Project 2 Report

Clayton Engle

CSCI 3327 Probability and Applied Statistics

Stockton University

4 May 2023

Contents

I Formulas **3**
0.1 Arithmetic Operations 4
0.2 Conditional Probability 5
0.3 Binomial Distribution 7
0.4 Geometric Distribution 8
0.4.1 Hypergeometric Distribution 8
0.5 Probability Mass Function 9
0.6 Poisson Distribution 9
0.7 Tchebyshev's Theorem 10
0.7.1 Continuous Uniform Distribution 10
0.7.2 Continuous Normal Distribution 11
0.8 Gamma Distribution 11
0.9 Beta Distribution 13
0.10 Marginal Probability Density Function 15
0.11 Conditional Probability Density Function 16
0.12 Distribution Function 17
0.13 Expected Value 18

II **20**
0.14 21
0.14.1 Binomial Distribution 23
0.14.2 Bayes' Theorem 24
0.14.3 Hypergeometric Distribution 24
0.14.4 Poisson Distribution 25
0.14.5 Tchebychev's Theorem 25

III Data Processing and Plotting **26**
0.15 The Apache Common Mathematics Library 27
0.15.1 The ApacheSmoother Class 27
0.15.2 The ApachePlotter Class 27
0.16 Working with Octave 28

0.17 Key Differences and Features 29

IV Sources **31**

0.18 Data Structures and Algorithms 32
0.19 Design Patterns 32
0.20 Math Formulas and Proofs 32

Part I

Formulas

In this section, all the formulas we used in class are typed up. In some cases I typed out the proof of the formulas to use as a reference in later sections when discussing how I implemented the solver for each formula in my program.

0.1 Arithmetic Operations

Mean:

$$\overline{mean(x)} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

In this formula, $\overline{x}$ represents the arithmetic mean, n represents the number of values in the set, and x_i represents the i-th value in the set. The formula represents the sum of all values in the set divided by the number of values in the set, which gives the average value of the set.

Median:

$$\tilde{x} = x_{\frac{n+1}{2}} \text{ if n is odd, } \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) \text{ if n is even} \tag{2}$$

In this formula, $\tilde{x}$ represents the arithmetic median, n represents the number of values in the set, and $x_{\frac{n+1}{2}}$ and $x_{\frac{n}{2}}$ represent the middle values in the set when n is odd or even, respectively.

Mode:

$$\text{mode}(x) = \underset{x_i}{\text{argmax}} \sum_{i=1}^{n} \delta(x - x_i) \tag{3}$$

In this formula, mode(x) represents the arithmetic mode, n represents the number of values in the set, x_i represents the i-th value in the set, and $\delta(x - x_i)$ is the Kronecker delta function, which equals 1 if $x = x_i$ and 0 otherwise. The mode is the value that appears most frequently in the set.

Variance:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{4}$$

In this formula, σ^2 represents the arithmetic variance, n represents the number of values in the set, x_i represents the i-th value in the set, and $\overline{x}$ represents the arithmetic mean of the set. The variance measures the spread of the set around the mean.

Similar to the mean, the variance calculation can be manipulated so that multiple processes can be executed at once. Understanding the bones of the variance formula will be essential for explaining the details of how we can do this, which will be discussed in a later section of the report. So, starting with the variance formula shown above, here is the proof:

Expanding the square, we get:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) \tag{5}$$

Using the linearity of summation, we can split this sum into three separate sums:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{2}{n}\sum_{i=1}^{n} x_i\overline{x} + \frac{1}{n}\sum_{i=1}^{n} \overline{x}^2 \tag{6}$$

The first sum can be rewritten in terms of the square of the individual deviations from the mean:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{n}\sum_{i=1}^{n} \overline{x}^2 - \frac{2}{n}\sum_{i=1}^{n} x_i\overline{x} \tag{7}$$

Simplifying the second and third terms, we get:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \overline{x}^2 - 2\overline{x}^2 \tag{8}$$

Simplifying further, we get:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 - \overline{x}^2 \tag{9}$$

Finally, using the definition of the arithmetic mean, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, we can substitute and simplify further:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2 \tag{10}$$

This completes the proof of the arithmetic variance formula.

Standard Deviation:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{11}$$

In this formula, σ represents the arithmetic standard deviation, n represents the number of values in the set, x_i represents the i-th value in the set, and $\overline{x}$ represents the arithmetic mean of the set. The standard deviation is the square root of the variance and also measures the spread of the set around the mean, but in the same units as the original data.

0.2 Conditional Probability

Combinations:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Permutations:

$$P_n^k = \frac{n!}{(n-k)!}$$

In both formulas, n represents the total number of items, and k represents the number of items to be selected. The notation $\binom{n}{k}$ represents the number of combinations of n items taken k at a time, and P_n^k represents the number of permutations of n items taken k at a time.

Probability of event A given event B:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

In this formula, $P(A \mid B)$ represents the probability of event A given that event B has occurred. $P(A \cap B)$ represents the probability that both events A and B occur, and $P(B)$ represents the probability that event B occurs.

