MTH102 Engineering Mathematics II

Lesson 7: Introduction to continuous random variables

Term: 2024

Outline

Outline

1 Continuous random variables

2 Mean and variance

3 Uniform distribution

Example 1

Suppose that the office hour is from 6:00 pm. to 8:00 pm. each Tuesday. A student is coming to ask questions during the office hour. Consider the arrival time of the student, then

$$S = \{\text{the time between 6:00 pm and 8:00 pm}\}.$$

Let X be the time difference in minutes between 6:00 pm and the arrival time, then

$$X(s) \in [0, 120], \ \forall s \in S.$$

Find the following probabilities:

(a) the student arrives between 7:00 pm and 7:10 pm;

(b) the student arrives at 7:00 pm.

Remark: if a sample space S is not discrete, for example $S = [a, b]$, then the probability of each outcome can not be listed one by one.

Probability density function

Definition

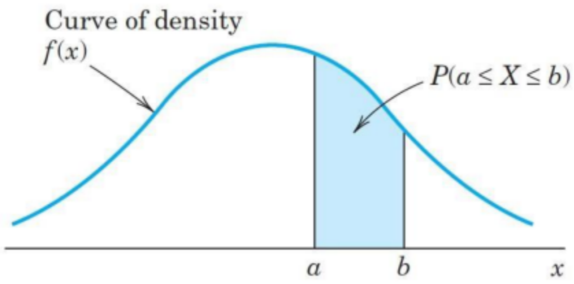
Let X be a random variable over the outcome space S which is an interval or union of intervals. X is called a **continuous random variable** *if there exists an integrable function $f(x)$ satisfying the following:*

(a) $f(x) \geq 0, \ \ x \in \mathbb{R}$;

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$;

(c) *If $a \leq b$, then*
$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

$f(x)$ *is called the* **probability density function (pdf)** *of X.*



Continuous random variable

For a continuous random variable X, the probability at a particular value $a \in \mathbb{R}$ is zero, i.e.

$$P(X = a) = 0.$$

Therefore,

$$P(X \leq a) = P(\{X < a\} \cup \{X = a\}) = P(X < a) + P(X = a) = P(X < a).$$

Consequently,

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

Cumulative distribution function

The **cumulative distribution function (cdf)** of a continuous random variable X with the pdf $f(x)$ is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R}.$$

Although $P(X \leq x) = P(X < x)$, we write $P(X \leq x)$ for consistency with the cdf of discrete random variables.

Properties:

- $F(x)$ is continuous, and for x values for which the derivative $F'(x)$ exists,

$$F'(x) = f(x).$$

- For any $(a, b) \subseteq S$,

$$P(a \leq X \leq b) = F(b) - F(a).$$

Example 2

Let X be a continuous random variable with pdf

$$f(x) = \begin{cases} 2x, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the cdf of X and sketch it.

(b) Find $P\left(\frac{1}{2} < X < \frac{3}{4}\right)$.

Sol: (a) $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$

When $0 < x < 1$,

$$\int_{-\infty}^{x} f(t)\, dt = \int_{0}^{x} 2t\, dt = 2 \times \frac{1}{2}\left[t^2\right]_0^x$$
$$= x^2$$

When $x \leq 0$, then

$$\int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{x} 0\, dt = 0.$$

When $x \geq 1$, then

$$\int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{0} f(t)\, dt + \int_{0}^{1} f(t)\, dt + \int_{1}^{x} f(t)\, dt$$

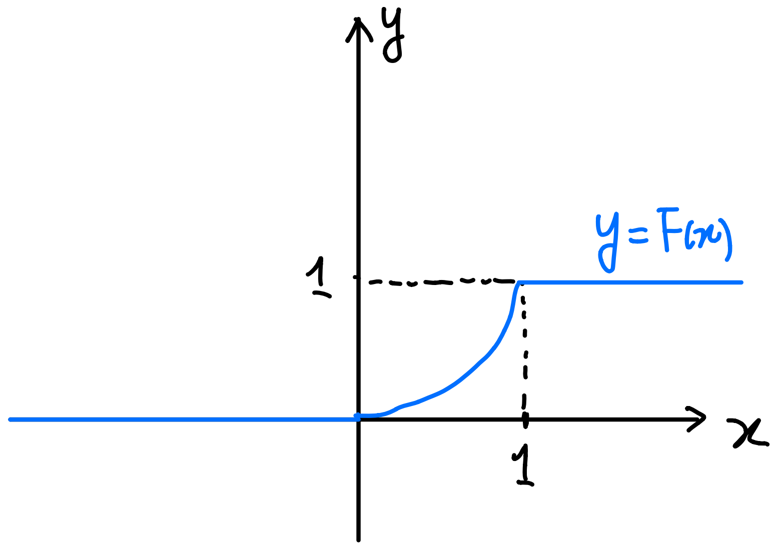
$$= \int_{-\infty}^{0} 0\, dt + \int_{0}^{1} 2t\, dt + \int_{1}^{x} 0\, dt$$

