二項分布とその極限
－正規分布とポアソン分布－
［改訂版］

渡邉 俊夫

二項分布

発生確率が p の事象を独立に n 回試行したとき、その事象が k 回起こる確率は

$$P(k) = \frac{n(n-1)\cdot\cdots\cdot(n-k+1)}{k(k-1)\cdot\cdots\cdot 1}p^k(1-p)^{n-k}$$

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

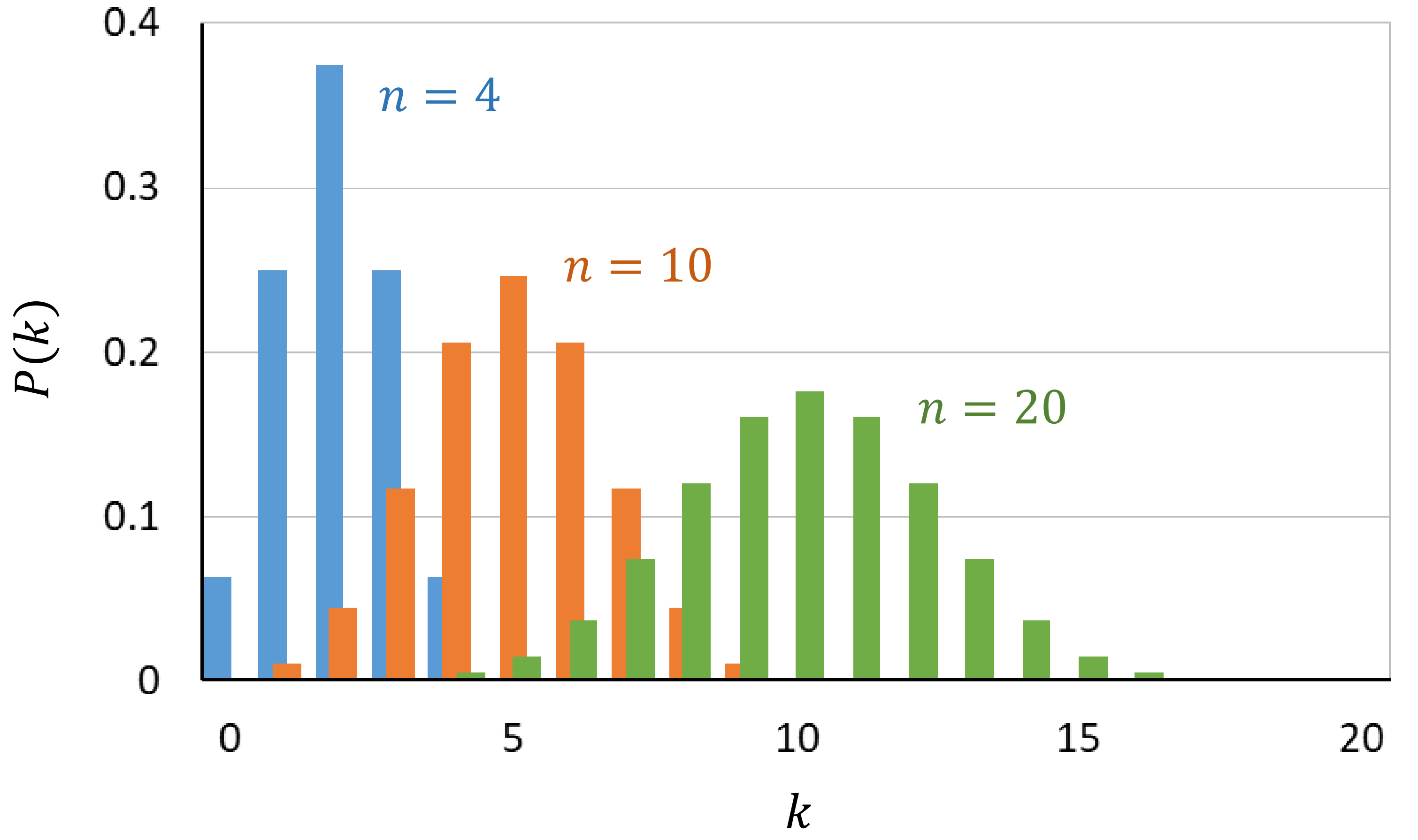
で表される。この離散確率分布を二項分布という。ここで、

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

