Module 4

Estimation vs. Inference

- Estimation provides a value for a parameter, such as the mean, calculated from sample data.
- Inference aims to understand the population distribution, including parameters like the standard error.

Parametric vs. Non-Parametric Approaches

- Parametric methods assume a specific distribution for the data, while non-parametric methods do not.
- The choice between these approaches depends on the data characteristics and the analysis goals.

Frequentist vs. Bayesian Statistics

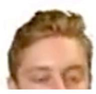
- Frequentist statistics focuses on long-run frequencies of events, while Bayesian statistics incorporates prior beliefs and updates them with new data.
- Understanding these differences is crucial for selecting appropriate statistical methods in machine learning applications.

Estimation: is the application of an algorithm, for example taking an average:

$$\overline{X} = \sum_{i=1}^{N} x_i / n$$

Inference: involves putting an accuracy on the estimate (e.g. standard error of an average):

$$[\sum_{i=1}^{N} \frac{(x_i - \overline{x})^2}{(n-1)}]^{1/2}$$



Customer Churn Data Overview

- The dataset includes customer characteristics, account types, revenue, satisfaction scores, and churn outcomes.

- Churn types are categorized into actively canceling and not renewing subscriptions.

Data Visualization Techniques

- A bar plot is created to show churn likelihood based on payment types, revealing that credit card users are less likely to churn.
- A categorical variable is generated to analyze churn by the number of months customers have been with the service.

Exploratory Data Analysis (EDA)

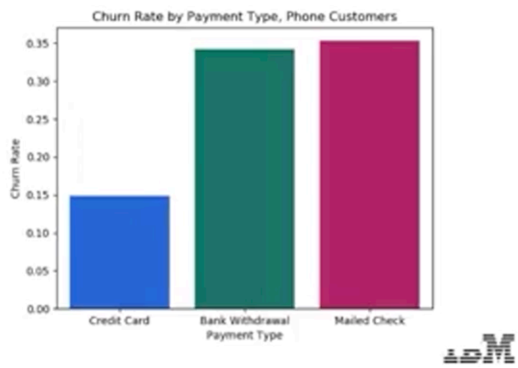
- A pair plot is used to examine relationships between customer tenure, gigabyte usage, total revenue, and customer lifetime value, distinguishing between churned and non-churned customers.
- A hexbin plot visualizes the relationship between customer tenure and monthly charges, indicating trends in customer retention based on these factors.

The session concludes with a preview of the next topic on parametric versus non-parametric statistics.

Code

```
# Examining churn data, churn value by
# payment type
sns.barplot(y='churn_value', x='payment',
            data=df_phone, ci=None)
plt.ylabel('Churn Rate')
plt.xlabel('Payment Type')
plt.title('Churn Rate by Payment Type,
          Phone Customers')
```

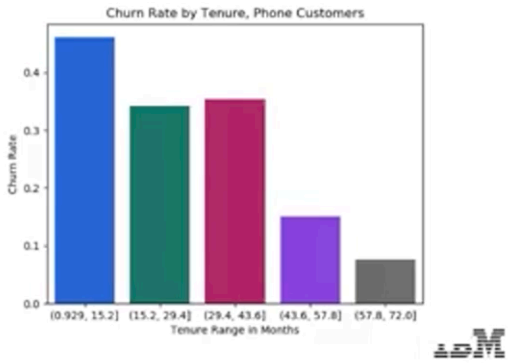
Output



Code

```
# Examining churn data, this time by
# tenure
sns.barplot(y='churn_value',
            x=pd.cut(df_phone.months,
                     bins=5),
           data=df_phone, ci=None)
plt.ylabel('Churn Rate')
plt.xlabel('Tenure Range in Months')
plt.title('Churn Rate by Tenure,
          Phone Customers')
```

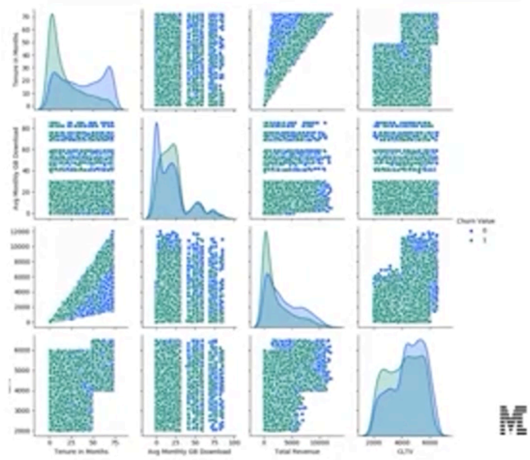
Output



Code

```
# Seaborn plot, feature correlations
pairplot = df_phone[['months',
                     'gb_mon',
                     'total_revenue',
                     'cltv',
                     'churn_value']]
sns.pairplot(pairplot,
             hue='churn_value')
```

