02-712 Week 9

Biological Modeling and Simulation

Aidan Jan

October 23, 2025

Probability and Sampling

Suppose we have a colony of fruit flies, some have white eyes and some have black eyes. Let S be our **sampling space** (set of outcomes), which contains n flies. We can label the flies as

$$S = \{0, 1, 2, \dots, n\}$$

An **event set** is a set of possible events. $E = S^*$. For example,

- $\{0\}$ = we only select the first fly
- $\{0, 2, 4, \dots\}$ = we select even number flies
- $\{0, 1, 2, \dots, n\}$ = we select all the flies.

A **probability density function** (pdf) is a function $P : E \to \mathbb{R}$, which maps the event set to real numbers. For example:

- $\forall e \in E, 0 \leq p \leq 1$
- $P(\emptyset) = 0$
- If the events are independent, we know that $P(e_1 \cup e_2 \cup \cdots \cup e_n) = P(e_1) + P(e_2) + \cdots + P(e_n)$ with $e_i \cap e_j = \emptyset \forall i, j$

e refers to events.

A **random variable** is a function $X : S \to \mathbb{R}$. We might declare that X is a random variable that returns 1 if at least one black-eyed fly is chosen, 0 otherwise. Therefore,

$$X(0) = 0$$
$$X(1) = 1$$
$$X(2) = 1$$

Using the probability density function, we can derive that $\Pr\{X = 0\} = \frac{1}{4}$, $\Pr\{X = 1\} = \Pr\{1\} + \Pr\{2\} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.

Discrete Random Variables

- A **Bernoulli** random variable is two disjoint events, A and B, such that $\Pr(A) = p$ and $\Pr(B) = 1 - p$. Basically, the outcome is either A or B.
- A **Binomial** random variable is a sum of n bernoulli events. $\sum n$ independently distributed Bernoulli trials, where
  - n = num trials

- p = probability per trial
- $S = \{0, 1, 2, \ldots, n\}$
- $x(i) = i$
- In our fly example, this would be like, if we chose n flies, and each fly has a p chance of having black eyes, then what is the chance we get k black-eyed flies?
- The formula: $\Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$

- A **Geometric** random variable is one where the sample space is $S = \{1, 2, \ldots, \infty\}$, and $x(i) = i$.
  - We want to know how many trials need to be done before we get one successful trial.
  - This means the formula is $\Pr\{X = k\} = (1-p)^{k-1} p$. Basically, $k-1$ failed trials (with their probabilities) followed by the single successful trial at the end.
  - For the fly example, how many flies do we need to check the eye color before we find one that has black eyes?

- A **Poisson** random variable is a set of events that is based on an exponential.
  - $S = \{0, 1, 2, \ldots, \infty\}$
  - $\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{Poisson}(\lambda)$
  - As an example, this would be like radioactive atom decay. They follow half-lives, and every time step will decay some percentage. The lambda is a parameter to the function represents the number of time steps in this case.
  - A poisson distribution is also what the binomial distribution would look like if $n = \infty$.

Continuous Random Variables

These use cumulative distribution functions (cdf) instead of pdf, which describes the probability that the value of X is less than a given value k. Additionally, since the distribution is continuous, the probability of landing on any exact value is zero. (It is one value, out of an infinite range of values).

There are some requirements for continuous random variables:

1. $f(x) \geq 0 \forall x$
2. $\int_{-\infty}^{\infty} f(u) \mathrm{d}u = 1$
3. For any (a, b), $\Pr(a \leq X \leq b) = \int_a^b f(u) \mathrm{d}u$

$f(X)$ represents the integral of the function. The graph represents the probability the value is less than a threshold value. $F(x_0)$ represents the probability that x is a given value. Some distributions are:

- **Uniform distribution:**

$$f(X) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x < b \\ 0 & x > b \end{cases}$$

  - The uniform distribution has a cdf graph that looks like a ramp from a to b, from $y = 0$ to $y = 1$.

