# STAT7630: Bayesian Statistics
# Lecture Slides # 14

Naive Bayes Classification
Chapter 14

Elvan Ceyhan
Department of Mathematics & Statistics
Auburn University

Fall 2024,
Updated: November, 2024

# Outline

# Goal of Classification

- The primary objective of classification is to predict the class membership of a categorical response variable $Y$ using a set of predictor variables $(X_1, X_2, \ldots, X_p)$.
- In logistic regression, this is achieved by modeling $Y$ as a binary response (e.g., $Y = 1$ or $Y = 0$) and classifying new observations based on their predictor values.
- Logistic regression, however, is limited to binary outcomes.
- For data sets where the response variable $Y$ contains more than two categories, more general classification methods are required to handle multiclass scenarios.

# Example of a Multicategory Response $Y$

- Consider a dataset containing three species of Antarctic penguins: *Adelie*, *Chinstrap*, and *Gentoo*.
- The classification objective is to assign a given penguin observation to one of these species using the following predictor variables:
  - $X_1$: Weight (binary, 1 if above average, 0 if below average)
  - $X_2$: Bill length (measured in mm)
  - $X_3$: Flipper length (measured in mm)
- The dataset (`penguins_bayes`) contains measurements for 344 penguins with known species labels:
  - 152 *Adelie*, 68 *Chinstrap*, and 124 *Gentoo*.

# Penguin Images

**Figure 1:** Adelie, Chinstrap, and Gentoo penguins.

# Possible Prior Specifications

- **Empirical Prior:** Assume the observed sample proportions represent the true population proportions.
    - This is a commonly used approach due to its data-driven nature.
- **Subjective Prior:** Specify prior probabilities based on expert knowledge or external information.
    - Useful when domain knowledge suggests deviations from sample proportions.
- **Noninformative Prior:** Assign equal prior probabilities to all classes.
    - Suitable only if the population proportions are expected to be roughly equal across categories.
    - May lead to suboptimal results when the true category proportions differ significantly.

# Naive Bayes Classification Compared to Logistic Regression

- **Logistic Regression:**
  - Effectively classifies binary response variables ($Y \in \{0, 1\}$).
  - Relies on a parametric model for the relationship between predictors and the log-odds of the response.
- **Naive Bayes Classification:**
  - Handles categorical response variables $Y$ with two or more categories seamlessly.
  - **Simplicity:** Based primarily on Bayes' Rule with minimal theoretical complexity.
  - **Computational efficiency:** Does not require iterative procedures like MCMC simulation.

# Outline

Naive Bayes Classification (NBC)

- NBC with One Categorical Predictor
- NBC with One Continuous Predictor
- NBC with Two Continuous Predictors

# Example of Naive Bayes Classification with One Categorical Predictor

- Use the categorical predictor *"above average weight"* ($X_1$) to classify a new penguin into one of three species.
- Dataset includes 342 penguins after excluding two with missing predictor values.

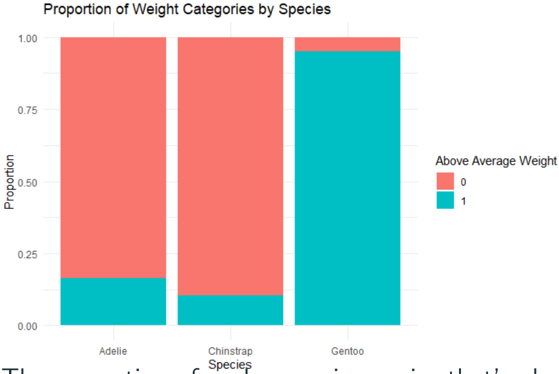


**Figure 2:** The proportion of each penguin species that's above average weight.

# Example of Naive Bayes Classification with One Categorical Predictor

- **Preliminary Observation:** The above bar plot reveals that the most likely species for below-average weight ($X_1 = 0$) is *Chinstrap*.
- **Question:** Should we classify any penguin with $X_1 = 0$ as *Chinstrap*?
- **Caution:**
  - Despite the bar plot, *Chinstrap* is the rarest species overall in the population.
  - Prior probabilities must be carefully considered before making a final classification.

