Quantitative Data Reasoning

Learning Objective

Probability – **Quantifying** Uncertainty

1. Understand probability concepts and their banking applications
2. Calculate and interpret probabilities for business decisions
3. Apply conditional probability and Bayes' theorem to real scenarios
4. Use probability to assess risk and make predictions
5. Build probability-based business models for decision-making

👀 **Our Focus:** Transform uncertainty into quantified risk assessments that drive intelligent business decisions.

💡 **Key Question:** "How can we measure the likelihood of business outcomes and use this knowledge to make better strategic decisions?"

Probability

From Uncertainty to Calculated Risk

The Banking Reality:

- **Every decision involves uncertainty** - Will customers default? Will campaigns succeed?
- **Risk is the core of banking** - Lending, investing, and strategic planning all require risk assessment
- **Probability quantifies uncertainty** - Transform "maybe" into "65% chance"
- **Better predictions = better profits** - Accurate risk assessment drives competitive advantage

Banking Applications of Probability:

🎯 **Marketing Campaigns:** What's the probability a customer will respond?

💰 **Credit Decisions:** What's the likelihood of loan default?

📊 **Product Development:** What's the chance of market acceptance?

🏢 **Branch Planning:** What's the probability of seasonal demand changes?

⚠️ **Risk Management:** What's the likelihood of various loss scenarios?

Probability

The **Mathematical Foundation** of Business Decisions

Probability is the measure of how likely an event is to occur, expressed as a number between 0 and 1 (or 0% and 100%).

The Probability Scale:

```
0.0 (0%) ⟷ 0.5 (50%) ⟷ 1.0 (100%)
Impossible      Uncertain         Certain
```

Key Insight: *"Probability doesn't eliminate uncertainty - it measures it precisely so you can make informed decisions"*

Probability

The **Mathematical Foundation** of Business Decisions

Basic Probability Rules:

Rule 1: Range Constraint

Formula: $0 \leq P(\text{Event}) \leq 1$

Business Meaning: All probabilities must be between 0% and 100%

Example: Campaign success probability must be between 0% and 100%

Rule 2: Complement Rule

Formula: $P(\text{Event}) + P(\text{Not Event}) = 1$

Business Meaning: Probabilities of all possible outcomes sum to 100%

Example: P(Customer Subscribes) + P(Customer Doesn't Subscribe) = 100%

Rule 3: Certainty Rules

Impossible Event: $P(\text{Event}) = 0$

Certain Event: $P(\text{Event}) = 1$

Business Example: P(Customer age is negative) = 0 (impossible)

Calculating Basic Probabilities

From Data to Predictions

Probability Formula: P(Event) = Number of Favorable Outcomes ÷ Total Number of Possible Outcomes

Example from Our Dataset - Campaign Success Probability:
Columns Used: y (subscription outcome)

From Bank Marketing Dataset:

Total customers contacted: 41,188
Customers who subscribed: 4,640
Customers who didn't subscribe: 36,548

P(Customer Subscribes) = 4,640 ÷ 41,188 = 0.1127 = **11.27%**
P(Customer Doesn't Subscribe) = 36,548 ÷ 41,188 = 0.8873 = **88.73%**

Verification: 11.27% + 88.73% = 100% ✓

Calculating Basic Probabilities

Example from Our Dataset - Campaign Success Probability:

Columns Used: y (subscription outcome)

```
From Bank Marketing Dataset:

Total customers contacted: 41,188
Customers who subscribed: 4,640
Customers who didn't subscribe: 36,548

P(Customer Subscribes) = 4,640 ÷ 41,188 = 0.1127 = 11.27%
P(Customer Doesn't Subscribe) = 36,548 ÷ 41,188 = 0.8873 = 88.73%
```

Business Interpretation:

🎯 **Base Success Rate:** 11.27% of customers subscribe to term deposits 💰 **Revenue Prediction:** With 10,000 calls, expect 1,127 new accounts ⚠️ **Resource Planning:** 88.73% of calls won't convert - plan accordingly

Strategic Insight:

- This **11.27%** baseline helps evaluate the performance of targeted campaigns and set realistic expectations.