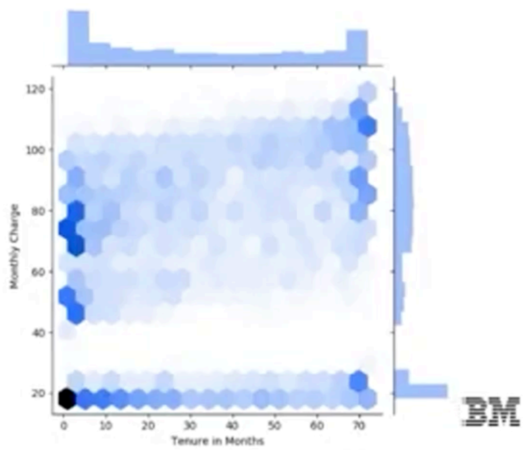
Output



Code

```
# Seaborn hexbin plot
sns.jointplot(x=df[labels['months']],
              y=df[labels['monthly']],
              kind='hex')
```

Output



Parametric Models

- Defined by a finite number of parameters and rely on strict assumptions about the data distribution.
- Examples include linear models and the normal distribution, which is characterized by mean and standard deviation.

Non-Parametric Models

- Do not rely on strict assumptions about data distribution, allowing for more flexibility.
- An example is using histograms to create a distribution based on actual data without assuming a specific distribution.

Common Distributions

- **Normal Distribution**: Most values cluster around the mean, with the central limit theorem stating that averages of samples will form a normal distribution.
- **Log-Normal Distribution**: Results from taking the logarithm of a variable, often seen in financial data.
- **Exponential Distribution**: Models the time until the next event occurs, with most values clustering at the lower end.
- **Poisson Distribution**: Represents the number of events occurring in a fixed interval, characterized by the average rate (lambda).

The **likelihood function** is related to probability and is a function of the **parameters** of the model:

$$\underbrace{\boxed{\mathcal{L}_n(\theta)}}_{\textbf{Function of the parameters}} = \Pi_{i=1}^{n} f(X_i, \theta)$$

Uniform Distribution

- Each value within a range has an equal chance of occurring, similar to rolling a fair die.
- All outcomes are equally likely, making it a straightforward distribution.

Normal (Gaussian) Distribution

- Values cluster around the mean, with extremes being less likely; the shape is bell-curved.
- The central limit theorem states that the distribution of sample averages approaches a normal distribution as sample size increases.

Log-Normal Distribution

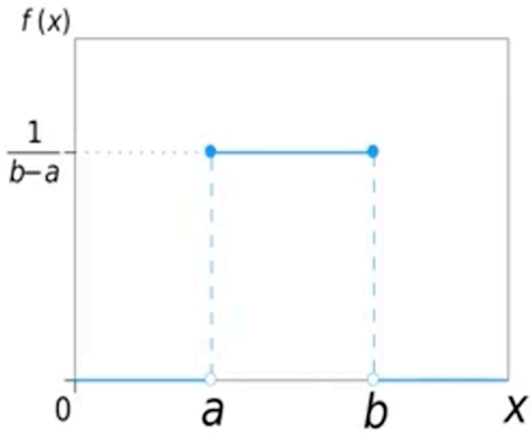
- If the logarithm of a variable is normally distributed, the original variable follows a log-normal distribution.
- Commonly seen in financial data, where most values cluster around a median with a long tail of outliers.

Exponential Distribution

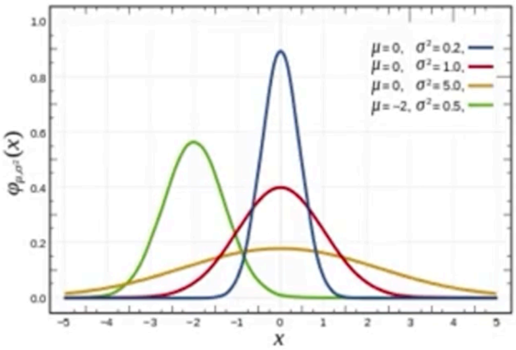
- Most values are concentrated on the left, representing the time until the next event occurs.
- Often used in scenarios like waiting times, where shorter intervals are more common.

