

6章 確率分布

6.1 離散的データの確率分布

ある確率変数の実現値がそれぞれの実現確率で生じる状態を確率分布といいます。例えば、確率変数をサイコロの目の値とすると、実現確率がそれぞれ $1/6$ の確率分布となります。確率分布にはこのように事象の数が有限なものから、1 時間に到着する客の数 (0 から ∞) のように、事象の数が理論上無限大のものもあります。

さて、ここではこのサイコロの例を用いて確率変数 X の期待値 $E(X)$ を求めてみます。期待値が確率変数の実現値にそれぞれの実現確率を掛けた合計として求められることは 2.3 節で説明した通りです。

$$\begin{aligned} E(X) &= \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 \\ &= \frac{21}{6} = 3.5 \end{aligned}$$

期待値は実現値の平均的な値ですから、確率変数の平均とも呼ばれます。5 章で述べた標本データの平均とは意味が異なりますが、何度も平均を取る操作を繰り返すと、その平均がここでの平均に近づいて行くことが知られています。

次に、平均からのずれの大きさを表わす、確率変数の分散 $V(X)$ を求めてみます。分散は実現値から平均を引いて 2 乗し、その値にそれぞれの実現確率を掛けた合計として求められます。

$$\begin{aligned} V(X) &= \frac{1}{6} \times (1-3.5)^2 + \frac{1}{6} \times (2-3.5)^2 + \cdots + \frac{1}{6} \times (6-3.5)^2 \\ &= \frac{17.5}{6} = 2.916667 \end{aligned}$$

これも標本データから求めた分散とは異なりますが、標本データの不偏分散を何度も求めてみるとその平均がここでの分散に近づくことが知られています。5 章で分散と不偏分散の 2 種類について述べた理由はここにあり、不偏分散がよく利用される理由もここにあります。

さて、平均と分散の式を一般的に表わしておきましょう。全事象の数を n 、確率変数を X として、事象 i に対応する確率変数の値 x_i が確率 p_i で実現されるとします。 n は有限の数の場合もあれば、無限大の場合もあります。この確率分布から確率変数 X の平均 $E(X)$ を求めます。

$$E(X) = \sum_{i=1}^n p_i x_i$$

さらに $E(X) = \mu$ として確率変数の分散を示します。

$$V(X) = \sum_{i=1}^n p_i (x_i - \mu)^2 = \sum_{i=1}^n p_i x_i^2 - \mu^2$$

分散は期待値の書式に従うと $E[(X - \mu)^2]$ と表わされることが分かります。

問題

確率変数の分散を表わす上式の最後の等号について証明せよ。

解答

$$\begin{aligned}\sum_{i=1}^n p_i (x_i - \mu)^2 &= \sum_{i=1}^n p_i (x_i^2 - 2\mu x_i + \mu^2) \\ &= \sum_{i=1}^n p_i x_i^2 - 2\mu \sum_{i=1}^n p_i x_i + \mu^2 \sum_{i=1}^n p_i \\ &= \sum_{i=1}^n p_i x_i^2 - \mu^2\end{aligned}$$

最後の変形には、 $\sum_{i=1}^n p_i x_i = \mu$ 、 $\sum_{i=1}^n p_i = 1$ を利用している。

6.2 連続的データの確率分布

連続的な確率変数の場合、 X がある値となる確率というのは意味がなく(0となります)、 X がある範囲内となる確率だけが意味を持ちます。例えば身長を表わす確率変数 X について、 $X = 170$ となることは意味がなく、 $170 \leq X \leq 171$ となる確率は決めることができるということです。ここではこのような連続的データの確率分布について見てみましょう。

連続的なデータの集計方法としてヒストグラムを描くということを4章で学びましたが、ここではデータの数をもっと大きくしていった場合のヒストグラムの極限の形を考えます。図6.1のようにデータ数を多く取って行くと、ヒストグラムはきめが細かくなり、ヒストグラムの上端を繋いだ線分は次第に滑らかになって行き、母集団の分布に近づいて行きます。

