**OCEA90 Activity: Interpreting climate histograms and probability density functions**

As we discussed in lectures, measurements of the atmosphere and ocean can be considered to be random variables, and hence can be described by a probability density function. In this activity we will solidify the introductory statistical concepts that were discussed in class using different examples.

1. In your own words, briefly explain how you would make a histogram for a random variable, such as measurements of temperature in the atmosphere. Hint: describe the four steps for making a histogram as introduced in week 2 lecture videos.

According to the week 2 lecture:

1. Step 1: We first found out the range first, which also means we need to find out the maximum and minimum value of temperature in the graph. The range is between maximum and minimum value.
2. Step 2: Divide the range in several bins, just like the example on pdf. We could set up each bin between levels of temperature.
3. Step 3: We could now start counting the occurrences of temperature in each bin.
4. Step 4: Add all the numbers in each bin and plot the number of occurrences (counts) within each bin against the corresponding bin values.

2. The table below presents the min January daily temperature in Montreal (Canada) in $^oC$.

| Day | Min T ($^oC$) | Day | Min T ($^oC$) |
|---|---|---|---|
| 1 | -11 | | |
| 2 | -15 | 17 | -17 |
| 3 | -14 | 18 | -20 |
| 4 | -3 | 19 | -16 |
| 5 | -1 | 20 | -25 |
| 6 | -12 | 21 | -21 |
| 7 | -16 | 22 | -21 |
| 8 | -10 | 23 | -23 |
| 9 | -3 | 24 | -13 |
| 10 | -13 | 25 | -6 |
| 11 | -18 | 26 | -10 |
| 12 | -18 | 27 | -13 |
| 13 | -19 | 28 | -8 |
| 14 | -19 | 29 | -5 |
| 15 | -18 | 30 | -1 |
| 16 | -8 | 31 | -1 |

a) Build a histogram of the data using bins such that
1: -25 ≤ T <-20
2: -20 ≤ T <-15
3: -15 ≤ T <-10
4: -10 ≤ T <-5
5: -5 ≤ T < 0

What are N, ΔT, M and $n_j$?

N = 31 Days
ΔT = 5°C
M = 5

| Bin | Temperature Range | Number between those domain (nj) |
|---|---|---|
| 1: -25 ≤ T <-20 | -25 ≤ T < -20 | 2 |
| 2: -20 ≤ T <-15 | -20 ≤ T < -15 | 5 |
| 3: -15 ≤ T <-10 | -15 ≤ T < -10 | 4 |
| 4: -10 ≤ T <-5 | -10 ≤ T < -5 | 5 |
| 5: -5 ≤ T < 0 | -5 ≤ T < 0 | 7 |

b) Verify that $N = \sum_{j=1}^{M} n_j = n_1 + n_2 + \cdots + n_M$

In this case: N = $\sum_{j=1}^{5} n_j = n_1 + n_2 + n_3 + n_4 + n_5$ = 2 + 5 + 4 + 5 + 7 = 23

c) Now let's convert the histogram constructed in a) into a probability density function for the same random variable. Calculate the probability density for each bin ($p_j$), where:

$p_j = n_j/(N\Delta T)$

$p_1 = n_1/(N\Delta T)$ = 2 / (31 * 5) = 0.013

$p_2 = n_2/(N\Delta T)$ = 5 / (31 * 5) = 0.032

$p_3 = n_3/(N\Delta T)$ = 4 / (31 * 5) = 0.026

$p_4 = n_4/(N\Delta T)$ = 5 / (31 * 5) = 0.032

$p_5 = n_5/(N\Delta T)$ = 7 / (31*5) = 0.045

d) Based on the probability density calculated in c), what is the probability that the January min daily temperature in Montreal lies between *-25℃ ≤ T < -20℃*, i.e. Prob{*-25℃ ≤ T < -20*}?
And what are Prob{*-20℃ ≤ T < -15*}, Prob{*-15℃ ≤ T < -10*}, Prob{*-10℃ ≤ T < -5*}, Prob{*-5℃ ≤ T < 0*}?

