

1. Univariate Continuous Random Variables

Week 1
January 4-8

1. What is 'random'?

There is no rigorous answer.

Convention: X denotes random variable (RV) and x denotes a value which X may take. We say that X is random if it may take value at random (notice, this is not a good definition because 'at random' is not defined precisely). In other words, we say X is random if there is no algorithm by which we can predict values of X . We say that x is a *realization* of X .

Example. X is household income in the US, and x is the income of your family.

Please give examples of other random variables and their realizations.

2. Distribution and density functions

Definition 2.1. *Distribution function of X is defined as*

$$F(x) = \Pr(X \leq x).$$

Sometimes $F(x)$ is called cumulative distribution function (cdf). Regarding RV, we may speak of *support* of X , as the set of all possible values X may take. We will study *continuous* RV with support = interval (which may be the entire real line, or all positive or nonnegative numbers). Now we can define continuous RV X such that $\Pr(a < X < b) > 0$ for all $a < b$ from the support set (interval). Continuous RV has a continuous distribution function (prove please).

Example (continued). The support of X is all nonnegative numbers (in fact, X is bounded but we may not go into detail because that bound is hard to establish). We may ask what is the probability that the income of an American family taken at random is less than \$50,000. This is

$$\Pr(X \leq 50,000) = F(50,000).$$

(Is X continuous indeed?)

Often, the continuity assumption is accepted for convenience.

Definition 2.2. *Density function of X is*

$$f(x) = \frac{dF(x)}{dx}. \tag{2.1}$$

Why is it called "density"? Because

$$\begin{aligned} f(x) &= \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\Pr(x < X \leq x + \Delta x)}{\Delta x}. \end{aligned} \quad (2.2)$$

The quantity under the limit shows how *dense* is X around x .

Distribution function presentation:

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

as follows from (2.1), Main Theorem of Calculus.

For continuous RV the density function is positive on its support (Why?).

3. Properties of distribution and density functions

Properties of the distribution function:

1. $\Pr(a \leq X \leq b) = F(b) - F(a)$ for any $a \leq b$.
2. $F(x)$ is a non-decreasing function, i.e. if $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$.
3. $F(+\infty) = 1$, i.e. $\lim_{x \rightarrow \infty} F(x) = 1$.
4. $F(-\infty) = 0$, i.e. $\lim_{x \rightarrow -\infty} F(x) = 0$.

Proof. Let's prove the first statement

$$\begin{aligned} \Pr(X \leq b) &= \Pr(a \leq X \text{ OR } a \leq X \leq b) \\ &= F(a) + \Pr(a \leq X \leq b), \end{aligned}$$

which implies $\Pr(a \leq X \leq b) = F(b) - F(a)$.

The second statement follows from the first one.

The third statement follows from the fact that $X < \infty$, and the fourth statement follows from the fact that $X > -\infty$.

We say that a function is a distribution function if it satisfies #2, #3, and #4.

Problem. Let F and G be two distribution functions. Is a linear combination, $\alpha F + (1 - \alpha)G$ a distribution function for any $0 \leq \alpha \leq 1$?

Solution. Property #2. Denote $H(x) = \alpha F(x) + (1 - \alpha)G(x)$. Let $x_1 \leq x_2$. We have

$$\begin{aligned} H(x_1) &= \alpha F(x_1) + (1 - \alpha)G(x_1) \\ &\leq \alpha F(x_2) + (1 - \alpha)G(x_2) = H(x_2). \end{aligned}$$

Property #3.

$$\begin{aligned} &\lim_{x \rightarrow \infty} (\alpha F(x) + (1 - \alpha)G(x)) \\ &= \alpha \lim_{x \rightarrow \infty} F(x) + (1 - \alpha) \lim_{x \rightarrow \infty} G(x) \\ &= \alpha 1 + (1 - \alpha)1 = 1 \end{aligned}$$

Property #4:

$$\begin{aligned} & \lim_{x \rightarrow -\infty} (\alpha F(x) + (1 - \alpha)G(x)) \\ &= \alpha \lim_{x \rightarrow -\infty} F(x) + (1 - \alpha) \lim_{x \rightarrow -\infty} G(x) \\ &= \alpha 0 + (1 - \alpha)0 = 0 \end{aligned}$$

Is the condition $0 \leq \alpha \leq 1$ crucial?

