Probability & Statistics (1)

Jointly Distributed Random Variables (I)

Asst. Prof. Chan, Chun-Hsiang

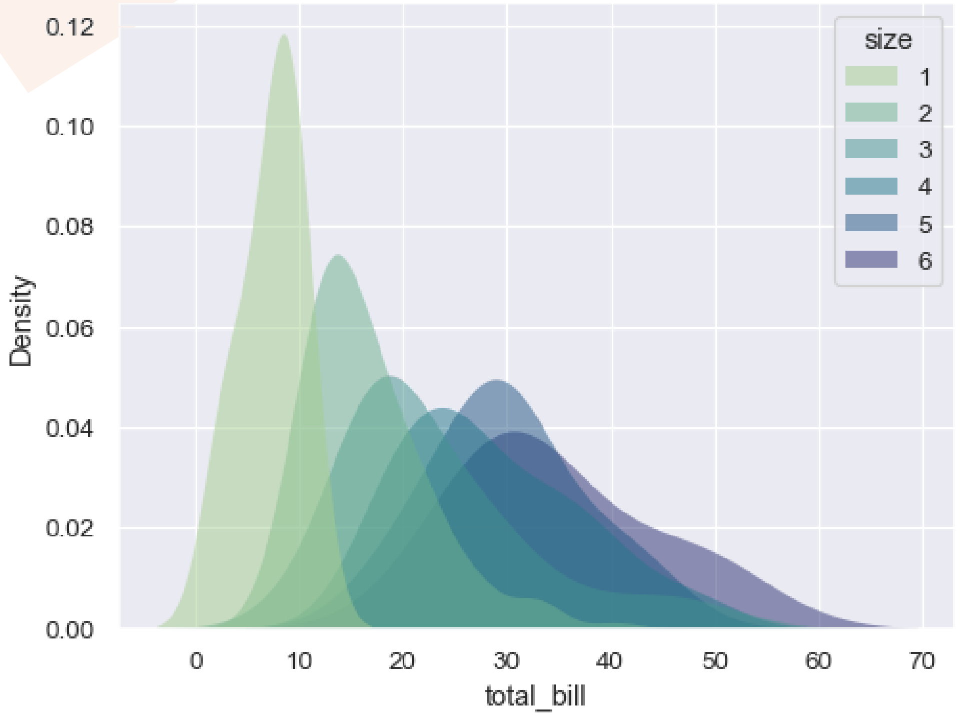
Master program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Intelligent Computing and Big Data, Chung Yuan Christian University, Taoyuan, Taiwan
Undergraduate program in Applied Artificial Intelligence, Chung Yuan Christian University, Taoyuan, Taiwan