$$= 0 + 2 \times \frac{1}{2}\left[t^2\right]_0^1 + 0$$

$$= 1$$

In conclusion,

$$F(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ x^2, & \text{if } 0 < x < 1 \\ 1, & \text{if } x \geq 1 \end{cases}$$



(2) $\mathbb{P}\left(\frac{1}{2} < X < \frac{3}{4}\right) = \mathbb{P}\left(\frac{1}{2} < X \leq \frac{3}{4}\right)$

$= \mathbb{P}\left(X \leq \frac{3}{4}\right) - \mathbb{P}\left(X \leqslant \frac{1}{2}\right)$

$= F\left(\frac{3}{4}\right) - F\left(\frac{1}{2}\right)$

$= \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2$, from Part (1).

$= \frac{5}{16}$

Example 3

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2, \\ o & \text{otherwise} \end{cases}$$

1. Determine the value of c.
2. Find the cdf of X.
3. Find $P(X > 1)$.

Sol : (1) Since $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

$$\Leftrightarrow \int_0^2 c(4x-2x^2)\,dx = 1$$

$$\Leftrightarrow c\left(4\times\frac{1}{2}\left[x^2\right]_0^2 - 2\times\frac{1}{3}\left[x^3\right]_0^2\right) = 1$$

$$\Leftrightarrow c\left(8-\frac{16}{3}\right) = 1$$

$$\Leftrightarrow \frac{8}{3}c = 1$$

$$\Leftrightarrow c = \frac{3}{8}$$

(2)

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

- When $x \leq 0$, $F(x) = \int_{-\infty}^{x} 0\,dt = 0$

- When $0 < x < 2$, $F(x) = \int_0^x \frac{3}{8}(4t-2t^2)\,dt$

$$= \frac{3}{8}\left(4\times\frac{1}{2}\left[t^2\right]_0^x - 2\times\frac{1}{3}\left[t^3\right]_0^x\right)$$

$$= \frac{3}{8}\left(2x^2 - \frac{2}{3}x^3\right)$$

- When $x \geq 2$, $F(x) = \int_{-\infty}^{0} f(t)\,dt + \int_0^2 f(t)\,dt + \int_2^{\infty} f(t)\,dt$

$$= \int_{-\infty}^{0} 0\,dt + \int_0^2 \frac{3}{8}(4t-2t^2)\,dt + \int_2^{\infty} 0\,dt$$

$= 0 + 1 + 0$, from part (1)

$= 1.$

Therefore,

$$F(x) = \begin{cases} 0, & \text{if} \quad x \leq 0 \\ \frac{3}{8}\left(2x^2 - \frac{2}{3}x^3\right), & \text{if} \quad 0 < x < 2 \\ 1, & \text{if} \quad x \geq 2 \end{cases}$$

(3) $\mathbb{P}(X > 1) = 1 - \mathbb{P}(X \leq 1)$, from complement rule

$= 1 - F(1)$

$= 1 - \frac{3}{8}\left(2 \times 1^2 - \frac{2}{3} \times 1^3\right)$, from Part (2)

$= \frac{1}{2}$

Example 4

The lifetime in hours of a certain kind of radio tube is a random variable having probability density function given by

$$f(x) = \begin{cases} 0 & x \leq 100, \\ \frac{100}{x^2} & x > 100. \end{cases}$$

What is the probability that exactly 2 of 5 such tubes in a radio set will have to be replaced within the first 150 hours of operation?

Sol: Let X be the lifetime of 1 radio tube, and Y be the number of replaced tubes among 5 tubes within the 1st 150 h of operation. Then

$$Y \sim b(5, p),$$

with
$$p = \mathbb{P}(X < 150) = \int_{-\infty}^{150} f(t)\,dt$$

$$= \int_{-\infty}^{100} 0\,dt + \int_{100}^{150} \frac{100}{t^2}\,dt$$

$$= 0 + 100 \times (-1) \left[\frac{1}{t}\right]_{100}^{150}$$

$$= 100 \times \left(\frac{1}{100} - \frac{1}{150}\right)$$

$$= 1 - \frac{2}{3} = \frac{1}{3}$$

Therefore,
$$\mathbb{P}(Y=2) = \binom{5}{2} \times \left(\frac{1}{3}\right)^2 \times \left(1-\frac{1}{3}\right)^{5-2}$$

$$= \frac{5 \times 4}{2} \times \frac{1}{9} \times \left(\frac{2}{3}\right)^3$$

$$= \frac{80}{243}$$

The probability that exactly 2 of 5 such cubes in a radio set have to be replaced within the 1st 150 h is $\frac{80}{243}$.

