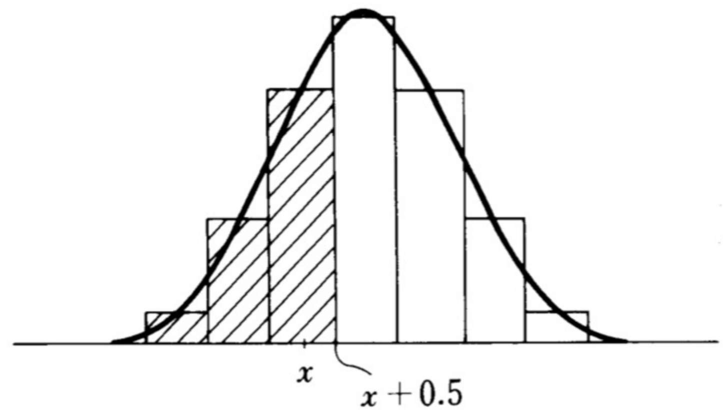
X は整数値のみをとる離散型確率変数で，Y は連続型確率変数とする．2項分布を正規分布で近似するときのように，離散変量 X の分布を連続変量 Y の分布で近似して確率の計算をするとき，

$$P(X=x) \cong P(x-0.5 < Y < x+0.5)$$

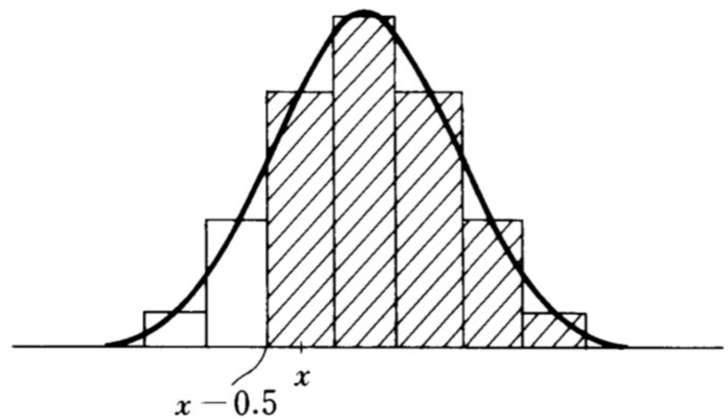


半整数補正

$$P(X \leqq x) \cong P(Y < x + 0.5)$$



$$P(X \geqq x) \cong P(Y > x - 0.5)$$



のように近似することを**半整数補正**（または**連続性の補正**）という．

二項分布

成功確率pのベルヌイ試行をn回行ったとき、x回成功したとすると、成功確率は$\hat{p} = \frac{x}{n}$と推定される。

成功回数xは二項分布し、その
平均はE[x]=np,
分散はVar[X]=np(1-p)
である。

成功確率$\hat{p}$の平均はE[$\hat{p}$]=E[$\frac{X}{n}$]=p,分散は
Var[$\hat{p}$]=V[$\frac{X}{n}$]=$\frac{p(1-p)}{n}$
となる。

- これより

$$z = \frac{\hat{p} - E[X]}{\sqrt{Var[X]}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)\ as\ n \to \infty$$

$$P[-z_{0.95} < z < z_{0.95}] = 0.95$$

$$P\left[-1.96 < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < 1.96\right] = 0.95$$

$$P\left[-1.96\sqrt{\frac{p(1-p)}{n}} < \hat{p} - p < 1.96\sqrt{\frac{p(1-p)}{n}}\right] = 0.95$$

例題 9 （2 項分布の正規近似）

硬貨を 400 回投げるとき，おもてが 180 回から 210 回まで出る確率を求めよ．

解 X をおもての出る数とすると，$X \sim B\left(400, \frac{1}{2}\right)$

$$n=400, \quad np=nq=400\times\frac{1}{2}=200\geqq 5, \quad npq=400\times\frac{1}{2}\times\frac{1}{2}=100$$

$n=400$ は十分大きいから，2 項分布 $B\left(400, \frac{1}{2}\right)$ は正規分布 $N(200, 10^2)$ で近似できる．

