**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**MÔN HỆ HỖ TRỢ QUYẾT ĐỊNH – IS254.P11**

**Đề tài: Policy Validation**

GVHD: ThS. Nguyễn Hồ Duy Trí

Nhóm sinh viên thực hiện: Nhóm 15

|  |  |
| --- | --- |
| Phan Nguyễn Lâm Hà | 21522030 |

🙡🙢 Tp. Hồ Chí Minh, 12/2024 🙠🙣

**LỜI CẢM ƠN**

Trước hết, nhóm chúng em xin gửi lời cảm ơn sâu sắc đến tập thể quý thầy cô trường Đại học Công nghệ Thông tin - Đại học Quốc gia TP.HCM và quý thầy cô khoa Hệ thống thông tin đã tạo điều kiện, giúp chúng em học tập và có được những kiến thức cơ bản làm tiền đề giúp chúng em hoàn thành được dự án này.

Đặc biệt, nhóm chúng em xin gửi lời cảm ơn chân thành và sâu sắc tới thầy Nguyễn Hồ Duy Trí (Giảng viên môn Hệ hỗ trợ quyết định IS254.P11). Nhờ sự hướng dẫn tận tình và chu đáo của thầy, nhóm chúng em đã học hỏi được nhiều kinh nghiệm và hoàn thành thuận lợi, đúng tiến độ cho dự án của mình.

Ngoài ra, chúng em cũng gửi lời cảm ơn đến tập thể lớp IS254.P11 khoảng thời gian qua đã đồng hành cùng nhau. Cảm ơn sự đóng góp của tất cả các bạn cho những buổi học luôn sôi nổi, thú vị và dễ tiếp thu.

Trong quá trình thực hiện tiểu luận, nhóm chúng em luôn giữ một tinh thần cầu tiến, học hỏi và cải thiện từ những sai lầm, tham khảo từ nhiều nguồn tài liệu khác nhau và luôn mong đạt được kết quả nhất có thể. Tuy nhiên, do vốn kiến thức còn hạn chế trong quá trình trau dồi từng ngày, nhóm chúng em không thể tránh được những sai sót, vì vậy chúng em mong rằng quý thầy cô sẽ đưa ra nhận xét một cách chân thành để chúng em học hỏi thêm kinh nghiệm nhằm mục đích phục vụ tốt các dự án khác trong tương lai. Xin chân thành cảm ơn quý thầy cô!

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

*……., ngày……...tháng……năm 2024*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên****)***

MỤC LỤC

[CHƯƠNG 1: LÝ DO CHỌN ĐỀ TÀI 6](#_Toc189227857)

[CHƯƠNG 2: TỔNG QUAN BỘ DỮ LIỆU 7](#_Toc189227858)

[2.1. Giới thiệu bộ dữ liệu 7](#_Toc189227859)

[2.2. Mô tả thuộc tính bộ dữ liệu 7](#_Toc189227860)

[2.3. Tiền xử lý dữ liệu 9](#_Toc189227861)

[CHƯƠNG 3: NỘI DUNG THUẬT TOÁN 12](#_Toc189227862)

[3.1. Policy Validation 12](#_Toc189227863)

[3.1.1. Khái niệm 12](#_Toc189227864)

[3.1.2. Mục tiêu 12](#_Toc189227865)

[3.1.3. Ví dụ tổng quan (Hệ thống tránh va chạm Collision Avoidance System) 12](#_Toc189227866)

[3.1.3.1. Bối cảnh 12](#_Toc189227867)

[3.1.3.2. Mục tiêu 12](#_Toc189227868)

[3.1.3.3. Phương thức 12](#_Toc189227869)

[3.2. Performance Metric Evaluation 13](#_Toc189227870)

[3.2.1. Khái niệm 13](#_Toc189227871)

[3.2.2. Mục đích của Performance Metric Evaluation 13](#_Toc189227872)

[3.2.3. Ứng dụng trong Policy Validation 14](#_Toc189227873)

[3.2.3.1. Xác định thước đo hiệu suất (Performance Metrics): 14](#_Toc189227874)

[3.2.3.2. Thu thập dữ liệu quỹ đạo (Trajectories): 14](#_Toc189227875)

[3.2.3.3. Quá trình Performance Metric Evaluation trong hệ thống 14](#_Toc189227876)

[3.2.4. Áp dụng 15](#_Toc189227877)

[3.2.4.1. Bộ dữ liệu 15](#_Toc189227878)

[3.2.4.2. Các bước thực hiện 15](#_Toc189227879)

[3.3. Rare Event Simulation 17](#_Toc189227880)

[3.3.1. Bối cảnh 17](#_Toc189227881)

[3.3.2. Importance Sampling 18](#_Toc189227882)

[3.3.2.1. Các bước thực hiện 18](#_Toc189227883)

[3.3.2.2. Ví dụ minh họa 19](#_Toc189227884)

[3.3.3. Đánh giá độ tin cậy của ước lượng 20](#_Toc189227885)

[3.3.4. Lợi ích và hạn chế 21](#_Toc189227886)

[3.3.5. Kết luận 21](#_Toc189227887)

[3.3.6. Áp dụng vào bộ dữ liệu 22](#_Toc189227888)

[3.4. Robustness Analysis 25](#_Toc189227889)

[3.4.1. Tầm quan trọng của Robustness Analysis 25](#_Toc189227890)

[3.4.2. Mô hình lập kế hoạch và overfiting 27](#_Toc189227891)

[3.4.3. Stress Testing 27](#_Toc189227892)

[3.4.4. Robust Dynamic Programming 27](#_Toc189227893)

[3.4.5. Kết luận 28](#_Toc189227894)

[3.4.6. Áp dụng vào bộ dữ liệu 28](#_Toc189227895)

[3.5. Trade Analysis 32](#_Toc189227896)

[3.5.1. Khái niệm 32](#_Toc189227897)

[3.5.2. Ứng dụng của phân tích đánh đổi 32](#_Toc189227898)

[3.5.3. Đường Pareto trong phân tích đánh đổi 33](#_Toc189227899)

[3.5.4. Ví dụ 33](#_Toc189227900)

[3.6. Adversarial Analysis 35](#_Toc189227901)

[3.6.1. Khái niệm 35](#_Toc189227902)

[3.6.2. Công thức phần thưởng đối nghịch 35](#_Toc189227903)

[3.6.3. Ứng dụng thực tế 36](#_Toc189227904)

[3.6.4. Thuật toán 38](#_Toc189227905)

[CHƯƠNG 4: TỔNG KẾT 40](#_Toc189227906)

[TÀI LIỆU THAM KHẢO 41](#_Toc189227907)

# LÝ DO CHỌN ĐỀ TÀI

Trong bối cảnh ngày càng có nhiều hệ thống ra quyết định tự động được ứng dụng trong các lĩnh vực quan trọng như xe tự hành, tài chính, chăm sóc y tế và quản lý công nghiệp, việc đảm bảo rằng các chiến lược ra quyết định (decision-making policies) hoạt động một cách chính xác và an toàn trong thực tế trở thành một yêu cầu cấp bách. Một chính sách ra quyết định dù được xây dựng cẩn thận thông qua mô hình hóa và tối ưu hóa vẫn có thể gặp những rủi ro tiềm ẩn khi được triển khai trong thế giới thực. Điều này đòi hỏi cần có một giai đoạn kiểm chứng (Policy Validation) để đánh giá và đảm bảo rằng chính sách đáp ứng được các tiêu chuẩn mong muốn.