は異なる n 個のうちから k 個を（重複なく）選ぶ組合せの数である。

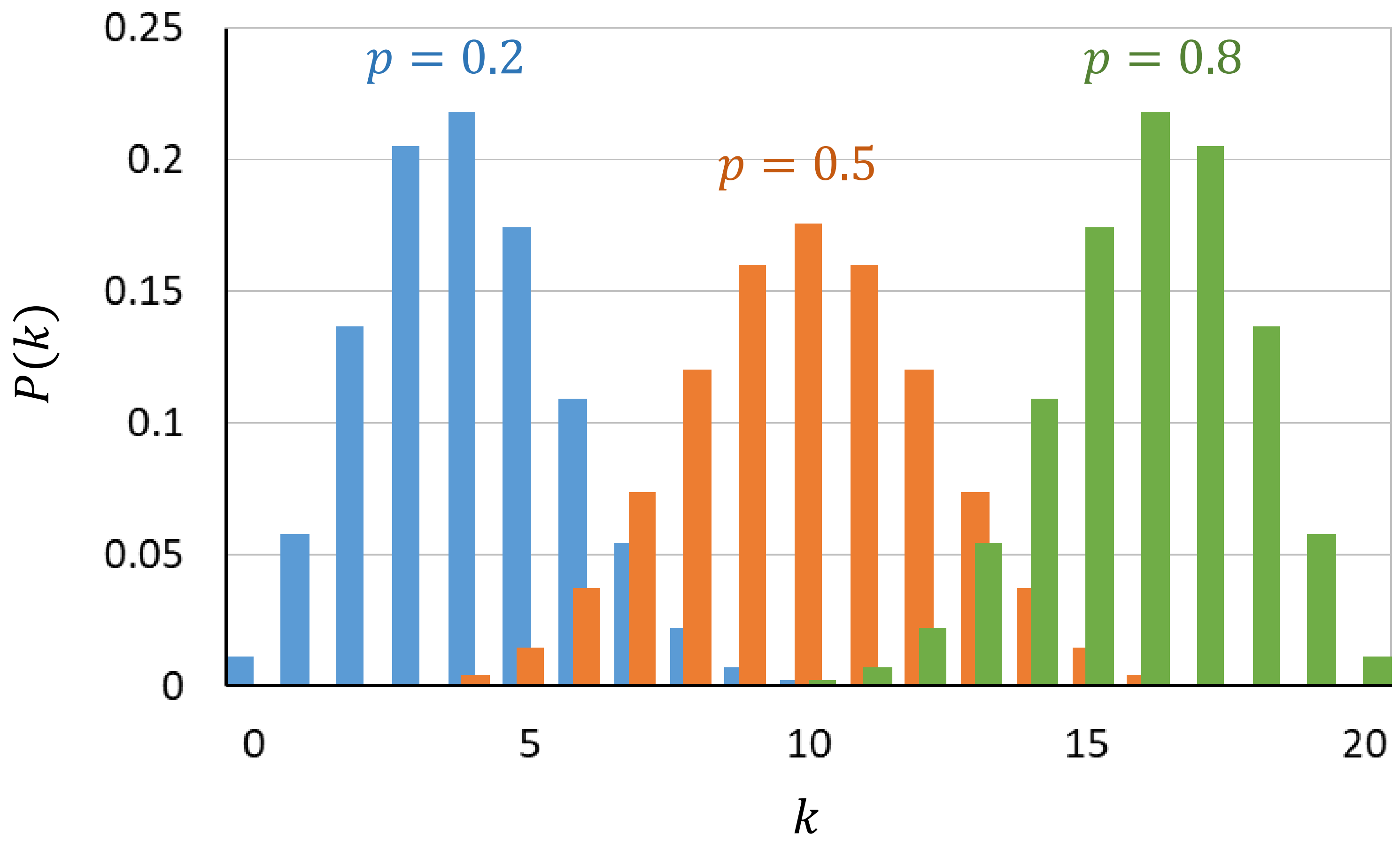
二項分布

$p=0.5$ のとき、$n=4,10,20$ の二項分布は下図のようになる。



二項分布

$n = 20$ のとき、$p = 0.2, 0.5, 0.8$ の二項分布は下図のようになる。



二項分布の確率の総和

二項分布の確率の総和は

$$\sum_{k=0}^{n} P(k) = \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= \left(p + (1-p)\right)^n = 1$$

