QBIO 401 HW 6- Hosseini

November 1, 2023

1 QBIO 401 Homework 6

1.1 Hirad Hosseini, Fall 2023

The format is the same as we discussed in class: rows are SNPs, columns 0:9 (Python notation, not including 9) are details about the SNPs, columns 9:90 (Python notation, not including 90) are individuals from Europe, and columns 90:179 (Python notation, not including 179) are individuals from Africa. Note: the INFO column is not informative for this assignment, since the file I have given you does not have all the columns and rows from the original data file.

1. It is possible that a SNP is polymorphic in the world-wide sample, but is not polymorphic in a given population sample. Count the number of SNPs in "abbgen1k.csv" for which the ALT allele does NOT have zero frequency in the sample of Europeans individuals. Also count this number for the sample of African individuals. Due to the out-of-Africa hypothesis, we expect that this number is greater for the African sample than the European sample. Is this what we observe?

[1]:
```
!pip install numpy
!pip install pandas
!pip install matplotlib
```

```
Requirement already satisfied: numpy in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (1.26.0)
Requirement already satisfied: pandas in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (2.1.1)
Requirement already satisfied: numpy>=1.22.4 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from pandas)
(1.26.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from pandas)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from pandas)
(2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from pandas)
(2023.3)
Requirement already satisfied: six>=1.5 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from python-
dateutil>=2.8.2->pandas) (1.16.0)
```

```
Requirement already satisfied: matplotlib in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (3.8.1)
Requirement already satisfied: contourpy>=1.0.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(1.1.1)
Requirement already satisfied: cycler>=0.10 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(4.43.1)
Requirement already satisfied: kiwisolver>=1.3.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(1.4.5)
Requirement already satisfied: numpy<2,>=1.21 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(1.26.0)
Requirement already satisfied: packaging>=20.0 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(23.1)
Requirement already satisfied: pillow>=8 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(10.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(2.8.2)
Requirement already satisfied: importlib-resources>=3.2.0 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from matplotlib)
(6.1.0)
Requirement already satisfied: zipp>=3.1.0 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from importlib-
resources>=3.2.0->matplotlib) (3.11.0)
Requirement already satisfied: six>=1.5 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
```

[2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[3]:
```
data = pd.read_csv("abbgen1k.csv")
data
```

[3]:

	CHROM	POS	ID	REF	ALT	QUAL	FILTER	\
0	22	16050115	rs587755077	G	A	100	PASS	
1	22	16050213	rs587654921	C	T	100	PASS	
2	22	16050783	rs587743568	A	G	100	PASS	
3	22	16050840	rs587616822	C	G	100	PASS	
4	22	16050958	rs587636807	A	T	100	PASS	
…	…	…	…	..	..	…	…	
15907	22	18097177	rs545154757	C	T	100	PASS	
15908	22	18097179	rs28539556	T	C	100	PASS	
15909	22	18097181	rs575863671	G	A	100	PASS	
15910	22	18097196	rs5747273	G	A	100	PASS	
15911	22	18097202	rs144996543	C	G	100	PASS	

	INFO	FORMAT	NA06984	…	\
0	AC=32;AF=0.00638978;AN=5008;NS=2504;DP=11468;E…	GT	0|0	…	
1	AC=38;AF=0.00758786;AN=5008;NS=2504;DP=15092;E…	GT	0|0	…	
2	AC=39;AF=0.00778754;AN=5008;NS=2504;DP=24717;E…	GT	0|0	…	
3	AC=26;AF=0.00519169;AN=5008;NS=2504;DP=23832;E…	GT	0|0	…	
4	AC=1;AF=0.000199681;AN=5008;NS=2504;DP=22326;E…	GT	0|0	…	
…	…	…	…	…	
15907	AC=16;AF=0.00319489;AN=5008;NS=2504;DP=15476;E…	GT	0|0	…	
15908	AC=270;AF=0.0539137;AN=5008;NS=2504;DP=15582;E…	GT	0|0	…	
15909	AC=19;AF=0.00379393;AN=5008;NS=2504;DP=15627;E…	GT	0|0	…	
15910	AC=1412;AF=0.281949;AN=5008;NS=2504;DP=15124;E…	GT	0|0	…	
15911	AC=42;AF=0.00838658;AN=5008;NS=2504;DP=14659;E…	GT	0|0	…	