Poisson Distribution

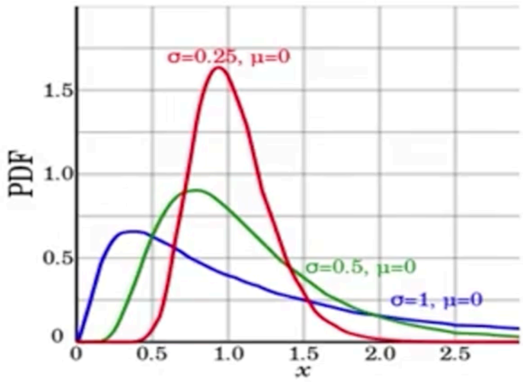
- Models the number of events occurring in a fixed interval of time, characterized by the average rate (lambda).
- Useful for predicting occurrences, such as the number of viewers in a given time frame.



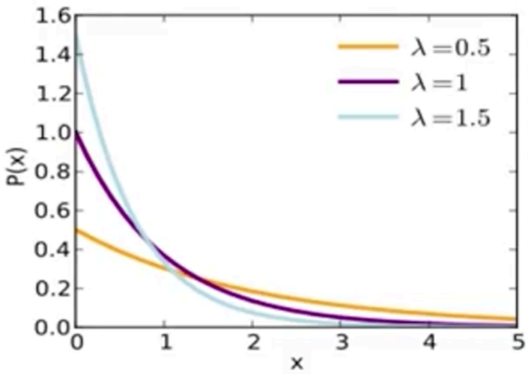
Uniform



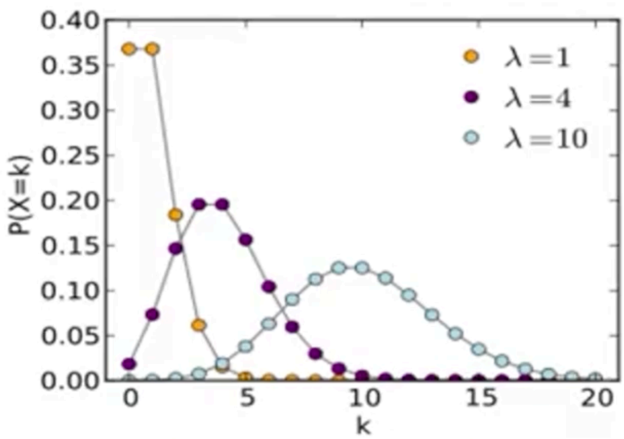
Gaussian / Normal



Log Normal



Exponential



Poisson

Understanding Hypothesis Testing

- A hypothesis is a statement about a population parameter, such as the mean in a given scenario.
- Two hypotheses are created: the null hypothesis (H0) and the alternative hypothesis (H1), with the null typically being a specific value.

Procedure of Hypothesis Testing

- Data from a sample is used to determine whether to accept the null hypothesis or reject it in favor of the alternative.
- It is common to say that you can reject the null hypothesis but never accept the alternative; instead, you proceed with the assumption that the alternative is true based on the test statistic.

Bayesian Approach to Hypothesis Testing

- In Bayesian interpretation, rather than establishing a decision boundary, posterior probabilities of both hypotheses are calculated to assess which is more likely.

Understanding Priors

- The prior probabilities for two hypotheses (fair coin vs. unfair coin) are set at 50/50, reflecting no initial preference.

- In real-world scenarios, prior knowledge may lead to a higher likelihood of selecting a fair coin.

Using Priors in Bayesian Analysis

- The posterior distribution is derived from the prior distribution and the likelihood of observed data.
- The likelihood ratio compares the probability of data under each hypothesis, adjusting the priors based on observed outcomes.

Key Takeaways

- The section covers the basics of hypothesis testing, emphasizing the role of prior distributions and likelihood ratios in Bayesian analysis.
- A coin toss example illustrates how Bayesian hypothesis testing operates, setting the stage for further exploration of frequentist approaches in the next section.

Priors: $P(H_1) = 1/2 = P(H_2) = 1/2$

Updating priors after seeing the data 3 heads (Bayes' Rule):

$$P(H_1|x) = \frac{P(x|H_1)P(H_1)}{P(x)}$$

$$\frac{P(H_1|x)}{P(H_2|x)} = \frac{P(H_1)P(x|H_1)}{P(H_2)P(x|H_2)}$$

Correlation vs. Causation

- Correlation indicates a relationship between two variables but does not imply that one causes the other.
- Understanding the underlying mechanisms is essential to avoid misinterpretation of data.

Role of Confounding Variables

- Confounding variables can influence both X and Y, leading to misleading correlations.
- Examples include the relationship between ice cream sales and drownings, both affected by temperature.

Spurious Correlations

- Spurious correlations occur when two variables appear related due to coincidence rather than a causal link.
- Examples include the correlation between the age of Miss America and murders by steam, which lacks a meaningful connection.