# Bayesian Inference Model for NBA Player Performance Prediction

Biya Brook

March 2025

## 1 Introduction

In this project, I apply Bayesian sequential updating—a mathematical method that combines different pieces of evidence—to predict how many points Jayson Tatum will score in a game. My model starts with his season average (27 points) and refines this prediction by adding information about the game location, opponent matchup, and his recent performance. This approach, using core concepts from CS109, systematically improves our prediction with each new piece of evidence, creating a normal distribution that tells us not just what Tatum might score, but how certain we can be about different outcomes. When compared to implied betting market odds, this posterior distribution can identify opportunities with positive expected value.

## 2 Methodology: Bayesian Inference Framework

At its core, my approach implements Bayes' theorem for continuous distributions. In CS109, we learned the Bayesian inference formula for probability density functions:

$$P(N = n|X = x) = \frac{f(X = x|N = n) \cdot P(N = n)}{f(X = x)} \tag{1}$$

In my sequential updating approach, I apply this framework to normal distributions, where $N$ represents the points Tatum will score (unknown parameter) and $X$ represents our observed evidence. When using normal distributions, the posterior takes a specific form:

$$\tau_0 = \frac{1}{\sigma_0^2} \quad \text{(prior precision)} \tag{2}$$

$$\tau_L = \frac{1}{\sigma_L^2} \quad \text{(likelihood precision)} \tag{3}$$

$$\tau_{post} = \tau_0 + \tau_L \quad \text{(posterior precision)} \tag{4}$$

$$\mu_{post} = \frac{\mu_0\tau_0 + \mu_L\tau_L}{\tau_{post}} \quad \text{(posterior mean)} \tag{5}$$

$$\sigma_{post}^2 = \frac{1}{\tau_{post}} \quad \text{(posterior variance)} \tag{6}$$

This solution emerges directly from the structure of the normal distribution, with precision (inverse of the variance) serving as the weighting factor for evidence (see Appendix A for full derivation).

## 3 Data Collection and Evidence Sources

I leveraged the NBA Stats API to collect data for Tatum's next game at home against the Lakers:

- **Season data**: All games Jayson Tatum from the current season (n=59)
- **Home/away splits**: Performance specifically in home games (n=29)
- **Opponent-specific data**: Performance against the Lakers in the last 3 years (n=5)
- **Recent form**: Last five games before the target matchup (n=5)

For each dataset, I created a normal distribution using the sample mean and standard deviation:

| Evidence Source | Mean | Std Dev | Precision |
|---|---|---|---|
| Season (Prior) | 26.93 | 7.90 | 0.0160 |
| Home Games | 27.20 | 12.36 | 0.0065 |
| vs Lakers | 27.60 | 10.45 | 0.0092 |
| Recent Form | 28.60 | 12.22 | 0.0067 |

Table 1: Input distributions for Bayesian inference

## 4 Sequential Bayesian Updating Process

My approach uses these four evidence sources along with the posterior mean and variance formulas using precision to sequentially update our beliefs about Tatum's scoring (see Appendix B for all calculations).

After all updates, the final posterior is $N|X \sim \mathcal{N}(27.43, 5.10^2)$. Figure 1 shows the evolution of these distributions, with each update narrowing our uncertainty.

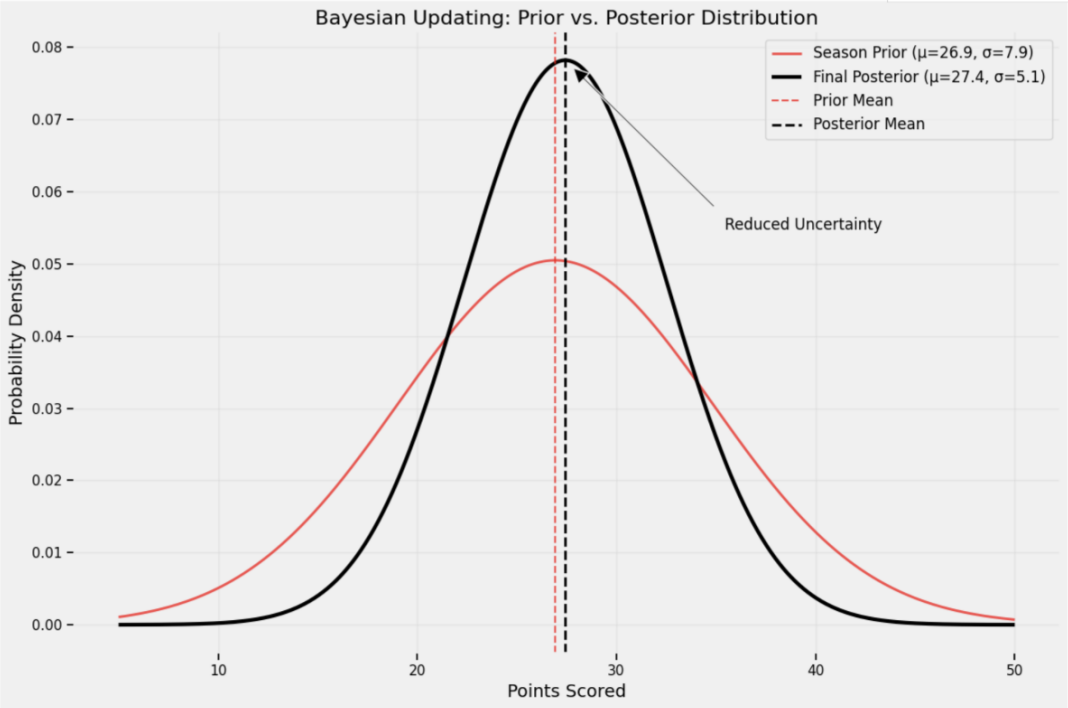


Figure 1: Sequential Bayesian inference showing the evolution of distributions

## 5 Application to Betting Markets using Expected Value

The final posterior distribution allows us to make probabilistic inferences about Tatum's performance against various scoring thresholds. For any threshold value $t$, we can calculate the probability that Tatum exceeds it using the CDF of my posterior distribution:

$$P(\text{Tatum} > t | X_1, X_2, X_3, X_4) = 1 - \Phi\left(\frac{t - 27.43}{5.10}\right) \tag{7}$$

