भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Stochastic Processes

Application of Metropolis–Hastings to Bayesian Logistic Regression, Test for Irreducibility, Sample Paths of a Random Process in Simulations

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

29 April 2025

Bayesian Logistic Regression using Challenger O-Ring Dataset

- Between 1981 and 1985, NASA launched 23 space-shuttle missions before the ill-fated Challenger flight on 28 Jan 1986

Bayesian Logistic Regression using Challenger O-Ring Dataset

- Between 1981 and 1985, NASA launched 23 space-shuttle missions before the ill-fated Challenger flight on 28 Jan 1986
- For every flight $i \in \{1, \ldots, 23\}$, engineers recorded

Bayesian Logistic Regression using Challenger O-Ring Dataset

- Between 1981 and 1985, NASA launched 23 space-shuttle missions before the ill-fated Challenger flight on 28 Jan 1986
- For every flight $i \in \{1, \ldots, 23\}$, engineers recorded
  - the ambient air-temperature at lift-off x_i (in °F), and

Bayesian Logistic Regression using Challenger O-Ring Dataset

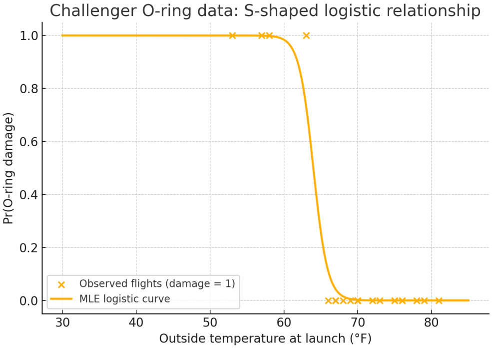
- Between 1981 and 1985, NASA launched 23 space-shuttle missions before the ill-fated Challenger flight on 28 Jan 1986
- For every flight $i \in \{1, \ldots, 23\}$, engineers recorded
  - the ambient air-temperature at lift-off x_i (in °F), and
  - $y_i = \mathbf{1}\{\text{whether at least one solid-rocket-booster O-ring was damaged during ascent}\}$

Bayesian Logistic Regression using Challenger O-Ring Dataset

- Between 1981 and 1985, NASA launched 23 space-shuttle missions before the ill-fated Challenger flight on 28 Jan 1986
- For every flight $i \in \{1, \ldots, 23\}$, engineers recorded
  - the ambient air-temperature at lift-off x_i (in °F), and
  - $y_i = \mathbf{1}\{\text{whether at least one solid-rocket-booster O-ring was damaged during ascent}\}$
- The lowest temperature available in the dataset was 53°F, whereas Challenger launched at 31°F despite warnings from engineers

Observed Empirical Relationship

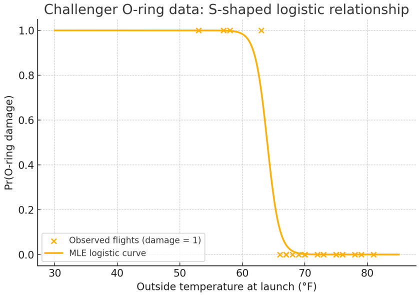
- Empirical relationship between probability of ring damage and temperature was observed to be as shown below.