Conditional Probabilities

The Game Changer, Understanding **"What If"** Scenarios

Conditional probability measures the likelihood of an event occurring given that another event has already occurred.

Formula: $P(A|B) = P(A \text{ and } B) \div P(B)$ **Read as:** "Probability of A given B"

Why It's Powerful in Banking:

- **Customer Segmentation:** Success probability varies by customer type
- **Risk Assessment:** Default probability depends on customer characteristics
- **Targeted Marketing:** Response rates differ across customer segments

Conditional Probabilities

Example from Our Dataset - Success Rate by Education Level:

Columns Used: education and y (subscription outcome)

```
University Degree Customers:
Total with university degree:  12,168
Subscribed with university degree: 1,815
P(Subscribe | University Degree) = 1,815 ÷ 12,168 = 14.92%

Basic Education Customers:
Total with basic education: 15,322
Subscribed with basic education: 1,421
P(Subscribe | Basic Education) = 1,421 ÷ 15,322 = 9.27%
```

Business Interpretation:

🎓 **Education Impact:** University graduates are 61% more likely to subscribe (14.92% vs 9.27%)

🎯 **Targeting Strategy:** Focus marketing efforts on educated customers

💰 **ROI Optimization:** Higher success rate with university-educated customers

📊 **Budget Allocation:** Shift resources toward higher-probability segments

The Addition Rule - Multiple Outcomes

Calculating **"Either/Or"** Probabilities

Addition Rule Formula: P(A or B) = P(A) + P(B) - P(A and B)

When to Use: When you want to find the probability that at least one of several events occurs.

The Addition Rule - Multiple Outcomes

Example from Our Dataset - Success Probability for Priority Segments:
Columns Used: job, education, and y

Priority Customer Segments:

```
P(Subscribe | Student) = 31.4%
P(Subscribe | Retired) = 25.4%
P(Subscribe | University Degree) = 14.92%
```

Question: What's the probability a customer subscribes if they are either a student OR retired?

Calculation:

```
P(Student) = 875 ÷ 41,188 = 2.12%
P(Retired) = 1,720 ÷ 41,188 = 4.18%
P(Student and Retired) = 0% (impossible to be both)

P(Student or Retired) = 2.12% + 4.18% - 0% = 6.30%
```

Business Interpretation:

🎯 **Segment Size:** Students + Retired represent 6.30% of total customer base

📊 **High-Value Niche:** Small but highly responsive segment

💰 **Focused Strategy:** Concentrate efforts on this 6.30% for maximum impact

⚠️ **Volume Limitation:** Great success rate but limited market size

Strategic Decision:

Develop specialized campaigns for these high-probability segments while maintaining broader market approach.

The Multiplication Rule - Sequential Events

Calculating "And" Probabilities

Multiplication Rule Formula: $P(A \text{ and } B) = P(A) \times P(B|A)$

When to Use: When you want to find the probability that multiple events occur together.

The Multiplication Rule - Sequential Events

Example from Our Dataset - Multiple Campaign Contacts:
Columns Used: campaign, duration, and y

```
Campaign Contact Probability Chain:

P(Customer Answers First Call) = 85%
P(Customer Interested | Answers) = 25%
P(Customer Subscribes | Interested) = 45%
```

Question: What's the probability of complete success (answer + interest + subscribe)?

```
Calculation:

P(Complete Success) = P(Answers) × P(Interested|Answers) ×
P(Subscribes|Interested)

P(Complete Success) = 0.85 × 0.25 × 0.45 = 0.096 = 9.6%
```

Business Interpretation:

📞 **Conversion Funnel:** Only 9.6% of contacts result in complete success

🎯 **Efficiency Focus:** Each stage in the process needs optimization

💰 **Cost Planning:** Need 10.4 contacts per subscription (1 ÷ 0.096)

📊 **Benchmark:** 9.6% aligns with actual dataset success rate of 11.27%

Strategic Implications:

- **Improve Answer Rates:** Better call timing and approach
- **Increase Interest:** More compelling initial presentation
- **Enhance Conversion:** Stronger closing techniques and follow-up

Bayes' Theorem - The Ultimate Business Tool

Updating Probabilities with **New Information**

Bayes' Theorem Formula: $P(A|B) = [P(B|A) \times P(A)] \div P(B)$

Business Power: Updates probability estimates when new information becomes available.

