Math of Machine Learning Olympiad 2021

May 2021

Problem 1

Find the general solution of the ODE system

$$\begin{cases} \dot{x} = x - y + z \\ \dot{y} = x + y - z, \\ \dot{z} = 2x - y. \end{cases}$$

Problem 2

Let $\Gamma = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, $P(x, y) = 4x^3 \sin y$, $Q(x, y) = x^3 + x^4 \cos y$. Compute the integral

$$\oint_\Gamma P(x, y) \mathrm{d}x + Q(x, y) \, \mathrm{d}y.$$

Problem 3

Let $\xi_1, \dots, \xi_n$ be independent identically distributed random variables taking their values in $\mathbb{N}$. Let $\mathbb{P}(\xi_1 \text{ is even}) = p \in (0, 1)$. Denote $\eta_n = \xi_1 + \dots + \xi_n$. Find the probability of the event that η_n is even.

Problem 4

Let $\xi_1, \dots, \xi_n, \dots,$ be a sequence of two-dimensional independent identically distributed $\mathcal{N}(0, \mathrm{I})$ random variables. Let

$$A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}.$$

Define X_∞ as the $L_2(\mathbb{P})$ limit of the sequence

$$X_n = \sum_{k=1}^{n} \left(\mathrm{I} - \frac{1}{10} A \right)^k \xi_k.$$

Find $\mathrm{Var}(X_\infty)$.

Problem 5

Suppose that we are given the map with n cities, and some of them are connected with roads. Let A be the corresponding $n \times n$ adjacency matrix. We assume that $A[i, j] = A[j, i]$, with values $A[i, j] \geqslant 0$ giving the travel time from city i to j if there is a road between them, and $A[i, j] = \infty$ otherwise. Assume also that

$A[i,i]=0$ for any $i\in\{1,\ldots,n\}$. Suggest an algorithm to find the minimal travel time between all city pairs $(i,j), i,j\in\{1,\ldots,n\}$. If there is no path between cities i and j, algorithm should return ∞. The number of points you get depends on the complexity of the suggested algorithm.

Problem 6

Let $\xi_1,\ldots,\xi_n,\ldots$ be a sequence of independent random variables such that $\xi_n \xrightarrow{\mathbb{P}} \xi$ as $n\to\infty$. Prove that ξ is degenerate, i.e. $\mathbb{P}(\xi=a)=1$ for some $a\in\mathbb{R}$.

Solutions

Problem 1

The characteristic polynomial is equal to

$$\det\begin{pmatrix}1-\lambda & -1 & 1\\ 1 & 1-\lambda & -1\\ 2 & -1 & -\lambda\end{pmatrix} = -(\lambda+1)(\lambda-1)(\lambda-2).$$

The matrix has three eigenvalues $\lambda_1=-1$, $\lambda_2=1$, and $\lambda_3=2$. The corresponding eigenvectors are $h_1=(1,-3,-5)^T$, $h_2=(1,1,1)^T$, and $h_3=(1,0,1)^T$, respectively. Then the general solution of the ODE system is given by

$$\begin{pmatrix}x(t)\\ y(t)\\ z(t)\end{pmatrix} = C_1e^{-t}\begin{pmatrix}1\\ -3\\ -5\end{pmatrix} + C_2e^{t}\begin{pmatrix}1\\ 1\\ 1\end{pmatrix} + C_3e^{2t}\begin{pmatrix}1\\ 0\\ 1\end{pmatrix}, \quad C_1,C_2,C_3\in\mathbb{R}.$$

Problem 2

We compute the integral in the counter-clockwise direction. Apply Green's formula:

$$\oint\limits_{x^2+y^2=1} P(x,y)\mathrm{d}x + Q(x,y)\,\mathrm{d}y = \iint\limits_{x^2+y^2\leqslant 1}\left(\frac{\partial Q(x,y)}{\partial x} - \frac{\partial P(x,y)}{\partial y}\right)\mathrm{d}x\mathrm{d}y = \iint\limits_{x^2+y^2\leqslant 1} 3x^2\,\mathrm{d}x\mathrm{d}y.$$

Using the polar coordinates $x=r\cos\varphi$, $y=r\sin\varphi$ where $0\leqslant r\leqslant 1$, $0\leqslant\varphi\leqslant 2\pi$, we obtain

$$\iint\limits_{x^2+y^2\leqslant 1} 3x^2\,\mathrm{d}x\mathrm{d}y = 3\int\limits_0^1 r^3\mathrm{d}r\int\limits_0^{2\pi}\cos^2\varphi\,\mathrm{d}\varphi = \frac{3}{4}\int\limits_0^{2\pi}\frac{1+\cos(2\varphi)}{2}\,\mathrm{d}\varphi = \frac{3\pi}{4}.$$