$$\text{Prob}\{-25°C \le T < -20°C\} = p_1$$

$$\text{Prob}\{-20°C \le T < -15°C\} = p_2$$

$$\text{Prob}\{-15°C \le T < -10°C\} = p_3$$

$$\text{Prob}\{-10°C \le T < -5°C\} = p_4$$

$$\text{Prob}\{-5°C \le T < 0°C\} = p_5$$

e) Based on the probabilities calculated in d), which temperature range is the most likely?

$$\text{Prob}\{-25°C \le T < -20°C\} = p_1$$

$$\text{Prob}\{-20°C \le T < -15°C\} = p_2$$

$$\text{Prob}\{-15°C \le T < -10°C\} = p_3$$

$$\text{Prob}\{-10°C \le T < -5°C\} = p_4$$

$$\text{Prob}\{-5°C \le T < 0°C\} = p_5$$

f) Verify that $\sum_{j=1}^{M} p_j \Delta T = 1$

$p_1 = n_1/(N\Delta T)$ = 2 / (31 * 5)

$p_2 = n_2/(N\Delta T)$ = 5 / (31 * 5)

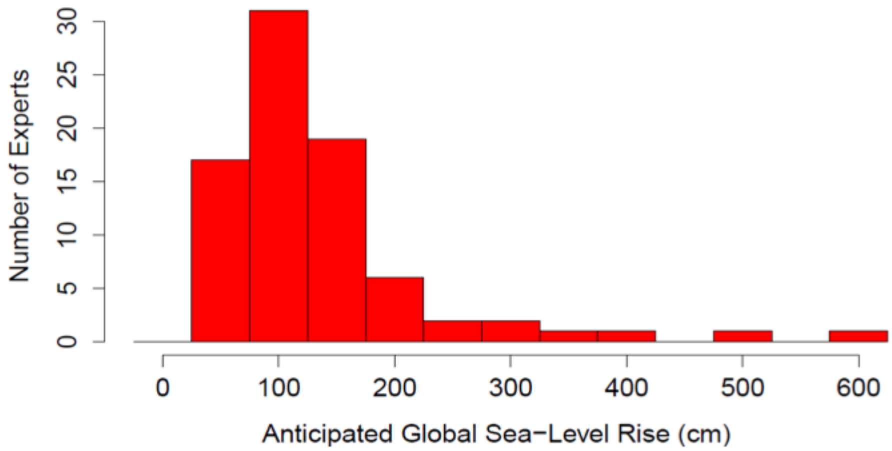
$p_3 = n_3/(N\Delta T)$ = 4 / (31 * 5)

$p_4 = n_4/(N\Delta T)$ = 5 / (31 * 5)

$p_5 = n_5/(N\Delta T)$ = 7 / (31*5)

(P1+ p2+ p3+ p4+ p5 ) * 5 = 1

3. The following histogram presents anticipated global sea level rise by 2100 from N= 90 experts (Norton et al. (2014)). The bins have a width of 50.




a) If 32 experts anticipate a global sea level rise between 75 and 125 cm, what is the probability that it will lie in that interval, i.e. what is Prob{*75cm ≤ SL < 125cm*}?

**Prob{*75cm ≤ SL < 125cm}* = *32 / 90 = 0.356***

b) In 1-2 sentences, explain how you would transform this histogram into a probability density function.

**Transform this histogram into a probability density function: Given the number of experts(N) is 90 and the width of the bin is 50 cm. This is provided by the probability density at each interval, which presents the sea level rise occurring within that specific range.**

c) What is the probability density ($p_j$) for the second bin (j=2) corresponding to the interval of 75-125 cm? Note that $n_2 = 32$.