になっている。ここで、二項定理

$$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} a^k b^{n-k}$$

を用いた。

二項分布の平均

二項分布における発生回数 k の平均は

$$\bar{k} = \sum_{k=0}^{n} kP(k) = \sum_{k=0}^{n} k \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n \cdot (n-1)!}{(k-1)!\,\left((n-1)-(k-1)\right)!} p \cdot p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$$= np \sum_{l=0}^{n-1} \frac{(n-1)!}{l!\,\left((n-1)-l\right)!} p^l (1-p)^{(n-1)-l}$$

$$= np \cdot \left(p + (1-p)\right)^{n-1} = np$$

である。

二項分布の2乗平均

二項分布における発生回数 k の2乗平均は

$$\overline{k^2} = \sum_{k=0}^{n} k^2 P(k) = \sum_{k=0}^{n} k(k-1)P(k) + \sum_{k=0}^{n} kP(k)$$

$$= \sum_{k=0}^{n} k(k-1)\frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k} + \bar{k}$$

$$= \sum_{k=2}^{n} \frac{n!}{(k-2)!\,(n-k)!}p^k(1-p)^{n-k} + np$$

二項分布の2乗平均

ここで、第1項は

$$\sum_{k=0}^{n} k(k-1)P(k) = \sum_{k=2}^{n} \frac{n!}{(k-2)!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=2}^{n} \frac{n(n-1)\cdot(n-2)!}{(k-2)!\big((n-2)-(k-2)\big)!} p^2 \cdot p^{k-2} (1-p)^{(n-2)-(k-2)}$$

$$= n(n-1)p^2 \sum_{l=0}^{n-2} \frac{(n-2)!}{l!\big((n-2)-l\big)!} p^l (1-p)^{(n-2)-l}$$

$$= n(n-1)p^2 \cdot \big(p+(1-p)\big)^{n-2} = n(n-1)p^2$$

となる。

二項分布の2乗平均

したがって、二項分布における発生回数 k の2乗平均は

$$\overline{k^2} = \sum_{k=0}^{n} k^2 P(k) = \sum_{k=0}^{n} k(k-1)P(k) + \sum_{k=0}^{n} kP(k)$$

$$= n(n-1)p^2 + np$$

$$= n^2p^2 + np(1-p)$$

である。

二項分布の分散

分散は「2乗平均」と「平均の2乗」との差で求められるから、
二項分布における発生回数 k の分散は

$$
\begin{aligned}
V(k) &= \overline{k^2} - \bar{k}^2 \\
&= \left(n^2p^2 + np(1-p)\right) - (np)^2 \\
&= np(1-p)
\end{aligned}
$$

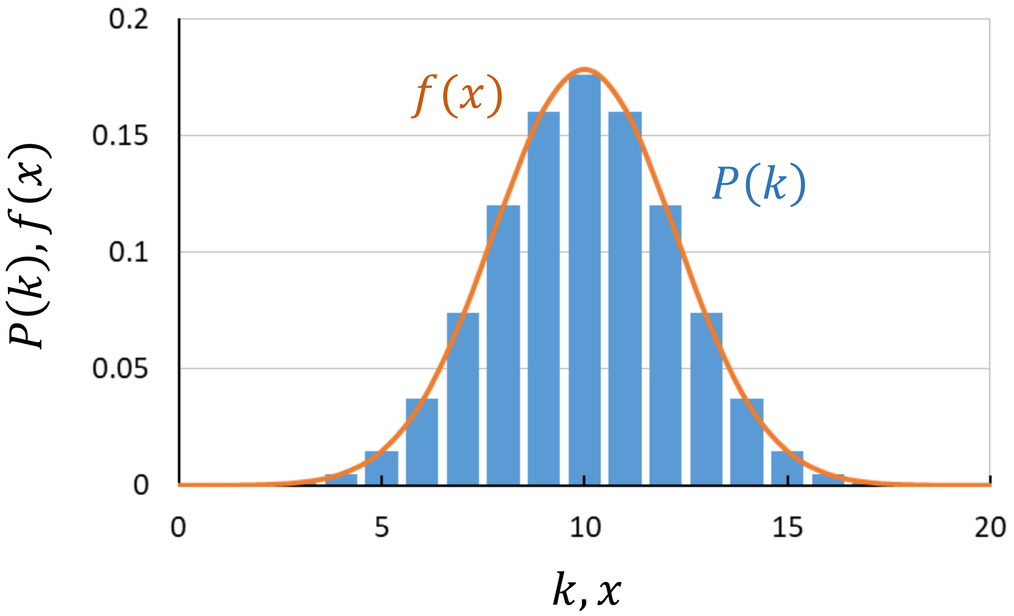
である。

正規分布

二項分布の確率は試行回数 n が大きく、かつ $\bar{k} = np \gg 1$ のとき、

$$P(k) \approx f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

に近づく。ここで、$f(x)$ は x を連続変数とする確率密度関数であり、$\mu = np$、$\sigma^2 = np(1-p)$ である。この確率分布を正規分布という。