Problem 3

Let ζ_n be the number of odd ξ_i's, that is, $\zeta_n=\sum\limits_{i=1}^{n}\mathbb{1}(\xi_i \text{ is odd})$. Note that η_n is even if and only if ζ_n is even. On the other hand, ζ_n has a binomial distribution $\mathrm{Binom}(n,1-p)$ as the sum of independent identically distributed Bernoulli random variables. This yields

$$\mathbb{P}(\eta_n \text{ is even}) = \mathbb{P}(\zeta_n \text{ is even}) = \sum_{k:0\leqslant 2k\leqslant n}\binom{n}{2k}(1-p)^{2k}p^{n-2k}.$$

It remains to compute the sum in the right hand side. Note that

$$1=(p+1-p)^n=\sum_{k=0}^{n}\binom{n}{k}(1-p)^kp^{n-k}$$

and

$$(2p-1)^n = (p-(1-p))^n = \sum_{k=0}^{n} \binom{n}{k} (-1)^k (1-p)^k p^{n-k}.$$

Then

$$\begin{aligned}\sum_{k:0\leqslant 2k\leqslant n} \binom{n}{2k} (1-p)^{2k} p^{n-2k} &= \frac{1}{2}\left(\sum_{k=0}^{n} \binom{n}{k} (1-p)^k p^{n-k} + \sum_{k=0}^{n} \binom{n}{k} (-1)^k (1-p)^k p^{n-k}\right) \\ &= \frac{1}{2}\left(1+(2p-1)^n\right).\end{aligned}$$

Problem 4

Since $\xi_1, \ldots, \xi_n, \ldots$ are independent,

$$\mathrm{Var}(X_n) = \sum_{k=1}^{n} \mathrm{Var}\left(\left(\mathrm{I} - \frac{1}{10}A\right)^k \xi_k\right)$$

Since A is symmetric,

$$\mathrm{Var}(X_n) = \sum_{k=1}^{n} \left(\mathrm{I} - \frac{1}{10}A\right)^k \mathrm{Var}(\xi_k) \left(\mathrm{I} - \frac{1}{10}A\right)^k = \sum_{k=1}^{n} \left(\mathrm{I} - \frac{1}{10}A\right)^{2k}.$$

It is easy to check that X_n indeed converges in $L_2(\mathbb{P})$ and

$$\mathrm{Var}(X_\infty) = \sum_{k=1}^{\infty} \left(\mathrm{I} - \frac{1}{10}A\right)^{2k}.$$

To compute this sum, note that $\mathrm{I} - \frac{1}{10}A = Q\Lambda Q^\top$, where $QQ^\top = Q^\top Q = \mathrm{I}$,

$$Q = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}, \quad \Lambda = \begin{pmatrix}0.5 & 0\\ 0 & 0.7\end{pmatrix}.$$

Hence,

$$\sum_{k=1}^{\infty} \Lambda^{2k} = \begin{pmatrix}\frac{0.25}{1-0.25} & 0\\ 0 & \frac{0.49}{1-0.49}\end{pmatrix} = \begin{pmatrix}\frac{1}{3} & 0\\ 0 & \frac{49}{51}\end{pmatrix} = \frac{1}{51}\begin{pmatrix}17 & 0\\ 0 & 49\end{pmatrix},$$

and

$$\mathrm{Var}(X_\infty) = Q\left(\sum_{k=1}^{\infty} \Lambda^{2k}\right) Q^\top = \begin{pmatrix}\frac{11}{17} & -\frac{16}{51}\\ -\frac{16}{51} & \frac{11}{17}\end{pmatrix}.$$

Problem 5

The simplest way to solve the problem is to use Floyd-Warshall algorithm. Let $B \in \mathbb{R}^{n\times n}$ be the shortest path matrix to be calculated. The algorithm runs in $\mathcal{O}(n^3)$ iterations with $\mathcal{O}(n^2)$ memory complexity. Pseudocode is provided in Algorithm 1.

Problem 6

Denote $\varphi_n(t)$ the characteristic function of ξ_n and $\varphi_\xi(t)$ the characteristic function of ξ. Since convergence in probability implies weak convergence, for any $t \in \mathbb{R}$,

$$\varphi_n(t) \to \varphi_\xi(t)\,, n \to \infty\,. \tag{1}$$

```
Input: Matrix A ∈ ℝ^{n×n}
Result: Matrix B ∈ ℝ^{n×n}
Initialize B[i, j] = A[i, j], i, j ∈ {1, ..., n};
for k = 1 to n do
    for i = 1 to n do
        for j = 1 to n do
            B[i, j] = min(B[i, j], B[i, k] + B[k, j]);
```

On the other hand,

$$\xi_n - \xi_{n+1} \xrightarrow{\mathbb{P}} 0\,, \quad n \to \infty\,.$$

Hence, for any $t \in \mathbb{R}$,

$$\varphi_{\xi_n - \xi_{n+1}}(t) = \mathbb{E}[\exp\left(it(\xi_n - \xi_{n+1})\right)] = \varphi_n(t)\varphi_{n+1}(-t) \to 1\,, n \to \infty\,.$$

This implies $|\varphi_n(t)| \to 1$. Combining with (1), $|\varphi_\xi(t)| = 1$ for any t, and thus ξ is degenerate.