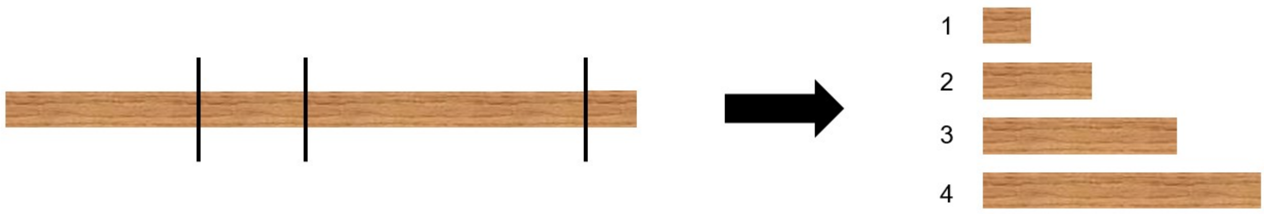
数直線上の $0, 1$ を両端とした長さ 1 の棒と 2 以上の整数 N があります。
区間 $[0, 1]$ 上の実数を一様ランダムに $(N-1)$ 点を独立に選び、この棒を N 個の断片に分割します。
断片を短い順に並べたとき k 番目 $(1 \leq k \leq N)$ の断片の長さを $S(N, k)$ とし、$S(N, k)$ の期待値を $E(N, k)$ とします。
例えば、$E(2, 1) = \frac{1}{4}$, $E(2, 2) = \frac{3}{4}$ です。

(1) $E(3, 1)$ を求めてください。

(2) $E(N, 1)$ を求めてください。ただし、N は $N \geqq 2$ をみたす整数とします。

(3) $E(N, k)$ を求めてください。ただし、N, k は $N \geq 2,\ 1 \leq k \leq N$ をみたす整数とします。

(4) 自由に問題を拡張して議論してください。

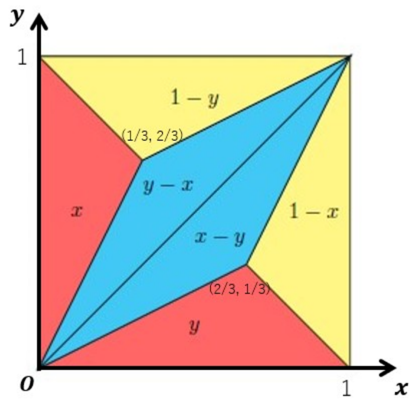


問題 (1) 解答

区間 $[0,1]$ から選んだ 2 つの乱数を x, y とおきます．このとき，3 つの断片の長さは

$$\min(x,y), \quad \max(x,y)-\min(x,y), \quad 1-\max(x,y)$$

となります．この 3 つのうちどれが最小になるか，および x, y の大小関係で $[0,1]^2$ を領域分けすると下図のようになります.（どれが最小になるかで色分けしています）



直線 $x=y$ での対称性に注意すれば，

$$\begin{aligned}
\frac{E(3,1)}{2} &= \iint_{0\le x\le y\le 1} \min(x, y-x, 1-y)dxdy \\
&= \iint_{1-y\le x, 1-y\le y-x, 0\le x\le y\le 1}(1-y)dxdy + \iint_{y-x\le x, y-x\le 1-y, 0\le x\le y\le 1}(y-x)dxdy \\
&\quad + \iint_{x\le y-x, x\le 1-y, 0\le x\le y\le 1} xdxdy \\
&= \int_{2/3}^{1}\left[\int_{1-y}^{2y-1}(1-y)dx\right]dy + \int_0^{1/3}\left[\int_x^{2x}(y-x)dy\right]dx + \int_{1/3}^{1}\left[\int_x^{\frac{x+1}{2}}(y-x)dy\right]dx + \int_0^{1/3}\left[\int_{2x}^{1-x}xdy\right]dx \\
&= \int_{2/3}^{1}(3y-2)(1-y)dy + \int_0^{1/3}\frac{x^2}{2}dx + \int_{1/3}^{1}\frac{(1-x)^2}{8}dx + \int_0^{1/3}(1-3x)xdx \\
&= \frac{1}{54} + \frac{1}{162} + \frac{1}{81} + \frac{1}{54} \\
&= \frac{1}{18}
\end{aligned}$$