Bayes' Theorem:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

In this formula, $P(A \mid B)$ represents the probability of event A given event B, $P(B \mid A)$ represents the probability of event B given event A, $P(A)$ represents the prior probability of event A, and $P(B)$ represents the prior probability of event B. The denominator $P(B)$ is known as the marginal probability of event B, which can be calculated using the law of total probability:

$$P(B) = \sum_i P(B \mid A_i) \cdot P(A_i)$$

Where A_i represents all possible events that are mutually exclusive and exhaustive. Bayes' Theorem can be generalized for any number of events. Let's assume that we have n mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_n$, and an event B that we are interested in. Then, Bayes' theorem can be generalized to:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)}$$

for $i = 1, 2, \ldots, n$.
Starting from the definition of conditional probability:

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

Using the law of total probability, we can express the joint probability $P(A_i \cap B)$ as:

$$P(A_i \cap B) = P(B \mid A_i) \cdot P(A_i)$$

Substituting this expression into the numerator of the previous equation, we get:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{P(B)}$$

Using the law of total probability again, we can express the marginal probability $P(B)$ as:

$$P(B) = \sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)$$

Substituting this expression into the denominator of the previous equation, we get:

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j) \cdot P(A_j)}$$

This is the generalized form of Bayes' theorem for n events. It allows us to update our beliefs about the probability of each event A_i given new evidence in the form of event B.

0.3 Binomial Distribution

Probability Mass Function:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{12}$$

where:

- $\binom{n}{k}$ is the binomial coefficient, representing the number of ways to choose k items out of n.
- p^k is p raised to the power of k, indicating the probability of k successes in n trials.
- $(1-p)^{n-k}$ is $(1-p)$ raised to the power of $(n-k)$, indicating the probability of $(n-k)$ failures in n trials.

Expected value of a binomial distribution:

$$E(X) = np \tag{13}$$

The variance of a binomial distribution is given by:

$$Var(X) = np(1-p) \tag{14}$$

Probability of a success at or before n trials:

$$P(X \leq n) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \tag{15}$$

Probability of a success after n trials:

$$P(X > n) = \sum_{k=n+1}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \tag{16}$$

0.4 Geometric Distribution

Probability mass function:

$$P(X = k) = (1-p)^{k-1} p$$

where: X is a random variable representing the number of trials until the first success p is the probability of success on each trial (and $q = 1 - p$ is the probability of failure) k is the number of trials (where $k = 1, 2, 3, \ldots$) The mean of a geometric distribution is:

$$E(X) = \frac{1}{p}$$

The variance of a geometric distribution is:

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

And the standard deviation of a geometric distribution is:

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{1-p}{p^2}}$$

0.4.1 Hypergeometric Distribution

The Hypergeometric distribution models the probability of drawing k objects of a certain type from a population of N objects, where K of the objects are of the type being sought, and the sampling is done without replacement.

0.5 Probability Mass Function

The probability mass function (pmf) of the Hypergeometric distribution is:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{Nn} \tag{17}$$

where X is the random variable representing the number of objects of the type being sought in a sample of size n, and $0 \leq k \leq K$.

Mean

The mean of the Hypergeometric distribution is:

$$\mu = \frac{nK}{N} \tag{18}$$

Variance

The variance of the Hypergeometric distribution is:

$$\sigma^2 = \frac{nK(N-K)(N-n)}{N^2(N-1)} \tag{19}$$

0.6 Poisson Distribution

The Poisson distribution models the probability of a given number of events occurring in a fixed interval of time or space, given that these events occur independently and at a constant rate.

Probability Mass Function

The probability mass function (pmf) of the Poisson distribution is:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{20}$$

where X is the random variable representing the number of events occurring in a fixed interval of time or space, and λ is the mean number of events in that interval.

Mean

The mean of the Poisson distribution is:

$$\mu = \lambda \tag{21}$$

Variance

The variance of the Poisson distribution is:

$$\sigma^2 = \lambda \tag{22}$$

0.7 Tchebyshev's Theorem

Tchebyshev's theorem provides a bound on the proportion of data that lies within a certain number of standard deviations of the mean, for any data distribution. Specifically, for any distribution with mean μ and standard deviation σ, and for any number $k > 0$, at least $1 - 1/k^2$ of the data lies within k standard deviations of the mean. Mathematically, this can be expressed as:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{23}$$

where X is a random variable with mean μ and standard deviation σ.

Note that this inequality is an upper bound, and that the proportion of data lying within k standard deviations of the mean could be higher than the bound given by Tchebyshev's theorem, depending on the specific distribution of the data.

0.7.1 Continuous Uniform Distribution

The continuous uniform distribution models a situation where a continuous random variable can take on any value within a given range, with all values being equally likely.

Probability Density Function

The probability density function (pdf) of the continuous uniform distribution is:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \ 0 \\ \text{otherwise} \end{cases} \tag{24}$$

where a and b are the lower and upper bounds of the range of possible values.