Việc kiểm chứng không chỉ giúp đánh giá hiệu suất của chính sách mà còn là cơ hội để phát hiện những hạn chế hoặc điểm yếu tiềm tàng, từ đó có thể điều chỉnh và sửa chữa kịp thời. Các công cụ và phương pháp phân tích trong Policy Validation như đánh giá chỉ số hiệu suất, mô phỏng các sự kiện hiếm, và phân tích độ bền vững giúp cải thiện độ tin cậy của hệ thống. Điều này đặc biệt quan trọng trong các môi trường phức tạp hoặc có yếu tố rủi ro cao, nơi mà một quyết định sai lầm có thể dẫn đến hậu quả nghiêm trọng.

Chọn đề tài này cũng xuất phát từ nhu cầu thực tế trong các ngành công nghiệp hiện đại. Chẳng hạn, trong lĩnh vực xe tự hành, việc đánh giá chính sách lái xe an toàn trước khi triển khai là yếu tố quyết định để giảm thiểu rủi ro tai nạn. Tương tự, trong tài chính, các chiến lược đầu tư cần được kiểm tra độ ổn định trong các kịch bản thị trường khác nhau. Trong y tế, các quyết định liên quan đến chẩn đoán và điều trị bệnh nhân cần được đảm bảo chính xác và tối ưu để bảo vệ tính mạng con người.

Mục tiêu của báo cáo này là trình bày các công cụ phân tích được đề xuất trong Policy Validation, từ việc đánh giá hiệu suất chính sách, phân tích các sự kiện hiếm, đến việc tìm kiếm các kịch bản thất bại tiềm năng. Thông qua đó, báo cáo hướng đến cung cấp một cái nhìn toàn diện về cách các phương pháp này có thể hỗ trợ việc tối ưu hóa, nâng cao tính an toàn và hiệu quả của các chiến lược ra quyết định trước khi áp dụng vào thực tế. Đây không chỉ là một bước đi quan trọng trong quy trình triển khai mà còn đóng vai trò quan trọng trong việc xây dựng lòng tin của người dùng và xã hội đối với các hệ thống tự động.

# TỔNG QUAN BỘ DỮ LIỆU

## Giới thiệu bộ dữ liệu

* Tên bộ dữ liệu: Student Performance Factors
* Tác giả: lainguyn123
* Bộ dữ liệu này cung cấp cái nhìn tổng quan về các yếu tố ảnh hưởng đến kết quả học tập của học sinh trong các kỳ thi. Nó bao gồm thông tin về thói quen học tập, mức độ chuyên cần, sự tham gia của phụ huynh, và các khía cạnh khác ảnh hưởng đến thành công học tập.
* Link dataset: <https://www.kaggle.com/datasets/lainguyn123/student-performance-factors>

A screenshot of a computer

Description automatically generated

## Mô tả thuộc tính bộ dữ liệu

Bộ dữ liệu Student Peformance Factor  gồm 6607 dòng và 20 cột:

|  |  |
| --- | --- |
| **Tên cột** | **Mô tả** |
| Hours\_Studied | Số giờ học mỗi tuần |
| Attendance | Tỷ lệ số buổi học đã tham gia (%) |
| Parental\_Involvement | Mức độ tham gia của phụ huynh vào việc học của học sinh (Thấp, Trung bình, Cao) |
| Access\_to\_Resources | Mức độ sẵn có của các tài nguyên giáo dục (Thấp, Trung bình, Cao) |
| Extracurricular\_Activities | Tham gia các hoạt động ngoại khóa (Có, Không) |
| Sleep\_Hours | Số giờ ngủ trung bình mỗi đêm |
| Previous\_Scores | Điểm số từ các kỳ thi trước đó |
| Motivation\_Level | Mức độ động lực của học sinh (Thấp, Trung bình, Cao) |
| Internet\_Access | Có sẵn kết nối Internet hay không (Có, Không) |
| Tutoring\_Sessions | Số buổi học gia sư mỗi tháng |
| Family\_Income | Mức thu nhập của gia đình (Thấp, Trung bình, Cao) |
| Teacher\_Quality | Chất lượng giáo viên (Thấp, Trung bình, Cao) |
| School\_Type | Loại trường học (Công lập, Tư thục) |
| Peer\_Influence | Ảnh hưởng từ bạn bè đến thành tích học tập (Tích cực, Trung tính, Tiêu cực) |
| Physical\_Activity | Số giờ hoạt động thể chất trung bình mỗi tuần |
| Learning\_Disabilities | Sự hiện diện của khuyết tật học tập (Có, Không) |
| Parental\_Education\_Level | Trình độ học vấn cao nhất của phụ huynh (Trung học, Đại học, Sau đại học) |
| Distance\_from\_Home | Khoảng cách từ nhà đến trường (Gần, Trung bình, Xa) |
| Gender | Giới tính của học sinh (Nam, Nữ) |
| Exam\_Score | Điểm số cuối cùng của kỳ thi |

## Tiền xử lý dữ liệu

Bước 1: Cài đặt thư viện cần thiết và cấu hình một số tùy chọn trong thư viện Pandas để có thể hiển thị dữ liệu đầy đủ hơn.

A screen shot of a computer program

Description automatically generated

Bước 2: Đọc dữ liệu từ file “StudentPerformanceFactors.csv"

A screenshot of a computer

Description automatically generated

Bước 3: Kiểm tra thông tin của từng cột trong bộ dữ liệu.

A screenshot of a computer program

Description automatically generated

Bước 4: Kiểm tra dữ liệu Null sau đó xóa những dòng dữ liệu bị Null và bị trùng lặp.

