

III-2-3.ポアソン分布

ポアソン分布とは、稀におこる現象がある条件の中で、 k 回起こる確率を表した離散確率分布曲線です。つまり、あらかじめ何かの情報で、ある条件でそれが起こる確率を期待値として知っていて、それが実際にその条件で何回 (k) 起きるか、 k の確率を表す曲線です。稀に起こるといのが一つのポイントで、ポアソン分布は一種の近似計算だから、頻繁の起こる現象には使えないのです。頻繁に起こる現象にポアソン分布を使わなければならない理由ありません。期待値を知っているならば、何も統計解析する必要はないだろうという意見もありそうです。しかし、世の中はそういうものではなくて、例えば、稀にしか起こらない地震が、今後 1 週間の内に、3 回以上起こる可能性はどのくらいあるとか、今後 10 年間で 3 人以上の人が、熊に食べられる可能性はどのくらいあるとか、いろいろ、使える場面はいろいろあります。ポアソン分布という名前は、それを考えた数学者、シメオン・ドニ・ポアソンの名前に由来します。私は、しばらく、ポアソンをフランス語の魚だと思っていて、確かに魚はめったに釣れないなと思っていました。しかし、めったに釣れないのは私だけで、上手な人は、結構連れますから、ポアソン分布は魚に由来していません。ポアソン分布の例として、紹介されるのは、多くの場合、ある一定時間に起こる現象に関するものが多く、実際、ポアソン分布の応用例として最初の報告されたのは、軍隊で一定期間中に馬に蹴られて死ぬ人の数の予測だったそうです（聞きかじりの知識で本当かどうか知りません。）。しかし、時間的に起こる確率でなくても、ポアソン分布的なものはあります。私が知っているのは、プランクトンの計数です。今は、様々なタイプのパーティクル・カウンターがありますから、一定量の液体の中の粒子の数を数えることは簡単です。形態の識別をしてくれる機械もあります。昔は、顕微鏡下で観察して、形態を識別して、数取器でカウントしました。計数板というものを使います。

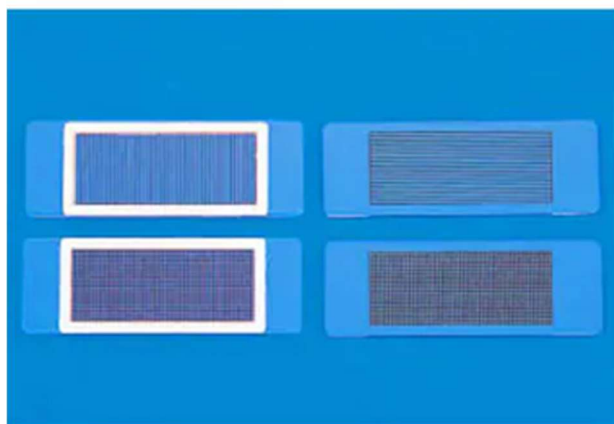


図18界線スライドグラス

様々な係数板がありますが、図 18 に示したものは、モノタローで売っている界線スライドグラスです。1 mm メッシュで深さ 1 mm です。これだと、1 区画の体積は、 0.001 mm^3 です。プランクトンの大きさや種類にもよりますが、出現数の少ないプランクトンの種類だと、全く出現しない区画がたくさんできます。かといって、1 ml 以上調べようとする、1000 区画もカウントしなければなりません。ポアソン分布にしたがうと考えて、最頻値を期待値として、それを 1000 倍して 1 ml 当たりの出現量の推定値とした方が良さそうです。

なお、界線スライドグラスの値段はバカ高いので、自作した方が良いでしょう。ただし、もっと細かい、細胞数のカウントなどだと、さすがに。血球など細胞の計数盤などを買わざるを得ないでしょう。

では、ポアソン分布を作ります。要は、二項分布の変形です。二項分布は

$$W(k) = \binom{n}{k} p^k q^{(n-k)}$$

ですが、ポアソン分布らしく、 $W(k)$ ではなくて、 $P(k)$ と書きます。

$$P(k) = \binom{n}{k} p^k q^{(n-k)}$$

ある事象が k 回起きる確率です。

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

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

$k!$ には p も n も含まれないので、これを外に取り出して、次の3つの部分に分ける

$$P(k) = \frac{1}{k!} AB$$

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

$$B = (1-p)^{(n-k)}$$

A と B について、 $n \rightarrow \infty$ の値を考える。

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

$$= n(n-1) \cdots (n-k+1) p^k$$

ここで無理やり、 $(np)^k$ を作ります

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

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

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0, \quad \lim_{n \rightarrow \infty} \frac{k-1}{n} = 0$$

だから

$$\lim_{n \rightarrow \infty} A = \frac{n!}{(n-k)!} p^k = (np)^k$$

二項分布では平均値（期待値 μ ）は、 $\mu = np$ だから

$$\lim_{n \rightarrow \infty} A = \mu^k$$

B について、

$$B = (1-p)^{(n-k)}$$

$\mu = np$ を使って、 p を消して n 、 k の式に変形。

$$\begin{aligned} B &= \left(1 - \frac{\mu}{n}\right)^{(n-k)} \\ &= \left(1 - \frac{\mu}{n}\right)^{\left(\frac{n}{\mu} - k\right)} \\ &= \frac{\left(1 - \frac{\mu}{n}\right)^{\frac{n}{\mu}}}{\left(1 - \frac{\mu}{n}\right)^k} \end{aligned}$$

$$\text{分母} \left(1 - \frac{\mu}{n}\right)^k \text{ については、} \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right) = 1$$

