

二項分布とその極限

—正規分布とポアソン分布—

渡邊 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

$$\begin{aligned} P(k) &= \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} p^k (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = {}_n C_k p^k (1-p)^{n-k} \end{aligned}$$

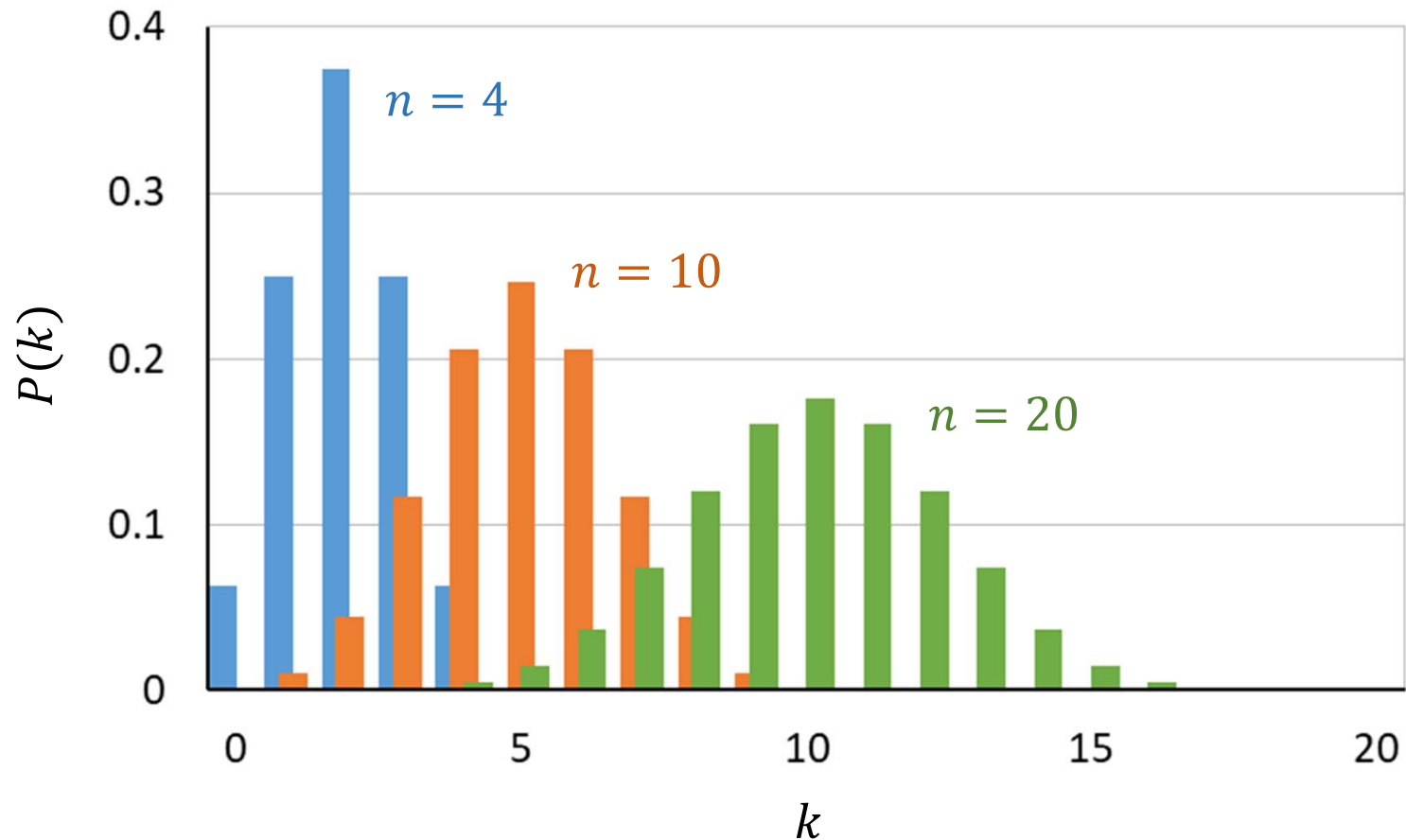
で表される。この離散確率分布を**二項分布**という。ここで、