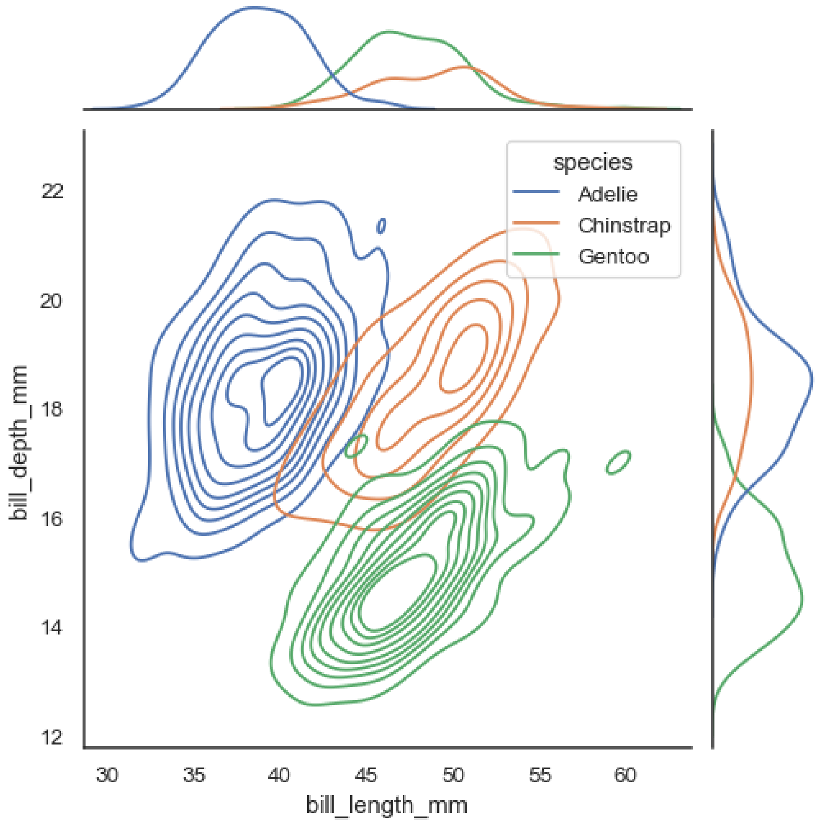
Outlines

1. Joint Distribution Functions
2. Independent Random Variables
3. Sums of Independent Random Variables
4. Conditional Distribution: Discrete Case
5. Conditional Distribution: Continuous Case
6. Order Statistics
7. Joint Probability Distribution of Functions of Random Variables
8. [#11] Assignment
9. Reference
10. Question Time

Joint Distribution Functions

- 之前我們都只有考慮到一個隨機變數，如果今天我們需要同時考慮兩個隨機變數結合起來的機率分布，這個時候我們就會使用到Joint Distribution Functions。





Joint Distribution Functions

- 假設我們今天給定兩個隨機變數： X與Y，那麼他們的joint cumulative probability distribution function可以被定義為:

$$F(a,b) = P\{X \le a, Y \le b\}, where -\infty < a, b < \infty$$

- 如果我們想要求X的distribution的話:

$$F_X(a) = P\{X \le a\} = P\{X \le a, Y \le \infty\} = P\left(\lim_{b\to\infty}\{X \le a, Y \le \infty\}\right)$$

$$= \lim_{b\to\infty} P\{X \le a, Y \le \infty\} = \lim_{b\to\infty} F(a,b) = F(a,\infty)$$

- 同理可證，如果我們要求Y的distribution的話:

$$F_Y(b) = P\{Y \le b\} = \lim_{a\to\infty} F(a,b) = F(\infty,b)$$

Joint Distribution Functions

- 通常我們會稱F_X與F_Y的distribution functions為X與Y的marginal distributions。

$$
\begin{aligned}
&P\{X > a, Y > b\} = 1 - P(\{X > a, Y > b\}^c) \\
&= 1 - P(\{X > a\}^c \cup \{Y > b\}^c) \\
&= 1 - P(\{X \le a\} \cup \{Y \le b\}) \\
&= 1 - [P\{X \le a\} + P\{Y \le b\} - P\{X \le a, Y \le b\}] \\
&= 1 - F_X(a) - F_Y(b) + F(a, b)
\end{aligned}
$$

Joint Distribution Functions

- 如果今天我們的隨機變數X與Y都是離散的時候，則X與Y 的joint probability mass function為

$$p(x,y) = P\{X = x, Y = y\}$$

- X的probability mass function可以從$p(x,y)$得出

$$p_X(x) = P\{X = x\} = \sum_{y:p(x,y)>0} p(x,y)$$

- 所以Y的probability mass function為:

$$p_Y(y) = P\{Y = y\} = \sum_{x:p(x,y)>0} p(x,y)$$

Joint Distribution Functions

- **範例一**

假設隨機從摸彩桶取出3張摸彩券，而這個桶子中有3張一獎，4張二獎與5張銘謝惠顧。我們令隨機變數X與Y分別為抽到一獎與二獎的張數，X與Y的joint PMF為: $p(i,j) = P\{X = i, Y = j\}$，試求其joint PMF的數值。

Solution:

$$p(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}; p(0,1) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}; p(0,2) = \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220}; p(0,3) = \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}$$

$$p(1,0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220}; p(2,0) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220}; p(3,0) = \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220};$$

Joint Distribution Functions

$$p(1,1)=\frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}}=\frac{30}{220};p(1,2)=\frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}}=\frac{18}{220};p(2,1)=\frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}}=\frac{12}{220};$$

Table 1 $P\{X=i,Y=j\}$

i \ j	**0**	**1**	**2**	**3**	**Row sum**=$P\{X=i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column sum=$P\{Y=j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

Joint Distribution Functions

- **範例二**

假設今天做人口抽樣調查，15%家庭中沒有小孩，20%有一個小孩，35%有兩個小孩，30%有三個小孩。假設小孩是男生或是女生的機率相等，那麼隨機抽樣一個家庭，B為小孩為男生的數量；G為小孩為女生的數量，則$P\{B=i,G=j\}$的機率為何?