A screenshot of a computer program

Description automatically generated

Bước 5: Chuyển đổi các cột dữ liệu kiểu String thành dạng số để tiến hành build model. A screenshot of a computer program

Description automatically generated

# NỘI DUNG THUẬT TOÁN

## Policy Validation

### Khái niệm

**Policy Validation** là quá trình kiểm tra và đánh giá một chính sách ra quyết định (policy) nhằm đảm bảo rằng nó hoạt động phù hợp với mục tiêu mong muốn trong bối cảnh thực tế. Quá trình này thường sử dụng các phương pháp mô phỏng và phân tích để xác định hiệu suất, tính an toàn, và khả năng thích ứng của chính sách trước khi triển khai

### Mục tiêu

* **Đánh giá hiệu suất:** Xác định các chỉ số như xác suất thành công hoặc rủi ro thất bại (ví dụ: xác suất xảy ra va chạm).
* **Phân tích tính bền vững:** Kiểm tra mức độ phù hợp của chính sách khi có sự khác biệt giữa mô hình mô phỏng và thực tế.
* **Phát hiện các điểm yếu:** Xác định những kịch bản hoặc yếu tố có thể dẫn đến thất bại.

### Ví dụ tổng quan (Hệ thống tránh va chạm Collision Avoidance System)

#### Bối cảnh

Hệ thống tránh va chạm là một thành phần quan trọng trong phương tiện tự hành. Mục tiêu là xây dựng một chính sách (policy) giúp phương tiện di chuyển an toàn, tránh va chạm với các vật thể hoặc phương tiện khác trong quá trình hoạt động.

#### Mục tiêu

Hệ thống tránh va chạm là một thành phần quan trọng trong phương tiện tự hành. Mục tiêu là xây dựng một chính sách (policy) giúp phương tiện di chuyển an toàn, tránh va chạm với các vật thể hoặc phương tiện khác trong quá trình hoạt động.

#### Phương thức

Để có thể đánh giá tổng quan chính sách về tính hiệu quả và khả thi, ta sẽ cùng đi vào các thuật toán con trong Policy Validation

## [Performance Metric Evaluation](https://docs.google.com/document/d/16_xRPILqZy-X9NrctpDa-_o41mwoYtdM/edit?pli=1#heading=h.yit9ekk9jxl1)

### Khái niệm

Performance Metric Evaluation là một quy trình trong Policy Validation được sử dụng để đo lường, đánh giá, và so sánh hiệu suất của các chính sách (policies) được thiết kế nhằm giải quyết một bài toán cụ thể, ví dụ như điều khiển, tối ưu hóa, hay ra quyết định trong hệ thống phức tạp.  Là việc định lượng hiệu suất của chính sách dựa trên các tiêu chí cụ thể (thước đo hiệu suất - performance metrics), như:

* Tỷ lệ thành công (Success Rate)
* Tỷ lệ thất bại (Failure Rate)
* Xác suất va chạm (Collision Probability)
* Thời gian hoàn thành nhiệm vụ (Completion Time)
* Chi phí trung bình (Average Cost)

Performance Metric Evaluation là một quy trình trong Policy Validation được sử dụng để đo lường, đánh giá, và so sánh hiệu suất của các chính sách (policies) được thiết kế nhằm giải quyết một bài toán cụ thể, ví dụ như điều khiển, tối ưu hóa, hay ra quyết định trong hệ thống phức tạp.  Là việc định lượng hiệu suất của chính sách dựa trên các tiêu chí cụ thể (thước đo hiệu suất - performance metrics), như:

* Tỷ lệ thành công (Success Rate)
* Tỷ lệ thất bại (Failure Rate)
* Xác suất va chạm (Collision Probability)
* Thời gian hoàn thành nhiệm vụ (Completion Time)
* Chi phí trung bình (Average Cost)

### Mục đích của Performance Metric Evaluation

* **Định lượng hiệu suất:** Đo lường cách một chính sách π thực hiện công việc dựa trên các thước đo định nghĩa trước.
* **So sánh chính sách:** Giúp so sánh giữa các chính sách π1,π2,…,πn để xác định chính sách tốt nhất trong một tập hợp.
* **Xác minh tính khả thi:** Kiểm tra xem chính sách có đáp ứng các tiêu chí cần thiết không (như tính an toàn hoặc tối ưu hóa).
* **Tối ưu hóa:** Xác định chính sách nào đạt hiệu suất cao nhất với chi phí thấp nhất.

### Ứng dụng trong Policy Validation

Khi xây dựng một policy π (ví dụ: điều khiển robot, tối ưu hóa mạng lưới, hoặc hệ thống tránh va chạm), chúng ta cần xác thực nó bằng cách kiểm tra hiệu suất qua mô phỏng hoặc triển khai thực tế. Performance Metric Evaluation thực hiện các bước sau:

#### Xác định thước đo hiệu suất (Performance Metrics):

Các thước đo có thể bao gồm:

* Xác suất thành công (fsuccess(𝜋)): Tần suất một chính sách hoàn thành nhiệm vụ.
* Chi phí trung bình (fcost(𝜋)): Tổng chi phí hành động do chính sách sinh ra.
* Độ tin cậy (Reliability): Tính ổn định của chính sách trong các điều kiện bất định.

#### Thu thập dữ liệu quỹ đạo (Trajectories):

Mô phỏng hoặc triển khai chính sách trong môi trường để thu thập dữ liệu quỹ đạo :

Trong đó:

* + st​: Trạng thái tại thời điểm t.
  + at​: Hành động được thực hiện từ st​.

#### Quá trình Performance Metric Evaluation trong hệ thống

1. Chạy chính sách π trên môi trường mô phỏng hoặc thực tế để thu thập dữ liệu quỹ đạo τ
2. Tính toán các thước đo hiệu suất, ví dụ:

* Tỷ lệ va chạm:
* Thời gian hoàn thành trung bình:

1. So sánh fcollision(𝜋), ftime(𝜋) giữa các chính sách để chọn ra chính sách tốt nhất.

### Áp dụng

#### Bộ dữ liệu

Bộ dữ liệu trong chương này được khởi tạo thủ công bằng công thức và nội dung sau:

* 1: Spam
* 0: Không spam

Sau đó sẽ tạo ra hai bộ dữ liệu, mỗi bộ dữ liệu chứa 1000 records với 2 chính là:

* true\_label: Bộ dữ liệu chứa data đúng
* model\_predictions: Bộ dữ liệu mô phỏng dự đoán của mô hình

#### Các bước thực hiện

Bước 1: Khởi tạo dữ liệu trong hàm main.

A black background with numbers

Description automatically generated

Bước 2: Định nghĩa hàm evaluate\_model\_performance().



Bước 3: Khởi tạo biến.



Trong đó:

* n\_samples: Số lượng mẫu trong tập dữ liệu
* bootstrap\_acciracies: Một danh sách rỗng để lưu trữ độ chính xác của mỗi mẫu bootstrap.