$${}_n C_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

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。

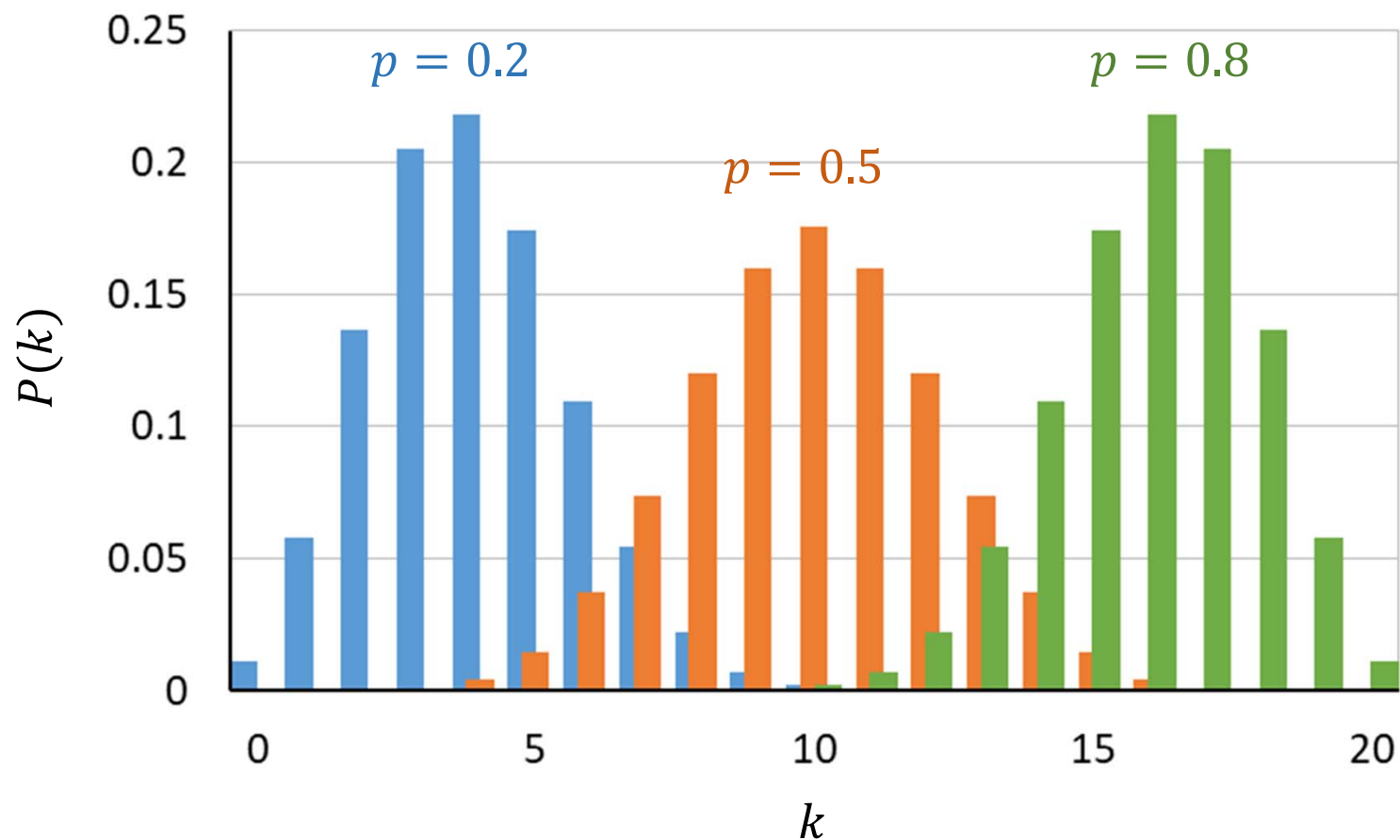
二項分布

$p = 0.5$ 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

$$\begin{aligned}\sum_{k=0}^n P(k) &= \sum_{k=0}^n \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k} \\ &= (p + (1-p))^n = 1\end{aligned}$$