	NA19143	NA19144	NA19146	NA19147	NA19113	NA19114	NA19256	NA19257	NA19117	\
0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
1	0|0	0|0	0|0	0|0	0|0	1|0	1|0	0|0	0|0	
2	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
3	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
4	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
…	…	…	…	…	…	…	…	…	…	
15907	0|0	0|0	0|0	0|1	0|0	0|0	0|0	0|0	0|0	
15908	0|0	0|1	0|0	0|0	0|0	0|0	0|0	1|1	0|0	
15909	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
15910	1|0	1|0	0|0	0|1	0|1	0|1	1|0	0|0	1|1	
15911	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	

	NA19118
0	0|0
1	0|0
2	0|0
3	0|0
4	0|0
…	…
15907	0|0

```
15908     0|0
15909     0|0
15910     0|0
15911     0|0

[15912 rows x 179 columns]
```

[4]:
```
data.shape
```

[4]:
```
(15912, 179)
```

[5]:
```
european_subset = data.iloc[:, 0:90]
european_subset
```

[5]:

	CHROM	POS	ID	REF	ALT	QUAL	FILTER	\
0	22	16050115	rs587755077	G	A	100	PASS	
1	22	16050213	rs587654921	C	T	100	PASS	
2	22	16050783	rs587743568	A	G	100	PASS	
3	22	16050840	rs587616822	C	G	100	PASS	
4	22	16050958	rs587636807	A	T	100	PASS	
…	…	…	…	..	..	…	…	
15907	22	18097177	rs545154757	C	T	100	PASS	
15908	22	18097179	rs28539556	T	C	100	PASS	
15909	22	18097181	rs575863671	G	A	100	PASS	
15910	22	18097196	rs5747273	G	A	100	PASS	
15911	22	18097202	rs144996543	C	G	100	PASS	

	INFO	FORMAT	NA06984	…	\
0	AC=32;AF=0.00638978;AN=5008;NS=2504;DP=11468;E…	GT	0|0	…	
1	AC=38;AF=0.00758786;AN=5008;NS=2504;DP=15092;E…	GT	0|0	…	
2	AC=39;AF=0.00778754;AN=5008;NS=2504;DP=24717;E…	GT	0|0	…	
3	AC=26;AF=0.00519169;AN=5008;NS=2504;DP=23832;E…	GT	0|0	…	
4	AC=1;AF=0.000199681;AN=5008;NS=2504;DP=22326;E…	GT	0|0	…	
…	…	…	…	…	
15907	AC=16;AF=0.00319489;AN=5008;NS=2504;DP=15476;E…	GT	0|0	…	
15908	AC=270;AF=0.0539137;AN=5008;NS=2504;DP=15582;E…	GT	0|0	…	
15909	AC=19;AF=0.00379393;AN=5008;NS=2504;DP=15627;E…	GT	0|0	…	
15910	AC=1412;AF=0.281949;AN=5008;NS=2504;DP=15124;E…	GT	0|0	…	
15911	AC=42;AF=0.00838658;AN=5008;NS=2504;DP=14659;E…	GT	0|0	…	

	NA12762	NA12763	NA12775	NA12776	NA12777	NA12778	NA12812	NA12814	NA12815	\
0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
1	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
2	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
3	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
4	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
…	…	…	…	…	…	…	…	…	…	

```
15907     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0
15908     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0
15909     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0
15910     0|1     0|0     0|0     0|0     1|1     1|0     0|0     1|0     0|0
15911     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0     0|0

      NA12827
0         0|0
1         0|0
2         0|0
3         0|0
4         0|0
…         …
15907     0|0
15908     0|0
15909     0|0
15910     1|1
15911     0|0

[15912 rows x 90 columns]
```

[6]:
```
def countPolymorphicSites(file_name: str) -> dict:
    data = pd.read_csv(file_name)
    europe_subset = data.iloc[:, 9:90]
    africa_subset = data.iloc[:, 90:179]
    africa_polymorphisms = sum(africa_subset.apply(checkIfPolymorphic, axis =␣
 ↪1))
    europe_polymorphisms = sum(europe_subset.apply(checkIfPolymorphic, axis =␣
 ↪1))
    return {"African subset non-polymorphic SNPs": africa_polymorphisms,
            "European subset non-polymorphic SNPs": europe_polymorphisms}

def checkIfPolymorphic(row): #checks if all columns in given SNP row are 0|0,␣
 ↪which would mean that there are no ALT genes
    #and thus the SNP is nonpolymorphic
    numNonpolymorphs = sum(row[0:] == "0|0")
    if numNonpolymorphs == row.count():
        return False
    else:
        return True
```

[7]:
```
print(countPolymorphicSites("abbgen1k.csv"))
```

```
{'African subset non-polymorphic SNPs': 12705, 'European subset non-polymorphic
SNPs': 7959}
```

In concordance with what we would expect under the out-of-Africa hypothesis, we see that the number of SNPs which are polymorphic is greater in the African subset of

the population compared to the European subset.