したがって，

$$E(3,1) = 2\cdot\frac{1}{18} = \frac{1}{9}$$

です．

問題 (2)(3) 解答

$[0,1]$ 上の独立同分布な一様乱数を $X_1,\dots,X_{N-1}$ とおきます．このとき $\{1,\dots,N-1\}$ の順列 $\{\sigma(1),\dots,\sigma(N-1)\}$ は $(N-1)!$ 通りありますが，その全てに対して $X_{\sigma(1)}<\cdots<X_{\sigma(N-1)}$ となるような確率は同様に確からしく，また，ある $i\neq j$ に対して $X_i=X_j$ となるような確率は 0 です．ここで，

$$D=\{(x_1,\dots,x_{N-1});0\le x_1\le\cdots\le x_{N-1}\le 1\}$$

$$D':=\{(y_1,\dots,y_{N-1})\in\mathbb{R}^{N-1};0\le y_i,\sum_i y_i\le 1\}.$$

と定めて，$(X_1,\dots,X_{N-1})$ を D 上の一様分布に従う $\mathbb{R}^{N-1}$-値に従う確率変数とします．

このとき，

$$\varphi(x_1,\dots,x_{N-1})=(x_1,x_2-x_1,x_3-x_2,\dots,x_{N-1}-x_{N-2})$$

とした場合，これは $\varphi:D\to D'$ の連続かつ線形な全単射となります．よって，ヤコビアンは場所によらず一定なため，$\varphi(X_1,\dots,X_{N-1})$ は D' 上の一様分布に従います．

次に D' 上の一様分布に従う $\mathbb{R}^{N-1}$-値確率変数 $(Y_1,\dots,Y_{N-1})$ を考えます．便宜上 $Y_N=1-\sum_{i=1}^{N-1}Y_i\in[0,1]$ であるとします．$(1,\dots,N)$ の $N!$ 個の並び替え $(\sigma(1),\dots,\sigma(N))$ に対して，$Y_{\sigma(1)}<\cdots<Y_{\sigma(N)}$ となるような確率は同様に確からしく，また，ある $i\neq j$ に対して $Y_i=Y_j$ となるような確率は 0 です．よって，

$$A=\{(x_1,\cdots,x_{N-1})\in D';0\le x_1\le\cdots\le x_{N-1}\le 1-\sum_{i=1}^{N-1}x_i\}$$

に対して，

$$E(N,k)=\left(\int\cdots\int_{D'\cap A}y_k dy_1\cdots dy_{N-1}\right)\cdot\left(\int\cdots\int_{D'\cap A}dy_1\cdots dy_{N-1}\right)^{-1}$$

となります．ここで，

$$\left(\int\cdots\int_{D'\cap A}dy_1\cdots dy_{N-1}\right)=\frac{1}{N!}\int\cdots\int_{D'}dy_1\cdots dy_{N-1}=\frac{1}{N!\cdot(N-1)!}$$

なので，

$$E(N,k)=\left(\int\cdots\int_{D'\cap A}y_k dy_1\cdots dy_{N-1}\right)\cdot N!\cdot(N-1)!$$

このとき，

$$\begin{aligned}\left(\int\cdots\int_{D'\cap A}y_k dy_1\cdots dy_{N-1}\right)&=\frac{1}{N!}\left(\int\cdots\int_{D'\cap A}y_k dy_1\cdots dy_{N-1}\right)\left(\int_0^\infty z^N e^{-z}dz\right)\\&=\frac{1}{N!}\int\cdots\int_{D'\cap A}\int_0^\infty zy_k z^{N-1}e^{-z}dzdy_1\cdots dy_{N-1}\end{aligned}$$

ここで，$zy_i=x_i(i=1,\dots,N-1),\ \ x_N=z(1-\sum_{i=1}^{N-1}y_i)$ と変数変換します．

このとき，$(x_1,\dots,x_N)\in B=\{(x_1,\dots,x_N);0\le x_1\le x_2\le\cdots\le x_N<\infty\}$ であり，

$dx_1 \cdots dx_N = z^{N-1} dz dy_1 \cdots dy_{N-1}$ です．よって，

$$\left(\int \cdots \int_{D' \cap A} y_k dy_1 \cdots dy_{N-1} \right) = \frac{1}{N!} \int \cdots \int_B x_k \exp(-x_1 - \cdots - x_N) dx_1 \cdots dx_N$$

です．よって，

$$\begin{aligned} E(N,k) &= (N-1)! \cdot \int \cdots \int_B x_k \exp(-x_1 - \cdots - x_N) dx_1 \cdots dx_N \\ &= (N-1)! \cdot \int_0^\infty dx_1 \int_{x_1}^\infty dx_2 \cdots \int_{x_{N-1}}^\infty dx_N \cdot x_k \cdot \exp(-x_1 - \cdots - x_N) \end{aligned}$$

です．

公式

$$\int_p^\infty e^{-cx}(ax+b)dx = \frac{1}{c} \cdot e^{-cp} \left(ap + b + \frac{a}{c} \right)$$