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

$$\begin{aligned}\bar{k} &= \sum_{k=0}^n kP(k) = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n \cdot (n-1)!}{(k-1)!((n-1)-(k-1))!} p \cdot p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{k'=0}^{n-1} \frac{(n-1)!}{k'!((n-1)-k')!} p^{k'} (1-p)^{(n-1)-k'} \\ &= np \cdot (p + (1-p))^{n-1} = np\end{aligned}$$

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

$$\begin{aligned}\overline{k^2} &= \sum_{k=0}^n k^2 P(k) = \sum_{k=0}^n k(k-1)P(k) + \sum_{k=0}^n kP(k) \\ &= \sum_{k=0}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \bar{k} \\ &= \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} + np\end{aligned}$$

二項分布の2乗平均

ここで、第1項は

$$\begin{aligned} \sum_{k=0}^n k(k-1)P(k) &= \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=2}^n \frac{n(n-1) \cdot (n-2)!}{(k-2)!((n-2)-(k-2))!} p^2 \cdot p^{k-2} (1-p)^{(n-2)-(k-2)} \\ &= n(n-1)p^2 \sum_{k'=0}^{n-2} \frac{(n-2)!}{k'!((n-2)-k')!} p^{k'} (1-p)^{(n-2)-k'} \\ &= n(n-1)p^2 \cdot (p + (1-p))^{n-2} = n(n-1)p^2 \end{aligned}$$

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

$$\begin{aligned}\overline{k^2} &= \sum_{k=0}^n k^2 P(k) = \sum_{k=0}^n k(k-1)P(k) + \sum_{k=0}^n kP(k) \\ &= n(n-1)p^2 + np \\ &= n^2p^2 + np(1-p)\end{aligned}$$

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

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

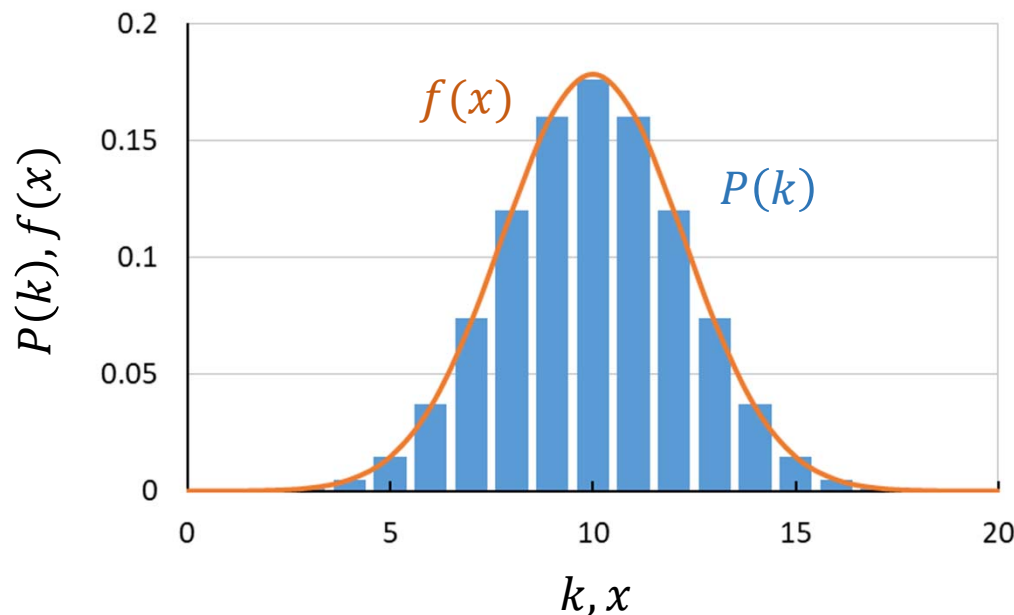
である。

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。



$$\begin{aligned}n &= 20 \\p &= 0.5 \\ \mu &= np = 10 \\ \sigma^2 &= np(1-p) = 5\end{aligned}$$