Mean

The mean of the continuous uniform distribution is:

$$\mu = \frac{a+b}{2} \tag{25}$$

Variance

The variance of the continuous uniform distribution is:

$$\sigma^2 = \frac{(b-a)^2}{12} \tag{26}$$

0.7.2 Continuous Normal Distribution

The continuous normal distribution, also known as the Gaussian distribution, is one of the most important and widely used probability distributions. It is commonly used to model natural phenomena such as measurement errors and biological variation.

Probability Density Function

The probability density function (pdf) of the continuous normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{27}$$

where μ is the mean and σ is the standard deviation.

Mean

The mean of the continuous normal distribution is simply the mean parameter μ:

$$\mu \tag{28}$$

Variance

The variance of the continuous normal distribution is the square of the standard deviation parameter σ^2:

$$\sigma^2 \tag{29}$$

Note that the normal distribution is symmetric around its mean, and that approximately 68

0.8 Gamma Distribution

The probability density function of a Gamma distribution with shape parameter α and scale parameter β is given by:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \tag{30}$$

To prove this formula, we will use the following integral:

$$\int_0^\infty x^{\alpha-1} e^{-x} dx = \Gamma(\alpha) \tag{31}$$

where $\Gamma(\alpha)$ is the gamma function. We can rewrite the Gamma distribution pdf as:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1)} \tag{32}$$

This allows us to simplify the denominator:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha+1}\Gamma(\alpha+1)} \cdot \frac{\beta^{\alpha+1}\Gamma(\alpha+1)}{\beta^{\alpha}\Gamma(\alpha)} \tag{33}$$

Using the definition of the gamma function, we can simplify this further:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha+1}} \cdot \frac{1}{\alpha} \tag{34}$$

This simplifies to the formula for the Gamma distribution pdf that we started with, which completes our proof.

The mean of a Gamma distribution with shape parameter α and scale parameter β is given by:

$$\mu = \alpha\beta \tag{35}$$

To prove this formula, we will use the definition of the mean of a continuous probability distribution:

$$\mu = \int_{-\infty}^{\infty} xf(x)dx \tag{36}$$

where $f(x)$ is the probability density function of the distribution. Using the formula for the Gamma distribution pdf, we can write the integral as:

$$\mu = \int_{0}^{\infty} x \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}dx \tag{37}$$

We can simplify this by factoring outWe can simplify this by factoring out β from the integral and substituting $u = x/\beta$:

$$\mu = \frac{1}{\beta^{\alpha}} \int_{0}^{\infty} u^{\alpha}e^{-u}\beta^{2}du \tag{38}$$

This simplifies to:

$$\mu = \frac{\beta}{\Gamma(\alpha)} \int_{0}^{\infty} u^{\alpha}e^{-u}udu \tag{39}$$

Using integration by parts, we can evaluate this integral as:

$$\mu = \frac{\beta}{\Gamma(\alpha)} \left[-u^{\alpha}e^{-u}\Big|_{0}^{\infty} + \alpha \int_{0}^{\infty} u^{\alpha-1}e^{-u}du \right] \tag{40}$$

The first term evaluates to zero, and the second term simplifies to:

$$\mu = \frac{\alpha\beta}{\Gamma(\alpha)} \int_{0}^{\infty} u^{\alpha-1}e^{-u}du \tag{41}$$

This integral is equal to $\Gamma(\alpha)$, so we have:

$$\mu = \alpha\beta \tag{42}$$

which completes our proof.

The variance of a Gamma distribution with shape parameter α and scale parameter β is given by:

$$\sigma^2 = \alpha\beta^2 \tag{43}$$

To prove this formula, we will use the definition of the variance of a continuous probability distribution:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \tag{44}$$

where $f(x)$ is the probability density function of the distribution, and μ is the mean. Using the formula for the Gamma distribution pdf and mean, we can write the integral as:

$$\sigma^2 = \int_{0}^{\infty} (x-\alpha\beta)^2 \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} dx \tag{45}$$

Expanding the square, we get:

$$\sigma^2 = \int_{0}^{\infty} x^2 \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} dx - 2\alpha\beta \int_{0}^{\infty} x \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} dx + \alpha^2\beta^2 \tag{46}$$

The first integral is the same as the mean squared:

$$\int_{0}^{\infty} x^2 \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} dx = \alpha(\alpha+1)\beta^2 \tag{47}$$

The second integral is equal to the mean times 2α:

$$\int_{0}^{\infty} x \cdot \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} dx = \alpha\beta \tag{48}$$

Substituting these values back into the original equation, we get:

$$\sigma^2 = \alpha(\alpha+1)\beta^2 - 2\alpha^2\beta^2 + \alpha^2\beta^2 \tag{49}$$

Simplifying, we get:

$$\sigma^2 = \alpha\beta^2 \tag{50}$$

which completes our proof.

0.9 Beta Distribution

The Beta distribution is a continuous probability distribution defined on the interval $[0, 1]$. It is often used to model the behavior of random variables that take on values in this interval. In this article, we will provide the formulas for the pdf, mean, and variance of the Beta distribution, and provide mathematical proofs to show that these formulas work correctly.