The allele frequency spectrum (see slide 64 in the Population Genetics lecture) shows the number of SNPs for which there are k haplotypes that have the ALT allele. For example, when k = 0 this is the number of SNPs such that there are no haplotypes with the ALT allele (all haplotypes have the REF allele), when k = 1 this is the number of SNPs such that exactly one haplotype has the ALT allele (and all other haplotypes have the REF allele), when k = 2 this is the number of SNPs such that exactly two haplotypes have the ALT allele (and all other haplotypes have the REF allele), … For the small example above, there is one SNP (SNP D) with k = 0 ALT alleles, there is one SNP (SNP A) with k = 1 ALT alleles, and there are two SNPs (SNPs B and C) with k = 3 ALT alleles. The allele frequency spectrum is then a barplot with the k numbers on the x-axis and the number of SNPs with k ALT allele on the y-axis. Note: do not plot k = 0, start with k = 1 (the k = 0 case are those sites that are not polymorphic, the number of SNPs with k = 0 ALT alleles is much higher than for all the other k numbers).

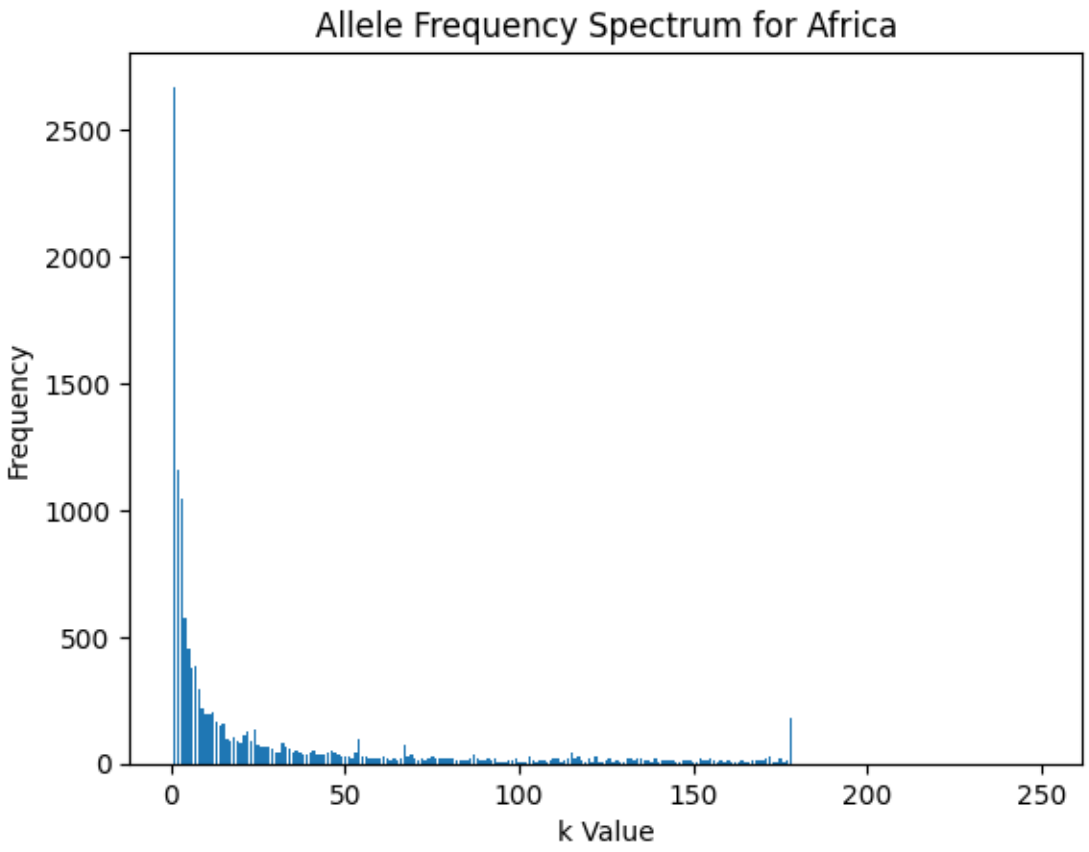
2. Generalize what you did in #1 to compute the allele frequency spectrum. Make two separate plots: the allele frequency spectrum for the European individuals and the allele frequency spectrum for the African individuals. Comment on any similarities or differences between the plots.