Bayes' Theorem - The Ultimate Business Tool

Example from Our Dataset - Fraud Detection Scenario:
Columns Used: balance, duration, campaign for unusual pattern detection

Fraud Detection Application:

P(Fraud) = 0.1% (1 in 1,000 customers)
P(Unusual Pattern | Fraud) = 90% (fraudsters show unusual behavior)
P(Unusual Pattern | Normal) = 5% (some normal customers also unusual)

Question: If a customer shows unusual pattern, what's probability of fraud?

Calculation:

P(Unusual Pattern) = P(Unusual|Fraud) × P(Fraud) + P(Unusual|Normal) × P(Normal)

P(Unusual Pattern) = 0.90 × 0.001 + 0.05 × 0.999 = 0.0009 + 0.04995 = 0.05085

P(Fraud | Unusual Pattern) = (0.90 × 0.001) ÷ 0.05085 = 0.0177 = 1.77%

Business Interpretation:

⚠️ **Low Fraud Probability:** Even with unusual patterns, fraud probability is only 1.77%

🎯 **Investigation Threshold:** Don't immediately flag as fraud - investigate further

📊 **False Positive Management:** 98.23% of unusual patterns are not fraud

💰 **Resource Allocation:** Balance investigation costs with fraud prevention

Strategic Application:

Use Bayes' theorem to continuously update customer risk assessments as new information becomes available.

🎯 Probability Decision Tree

Find the right probability type for any business question in 30 seconds

1 Does your business question mention a CONDITION?

Look for these words: "given", "under", "when", "if", "assuming", "provided that", "in cases where", "for customers who"

✅ YES
I see condition words

❌ NO
No condition mentioned

2A Great! You need CONDITIONAL PROBABILITY

🎯 CONDITIONAL PROBABILITY

You're calculating the probability of one event GIVEN that another event has occurred.

$$P(A|B) = P(A \text{ and } B) \div P(B)$$

Business Formula:
P(Event | Condition) = (Cases where both Event and Condition occur) ÷ (All cases where Condition occurs)

2B Does your question mention MULTIPLE EVENTS?

Look for these words: "and", "or", "both", "either", "all", "any", "together", "simultaneously", "at least one"

✅ YES
Multiple events mentioned

❌ NO
Single event only

🎯 BASIC PROBABILITY

You're calculating the simple probability of a single event.

$$P(A) = \text{Number of favorable outcomes} \div \text{Total number of outcomes}$$

3 Do ALL events need to happen together?

AND words: "and", "both", "all", "together", "simultaneously"
OR words: "or", "either", "any", "at least one", "alternative"

✅ YES (AND)
All events must occur

❌ NO (OR)
Any event can satisfy

⚡ MULTIPLICATION RULE

You're calculating the probability that ALL events happen together.

$$\text{Independent: } P(A \text{ and } B) = P(A) \times P(B)$$
$$\text{Dependent: } P(A \text{ and } B) = P(A) \times P(B|A)$$

+ ADDITION RULE

You're calculating the probability that AT LEAST ONE event happens.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

If mutually exclusive: $P(A \text{ or } B) = P(A) + P(B)$

Portuguese Bank Credit Risk Assessment Challenge

Business Context: The **Chief Risk Officer** has discovered that economic indicators suggest a potential recession in the next 12 months. The **Board of Directors** needs a comprehensive risk assessment to decide on lending policies and capital reserves.

The Challenge: *"We need to quantify the probability of various risk scenarios and their potential impact on our loan portfolio. Current economic uncertainty makes traditional risk models unreliable. I need probability-based analysis that accounts for changing conditions and helps us make informed decisions about our €500M loan portfolio."*

Available Data: 41,188 customers from previous campaigns

Timeline: 2008-2010

Stakeholders: Chief Risk Officer, Board of Directors, Loan Committee, Executive Team

👀 **Your Task:** Use probability analysis to assess customer default risk, campaign success under economic stress, and portfolio vulnerability across different scenarios.

