Master 2 Computer Science
RL Course M2 AI

Mathematical Reminders

Akka Zemmari

Introduction

Probability Theory Reminders

Statistical Inference

Optimization and Differential Calculus

Fixed Point and Its Usage in Reinforcement Learning

Introduction

Discleamer

This chapter aims to introduce the mathematical concepts, notions, and notations that we will use throughout this course.

It is clearly a reminder and not a complete course. Each concept discussed here would deserve several chapters to be fully developed, but the focus is on fundamentals, definitions, and some useful theorems (without proofs) for the following.

Probability Theory Reminders

Random Variables

- **Definition**: A random variable (r.v.) is a function that assigns a real number to each possible outcome of a random experiment.
  There are two main types of random variables:
  - **Discrete**: Takes a finite or countable number of values (e.g., rolling a die).
  - **Continuous**: Takes an uncountable set of values, usually intervals of real numbers (e.g., measuring a person's height).
- **Notation**: Let X be a r.v.. We denote $\mathbb{P}r\left(X = x\right)$ the probability that X takes the value x (discrete) or $f_X(x)$ the probability density function (continuous).

Discrete Probability Distributions

- **Definition**: The probability distribution of a discrete r.v. is a list of probabilities associated with each possible value.
- **Common Examples**:
  - **Bernoulli**: $X \sim \mathrm{B}(p)$, with $\mathbb{P}r\,(X = 1) = p$ and $P(X = 0) = 1 - p$.
  - **Binomial**: $X \sim \mathrm{Bin}(n, p)$, with $\mathbb{P}r\,(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
  - **Poisson**: $X \sim \mathrm{Poisson}(\lambda)$, with $\mathbb{P}r\,(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$.
- **Link with RL**: Discrete distributions often model rewards (distributed discretely) or choices of actions in an environment.

Continuous Probability Distributions

- **Definition**: The probability distribution of a continuous r.v. is described by a probability density function (pdf) $f_X(x)$, where $\mathbb{P}r\left(a \leq X \leq b\right) = \int_a^b f_X(x)dx$.
- **Common Examples**:
  - **Normal (Gaussian)**: $X \sim \mathcal{N}(\mu, \sigma^2)$, with $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
  - **Exponential**: $X \sim \text{Exp}(\lambda)$, with $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.
  - **Uniform**: $X \sim \text{Uniform}(a, b)$, with $f_X(x) = \frac{1}{b-a}$ for x in $[a, b]$.
- **Link with RL**: useful for modeling continuous rewards or actions in continuous environments (e.g., continuous control).

Cumulative Distribution Function (CDF)

- **Definition**: The cumulative distribution function of a r.v. X, discrete or continuous, is defined by $F_X(x) = \mathbb{P}r\,(X \leq x)$.
- **Properties**:
    - **Monotonic**: $F_X(x)$ is a non-decreasing function.
    - **Limits**: $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.
- **Use in RL**: It allows evaluating the cumulative probability of obtaining a reward or being in a certain state.

Expectation, Variance, and Moments

Expectation

- **Definition**: The expectation (or mean) of a r.v. is the average value that this r.v. takes over a large number of realizations of the random experiment.
- **Formulation**:
  - For a discrete r.v. X with probability function $\mathbb{P}r\,(X = x_i)$:

$$\mathbb{E}[X] = \sum_i x_i \cdot \mathbb{P}r\,(X = x_i)$$

  - For a continuous r.v. X with probability density $f_X(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$$

Expectation, Variance, and Moments

Expectation

- **Properties**:
  - Linearity: $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ for constants a and b.
  - Sum: If X and Y are two random variables, then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- **Link with RL**: Expectation is used to calculate the expected value of a reward, state, or action. It is fundamental for policy evaluation in RL, notably in Bellman equations.

Expectation, Variance, and Moments

Variance

- **Definition**: Variance measures the dispersion of the values of a random variable around its expectation.
- **Formulation**:
  - For a random variable X:

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

  - For a discrete variable:

$$\mathrm{Var}(X) = \sum_i (x_i - \mathbb{E}[X])^2 \cdot \mathbb{P}r\left(X = x_i\right)$$

  - For a continuous variable:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 \cdot f_X(x)\, dx$$

Expectation, Variance, and Moments

Variance

- **Properties**:
  - Non-negativity: $\text{Var}(X) \geq 0$, with $\text{Var}(X) = 0$ if and only if X is a constant.
  - Relation to expectation: $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- **Link with RL**: Variance is useful for evaluating the uncertainty or variability of rewards and transitions in an environment.