[8]:
```
def extractAlleleFreqSpect(file_name: str): #subset dataframe into europe &␣
 ↪africa, then apply getALleleFreq function to return k-counts
    data = pd.read_csv(file_name)
    europe_subset = data.iloc[:, 9:90]
    african_subset = data.iloc[:, 90:179]
    africanKCounts = [0]*250
    europeanKCounts = [0]*250
    african_subset.apply(lambda row: getAlleleFreq(row, africanKCounts), axis =␣
 ↪1)
    europe_subset.apply(lambda row: getAlleleFreq(row, europeanKCounts), axis =␣
 ↪1)
    return africanKCounts[1:], europeanKCounts[1:]

def getAlleleFreq(row, kCountsList): #sums each SNP row by the number of ALT␣
 ↪haplotypes
    alleleFreq = 2 * sum(row[0:] == "1|1") + sum(row[0:] == "1|0") + sum(row[0:
 ↪] == "0|1")
    kCountsList[alleleFreq] += 1

def plotAlleleFreqSpect(alleleFreqList, subpopName): #utilize pyplot to␣
 ↪construct allele frequency spectra
    plt.bar(range(1,250), alleleFreqList)
    plt.xlabel("k Value")
    plt.ylabel ("Frequency")
    plt.title(f"Allele Frequency Spectrum for {subpopName}")
```