Bước 4: Lấy Mẫu Bootstrap.

A screen shot of a computer program

Description automatically generated

Trong đó:

* Lặp n\_bootstrap lần để tạo các mẫu bootstrap.
* Trong mỗi lần lặp:
* Tạo các chỉ số ngẫu nhiên có phục hồi.
* Tạo các mẫu bootstrap cho true\_labels và predictions sử dụng các chỉ số này.
* Tính độ chính xác của mẫu bootstrap.
* Thêm độ chính xác vào danh sách bootstrap\_accuracies.

Bước 5: Tính các metric.

A screen shot of a computer code

Description automatically generated

Trong đó:

* mean\_accuracy: Độ chính xác trung bình của tất cả các mẫu bootstrap.
* std\_accuracy: Độ lệch chuẩn của các độ chính xác.
* se: Sai số chuẩn của độ chính xác trung bình.

Bước 6: Tính khoảng tin cậy 95%.

A screenshot of a computer

Description automatically generated

Tính giới hạn dưới và giới hạn trên của khoảng tin cậy 95% sử dụng sai số chuẩn và giá trị Z cho mức tin cậy 95% (1.96).

Bước 7: Tính sai số tương đối và cho ra kết quả.

A computer screen with text

Description automatically generated

Bước 8: Xuất ra kết quả

A screen shot of a computer code

Description automatically generated

## Rare Event Simulation

### Bối cảnh

Trong nhiều hệ thống như xe tự hành, hàng không, hay bảo mật mạng, có những rủi ro mà chúng ta muốn ước tính chính xác xác suất xảy ra để đảm bảo an toàn hoặc quản lý tài nguyên.

Ví dụ: Hệ thống tránh va chạm: Xác suất va chạm giữa hai phương tiện tự hành là rất nhỏ, nhưng nếu xảy ra, hậu quả có thể nghiêm trọng. ài chính: Xác suất một sự kiện khủng hoảng tài chính là rất thấp trong một năm, nhưng nếu xảy ra, nó có thể gây ra ảnh hưởng to lớn đến nền kinh tế.

Những sự kiện này xảy ra rất hiếm nhưng lại có tác động lớn. Nếu sử dụng phương pháp Monte Carlo cơ bản - tức là lấy mẫu ngẫu nhiên từ không gian trạng thái, thì chúng ta sẽ cần rất nhiều mẫu để ước lượng chính xác xác suất của các sự kiện này, gây ra sự lãng phí tài nguyên tính toán.

### Importance Sampling

Phương pháp này thực hiện hai bước chính:

* Lấy mẫu từ phân phối đề xuất P': Thay vì lấy mẫu từ phân phối thực P, chúng ta chọn một phân phối khác P' để lấy mẫu. Phân phối P' được lựa chọn sao cho tăng tần suất xuất hiện của các sự kiện mà ta quan tâm (các sự kiện hiếm), giúp thu thập nhiều dữ liệu liên quan đến sự kiện hiếm hơn.
* Điều chỉnh trọng số cho mỗi mẫu: Do lấy mẫu từ phân phối khác với phân phối thực, chúng ta cần điều chỉnh lại các mẫu để đảm bảo rằng các mẫu phản ánh đúng bản chất của phân phối thực. Việc điều chỉnh được thực hiện thông qua việc tính trọng số của từng mẫu dựa trên tỷ lệ của xác suất từ phân phối thực P so với phân phối đề xuất P'.

#### Các bước thực hiện

Thuật toán Importance Sampling chia làm 4 bước:

1. Chọn phân phối đề xuất P'

* Phân phối đề xuất P' cần được chọn sao cho khả năng tạo ra các sự kiện hiếm cao hơn phân phối thực P .
* Ví dụ, nếu chúng ta muốn ước tính xác suất va chạm trong hệ thống xe tự hành, P'có thể ưu tiên các trạng thái mà các phương tiện ở gần nhau hoặc tốc độ cao hơn.
* Điều này tăng khả năng xuất hiện của các tình huống mà sự kiện hiếm (va chạm) dễ xảy ra.

1. Lấy mẫu từ P'

* Lấy một tập mẫu từ phân phối đề xuất P'
* Mỗi mẫu ri đại diện cho một trạng thái hoặc hành động trong không gian trạng thái của hệ thống.

1. Tính trọng số cho mỗi mẫu

* Trọng số của một mẫu ri được tính bằng công thức:

Trong đó:

* P(ri) là x

Xác suất của mẫu ri theo phân phối thực P.

* P'(ri) là xác suất của mẫu ri theo phân phối thực P'.
* Ý nghĩa của việc tính trọng số là để "hiệu chỉnh" lại mẫu từ phân phối đề xuất sao cho phù hợp với phân phối thực. Nếu P và P' rất khác nhau, trọng số này sẽ lớn hoặc nhỏ hơn nhiều so với 1, giúp phản ánh sự khác biệt giữa các phân phối.

1. Tính toán kỳ vọng hoặc xác suất ước tính

Sau khi đã lấy được các mẫu và tính trọng số, chúng ta có thể ước lượng kỳ vọng của một hàm f(r) (ví dụ, xác suất xảy ra sự kiện) như sau:

Trong đó:

* f(ri) là giá trị của hàm f tại mẫu ri.
* w(ri) là trọng số điều chỉnh cho mỗi mẫu.

#### Ví dụ minh họa

Giả sử chúng ta có một hệ thống dự đoán khả năng va chạm của xe tự hành. Trong thực tế, xác suất xảy ra va chạm rất thấp, có thể chỉ ở mức . Dưới đây là các bước ứng dụng thuật toán Importance Sampling:

* Chọn phân phối đề xuấtP':
* Chọn một phân phối P' ưu tiên các tình huống mà khoảng cách giữa các xe là nhỏ, hoặc khi các xe có tốc độ cao hơn bình thường.
* Phân phối này không phản ánh hoàn toàn tình hình thực tế (vì trong thực tế khoảng cách giữa các xe thường an toàn hơn), nhưng nó giúp tạo ra nhiều mẫu gần với các tình huống có khả năng va chạm hơn.
* Lấy mẫu từ P'
* Lấy một tập mẫu từ P', mỗi mẫu là một trạng thái của hệ thống (ví dụ, vị trí và tốc độ của các xe trong một khoảng thời gian cụ thể).
* Số lượng mẫu lấy từ P' có thể ít hơn rất nhiều so với số mẫu cần thiết nếu lấy trực tiếp từ P.
* Tính trọng số:
* Với mỗi mẫu ri, tính trọng số để đảm bảo các mẫu được hiệu chỉnh theo phân phối gốc.
* Nếu một mẫu ở gần sự kiện hiếm (tức là khoảng cách giữa các xe nhỏ), trọng số sẽ điều chỉnh để đảm bảo sự kiện này không bị đánh giá quá cao trong ước lượng cuối cùng.
* Ước tính xác suất:

Sau khi đã có trọng số cho các mẫu, tính toán xác suất ước tính cho sự kiện va chạm bằng cách lấy trung bình có trọng số:

Trong đó 1va chạm ri là hàm chỉ báo sự kiện va chạm, trả về 1 nếu xảy ra va chạm, và 0 nếu không va chạm.