- **Normal (Gaussian) distribution:**

$$N(\mu, \sigma^2)$$

  - This is the classic bell curve function. The center of the distribution is μ, and the standard deviation is σ.
  - There is no closed-form solution for the cdf, but it would be the integral of the function:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Exponential distribution:**

$$\text{Exp}(\lambda)$$

$$f(X) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

$$F(x_0) = \begin{cases} 0 & x_0 < 0 \\ 1 - e^{-\lambda x_0} & x_0 \geq 0 \end{cases}$$

Joint Distributions

Sometimes we want to model multiple random variables instead of a single one. We may look at two variables instead:

$$\Pr\{X = x_i, Y = y_i\} = f_{xy}(x_i, y_i)$$

$$\Pr\{x_0 \leq X \leq x_1, y_0 \leq Y \leq Y_1\} = \int_{x=x_0}^{x_1} \int_{y=y_0}^{y_1} f_{xy}(x_i, y_i) \mathrm{d}y \mathrm{d}x$$

Conditional probabilities also fall into this category:

$$\Pr\{X = x_i | Y = y_i\}, \quad \text{where } f_{xy}(x_i, y_i)$$

We also have Moments:

- Moments quantify a distribution's location, shape and scale. It is an abstraction of expected value.
- The first moment is the mean ($\mathbb{E}(x)$)
- The second moment is the variance ($\mathbb{E}(x)^2$)
- etc.

$$\mathbb{E}(x) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

$$\text{Var}(x) = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x - \left(\int_{-\infty}^{\infty} x f(x) \mathrm{d}x \right)^2$$

Sampling

Discrete Uniform Random Variables

A discrete uniform random variable is basically a random number generator. This random variable represents a random number between the range of two numbers, often from 0 to some integer.

Uniform Random Variables

If we want go from random integers to random floats in a range, we can do the following:

- Suppose we want a something uniform between 0 and 1. All we have to do is take the random number we generate, and divide it by the range. It would give a number between 0 and 1 that has decent precision if the original max integer is large.
- To get a value between an arbitrary range $[a, b)$, then divide the random integer by $(b - a)$, then add a.

Fundamental Transformation Law of Probability

Given X and density $f(X)$ and the function $y(X) : \mathbb{R} \to \mathbb{R}$, $y(X)$ has density $g(y)$ where $g(y) = f(x)\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|$

Transformation Method

Let $f(x) = u(0, 1]$. u is the uniform distribution. Let $g(y)$ be the target density.

Let $F(x) = 1$ on $(0, 1]$.

If $\frac{\mathrm{d}x}{\mathrm{d}y} = g(y)$, then it implies $\mathrm{d}x = g(y)\mathrm{d}y$.

$$x = \int \mathrm{d}x = \int g(y)\mathrm{d}y = G(y)$$

In summary, there are four steps:

1. Integrate x
2. Invert $G^{-1}(u)$
3. Sample: $x = u(0, 1]$
4. Transform: $y = G^{-1}(x)$

What we are really doing here is if the CDF is a curve on a graph (it will range from 0 to 1 in the x (vertical)-axis), we are using the curve and the values x_0 and x_1 to estimate what range of y (horizontal axis) would generate those values.

- What if we can't invert G?
  - We know that $y = G^{-1}(x)$. Therefore, $G(y) = x$, and $G(y) - x = 0$. From here, we can input this into a zero finder and solve.
- What if we can't integrate $g(y)$ to get $G(y)$?
  - We can use one of the many numerical integration methods from before.

The Rejection Method

Suppose we want to find a density function that we don't know the function of, but we have a sample curve with a lot of noise. How do we find the cdf?

- One way is with the rejection method, where we choose a function $h(x)$ that is strictly greater than our function $g(x)$ for all of x. Then, let $A = \int_{-\infty}^{\infty} h(u)\mathrm{d}u$, and $F(x) = \frac{1}{A}h(x)$.
- Then, pick a point below the envelope function, and then test if that point happens to also be under $g(x)$. If it is, then we consider it as a fair random sample under $g(x)$. Otherwise, we throw away the point and try again.
- This is an easier method than Transformation, but it would take a lot of time if the envelope is bad, since most of the sampled points will fail.
- We can rate how good the envelope is by looking at the area under the envelope. The probability of a good pick is the reciprocal of that. (This is because the area of the pdf/cdf must be equal to 1, so we are comparing the two areas.)