The probability density function of a Beta distribution with shape parameters α and β is given by:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{51}$$

where $B(\alpha,\beta)$ is the beta function, defined as:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \tag{52}$$

To prove the correctness of the formula for the pdf, we need to show that it satisfies the properties of a probability density function. Namely, the integral of the pdf over the interval $[0,1]$ must be equal to 1. We have:

$$\begin{aligned}\int_0^1 f(x;\alpha,\beta)dx &= \int_0^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\ &= \frac{1}{B(\alpha,\beta)}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{B(\alpha,\beta)}{B(\alpha,\beta)} \\ &= 1\end{aligned}$$

where the last equality follows from the definition of the beta function.

The mean of a Beta distribution with shape parameters α and β is given by:

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{53}$$

To prove the correctness of this formula, we need to show that it satisfies the definition of the mean. Namely, we need to show that:

$$\int_0^1 xf(x;\alpha,\beta)dx = \mu \tag{54}$$

where $f(x;\alpha,\beta)$ is the pdf of the Beta distribution. We have:

$$\begin{aligned}\int_0^1 xf(x;\alpha,\beta)dx &= \int_0^1 x\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\ &= \frac{1}{B(\alpha,\beta)}\int_0^1 x^{\alpha}(1-x)^{\beta-1}dx \\ &= \frac{1}{B(\alpha,\beta)}B(\alpha+1,\beta) \\ &= \frac{\alpha}{\alpha+\beta}\end{aligned}$$

where the third equality follows from a change of variables, and the fourth equality follows from the definition of the beta function.

The variance of a Beta distribution with shape parameters α and β is given by:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{55}$$

To prove the correctness of this formula, we need to show that it satisfies the definition of the variance. Namely, we need to show that:

$$\int_0^1 (x-\mu)^2 f(x;\alpha,\beta)dx = \sigma^2 \tag{56}$$

where $f(x;\alpha,\beta)$ is the pdf of the Beta distribution, and μ is the mean. We have:

$$\begin{aligned}
\int_0^1 (x-\mu)^2 f(x;\alpha,\beta)dx &= \int_0^1 (x-\frac{\alpha}{\alpha+\beta})^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 (x-\frac{\alpha}{\alpha+\beta})^2 x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 (x^2-\frac{2\alpha}{\alpha+\beta}x+\frac{\alpha^2}{(\alpha+\beta)^2})x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1}dx - \frac{2\alpha}{\alpha+\beta}\int_0^1 \\
& x^{\alpha}(1-x)^{\beta-1}dx + \frac{\alpha^2}{(\alpha+\beta)^2}\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{2\alpha^2}{(\alpha+\beta)^2(\alpha+\beta+1)} + \\
& \frac{\alpha^2}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

where the third equality follows from a change of variables, and the fourth equality follows from a combination of integration by parts and the definition of the beta function.

0.10 Marginal Probability Density Function

Suppose we have a joint probability density function $f_{X,Y}(x,y)$ for two continuous random variables X and Y. The marginal probability density function of X is obtained by integrating $f_{X,Y}(x,y)$ over all possible values of Y:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \tag{57}$$

Similarly, the marginal probability density function of Y is obtained by integrating $f_{X,Y}(x,y)$ over all possible values of X:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \tag{58}$$

Proof:

To derive the formula for the marginal probability density function of X, we integrate $f_{X,Y}(x,y)$ over all possible values of Y:

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\
&= \lim_{b\to\infty} \int_{-b}^{b} f_{X,Y}(x,y)\,dy \qquad \text{(using symmetry of the integral)} \\
&= \lim_{b\to\infty} \int_{-b}^{b} \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\,dy \\
&= \lim_{b\to\infty} \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\,dy \\
&= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy
\end{aligned}$$

Similarly, to derive the formula for the marginal probability density function of Y, we integrate $f_{X,Y}(x,y)$ over all possible values of X.

0.11 Conditional Probability Density Function

The conditional probability density function of Y given $X = x$ is denoted by $f_{Y|X}(y|x)$ and is defined as:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{59}$$

Similarly, the conditional probability density function of X given $Y = y$ is denoted by $f_{X|Y}(x|y)$ and is defined as:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{60}$$

Proof:

To derive the formula for the conditional probability density function of Y given $X = x$, we start with the definition of conditional probability:

$$f_{Y|X}(y|x) = P(Y = y|X = x) \qquad = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

where the last equality follows from the definition of joint probability density function and marginal probability density function. Similarly, to derive the formula for the conditional probability density function of X given $Y = y$, we start with the definition of conditional probability:

$$f_{X|Y}(x|y) = P(X = x|Y = y) \qquad = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

where the last equality follows from the definition of joint probability density function and marginal probability density function.