$n = 20$
$p = 0.5$

$\mu = np = 10$
$\sigma^2 = np(1-p) = 5$

二項分布から正規分布へ

二項分布

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

を、発生確率 p の代わりに、偏差（平均との差） $t = k - \mu = k - np$ で表すと

$$p = \frac{k-t}{n}$$

より

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

二項分布から正規分布へ

$$= \frac{n!}{k!\,(n-k)!}\frac{1}{n^n}(k-t)^k(n-k+t)^{n-k}$$

$$= \frac{n!}{k!\,(n-k)!}\frac{k^k(n-k)^{n-k}}{n^n}\left(1-\frac{t}{k}\right)^k\left(1+\frac{t}{n-k}\right)^{n-k}$$

$$= \frac{n!}{n^n}\cdot\frac{k^k}{k!}\cdot\frac{(n-k)^{n-k}}{(n-k)!}\left(1-\frac{t}{k}\right)^k\left(1+\frac{t}{n-k}\right)^{n-k}$$

となる。

二項分布から正規分布へ

$n \gg 1,\ k \approx \bar{k} = np \gg 1,\ n - k \gg 1$ かつ $t \ll k,\ t \ll n - k$ の場合を考える。

$N \gg 1$ で成り立つスターリングの近似（付録を参照）

$$N! = \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$$

および $x \ll 1$ で成り立つ展開

$$(1+x)^a = \exp[\log(1+x)^a] = \exp[a\log(1+x)]$$
$$= \exp\left[a\left(x - \frac{x^2}{2} + \cdots\right)\right]$$

を用いると

二項分布から正規分布へ

$$P(t) \cong \frac{\sqrt{2\pi n}}{e^n} \cdot \frac{e^k}{\sqrt{2\pi k}} \cdot \frac{e^{n-k}}{\sqrt{2\pi(n-k)}}$$

$$\times \exp\left[k\left(-\frac{t}{k} - \frac{1}{2}\left(\frac{t}{k}\right)^2 - \cdots\right)\right]$$

$$\times \exp\left[(n-k)\left(\frac{t}{n-k} - \frac{1}{2}\left(\frac{t}{n-k}\right)^2 + \cdots\right)\right]$$

$$\cong \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \exp\left[-t - \frac{t^2}{2k}\right] \cdot \exp\left[t - \frac{t^2}{2(n-k)}\right]$$

二項分布から正規分布へ

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \exp\left[-\frac{t^2}{2k} - \frac{t^2}{2(n-k)}\right]$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \exp\left[-\frac{nt^2}{2k(n-k)}\right]$$

ここで、$\dfrac{k(n-k)}{n} \approx \dfrac{pn(n-pn)}{n} = np(1-p) = \sigma^2$ より

$$P(t) \cong \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{t^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2}$$

二項分布から正規分布へ

偏差（平均との差）$t = k - \mu$ を k に戻せば

$$P(k) \cong \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(k-\mu)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}\sigma} e^{-(k-\mu)^2/2\sigma^2}$$

となる。

さらに、離散変数 k を連続変数 x に変換すれば

$$P(k) \approx f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

となり、正規分布が得られる。

正規分布の確率の総和

正規分布の確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

において、$X = (x-\mu)/\sigma$ とおくと

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx \right)^2 = \left(\int_{-\infty}^{\infty} e^{-X^2/2} \sigma dX \right)^2$$

$$= \sigma^2 \int_{-\infty}^{\infty} e^{-X^2/2} dX \int_{-\infty}^{\infty} e^{-Y^2/2} dY = \sigma^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(X^2+Y^2)/2} dX\, dY$$

$$= \sigma^2 \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2/2} r dr = \sigma^2 \cdot 2\pi \left[\left(-e^{-r^2/2} \right) \right]_0^{\infty} = 2\pi\sigma^2$$