二項分布から正規分布へ

二項分布

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

より

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

であるから

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

二項分布から正規分布へ

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

より

$$\begin{aligned} \frac{P(k+1)}{P(k)} - 1 &= \frac{(n-k)p}{(k+1)(1-p)} - 1 = \frac{(n-k)p - (k+1)(1-p)}{(k+1)(1-p)} \\ &= \frac{(n+1)p - (k+1)}{(k+1)(1-p)} = -\frac{(k+1) - (n+1)p}{(k+1)(1-p)} \end{aligned}$$

ここで、 $n \gg 1, k \gg 1$ とすると

$$\frac{P(k+1)}{P(k)} - 1 \approx -\frac{k - np}{k(1-p)}$$

二項分布から正規分布へ

$$\frac{P(k+1)}{P(k)} - 1 \approx -\frac{k - np}{k(1-p)}$$

$\delta = k - np$ とおき、 $|\delta| \ll np$ とすると

$$\frac{P(k+1)}{P(k)} - 1 \approx -\frac{\delta}{(np + \delta)(1-p)} = -\frac{\delta}{np \left(1 + \frac{\delta}{np}\right) (1-p)}$$

$$= -\frac{\delta}{np(1-p)} \left(1 - \frac{\delta}{np} + \dots\right)$$

$$\approx -\frac{\delta}{np(1-p)} = -\frac{k - np}{np(1-p)}$$

二項分布から正規分布へ

$$\frac{P(k+1)}{P(k)} - 1 \approx -\frac{k-np}{k(1-p)} \approx -\frac{k-np}{np(1-p)}$$

二項分布の平均 $\mu = np$ 、分散 $\sigma^2 = np(1-p)$ を用いると

$$\frac{P(k+1)}{P(k)} - 1 \approx -\frac{k-\mu}{\sigma^2}$$

$$\therefore \frac{P(k+1) - P(k)}{P(k)} \approx -\frac{k-\mu}{\sigma^2}$$

離散分布 $P(k)$ を連続分布 $f(x)$ に変換すると

$$\frac{1}{f(x)} \cdot \frac{f(x+\Delta x) - f(x)}{\Delta x} = -\frac{x-\mu}{\sigma^2}$$

二項分布から正規分布へ

$$\frac{1}{f(x)} \cdot \frac{f(x + \Delta x) - f(x)}{\Delta x} = -\frac{x - \mu}{\sigma^2}$$

$\Delta x \rightarrow 0$ とすると

$$\frac{1}{f(x)} \cdot \frac{df(x)}{dx} = -\frac{x - \mu}{\sigma^2}$$

両辺を積分して、 $x = \mu$ で $f(\mu) = C$ とすると

$$\log f(x) - \log C = -\frac{(x - \mu)^2}{2\sigma^2}$$

$$\therefore f(x) = C e^{-(x-\mu)^2/2\sigma^2}$$

となる。

二項分布から正規分布へ

$f(x) = C e^{-(x-\mu)^2/2\sigma^2}$ の定数 C は

$$\int_{-\infty}^{\infty} f(x) dx = C \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = C\sigma \int_{-\infty}^{\infty} e^{-X^2/2} dX = C\sigma I = 1$$

より定まる。ここで、 $X = (x - \mu)/\sigma$ であり、定積分 I の2乗は

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-X^2/2} dX \right)^2 = \int_{-\infty}^{\infty} e^{-X^2/2} dX \int_{-\infty}^{\infty} e^{-Y^2/2} dY \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(X^2+Y^2)/2} dX dY = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2/2} r dr = 2\pi \left[(-e^{-r^2/2}) \right]_0^{\infty} = 2\pi \end{aligned}$$

二項分布から正規分布へ

$$\int_{-\infty}^{\infty} f(x) dx = C \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = C\sigma \int_{-\infty}^{\infty} e^{-X^2/2} dX = 1$$

および

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

より、 $C = \frac{1}{\sqrt{2\pi}\sigma}$ であるから

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

となり、正規分布が得られる。

正規分布の平均

正規分布の平均は

$$\begin{aligned}\bar{x} &= \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu) e^{-(x-\mu)^2/2\sigma^2} dx + \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \mu e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X e^{-X^2/2} dX + \mu \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-X^2/2} dX \\ &= \mu\end{aligned}$$

