Probability in Computer Science Cheatsheet

Comprehensive Guide with Python Examples

---

1 Introduction to Probability

Probability is the measure of the likelihood that an event will occur. It is widely used in computer science for applications like machine learning, randomized algorithms, and network modeling.

Basic Definitions

1. **Sample Space (S)**: The set of all possible outcomes of an experiment. Example: Flipping a coin: $S = \{\text{Heads, Tails}\}$

2. **Event (A)**: A subset of the sample space. Example: Rolling an even number with a die: $A = \{2, 4, 6\}$

3. **Probability of an Event ($P(A)$)**:

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

Example: Rolling a 4 with a die: $P(4) = \frac{1}{6}$

—

2 Key Rules of Probability

2.1 Addition Rule (Union of Events)

If A and B are two events:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- If A and B are mutually exclusive ($P(A \cap B) = 0$):

$$P(A \cup B) = P(A) + P(B)$$

2.2 Multiplication Rule (Intersection of Events)

If A and B are two events:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

- If A and B are independent ($P(B|A) = P(B)$):

$$P(A \cap B) = P(A) \cdot P(B)$$

2.3 Complement Rule

The probability of the complement of an event (A') is:

$$P(A') = 1 - P(A)$$

Python Example: Basic Probability Rules

```python
# Probability Example: Dice Rolls
favorable_outcomes = 1  # Rolling a 4
total_outcomes = 6
P_4 = favorable_outcomes / total_outcomes
print(f"P(Rolling 4): {P_4}")  # Output: 0.1667
```

—

3 Conditional Probability and Bayes' Theorem

3.1 Conditional Probability

The probability of A given that B has occurred:

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

- Used in scenarios like diagnosing diseases or spam email detection.

3.2 Bayes' Theorem

Relates $P(A|B)$ and $P(B|A)$:

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

- Crucial for Bayesian inference in machine learning.

Python Example: Conditional Probability

```python
# Conditional Probability Example
P_B = 0.6  # Probability of B
P_A_and_B = 0.3  # Probability of A and B
P_A_given_B = P_A_and_B / P_B
print(f"P(A | B): {P_A_given_B}")  # Output: 0.5
```

—

4 Random Variables and Expectation

4.1 Random Variables

A random variable (X) is a function that assigns numerical values to outcomes of a random experiment. - **Discrete**: Takes specific values (e.g., rolling a die). - **Continuous**: Takes any value in a range (e.g., height, weight).

4.2 Expectation (Mean)

The expected value of a random variable (X) is the weighted average of all possible values:

$$E(X) = \sum_x x \cdot P(X = x)$$

4.3 Variance

The variance measures the spread of a random variable:

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

Python Example: Expectation and Variance

```python
# Discrete Random Variable Example
values = [1, 2, 3]
probabilities = [0.2, 0.5, 0.3]
expected_value = sum(x * p for x, p in zip(values, probabilities))
variance = sum((x**2) * p for x, p in zip(values, probabilities)) - expected_value**2
print(f"E(X): {expected_value}, Var(X): {variance}")
# Output: E(X): 2.1, Var(X): 0.49
```

—

5 Common Probability Distributions

5.1 Binomial Distribution

Used for n independent trials with success probability p:

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

- Mean: $\mu = np$ - Variance: $\sigma^2 = np(1 - p)$

Python Example: Binomial Distribution

```python
from math import comb

n = 10  # Number of trials
k = 3   # Number of successes
p = 0.5 # Probability of success
P_k = comb(n, k) * (p**k) * ((1-p)**(n-k))
print(f"P(X = {k}): {P_k}")
```

5.2 Poisson Distribution

Models the number of events in a fixed interval:

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

- Mean: $\mu = \lambda$ - Variance: $\sigma^2 = \lambda$

Python Example: Poisson Distribution

```python
from math import exp, factorial

lam = 4  # Average rate
k = 2    # Number of occurrences
P_k = (lam**k * exp(-lam)) / factorial(k)
print(f"P(X = {k}): {P_k}")
```

—

6 Applications in Computer Science

- **Machine Learning:** Bayesian networks, Naive Bayes classifiers.
- **Randomized Algorithms:** Monte Carlo simulations.
- **Big Data:** Sampling techniques and error estimation.
- **Networking:** Traffic modeling, reliability, and packet loss analysis.

This cheatsheet provides a compact summary. For deeper understanding, explore textbooks like *Probability and Computing* by Mitzenmacher and Upfal.