0.12 Distribution Function

The joint distribution function of X and Y is denoted by $F_{X,Y}(x, y)$ and is defined as:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v), du, dv \tag{61}$$

The marginal distribution function of X is obtained by integrating $F_{X,Y}(x, y)$ over all possible values of Y:

$$F_X(x) = \int_{-\infty}^{\infty} F_{X,Y}(x, y), dy = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v), du, dv \tag{62}$$

Similarly, the marginal distribution function of Y is obtained by integrating $F_{X,Y}(x, y)$ over all possible values of X:

$$F_Y(y) = \int_{-\infty}^{\infty} F_{X,Y}(x, y), dx = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(u, v), du, dv \tag{63}$$

The conditional distribution function of Y given $X = x$ is denoted by $F_{Y|X}(y|x)$ and is defined as:

$$F_{Y|X}(y|x) = P(Y \leq y|X = x) = \int_{-\infty}^{y} \frac{f_{X,Y}(x, v)}{f_X(x)}, dv \tag{64}$$

Proof:

To derive the formula for the joint distribution function of X and Y, we use the definition of joint probability:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \qquad = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v), du, dv$$

where the last equality follows from the definition of double integral. The marginal distribution functions of X and Y can be obtained by integrating the joint distribution function over all possible values of the other variable. For example, to derive the formula for the marginal distribution function of X, we integrate $F_{X,Y}(x, y)$ over all possible values of Y:

$$F_X(x) = \int_{-\infty}^{\infty} F_{X,Y}(x, y), dy \quad = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{X,Y}(u, v), du, dv \ = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v), dv, du$$

where the last equality follows from the symmetry of the integral. Similarly, the formula for the marginal distribution function of Y can be obtained by integrating $F_{X,Y}(x, y)$ over all possible values of X.

Finally, to derive the formula for the conditional distribution function of Y given $X = x$, we start with the definition of conditional probability:

$$\begin{aligned} F_{Y|X}(y|x) &= P(Y \leq y|X = x) \\ &= \frac{P(X \leq x, Y \leq y)}{P(X = x)} \\ &= \frac{\int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v), du, dv}{\int_{-\infty}^{\infty} f_{X,Y}(x, v), dv} \\ &= \frac{\int_{-\infty}^{y} f_{X,Y}(x, v), dv}{f_X(x)} \\ &= \int_{-\infty}^{y} \frac{f_{X,Y}(x, v)}{f_X(x)}, dv \end{aligned}$$

where the last equality follows from the definition of conditional probability density function. This completes the proof.

0.13 Expected Value

Suppose we have n continuous random variables $X_1, X_2, \ldots, X_n$, with joint probability density function $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$. Let $g(X_1, X_2, \ldots, X_n)$ be a real-valued function of these random variables. Then, the expected value of $g(X_1, X_2, \ldots, X_n)$, denoted by

$E[g(X_1, X_2, \ldots, X_n)]$, is given by:

$$E[g(X_1, X_2, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \cdots dx_n \tag{65}$$

Proof:

The expected value of $g(X_1, X_2, \ldots, X_n)$ is given by:

$$\begin{aligned} E[g(X_1, X_2, \ldots, X_n)] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) \frac{f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)}{f_{X_1, X_2, \ldots, X_{n-1}}(x_1, x_2, \ldots, x_{n-1})} f_{X_1, X_2, \ldots, X_{n-1}}(x_1, x_2, \ldots \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) \frac{f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)}{f_{X_1, X_2, \ldots, X_{n-1}}(x_1, x_2, \ldots, x_{n-1})}, dx_1 dx_2 \cdot \right. \end{aligned}$$

where the last equality follows from the definition of joint probability density function. This completes the proof.

Part II

0.14

In this section, we look at applications of probability on a real life data set.

Table 1: 2000 to 2020 New Jersey General Election for United States House of Representatives: Voting Splits by Party

District	Avg. Democrat (%)	Dσ	Avg. Republican (%)	Rσ
1	66.2	3.8	33.8	3.8
2	41.7	9.7	58.3	9.7
3	58.1	7.3	41.9	7.3
4	60.8	6.3	39.2	6.3
5	59.6	6.9	40.4	6.9
6	48.6	7.4	51.4	7.4
7	55.5	7.1	44.5	7.1
8	64.9	5.7	35.1	5.7
9	67.1	5.7	32.9	5.7
10	56.3	7.1	43.7	7.1
11	53.8	7.5	46.2	7.5
12	63.9	6.1	36.1	6.1

Given the data in the table, the goal is to find the probability that Republicans win District 4 and Democrats win every other district. It is assumed that the percentage of votes for each party in each district follows a normal distribution. The CDF of a normal distribution with mean μ and standard deviation σ is given by:

$$\begin{aligned}\Phi(x) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}dt \\ &= \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right],\end{aligned}$$

where $\operatorname{erf}(x)$ is the error function.

To find the probability that Republicans win District 4, the CDF for the normal distribution is used:

$$P(R_4) = \Phi\left(\frac{39.2 - 60.8}{6.3}\right)$$

Next, the probability that Democrats win every other district needs to be calculated. For each district, the probability is calculated by finding the CDF of the normal distribution for the average Democrat percentage minus the average Republican percentage, divided by the standard deviation of the Democrat percentage:

$$P(D_i) = \Phi\left(\frac{\text{Avg. Democrat } (\%)_i - \text{Avg. Republican } (\%)_i}{D\sigma_i}\right)$$

Now, the probabilities for all districts can be multiplied together to find the desired probability:

$$P(R_4 \cap D_{1,2,3,5,6,7,8,9,10,11,12}) = P(R_4) \prod_{i \neq 4} P(D_i)$$

Here, μ_i and σ_i are the mean and standard deviation values for district i, respectively, and $\Phi(x)$ is the CDF of the standard normal distribution.