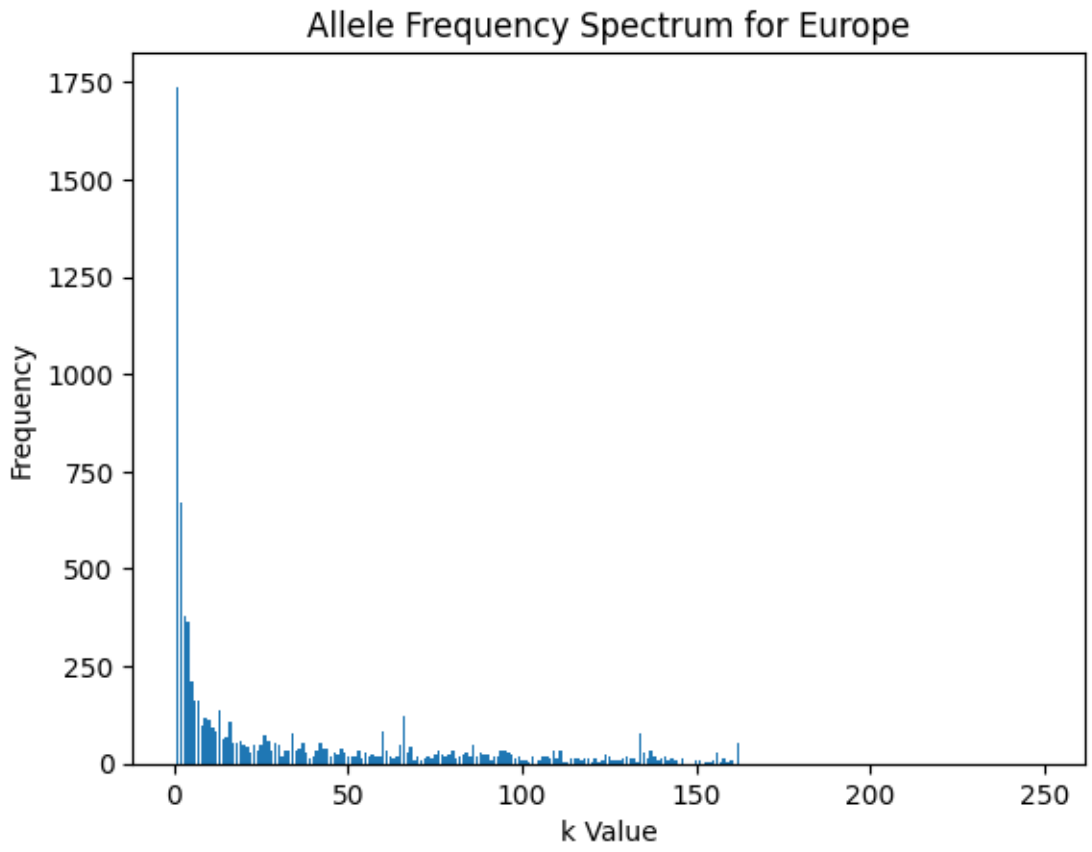
[9]:
```
africa, europe = extractAlleleFreqSpect("abbgen1k.csv")
print(africa, europe)
```

```
[2672, 1156, 1046, 577, 456, 377, 383, 297, 219, 198, 198, 200, 163, 151, 161,
93, 87, 102, 88, 84, 110, 126, 92, 133, 77, 69, 67, 65, 61, 40, 45, 83, 66, 62,
42, 54, 45, 35, 37, 44, 52, 39, 33, 36, 44, 49, 43, 34, 26, 27, 26, 24, 45, 98,
25, 28, 23, 21, 19, 23, 26, 18, 14, 17, 12, 18, 76, 31, 34, 22, 15, 17, 14, 19,
29, 20, 18, 17, 23, 21, 20, 10, 15, 10, 16, 18, 33, 18, 11, 16, 19, 10, 21, 7,
9, 9, 13, 16, 23, 6, 9, 6, 31, 11, 6, 14, 10, 8, 14, 17, 18, 9, 11, 19, 40, 17,
25, 12, 6, 24, 4, 28, 6, 6, 12, 18, 9, 15, 7, 9, 21, 18, 12, 20, 20, 15, 10, 4,
19, 6, 14, 11, 12, 12, 4, 9, 10, 15, 11, 9, 6, 19, 13, 14, 17, 10, 6, 10, 2, 16,
2, 9, 5, 13, 9, 6, 13, 14, 11, 11, 24, 28, 8, 3, 20, 8, 15, 183, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] [1738, 668, 379, 364, 210, 163, 163, 95,
119, 111, 91, 85, 137, 61, 66, 107, 54, 52, 57, 46, 43, 27, 48, 32, 46, 74, 56,
34, 51, 48, 20, 35, 32, 77, 31, 39, 52, 29, 15, 16, 35, 53, 36, 39, 20, 30, 24,
39, 26, 19, 18, 16, 32, 13, 26, 18, 25, 17, 20, 81, 34, 16, 12, 16, 47, 120, 30,
44, 10, 16, 10, 12, 19, 11, 21, 31, 24, 18, 21, 35, 15, 18, 21, 30, 19, 49, 18,
28, 23, 21, 10, 17, 16, 33, 34, 28, 25, 14, 20, 10, 9, 6, 16, 0, 9, 16, 17, 13,
34, 15, 31, 5, 5, 13, 14, 11, 10, 11, 12, 5, 11, 5, 7, 23, 20, 10, 10, 8, 13,
20, 13, 14, 5, 80, 28, 12, 35, 18, 9, 15, 16, 7, 15, 7, 1, 12, 1, 1, 1, 10, 9,
0, 3, 2, 7, 27, 4, 12, 3, 8, 0, 51, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

[10]:
```
plotAlleleFreqSpect(africa, "Africa")
```



[11]:
```
plotAlleleFreqSpect(europe, "Europe")
```



Both plots demonstrate a geometric distribution which has a high frequency for lower k counts and quickly declines to relativley low frequencies for higher k counts. For the African allele frequency spectrum, we observe a higher initial frequency at k = 1 than the European spectrum. We also observe higher possible k values for the African subset as the plot shows some non-zero frequencies near k = 175 while Europe has k frequencies near k = 160. Overall, there appears to be higher frequency counts for Africa compared to Europe which may be explained in part by the higher number of polymorphic sites observed for Africa (as observed in Question 1).

Next I want you to compute the bootstrap 95% confidence interval for the average pairwise diversity statistic. Below are instructions to compute the average pairwise diversity statistic: a. Randomly pick two haplotypes. For our small example, say we pick the left haplotype of NA1 (1,0,1,0) and the right haplotype of NA2 (0,0,1,0). This pair of haplotypes disagree at the first SNP but agree at the other three SNPs. So = 0.25 (the fraction of SNPs for which the pair disagrees). b. Repeat step (a) 50 times (randomly selecting two haplotypes each time). Take the average of these 50 values. This is our estimate of the average pairwise diversity. To get the bootstrap 95% confidence interval repeat steps (a) and (b) 100 times, then sort the 100 average pairwise diversity estimates, the estimates at the 2nd and the 98th positions make the bootstrap 95% confidence interval. Note: You can't use the "myboot" function from lecture because that function used a vector as input, while for this problem you need two-dimensional matrices.

3. I want you to compute three separate bootstrap 95% confidence intervals for the average pairwise diversity statistic: one within the European samples, one within the African samples, and one between the European and African samples. To get the three confidence intervals, you have to repeat the entire procedure above three times: to get the confidence intervals within the European samples all the randomly chosen haplotypes are from Europeans, to get the confidence interval within the African samples all the randomly haplotypes are from Africans, and to get the confidence interval between the European and African samples all the randomly chosen pairs of haplotypes contain one from a European and one from an African. Due to the out-of-Africa hypothesis, we expect that the average pairwise diversity within Europe will be less than the other two. Is this what we observe?

Note: You cannot eliminate all loops for this problem, but do try to eliminate unnecessary loops (which can really slow your code down)

[12]:
```
def calculateAvgPairwiseDiversity(popDataframe, reps: int, mixedSamples: bool)␣
 ↪-> float: #obtains average pi value across 50 reps
    sumPiValues = 0
    for replication in range(reps):
        if mixedSamples == False: #checks if we want Afro-European sampling
            haplotype_pairs = popDataframe.sample(n = 2, axis = 1, replace =␣
 ↪True)
        else:
            haplotype_pairs = pd.concat([popDataframe.iloc[:, 9:90].sample(n =␣
 ↪1, axis = 1, replace = True), popDataframe.iloc[:, 90:179].sample(n = 1,␣
 ↪axis = 1, replace = True)], axis = 1)
        hapOneCol = np.random.randint(2) #selects one of either maternal or␣
 ↪paternal haplotype
        hapOne = haplotype_pairs.iloc[:,0].apply(lambda row: row.
 ↪split("|")[hapOneCol].strip()).to_numpy()
        hapTwoCol = np.random.randint(2)
        hapTwo = haplotype_pairs.iloc[:,1].apply(lambda row: row.
 ↪split("|")[hapTwoCol].strip()).to_numpy()
        numPairwiseMatches = sum(hapOne != hapTwo) # Numpy's != operator␣
 ↪returns boolean vector where True corresponds to rows
        # where all columns match. We sum this value to obtain all the SNPs␣
 ↪that match for our two randomly sampled haplotypes.
        repPiValue = float(numPairwiseMatches)/len(hapOne)
        sumPiValues += repPiValue
    averagePiValue = float(sumPiValues)/reps #obtain average Pi value across␣
 ↪all reps
    return averagePiValue