**Density ($p_j$) = 32/ (90 * 50) = 32 / 4500 = 0.007**

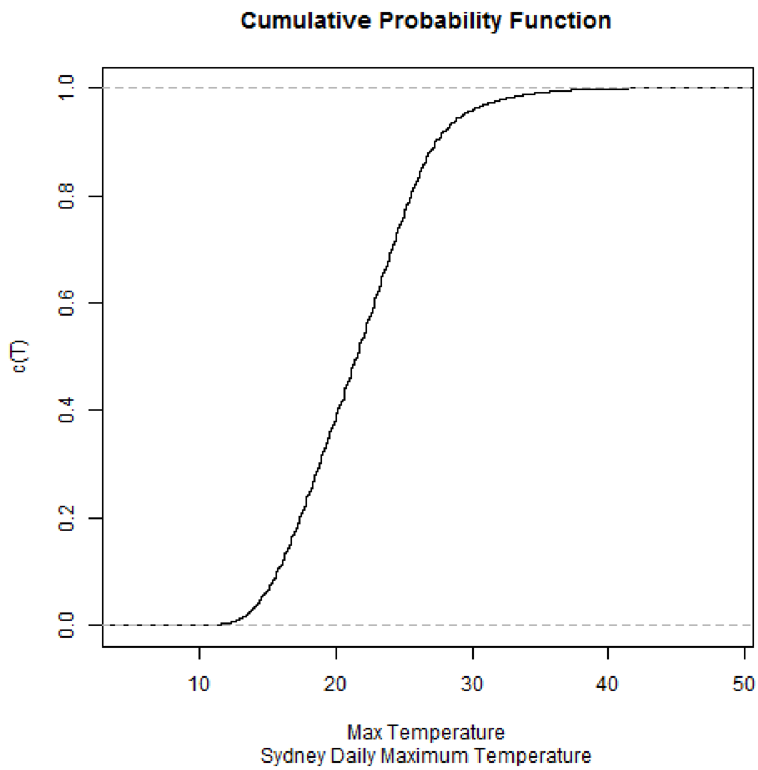
d) Based on the shape of the histogram and the experts' judgment, is it likely that global sea level rise will be > 300cm? > 500cm? Discuss your reasoning here in ~1-3 sentences.

**Based on the shape of the histogram, it seems unlikely that the global sea level rise will exceed 300 cm. Most experts predict that the rise will be, than 300 cm. Furthermore according to the histogram there is no anticipation from any experts for a rise, than 500 cm.**

4. The following figure is showing the CDF of the Sydney daily max temperature discussed earlier this week.

a) In your own words, explain what a cumulative probability density function shows.

The cumulative distribution function(CDF) is a way to show and observe climate variables or values such as the temperature. it provides a visual representation of the probability that a random variable takes on a value less than or equal to a given point.




b) Using the plotted CDF, estimate the median of the distribution and explain clearly how you did this.

The median daily max temp at Sydney is 21.6°C based on lecture notes. The median corresponds to the temperature at which the cumulative probability distribution function (CDF) has a value of 0.5.

c) Using the plotted CDF above, estimate the probability that daily max temperature in Sydney will be less than or equal 30°C? larger than 30°C? between 20°C and 30°C?

The daily max temperature in Sydney will less than or equal to 30°C. The value is approximately 0.9

The daily max temperature in Sydney will large than 30°C. The value is approximately 1

The daily max temperature in Sydney between 20°C and 30°C: Since 30°C is 0.94 and 20°C is 0.4. So 0.94 - 0.4 = 0.54.