Assuming that the outcomes in each district are independent, the probability that Democrats win in all 12 districts in the same election can be calculated as the product of the probabilities of a Democratic win in each district. Mathematically, this can be written as:

$$P(\text{Democrats win all 12 districts}) = \prod_{i=1}^{12} P(D_i)$$

The mean and standard deviation values from the table are used to calculate the probability of a Democratic win in each district and then use the product formula above to calculate the probability that Democrats win in all 12 districts in the same election. Specifically:

$$P(\text{Democrats win all 12 districts}) = \prod_{i=1}^{12} \Phi\left(\frac{0.5 - \mu_i}{\sigma_i\sqrt{2}}\right)$$

Using my StatsLibrary java class to speed up the calculations, the probability that Democrats win in all 12 congressional districts in the same election is approximately 0.28%

In a similar manner, the probability of Republicans winning any single district and Democrats winning all 11 others in the same election can be calculated.

For each district i, calculate the probability of Republicans winning the district $P(R_i)$ and Democrats winning all other districts:

$$P(R_i \cap D_{j \neq i}) = P(R_i) \prod_{j \neq i} P(D_j)$$

Then, sum the probabilities for all districts:

$$P(\text{Republicans win any single district}) = \sum_{i=1}^{12} P(R_i \cap D_{j \neq i})$$

Using the previously calculated probabilities for each district, the sum can be calculated:

$$P(\text{Republicans win any single district}) \approx$$

$$\mathrm{P}(\mathrm{R}_1 \cap D_{2,3,4,5,6,7,8,9,10,11,12}) + P(R_2 \cap D_{1,3,4,5,6,7,8,9,10,11,12}) + \cdots + P(R_{12} \cap D_{1,2,3,4,5,6,7,8,9,10,11})$$

In 2018, Democrats won all but one district in New Jersey, with the voter splits provided in the table below.

Table 2: 2018 New Jersey General Election for United States House of Representatives: Voting Splits by Party

District	Democrat (%)	Republican (%)	Winner
1	64.4	33.2	D
2	52.3	46.4	D
3	50.1	48.7	D
4	55.8	43.3	D
5	56.7	42.6	D
6	63.7	35.8	D
7	51.7	47.4	D
8	77.2	22.4	D
9	71.3	28.4	D
10	87.2	12.6	D
11	56.3	42.3	D
12	66.6	31.9	D

Overall, the Democrats received 54.2% of the statewide popular vote, while the Republicans received 44.2%.

In the next few sections, we will examine how the concepts we learned in class can be applied to this data by proposing a question and using each probability distribution or theorem to set up and solve the problem, deriving meaningful information about the data.

Question: Given the historical voting data, what is the likelihood of the Democratic party winning a majority of the districts in the next election?

To answer this question, we can employ various probability distributions and theorems.

0.14.1 Binomial Distribution

Using the Binomial distribution, we estimate the probability of Democrats winning a majority of the districts. We calculate the probability of Democrats winning at least 7 out of 12 districts.

Let X be the number of districts won by the Democratic party. We assume that the probability of winning an election in each district remains constant and equal to the historical average. For simplicity, we round the average percentage of votes to the nearest integer. The probability mass function of the Binomial distribution is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where $n = 12$ (total number of districts), k is the number of districts won by the Democratic party, and p is the probability of winning a district, given by the historical average.

To find the probability of Democrats winning at least 7 districts, we calculate the sum of probabilities for $k = 7, 8, \ldots, 12$:

$$P(X \geq 7) = \sum_{k=7}^{12} \binom{12}{k} p^k (1-p)^{12-k}$$

Calculating this for each district and summing the probabilities gives us the likelihood of the Democratic party winning a majority of the districts.

0.14.2 Bayes' Theorem

We can use Bayes' theorem to update our beliefs about the probability of the Democratic party winning a majority of districts based on the 2018 election results.

Let A represent the event of Democrats winning a majority of districts, and B represent the 2018 election results. We are interested in finding $P(A|B)$. According to Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

We estimate $P(A)$ using the historical data, and $P(B|A)$ as the likelihood of observing the 2018 results, given that Democrats win a majority of districts. We can estimate $P(B)$ by considering all possible outcomes.

Computing $P(A|B)$ gives us an updated probability of the Democratic party winning a majority of districts based on the 2018 election results.

0.14.3 Hypergeometric Distribution

Assuming there is a fixed number of upcoming elections, we can use the Hypergeometric distribution to estimate the probability of Democrats winning a majority of districts.