である。

正規分布の分散

正規分布の分散は

$$\begin{aligned} V(x) &= \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X^2 e^{-X^2/2} dX = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X \cdot X e^{-X^2/2} dX \\ &= \frac{\sigma^2}{\sqrt{2\pi}} [X \cdot (-e^{-X^2/2})]_{-\infty}^{\infty} - \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-e^{-X^2/2}) dX \\ &= \sigma^2 \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-X^2/2} dX = \sigma^2 \end{aligned}$$

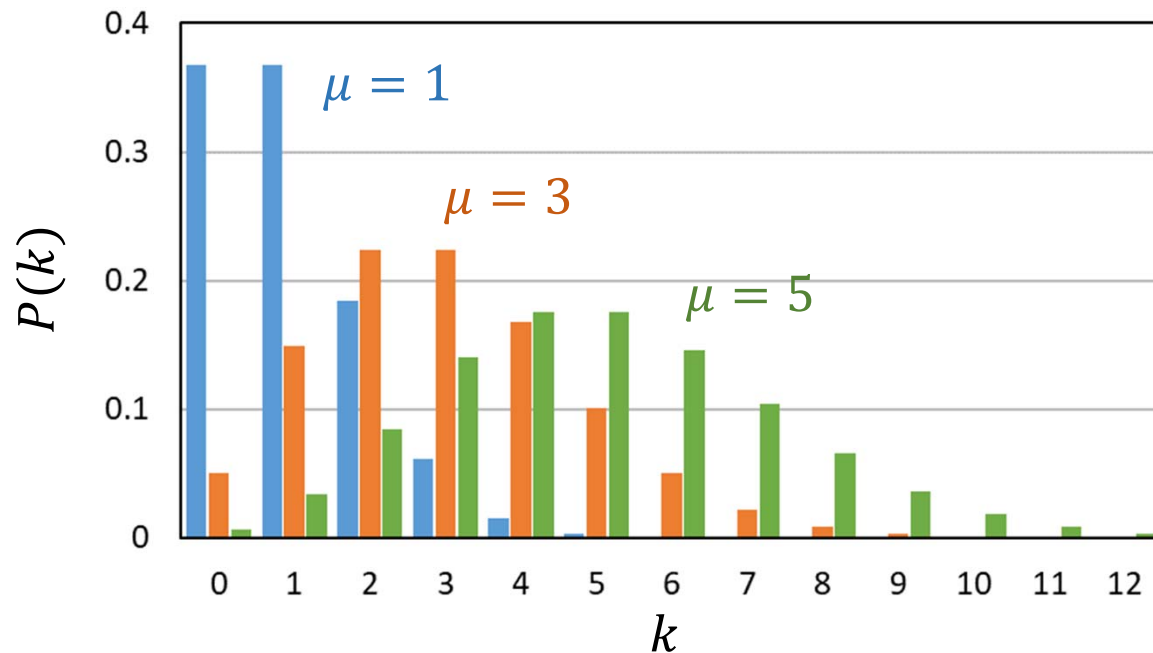
である。

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

$$\begin{aligned} P(k) &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \end{aligned}$$

において $p \ll 1$ とすると、 $n \gg k$ だから

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

ここで、 $\mu = np$ とおくと

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

二項分布からポアソン分布へ

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

において、 e の定義

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$$

を用いると

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

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

$$\begin{aligned}\bar{k} &= \sum_{k=0}^{\infty} kP(k) = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} \\ &= e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \\ &= \mu\end{aligned}$$

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

$$\begin{aligned}\overline{k^2} &= \sum_{k=0}^{\infty} k^2 P(k) = \sum_{k=0}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu} \\ &= e^{-\mu} \sum_{k=0}^{\infty} k(k-1) \frac{\mu^k}{k!} + e^{-\mu} \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} \\ &= e^{-\mu} \sum_{k=2}^{\infty} \frac{\mu^k}{(k-2)!} + e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} \\ &= \mu^2 + \mu\end{aligned}$$

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。

参考文献

- ・薩摩順吉「確率・統計」(理工系の数学入門コース)岩波書店、1989.