Solution:

$$P\{B=0,G=0\}=P\{no\ children\}=0.15$$

Joint Distribution Functions

Table $P\{B = i, G = j\}$

i \ j	0	1	2	3	Row Sum $= P\{B = j\}$
0					0.3750
1					0.3875
2					0.2000
3					0.0375
Column Sum $= P\{G = j\}$	0.3750	0.3875	0.2000	0.0375	

Joint Distribution Functions

- We say that X and Y are jointly continuous if there exists a function $f(x,y)$, defined for all real x and y, having the property that, for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane),

$$P\{(X,Y) \in C\} = \iint_{(x+y)\in C} f(x,y)dxdy$$

- The function $f(x,y)$ is called the joint probability density function of X and Y. If A and B are any set of real numbers, then, by defining $C = \{(x,y): x \in A, y \in B\}$

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)dxdy$$

Joint Distribution Functions

$$F(a,b) = P\{X \in (-\infty, a], Y \in (-\infty, b]\} = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)dxdy$$

Differentiation

$$f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$$

$$P\{a < X < a + da, b < Y < b + db\} = \int_{-\infty}^{b+db}\int_{-\infty}^{a+da} f(x,y)dxdy$$

$$\approx f(a,b)dadb$$

Joint Distribution Functions

$$P\{X \in A\} = P\{X \in A, Y \in (-\infty, \infty)\} = \int_A \int_{-\infty}^{\infty} f(x, y)dydx = \int_A f_X(x)dx$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy\,; \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Joint Distribution Functions

- **範例三**

給定X與Y的joint density function為

$$f(x,y)=\begin{cases}2e^{-x}e^{-2y}, 0<x<\infty, 0<y<\infty\\ 0, otherwise\end{cases}$$

試問 $(a)\ P\{X>1,Y<1\}$; $(b)\ P\{X<Y\}$; $(c)\ P\{X<a\}$

Solution:

$$P\{X>1,Y<1\}=\int_0^1\int_1^\infty 2e^{-x}e^{-2y}\,dxdy=\int_0^1 2e^{-2y}\left(-e^{-x}\Big|_1^\infty\right)\mathrm{dy}$$

$$=\mathrm{e}^{-1}\int_0^1 2e^{-2y}\,dy=e^{-1}(1-e^{-2})$$

Joint Distribution Functions

(b)

$$P\{X < Y\} = \iint_{(x,y):x<y} 2e^{-x}e^{-2y}\,dxdy = \int_0^\infty \int_0^y 2e^{-x}e^{-2y}\,dxdy$$

$$= \int_0^\infty 2e^{-y}(1 - e^{-y})dy = \int_0^\infty 2e^{-y}dy - \int_0^\infty 2e^{-3y}dy = 1 - \frac{2}{3} = \frac{1}{3}$$

(c)

$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-x}e^{-2y}\,dydx = \int_0^a e^{-x}dx = 1 - e^{-a}$$

Joint Distribution Functions

- **範例四**

X 與 Y的joint probability定義如下:

$$f(x,y) = \begin{cases} e^{-(x+y)}, 0 < x < \infty, 0 < y < \infty \\ 0, otherwise \end{cases}$$

試問隨機變數X/Y的PDF為何?

Joint Distribution Functions

Solution:

令$a > 0$

$$F_{\frac{X}{Y}}(a) = P\left\{\frac{X}{Y} \le a\right\} = \iint_{\frac{x}{y}\le a} e^{-(x+y)}dxdy = \int_0^\infty \int_0^{ay} e^{-(x+y)}dxdy$$

$$= \int_0^\infty (1 - e^{-ay})e^{-y}dy = \{-e^{-y} + \frac{e^{-(a+1)y}}{a+1}\}$$

Differentiation shows that the density function of X/Y is given by

$$f_{\frac{X}{Y}}(a) = \frac{1}{(a+1)^2}, where\ 0 < a < \infty$$

Joint Distribution Functions

We can also define joint probability distributions for n random variables in exactly the same manner as we did for $n = 2$. For instance, the joint cumulative probability distribution function $F(a_1, a_2, \ldots, a_n)$ of the n random variables $X_1, X_2, \ldots, X_n$ is defined by