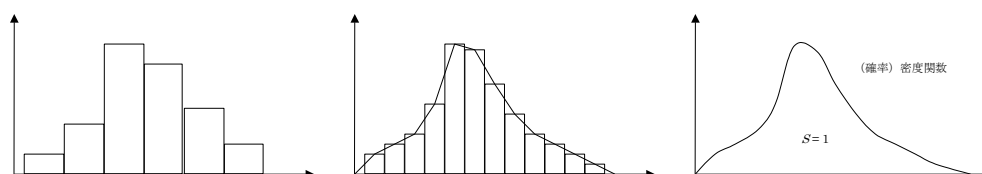


図 6.1 ヒストグラムの極限と密度関数

しかし、多くの場合母集団はデータ数が無限大であると考えていますので、グラフに描く場合縦軸の値としてそのままの度数を書いていくわけにもいきません。そこでグラフの形をそのままにして縦軸の目盛を、グラフの全面積が1となるように付けてやることにします。例えば以下の図6.2を見て下さい。

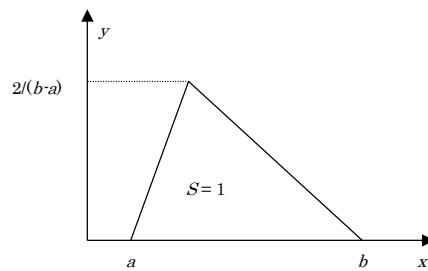


図 6.2 縦軸の目盛の付け方

この場合、全面積を 1 とすると、図の頂点は $2/(b-a)$ という値になることは容易に分かります。このように、ヒストグラムの極限形で、全面積が 1 となるように縦軸目盛を取った関数を、確率密度関数（または単に密度関数）と呼びます。

さて、この確率密度関数を用いてどんなことが計算できるのでしょうか。今、確率密度関数を $y = f(x)$ と表わします。ここではこれを使った計算を見てみましょう。

まず、確率密度関数の面積について考えてみます。ヒストグラムで面積は確率変数 X がある範囲に入る度数を表わしていて、全度数でそれを割ることは X がその範囲に入る割合を表わしていました。今、データ数を限りなく大きくし、全面積を 1 にしていますので、ある部分の面積はそれ自身で全度数に対する割合、即ち確率を表わしています。図 6.3 を見て下さい。

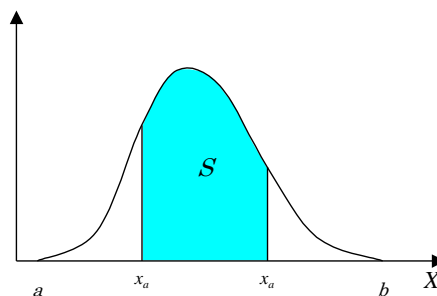


図 6.3 面積と確率

確率変数 X は a から b の範囲を取りますが、 x_a から x_b の間の確率密度関数の面積を S とすると、 S は X がこの範囲の値となる確率になります。この確率は範囲を明記して、 $P(x_a \leq X \leq x_b)$ のようにも書かれます。この部分の面積はまた、積分を使って以下のように表わすこともできます。ここで積分は単に面積を表わすものと理解しておいてもらえば結構です。

$$\text{確率} \quad P(x_a \leq X \leq x_b) = S = \int_{x_a}^{x_b} f(x) dx$$

特に、確率密度関数 $f(x)$ に対して、 a から x までの面積を確率分布関数または単に分布関数と言い、大文字を使って $F(x)$ と書きます。

$$\text{分布関数} \quad F(x) = P(a \leq X \leq x) = \int_a^x f(x) dx$$

これは度数分布表の累積相対度数に相当しています。また、分布関数の微分が確率密度関数になることは容易に分かると思います。また全領域の面積については、全確率が 1 となることを用いて、以下のようになります。

$$\text{全確率} \quad F(b) = P(a \leq X \leq b) = \int_a^b f(x)dx = 1$$

ここに積分する領域を a から b としましたが、 $a \rightarrow -\infty$ 、 $b \rightarrow \infty$ であっても構いません。実際、次章で述べる正規分布は $-\infty$ から ∞ の範囲をとる分布です。

面積と積分について一言簡単にふれておきます。 dx を微小な幅とすると、確率変数が x から $x+dx$ の間に入る確率、即ち確率密度関数の面積は、近似的に図 6-4 における網掛け部分の面積で表わされます。