正規分布の確率の総和

したがって、正規分布の確率の総和は

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}dx = \frac{1}{\sqrt{2\pi}\sigma}\cdot\sqrt{2\pi\sigma^2} = 1$$

になっている。

正規分布の平均

正規分布の平均は

$$\bar{x} = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} xe^{-(x-\mu)^2/2\sigma^2}dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu)e^{-(x-\mu)^2/2\sigma^2}dx + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mu e^{-(x-\mu)^2/2\sigma^2}dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} Xe^{-X^2/2}dX + \mu\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-X^2/2}\,dX$$

$$= \mu$$

である。

正規分布の分散

正規分布の分散は

$$V(x) = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X^2 e^{-X^2/2} dX = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X \cdot X e^{-X^2/2} dX$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left[X \cdot \left(-e^{-X^2/2} \right) \right]_{-\infty}^{\infty} - \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(-e^{-X^2/2} \right) dX$$

$$= \sigma^2 \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-X^2/2} dX = \sigma^2$$

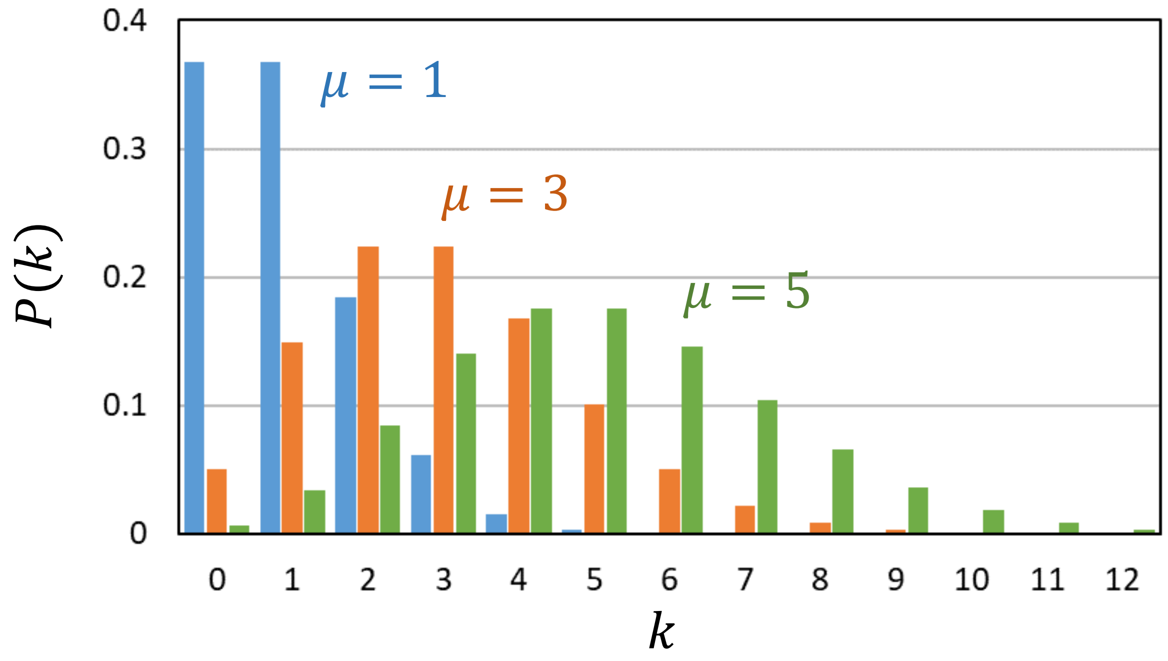
である。

ポアソン分布

二項分布の確率は試行回数 n が大きく、かつ発生確率 p が小さいとき、

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

に近づく。ここで、$\mu = np$ である。この確率分布を**ポアソン分布**という。



二項分布からポアソン分布へ

二項分布

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

を、発生確率 p の代わりに、平均値 $\mu = np$ で表すと

$$p = \frac{\mu}{n}$$

より

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

$$= \frac{n(n-1)\cdot \cdots \cdot 1}{k!} \frac{\mu^k}{n^k} \left(1 - \frac{\mu}{n}\right)^{n-k}$$

二項分布からポアソン分布へ

$$= \frac{n(n-1)\cdot\cdots\cdot(n-k+1)}{n^k}\cdot\frac{\mu^k}{k!}\left(1-\frac{\mu}{n}\right)^{n-k}$$

$$= 1\cdot\left(1-\frac{1}{n}\right)\cdot\cdots\cdot\left(1-\frac{k-1}{n}\right)\cdot\frac{\mu^k}{k!}\left(1-\frac{\mu}{n}\right)^{n-k}$$