Y を平均が 200 で，分散が 10^2 の正規変量とすると，

$$Z=\frac{Y-200}{10} \sim N(0, 1)$$

よって求める確率は，半整数補正により

$$\begin{aligned}
P(180 \leqq X \leqq 210) &\fallingdotseq P(179.5 < Y < 210.5) \\
&= P\left(\frac{179.5-200}{10} < Z < \frac{210.5-200}{10}\right) \\
&= P(-2.05 < Z \leqq 1.05) \\
&= \Phi(1.05)+\Phi(2.05)-1 \\
&= 0.8531+0.9798-1=\mathbf{0.8329}
\end{aligned}$$

例題 10　(2項分布の正規近似)

ある人があるゲームに勝つ確率を $\frac{1}{3}$, 負ける確率を $\frac{2}{3}$ とする. ゲームに勝てば1000円得をし, 負ければ250円損をする. この人がこのゲームを20回行うとき, 少なくとも3000円の得をする確率を求めよ.

解　この人がゲームに勝つ回数を x とすると，20 回のゲームによるこの人の利益は

$$1000x-250(20-x)$$

で，これが 3000 より大きいことから

$$1000x-250(20-x)\geqq 3000$$

$$x\geqq 6.4$$

よって，少なくとも 3000 円の得をするには，20 回中 7 回以上ゲームに勝たねばならない．その確率は $n=20$, $p=\dfrac{1}{3}$ の 2 項分布より

$$\sum_{x=7}^{20} {}_{20}C_x\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{20-x}$$

この確率の計算には 2 項分布の正規近似を使う．

$$\mu=np=20\times\frac{1}{3}\fallingdotseq 6.67,$$

$$\sigma=\sqrt{npq}=\sqrt{20\times\frac{1}{3}\times\frac{2}{3}}\fallingdotseq 2.11,$$

$$Z=\frac{X-\mu}{\sigma}=\frac{X-6.67}{2.11},$$

$$\frac{6.5-6.67}{2.11}=-0.08,\qquad \frac{20.5-6.67}{2.11}=6.55$$

であるから，

$$P(X\geqq 7)=\sum_{x=7}^{20} {}_{20}C_x\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{20-x}$$

$$\fallingdotseq\int_{-0.08}^{6.55}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\fallingdotseq\Phi(0.08)=\mathbf{0.532}$$

テレビ視聴率

テレビ視聴率は，視聴率の高い番組ほど多くの視聴者が見ているので，広告宣伝の効果が高く影響力が 強いと考えられている．このため，視聴率の高さが広告宣伝費用に反映されるので，テレビ会社は高い視聴率を 得ようとして番組を製作している．

ある調査会社のデータによると，関東地区では 600 世帯を対象にしているようである． NHK 大河ドラマの関東地区世帯視聴率は26.2%であった． 真の世帯視聴率の 95 %信頼区間を求めよ．

解答：

p の分散推定値は $s^2 = p^\wedge(1 - p^\wedge)/n = 0.262*(1 - 0.262)/600 = 0.262*0.738/600 = 0.00032226$

p の標準偏差の推定値（標準誤差）は $s = \sqrt{0.00032226} = 0.01795$

95 %信頼区間の幅は $d = z_0 s = 1.96*0.01795 = 0.035$

下限は，$p^\wedge - d = 0.262 - 0.035 = 0.227$，　上限は，$p^\wedge + d = 0.262 + 0.035 = 0.297$

よって，<u>$0.227 < p < 0.297$</u> である．

カイ二乗検定

N回のベルヌイ試行でX回成功した時に、成功率がp_0であるという。

- 帰無仮説：$p = p_0$
- 対立仮説：$p \neq p_0$

$$z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \sim N(0,1)\ as\ n \to \infty$$

$$z^2 = \frac{(X - np_0)^2}{np_0(1 - p_0)} \sim \chi^2(1)\ as\ n \to \infty$$

	成功	失敗
観測度数	X	$n-X$
期待度数	np_0	$n\,(1\text{-}p_0)$

$$\chi^2 = \frac{(X-np_0\,)^2}{np_0} + \frac{(n-X-n(1-p_0)\,)^2}{n(1-p_0)}$$

$$\chi^2 = \frac{(X-np_0\,)^2}{np_0(1-p_0)} = z^2 \sim \chi^2(1) \quad as\; n \to \infty$$

