Random Points on $[0,1]$

1 問題

$[0,1]$ 区間上から一様ランダムに点を N 個取ることを考える．以下の期待値を答えよ．なお，**区間**とは点をちょうど 2 つ含む閉区間のうち極小のものを指す (区間は $N-1$ 個ある).

(1) 重心.

(2) 最大値.

(3) 最近点対の距離

(4) 全点対の距離の平均

(5) 区間の距離のうち k 番目に大きい物の長さ

(6) 区間の距離のうち平均より大きい物の個数

1　重心

i 番目にとってきた点の位置を X_i とする．$X_1, \ldots, X_N$ は $[0,1]$ 一様分布に従う独立同分布である．

$$
\begin{aligned}
& E\left[\frac{X_1 + \cdots + X_N}{N}\right] \\
& = \frac{E[X_1] + \cdots + E[X_N]}{N} \\
& = \frac{1}{2}
\end{aligned}
$$

2　最大値

最大値が k 以下になる確率を $P(k)$ とおく．X_i が k 以下になる確率は k なので，$P(k) = k^N$ である．確率変数 M を $M = \max(X_1, \dots, X_N)$ で定義すると，$P(k)$ は M の累積密度関数に相当する．微分すると確率密度関数 $p(k)$ がわかる．

$$p(k) = \frac{\mathrm{d}P}{\mathrm{d}k}(k) = Nk^{N-1}$$

求めたい値は $\int_0^1 kp(k)\mathrm{d}k$ なので，あとは計算するだけ．

$$\begin{aligned}
&\int_0^1 kp(k)\mathrm{d}k \\
&= \int_0^1 Nk^N \mathrm{d}k \\
&= \left[\frac{N}{N+1}k^{N+1}\right]_0^1 \\
&= \frac{N}{N+1}
\end{aligned}$$

3　最近点対の距離

最近点対の距離が k 以上になる確率を $P(k)$ とおく．[0,1] から最近点対の距離が k 以上になるように点を N 個選ぶ操作は，[0, 1 - (N-1)k] から点を N 個選ぶ操作と対応する (各区間から k だけ抜くことを考えれば良い)．よって

$$P(k) = (1 - (N-1)k)^N$$

最近点対の距離を M とすると P は M の累積密度関数に相当する．よって P を微分すれば M の確率密度関数 p がもとまる．ここで $P(k)$ が M が k **以上**となる確率であることに注意しなければならない．つまり，符号をひっくり返す必要がある．

$$p(k) = -\frac{\mathrm{d}P}{\mathrm{d}k}(k)$$

求めたい値は $\int_0^{\frac{1}{N-1}} kp(k)\mathrm{d}k$ なので，あとは計算するだけ．部分積分をすると便利.

$$\begin{aligned}
&\int_0^{\frac{1}{N-1}} kp(k)\mathrm{d}k \\
&= [-kP(k)]_0^{\frac{1}{N-1}} + \int_0^{\frac{1}{N-1}} P(k)\mathrm{d}k \\
&= \int_0^{\frac{1}{N-1}} P(k)\mathrm{d}k \\
&= \left[\frac{(1-(N-1)x)^{N+1}}{1-N^2}\right]_0^{\frac{1}{N-1}} \\
&= \frac{1}{N^2-1}
\end{aligned}$$

4　全点対の距離の平均

$$
\begin{aligned}
&E\left[\sum_{i=1}^{N}\sum_{j=i+1}^{N}|X_i - X_j|\right]\frac{2}{N(N-1)} \\
&= \sum_{i=1}^{N}\sum_{j=i+1}^{N} E\left[|X_i - X_j|\right]\frac{2}{N(N-1)}
\end{aligned}
$$

ここで，$E\left[|X_i - X_j|\right]$ は i, j によらないので

$$
\begin{aligned}
&= E[|X_1 - X_2|] \\
&= \frac{1}{3}
\end{aligned}
$$

5 k-th 点対の距離

$[0,1]$ から N 点取った時の k 番目に短い区間の大きさの期待値を $f_k(N)$ とする．問 3 で求めた通り $f_N(1)=\dfrac{1}{N^2-1}$ である．
k 番目に短い区間の大きさの期待値は，$N-1$ 個の区間全てから一番短い区間の大きさを引くことを考えると，

$$\begin{aligned}
&f_k(N)\\
&=f_1(N)+(1-(N-1)f_1(N))f_{k-1}(N-1)\\
&=\frac{1}{N^2-1}+\frac{N}{N+1}f_{k-1}(N-1)
\end{aligned}$$

具体的に計算してみると．

$$\begin{aligned}
f_1(N) &= \frac{1}{N^2-1}\\
&= \frac{1}{N+1}\frac{1}{N-1}\\
f_2(N) &= \frac{1}{N^2-1}+\frac{N}{N+1}f_1(N-1)\\
&= \frac{1}{N^2-1}+\frac{N}{N+1}\frac{1}{N}\frac{1}{N-2}\\
&= \frac{1}{N^2-1}+\frac{1}{(N+1)(N-2)}\\
&= \frac{1}{N+1}\left(\frac{1}{N-1}+\frac{1}{N-2}\right)\\
f_3(N) &= \frac{1}{N^2-1}+\frac{N}{N+1}f_2(N-1)\\
&= \frac{1}{N^2-1}+\frac{N}{N+1}\frac{1}{N}\left(\frac{1}{N-2}+\frac{1}{N-3}\right)\\
&= \frac{1}{N+1}\left(\frac{1}{N-1}+\frac{1}{N-2}+\frac{1}{N-3}\right)\\
&\ \vdots\\
f_k(N) &= \frac{1}{N+1}\left(\frac{1}{N-1}+\frac{1}{N-2}+\cdots+\frac{1}{N-k}\right)
\end{aligned}$$

6　平均より長い点対の個数

最小値と最大値を L, R とする．R, L が決まっているきの条件付き期待値を考える．スケールが違うだけなので，平均より大きいものの個数は条件なしの期待値と等しい．よって，$L = 0, R = 1$ の場合のみ考えれば良い．これは区間を $N-2$ 箇所で切って，できる $N-1$ 個の区間のうち長さが $\frac{1}{N-2}$ 以上のものの個数である．$l = \frac{1}{N-2}$ とする．一番左の区間が長さ l 以上になる確率は $(1-l)^{N-2}$ である．ある切る場所について，そこから右に長さ l に何もなく，$[0,1]$ 区間の端が来ることもない確率は $(1-l)^{N-2}$ である．よって $N-2$ 個の区間についてそれが長さ l 以上になる確率は $(1-l)^{N-2}$．よってそのような区間の個数の期待値は $(N-2)(1-l)^{N-2} = (N-2)(1-\frac{1}{N-2})^{N-2}$ になる．

ちなみに $N \to \infty$ のとき $\frac{\text{答え}}{N}$ は e^{-1} に収束する．