Let N be the total number of upcoming elections, K be the number of elections where the Democratic party has a historical advantage (i.e., more than 50

The probability mass function of the Hypergeometric distribution is given by:

$$P(X = n) = \frac{\binom{K}{n}\binom{N-K}{n-N}}{\binom{N}{n}}$$

To find the probability of Democrats winning a majority of districts, we calculate the sum of probabilities for $n \geq \frac{N}{2}$:

$$P(X \geq \frac{N}{2}) = \sum_{n=\frac{N}{2}}^{K} \frac{\binom{K}{n}\binom{N-K}{n-N}}{\binom{N}{n}}$$

Calculating this for the given data gives us the likelihood of the Democratic party winning a majority of districts in a fixed number of upcoming elections.

0.14.4 Poisson Distribution

We can use the Poisson distribution to estimate the number of districts the Democratic party would win in the next election, assuming a constant winning rate. Let λ be the average number of districts won by the Democratic party, calculated using the historical data. The probability mass function of the Poisson distribution is given by:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

To find the probability of Democrats winning a majority of districts, we calculate the sum of probabilities for $k \geq 7$:

$$P(X \geq 7) = \sum_{k=7}^{12} \frac{e^{-\lambda}\lambda^k}{k!}$$

Calculating this for the given data gives us the likelihood of the Democratic party winning a majority of districts, assuming a constant winning rate.

0.14.5 Tchebychev's Theorem

Using Tchebychev's theorem, we can estimate the likelihood of election outcomes that deviate significantly from the historical averages. Let μ be the historical average percentage of votes for the Democratic party, and σ be the standard deviation. Tchebychev's theorem states that:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

To find the probability of Democrats winning a majority of districts, we can set k such that $\mu + k\sigma \geq 50\%$. Then, we calculate the upper bound for the probability of winning a majority of districts using Tchebychev's theorem:

$$P(X \geq 50\%) \leq \frac{1}{k^2}$$

Calculating this for the given data gives us an upper bound on the likelihood of the Democratic party winning a majority of districts, considering significant deviations from the historical averages.

Part III

Data Processing and Plotting

0.15 The Apache Common Mathematics Library

The ApacheSalter Class

The ApacheSalter class utilizes the Apache Commons Math library to generate salt values from a normal distribution based on the input x-coordinates. A NormalDistribution object is created with a default mean of 0 and standard deviation of 1. The DescriptiveStatistics class calculates the mean and standard deviation of the input x-coordinates. For each element in the xCoords array, a random salt value is generated from a normal distribution with the same mean and standard deviation using normalDistribution.inverseCumulativeProbability(1 - (1 - CONFIDENCE LEVEL) / 2) * std + mean. The inverseCumulativeProbability method returns a random value from the normal distribution with the specified cumulative probability, ensuring a value within a 99

0.15.1 The ApacheSmoother Class

The Smoother class provides two main methods: smooth and savitzkyGolayCoefficients. The smooth method accepts two arrays x and y representing the x-coordinates and y-values of a dataset, a filter window size windowSize, and a polynomial order polynomialOrder for the Savitzky-Golay filter. The method checks for input validity, creates a new array smoothedY to store the smoothed y-values, and calculates the Savitzky-Golay coefficients using the savitzkyGolayCoefficients method. The filter is applied by looping through each element of y, starting from the $(windowSize/2)^{th}$ element, and calculating the smoothed y-value by taking the weighted average of the y-values in the filter window using the calculated coefficients. The smoothed y-value is stored in the corresponding index of smoothedY. Finally, the method fills in the edges of the data with the original y-values and returns the smoothed y-values in the smoothedY array.

The savitzkyGolayCoefficients method accepts a filter window size windowSize and a polynomial order polynomialOrder for the Savitzky-Golay filter. The method checks that windowSize is odd and throws an IllegalArgumentException if not. It creates a Vandermonde matrix based on the window size and polynomial order and computes the pseudoinverse of the matrix. The first row of the pseudoinverse is returned as the Savitzky-Golay coefficients in an array.

0.15.2 The ApachePlotter Class

The Plotter class includes an interpolate method that generates an array of (x, y) pairs for a function using cubic spline interpolation. This method takes a PolynomialSplineFunction as input and returns an array of (x, y) pairs for the function. A new method, newInterpolationPlot, calls interpolate and writeToFile to generate the data and write it to a CSV file with the specified filename.

The cubic spline interpolation is implemented using the SplineInterpolator class from the Apache Commons Math library. The interpolate method accepts an array of (x, y) pairs and converts them into separate arrays of x and y values, which are then passed into the

SplineInterpolator object to generate a PolynomialSplineFunction. The interpolate method returns this PolynomialSplineFunction.

0.16 Working with Octave

This program generates a sinusoidal function with 100 data points, salts the y-values with random noise, and smooths the salted y-values using a moving average filter. The original points, salted points, and smoothed points are then plotted.

Generating the Sinusoidal Function

The program generates the x and y ordered pairs for the sinusoidal function using the linspace and sin functions:

```
x = linspace(0, 2*pi, 100);
y = sin(x);
```

The linspace function creates an evenly spaced vector of 100 points between 0 and 2π, which is used as the x-values. The sin function is then applied to the x-values to generate the corresponding y-values.

Salting the Y-Values

The program salts the y-values by adding random noise to the original y-values using the rand function and the specified salt range:

```
salt_range = 0.2;
salted_y = y + salt_range * rand(size(y)) - salt_range/2;
```

The rand function generates a vector of random values between 0 and 1 with the same size as the y-values. The salt range is then applied to the random values, centered around 0, and added to the original y-values.