（ただし $p \geq 0, c > 0$）を用いることでこの N 重積分を内側から順に計算することができます．

$$\int_{x_{N-1}}^\infty \exp(-x_N) dx_N = \exp(-x_{N-1})$$

$$\int_{x_{N-2}}^\infty \exp(-2x_{N-1}) dx_{N-1} = \frac{1}{2} \exp(-2x_{N-2})$$

$\vdots$

$$\int_{x_k}^\infty \frac{\exp(-(N-k)x_{k+1})}{(N-1-k)!} dx_{k+1} = \frac{\exp(-(N-k)x_k)}{(N-k)!}$$

$$\int_{x_k-1}^\infty \frac{\exp(-(N-k+1)x_k)}{(N-k)!} x_k dx_k = \frac{\exp(-(N-k+1)x_{k-1})}{(N-k+1)!} \left(x_{k-1} + \frac{1}{N-k+1} \right)$$

$$\int_{x_{k-2}}^\infty \frac{\exp(-(N-k+2)x_{k-1})}{(N-k+1)!} \cdot \left(x_{k-1} + \frac{1}{N-k+1} \right) dx_{k-1} = \frac{\exp(-(N-k+2)x_{k-2})}{(N-k+2)!} \left(x_{k-2} + \sum_{l=1}^{2} \frac{1}{N-k+l} \right)$$

$\vdots$

$$\int_0^\infty \frac{\exp(-Nx_1)}{(N-1)!} \left(x_1 + \sum_{l=1}^{k-1} \frac{1}{N-k+l} \right) dx_1 = \frac{1}{N!} \left(\sum_{l=1}^{k} \frac{1}{N-k+l} \right)$$

よって，

$$\begin{aligned} E(N,k) &= (N-1)! \cdot \frac{1}{N!} \left(\sum_{l=1}^{k} \frac{1}{N-k+l} \right) \\ &= \frac{1}{N} \left(\frac{1}{N-k+1} + \cdots + \frac{1}{N} \right) \end{aligned}$$

特に $k = 1$ のとき

$$E(N,1) = \frac{1}{N^2}$$

問題 (2)(3) 別解

前章の問題 (2)(3) の解答は数学科かそれに近い専攻でないと少々厳しいかもしれませんので，組み合わせ論的な考え方も挙げておきます．論証の厳密さについてはご容赦ください．

問題 (2)(3) の別解に入る前に，準備として，「$a+b+c+d=100$ をみたす非負整数 (a,b,c,d) の組は何通りあるか」という問題を考えてみましょう．これは，◯ 100 個と仕切り｜3 個を横一列に並べる組み合わせの数に等しく，${}_{103}C_3$ 通りとなります．

次に，「$a+b+c+d=100$ をみたす 5 以上の整数 (a,b,c,d) の組は何通りあるか」という問題を考えてみましょう．これは，$a'=a-5, b'=b-5, c'=c-5, d'=d-5$ とおけば「$a'+b'+c'+d'=80$ をみたす非負整数 (a',b',c',d') の組は何通りあるか」という問題に置き換えられますので，◯ 80 個と仕切り｜3 個を横一列に並べる組み合わせの数に等しく，${}_{83}C_3$ 通りとなります．a,b,c,d に予め 5 ずつ振り分けることで，◯の個数が $5\times4=20$ 個減りました．

以下の別解では，このような考え方を用います．

——

以降，数直線上の区間 $[0,1]$ において，0 側を「左」と呼ぶことにします．

断片の長さを短い順に $l_1, l_2, \cdots, l_N$ としたとき，これらの並び替えは $N!$ 通りありますが，どの並び替えについても断片の長さが左から順にその並びになる確率密度[1]は等しいため，最も左の断片が最も短くなるときの条件付き期待値を求めればよいです．

最も左の（最も短い）断片の長さを x とおくと，残りの断片の長さは

$$x+\alpha_2,\ x+\alpha_3,\ \cdots, x+\alpha_N \quad (\alpha_i \geqq 0)$$

であり，

$$\alpha_2+\alpha_3+\cdots+\alpha_N = 1-Nx$$

であるため，（N 個の断片に x を振り分けた後に）長さ $(1-Nx)$ の棒を $(N-1)$ 個に分ける（つまり長さ $(1-Nx)$ の区間から $(N-2)$ 個の点をランダムに選ぶ）ことに等しく，最も左の（最も短い）断片の長さが x である確率密度は $(1-Nx)^{N-2}$ に比例します．