Expectation, Variance, and Moments

Covariance and Correlation

- **Definition of Covariance**: Covariance measures how two random variables vary together.

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- **Correlation**: Correlation is a normalized version of covariance, measuring the strength and direction of the linear relationship between two random variables.

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$

Expectation, Variance, and Moments

Covariance and Correlation

- **Properties**:
  - $\text{Cov}(X, Y) > 0$: Indicates that X and Y tend to increase together.
  - $\text{Cov}(X, Y) < 0$: Indicates that X increases when Y decreases.
  - $\rho_{X,Y}$ ranges between -1 (perfect negative correlation) and 1 (perfect positive correlation).
- **Link with RL**: Covariance and correlation can be used to understand the relationships between different variables in an environment, such as actions and rewards, or successive states.

Expectation, Variance, and Moments

Higher-Order Moments

- **Moments**: The k-th order moments of a random variable X are given by $\mathbb{E}[X^k]$.
  - **First moment**: The expectation $\mathbb{E}[X]$, representing the mean.
  - **Second moment**: $\mathbb{E}[X^2]$, related to the variance.
  - **Third moment**: Skewness, measuring the asymmetry of the distribution relative to its mean.
  - **Fourth moment**: Kurtosis, measuring the concentration of values around the mean (the "tail size" of the distribution).
- **Link with RL**: Higher-order moments can be used to more finely characterize reward distributions or state transitions.

Expectation, Variance, and Moments

Applications in Reinforcement Learning

- **Policy Evaluation**: Use of expectation to calculate the expected value of policies.
- **Risk Management**: Use of variance and higher-order moments to assess and minimize uncertainty in outcomes in stochastic environments.
- **Policy Optimization**: Adjustment of policies based on the moments of reward distributions, to maximize expected reward while minimizing risk.

Inequalities and Limit Theorems

Markov Inequality

- **Statement**: For a non-negative random variable X and a threshold $a > 0$, Markov's inequality states that:

$$\mathbb{P}r\,(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- **Application**: Used to obtain bounds on the probabilities of rare events.
- **Link with RL**: Can be used to evaluate the probabilities of large errors in value estimates.

Inequalities and Limit Theorems

Chebyshev's Inequality

- **Statement**: For a random variable X with expectation μ and variance σ^2, Chebyshev's inequality states that for any $k > 0$:

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

- **Application**: Used to obtain bounds on deviations from the mean.
- **Link with RL**: Used to assess the stability of value estimates and potential deviation from the expected value.

Inequalities and Limit Theorems

Law of Large Numbers

- **Statement**: The law of large numbers states that the empirical average of a large number of independent and identically distributed random variables converges to their expectation.
- **Formulation**:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}[X] \quad \text{as } n \to \infty$$

- **Application**: Ensures that estimates based on samples become accurate with a large number of samples.
- **Link with RL**: Ensures that average reward estimates converge to their true value as more samples are collected.

Inequalities and Limit Theorems

Central Limit Theorem

- **Statement**: The central limit theorem states that the sum (or average) of many independent identically distributed random variables tends towards a normal distribution, regardless of the original distribution of the variables.
- **Formulation**:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i]) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } n \to \infty$$

- **Application**: Used for probabilistic approximations and to derive confidence intervals.
- **Link with RL**: Enables predictions on the distribution of value estimates, facilitating policy performance analysis.

Statistical Inference

Parametric Estimation

Point Estimators

- **Definition**: A point estimator is a statistic computed from the sample and used to estimate an unknown population parameter (e.g., mean or variance).
- **Properties of Estimators**:
  - **Bias**: An estimator is biased if its expectation is not equal to the parameter being estimated.
  - **Consistency**: An estimator is consistent if, as the sample size increases, the estimation converges in probability to the parameter being estimated.
  - **Efficiency**: An estimator is efficient if it has the smallest variance among all unbiased estimators.
- **Link with RL**: Point estimators are used in evaluating policy performance and for estimating value functions.

Parametric Estimation

Confidence Interval Estimation

- **Definition**: A confidence interval is an interval constructed from sample data such that, under certain assumptions, it contains the value of the unknown parameter with a given probability (confidence level).
- **Calculating a Confidence Interval**:
  - For the mean μ of a normal distribution with known variance σ^2, a $1-\alpha$ confidence interval is given by:

    $$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

    where $\bar{X}$ is the sample mean and $z_{\alpha/2}$ is the quantile of the normal distribution.
- **Link with RL**: Confidence intervals are used to quantify uncertainty in policy value estimates and in decisions

Hypothesis Testing

Hypothesis Formulation

- **Null Hypothesis** H_0: This is the hypothesis to be tested, often formulated as a hypothesis of no effect or status quo.
- **Alternative Hypothesis** H_1: This is the hypothesis accepted if the data provides sufficient evidence against H_0.
- **Significance Levels and Errors**:
  - **Type I Error** (α): Rejecting H_0 when it is true.
  - **Type II Error** (β): Failing to reject H_0 when it is false.
- **Link with RL**: Hypothesis tests can be used to compare different policies or validate reward models.