Outline

Mean of continuous random variable

In lesson 5, we have defined the mean of a discrete random variable X by

$$E[X] = \sum_{x} xP(X = x).$$

Definition

If X is a continuous random variable having a probability density function $f(x)$, then the **mean***, or the* **expectation***, of X, denoted by $E(X)$, is defined by*

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

Remark. $E[X]$ may not exist (the random variable given in Example 4).

Example 5 (continued from Example 2)

Find the mean of X when the pdf of X is

$$f(x) = \begin{cases} 2x, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Sol : $E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_0^1 2x^2\, dx$

$= 2 \times \frac{1}{3} \left[x^3\right]_0^1$

$= \frac{2}{3}$

Expectation of a function of a random variable

Lesson 5

In ~~week 10~~; we have defined the mean of a function of a discrete random variable X, i.e. for any real-valued function g,

$$E[g(X)] = \sum_{x} g(x)P(X = x).$$

Definition

If X is a continuous random variable having a probability density function $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Example 6 (continued from Example 2)

Find $E[e^X]$ when the pdf of X is

$$f(x) = \begin{cases} 2x, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Sol: $E(e^X) = \int_{-\infty}^{+\infty} e^x f(x)\,dx = 2\int_0^1 x e^x\,dx$

$= 2[xe^x]_0^1 - 2\int_0^1 e^x\,dx = 2e - 2[e^x]_0^1$

$= \cancel{2e} - \cancel{2}e + 2$

$= 2$

IBP: $u = x$, $v' = e^x$

$u' = 1$, $v = e^x$

IBP: Integration by Parts

Properties of mean

Let $X_1, X_2, \ldots, X_n$ be random variables. The mean of a sum of random variables equals the sum of the mean of each random variable, i.e.

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

Moreover, let $a_1, a_2, \ldots, a_n$ be constant. Then

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n).$$

Example 7

Sol: Let X_i count whether the cap stops at the i-th station.

So $X_i \sim \text{Bernoulli}(p)$, with

$$p = \mathbb{P}(\{\text{at least 1 passenger get off at station } i\}) = 1 - \left(\frac{4}{5}\right)^4$$

i.e. $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1 - p$. And $\mathbb{E}(X_i) = p$, for $i = 1, \cdots, 5$.

A cab drives 4 passengers to their destinations. There are 5 potential cab stops during the route. If there is no one saying that he/she will get off when the cab is approaching one stop, the driver will continue driving without stop. Assume that for each passenger, it is equally likely that he/she will get off at any of the 5 cab stops. In average, how many times will the cab take stops?

Then the number of times in average that the cab takes stops is

$$\mathbb{E}(X_1 + \cdots + X_5) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_5) = 5\,\mathbb{E}(X_1) = 5p = 5 \times \left[1 - \left(\frac{4}{5}\right)^4\right]$$
$$= \frac{369}{125} \approx 2.952$$

Definition of variance

Definition

The variance of a random variable X, denoted by $Var(X)$, is defined as the mean of the function of X with $g(X) = (X - E[X])^2$, i.e.

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx.$$

In the practice, it is more convenient to compute the variance via the following equivalent formula

$$Var(X) = E[X^2] - (E[X])^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - (E[X])^2.$$

- The variance is always nonnegative.
- The square root of $Var(X)$, i.e. $\sqrt{Var(X)}$, is called the *standard deviation* of X.