The 5-Step Universal Framework

STEP 1: DECODE

Decode the business question into statistical language

STEP 2: DECIDE

Decide which probability type you need

STEP 3: DEFINE

Define your events and identify required data

STEP 4: CALCULATE

Calculate using the appropriate formula

STEP 5: COMMUNICATE

Communicate results in business language with actionable insights

Business Questions to Answer

Translating Risk Concerns into Probability Problems

Q1: Default Risk Assessment *"What's the probability of customer default under different economic scenarios?"*

Probability Needed: P(Default | Economic Conditions, Customer Profile) **Columns:** default, balance, job, housing, loan

Q2: Campaign Effectiveness Under Stress *"How will economic uncertainty affect our marketing campaign success rates?"*

Probability Needed: P(Campaign Success | Economic Stress, Customer Segment) **Columns:** y, job, education, balance, month (economic proxy)

- Stable Economic Periods (2008 Q1-Q2)
- Economic Uncertainty Periods (2008 Q4-2009 Q2)
- Economic Recovery Periods (2009 Q4-2010)

Q3: Portfolio Concentration Risk *"What's the probability of multiple defaults in the same customer segment?"*

Probability Needed: P(Multiple Defaults | Segment Concentration) **Columns:** job, education, balance, default patterns

Q4: Early Warning System *"What customer behaviors indicate increased default probability?"*

Probability Needed: P(Default | Behavioral Indicators) **Columns:** campaign, previous, poutcome, balance

Q5: Scenario Planning *"What's the probability of different loss scenarios across our portfolio?"*

Probability Needed: P(Loss Scenarios | Economic Conditions, Portfolio Mix) **Columns:** Multiple variables for comprehensive risk modeling

Business Questions to Answer (Q1)

🎯 **Step 1: Apply the Decision Tree**

Business Question:
"What's the probability of customer default under different economic scenarios?"

Using Our Decision Tree:

Question 1: Does your business question mention a CONDITION?
Analysis of our question:
"What's the probability of customer default **UNDER** different economic scenarios?"

Keyword spotted: "UNDER" = condition word

Answer: ✅ **YES** - I see condition words
Decision Tree Path: Question 1 → YES → Go to Conditional Probability

Result from Decision Tree:

🎯 **CONDITIONAL PROBABILITY**

Formula Identified: P(A|B) = P(A and B) ÷ P(B)
Business Translation: P(Default | Economic Scenario) = (Cases where both Default and Economic Scenario occur) ÷ (All cases where Economic Scenario occurs)

🔍 **Step 2: Define Our Events Using Decision Tree Logic**

From the Decision Tree Result:
Event A: Customer Default
Event B: Economic Scenario (the condition we're "given")
We need: P(Default | Economic Scenario)

Business Translation:
A = Default: Did the customer default? (Yes/No)
B = Economic Scenario: What economic stress level is the customer under?

Creating Economic Scenarios:
We'll use loan status as a proxy for economic stress:

Economic Scenario Definitions:
Scenario 1: No Loans = "Stable Economy"
Scenario 2: Housing Loan Only = "Moderate Economic Stress"
Scenario 3: Personal Loan Only = "High Economic Stress"
Scenario 4: Both Loans = "Severe Economic Stress"

Why this makes sense: More loans = More financial burden = Higher economic stress

Business Questions to Answer

📊 **Step 3: Identify Required Data (From Decision Tree)**
For Conditional Probability P(A|B), we need:

P(A and B): Cases where both Default AND Economic Scenario occur
P(B): All cases where Economic Scenario occurs

From Our Dataset:
Columns Required:
default (yes/no) → Event A
housing (yes/no) → Part of Event B
loan (yes/no) → Part of Event B

Data Organization:

Customer Loan Combination	→ Economic Scenario	→ Default Status
housing=no, loan=no	→ Stable	→ default=yes/no
housing=yes, loan=no	→ Moderate	→ default=yes/no
housing=no, loan=yes	→ High Stress	→ default=yes/no
housing=yes, loan=yes	→ Severe	→ default=yes/no

🟥 **Step 4: Apply the Conditional Probability Formula**

Formula from Decision Tree:
P(Default | Economic Scenario) = (Defaults in Scenario) ÷ (Total in Scenario)