def bootstrapPairwiseDiversity(popDataframe, numTrials: int, mixedSamples:␣
 ↪bool) -> list:
    bootstrapValues = []
    for trial in range(numTrials): #conducts n trials of k reps
```

```
        bootstrapValues.append(calculateAvgPairwiseDiversity(popDataframe, 50,␣
↪mixedSamples))
    bootstrapValues.sort() #sorts average trial Pi values from smallest to␣
↪largest
    return [bootstrapValues[2], bootstrapValues[98]] #selects values␣
↪corresponding to 95% confidence interval
```

[13]:
```
data = pd.read_csv("abbgen1k.csv")
african_subset = data.iloc[:, 9:90]
print("African subset CI: " + str(bootstrapPairwiseDiversity(african_subset,␣
↪100, False)))
```

```
African subset CI: [0.08894419306184016, 0.09707642031171446]
```

[14]:
```
european_subset = data.iloc[:, 90:179]
print("European subset CI: " + str(bootstrapPairwiseDiversity(european_subset,␣
↪100, False)))
```

```
European subset CI: [0.10880718954248364, 0.11700854700854696]
```

[15]:
```
afroeuropean_subset = data
print("Afro-European subset CI: " +␣
↪str(bootstrapPairwiseDiversity(afroeuropean_subset, 100, True)))
```

```
Afro-European subset CI: [0.11547762694821519, 0.1211236802413273]
```

In concordance with the out-of-Africa hypothesis, we do observe a lower average pairwise diversity within Europe compared to the Africa and Europe-Africa population compositions based on our bootstrapping results. The European and Europe-Africa population compositions have similar overlapping confidence intervals whereas the African sample clearly has a lower confidence interval than both.

[16]:
```
!pip install nbconvert[webpdf]
```

```
Collecting nbconvert[webpdf]
  Downloading nbconvert-7.10.0-py3-none-any.whl.metadata (7.8 kB)
Collecting beautifulsoup4 (from nbconvert[webpdf])
  Downloading beautifulsoup4-4.12.2-py3-none-any.whl (142 kB)
                           143.0/143.0
kB 1.9 MB/s eta 0:00:00a 0:00:01
Collecting bleach!=5.0.0 (from nbconvert[webpdf])
  Downloading bleach-6.1.0-py3-none-any.whl.metadata (30 kB)
Collecting defusedxml (from nbconvert[webpdf])
  Downloading defusedxml-0.7.1-py2.py3-none-any.whl (25 kB)
Requirement already satisfied: importlib-metadata>=3.6 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
nbconvert[webpdf]) (6.0.0)
Collecting jinja2>=3.0 (from nbconvert[webpdf])
```

