Conjugate Priors:

Imagine you're trying to update your beliefs using Bayesian methods, and you have a specific form for your initial belief (prior) and for the likelihood of observing data. In some fortunate cases, the product of these two functions simplifies, making the calculation easier. When this happens, we say that the prior is a "conjugate" prior for the likelihood.

- **Prior Function $f(\theta)$:**
  - **Example:** Consider you're estimating the probability of success θ in a sequence of Bernoulli trials. You might have a prior belief that the success rate follows a Beta distribution, which could be expressed as $f(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.
- **Likelihood Function $(g_x(\theta)$:**
  - **Example:** For each trial, the likelihood of observing a success or failure might follow a Bernoulli distribution. The likelihood function $g_x(\theta)$ could be expressed as $g_x(\theta) \propto \theta^k(1-\theta)^{n-k}$, where k is the number of successes observed in n trials.
- **Conjugate Prior:**
  - **Simplification:** If the product of $f(\theta)$ and $g_x(\theta)$ simplifies into a known distribution, we call $f(\theta)$ a conjugate prior for the likelihood $g_x(\theta)$.
  - **Example:** In the case mentioned above, if the product of Beta prior and Bernoulli likelihood simplifies into another Beta distribution, we say that the Beta distribution is a conjugate prior for the Bernoulli likelihood.

Benefits of Conjugate Priors:

- **Analytical Solutions:** Having a conjugate prior makes Bayesian updates analytically solvable. This means you can update your beliefs without resorting to numerical methods like grid computations.
- **Mathematical Convenience:** The conjugate priors offer a convenient mathematical form that allows for simpler calculations, reducing the computational complexity.

In summary, conjugate priors are special cases where the product of the prior and likelihood functions simplifies, making Bayesian updates more straightforward and computationally efficient. They are particularly useful in cases where analytical solutions are preferred over numerical methods.

Common Conjugate Priors

1. Binomial and beta distributions
2. Gamma and Poisson distributions
3. Multinomial and Dirichlet distributions

Problem 1

Suppose we observe 8 heads in 10 flips of a coin. If the prior distribution for the probability of heads is a beta distribution with parameters $\alpha = 3$ and $\beta = 5$, what is the posterior distribution for the probability of heads?

Solution

$$f(\theta|x) \sim Beta(11, 7)$$

Problem 2

A machine produces widgets with a Poisson distribution, where the rate parameter λ is unknown. Before observing any widgets, you believe that λ follows a gamma distribution with parameters $\alpha = 2$ and $\beta = 1$. You observe 5 widgets in the first hour. What is the posterior distribution of λ?

Solution

We are given that the machine produces widgets with a Poisson distribution, where the rate parameter λ is unknown. The prior distribution for λ is given as $\lambda \sim \text{Gamma}(\alpha = 2, \beta = 1)$.

Let X be the number of widgets produced in the first hour. We are given that $X = 5$.

The posterior distribution of λ is given by:

$$\begin{aligned} f(\lambda \mid X) &\propto f(X \mid \lambda) \cdot f(\lambda) \\ &= \frac{e^{-\lambda}\lambda^X}{X!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \\ &\propto e^{-\lambda}\lambda^X\lambda^{\alpha-1}e^{-\beta\lambda} \\ &= e^{-\lambda(\beta+1)}\lambda^{X+\alpha-1} \\ &\propto \text{Gamma}(\alpha+X, \beta+1) \end{aligned}$$

Therefore, the posterior distribution of λ is $\lambda \sim \text{Gamma}(\alpha = 7, \beta = 2)$.

Problem 3

While traveling through the Mushroom Kingdom, Mario and Luigi find some rather unusual coins. They agree on a prior of f(θ) ~ beta(5,5) for the probability of heads, though they disagree on what experiment to run to investigate θ further.

a) Mario decides to flip a coin 5 times. He gets four heads in five flips.

b) Luigi decides to flip a coin until the first tails. He gets four heads before the first tail.

Show that Mario and Luigi will arrive at the same posterior on θ, and calculate this posterior.