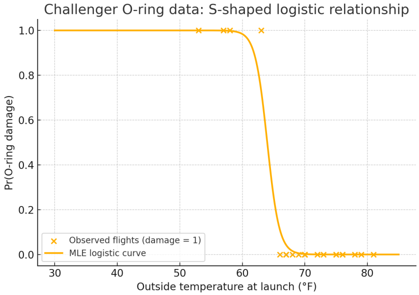
Regression Model

- For data point i, let

$$\eta_i = \beta_0 + \beta_1 x_i.$$



Regression Model



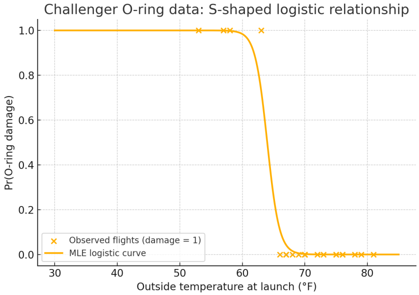
- For data point i, let

$$\eta_i = \beta_0 + \beta_1 x_i.$$

- Taking $\beta = (\beta_0, \beta_1)$, let

$$\pi_i = \text{Logit}(\eta_i) = \frac{\exp \eta_i}{1 + \exp \eta_i}.$$

Regression Model



- For data point i, let

$$\eta_i = \beta_0 + \beta_1 x_i.$$

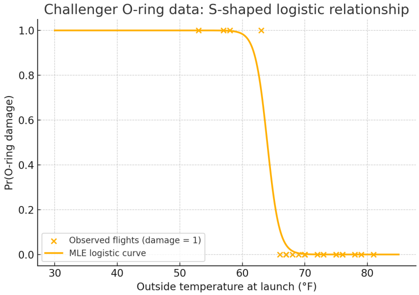
- Taking $\beta = (\beta_0, \beta_1)$, let

$$\pi_i = \text{Logit}(\eta_i) = \frac{\exp \eta_i}{1 + \exp \eta_i}.$$

- Conditional distribution of one output data point:

$$\mathbb{P}(y_i = 1 \mid x_i, \beta) = \pi_i = 1 - \mathbb{P}(y_i = 0 \mid x_i, \beta).$$

Regression Model



- For data point i, let

$$\eta_i = \beta_0 + \beta_1 x_i.$$

- Taking $\beta = (\beta_0, \beta_1)$, let

$$\pi_i = \mathrm{Logit}(\eta_i) = \frac{\exp \eta_i}{1 + \exp \eta_i}.$$

- Conditional distribution of one output data point:

$$\mathbb{P}(y_i = 1 \mid x_i, \beta) = \pi_i = 1 - \mathbb{P}(y_i = 0 \mid x_i, \beta).$$

- Data points $\{(x_i, y_i)\}_{i=1}^{23}$ are IID

Prior, Likelihood, and Posterior

- $\beta = (\beta_0, \beta_1)$ unknown. A reasonable prior is

$$\beta_0, \beta_1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \qquad g(\beta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}\right).$$

Prior, Likelihood, and Posterior

- $\beta = (\beta_0, \beta_1)$ unknown. A reasonable prior is

$$\beta_0, \beta_1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \qquad g(\beta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}\right).$$

- Likelihood function is given by

$$p(\mathbf{y} \mid \mathbf{x}, \beta) = \prod_{i=1}^{23} p(y_i \mid x_i, \beta) = \prod_{i=1}^{23} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}.$$

Prior, Likelihood, and Posterior

- $\beta = (\beta_0, \beta_1)$ unknown. A reasonable prior is

$$\beta_0, \beta_1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \qquad g(\beta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}\right).$$

- Likelihood function is given by

$$p(\mathbf{y} \mid \mathbf{x}, \beta) = \prod_{i=1}^{23} p(y_i \mid x_i, \beta) = \prod_{i=1}^{23} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}.$$

- Posterior function is given by

$$f(\beta \mid \mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x}, \beta)\, g(\beta)}{C}, \qquad C = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{x}, \beta')\, g(\beta')\, \mathrm{d}\beta_0'\, \mathrm{d}\beta_1'.$$

Prior, Likelihood, and Posterior

- $\beta = (\beta_0, \beta_1)$ unknown. A reasonable prior is

$$\beta_0, \beta_1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \qquad g(\beta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\beta_0^2 + \beta_1^2}{2\sigma^2}\right).$$

- Likelihood function is given by

$$p(\mathbf{y} \mid \mathbf{x}, \beta) = \prod_{i=1}^{23} p(y_i \mid x_i, \beta) = \prod_{i=1}^{23} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}.$$

- Posterior function is given by

$$f(\beta \mid \mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x}, \beta)\, g(\beta)}{C}, \qquad C = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{x}, \beta')\, g(\beta')\, \mathrm{d}\beta_0'\, \mathrm{d}\beta_1'.$$

