1. Let X be a random variable with probability density function $f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$.

A sample of size 3 is randomly selected from this distribution. Let Y be a random variable representing the median value from the sample; calculate the variance of Y.

解: (Ross S. M. "A First Course in Probability", Tenth Edition, Pearson. Section 6.6 Order Statistics, c.f. Exam P syllabus)

Let $X_1, X_2, \ldots, X_n$ be n i.i.d. continuous r.v. having a common density function f and distribution F. Define

$$\begin{aligned} X_{(1)} &= \text{smallest of } X_1,\, X_2,\, \ldots,\, X_n \\ X_{(2)} &= \text{second smallest of } X_1,\, X_2,\, \ldots,\, X_n \\ &\vdots \\ X_{(j)} &= j\text{-th smallest of } X_1,\, X_2,\, \ldots,\, X_n \\ &\vdots \\ X_{(n)} &= \text{largest of } X_1,\, X_2,\, \ldots,\, X_n \end{aligned}$$

The ordered values $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ are known as the *order statistics* corresponding to the r.v.s $X_1, X_2, \ldots, X_n$. The joint pdf of the order statistics is obtained by noting that $X_{(1)}, X_{(2)}, X_{(n)}$ will take on the values $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$ iff for some permutation $(i_1, i_2, \ldots, i_n)$ of $(1, 2, \ldots, n)$, $X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}$. Since for any permutation $(i_1, i_2, \ldots, i_n)$ of $(1, 2, \ldots, n)$,

$$\begin{aligned} &\mathsf{P}\Big\{x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2},\ x_{i_2} - \frac{\varepsilon}{2} < X_2 < x_{i_2} + \frac{\varepsilon}{2},\ \ldots,\ x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2}\Big\} \\ \approx\ & \varepsilon^n f_{X_1, X_2, \ldots, X_n}(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) \\ =\ & \varepsilon^n f(x_{i_1}) \cdot f(x_{i_2}) \cdots f(x_{i_n}) \\ =\ & \varepsilon^n f(x_1) \cdot f(x_2) \cdots f(x_n) \end{aligned}$$

It follows that, for $x_1 < x_2 < \ldots < x_n$,

$$\begin{aligned} &\mathsf{P}\Big\{x_1 - \frac{\varepsilon}{2} < X_{(1)} < x_1 + \frac{\varepsilon}{2},\ x_2 - \frac{\varepsilon}{2} < X_{(2)} < x_2 + \frac{\varepsilon}{2},\ \ldots,\ x_n - \frac{\varepsilon}{2} < X_{(n)} < x_n + \frac{\varepsilon}{2}\Big\} \\ \approx\ & n!\, \varepsilon^n f(x_1) \cdot f(x_2) \cdots f(x_n) \end{aligned}$$

Dividing by ε^n and letting $\varepsilon \to 0$ yields

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_1, x_2, \ldots, x_n) = n!\, f(x_1) \cdot f(x_2) \cdots f(x_n), \quad x_1 < x_2 < \ldots < x_n$$

The density function of $X_{(j)}$ can be obtained as follows: in order for $X_{(j)}$ to equal x, it is necessary for $j-1$ of the n values $X_1, X_2, \ldots, X_n$ to be less than x, $n-j$ of them to be greater than x, and exactly 1 of them to equal x; the probability of this group is $F(x)^{j-1}(1-F(x))^{n-j} f(x)$, and there are $\dfrac{n!}{(j-1)!(n-j)!}$ such groups, so the probability density function of $X_{(j)}$ is given by

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!\,(j-1)!}\, F(x)^{j-1}(1-F(x))^{n-j} f(x)$$

Now $n = 3$, $j = 2$, so $f(y)$, the pdf of Y, is simply $\dfrac{3!}{(3-2)!\,(2-1)!}\,(y^2)^{2-1}(1-y^2)^{3-2}\,2y = 6\,y^2 \cdot (1-y^2) \cdot 2y = 12(y^3 - y^5)$, $\mathsf{E}\{Y\} = \displaystyle\int_0^1 y \cdot f(y)\,\mathrm{d}y = 12\int_0^1 y^4 - y^6\,\mathrm{d}y = 12\left(\frac{1}{5} - \frac{1}{7}\right) = \frac{24}{35}$, $\mathsf{E}\{Y^2\} = \displaystyle\int_0^1 y^2 \cdot f(y)\,\mathrm{d}y = 12\int_0^1 y^5 - y^7\,\mathrm{d}y = 12\left(\frac{1}{6} - \frac{1}{8}\right) = \frac{1}{2}$, so $\operatorname{var} Y = \mathsf{E}\{Y^2\} - (\mathsf{E}\{Y\})^2 = \dfrac{1}{2} - \dfrac{24^2}{35^2} = 0.029795918$.