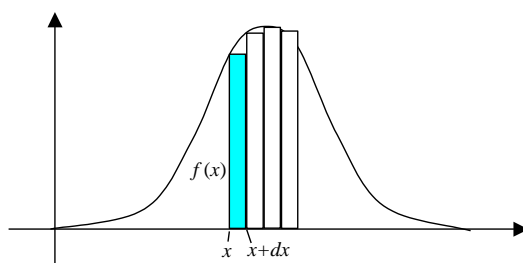


図 6-4 積分と面積

これは高さ $f(x)$ と幅 dx の長方形ですから、面積は $f(x)dx$ となり、この微小な量の合計が積分として与えられています。模式的に書くと以下のような感じです。

$$\int_a^b f(x)dx = f(x)dx \text{ の合計}$$

この教科書では積分について、詳しく説明する予定はありません。きちんと学びたい方は、「基礎からの数学」を勉強して下さい。

次に、確率変数 X の平均 $E(X)$ を求めてみましょう。 dx を微小な幅とすると、確率変数が x から $x+dx$ の間に入る確率は、近似的に $f(x)dx$ で表わされますから、その時の X の値を x で代表させると、平均は実現値 \times 確率の合計ですから、 $x \cdot f(x)dx$ の合計で表わされます。この合計は上に述べたように積分を意味しており、以下のように表わされることになります。

$$\text{確率変数の平均} \quad E(X) = \int_a^b xf(x)dx$$

我々はこの平均を μ という記号で表わすことにします。

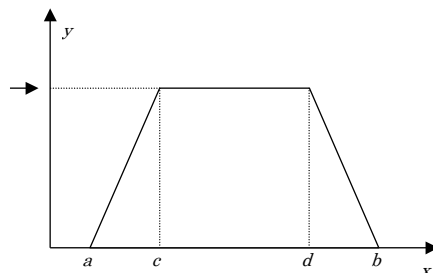
次に確率変数 X の分散 $V(X)$ を求めてみましょう。定義から分散は $(X - \mu)^2$ の期待値として表わされ、 $(x - \mu)^2 \cdot f(x)dx$ の合計になりますので、以下の積分になります。

$$\text{確率変数の分散} \quad V(X) = \int_a^b (x - \mu)^2 f(x)dx$$

この分散についても、表式の簡単化のために、 σ^2 という記号で表わすことにします。

問題

確率密度関数が以下の形状をしていたとすると、台形の上底の位置の目盛はいくらになるか。



解答

$$2/(b+d-a-c)$$

問題（省略しても結構です）

確率密度関数が $0 \leq x \leq \pi$ の範囲で $y = a \sin x$ で与えられるとき、以下の問いに答えよ。

- 1) a の値を求めよ。
- 2) 確率変数の平均 $E(X)$ を求めよ。

解答

$$1) \int_0^\pi a \sin x dx = a[-\cos x]_0^\pi = 2a = 1 \quad \text{より、} a = 1/2$$

$$2) E(X) = 1/2 \int_0^\pi x \sin x dx = 1/2[-x \cos x]_0^\pi + 1/2 \int_0^\pi \cos x dx = \pi/2$$

6.3 確率変数の平均と分散の一般的性質 [Skip OK]

ここでは確率変数の平均と分散について、その性質をまとめてみましょう。証明については、離散的な確率変数の場合についてだけ行ないます。連続的な確率変数については、積分を使うのでなじみが薄いと思いますし、内容は離散的な場合と基本的に同じですので、省略します。

6.3.1 線形変換

平均 μ 、分散 σ^2 の確率変数 X に対して、 $X' = cX + d$ という線形変換によって新しい確率変数 X' を作ります。ここで、 c と d は任意の定数とします。この新しい確率変数 X' に対して、平均と分散はどうなるか、考えてみましょう。

最初は平均です。

$$\begin{aligned}
 E(X') &= \sum_{i=1}^n (cx_i + d)p_i = c \sum_{i=1}^n x_i p_i + d \sum_{i=1}^n p_i \\
 &= cE(X) + d = c\mu + d
 \end{aligned}$$