Example 8 (continued from Example 2)

Find the variance and standard deviation of X when the pdf of X is

$$f(x) = \begin{cases} 2x, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Sol: $\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx = 2\int_0^1 x^3\, dx = \frac{2}{4}\left[x^4\right]_0^1 = \frac{1}{2}$

So $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2$, by Example 5

$= \frac{1}{18}$

And then stand deviation of $X = \sqrt{\text{Var}(X)} = \frac{1}{3}\frac{\sqrt{2}}{2} = \frac{\sqrt{2}}{6}$

Property of variance

Let X be a random variable, a, b, and c be constants. Consider $Y = aX + b$ as a linear function of X. Then

- $Var(c) = 0$.
-
$$Var(Y) = Var(aX + b) = a^2 Var(X).$$

Outline

Uniform random variable

A random variable X is said to be *uniformly* distributed over the interval (a, b) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

The cdf of X is

$$F(x) = \begin{cases} 0 & x \leq a, \\ \frac{x-a}{b-a} & a < x < b, \\ 1 & x \geq b. \end{cases}$$

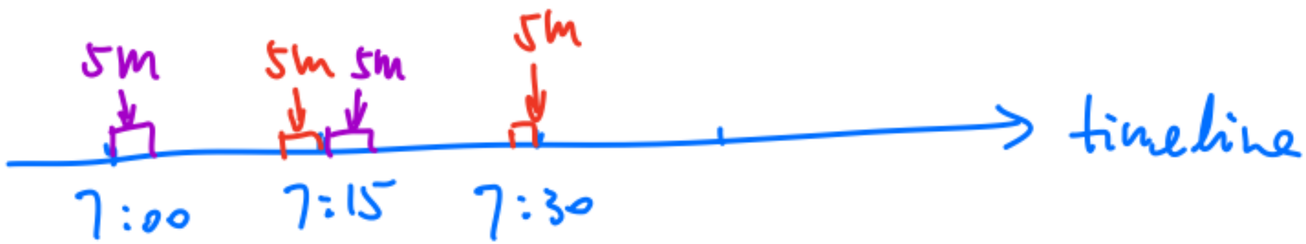
The mean and variance of X are

$$E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}.$$

Example 9

Buses arrive at a specified stop at 15-minute intervals starting at 7:00 am. That is, they arrive at 7:00, 7:15, 7:30, 7:45 and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7:00 and 7:30, find the probability that he waits

1. less than 5 minutes for a bus;
2. more than 10 minutes for a bus.



Sol: Suppose X be the number of minutes past 7 that the passenger arrivals at the stop. Then $X \sim \text{Uniform}[0, 30]$.

Since the buses arrive every 15 m starting at 7:00 am,

{passenger will wait less than 5 m} = {he arrives during 7:10 to 7:15} ∪ {he arrives during 7:25 to 7:30}.

(a) $\mathbb{P}(\{\text{passenger will wait less than 5 m}\})$

$$= \mathbb{P}(10 < X < 15) + \mathbb{P}(25 < X < 30)$$

$$= F(15) - F(10) + F(30) - F(25)$$

$$= \frac{15}{30} - \frac{10}{30} + \frac{30}{30} - \frac{25}{30}$$

$$= \frac{1}{3}$$

(b) {passenger will wait more than 10 m}

= {passenger arrives at stop during 7:00 to 7:05} ∪ {―――――――― 7:15 to 7:20}

So $\mathbb{P}(\{\text{passenger will wait more than 10 m}\})$

$$= \mathbb{P}(0 < X < 5) + \mathbb{P}(15 < X < 20)$$

$$= F(5) - F(0) + F(20) - F(15) = \frac{5}{30} - 0 + \frac{20}{30} - \frac{15}{30} = \frac{1}{3}.$$

Example 10

A stick of length 1 is split randomly at a point X that is uniformly distributed over $(0, 1)$. Suppose a point p on the stick is located p unit of distance away from the stick's left endpoint and $0 \leq p \leq 1$. Determine the expected length of the piece with the stick's left endpoint that contains the point p.

Sol : It is known $X \sim \text{Uniform}(0,1)$.

Let Y be the length of the piece that contains point p.

Consider the indecator function $\mathbb{1}_A$ for an event A as follows:

$$\mathbb{1}_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs}. \end{cases}$$

Then let $A = \{X > p\}$, we have

$$Y = X \, \mathbb{1}_{\{X > p\}} .$$

Therefore,

$$\mathbb{E}(Y) = \mathbb{E}(X \, \mathbb{1}_{\{X>p\}})$$

$$= \int_p^{+\infty} x f(x) \, dx$$

$$= \int_p^1 x \cdot 1 \, dx \quad , \quad \text{since } X \sim \text{Unif}(0,1)$$

$$f(x) = \begin{cases} 1, & \text{if } x \in (0,1) \\ 0, & \text{otherwise} \end{cases}$$

$$= \frac{1}{2} \left[x^2\right]_p^1$$

$$= \frac{1}{2}(1 - p^2)$$

<u>In conclusion</u>, the expected length of piece that contains the point p is $\frac{1}{2}(1-p^2)$.