- **No closed-form expression for C**

Setting up Metropolis–Hastings

- Set Q matrix (kernel) as
$$Q(\beta, \cdot) = \mathcal{N}(\beta, \Sigma),$$
for some fixed covariance matrix Σ

Setting up Metropolis–Hastings

- Set Q matrix (kernel) as

$$Q(\beta, \cdot) = \mathcal{N}(\beta, \Sigma),$$

  for some fixed covariance matrix Σ
- Q is symmetric

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Setting up Metropolis–Hastings

- Set Q matrix (kernel) as
$$Q(\beta, \cdot) = \mathcal{N}(\beta, \Sigma),$$
for some fixed covariance matrix Σ
- Q is symmetric
- The acceptance probabilities are then given by
$$A_{\beta,\beta'} = \min\left\{1, \frac{f(\beta' \mid \mathbf{x}, \mathbf{y}) \cdot Q(\beta', \beta)}{f(\beta \mid \mathbf{x}, \mathbf{y}) \cdot Q(\beta, \beta')}\right\} = \min\left\{1, \frac{p(\mathbf{y} \mid \mathbf{x}, \beta') \cdot g(\beta')}{p(\mathbf{y} \mid \mathbf{x}, \beta) \cdot g(\beta)}\right\}.$$

Setting up Metropolis–Hastings

- Set Q matrix (kernel) as
$$Q(\beta, \cdot) = \mathcal{N}(\beta, \Sigma),$$
for some fixed covariance matrix Σ
- Q is symmetric
- The acceptance probabilities are then given by
$$A_{\beta,\beta'} = \min\left\{1, \frac{f(\beta' \mid \mathbf{x}, \mathbf{y}) \cdot Q(\beta', \beta)}{f(\beta \mid \mathbf{x}, \mathbf{y}) \cdot Q(\beta, \beta')}\right\} = \min\left\{1, \frac{p(\mathbf{y} \mid \mathbf{x}, \beta') \cdot g(\beta')}{p(\mathbf{y} \mid \mathbf{x}, \beta) \cdot g(\beta)}\right\}.$$
- For any bounded, continuous function h, we have by the ergodic theorem that
$$\frac{1}{n}\sum_{k=1}^{n} h(X_k) \quad \xrightarrow{\text{a.s.}} \quad \int h(\beta) f(\beta \mid \mathbf{x}, \mathbf{y}) \, \mathrm{d}\beta,$$
where $\{X_k\}_{k=0}^{\infty}$ is a DTMC with stationary kernel $f(\cdot \mid \mathbf{x}, \mathbf{y})$

Other Tidbits

Test for Irreducibility

Theorem (Test for Irreducibility)

A row stochastic matrix P of size $d \times d$ is irreducible if and only if

$$\sum_{k=0}^{d-1} P^k > \mathbf{0},$$

where $P^0 = I$, $\mathbf{0}$ denotes the all-zeros matrix of size $d \times d$, and the inequality is entry-wise.

A simple algorithm to check for irreducibility:

- Start with $M_0 = I$.

Test for Irreducibility

Theorem (Test for Irreducibility)

A row stochastic matrix P of size $d \times d$ is irreducible if and only if

$$\sum_{k=0}^{d-1} P^k > \mathbf{0},$$

where $P^0 = I$, $\mathbf{0}$ denotes the all-zeros matrix of size $d \times d$, and the inequality is entry-wise.

A simple algorithm to check for irreducibility:

- Start with $M_0 = I$.
- Recursively define

$$M_{\ell+1} = I + PM_\ell.$$

Test for Irreducibility

Theorem (Test for Irreducibility)

A row stochastic matrix P of size $d \times d$ is irreducible if and only if

$$\sum_{k=0}^{d-1} P^k > \mathbf{0},$$

where $P^0 = I$, $\mathbf{0}$ denotes the all-zeros matrix of size $d \times d$, and the inequality is entry-wise.