Hypothesis Testing

- **Parametric Tests**: Based on specific assumptions about the distribution of data (e.g., Student's t-test, ANOVA).
- **Non-Parametric Tests**: No strong assumptions about the data distribution (e.g., Wilcoxon test, Kruskal-Wallis test).
- **Link with RL**: Parametric tests are used when data assumptions are met, while non-parametric tests offer an alternative when these assumptions are not verified.

Hypothesis Testing

Test Power and Sample Size

- **Test Power**: Probability of rejecting H_0 when H_1 is true. Power depends on the sample size, the effect to be detected, and the significance level.
- **Calculating Sample Size**: For a test with power $1 - \beta$ and significance level α, the sample size can be determined to ensure that the test detects an effect of a given size.
- **Link with RL**: Sufficient test power is essential to ensure that optimal policies are detected and validated with statistical confidence.

Optimization and Differential Calculus

Basic Concepts in Optimization

Optimization Problems

- **Definition**: An optimization problem involves finding the values of a set of variables that minimize or maximize an objective function, often under constraints.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{or} \quad \max_{x \in \mathbb{R}^n} f(x)$$

- **Constrained Optimization**: Optimization under constraints that can be equalities or inequalities.
- **Link with RL**: Optimization problems appear in reinforcement learning in the form of maximizing cumulative rewards and optimizing value functions.

Basic Concepts in Optimization

Solution Methods

- **Exact Methods**:
    - **Gradient Method**: Uses the gradient of the objective function to find a local optimal point by following the direction of maximum descent.

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

    - **Newton's Method**: Uses second derivative information (Hessian) to accelerate convergence to a critical point.

$$x_{k+1} = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$$

Basic Concepts in Optimization

Solution Methods

- **Approximate Methods**:
  - **Stochastic Gradient Descent (SGD)**: A variant of the gradient method where only stochastic estimates of the gradient are used, allowing optimization of non-convex functions and large-scale problems.
- **Link with RL**: GD methods are used in algorithms like Q-learning and Policy Gradient to optimize value functions and policies.

Differential Calculus

Derivatives and Gradients

- **Derivative of a Function**: The derivative of $f(x)$ with respect to x is the limit of the rate of change of the function as x changes.

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

- **Gradient**: For a function $f(x_1, \ldots, x_n)$, the gradient is a vector containing the partial derivatives with respect to each variable.

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

- **Link with RL**: Gradients are used to adjust parameters in RL, such as policy gradient methods.

Differential Calculus

Derivatives of Vector Functions

- **Jacobian**: This is the matrix of partial derivatives of a vector function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, containing all partial derivatives with respect to each input variable.

$$J_{\mathbf{f}}(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

- **Link with RL**: Jacobians are relevant in neural networks and in calculating gradients for multidimensional outputs in deep RL algorithms.

Fixed Point and Its Usage in Reinforcement Learning

Introduction to Fixed Point

A fixed point of a function is a point that remains unchanged when the function is applied to it. Formally, if f is a function, then x is a fixed point of f if:

$$f(x) = x.$$

Fixed points appear in various contexts in mathematics, optimization, and dynamical systems modeling.

In reinforcement learning (RL), fixed points are fundamental because they allow us to formulate and solve equations that define the optimal values of states or policies.

Banach Fixed Point Theorem

The Banach Fixed Point Theorem, also known as the Contraction Mapping Theorem, is a key mathematical tool in the analysis of fixed points. It states that any contraction mapping defined on a complete metric space has a unique fixed point, and repeated application of the function converges to this fixed point.

Theorem: Let (X, d) be a complete metric space, and let $T : X \to X$ be a contraction (i.e., there exists a $0 < c < 1$ such that $d(T(x), T(y)) \leq cd(x, y)$ for all $x, y \in X$). Then T has a unique fixed point, and for any $x_0 \in X$, the sequence defined by $x_{n+1} = T(x_n)$ converges to this fixed point.

Usage of Fixed Points in RL

In RL, fixed points are crucial for understanding and solving Bellman equations, which are at the core of RL algorithms such as policy evaluation, value iteration, and $Q-$learning.