Note - this question shows that the Beta distribution has 2 conjugate priors - the binomial and the geometric distributions.

Solution

Given that the prior distribution is $f(\theta) \sim \text{Beta}(5,5)$, the posterior distribution can be calculated for the two experiments as follows:

(a) Mario flips a coin 5 times and gets 4 heads. The likelihood function for this experiment is given by:

$$f(X|\theta) = \theta^4(1-\theta)^1$$

The posterior distribution can be calculated as:

$$\begin{aligned} f(\theta|X) &\propto f(X|\theta) \cdot f(\theta) \\ &\propto \theta^4(1-\theta)^1 \cdot \theta^{5-1}(1-\theta)^{5-1} \\ &\propto \theta^{9-1}(1-\theta)^{6-1} \end{aligned}$$

Therefore, the posterior distribution of θ is $\theta \sim \text{Beta}(9,6)$.

(b) Luigi flips a coin until the first tails, and gets 4 heads before the first tail. The likelihood function for this experiment is given by:

$$f(X|\theta) = \theta^4(1-\theta)^1$$

The posterior distribution can be calculated as:

$$\begin{aligned} f(\theta|X) &\propto f(X|\theta) \cdot f(\theta) \\ &\propto \theta^4(1-\theta)^1 \cdot \theta^{5-1}(1-\theta)^{5-1} \\ &\propto \theta^{9-1}(1-\theta)^{6-1} \end{aligned}$$

Therefore, the posterior distribution of θ is also $\theta \sim \text{Beta}(9,6)$.

Hence, we can see that Mario and Luigi will arrive at the same posterior on θ, which is given by $\theta \sim \text{Beta}(9,6)$.

Markov Chains

A Markov chain is like a system where you move from one state to another, but the future state depends only on your current state and not on how you arrived there.

States:

Imagine you're playing a board game with different spaces labeled A, B, and C. Each space is a "state" in our Markov chain.

Transition Probabilities:

From any space, you can move to another space with certain probabilities. For example, if you're currently on space A, there's a 30% chance you'll move to space B and a 70% chance you'll stay on A.

Memoryless Property:

The key idea is that your next move only depends on your current space, not on the sequence of spaces you've been through. It's like a game where your next move is determined by the roll of a fair die, and not by how you got to your current position.

Example:

Board Game Example:

- You start on space A.
- From A, you roll a die to determine your next move. If it's a 1 or 2, you move to B; if it's a 3, 4, 5, or 6, you stay on A.
- From B, you have a 50-50 chance of moving to A or C.
- From C, you always move back to A.

Sequence of Moves:

If you're on A, you might move to B and then back to A in the next turn. The probability of moving from A to B and then back to A is determined only by the probabilities associated with A and B, not by how many turns it took to get there.

Markov Property Explained with a Board Game:

Imagine you're playing a board game where you move from one space to another based on a roll of a fair six-sided die. The spaces are labeled A, B, and C.

1. **Markov Property in Action:**
   - You start on space A.
   - Your next move depends only on your current space, not on how you got there. For example, if you're on A, you roll a die. If it's a 1 or 2, you move to B; if it's 3, 4, 5, or 6, you stay on A.
2. **Memoryless Transition:**
   - Let's say you moved from A to B. Now, your next move is determined by your current position (B) and not by the fact that you just came from A. It's like rolling the die independently of your past moves.
3. **Formalizing the Markov Property:**
   - If X_n represents your position at time n (where n is the turn number), the Markov property says that the probability of your position at the next turn X_{n+1} depends only on your current position X_n.

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

   - This equation means that knowing the entire history of your moves (past states) is unnecessary to predict your next position; just your current position is enough.

In summary, the Markov property simplifies the prediction of future states by focusing on the current state, making it a handy concept in various fields, from board games to real-world systems.

Transition Matrix

A transition matrix is a way to represent the possible transitions between different states in a system, often associated with a Markov chain. In simpler terms, it's like a roadmap showing the probabilities of moving from one condition or situation to another.