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}$$

Further, the n random variables are said to be jointly continuous if there exists a function $f(x_1, x_2, \ldots, x_n)$, called the joint probability density function, such that, for any set C in n-space,

$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \iint_{(x_1, x_2, \ldots, x_n) \in C} \cdots \int f(x_1, \ldots x_n) dx_1 dx_2 \ldots dx_n$$

Joint Distribution Functions

In particular, for any n sets of real numbers A_1, A_2, …, A_n

$$P\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\}$$

$$= \int_{A_n} \int_{A_{n-1}} \dots \int_{A_1} f(x_1, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Independent Random Variables

隨機變數X與Y彼此為獨立(independent)，則任何兩個集合的實數A與B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

令$E_A = \{X \in A\}, E_B = \{Y \in B\}$，因此對於所有的$a, b$

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

因此，X與Y的joint distribution function F，且X與Y相互獨立

$$F(a, b) = F_X(a)F_Y(b), for\ all\ a, b$$

當今天為離散隨機變數的時候

$$p(x, y) = p_X(x)p_Y(y), for\ all\ a, b$$

Independent Random Variables

$$P\{X \in A, Y \in B\} = \sum_{y \in B} \sum_{x \in A} p(x, y) = \sum_{y \in B} \sum_{x \in A} p_X(x)\, p_Y(y)$$

$$= \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) = P\{Y \in B\} P\{X \in A\}$$

$$f(x, y) = f_X(x) f_Y(y), for\ all\ x, y$$

Independent Random Variables

- **範例五**

假設有$n+m$次獨立試驗，成功的機會為p。X為在前面n次試驗中成功的次數，Y為在後面m次試驗中成功的次數。X與Y相互獨立。

$$P\{X=x,Y=y\}=\binom{n}{x}p^x(1-p)^{n-x}\binom{m}{y}p^y(1-p)^{m-y}$$

$$=P\{X=x\}P\{Y=y\}$$

相對地，X與Y為相互獨立，其中Z為在$n+m$有多少次成功。

Independent Random Variables

- **範例六**

給一段的時間區間中，進去郵局的男女性別機率為p與$1-p$為一個 Poisson random variable，其參數為λp與$\lambda(1-p)$。

Solution:

令隨機變數X與Y代表走進郵局的男女性分別數量。

$$P\{X=i,Y=j\}$$
$$=P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\}$$
$$+P\{X=i,Y=j|X+Y\neq i+j\}P\{X+Y\neq i+j\}$$

According to …

$$P(E)=P(E|F)P(F)+P(E|F^c)P(F^c)$$

Independent Random Variables

Since $P\{X = i, Y = j | X + Y \neq i + j\} = 0$

$$P\{X = i, Y = j\} = P\{X = i, Y = j | X + Y = i + j\} P\{X + Y = i + j\}$$

Then,

$$P\{X + Y = i + j\} = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

Given that $i + j$ people do enter the post office, since each person entering will be male with probability p, it follows that the probability that exactly i of them will be male.

$$P\{X = i, Y = j | X + Y = i + j\} = \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

Independent Random Variables

$$P\{X=i, Y=j\} = \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

$$= e^{-\lambda} \frac{(\lambda p)^i}{i!\,j!} [\lambda(1-p)]^j$$

$$P\{X=i\} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!}$$

Similarly,

$$P\{Y=j\} = e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!}$$

Independent Random Variables

- **範例七**

一對情侶約去看世足，假設他們到達時間的時間為time uniformly distribution從中午12點到下午1點，試問兩個到達時間差距超過十分鐘的機率:

Solution:

$$2P\{X+10<Y\}=2\iint_{x+10<y} f(x,y)dxdy=2\iint_{x+10<y} f_X(x)f_Y(y)dxdy$$

$$=2\int_{10}^{60}\int_{0}^{y-10}\left(\frac{1}{60}\right)^2 dxdy=\frac{2}{60^2}\int_{10}^{60}(y-10)dy=\frac{25}{36}$$

Independent Random Variables

- **Proposition 1**

The continuous (discrete) random variables X and Y are independent if and only if their joint probability density function can be expressed.

$$f(x, y) = h(x)g(x), where \ -\infty < x < \infty, -\infty < y < \infty$$

Proof:

Let us give the proof in the continuous case. First, note that independence implies that the joint density is the product of the marginal densities of X and Y, so the preceding factorization will hold when the random variables are independent.