```
  Downloading Jinja2-3.1.2-py3-none-any.whl (133 kB)
                             133.1/133.1
kB 5.2 MB/s eta 0:00:00
Requirement already satisfied: jupyter-core>=4.7 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
nbconvert[webpdf]) (5.3.0)
Collecting jupyterlab-pygments (from nbconvert[webpdf])
  Downloading jupyterlab_pygments-0.2.2-py2.py3-none-any.whl (21 kB)
Collecting markupsafe>=2.0 (from nbconvert[webpdf])
  Downloading MarkupSafe-2.1.3-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86
_64.whl.metadata (3.0 kB)
Collecting mistune<4,>=2.0.3 (from nbconvert[webpdf])
  Downloading mistune-3.0.2-py3-none-any.whl.metadata (1.7 kB)
Collecting nbclient>=0.5.0 (from nbconvert[webpdf])
  Downloading nbclient-0.8.0-py3-none-any.whl.metadata (7.8 kB)
Collecting nbformat>=5.7 (from nbconvert[webpdf])
  Downloading nbformat-5.9.2-py3-none-any.whl.metadata (3.4 kB)
Requirement already satisfied: packaging in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
nbconvert[webpdf]) (23.1)
Collecting pandocfilters>=1.4.1 (from nbconvert[webpdf])
  Downloading pandocfilters-1.5.0-py2.py3-none-any.whl (8.7 kB)
Requirement already satisfied: pygments>=2.4.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
nbconvert[webpdf]) (2.15.1)
Collecting tinycss2 (from nbconvert[webpdf])
  Downloading tinycss2-1.2.1-py3-none-any.whl (21 kB)
Requirement already satisfied: traitlets>=5.1 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
nbconvert[webpdf]) (5.7.1)
Collecting playwright (from nbconvert[webpdf])
  Downloading playwright-1.39.0-py3-none-manylinux1_x86_64.whl.metadata (3.6 kB)
Requirement already satisfied: six>=1.9.0 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
bleach!=5.0.0->nbconvert[webpdf]) (1.16.0)
Collecting webencodings (from bleach!=5.0.0->nbconvert[webpdf])
  Downloading webencodings-0.5.1-py2.py3-none-any.whl (11 kB)
Requirement already satisfied: zipp>=0.5 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from importlib-
metadata>=3.6->nbconvert[webpdf]) (3.11.0)
Requirement already satisfied: platformdirs>=2.5 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from jupyter-
core>=4.7->nbconvert[webpdf]) (3.10.0)
Requirement already satisfied: jupyter-client>=6.1.12 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
nbclient>=0.5.0->nbconvert[webpdf]) (8.1.0)
Collecting fastjsonschema (from nbformat>=5.7->nbconvert[webpdf])
```

```
  Downloading fastjsonschema-2.18.1-py3-none-any.whl.metadata (2.0 kB)
Collecting jsonschema>=2.6 (from nbformat>=5.7->nbconvert[webpdf])
  Downloading jsonschema-4.19.2-py3-none-any.whl.metadata (7.9 kB)
Collecting soupsieve>1.2 (from beautifulsoup4->nbconvert[webpdf])
  Downloading soupsieve-2.5-py3-none-any.whl.metadata (4.7 kB)
Collecting greenlet==3.0.0 (from playwright->nbconvert[webpdf])
  Downloading greenlet-3.0.0-cp39-cp39-manylinux_2_24_x86_64.manylinux_2_28_x86_
64.whl.metadata (3.8 kB)
Collecting pyee==11.0.1 (from playwright->nbconvert[webpdf])
  Downloading pyee-11.0.1-py3-none-any.whl.metadata (2.7 kB)
Requirement already satisfied: typing-extensions in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from
pyee==11.0.1->playwright->nbconvert[webpdf]) (4.7.1)
Collecting attrs>=22.2.0 (from
jsonschema>=2.6->nbformat>=5.7->nbconvert[webpdf])
  Downloading attrs-23.1.0-py3-none-any.whl (61 kB)
                             61.2/61.2 kB
4.5 MB/s eta 0:00:00
Collecting jsonschema-specifications>=2023.03.6 (from
jsonschema>=2.6->nbformat>=5.7->nbconvert[webpdf])
  Downloading jsonschema_specifications-2023.7.1-py3-none-any.whl.metadata (2.8
kB)
Collecting referencing>=0.28.4 (from
jsonschema>=2.6->nbformat>=5.7->nbconvert[webpdf])
  Downloading referencing-0.30.2-py3-none-any.whl.metadata (2.6 kB)
Collecting rpds-py>=0.7.1 (from
jsonschema>=2.6->nbformat>=5.7->nbconvert[webpdf])
  Downloading
rpds_py-0.10.6-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(3.7 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from jupyter-
client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (2.8.2)
Requirement already satisfied: pyzmq>=23.0 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from jupyter-
client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (25.1.0)
Requirement already satisfied: tornado>=6.2 in
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-packages (from jupyter-
client>=6.1.12->nbclient>=0.5.0->nbconvert[webpdf]) (6.3.3)
Downloading bleach-6.1.0-py3-none-any.whl (162 kB)
                           162.8/162.8 kB
8.3 MB/s eta 0:00:00
Downloading
MarkupSafe-2.1.3-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (25
kB)
Downloading mistune-3.0.2-py3-none-any.whl (47 kB)
                           48.0/48.0 kB
2.9 MB/s eta 0:00:00
```