Problem. Prove that the function

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

is a distribution function (this is a special case of *exponential* distribution, to be considered later).

Solution. Check #2: if $x_1 \leq x_2$ then $-x_1 \geq -x_2$ and

$$e^{-x_1} \geq e^{-x_2}$$

and

$$F(x_1) = 1 - e^{-x_1} \leq 1 - e^{-x_2} = F(x_2).$$

Check #4: $F(-\infty) = 0$, check #3:

$$\lim_{x \rightarrow \infty} F(x) = 1 - e^{-\infty} = 1 - 0 = 1$$

Please, give examples of a function which is not a distribution function.

Properties of the density function:

1. Density is nonnegative, $f(x) \geq 0$.
2. The area under density is 1, i.e.

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

- 3.

$$\Pr(a \leq X \leq b) = \int_a^b f(x) dx$$

Proof. $f(x) \geq 0$ follows from the fact that $F(x)$ is increasing. The second and the third statements follow from the properties of the distribution function: $F(\infty) = 1$ and $F(-\infty) = 0$.

We say that a function is a density function if it satisfies #1 and #2.

Problem. Let f and g be two density functions. Is a linear combination, $\alpha f + (1 - \alpha)g$ a density function for any $0 \leq \alpha \leq 1$?

Solution. Denote $h(x) = \alpha f(x) + (1 - \alpha)g(x)$. Check property #1:

$$\begin{aligned} h(x) &= \alpha f(x) + (1 - \alpha)g(x) \\ &\geq \alpha 0 + (1 - \alpha)0 = 0 \end{aligned}$$

Check property #2:

$$\begin{aligned}
 & \int_{-\infty}^{\infty} h(x) dx \\
 &= \alpha \int_{-\infty}^{\infty} f(x) dx + (1 - \alpha) \int_{-\infty}^{\infty} g(x) dx \\
 &= \alpha 1 + (1 - \alpha) 1 = 1.
 \end{aligned}$$

Is $0 \leq \alpha \leq 1$ crucial?

Problem. Find a such that

$$f(x) = \frac{a}{1 + x^2}$$

is a density function (this is the density of *Cauchy* distribution). Find the distribution function.

Solution. Check #1: $f(x) > 0$ for all x . To find a we have to calculate

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \frac{1}{1 + x^2} dx \\
 &= \left(\text{Arc tan}(x) \Big|_{-\infty}^{\infty} \right) \\
 &= \left(\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right) \\
 &= \left(\frac{\pi}{2} + \frac{\pi}{2} \right) \\
 &= \pi.
 \end{aligned}$$

That means

$$a = \frac{1}{\pi}.$$

The distribution function is

$$\begin{aligned}
 F(x) &= \frac{1}{\pi} \int_{-\infty}^x \frac{1}{1 + t^2} dt \\
 &= \frac{1}{\pi} \left(\text{Arc tan}(x) + \frac{\pi}{2} \right) \\
 &= \frac{1}{\pi} \text{Arc tan}(x) + \frac{1}{2}.
 \end{aligned}$$

Mode is where the density attains its maximum. We call density $f(x)$ *unimodal* if it has one mode, i.e. $f(x)$ is increasing at left and decreasing at right of the mode. Otherwise, we call the density *multimodal*. Let m denote the mode, we call X *symmetric* if

$$\Pr(X - m \leq -a) = \Pr(X - m \geq a).$$

For a symmetric distribution: $f(m - x) = f(x + m)$ and $F(m - x) = 1 - F(x + m)$. Illustrate geometrically.

Prove, that the Cauchy distribution is symmetric, and therefore unimodal with mode=0.

Exponential distribution is not symmetric.

4. Quantiles, quartiles, and median

Definition 4.1. The p th quantile ($0 < p < 1$), x_p is such that

$$F(x_p) = p.$$

The $\frac{1}{4}$ th quantile is called the lower quartile, the $\frac{3}{4}$ th quantile is called the upper quartile. Median is the $\frac{1}{2}$ th quantile, i.e.

$$F(\text{median}) = \frac{1}{2}.$$

Percentile is a quantile expressed in per cents (e.g. 75 percentile, 25 percentile, etc.).

Quantiles and quartiles are used to characterize the range of the distribution.