χ^2値は、試行回数nが大きくなるにつれて帰無仮説のもとで自由度1のχ^2分布に従うと近似できる

例題A 君と B 君が将棋を行った. 10 局やったところ, A 君の 7 勝 3 敗であった. A 君と B 君で将棋の強さに 違いがあるか検定せよ。また, 30 局やって, A 君の 21 勝 9 敗であったとき（勝率は 7 割で先ほどと同じ）ではどうか.

帰無仮説として, A 君と B 君の将棋の強さが等しい, とする.
すなわち, A 君の勝率が 0.5 であるとする. よって, H_0: $p = 0.5$である。帰無仮説のもとでA君の勝ち負けの期待度数は、5勝5敗であるので

	勝ち	負け
観測度数	7	3
期待度数	5	5

$$\chi^2 = \frac{(7-5)^2}{5} + \frac{(3-5)\ )^2}{5} = 1.6$$

二項分布を用いると

例題A 君と B 君が将棋を行った．10 局やったところ，A 君の 7 勝 3 敗であった．A 君と B 君で将棋の強さに 違いがあるか検定せよ。また，30 局やって，A 君の 21 勝 9 敗であったとき（勝率は 7 割で先ほどと同じ）ではどうか．

帰無仮説として，A 君と B 君の将棋の強さが等しい，とする．
すなわち，A 君の勝率が 0.5 であるとする． よって，H_0：p ＝ 0.5である。帰無仮説のもとでA君の勝ち負けの期待度数は、5勝5敗であるので

（b） 粒子の数は $\lambda=2.07$ のポアソン分布に従うという帰無仮説の下で，期待度数を次のように求める．

値 x	0	1	2	3	4	5	6 以上	計
確率 $e^{-2.07}\dfrac{2.07^x}{x!}$	0.13	0.26	0.27	0.19	0.10	0.04	0.01	1.0
期待度数＝300×確率	39	78	81	57	30	12	3	300
観測度数	38	75	89	54	20	19	5	300

これから

$$\chi^2=\frac{(38-40.6)^2}{40.6}+\frac{(75-81.2)^2}{81.2}+\cdots+\frac{(5-5.0)^2}{5.0}=9.39$$

$\alpha=0.05$, $\nu=7-1=6$ より $\chi^2_{0.05}(6)=12.59$. $\chi^2=9.39$ はこの値を超えないから，仮説は採択である．よってこのデータはポアソン分布に**適合している**．

例題 3 （ポアソン分布の適合度検定）

大気中に浮遊するある微小な物質の量を推定するため，空間内にいくつかの点を選び，その点のまわりの単位体積内の粒子数を計測する．いま300点を選んで観測した結果，つぎのデータが得られた．

（a） 粒子の数の平均と分散を求めよ．

（b） このデータにポアソン分布をあてはめ，その適合性を調べよ．

粒子の数	0	1	2	3	4	5	6以上	計
観測度数	38	75	89	54	20	19	5	300

解　(a)　粒子数を x，度数を f とすると，平均は

$$\bar{x}=\frac{\sum x_i f_i}{n}=\frac{0\times 38+1\times 75+\cdots+6\times 5}{300}=\frac{31}{15}\fallingdotseq \mathbf{2.07}$$

分散は

$$s^2=\frac{\sum x_i{}^2 f_i}{n}-\bar{x}^2=\frac{0^2\times 38+1^2\times 75+\cdots+6^2\times 5}{300}-\left(\frac{31}{15}\right)^2\fallingdotseq \mathbf{2.04}$$

平均と分散がほぼ等しいので，粒子の数の分布としてポアソン分布が予想される．