# Bayes' Rule for Classification with One Categorical Predictor

**Bayes' Rule:** The probability that a categorical response takes value $y^*$, given a particular value of the categorical predictor $X_1$, is computed as:

$$p(y^* \mid x_1) = \frac{\text{prior} \times \text{likelihood}}{\text{normalizing constant}} = \frac{p(y^*)L(y^* \mid x_1)}{p(x_1)}$$

**Normalizing Constant:** The denominator $p(x_1)$ is given by:

$$p(x_1) = \sum_{\text{all } y} p(y)L(y \mid x_1)$$

**Expanded Form:**

$$p(x_1) = p(y = A)L(y = A \mid x_1) + p(y = C)L(y = C \mid x_1) + p(y = G)L(y = G \mid x_1)$$

**Interpretation:**

- $p(y^*)$: Prior probability of the class $y^*$.
- $L(y^* \mid x_1)$: Likelihood of observing $x_1$ given class $y^*$.
- $p(x_1)$: Overall probability of observing $x_1$ across all classes.

# Examples of Calculations for Naive Bayes Classification

## Example: Probability Calculation for Adelie Penguins

- Below is a table from R output with counts broken down by species and weight category.
- For a penguin with below average weight ($X_1 = 0$), the probability that it is an Adelie ($y = A$) is:

$$p(y = A \mid x_1 = 0) = \frac{126}{193} \approx 0.6528$$

**Table 1:** Species Counts by Group

| Species \ $X_1$ | 0 | 1 | Total |
|---|---|---|---|
| Adelie | 126 | 25 | 151 |
| Chinstrap | 61 | 7 | 68 |
| Gentoo | 6 | 117 | 123 |
| Total | 193 | 149 | 342 |

# Examples of Calculations (continued)

**Bayes' Rule Components:**

$$p(y = A) = \frac{151}{342}, \quad p(y = C) = \frac{68}{342}, \quad p(y = G) = \frac{123}{342}$$

$$L(y = A \mid x_1 = 0) = \frac{126}{151} \approx 0.8344,$$
$$L(y = C \mid x_1 = 0) = \frac{61}{68} \approx 0.8971$$
$$L(y = G \mid x_1 = 0) = \frac{6}{123} \approx 0.0488$$

**Normalizing Constant:**

$$p(x_1 = 0) = \frac{151}{342} \cdot \frac{126}{151} + \frac{68}{342} \cdot \frac{61}{68} + \frac{123}{342} \cdot \frac{6}{123} = \frac{193}{342}$$

# Examples of Calculations (continued)

**Posterior Probability for Adelie Penguins:**

$$p(y = A \mid x_1 = 0) = \frac{p(y = A) \cdot L(y = A \mid x_1 = 0)}{p(x_1 = 0)} = \frac{\left(\frac{151}{342}\right) \times \left(\frac{126}{151}\right)}{\frac{193}{342}} \approx 0.6528$$

**Conclusion:** These calculations confirm Bayes' Rule:

$$p(y^* \mid x_1) = \frac{p(y^*)L(y^* \mid x_1)}{p(x_1)}$$

**Other Posterior Probabilities:**

$$p(y = C \mid x_1 = 0) \approx 0.3161$$

$$p(y = G \mid x_1 = 0) \approx 0.0311$$

**Conclusion:** These results illustrate the application of Bayes' Rule for calculating posterior probabilities using a categorical predictor.

# Conclusions

- **Highest Posterior Probability:** The category with the highest posterior probability is **"Adelie"**.
- Even though the proportion of Chinstraps below average weight exceeds that of Adelies, the prevalence of Adelies in the population makes it more likely that a random below-average-weight penguin is an Adelie.
- This outcome reflects the prior probabilities set to match species proportions in the sample:

$$p(y = A),\ p(y = C),\ p(y = G).$$

- **Alternative Priors:** Using different priors, such as $p(y = A) = p(y = C) = p(y = G) = \frac{1}{3}$, would yield different posterior probabilities.
- The current approach, using sample proportions as priors, is likely the most reasonable choice for this analysis.