Problem. Find the p th quantile of the exponential distribution defined by the distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{if } x \geq 0 \end{cases}.$$

Also, find median, the lower and the upper quartile. Find a and b such that $\Pr(a < X < b) = 0.5$

Solution. We need to solve equation

$$1 - e^{-x} = p$$

which yields the p th quantile

$$x_p = -\ln(1 - p).$$

The median is

$$x_{.5} = -\ln(1 - .5) = 0.69315.$$

The lower quartile is

$$x_{.25} = -\ln(1 - .25) = .28768.$$

The upper quartile is

$$x_{.75} = -\ln(1 - .75) = 1.3863.$$

Since $\Pr(x_{.25} < X) = .25$ and $\Pr(x_{.75} > X) = .25$ we have

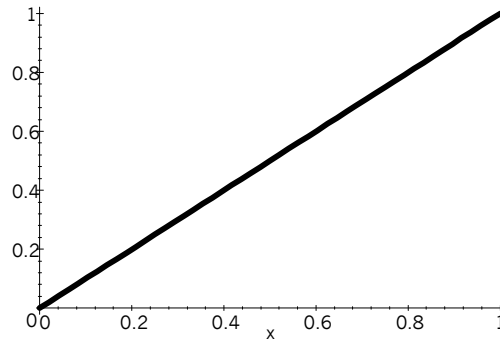
$$\Pr(x_{.25} < X < x_{.75}) = .5$$

that means $a = x_{.25}$ and $b = x_{.75}$.

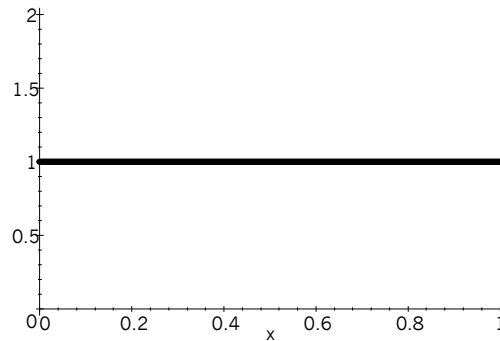
5. Uniform distribution

The simplest continuous random variable is

Uniform random variable X which takes values on the interval $[a, b]$ with 'equal probability'. If X is a uniform random variable on $[a, b]$ we may interpret it as an outcome "choose a number at random between a and b ". The latter means that the probability to fall within the interval $[c, d]$ where $a \leq c \leq d \leq b$ is proportional to the length of the interval $(d - c)$ and does not depend on the specific location of that interval.



Distribution function of a uniform random variable, $U(0, 1)$



Density (constant) of the uniform distribution $U(0, 1)$.

Sometimes the uniform distribution is referred to as 'rectangular' and denoted $R(a, b)$. The density of the uniform distribution is

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b \end{cases}$$

and the distribution function is

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases} \quad (5.1)$$

To show that X has the uniform distribution on (a, b) we use 'tilde', and write $X \sim U(a, b)$.

Problem. Find the lower quartile for $X \sim U(a, b)$.

Solution. We solve

$$F(x) = \frac{x-a}{b-a} = \frac{1}{4}$$

with the solution

$$x = a + \frac{b-a}{4} = \frac{3a+b}{4}$$

$$\text{The lower quartile} = \frac{3a+b}{4}.$$

Median:

$$\frac{x - a}{b - a} = \frac{1}{2}$$

with the solution

$$x = a + \frac{b - a}{2} = \frac{a + b}{2}$$
$$\text{median} = \frac{a + b}{2}.$$

For uniform RV

$$\Pr(c < X < d) = \frac{d - c}{b - a}$$

where $a < c < d < b$.

Problem. *Robber and police.*

The robber knows that the alarm starts as soon he gets into the bank. The robber also knows that he needs 2 minutes to finish off his business. Police comes within 5 minutes after alarm starts. Assuming that the distribution of the police arrival is uniform what is the probability that the robber will not be caught?

Solution. Let X denotes the RV – time of police arrival after alarm starts. We know that $X \sim U(0, 5)$. The probability that the robber will be caught is

$$\Pr(X \leq 2) = F(2)$$
$$= \frac{x - 0}{5 - 0} = \frac{2}{5},$$

since $x = 2$. The probability that the robber will not be caught is

$$\Pr(X > 2)$$
$$= 1 - \Pr(X \leq 2)$$
$$= 1 - \frac{2}{5}$$
$$= \frac{3}{5}.$$

6. Homework (due Wednesday, January 13)

Maximum number of points is 38.