よって，確率密度関数を定数 c を用いて

$$f(x) = c(1-Nx)^{N-2}$$

とおけば，

1. 確率変数 X の確率密度関数を $f_X(x)$ とすれば，$P(a \leqq X \leqq b) = \int_a^b f_X(x)dx$ と書けます．
ここでは，N 変数の確率密度関数を考えます．

$$
\begin{aligned}
\int_0^{\frac{1}{N}} f(x)dx &= cN^{N-2}\int_0^{\frac{1}{N}}\left(\frac{1}{N}-x\right)^{N-2}dx \\
&= cN^{N-2}\int_0^{\frac{1}{N}} x^{N-2}dx \\
&= \frac{c}{N(N-1)} = 1
\end{aligned}
$$

より，

$$
c = N(N-1)
$$

を得ます．よって，

$$
\begin{aligned}
E(N,1) &= \int_0^{\frac{1}{N}} xf(x)dx \\
&= N^{N-1}(N-1)\int_0^{\frac{1}{N}} x\left(\frac{1}{N}-x\right)^{N-2}dx \\
&= N^{N-1}(N-1)\int_0^{\frac{1}{N}}\left(\frac{1}{N}-x\right)x^{N-2}dx \\
&= N^{N-1}(N-1)\left(\frac{1}{N(N-1)}\left[x^{N-1}\right]_0^{\frac{1}{N}} - \frac{1}{N}\left[x^N\right]_0^{\frac{1}{N}}\right) \\
&= \frac{1}{N} - \frac{N-1}{N^2} \\
&= \frac{1}{N^2}
\end{aligned}
$$

となり，問題 (2) の答えが求まりました．

次に，

$$
E(N,k) = \sum_{i=1}^{k}\frac{1}{N(N-i+1)} \tag{1}
$$

であることを帰納法を用いて示します．
$N=2$ のときは，問題文より (1) 式は成立しています．
$N=m-1$ のとき成立していると仮定して，$N=m$ のときを考えます．
先ほどと同様の考え方により，最も左の断片の長さが最も短く x であるとき，残りの断片の長さは

$$
x+\alpha_2,\ x+\alpha_3,\cdots,x+\alpha_m \quad (\alpha_i \geqq 0)
$$

であり，

$$
\alpha_2+\alpha_3+\cdots+\alpha_m = 1-mx
$$

であるため，(m 個の断片に x を振り分けた後に）長さ $(1-mx)$ の棒を $(m-1)$ 個に分けることに等しいです．また，(2) の解答より x の期待値は $\frac{1}{m^2}$ なので，帰納法の仮定と合わせて，$2 \leqq k \leqq m$ のとき，

$$
\begin{aligned}
E(m,k) &= \frac{1}{m^2} + (1 - m \cdot \frac{1}{m^2})E(m-1, k-1) \\
&= \frac{1}{m^2} + \frac{m-1}{m} \sum_{i=1}^{k-1} \frac{1}{(m-1)(m-i)} \\
&= \frac{1}{m^2} + \sum_{i=1}^{k-1} \frac{1}{m(m-i)} \\
&= \sum_{i=1}^{k} \frac{1}{m(m-i+1)}
\end{aligned}
$$

となり，$N = m$ のときも (1) 式は成立します．
以上より，任意の N, k において (1) 式が成立することが示されました．

問題 (3) 別解

(3) の解答はシンプルな形式をしていますが，この表式の意味づけを示唆する別解もいただきました．下記の $(N-1)$ 次元立体内に一様分布するかについては議論が必要ですが，とても美しい方針のため，概略を紹介します．

———

各断片の長さを左から順に $x_1, x_2, \cdots, x_N$ とします．$\sum_{i=1}^{N} x_i = 1 \quad (x_i \geqq 0)$ であるため，$\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$ は

$$(1, 0, \cdots, 0), (0, 1, \cdots, 0), \cdots, (0, 0, \cdots, 1)$$

の N 点を頂点とする $(N-1)$ 次元単体上に分布します．

この単体上において，$x_1 \leqq x_2 \leqq \cdots \leqq x_N$ をみたす領域は，

$$(0, \cdots, 0, 1), (0, \cdots, \frac{1}{2}, \frac{1}{2}), \cdots, (\frac{1}{N}, \cdots, \frac{1}{N}, \frac{1}{N})$$

の N 個の頂点で囲まれた $(N-1)$ 次元の立体であり，$E(N,k)$ はこの立体の重心の第 k 成分と一致するので，

$$E(N,k) = \sum_{i=1}^{k} \frac{1}{N(N-i+1)}$$

問題 (4)

— 初めて $\frac{1}{N}$ 未満の長さの断片が現れるまでに選ぶ乱数の個数の期待値を計算
— $S(N,k)^2$ の期待値を計算
— シミュレーションを行って一致しているか確認
— 離散化した場合の極限を計算
など，さまざまな拡張をいただきました．