# Outline

# Example of Naive Bayes Classification with One Continuous Predictor

- Now consider classification based on a **continuous predictor**.
- For instance, let $X_2 =$**bill length (in mm)** be the predictor used to classify penguin species.
- Suppose an observed penguin has a **bill length of 50 mm**.
- **Observation:** The below plot indicates that this bill length would be *extremely uncommon* for an Adelie.

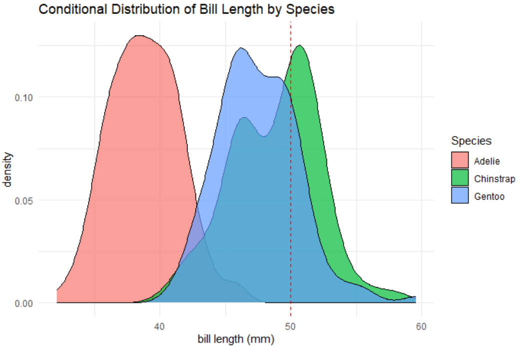


**Figure 3:** Density plots of the bill lengths (mm) observed among three penguin species.

# Naive Bayes Classification with One Continuous Predictor

- For a **continuous predictor**, the Naive Bayes approach assumes:

$$X_2|(Y = A) \sim N(\mu_A, \sigma_A^2),$$
$$X_2|(Y = C) \sim N(\mu_C, \sigma_C^2),$$
$$X_2|(Y = G) \sim N(\mu_G, \sigma_G^2)$$

- This assumption implies the predictor follows a separate **conditional normal distribution** for each response category.
- While this approach is somewhat restrictive, it is appropriate here, as suggested by the estimated density plots for **bill length**.
- The **means** and **variances** of these normal distributions are typically set to the **sample means and variances** for each species in the data.

# Using Bayes' Rule to Get Posterior Probabilities for Each Category

- To calculate the posterior probability of an observation belonging to a category $y^*$, we use Bayes' Rule:

$$p(y^*|x_2) = \frac{p(y^*)L(y^*|x_2)}{p(x_2)} = \frac{p(y^*)L(y^*|x_2)}{\sum_{\text{all } y} p(y)L(y|x_2)}$$

- Example calculations for $x_2 = 50$ mm (see `R` code for normal density values):

$$p(x_2 = 50) = \frac{151}{342} \cdot 0.0000212 + \frac{68}{342} \cdot 0.112 + \frac{123}{342} \cdot 0.09317 \approx 0.05579$$

- **Posterior probabilities:**

$$p(y = A|x_2 = 50) = \frac{\frac{151}{342} \cdot 0.0000212}{0.05579} \approx 0.0002$$

$$p(y = C|x_2 = 50) \approx 0.3992, \quad p(y = G|x_2 = 50) \approx 0.600$$

- **Conclusion:** The observation is most likely to belong to the category $G$ (Gentoo).

## Conclusions

- For a penguin with a bill length of 50 mm, the category with the highest posterior probability is **“Gentoo”**.
- The predominance of Gentoos in the population contributes to their higher posterior probability, even though bill length values of 50 mm are less common among Gentoos compared to Chinstraps.
- This highlights the importance of incorporating prior probabilities in Bayesian classification to account for population-level proportions.

# Outline

# Naive Bayes Classification with Two Continuous Predictors

- The Naive Bayes Classification framework can accommodate multiple predictors.
- For the penguin example, incorporating both $X_2 =$ *bill length* and $X_3 =$ *flipper length* may improve classification accuracy.

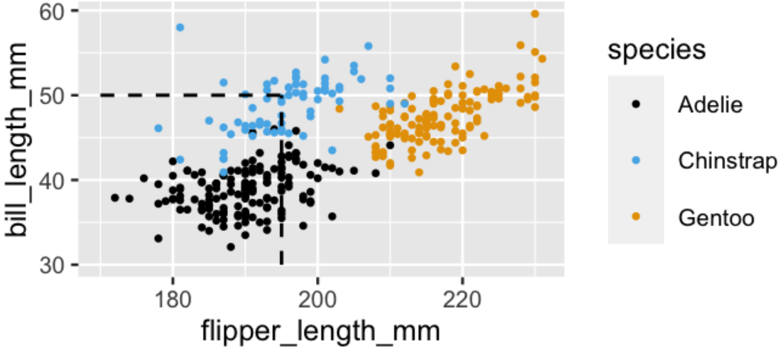


**Figure 4:** Density plots of the bill lengths (mm) and flipper lengths (mm) among our three penguin species.

# Naive Bayes Classification with Two Continuous Predictors

- Using Bayes' Rule, the likelihood component $L(y|x_2, x_3)$ is simplified with the naive assumption of (conditional) independence:

$$L(y|x_2, x_3) = f(x_2, x_3|y) = f(x_2|y)f(x_3|y).$$

- This assumption, however, may not hold in practice. For example, in the penguin dataset, $X_2$ and $X_3$ exhibit a **positive association** (as seen in the scatterplot), indicating dependence.

# Calculations for Naive Bayes Classification with Two Continuous Predictors

- Consider a new penguin with bill length $X_2 = 50$ and flipper length $X_3 = 195$.
- Compute the posterior probabilities for each category:

$$p(y = A)L(y = A|x_2 = 50, x_3 = 195) = \frac{151}{342} \cdot 0.0000212 \cdot 0.04554$$

$$p(y = C)L(y = C|x_2 = 50, x_3 = 195) = \frac{68}{342} \cdot 0.112 \cdot 0.05541$$

$$p(y = G)L(y = G|x_2 = 50, x_3 = 195) = \frac{123}{342} \cdot 0.09317 \cdot 0.0001934$$

- Compute the normalizing constant:

$$\sum_{\text{all } y} p(y)L(y|x_2 = 50, x_3 = 195) \approx 0.001241$$

# Continued Calculations for Naive Bayes Classification

- Posterior probability for $y = A$:

$$p(y = A|x_2 = 50, x_3 = 195) = \frac{\frac{151}{342} \cdot 0.0000212 \cdot 0.04554}{0.001241} \approx 0.0003$$

- Similarly, compute the posterior probabilities for the remaining categories:

$$p(y = C|x_2 = 50, x_3 = 195) \approx 0.9944$$

$$p(y = G|x_2 = 50, x_3 = 195) \approx 0.0052$$

- **Observations:**
  - The category with the highest posterior probability is $y = C$ (Chinstrap).
  - The classification reflects the contribution of both predictors, $X_2$ (bill length) and $X_3$ (flipper length), weighted by their respective likelihoods and prior probabilities.

# Conclusions

- This penguin is almost certainly classified as a **Chinstrap**.
- The combination of **bill length** and **flipper length** aligns strongly with the characteristics of Chinstrap penguins for this set of variables.
- The Naive Bayes classifier effectively integrates multiple predictors to enhance classification accuracy, even under the assumption of independence.

## Doing It the Easy Way: The `naiveBayes` Function

- To streamline the process and avoid tedious calculations, we can leverage the `naiveBayes` function from the `e1071` package in R.
- This function:
  - Automatically computes prior category probabilities based on observed category proportions in the sample (the preferred approach).
  - Efficiently predicts the class of a "new" observation with specified predictor values.
- See the `R` example for practical implementation and predictions for new penguin data.

# Assessing the Performance of Naive Bayes Classification

- Tools for evaluating classification accuracy are similar to those covered in Chapter 13:
  - **Confusion Matrix:** Provides a summary of prediction outcomes (e.g., true positives, false positives).
  - **Cross-Validation:** Offers robust estimates of classification accuracy by splitting the data into training and validation sets.
- For multiple potential predictor variables:
  - Develop several classification models.
  - Compare performance using confusion matrices and cross-validation metrics.
- **Practical Implementation:** See `R` examples for applying these techniques on the penguins dataset.

# Naive Bayes vs. Logistic Regression

- For **categorical responses with more than two categories**:
  - Logistic regression is not applicable.
  - Other generalized linear models exist but are beyond the scope of this class.
- When the response is **binary** (two categories):
  - **Advantages of Logistic Regression:**
    - Provides insights via regression coefficients about the relationship between the response and predictors.
  - **Simplifying Assumptions of Naive Bayes:**
    - Predictors are normally distributed.
    - Predictors are independent of one another.
    - These assumptions may not hold in reality.
- Knowing both tools equips us to handle various classification scenarios effectively.