### Đánh giá độ tin cậy của ước lượng

Để đánh giá độ tin cậy của ước lượng, phương pháp này cũng sử dụng sai số chuẩn (standard error) và khoảng tin cậy (confidence interval). Điều này giúp đảm bảo rằng ước lượng không chỉ là không chệch mà còn đáng tin cậy, đặc biệt là trong các ứng dụng đòi hỏi tính chính xác cao như hệ thống an toàn giao thông hoặc hệ thống tài chính.

**Khoảng tin cậy:**

Để ước lượng mức độ chính xác của ước lượng xác suất sự kiện hiếm, ta có thể sử dụng khoảng tin cậy cho kết quả. Ví dụ, nếu ta ước lượng xác suất với độ tin cậy 95%, khoảng tin cậy có thể được xác định như sau:

Trong đó:

* p là ước lượng xác suất từ mẫu.
* Z là hệ số liên quan đến độ tin cậy mong muốn (1.96 cho độ tin cậy 95%).
* n là kích thước mẫu.
* là độ lệch chuẩn của ước lượng xác suất.

### Lợi ích và hạn chế

* Lợi ích:
* Tăng hiệu quả tính toán: Giảm số lượng mẫu cần thiết để đạt được độ chính xác cao, tiết kiệm tài nguyên.
* Tập trung vào sự kiện quan trọng: Giúp chúng ta tập trung vào các sự kiện hiếm mà không cần phải kiểm tra toàn bộ không gian mẫu.
* Hạn chế:
* Phụ thuộc vào phân phối đề xuất: Nếu phân phối đề xuất không được chọn hợp lý, ước lượng có thể thiếu chính xác hoặc bị lệch.
* Khó khăn trong việc xây dựng phân phối đề xuất: Để xây dựng phân phối P' tối ưu đòi hỏi kiến thức chuyên môn sâu về hệ thống và các yếu tố tác động đến sự kiện hiếm.

### Kết luận

Rare Event Simulation với thuật toán Importance Sampling là một phương pháp mạnh mẽ cho phép chúng ta ước lượng xác suất của các sự kiện rất hiếm, tiết kiệm thời gian và tài nguyên tính toán so với phương pháp lấy mẫu trực tiếp. Bằng cách lựa chọn phân phối đề xuất phù hợp và điều chỉnh trọng số cho các mẫu, chúng ta có thể đạt được các ước lượng không chệch và đáng tin cậy cho các ứng dụng yêu cầu độ chính xác cao như hệ thống an toàn, tài chính, và các hệ thống tự động khác.

### Áp dụng vào bộ dữ liệu

* Vẽ biểu đồ histogram của cột Exam\_Score

A graph of a person with a blue line

Description automatically generated

* Giả định rare event là điểm từ 55 đến 63( nhỏ hơnphân vị thứ 10), tạo cột rare\_event chứa 1 và 0 để xác định dòng có phải rare event không.

A screen shot of a computer program

Description automatically generated

* Vẽ biểu đồ correlations để quan sát độ tương quan giữa các cột.

A chart of numbers and symbols

Description automatically generated with medium confidence

* Danh sách selected\_features chứa tên các thuộc tính có mối tương quan đáng kể (lớn hơn 0.1 hoặc nhỏ hơn -0.1) với điểm thi Exam\_Score.

A screen shot of a computer program

Description automatically generated

* Sử dụng Linear Regression để dự đoán Exam\_Score.

A computer screen with text and numbers

Description automatically generated with medium confidence

* Cột predicted\_score là điểm Exam\_Score theo dự đoán.

A black background with white text

Description automatically generated

* MSE và R-squared của predicted\_score.

A screen shot of a computer code

Description automatically generated

A screen shot of a computer code

Description automatically generated

* Tính xác suất sự kiện hiếm thực tế trong toàn bộ dữ liệu.

A screen shot of a computer

Description automatically generated

* Phân phối đề xuất: Các học sinh có điểm dự đoán thấp hơn dưới ngưỡng phân vị thứ 10, sau đó lấy 500 mẫu ngẫu nhiên.

A screenshot of a computer

Description automatically generated

* p\_proposal\_rare\_event: Xác suất sự kiện hiếm trong phân phối đề xuất. Tính trọng số cho các mẫu trong phân phối đề xuất.

A screenshot of a computer

Description automatically generated

* Ước lượng xác suất của sự kiến hiếm

A screen shot of a computer program

Description automatically generated

Khoảng tin cậy 95% cho biết rằng, với độ tin cậy 95%, xác suất thực của sự kiện hiếm nằm trong khoảng 0.99% đến 16.82%.

## Robustness Analysis

### Tầm quan trọng của Robustness Analysis

Trước khi triển khai một hệ thống vào thực tế, điều quan trọng là phải nghiên cứu độ bền vững của hệ thống đối với các sai sót trong quá trình mô hình hóa. Việc này đảm bảo hệ thống hoạt động ổn định ngay cả khi mô hình không hoàn toàn khớp với thực tế.

Chúng ta có thể áp dụng các công cụ như policy evaluation và importance sampling, nhưng thay vào đó, cần đánh giá các policy trên các môi trường khác với mô hình được giả định trong quá trình tối ưu hóa.

A graph of a function

Description automatically generated

Hình 1: Ảnh hưởng của mô hình trong độ đo thực tế với h (m/s2) trong khi mô hình là 1 m/s2

Ta có thể thấy ở hình trên, hlimit chạy từ giá trị ) 0 cho đến 2:

* Tại vị trí 0 thì tỷ lệ va chạm (collision probability) đạt ngưỡng cao nhất (0.1)
* Điểm cân bằng (hlimit = 1) mà tại đó là khoảng giao giữa sự tăng giảm gia tốc đồng thời cũng là sự giao của tỷ lệ va chạm. Khi đường nét xanh qua giá trị này, ta có thể thấy tỷ lệ va chạm giảm mạnh. Điều đó cho thấy điểm tối ưu của mô hình sẽ là 1.
* + Khi hlimit chạy từ 1 cho đến 2, tỷ lệ va chạm giảm rõ rệt và gần như đạt ngưỡng thấp nhất (0.0001) từ 2 trở đi.Thể hiện rõ sự đối nghịch hoàn toàn giữa tỷ lệ va chạm và gia tốc tối thiểu.