1. (5 points). *Logistic* distribution is defined via the distribution function as

$$F(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

where α and β are any numbers (parameters) and β is positive.

- 1.1. Show that $F(x)$ is, in fact, a distribution function.
- 1.2. Find the density and show that X is symmetric.
- 1.3. Sketch the graph of the distribution and the density functions.
- 1.4. Find the median and derive the formula for the p th quantile.
- 1.5. Calculate the probability $\Pr(\beta X + 1 > 0)$ given $\alpha = 0.5$.

Solution.

1.1. Check whether $F(x)$ is a nondecreasing function. Let $x_1 \leq x_2$. Then, since $\beta > 0$ we have $\alpha + \beta x_1 \leq \alpha + \beta x_2$, and therefore $\exp(\alpha + \beta x_1) \leq \exp(\alpha + \beta x_2)$. This implies $F(x_1) \leq F(x_2)$. Check $F(\infty) = 1$: Rewrite $F(x) = 1/(1 + \exp(-\alpha - \beta x))$. Since $\beta > 0$ we have $-\alpha - \beta x \rightarrow -\infty$ when $x \rightarrow \infty$ which implies $\exp(-\alpha - \beta x) \rightarrow 0$. This implies $F(\infty) = 1$. Check $F(-\infty) = 0$. When $x \rightarrow -\infty$ we have $\alpha + \beta x \rightarrow -\infty$ and $\exp(\alpha + \beta x) \rightarrow 0$. This implies $F(-\infty) = 0$.

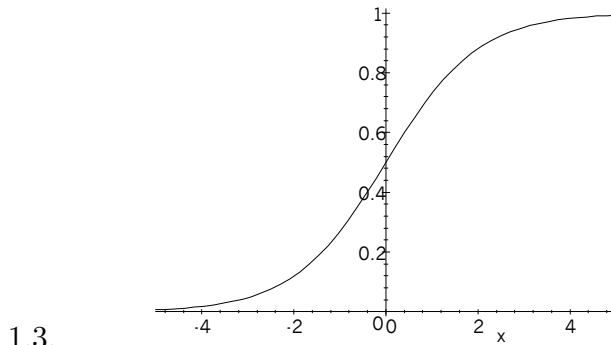
1.2. Take derivative to obtain the density

$$f(x) = \beta \frac{e^{\alpha + \beta x}}{(1 + e^{\alpha + \beta x})^2}.$$

Find the mode, $f(x) = \max$. Take derivative of $f(x)$ and equate it to zero which leads to $\exp(\alpha + \beta x) = 1$. The mode is $m = -\alpha/\beta$. In order to show that $f(x)$ is symmetric we must show that $f(m - x) = f(x + m)$. We have $\exp(\alpha + \beta(-\alpha/\beta - x)) = \exp(-\beta x)$, i.e.

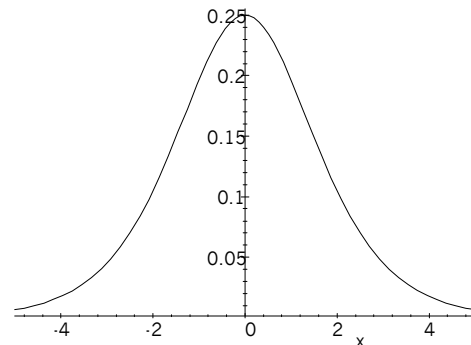
$$f(x - m) = \beta \frac{e^{-\beta x}}{(1 + e^{-\beta x})^2} = \beta \frac{e^{\beta x}}{(1 + e^{\beta x})^2}.$$

Similarly for $f(x + \beta)$, we have $\exp(\alpha + \beta(x - \alpha/\beta)) = e^{\beta x}$. So that $f(x - m) = f(x + m)$.



1.3.

Typical distribution function



Typical density function

1.4. Let p be within $(0, 1)$. The p th quantile is the solution to the equation $F(x) = p$ that leads to

$$x_p = \frac{\ln(\frac{p}{1-p}) - \alpha}{\beta}.$$

The median corresponds to $p = 0.5$, i.e. median for the logistic distribution is $-\alpha/\beta$. We can come to the same conclusion noticing that $f(x)$ is symmetric around the mode.