2行目に移る前に、全確率が1であるという式 $\sum_{i=1}^n p_i = 1$ を用いました。

次は分散ですが、これには X' の平均が $c\mu + d$ であるという上の結果を用います。

$$\begin{aligned}
 V(X') &= \sum_{i=1}^n [(cx_i + d) - (c\mu + d)]^2 p_i \\
 &= c^2 \sum_{i=1}^n (x_i - \mu)^2 p_i = c^2 V(X) = c^2 \sigma^2
 \end{aligned}$$

これから以下の結論を得ます。これは連続的な確率変数に対しても成り立つ、非常に重要な関係です。

確率変数 X の平均 $E(X)$ と分散 $V(X)$ が、 $E(X) = \mu$ 、 $V(X) = \sigma^2$ で与えられるとき、確率変数 $X' = cX + d$ の平均 $E(X')$ と分散 $V(X')$ は以下で与えられる。

$E(X') = c\mu + d$	$V(X') = c^2 \sigma^2$
--------------------	------------------------

問題

確率変数 X の平均が 170.3、分散が 64.62 のとき、以下の確率変数 X' の平均と分散を求めよ。

- 1) $X' = 2X + 100$
2) $X' = 0.1X - 10$

解答

- 1) $E(X') = 440.6$, $V(X') = 258.48$
2) $E(X') = 7.03$, $V(X') = 0.6462$

問題

確率変数 X の平均が μ 、分散が σ^2 のとき、どのような線形変換 $X' = cX + d$ を考えれば、平均を 0、分散を 1 にできるか。

解答

$$X' = \frac{X - \mu}{\sigma} \quad \text{即ち、} c = \frac{1}{\sigma}, \quad d = -\frac{\mu}{\sigma} \text{ とする。}$$

6.3.2 確率変数の合成

これまで、1つの確率変数に対して、線形変換を考えてきましたが、ここでは互いに独立な2つの確率変数 X_1, X_2 について、その和 $X = X_1 + X_2$ を新しい確率変数として平均と分散を求めてみましょう。ここでは計算が多少厄介になりますので後に回して、結果だけをまとめておきます。

確率変数 X_1, X_2 について、その平均と分散が

$$E(X_1) = \mu_1, E(X_2) = \mu_2 \quad V(X_1) = \sigma_1^2, V(X_2) = \sigma_2^2$$

のように与えられているとき、確率変数 $X = X_1 + X_2$ の平均と分散は以下のように与えられる。

$$\boxed{E(X) = \mu_1 + \mu_2 \quad V(X) = \sigma_1^2 + \sigma_2^2}$$

以下では興味のある読者のために離散的な場合の証明を与えておきます。

互いに独立な確率変数 X_a ($a=1, 2$) の実現値と生起確率をそれぞれ $x_i^{(a)}$, $p_i^{(a)}$ としますと、平均は以下のように計算されます。

$$\begin{aligned} E(X) &= \sum_{i=1}^m \sum_{j=1}^n p_i^{(1)} p_j^{(2)} (x_i^{(1)} + x_j^{(2)}) = \sum_{i=1}^m p_i^{(1)} x_i^{(1)} \sum_{j=1}^n p_j^{(2)} + \sum_{i=1}^m p_i^{(1)} \sum_{j=1}^n p_j^{(2)} x_j^{(2)} \\ &= \sum_{i=1}^m p_i^{(1)} x_i^{(1)} + \sum_{j=1}^n p_j^{(2)} x_j^{(2)} = E(X_1) + E(X_2) = \mu_1 + \mu_2 \end{aligned}$$

ここに独立性とは、それぞれの確率変数について値が $x_i^{(1)}, x_j^{(2)}$ となる確率が $p_i^{(1)} \times p_j^{(2)}$ のように互いの確率の積で与えられることです。