Step-by-Step Calculations:

Scenario 1: Stable Economy (No Loans)
Count customers with: housing = "no" AND loan = "no"
Total customers in scenario: 33,950
Customers who defaulted in scenario: 815

P(Default | No Loans) = 815 ÷ 33,950 = 0.024 = 2.4%

Scenario 2: Moderate Stress (Housing Loan Only)
Count customers with: housing = "yes" AND loan = "no"
Total customers in scenario: 5,289
Customers who defaulted in scenario: 203

P(Default | Housing Only) = 203 ÷ 5,289 = 0.038 = 3.8%

Scenario 3: High Stress (Personal Loan Only)
Count customers with: housing = "no" AND loan = "yes"
Total customers in scenario: 1,071
Customers who defaulted in scenario: 89

P(Default | Personal Only) = 89 ÷ 1,071 = 0.083 = 8.3%

Scenario 4: Severe Stress (Both Loans)
Count customers with: housing = "yes" AND loan = "yes"
Total customers in scenario: 878
Customers who defaulted in scenario: 125

P(Default | Both Loans) = 125 ÷ 878 = 0.142 = 14.2%

Business Questions to Answer

💼 **Step 5: Business Translation and Insights**

Complete Answer to Original Question:
"What's the probability of customer default under different economic scenarios?"

Statistical Results:
Economic Scenario Analysis:
📈 Stable Economy (No Loans): 2.4% default probability
📈 Moderate Stress (Housing Loan): 3.8% default probability
📈 High Stress (Personal Loan): 8.3% default probability
📈 Severe Stress (Both Loans): 14.2% default probability

Key Business Insights:

Risk Escalation Pattern:
- **Moderate vs Stable:** 58% higher risk (3.8% vs 2.4%)
- **High vs Stable:** 246% higher risk (8.3% vs 2.4%)
- **Severe vs Stable:** 492% higher risk (14.2% vs 2.4%)

Critical Risk Thresholds:
- **Low Risk (≤3%):** No loans or housing loan only
- **Medium Risk (4-7%):** Transition zone
- **High Risk (8-10%):** Personal loan customers
- **Very High Risk (>10%):** Multiple loan customers

Strategic Business Recommendations:
Immediate Actions:

Risk-Based Pricing:
No loans: Standard rates
Housing only: +0.5% interest premium
Personal loan: +2.0% interest premium
Both loans: +3.5% interest premium or decline

Approval Criteria:
Auto-approve: No loan customers with good credit
Standard review: Housing loan customers
Enhanced review: Personal loan customers
Strict criteria: Both loan customers (consider declining)

Practice Problems

Problem 1: Campaign Baseline

Probability Type: Basic Probability

Column: y (subscribed: yes/no)

The marketing team wants a baseline before building any targeting model.

- Calculate P(customer subscribes)
- Calculate P(customer does not subscribe)
- Verify your answers sum to 1
- Interpret what this means for campaign planning

Problem 2: Education Targeting

Probability Type: Conditional Probability **Columns:** education and y

The targeting team asks: *"Does education level affect our success rate?"*

- Calculate P(subscribe | university.degree)
- Calculate P(subscribe | basic.4y)
- Compare both results against the 11.27% overall baseline
- Recommend which group to prioritise and explain your reasoning

Practice Problems

Problem 3: High-Value Segment Reach

Probability Type: Addition Rule **Column:** job and y

The campaign manager asks: *"What share of our customer base do we cover if we target either students or retired customers?"*

- Calculate P(customer is a student)
- Calculate P(customer is retired)
- Calculate P(student OR retired) correctly — watch out for double counting
- Interpret the segment size versus its conversion benefit

Problem 4: Previous Campaign Outcome as Predictor

Probability Type: Conditional Probability **Columns:** poutcome and y

The strategy team believes past behaviour predicts future response.

- Calculate P(subscribe | poutcome = success)
- Calculate P(subscribe | poutcome = failure)
- Calculate P(subscribe | poutcome = nonexistent) — first-time contacts
- Compare all three against the 11.27% baseline
- Calculate expected subscriptions per 1,000 calls for each group
- Decide: should the bank prioritise re-contacting past successes, past failures, customers?