A simple algorithm to check for irreducibility:

- Start with $M_0 = I$.
- Recursively define
$$M_{\ell+1} = I + PM_\ell.$$
- If $M_\ell > \mathbf{0}$ for any $1 \leq \ell \leq d-1$, then P is irreducible.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Sample Path of a Random Process in Simulations

```python
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.animation import FuncAnimation
from scipy.stats import bernoulli, norm, gaussian_kde
from IPython.display import HTML

# Parameters of Bernoulli distribution
p = 0.5
mu = p
sigma = np.sqrt(p * (1 - p))

# Sample Size Progression
x_range = np.linspace(-4, 4, 500)
x_val = np.pi / np.sqrt(3)
sample_sizes = np.linspace(10, 1000, 50, dtype=int)

# True standard normal PDF
true_pdf = norm.pdf(x_range)

# Initialize figure
fig, ax = plt.subplots(figsize=(10, 6))
line_kde, = ax.plot([], [], label='Estimated KDE PDF', color='blue')
line_true, = ax.plot(x_range, true_pdf, '--', label='Standard Normal', color='black')
vline = ax.axvline(x_val, color="gray", linestyle=":")
title = ax.set_title('')
ax.set_xlim(-4, 4)
ax.set_ylim(0, 0.5)
ax.legend()


# Update function for animation
def update(n):
    samples = bernoulli.rvs(p, size=(10000, n))
    means = np.mean(samples, axis=1)
    Z = (means - mu) / (sigma / np.sqrt(n))
    kde = gaussian_kde(Z)
    kde_vals = kde(x_range)
    line_kde.set_data(x_range, kde_vals)
    est_pdf_at_x = kde(x_val).item()
    true_pdf_at_x = norm.pdf(x_val)
    title.set_text(f"n={n}, PDF@x≈{x_val:.2f} ≈ {est_pdf_at_x:.4f}, N(0,1) PDF@x ≈ {true_pdf_at_x:.4f}")
    return line_kde, line_true, vline, title

# Create animation
ani = FuncAnimation(fig, update, frames=sample_sizes, blit=True, interval=200)
HTML(ani.to_jshtml())
```

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Sample Path of a Random Process in Simulations

```python
# Generate an array of 10 numbers according to a unit Gaussian distribution
rng.standard_normal(10)
```

```
array([ 0.52788774, -0.89551364,  0.78844006,  0.19032279,  0.92833631,
        0.00333179, -1.08593089, -0.52056583,  1.06936692, -0.51129284])
```

Question

Can we assume that the generated samples correspond to a given $\omega \in \Omega$?
That is, can we assume that

$$X_1(\omega) = 0.52788774, \;\; X_2(\omega) = -0.89551364, \ldots$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Back to Basics

$$(\Omega, \mathscr{F}, \mathbb{P}) \quad + \quad X \qquad \longrightarrow \qquad F_X$$

$$\begin{tikzcd}
(\Omega, \mathscr{F}, \mathbb{P}) \arrow[r, "X"] & (\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P}_X) \arrow[d] \\
X^{-1}((-\infty, x]) \arrow[u] & (-\infty, x] \arrow[l, "\text{pre-image}"']
\end{tikzcd}$$

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X^{-1}((-\infty, x])), \quad x \in \mathbb{R}$$

Back to Basics

$$(\Omega, \mathscr{F}, \mathbb{P}) \quad + \quad X \quad \longrightarrow \quad F_X$$

$$\begin{tikzcd}
(\Omega, \mathscr{F}, \mathbb{P}) \arrow[r, "X"] & (\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P}_X) \arrow[d] \\
X^{-1}((-\infty, x]) \arrow[u] & (-\infty, x] \arrow[l, "\text{pre-image}"']
\end{tikzcd}$$

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X^{-1}((-\infty, x])), \quad x \in \mathbb{R}$$

Proposition (From CDF to Probability Space and Random Variable)

Given a function $F : \mathbb{R} \to [0, 1]$ satisfying the properties of a CDF, there exists $(\Omega', \mathscr{F}', \mathbb{P}')$ and a random variable $X' : \Omega' \to \mathbb{R}$ such that $F_{X'} = F$.