Imagine a board game with three spaces labeled A, B, and C. A player rolls a fair six-sided die, and the game follows these rules:

1. If you're on A and roll a 1 or 2, you move to B.
2. If you're on A and roll a 3, 4, 5, or 6, you stay on A.
3. If you're on B, you have a 50-50 chance of moving back to A or moving to C.
4. If you're on C, you always move back to A.

Now, we can represent these transitions with a transition matrix. Let's denote the states as A, B, and C. The rows of the matrix represent the current state, and the columns represent the possible next states. The numbers in the matrix represent the probabilities of transitioning from the current state to the next state.

$$\text{Transition Matrix}(P):$$

P =

$$\begin{bmatrix} P(A \to A) & P(A \to B) & P(A \to C) \\ P(B \to A) & P(B \to B) & P(B \to C) \\ P(C \to A) & P(C \to B) & P(C \to C) \end{bmatrix}$$

Now, filling in the probabilities based on our board game rules:

P =

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0.5 & 0 & 0.5 \\ 1 & 0 & 0 \end{bmatrix}$$

Here, each row represents the current state, and the entries in each row show the probabilities of moving to the other states.

- If you're on A, there's a $\frac{1}{3}$ chance you stay on A, $\frac{1}{3}$ chance you move to B, and $\frac{1}{3}$ chance you move to C.
- If you're on B, there's a 0.5 chance you move to A and a 0.5 chance you move to C.
- If you're on C, you always move back to A.

This matrix helps us visualize and calculate the probabilities of transitioning between states in a system, making it a useful tool in understanding and analyzing Markov processes.

Problem

Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 nonempty urns after 9 balls have been distributed?

Solution

There are only 9 possible configurations of urns based on the problem statement. However, in the Markov chain modeling, we only consider the number of non-empty urns. Hence, the states are defined as follows:

- State 0: 0 non-empty urns
- State 1: 1 non-empty urn
- State 2: 2 non-empty urns
- State 3: 3 non-empty urns
- State 4: 4 or more non-empty urns (this is a collapsed state that includes all configurations with 4, 5, 6, 7, or 8 non-empty urns)

Note that there are 5 possible states in total.

Transition Matrix:

The transition matrix P represents the probabilities of moving from one state to another. Since each ball is equally likely to be put in any urn, the transition probabilities are straightforward: [P =

$$\begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} \\ p_{30} & p_{31} & p_{32} & p_{33} & p_{34} \\ p_{40} & p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

]

Explanation for each element:

[P =

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{8} & \frac{7}{8} & 0 & 0 & 0 \\ 0 & \frac{2}{8} & \frac{6}{8} & 0 & 0 \\ 0 & 0 & \frac{3}{8} & \frac{5}{8} & 0 \\ 0 & 0 & 0 & \frac{4}{8} & \frac{4}{8} \end{bmatrix}$$

]

Initial State Vector:

Let's assume that initially, all urns are empty. So, the initial state vector (X_0) is: [X_0 =

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

]

Transition After 9 Balls:

To find the state after 9 balls, we calculate $X_9 = P^9 \cdot X_0$

After obtaining X_9, the probability of having exactly 3 non-empty urns is given by X_9 in State 3.

$P(\text{3 non-empty urns after 9 balls}) = X_9[3]$

Now, let's calculate X_9.

```
import numpy as np

# Transition Matrix
P = np.array([
    [0, 1, 0, 0, 0],
    [1/8, 7/8, 0, 0, 0],
    [0, 2/8, 6/8, 0, 0],
    [0, 0, 3/8, 5/8, 0],
    [0, 0, 0, 4/8, 4/8]
])

# Initial State Vector
X_0 = np.array([1, 0, 0, 0, 0])

# Calculate X_9
X_9 = np.linalg.matrix_power(P, 9).dot(X_0)

# Probability of 3 non-empty urns after 9 balls
prob_3_nonempty_urns = X_9[3]

print(f"Probability of having exactly 3 non-empty urns after 9 balls: {prob_3_nonempty_urns:.4f}")
```

```
Probability of having exactly 3 non-empty urns after 9 balls: 0.0838
```