B 分子については、ネイピア数 e が t 次のような超越関数であることを知っていれば、

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

μ を一つの単位数と考えて、 $-\frac{\mu}{n}$ を $\frac{1}{n}$ と書き換えて、

$$\left(1 - \frac{\mu}{n}\right)^{\frac{n}{\mu}} = e^{-\mu}$$

となります。解法はネイピア数を知っていれば簡単です。ただ、これを思いつくのは、ポアソン分布が指数関数になっていることを知っているからです。私もそうでした（無理に指数関数を作ろうとした。）、何の事前情報もなしにこの解放を思いつく人は少ないでしょう（数学に慣れている人を除く）。とにかく、必要ならば、ブログの統計学 III-3-2.（ネイピア数）の説明を読めばよいので、知っていることにして

$$\lim_{n \rightarrow \infty} B = \frac{\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^{\frac{n}{\mu}}}{\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^k} = \frac{e^{-\mu}}{1} = e^{-\mu}$$

3つを合わせて、

$$\lim_{n \rightarrow \infty} P(k) = \frac{1}{k!} \lim_{n \rightarrow \infty} A \lim_{n \rightarrow \infty} B = \frac{1}{k!} \mu^k e^{-\mu}$$

となります。期待値 λ の事象が、ある制限内で起こるということを無限回繰り返した時に、描かれる確率分布はこのように表せます。この計算では、二項分布の平均値の期待値

として μ という記号を使ってきましたが、ポアソン分布の場合、その期待値は、何らかの情報によって外から与えられる（例えば10年間で熊に喰われたか人の数）ので、 μ を λ に書き換えて

$$P(k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

と書きます。ネイピア数の計算の仕方はいくつかあるのですが、 $e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$ を知っていれば、後はちょっとしたテクニックを使った式の書き換えだけで導出できます。

ポアソン分布では、平均値に依存する形で分散が変化します。2項分布だって分散が平均値によって変わります。ポアソン分布は np が一定で、 p が小さくなれば n が大きくなる。 n が大きくなれば分布が平均値を中心に尖ってくるのだから、当たり前だろうと言われればそれまでです。しかし、完全なポアソン分布では、 k の分散の値は平均の値と同じになります。この性質は、データの分散がポアソンのであるかそうでないかを考えるときに重要です。その証明をします。二乗の期待値から期待値の二乗を引くと分散になるという、2項分布で使った式から出発します。

$$V(k) = E(k^2) - (E(k))^2$$

ポアソン分布は二項分布の極限ですから、この式は成り立ちます。

ポアソン分布ではデータが k である確率は、

$$P(k) = \frac{1}{k!} \lambda^k e^{-\lambda}$$

ですから、 k^2 の期待値は

$$E_{k^2} = \sum_{k=0}^{\infty} k^2 P(k) = \sum_{k=0}^{\infty} k^2 \frac{1}{k!} \lambda^k e^{-\lambda}$$

ここから、1段階ずつ k を減らした無限総和の式を作っていきます。

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

第2項は $k = k-1$ とした時の確率の総和だから1。

$$E_{k^2} = \sum_{k=0}^{\infty} \frac{k-1}{(k-1)!} \lambda^k e^{-\lambda} + \lambda$$

第1項について同じ手法を繰り返して、

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

シグマの中は $k = k-1$ の時の確率の総和だから 1、したがって

$$E_{k^2} = \lambda^2 + \lambda$$

そもそも

$$E_k = \lambda$$

だから

$$(E_k)^2 = \lambda^2$$

したがって

$$\begin{aligned} V_{(k)} &= E_{(k^2)} - (E_{(k)})^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

となります。ポアソン分布では、分散は期待値と同じになります。これを分散が平均値と等しくなると覚えてもまあ良いのかなと思います。

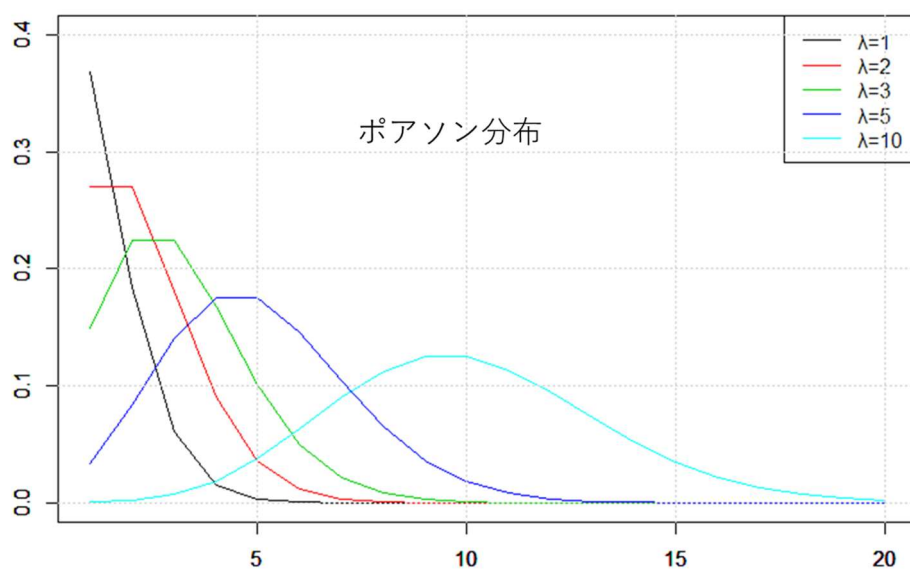


図 19 ポアソン分布