Exercise

Let X be the result of an experiment, and X is uniformly distributed over $(2, 5)$. The experiment has been carried out 3 times independently, what is the probability that it happens at least twice that the result is larger than 3?

<u>Sol</u>: Let Y be the number of times that the result is larger than 3. Since $X \sim \text{Uniform}(2,5)$, we have $Y \sim b(3,p)$, with

$$
\begin{aligned}
p = \mathbb{P}(X > 3) &= 1 - \mathbb{P}(X \leq 3) \\
&= 1 - F(3) \\
&= 1 - \frac{3-2}{5-2} \\
&= \frac{2}{3}
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathbb{P}(Y \geq 2) &= 1 - \mathbb{P}(Y < 2) \\
&= 1 - \mathbb{P}(Y \leq 1) \\
&= 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) \\
&= 1 - \binom{3}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^3 - \binom{3}{1}\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^2 \\
&= 1 - \overset{1}{\cancel{3}} \times 1 \times \frac{1}{27} - 3 \times \frac{2}{3} \times \frac{1}{9} \\
&= 1 - \frac{7}{27} \\
&= \frac{20}{27}
\end{aligned}
$$

<u>In conclusion</u>, the probability is $\frac{20}{27}$ that at least twice the result is larger than 3.

Example 11

A bakery sells rolls in units of a dozen. The demand X (in 1000 units) for rolls has a uniform distribution on $[2, 4]$. It costs \$2 to make a unit that sells for \$5 on the first day when the rolls are fresh. Any leftover units are sold on the second day for \$1. How many units should be made to maximize the expected value of the profit?

Sol : Demand $X \sim \text{Uniform}(2,4)$.

then X has density $f(t) = \begin{cases} \frac{1}{2}, & \text{if } 2 < t < 4 \\ 0, & \text{otherwise} \end{cases}$

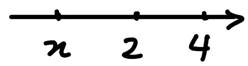
Let x be the units of rolls made. Then we have the profit

$$Y = \begin{cases} (5-2)X + (1-2)(x-X), & \text{if } X < x \\ (5-2)x, & \text{if } X \geq x \end{cases}$$

$$= \begin{cases} 4X - x, & \text{if } X < x \\ 3x, & \text{if } X \geq x \end{cases}$$

$$\mathbb{E}(Y) = \int_{-\infty}^{x} (4t - x) f(t)\, dt + \int_{x}^{+\infty} 3x f(t)\, dt$$

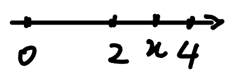
- When $x \leq 2$,



$$\int_{-\infty}^{x} (4t - x) f(t)\, dt + \int_{x}^{+\infty} 3x f(t)\, dt$$

$$= 0 + \int_{2}^{4} \frac{3x}{2}\, dt = \frac{3}{2} x (4-2) = 3x$$, it attains the maximum 6($), when $x = 2$ (dozens) 1000.

- When $2 < x < 4$,



$$\int_{-\infty}^{x} (4t - x) f(t)\, dt + \int_{x}^{+\infty} 3x f(t)\, dt$$

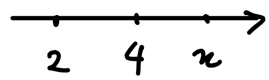
$$= \int_{2}^{x} \frac{4t - x}{2}\, dt + \int_{x}^{4} \frac{3x}{2}\, dt$$

$$= 2 \times \frac{1}{2} \left[t^2 \right]_2^x - \frac{x}{2}(x-2) + \frac{3}{2} x (4 - x)$$

$$= -x^2 + 7x - 4 = -\left(x - \frac{7}{2}\right)^2 + \frac{33}{4}.$$

it attains the maximum $\frac{33}{4}$ ($), when $x = \frac{7}{2}$ (dozens) 1000.

· When $x \geqslant 4$



$$\int_{-\infty}^{x} (4t-x) f(t)\,dt + \int_{x}^{+\infty} 3x f(t)\,dt$$

$$= \int_{2}^{4} \frac{4t-x}{2}\,dt + 0$$

$$= 2\times\frac{1}{2}\left[t^2\right]_2^4 - \frac{x}{2}(4-2)$$

$$= 12 - x.$$

it attaining the maximum 8 (\$), when $x = 4$ (1000 dozens).

In comparison the 3 cases above, we have the maximum profits maybe gained are respectively 6 (\$) < 8 (\$) $< \frac{33}{4}$ (\$) when $x = 2, 4, \frac{7}{2}$ (1000 dozens).

In conclusion, ~~$\frac{7}{2}$ units~~ 3500 dozens should be made to maximize the expected value of profit.