Significance of Proposition

- In most practical scenarios, we are not given $(\Omega, \mathscr{F}, \mathbb{P})$ and X to begin with

Indian Institute of Technology Hyderabad

Significance of Proposition

- In most practical scenarios, we are not given $(\Omega, \mathscr{F}, \mathbb{P})$ and X to begin with
- Despite this restriction, we generate samples from a custom CDF on computer, and treat these samples as output of a random variable (process)

Significance of Proposition

- In most practical scenarios, we are not given $(\Omega, \mathscr{F}, \mathbb{P})$ and X to begin with
- Despite this restriction, we generate samples from a custom CDF on computer, and treat these samples as output of a random variable (process)
- The natural question is: is this treatment justified?

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Significance of Proposition

- In most practical scenarios, we are not given $(\Omega, \mathscr{F}, \mathbb{P})$ and X to begin with
- Despite this restriction, we generate samples from a custom CDF on computer, and treat these samples as output of a random variable (process)
- The natural question is: is this treatment justified?
- The preceding theorem says: **YES!**

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F
- We would like to associate each sample to one random variable (say x_i to a RV X_i), and treat the samples as the outputs of the random variables

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F
- We would like to associate each sample to one random variable (say x_i to a RV X_i), and treat the samples as the outputs of the random variables
- However, it would not be desirable to have

$$x_1 = X_1(\omega_1), \quad x_2 = X_2(\omega_2), \ldots,$$

because there exists no theory for analysing the behaviour of such samples

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F
- We would like to associate each sample to one random variable (say x_i to a RV X_i), and treat the samples as the outputs of the random variables
- However, it would not be desirable to have

$$x_1 = X_1(\omega_1), \quad x_2 = X_2(\omega_2), \ldots,$$

  because there exists no theory for analysing the behaviour of such samples
- What we desire is

$$x_1 = X_1(\omega), \quad x_2 = X_2(\omega), \ldots,$$

  a scenario wherein all samples correspond to a given ω

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F
- We would like to associate each sample to one random variable (say x_i to a RV X_i), and treat the samples as the outputs of the random variables
- However, it would not be desirable to have

$$x_1 = X_1(\omega_1), \quad x_2 = X_2(\omega_2), \ldots,$$

  because there exists no theory for analysing the behaviour of such samples
- What we desire is

$$x_1 = X_1(\omega), \quad x_2 = X_2(\omega), \ldots,$$

  a scenario wherein all samples correspond to a given ω
- Is there a feasible fit for the latter viewpoint?

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F
- We would like to associate each sample to one random variable (say x_i to a RV X_i), and treat the samples as the outputs of the random variables
- However, it would not be desirable to have

$$x_1 = X_1(\omega_1), \quad x_2 = X_2(\omega_2), \ldots,$$

  because there exists no theory for analysing the behaviour of such samples
- What we desire is

$$x_1 = X_1(\omega), \quad x_2 = X_2(\omega), \ldots,$$

  a scenario wherein all samples correspond to a given ω
- Is there a feasible fit for the latter viewpoint?

From Single Random Variable to a Random Process

- Suppose that we generate n samples $x_1, x_2, \ldots, x_n$, all from a custom CDF F
- We would like to associate each sample to one random variable (say x_i to a RV X_i), and treat the samples as the outputs of the random variables
- However, it would not be desirable to have

$$x_1 = X_1(\omega_1), \quad x_2 = X_2(\omega_2), \ldots,$$

  because there exists no theory for analysing the behaviour of such samples
- What we desire is

$$x_1 = X_1(\omega), \quad x_2 = X_2(\omega), \ldots,$$

  a scenario wherein all samples correspond to a given ω
- Is there a feasible fit for the latter viewpoint? **YES!**

Frame Title

Proposition (From CDF to Probability Space and Random Variable)

Given a function $F : \mathbb{R} \times \mathbb{R} \times \cdots \to [0, 1]$ satisfying the properties of a joint CDF, there exists $(\Omega', \mathscr{F}', \mathbb{P}')$ and a random process $X' : \Omega' \to \mathbb{R} \times \mathbb{R} \times \cdots$ such that $F_{X'} = F$.