Smoothing the Salted Y-Values

The program smooths the salted y-values using a moving average filter of size 5, which is implemented using the filter function with a numerator of ones(1, window size)/window size and a denominator of 1:

```
window_size = 5;
smoothed_y = filter(ones(1, window_size)/window_size, 1, salted_y);
```

The ones function creates a vector of ones with a length of window size, which is used as the numerator of the filter. The denominator is set to 1, indicating that there is no feedback in the filter. The filter function applies the filter to the salted y-values to generate the smoothed y-values.

Plotting the Results

Finally, the program plots the original points, salted points, and smoothed points using the subplot function to create a 3-by-1 grid of plots:

```
subplot(3, 1, 1);
plot(x, y, '-');
title('Original Points');
xlabel('x');
ylabel('y');

subplot(3, 1, 2);
plot(x, salted_y, '-');
title(sprintf('Salted Points (range = %f)', salt_range));
xlabel('x');
ylabel('y');

subplot(3, 1, 3);
plot(x, smoothed_y, '-');
title(sprintf('Smoothed Points (window size = %d)', window_size));
xlabel('x');
ylabel('y');
```

The subplot function creates a 3-by-1 grid of plots, where the first subplot shows the original points, the second subplot shows the salted points, and the third subplot shows the smoothed points. The plot function is used to create line plots of the x-values and the corresponding y-values, and the title, xlabel, and ylabel functions are used to label the plots.

0.17 Key Differences and Features

Java and Octave are programming languages that exhibit distinct technical characteristics, which significantly impact their respective capacities for handling mathematical computations and data processing tasks. To better appreciate these intricacies, a deeper exploration of their specific attributes and their effects on aptitude for these tasks is necessary.

Java, a versatile general-purpose language, is renowned for its utility in enterprise and web development. It employs a statically typed system, where variables are assigned types at compile time, enhancing code safety and portability. This feature minimizes potential errors that may arise during execution. Java's extensive ecosystem of libraries, such as the Apache Commons Math library, provides developers with a wide range of mathematical functions and tools, further extending its capabilities.

In contrast, Octave, a high-level language specifically designed for numerical computing and data analysis, functions as an interpreted language, eliminating the need for compilation prior to code execution. Its dynamically typed nature allows variables to change types at runtime, offering flexibility that Java cannot provide. Octave's built-in package system includes a vast array of mathematical and data processing functions, underscoring its suitability for these particular applications.

Java's robust type system and safety features contribute to its reliability and robustness, effectively mitigating errors and facilitating code portability across various platforms. However, these advantages come at the cost of increased code complexity and decreased ease of use, particularly for tasks involving extensive mathematical computations. Additionally, the need for a compiler in Java's development process can reduce efficiency and complicate the handling of large datasets.

On the other hand, Octave's specialized focus on mathematical computations and data analysis results in a user-friendly experience for these specific tasks, owing to its dynamic typing and interactive environment. However, Octave may not perform as optimally as Java for certain operations requiring low-level optimizations. Furthermore, Octave's interpreted nature may lead to slower performance compared to Java when faced with large-scale computations.

In summary, Java and Octave each possess unique strengths and weaknesses in the context of mathematical and data processing applications. Java's strong type system and safety features make it ideal for enterprise and web development, whereas Octave's dynamic typing and comprehensive set of built-in functions cater to scientific computing and data analysis. The optimal choice of programming language depends on the specific requirements and objectives of the project, necessitating a thorough assessment of each language's capabilities and limitations.

Part IV

Sources

0.18 Data Structures and Algorithms

Aggarwal, C. C. (2016). Hashing: Fundamentals and Applications. Springer International Publishing.

Compiler Explorer. (n.d.). Compiler Explorer. https://godbolt.org/

Cormen, T. (2011, September 8). Red-black trees [Video]. Introduction to Algorithms. Massachusetts Institute of Technology. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/video-lectures/lecture-10-red-black-trees-rotations-insertions-deletions/

Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C. (2009). Introduction to Algorithms (3rd ed.). MIT Press.

Google. (2023). FarmHash source code. In GitHub. https://github.com/google/farmhash/tree/master/

Tim Roughgarden. (2018, January 9). Hash Tables [Video]. Algorithms Specialization. Stanford University. https://www.coursera.org/lecture/algorithms-divide-conquer/hash-tables-C3qUt

Wikipedia. (2023, February 21). MurmurHash. In Wikipedia. https://en.wikipedia.org/wiki/MurmurH

0.19 Design Patterns

Freeman, E., Freeman, E. (2004). Head First Design Patterns. O'Reilly Media, Inc.

University of Waterloo. (n.d.). Software Design Patterns. Retrieved from https://www.youtube.com/pl

0.20 Math Formulas and Proofs

Wackerly, D. D., Mendenhall, W., Scheaffer, R. L. (2008). Mathematical statistics with applications (7th ed.). Brooks/Cole, Cengage Learning.