=> Qua phân tích ta rút ra được rằng đại lượng gia tốc tối thiểu biến thiên theo tỷ lệ va chạm. Việc xây dựng mô hình chỉ đáp ứng cho giá trị gia tốc đầu vào hlimit = 1 là không hề tối ưu cho mô hình. Cần việc xây dựng một mô hình bền vững hơn. Điều đó thể hiện sự quan trọng của Robustness Analysis trong việc đánh giá Policy

Ngoài ra, chúng ta cũng có thể nghiên cứu độ nhạy của các chỉ số (metrics) đối với các giả định mô hình trong không gian trạng thái. Nếu các chỉ số duy trì hiệu suất ổn định trong các thay đổi hợp lý của môi trường, điều này tạo sự tự tin rằng hệ thống sẽ hoạt động như mong đợi khi triển khai.

### Mô hình lập kế hoạch và overfiting

Một cách tiếp cận phổ biến là sử dụng mô hình lập kế hoạch đơn giản hơn nhằm:

* Tránh overfitting với các giả định mô hình không phản ánh đúng thực tế.
* Giảm độ phức tạp tính toán, giúp quá trình tối ưu hóa trở nên hiệu quả hơn.

Ví dụ: Sử dụng một mô hình động học máy bay đơn giản với không gian trạng thái thấp và rời rạc để tạo policy tránh va chạm. Đánh giá chính sách này trên một mô phỏng liên tục, chi tiết và có độ trung thực cao.

Một mô hình lập kế hoạch đơn giản thường có khả năng chống chịu tốt hơn trước những sai lệch trong các đánh giá thực tế.

### Stress Testing

Đánh giá các policy trên nhiều mô hình đánh giá khác nhau, đặc biệt là trong các tình huống cực đoan, thường được gọi là stress testing.

Ví dụ:

Trong bài toán tránh va chạm (collision avoidance):

* Các kịch bản cực đoan có thể bao gồm tình huống hai máy bay đang hội tụ với tốc độ tăng độ cao lớn.
* Mặc dù không tối ưu hóa hệ thống cho các kịch bản phi thực tế này, việc hiểu các tình huống có thể dẫn đến lỗi hệ thống rất hữu ích trong giai đoạn thiết kế.

### Robust Dynamic Programming

Nếu policy quá nhạy cảm với các giả định mô hình, ta có thể sử dụng robust dynamic programming.

Thay vì sử dụng một mô hình chuyển trạng thái cụ thể, phương pháp này:

* Định nghĩa một tập hợp các mô hình chuyển trạng thái và reward.
* Sửa đổi phương trình Bellman để xem xét tính bền vững của các policy trước các mô hình khác nhau.

Công thức Bellman đã được update:

Trong đó:

* : Giá trị của trạng thái sss tại bước k+ 1
* min(i): Xem xét kịch bản xấu nhất trong tập hợp các mô hình i
* max(a): Tìm hành động a tốt nhất (tức là hành động tối ưu) tại trạng thái s dựa trên giá trị hàm phần thưởng và giá trị của các trạng thái tiếp theo.
* : phần thưởng khi thực hiện hành động a tại trạng thái s cho mô hình i.
* γ: Hệ số chiết khấu (discount factor), kiểm soát ảnh hưởng của phần thưởng trong tương lai.
* Xác suất chuyển từ trạng thái s sang trạng thái s′ khi thực hiện hành động a.
* giá trị của trạng thái s′ ′ tại bước k.

### Kết luận

Robustness Analysis là quy trình kiểm tra độ bền vững của một mô hình trước các sai lệch hoặc biến động trong dữ liệu và giả định ban đầu.

* Sử dụng các mô hình lập kế hoạch đơn giản giúp giảm thiểu rủi ro overfitting và đảm bảo khả năng chịu đựng của hệ thống.
* Stress testing và robust dynamic programming là những phương pháp hữu ích để đảm bảo hệ thống hoạt động ổn định trong các môi trường đa dạng và không hoàn hảo.

### Áp dụng vào bộ dữ liệu

* Chia dữ liệu thành hai tập Feature và Target. Sau đó sử dụng Linear Regression để build model và tiến hành dữ đoán kết quả.

A screenshot of a computer program

Description automatically generated

* Đánh giá độ ổn định của mô hình trước và sau khi thêm dữ liệu nhiễu bằng độ đo Mean Squared Error

A screenshot of a computer program

Description automatically generated

=> MSE có sự tăng nhẹ (từ 5.414 lên 5.474), cho thấy rằng mô hình không hoàn toàn ổn định khi có nhiễu. Mặc dù sự thay đổi không quá lớn, nhưng nó vẫn cho thấy rằng nhiễu làm giảm hiệu quả của mô hình.

* Áp dụng công thức Bellman đã được update:

A screen shot of a computer program

Description automatically generated

Trong bài toán, muốn xác định giá trị kỳ vọng tối ưu cho các trạng thái học tập của học sinh (Low, Medium, High) bằng cách thực hiện các hành động như tăng thời gian học (Increase\_Hours) hoặc cải thiện việc tham gia học tập (Improve\_Attendance), đồng thời đối mặt với sự không chắc chắn từ hai mô hình khác nhau.

Quá trình tính toán bắt đầu với giá trị kỳ vọng khởi tạo là 0 cho mỗi trạng thái. Ở mỗi vòng lặp, thuật toán cập nhật giá trị này bằng cách:

* Tính phần thưởng và giá trị kỳ vọng cho mỗi hành động dựa trên các mô hình khác nhau.
* Lấy giá trị nhỏ nhất từ các mô hình để đảm bảo phương pháp tiếp cận an toàn, giảm thiểu rủi ro từ mô hình xấu nhất.
* Sựa chọn hành động tốt nhất mang lại giá trị lớn nhất cho trạng thái hiện tại.

Sau 10 vòng lặp, thuật toán hội tụ, trả về giá trị kỳ vọng tối ưu cho mỗi trạng thái. Kết quả này có thể dùng để hướng dẫn hành động nào sẽ giúp cải thiện động lực học tập hiệu quả nhất trong môi trường không chắc chắn.

## Trade Analysis

### Khái niệm

Khi phát triển một hệ thống, có rất nhiều nhiệm vụ, mục tiêu. Các mục tiêu đôi  khi

Phân tích đánh đổi (trade analysis) là một phương pháp giúp đánh giá và tối ưu hóa các hệ thống có nhiều mục tiêu mâu thuẫn. Trong các hệ thống tự động, một ví dụ điển hình là sự cân bằng giữa an toàn và hiệu suất.