分散については、平均の値 $\mu_1 + \mu_2$ を利用して以下のように計算できます。

$$\begin{aligned} V(X) &= \sum_{i=1}^m \sum_{j=1}^n p_i^{(1)} p_j^{(2)} (x_i^{(1)} + x_j^{(2)} - \mu_1 - \mu_2)^2 \\ &= \sum_{i=1}^m \sum_{j=1}^n p_i^{(1)} p_j^{(2)} [(x_i^{(1)} - \mu_1)^2 + (x_j^{(2)} - \mu_2)^2 + 2(x_i^{(1)} - \mu_1)(x_j^{(2)} - \mu_2)] \\ &= \sum_{i=1}^m p_i^{(1)} (x_i^{(1)} - \mu_1)^2 \sum_{j=1}^n p_j^{(2)} + \sum_{i=1}^m p_i^{(1)} \sum_{j=1}^n p_j^{(2)} (x_j^{(2)} - \mu_2)^2 \\ &\quad + 2 \sum_{i=1}^m p_i^{(1)} (x_i^{(1)} - \mu_1) \sum_{j=1}^n p_j^{(2)} (x_j^{(2)} - \mu_2) \\ &= V(X_1) + V(X_2) = \sigma_1^2 + \sigma_2^2 \end{aligned}$$

問題

上記確率変数の合成で、 $X = c_1 X_1 + c_2 X_2$ の場合、平均と分散はどうなるか。

解答

$$E(c_i X_i) = c_i \mu_i, \quad V(c_i X_i) = c_i^2 \sigma_i^2 \quad (i=1, 2) \text{ より、}$$

$$E(X) = c_1 \mu_1 + c_2 \mu_2, \quad V(X) = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2$$

問題

平均 μ , 分散 σ^2 の互いに独立な確率変数 X_i ($i=1, 2, \dots, n$) を使って、標本平均に相当する新しい確率変数 $X = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ を作った。新しい確率変数の平均、分散、標準偏差を求めよ。

解答

$$\text{平均 } \mu, \text{ 分散 } \frac{\sigma^2}{n}, \text{ 標準偏差 } \frac{\sigma}{\sqrt{n}}$$

これによって、データ数を多くすると推測の精度が上がることが分る。また、後に述べる中心極限定理から X の分布型が $n \rightarrow \infty$ で正規分布に近づくことも知られている。

問題

ある工事は独立な3つの工程から成り立っている。各工程の工期の平均と標準偏差は以下で与えられる。この工事全体の工期の平均と分散と標準偏差はいくらか。

	平均	標準偏差
工程 1	10.3	3.2
工程 2	6.5	1.5
工程 3	9.4	2.8

解答

$$\text{平均} \quad 10.3 + 6.5 + 9.4 = 26.2$$

$$\text{分散} \quad 3.2^2 + 1.5^2 + 2.8^2 = 20.33$$

$$\text{標準偏差} \quad \sqrt{20.33} = 4.50888$$

6.4 確率分布の例

この節では、いろいろな分野で実際によく利用される確率分布の例をまとめようと思います。

6.4.1 離散的な確率分布

2 項分布

ある事象 A の実現確率が p で与えられるとき、 n 回の独立な試行で事象 A が起こる回数の確率分布が2項分布です。事象 A が r 回起こる確率 $P(r)$ は以下で与えられます。

$$P(r) = {}_n C_r p^r (1-p)^{n-r}$$

さて、今事象 A が最初から連続して r 回起こり、残りの $n-r$ 回は起こらなかったとします。これも r 回起こった1つの例です。この場合の実現確率は $p^r (1-p)^{n-r}$ となります。しかし、最初の問題は事象 A が r 回起こったというだけですから、 n 回の試行のうちどこで事象 A が起きたのか、その場合の数だけこの確率に掛けてやらなければなりません。 n 回中どこで起こったのか r 回取り出す場合の数は、1章で学んだ組み合わせで ${}_n C_r$ となります。

これらをまとめて、上の確率が求まります。

この確率を用いて、2項分布の平均は np 、分散は $np(1-p)$ であることが計算できます。

$$E(X) = \sum_{r=0}^n r P(r) = np$$

$$V(X) = \sum_{r=0}^n (r - np)^2 P(r) = np(1 - p)$$

また、 p を有限にして試行回数 n を十分大きくすると (r も大きくなります)、事象 A の起こる回数は平均 np 、分散 $np(1 - p)$ の正規分布 (後述) に近づくことも知られています。

問題

10 回じゃんけんを行い、勝ち数が 0~10 となる各々の確率を求めよ。但し、あいこの場合は決着がつくまで行うものとする。

解答

$P(r) = {}_{10}C_r (1/2)^{10}$ より、

勝ち数	確率	勝ち数	確率
0	0.000977	6	0.205078
1	0.009766	7	0.117188
2	0.043945	8	0.043945
3	0.117188	9	0.009766
4	0.205078	10	0.000977
5	0.246094		