This can also be done with discrete random variables.

Markov Models

A Markov model is a finite state system.

- To move between the states, we have **transition probabilities**. E.g., state 1 has a 50% chance of transitioning to itself, and 50% chance of transitioning to state 2. State 2 has percent chance of going to itself, 1, and 3. etc.
- All the edges leaving a given state must add up to 1 since they represent probabilities.

Suppose we started at a given state, and just simulated moving between states. This series of states, e.g. $s_1, s_2, s_3, s_3, s_3, s_4, s_2$, is known as a **Markov chain**.

- Given a Markov chain, we can calculate the probability of that chain occurring by multiplying the state transition probabilities.

To define a Markov model, we need the following parts:

- State set $Q = \{q_1, q_2, \ldots, q_n\}$. These are our states.
- Starting density $\Pr\{q(0) = q_i\} = p_i$. Decides the probability of starting the chain at a given state.
- Transition matrix $\Pr\{q(n+1) = q_j | q(n) = q_i\} = p_{ij}$

From here, we can simulate the Markov model. The model above is a **1st order** Markov model, since we only consider the previous state.

- There also exist k-th order Markov models, which consider more previous states for the transition probabilities.

Simulating Markov Models

Starting at the first step, we have

$$\begin{bmatrix}\Pr\{q(0) = q_1\}\\ \Pr\{q(0) = q_2\}\end{bmatrix} = \begin{bmatrix}p_1\\ p_2\end{bmatrix}$$

This is simply the starting density, since that defines which state we start in. Now, how do we find the probabilities for step 2?

$$\begin{bmatrix}\Pr\{q(1) = q_1\}\\ \Pr\{q(1) = q_2\}\end{bmatrix} = \begin{bmatrix}(p_1 p_{11} + p_2 p_{21})\\ (p_2 p_{22} + p_1 p_{12})\end{bmatrix}$$

What this means is that the probability we are in state 1 at step 2 is the sum of

- Probability we were in state 1 in step 1, times the probability we stay in state 1
- Probability we were in state 2 in step 1, times the probability we transitioned from 2 to 1.

The same logic can be applied to find the probability we are in state 2 at step 2. Continuing this process, we can find the probabilities at any step. This is basically a series of matrix multiplications when implemented.

Markov Model End Behavior

We can have three different types of model.

- Type 1: the model converges. In a state machine representation, exactly one of the states has a transition probability of 1 to itself. This means that once this node is reached, it stays there. In other words, starting at any other state (assuming the graph is connected) will eventually end up there.
- Type 2: the model diverges. In a state machine representation, we have multiple sink nodes. In an infinite amount of time, starting at any state would result in one of the sink states. From this, we can gain a **stationary density** vector, which stores the probabilities of *ending* at each state. For non-sink nodes, this probability is always 0.

- Type 3: the model oscillates. In a state machine representation, we have no sink nodes. Therefore, at an infinite amount of time, the current state may be any state. We also can derive a stationary density by seeing the values the probabilities converge to.

The stationary densities can be computed by finding the eigenvectors of the transition matrix. The eigenvector is the ending state density. The lambda should be equal to 1, or otherwise it does not make sense.

- The only case with convergence is when there is exactly one eigenvector with an eigenvalue of 1 or -1.

Kolmogourov Criterion

For any cycle $q_1, q_2, \ldots, q_k$, if the following is true:

$$p_{12} \times p_{23} \times p_{34} \times \cdots \times p_{k-1,k} \times p_{k,1} = p_{k,k-1} \times p_{k-1,k-2} \times \cdots \times p_{1,k}$$

e.g., any cycle returns the same probability going in reverse direction, then there is a unique stationary density, and we have detailed balance.

- Detailed balance means $\forall i, j \quad \pi_i p_{ij} = \pi_j \pi_{ji}$, where π represents the stationary density.

Mixing Time

The mixing time is essentially an index for how long it takes to reach the stationary density. Or, how many steps need to be taken before all the probabilities converge?

- This is related to the eigenvalues of the mixing time.
- It is the **ratio** of the two **largest eigenvalues.**