* **An toàn**: Đại diện cho các yếu tố như xác suất va chạm hoặc rủi ro liên quan.
* **Hiệu suất**: Bao gồm các yếu tố như tốc độ phản hồi, số lần thay đổi khuyến nghị (advisory changes), hoặc mức độ can thiệp.

Khi thiết kế các chính sách, việc tối ưu hóa một mục tiêu thường đi kèm với chi phí làm giảm hiệu quả của mục tiêu khác. Phân tích đánh đổi được thực hiện để hiểu và lựa chọn các giải pháp phù hợp với yêu cầu thực tế của hệ thống.

### Ứng dụng của phân tích đánh đổi

1. **Tối ưu hóa hệ thống**: Phân tích đánh đổi giúp xác định giải pháp tốt nhất để cân bằng giữa các mục tiêu, ví dụ như tăng cường an toàn mà không làm giảm hiệu suất quá mức.
2. **So sánh các phương pháp**: Bằng cách vẽ các đường đánh đổi (trade-off curves), người thiết kế có thể so sánh hiệu quả của các phương pháp khác nhau, như sử dụng lập trình động so với các chính sách đơn giản dựa trên ngưỡng.
3. **Loại bỏ các giải pháp kém hiệu quả**: Những chính sách không hiệu quả (dominated policies) – tức bị áp đảo ở cả hai chỉ số – có thể dễ dàng loại bỏ để tập trung vào các giải pháp tối ưu hơn

Chúng ta có thể nhìn qua 1 ví dụ về việc ứng dụng 2 mô hình khác nhau và sẽ thấy được sự

### Đường Pareto trong phân tích đánh đổi

Khái niệm: Đường Pareto là một công cụ quan trọng trong phân tích đánh đổi, đại diện cho tập hợp các giải pháp Pareto tối ưu. Một giải pháp Pareto tối ưu là giải pháp mà không có bất kỳ cách nào để cải thiện một mục tiêu mà không làm xấu đi mục tiêu khác.

* **Biên Pareto (Pareto Frontier)**: Tập hợp các điểm biểu diễn các giải pháp không bị chi phối (dominated).
* **Giải pháp chi phối (Dominated Solutions)**: Các giải pháp kém hiệu quả hơn ở cả hai mục tiêu so với ít nhất một giải pháp khác.

Trong không gian hai chiều, đường Pareto giúp trực quan hóa và xác định các giải pháp hiệu quả.

### Ví dụ

Một ví dụ cụ thể về đường Pareto trong một hệ thống tránh va chạm. Biểu đồ này biểu diễn sự đánh đổi giữa xác suất va chạm và số lần thay đổi khuyến nghị.

A graph with blue dots

Description automatically generated

Hình 2: Biểu đồ Pareto biểu diễn sự đánh đổi giữa xác suất va chạm và số lần thay đổi khuyến nghị.

1. Ý nghĩa của biểu đồ:

* Trục hoành: Số lần thay đổi khuyến nghị (Advisory changes) – đại diện cho hiệu suất.
* Trục tung: Xác suất va chạm (Collision probability) – đại diện cho an toàn.
* Đường màu xanh: Biên Pareto gần đúng, biểu diễn tập hợp các giải pháp tối ưu.

1. Quan sát:

* Giai đoạn đầu (phần dốc của đường): Tăng nhẹ số lần thay đổi khuyến nghị giúp giảm mạnh xác suất va chạm, thể hiện mức cải thiện lớn về an toàn.
* Giai đoạn sau (phần phẳng của đường): Khi số lần thay đổi vượt quá khoảng 1.5, lợi ích trong việc giảm xác suất va chạm giảm dần (hiện tượng lợi nhuận giảm dần).
* Chính sách chi phối và không chi phối: Các điểm nằm bên trái hoặc bên dưới đường Pareto là chính sách bị chi phối (dominated), vì có các giải pháp tốt hơn trên đường Pareto.

1. Ứng dụng thực tế:

* Các giải pháp trên đường Pareto phù hợp với các mức độ ưu tiên khác nhau, từ nhấn mạnh an toàn đến tối ưu hiệu suất.
* Người thiết kế cần cân nhắc yêu cầu thực tế của hệ thống để chọn vị trí phù hợp trên đường Pareto, chẳng hạn:
* Nếu cần giảm xác suất va chạm ở mức tối thiểu, chọn chính sách gần trục tung.
* Nếu cần giảm tối đa số lần thay đổi khuyến nghị, chọn chính sách gần trục hoành.

**Kết luận**

Đường Pareto là công cụ quan trọng trong phân tích đánh đổi, giúp tập trung vào các chính sách hiệu quả và loại bỏ các giải pháp kém hiệu quả. Hình 14.5 minh họa rõ ràng cách hệ thống có thể cân bằng giữa an toàn và hiệu suất. Bằng cách dựa vào đường Pareto, người thiết kế có thể đưa ra các quyết định phù hợp với mục tiêu vận hành của hệ thống, tối ưu hóa kết quả mà không hy sinh quá mức bất kỳ mục tiêu nào.

## Adversarial Analysis

### Khái niệm

Phân tích đối nghịch là cách nghiên cứu độ mạnh (robustness) của một chính sách từ góc độ "đối thủ".

Tại mỗi bước thời gian, có một đối thủ sẽ chọn trạng thái tiếp theo sau khi áp dụng hành động từ chính sách hiện tại.

Đối thủ có 2 mục tiêu cần cân bằng:

* Cực tiểu hóa phần thưởng của chúng ta.
* Cực đại hóa xác suất của quỹ đạo kết quả theo mô hình chuyển tiếp.

### Công thức phần thưởng đối nghịch

Trong đó:

* π là chính sách của chúng ta.
* R là hàm phần thưởng gốc.
* T là mô hình chuyển tiếp gốc..
* λ là tham số kiểm soát độ quan trọng của việc cực đại hóa xác suất quỹ đạo.
* Đối thủ cố gắng tối đa hóa tổng của:
* Phần thưởng tiêu cực của chúng ta (-R).
* λ lần logarit của xác suất quỹ đạo.

### Ứng dụng thực tế

Trong bài toán tránh va chạm máy bay:

* Cần cân bằng giữa xác suất va chạm.
* Với các thông số khác như số lượng thay đổi cố vấn.
* Có thể tính toán bằng cách đánh giá chính sách chính xác.

Phương pháp này giúp kiểm tra độ mạnh của chính sách trong điều kiện xấu nhất, khi có một "đối thủ" cố tình chọn các trạng thái bất lợi nhất có thể.

**Biểu đồ:**

A graph with different colored lines

Description automatically generated

Hình 3: Biểu đồ phân tích sự đánh đổi giữa an toàn và hiệu quả vận hành khi thay đổi các tham số của các hệ thống tránh va chạm khác nhau.