問題

n 回の試行中、事象 A が r 回起こる確率が $P(r) = {}_nC_r p^r (1 - p)^{n-r}$ で与えられる 2 項分布の平均が np となることを証明せよ。

解答

$$\begin{aligned} E(X) &= \sum_{r=0}^n r P(r) = \sum_{r=0}^n r {}_nC_r p^r (1 - p)^{n-r} = \sum_{r=1}^n r \frac{n!}{r!(n-r)!} p^r (1 - p)^{n-r} \\ &= \sum_{r=1}^n np \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} (1 - p)^{n-r} = np \sum_{r'=0}^{n-1} {}_{n-1}C_{r'} p^{r'} (1 - p)^{n-1-r'} \\ &= np \end{aligned}$$

ここに、 $r' = r - 1$ としています。興味のある人は分散についてもやってみて下さい。

多項分布

2 項分布を拡張して、事象 A_i ($i = 1, 2, \dots, s$) が実現確率 p_i で起こる場合を考えてみます。このとき、 n 回の試行で事象 A_i が r_i 回起こる確率 $P(r_1, r_2, \dots, r_s)$ を求めると以下ようになります。

$$P(r_1, r_2, \dots, r_s) = \frac{n!}{r_1! r_2! \dots r_s!} p_1^{r_1} p_2^{r_2} \dots p_s^{r_s}$$

但し、 $r_1 + r_2 + \dots + r_s = n$, $p_1 + p_2 + \dots + p_s = 1$

これは 2 項分布の拡張で多項分布といいます。2 項分布は $s = 2$ として、事象 A_1 として A が起こる場合、事象 A_2 として A が起こらない場合を考えることに相当します。

ポアソン分布

これは待ち行列と呼ばれるオペレーションズリサーチの分野でよく用いられる確率分布です。待ち行列とは文字通り、サービス窓口に客が並ぶ行列で、実際の窓口もあれば、コンピュータ上でのジョブの処理のような目に見えないものもあります。客の到着がランダムで、それまでの到着状況に依存しない場合、単位時間に到着する人数の確率分布がポアソン分布です。単位時間あたりの客の平均到着数が λ で与えられるとき、単位時間に r 人到着する確率 $P(r)$ は以下で与えられます。

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

先の2項分布を元に、 $np = \lambda$ （一定）として、 n を十分大きくした場合に2項分布とポアソン分布が近似的に一致することが知られています。

問題

1時間に平均5人ランダムに（ポアソン分布に従って）到着するとき、0～10人の到着する確率は各々いくらか求めよ。

解答

到着数	確率	到着数	確率
0	0.006738	6	0.146223
1	0.033690	7	0.104445
2	0.084224	8	0.065278
3	0.140374	9	0.036266
4	0.175467	10	0.018133
5	0.175467		

6.4.2 連続的な確率分布

一様分布

確率密度関数が $f(x) = c$ （一定）という形で与えられる確率分布を一様分布といいます。これは確率変数 X が、 $x \leq X \leq x + \Delta x$ の値をとる確率が x によらず、 $c\Delta x$ で与えられる分布です。確率変数の変数域が $a \leq X \leq b$ の場合、密度関数の高さは、 $c = 1/(b - a)$ になることは容易に分かると思います。

正規分布

正規分布は統計学で最も重要な分布で、偶発的なデータのゆらぎによって生じる確率分布です。十分理解してもらうために、次章で詳しく説明しようと思いますので、ここでは省略させていただきます。また、推定や検定といった統計学の分野でよく利用される、 χ^2 分布、F分布、t分布などの確率分布は、正規分布から派生するもので、これについても後ほど説明します。

指数分布

指数分布は前に述べたポアソン分布と対を成す分布で、単位時間あたりの客の到着数がポアソン分布に従う場合に、その客の到着時間間隔の確率分布が指数分布です。指数分布の確率密度関数は、単位時間あたりの平均到着数を λ とすると、以下のように与えられます。

$$f(x) = \lambda e^{-\lambda x}$$

この確率分布の平均、即ち到着時間間隔の平均は、当然単位時間当りの到着数の逆数、 $1/\lambda$ になります。