1.5. We have

$$\begin{aligned} \Pr(\beta X + 1 > 0) &= \Pr(X > -1/\beta) = 1 - \Pr(X \leq -1/\beta) = 1 - F(-1/\beta) \\ &= 1 - \frac{e^{\alpha-1}}{1 + e^{\alpha-1}} = \frac{1}{1 + e^{\alpha-1}} = \frac{1}{1 + e^{-.5}} = .62246 \end{aligned}$$

The answer is .62246.

2. (4 points). X is a CRV with the support set (a, b) . The density function has a triangle shape:

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ h(x - a) & \text{if } a \leq x \leq (a + b)/2 \\ h(b - x) & \text{if } (a + b)/2 < x \leq b \\ 0 & \text{if } x > b \end{cases}$$

Find h . Find the distribution function and sketch the graph. Find q and g such that $\Pr(q < X < g) = 0.5$.

Solution. We find h from the condition $\int f(x)dx = 1$. Since $f(x)$ has a triangle form on (a, b) the area under f is the area of the triangle with the base $(b - a)$. The height of the triangle is $h((a+b)/2 - a) = h(b-a)/2$. Hence, the area must be $h(b-a)^2/4 = 1$ and we obtain $h = 4/(b-a)^2$. The distribution function has the form

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{2(x-a)^2}{(b-a)^2} & \text{if } a \leq x \leq \frac{a+b}{2} \\ 1 - \frac{2(b-x)^2}{(b-a)^2} & \text{if } \frac{a+b}{2} < x \leq b \\ 1 & \text{if } b < x \end{cases}.$$

The graph consists of a coupled continuously attached piecewise parabolas. For the lower and the upper quantiles $x_{.25} = q$ and $x_{.75} = g$ we have $\Pr(q < X < g) = .5$. The lower quartile is satisfies the equation $2(x-a) = \frac{1}{4}(b-a)^2$, i.e. $q = x_{.25} = a + \frac{1}{8}(b-a)^2$ and the upper quartile is $g = x_{.75} = b - \frac{1}{8}(b-a)^2$.

3. (5 points). Let $F(x) = 1 - \exp(-\alpha x^\beta)$ for $x \geq 0$ (and $F(x) = 0$ for $x < 0$) where α and β are positive parameters. Show that $F(x)$ is a distribution function and find the density. Find the p th quantile. Find a and b such that $\Pr(a < X < b) = 0.8$. Are a and b unique?

Solution. If $x_1 \leq x_2$ then $\alpha x_1^\beta \leq \alpha x_2^\beta$ since α and β are positive. It implies that $\exp(-\alpha x_1^\beta) \geq \exp(-\alpha x_2^\beta)$ and $F(x_1) = 1 - \exp(-\alpha x_1^\beta) \leq 1 - \exp(-\alpha x_2^\beta) = F(x_2)$, the condition on monotonicity is proved. Now we show that $F(\infty) = 1$. Indeed, when $x \rightarrow \infty$ then $-\alpha x^\beta \rightarrow -\infty$ and $F(x) = 1 - \exp(-\alpha x^\beta) \rightarrow 1$. At last we show that $F(-\infty) = 0$. Indeed, then $-\alpha x^\beta \rightarrow 0$ and $F(x) = 1 - \exp(-\alpha x^\beta) \rightarrow 1 - 1 = 0$.

The p th quantile is the solution to the equation $F(x) = 1 - \exp(-\alpha x^\beta) = p$, that means $-\alpha x^\beta = \log(1-p)$ and $x_p = (-\alpha^{-1} \log(1-p))^{1/\beta}$. To find a and b we define quantiles $a = x_{.1}$ and $b = x_{.9}$. Then $\Pr(x_{.1} < X < x_{.9}) = .8$. The values of a and b are not unique because we could take $a = x_{.05}$ and $b = x_{.85}$.

4. (3 points). A line segment of length a is cut once at random. What is the probability that the left piece is more than twice the length of the right piece?