5. Based on the average January temperature in Montreal (Canada) presented below, order the temperature values into ascending order and calculate the empirical cumulative probability distribution $F(T_{(j)})$ for each temperature value $T_{(j)}$.

| Year | T (°C) |
|---|---|
| 1 | -7.0 |
| 2 | -8.5 |
| 3 | -8.9 |
| 4 | -8.0 |
| 5 | -7.9 |
| 6 | -9.5 |
| 7 | -8.4 |
| 8 | -7.5 |
| 9 | -6.1 |
| 10 | -7.2 |

Make them in order first:

-9.5, -8.9, -8.5, -8.0, -8.4, -8.0, -7.9, -7.5, -7.2, -7.0, -6.1

For -9.5: F(-9.5) = 1/10 = 0.1

For -8.9: F(-8.9) = 2/10 = 0.2

For -8.5: F(-8.5) = 3/10 = 0.3

For -8.4: F(-8.4) = 4/10 = 0.4

For -8.0: F(-8.0) = 5/10 = 0.5

For -7.9: F(-7.9) = 6/10 = 0.6

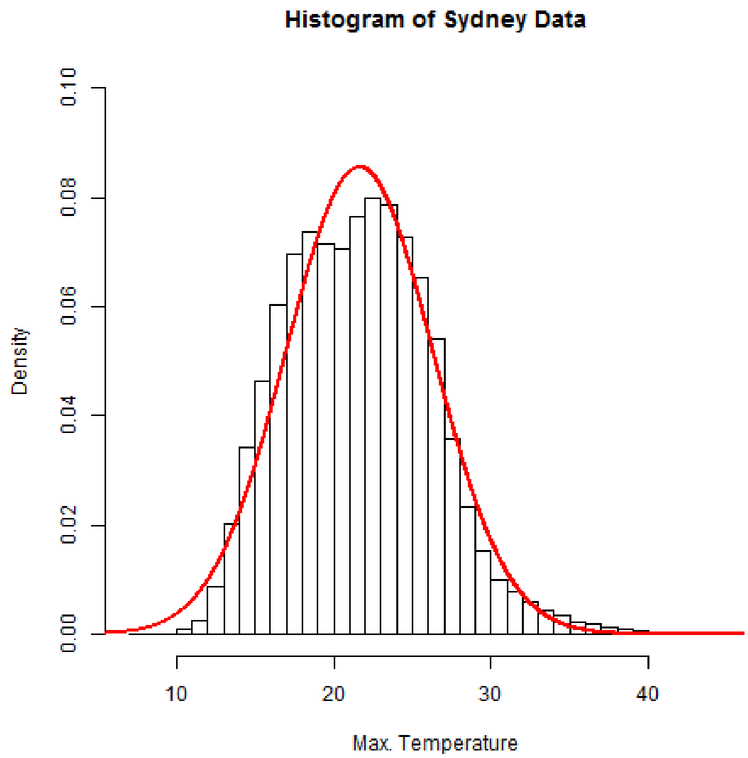
For -7.5: F(-7.5) = 7/10 = 0.7

For -7.2: F(-7.2) = 8/10 = 0.8

For -7.0: F(-7.0) = 9/10 = 0.9

For -6.1: F(-6.1) = 10/10 = 1.0

6. In the figure below, the Sydney daily max temperature PDF is presented with a Normal distribution superposed – red smooth curve.




a) How was that Normal curve calculated? Provide the equation and mention which statistics were calculated from the data to fit the curve.

Normal curve is calculated by the normal distribution, I also do some research online, it also called Normal probability density formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

| Normal probability density formula | Explanation |
|---|---|
| $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | • $f(x)$ = probability<br>• $x$ = value of the variable<br>• μ = mean<br>• σ = standard deviation<br>• $\sigma^2$ = variance |

b) Does the Normal distribution describe the Sydney max temperature well? Justify your answer.
No, I think the Normal distribution describe the sydney max temperature not quiet well. The beginning part and ending part is not measure well.

c) If the PDF of Sydney daily maximum temperature was described by a normal distribution, what could you say about the mean, the mode and the median temperatures?

Based on the lecture silde
The Mean: It represent the average temperature.

The Mode: It corresponds to the peak of the symmetric bell-shaped curve.

The Median:It is located at the center of the distribution, dividing it into two equal halves.