```
Downloading nbclient-0.8.0-py3-none-any.whl (73 kB)
                           73.1/73.1 kB
3.0 MB/s eta 0:00:00
Downloading nbformat-5.9.2-py3-none-any.whl (77 kB)
                           77.6/77.6 kB
4.4 MB/s eta 0:00:00
Downloading nbconvert-7.10.0-py3-none-any.whl (256 kB)
                           256.7/256.7 kB
13.5 MB/s eta 0:00:00
Downloading playwright-1.39.0-py3-none-manylinux1_x86_64.whl (35.5 MB)
                           35.5/35.5 MB
32.8 MB/s eta 0:00:0000:0100:01
Downloading
greenlet-3.0.0-cp39-cp39-manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl (610
kB)
                           610.4/610.4 kB
26.6 MB/s eta 0:00:00
Downloading pyee-11.0.1-py3-none-any.whl (15 kB)
Downloading jsonschema-4.19.2-py3-none-any.whl (83 kB)
                           83.6/83.6 kB
5.2 MB/s eta 0:00:00
Downloading soupsieve-2.5-py3-none-any.whl (36 kB)
Downloading fastjsonschema-2.18.1-py3-none-any.whl (23 kB)
Downloading jsonschema_specifications-2023.7.1-py3-none-any.whl (17 kB)
Downloading referencing-0.30.2-py3-none-any.whl (25 kB)
Downloading
rpds_py-0.10.6-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2 MB)
                           1.2/1.2 MB
19.7 MB/s eta 0:00:00a 0:00:01
Installing collected packages: webencodings, fastjsonschema, tinycss2,
soupsieve, rpds-py, pyee, pandocfilters, mistune, markupsafe, jupyterlab-
pygments, greenlet, defusedxml, bleach, attrs, referencing, playwright, jinja2,
beautifulsoup4, jsonschema-specifications, jsonschema, nbformat, nbclient,
nbconvert
Successfully installed attrs-23.1.0 beautifulsoup4-4.12.2 bleach-6.1.0
defusedxml-0.7.1 fastjsonschema-2.18.1 greenlet-3.0.0 jinja2-3.1.2
jsonschema-4.19.2 jsonschema-specifications-2023.7.1 jupyterlab-pygments-0.2.2
markupsafe-2.1.3 mistune-3.0.2 nbclient-0.8.0 nbconvert-7.10.0 nbformat-5.9.2
pandocfilters-1.5.0 playwright-1.39.0 pyee-11.0.1 referencing-0.30.2 rpds-
py-0.10.6 soupsieve-2.5 tinycss2-1.2.1 webencodings-0.5.1
```

```
[ ]: !jupyter nbconvert --to pdf QBIO\ 401\ HW\ 6-\ Hosseini.ipynb
```

```
[NbConvertApp] Converting notebook QBIO 401 HW 6- Hosseini.ipynb to pdf
/home/hiradh/miniconda3/envs/rna/lib/python3.9/site-
packages/nbconvert/utils/pandoc.py:51: RuntimeWarning: You are using an
unsupported version of pandoc (2.9.2.1).
Your version must be at least (2.14.2) but less than (4.0.0).
```

```
Refer to https://pandoc.org/installing.html.
Continuing with doubts…
  check_pandoc_version()
[NbConvertApp] Support files will be in QBIO 401 HW 6- Hosseini_files/
[NbConvertApp] Making directory ./QBIO 401 HW 6- Hosseini_files
[NbConvertApp] Writing 64642 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex', '-quiet']
```