**Giải thích biểu đồ:**

* Trục x: Số lượng thay đổi cố vấn (Advisory changes).
* Trục y: Xác suất va chạm (Collision probability).
* Có 3 đường cong khác nhau:
* Đường xanh lá: Simple(θshresh) - Chính sách đơn giản với tham số ngưỡng.
* Đường xanh dương: Simple(θreach) - Chính sách đơn giản với tham số tiếp cận.
* Đường tím: Optimal(β) - Chính sách tối ưu với tham số β.

**Phân tích đặc điểm:**

* Tất cả các đường đều cho thấy sự đánh đổi giữa xác suất va chạm và số lượng thay đổi cố vấn.
* Khi số lượng thay đổi cố vấn tăng, xác suất va chạm giảm.
* Chính sách tối ưu (đường tím) cho kết quả tốt nhất: với cùng số lượng thay đổi cố vấn, nó có xác suất va chạm thấp hơn.

**Ý nghĩa của tham số β:**

* Khi β gần bằng 1: Hệ thống rất an toàn. Nhưng phải chấp nhận nhiều thay đổi cố vấn hơn.
* Khi β tiến về 0:  Độ an toàn giảm. Nhưng ít thay đổi cố vấn hơn.
* Chi phí của việc thay đổi cố vấn là -(1-β).

**Kết luận:**

* Chính sách tối ưu vượt trội hơn các chính sách đơn giản có tham số.
* Với một mức độ an toàn cụ thể, chính sách tối ưu cần ít thay đổi cố vấn hơn so với các chính sách đơn giản.
* Có thể điều chỉnh sự cân bằng giữa an toàn và số lượng thay đổi cố vấn thông qua tham số β.

### Thuật toán

A screenshot of a math equation

Description automatically generated

Hình 4: Hàm adversarial.

1. Mục đích của thuật toán:

* Chuyển đổi bài toán gốc thành bài toán đối nghịch (adversarial problem).
* Tạo ra một MDP (Markov Decision Process) mới với phần thưởng đối nghịch.

1. Các điều kiện tiên quyết:

* Không gian trạng thái và hành động phải rời rạc.
* Có thể giải quyết bằng các thuật toán quy hoạch động.
* Giải pháp là chính sách ánh xạ từ trạng thái sang trạng thái.

1. Giải thích từng phần:

* Input parameters:
* 𝒫: MDP gốc.
* π: Chính sách cần đánh giá.
* λ: Tham số kiểm soát độ quan trọng của xác suất quỹ đạo.
* Khởi tạo:
* Lấy các thành phần từ MDP gốc (𝒮', T', R', γ).
* Tạo ma trận phần thưởng R' và ma trận chuyển tiếp T' mới.
* Vòng lặp chính:
* Với mỗi trạng thái s và hành động a
* Tính phần thưởng đối nghịch:
* Cập nhật xác suất chuyển tiếp: T'[s,a,a] = 1

1. Đặc điểm quan trọng:

* Bài toán là tất định (deterministic): mỗi hành động dẫn đến một trạng thái cụ thể.
* Có thể giải quyết như một bài toán tìm kiếm.

1. Xử lý với không gian lớn:

* Nếu bài toán có số chiều cao hoặc liên tục.
* Có thể sử dụng các kỹ thuật xấp xỉ từ chương.

1. Output:

Trả về một MDP mới với:

* Ma trận chuyển tiếp T' đã cập nhật
* Ma trận phần thưởng R' đối nghịch
* Hệ số chiết khấu γ giữ nguyên

Thuật toán này quan trọng trong việc đánh giá độ mạnh của chính sách bằng cách mô phỏng "kịch bản xấu nhất" thông qua việc tạo ra một đối thủ cố tình chọn các trạng thái bất lợi.

# TỔNG KẾT

Xác thực chính sách là một bước quan trọng trong quá trình tối ưu hóa chính sách, giúp đánh giá hiệu quả của các chính sách trong môi trường thực tế. Có nhiều phương pháp để thực hiện xác thực chính sách, bao gồm kỹ thuật lập trình động (dynamic programming) hoặc các phương pháp thông qua việc lấy mẫu từ các cuộn (sampling rollouts). Để đánh giá độ tin cậy của các chỉ số hiệu suất, chúng ta có thể sử dụng sai số chuẩn (standard error), khoảng tin cậy (confidence intervals), hoặc các phương pháp Bayes như đã đề cập trước đây.

Đối với các sự kiện hiếm, việc ước lượng xác suất có thể được thực hiện hiệu quả hơn thông qua phương pháp lấy mẫu quan trọng (importance sampling), trong đó lấy mẫu từ một phân phối thay thế và trọng số kết quả một cách thích hợp. Tuy nhiên, vì mô hình được sử dụng trong tối ưu hóa có thể không phản ánh chính xác thế giới thực, điều quan trọng là phải nghiên cứu độ nhạy của chính sách đối với các giả định mô hình.

Lập trình động mạnh mẽ (robust dynamic programming) có thể giúp cải thiện khả năng chống chịu với sự không chắc chắn của mô hình bằng cách tối ưu hóa với một tập hợp các mô hình chuyển trạng thái và phần thưởng khác nhau. Phân tích sự trao đổi (trade analysis) giúp xác định cách cân bằng nhiều mục tiêu hiệu suất trong quá trình tối ưu hóa chính sách. Cuối cùng, phân tích đối kháng (adversarial analysis) liên quan đến việc đối kháng chọn lựa trạng thái mà chúng ta chuyển đến tại mỗi bước, với mục tiêu tối thiểu hóa mục tiêu của chúng ta trong khi tối đa hóa khả năng xảy ra của quỹ đạo.

Việc kết hợp các phương pháp này không chỉ giúp tối ưu hóa chính sách mà còn đảm bảo chính sách hoạt động tốt trong các tình huống không chắc chắn và các điều kiện thực tế khác nhau.

# TÀI LIỆU THAM KHẢO

1. The MIT Press, Massachusetts Institute of Technology. (2024, June 18). Book Details - MIT Press. MIT Press. <https://mitpress.mit.edu/9780262047012/algorithms-for-decision-making/>
2. Asmussen, S., & Glynn, P. W. (2007). Stochastic Simulation: Algorithms and analysis. In *Applications of mathematics*. <https://doi.org/10.1007/978-0-387-69033-9>
3. Arief, M., Bai, Y., Ding, W., He, S., Huang, Z., Lam, H., & Zhao, D. (2021, November 3). Certifiable Deep Importance sampling for Rare-Event simulation of Black-Box systems. arXiv.org. <https://arxiv.org/abs/2111.02204>