Independent Random Variables

Now, suppose that

$$f_{X,Y}(x,y) = h(x)g(y)$$

Then,

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = \int_{-\infty}^{\infty} h(x)\,dx \int_{-\infty}^{\infty} g(y)\,dy = C_1C_2$$

where $C_1 = \int_{-\infty}^{\infty} h(x)dx$ and $C_2 = \int_{-\infty}^{\infty} g(y)\,dy$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = C_2h(x)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = C_1g(y)$$

Since $C_1C_2 = 1$, it follows that $f(x,y) = f_X(x)f_Y(y)$

Independent Random Variables

- 範例八

If the joint density function of X and Y is

$$f(x, y) = 6e^{-2x}e^{-3y}, where\ 0 < x < \infty, 0 < y < \infty$$

And is equal to 0 outside this region, are the random variables independent? What if the joint density function is

$$f(x, y) = 24xy, where\ 0 < x < 1, 0 < y < 1, 0 < x + y < 1$$

and is equal to 0 otherwise.

Independent Random Variables

Solution:

In the first instance, the joint density function factors, and thus the random variables, are independent (with one being exponential with rate 2 and the other exponential with rate 3. In the second instance, because the region in which the joint density is nonzero cannot be addressed in the form $x \in A, y \in B$, the joint density does not factor, so the random variables are note independent.

$$I(x,y) = \begin{cases} 1, if\ 0 < x < 1, 0 < y < 1, 0 < x + y < 1 \\ 0, otherwise \end{cases}$$

And writing $f(x,y) = 24xyI(x,y)$

Independent Random Variables

- We need to define more than two random variables. In general, the n random variables $X_1, X_2, \dots, X_n$ are said to be independent if, for all sets of real numbers $A_1, A_2, \dots, A_n$,

$$P\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

As a before, it can be shown that this condition is equivalent to

$$P\{X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n\} = \prod_{i=1}^{n} P\{X_i \leq a_i\}, for\ all\ a_1, a_2, \dots, a_n$$

Independent Random Variables

- **範例九**

令X, Y, Z為獨立與uniform distributed over (0,1). Compute $P\{X \geq YZ\}$.

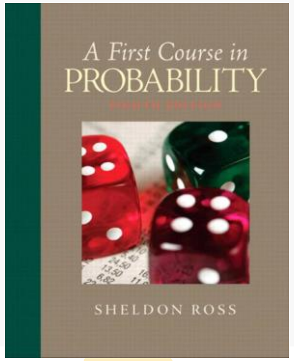
Solution:

Since $f_{X,Y,Z}(x,y,z) = f_Z(x)f_Z(y)f_Z(z) = 1, where\ 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$

$$P\{X \geq YZ\} = \iiint_{x \geq yz} f_{X,Y,Z}(x,y,z)dxdydz$$

$$= \int_0^1 \int_0^1 \int_{yz}^1 dxdydz = \int_0^1 \int_0^1 (1 - yz)dydz = \int_0^1 \left(1 - \frac{z}{2}\right) dz = \frac{3}{4}$$

[#11] Assignment



- Selected Problems from Sheldon Ross Textbook [1].

6.2. Suppose that 3 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let X_i equal 1 if the ith ball selected is white, and let it equal 0 otherwise. Give the joint probability mass function of

(a) X_1, X_2;

(b) X_1, X_2, X_3.

6.3. In Problem 2, suppose that the white balls are numbered, and let Y_i equal 1 if the ith white ball is selected and 0 otherwise. Find the joint probability mass function of

(a) Y_1, Y_2;

(b) Y_1, Y_2, Y_3.

6.10. The joint probability density function of X and Y is given by

$$f(x, y) = e^{-(x+y)} \qquad 0 \leq x < \infty, 0 \leq y < \infty$$

Find (a) $P\{X < Y\}$ and (b) $P\{X < a\}$.

[1] Sheldon Ross. A First of Course in Probability. 8th edition.

Reference

Ross, S. (2010). *A first course in probability*. Pearson.

The End

If you have any questions, please do not hesitate to ask me.

Thank you for your attention))