$p \ll 1$ の場合、$k \approx \bar{k} = \mu = np \ll n$ だから

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

$x \ll 1$ で成り立つ近似式 $1 - x \cong e^{-x}$ を用いると

$$P(k) \cong \frac{\mu^k}{k!}\left(e^{-\mu/n}\right)^{n} = \frac{\mu^k}{k!}e^{-\mu}$$

となり、ポアソン分布が得られる。

ポアソン分布の確率の総和

ポアソン分布の確率の総和は

$$\sum_{k=0}^{\infty} P(k) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = 1$$

になっている。ここで、e^x のマクローリン展開

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

を用いた。

ポアソン分布の平均

ポアソン分布における発生回数 k の平均は

$$\bar{k} = \sum_{k=0}^{\infty} kP(k) = \sum_{k=0}^{\infty} k\frac{\mu^k}{k!}e^{-\mu} = e^{-\mu}\sum_{k=0}^{\infty} k\frac{\mu^k}{k!}$$

$$= e^{-\mu}\sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} = \mu e^{-\mu}\sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!}$$

$$= \mu$$

である。

ポアソン分布の2乗平均

ポアソン分布における発生回数 k の2乗平均は

$$\overline{k^2} = \sum_{k=0}^{\infty} k^2 P(k) = \sum_{k=0}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu}$$

$$= e^{-\mu} \sum_{k=0}^{\infty} k(k-1) \frac{\mu^k}{k!} + e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!}$$

$$= e^{-\mu} \sum_{k=2}^{\infty} \frac{\mu^k}{(k-2)!} + e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!}$$

$$= \mu^2 + \mu$$

である。

ポアソン分布の分散

分散は「2乗平均」と「平均の2乗」との差で求められるから、
ポアソン分布における発生回数 k の分散は

$$
\begin{aligned}
V(k) &= \overline{k^2} - \bar{k}^2 \\
&= (\mu^2 + \mu) - \mu^2 \\
&= \mu
\end{aligned}
$$

である。

まとめ

- 発生確率が p の事象を独立に n 回試行したとき、その事象が k 回起こる確率は、二項分布

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

にしたがう。二項分布における発生回数 k の平均は $\bar{k} = np$、分散は $V(k) = np(1-p)$ である。

- 二項分布の確率は試行回数 n が大きく、かつ $\bar{k} = np \gg 1$ のとき、x を連続変数とする正規分布の確率密度関数に近づき、

$$P(k) \approx f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

で表される。ここで、$\mu = np$、$\sigma^2 = np(1-p)$ である。

まとめ

- 二項分布の確率は試行回数 n が大きく、かつ発生確率 p が小さいとき、平均 $\bar{k} = np = \mu$ のポアソン分布

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

に近づく。ポアソン分布における発生回数 k の分散は $V(k) = \mu$ である。

付録:スターリングの近似

ガンマ関数の積分表示より

$$n! = \Gamma(n+1) = \int_0^\infty t^n \mathrm{e}^{-t}\mathrm{d}t = \int_0^\infty \mathrm{e}^{-t+n\log t}\mathrm{d}t = \int_0^\infty \mathrm{e}^{-f(t)}\mathrm{d}t$$

ここで、$f(t) = t - n\log t$ である。

$f(t)$ を $t = n$ でテイラー展開すると

$$f'(t) = 1 - \frac{n}{t},\quad f''(t) = \frac{n}{t^2},\quad f'''(t) = -\frac{2n}{t^3}$$

より

$$f(t) = f(n) + \frac{1}{2n}(t-n)^2 - \frac{1}{3n^2}(t-n)^3 + \cdots$$

付録:スターリングの近似

$n \gg 1$ に対して、$\mathrm{e}^{-f(t)}$ は $t = n$ の付近でのみ大きな値をもつから

$$n! = \int_0^\infty \mathrm{e}^{-f(t)}\mathrm{d}t \approx \mathrm{e}^{-f(n)} \int_{-\infty}^\infty \mathrm{e}^{-(t-n)^2/2n}\mathrm{d}t$$

$$= \mathrm{e}^{-n+n\log n}\sqrt{2\pi n} = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

を得る。

参考文献:小野寺嘉孝「物理のための応用数学」裳華房、1988