Solution. Let us denote X the length of the left piece. We know that $X \sim U(0, a)$. The length of the right piece is $a - X$. The needed probability is $\Pr(2X > a - X) = \Pr(X > a/3) = 1 - \Pr(X \leq a/3) = 1 - (a/3)/a = 2/3$. The answer is $2/3$.

5. (7 points). If F and G are distribution functions, show that:

5.1. FG is a distribution function.

5.2. F^m is a distribution function where $m > 0$.

5.3. $F^m G^n$ is a distribution function where $m > 0$ and $n > 0$.

5.4. If f and g are density functions is fg a density function? Give counterexample, if not.

Solution.

5.1. The monotonicity follows from the fact that if $0 \leq F(x_1) \leq F(x_2)$ and $0 \leq G(x_1) \leq G(x_2)$ then $F(x_1)G(x_1) \leq F(x_2)G(x_2)$. Also as we see, $\lim_{x \rightarrow \infty} F(x)G(x) = 1$ if $\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} G(x) = 1$. Similarly $\lim_{x \rightarrow -\infty} F(x)G(x) = 0$.

5.2. $0 \leq F(x_1) \leq F(x_2)$ implies $0 \leq F^m(x_1) \leq F^m(x_2)$ for a positive m . And $\lim_{x \rightarrow \infty} F^m(x) = 1$ if $\lim_{x \rightarrow \infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F^m(x) = 0$ if $\lim_{x \rightarrow -\infty} F(x) = 0$.

5.3. follows from 5.1 and 5.2.

5.4. No. Counterexample: let $X \sim U(0, 1)$ and $Y \sim U(2, 3)$ then the product of the densities is zero which is not a density.

6. (4 points). We say that RV X is greater than Y if $\Pr(X > c) > \Pr(Y > c)$ for any real c . Show that in this case $F_Y(x) > F_X(x)$ for any x . Why a similar density-based definition is wrong: X is greater than Y if $f_X(x) < f_Y(x)$ for any x where f is the density.

Solution. We observe that $\Pr(X > c) = 1 - F_X(c)$ where F_X is the distribution function. Then the condition $\Pr(X > c) > \Pr(Y > c)$ is rewritten as $1 - F_X(c) > 1 - F_Y(c)$ that implies $F_X(c) < F_Y(c)$.

It cannot be $f_X(x) < f_Y(x)$ for all x because then $1 = \int f_X(x) < \int f_Y(x)$ which implies $\int f_Y(x) > 1$.

7. (6 points). The Dartmouth campus daily water consumption (in thousands of liters) is a random variable whose probability density is given by

$$f(x) = \begin{cases} \frac{1}{9}xe^{-\frac{x}{9}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}.$$

Assume the daily water capacity in Dartmouth campus is 9 thousand liters. What is the probability that at a given day Dartmouth runs out of water? What is the probability to run out of water at least one day during the year (365 days/year)?

Solution. If X is the RV water consumption the probability that Dartmouth runs out of water is $\Pr(X > 9) = 1 - \Pr(X \leq 9) = 1 - \int_0^9 f(x)dx = \frac{1}{9} \int_0^9 xe^{-\frac{x}{9}}dx$. We use integration by part ($\int u dv = uv - \int v du$) to find the integral

$$\int_0^9 xe^{-\frac{x}{9}}dx = 9 - 36e^{-3} = 7.2077$$

and the answer is $\Pr(X > 9) = 1 - 7.2077/9 = .8$.

The probability to run out of water at least one day during the whole year is $1 - \Pr^{365}(X < 9) = 1 - (1 - \Pr(X > 9))^{365}$ because $\Pr(X < 9)$ is the probability to have water consumption less than 9 at **one** day (here we assume that the daily water consumption is independent). The answer is

$$1 - .2^{365} = 1 - 7.5153 \times 10^{-256} = 1.$$

We conclude that with probability 1 Dartmouth runs out of water at least one day during the year.

8. (4 points). A segment of length 0.6 is dropped at random on real line. What is the probability to cover at least one integer number?

Solution. Let X be the coordinate of the left end and $\lfloor X \rfloor$ be the maximum integer less than X . Denote $Y = X - \lfloor X \rfloor$. Since the segment is dropped at random $Y \sim U(0, 1)$. The piece will not cover an integer if $Y < .4$. The probability of the latter event is 0.4. Hence, the probability to cover an integer is $1 - 0.4 = 0.6$.