To apply this model to real betting markets, I used the Odds API to collect current betting lines and their associated odds from DraftKings. I compared my model's probability estimates with the bookmakers' implied probabilities to identify any positive expected value betting opportunities.

| | line | outcome | decimal_odds | american_odds | model_probability | implied_probability | edge | expected_value | roi |
|---|---|---|---|---|---|---|---|---|---|
| **5** | 28.5 | Under | 1.869565 | -115 | 58.32 | 53.49 | 4.83 | 9.04 | 9.04 |
| **1** | 19.5 | Over | 1.110011 | -909 | 93.99 | 90.09 | 3.90 | 4.33 | 4.33 |
| **0** | 17.5 | Over | 1.059988 | -1667 | 97.42 | 94.34 | 3.07 | 3.26 | 3.26 |
| **2** | 24.5 | Over | 1.380228 | -263 | 71.69 | 72.45 | -0.76 | -1.04 | -1.04 |
| **3** | 28.5 | Over | 1.869565 | -115 | 41.68 | 53.49 | -11.81 | -22.08 | -22.08 |
| **4** | 28.5 | Over | 1.869565 | -115 | 41.68 | 53.49 | -11.81 | -22.08 | -22.08 |
| **6** | 29.5 | Over | 2.050000 | +105 | 34.23 | 48.78 | -14.55 | -29.82 | -29.82 |
| **7** | 34.5 | Over | 3.550000 | +255 | 8.29 | 28.17 | -19.88 | -70.58 | -70.58 |
| **8** | 39.5 | Over | 7.500000 | +650 | 0.90 | 13.33 | -12.43 | -93.26 | -93.26 |
| **9** | 44.5 | Over | 17.000000 | +1600 | 0.04 | 5.88 | -5.84 | -99.30 | -99.30 |
| **10** | 49.5 | Over | 41.000000 | +4000 | 0.00 | 2.44 | -2.44 | -99.97 | -99.97 |

Figure 2: Comparison of model probabilities vs. implied bookmaker probabilities

As shown in Figure 2, this analysis revealed several potential betting opportunities across different thresholds, such as under 28.5 points and over 19.5 points.

## 6 Conclusion

This project demonstrates Bayesian inference in action—applying key concepts from CS109 to a real-world prediction challenge. By sequentially updating our beliefs with multiple evidence sources, we created a model that:

- Weights each piece of evidence according to its precision
- Systematically reduces uncertainty with each update
- Generates calibrated probabilities for performance predictions

While it's very difficult to outperform sportsbooks' sophisticated models, this approach suggests that a deeply focused, player-specific analysis can potentially identify edges. By examining factors unique to Tatum's performance context, we found several thresholds where our model diverged from market probabilities.

This demonstrates the power of Bayesian methods in specialized contexts, where detailed evidence integration can reveal inefficiencies. Beyond sports, this same methodology applies to any field requiring belief updates from multiple evidence sources with varying reliability—from medical diagnosis to financial forecasting.

# A Derivation of Precision-Weighted Formulas

The precision parameter $\tau$ emerges naturally from the standard Gaussian PDF:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8}$$

By substituting $\tau = \frac{1}{\sigma^2}$, we can rewrite this as:

$$f(x) = \frac{\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{\tau(x-\mu)^2}{2}} \tag{9}$$

When multiplying PDFs as we do in Bayes' theorem, we multiply:

$$f(n|x) \propto f(x|n)f(n) \tag{10}$$

$$\propto \exp\left(-\frac{\tau_L(x-n)^2}{2}\right) \cdot \exp\left(-\frac{\tau_0(n-\mu_0)^2}{2}\right) \tag{11}$$

$$= \exp\left(-\frac{\tau_L(x-n)^2 + \tau_0(n-\mu_0)^2}{2}\right) \tag{12}$$

Expanding the terms with $n$:

$$-\frac{\tau_L(x-n)^2 + \tau_0(n-\mu_0)^2}{2} = -\frac{\tau_L(x^2 - 2xn + n^2) + \tau_0(n^2 - 2n\mu_0 + \mu_0^2)}{2} \tag{13}$$

$$= -\frac{\tau_L x^2 - 2\tau_L xn + \tau_L n^2 + \tau_0 n^2 - 2\tau_0 n\mu_0 + \tau_0\mu_0^2}{2} \tag{14}$$

$$\tag{15}$$

Collecting terms with $n^2$ and $n$:

$$= -\frac{(\tau_L + \tau_0)n^2 - 2(\tau_L x + \tau_0\mu_0)n + \tau_L x^2 + \tau_0\mu_0^2}{2} \tag{16}$$

Completing the square:

$$= -\frac{(\tau_L + \tau_0)\left(n^2 - \frac{2(\tau_L x + \tau_0\mu_0)}{(\tau_L + \tau_0)}n\right) + \tau_L x^2 + \tau_0\mu_0^2}{2} \tag{17}$$

$$= -\frac{(\tau_L + \tau_0)\left(n - \frac{\tau_L x + \tau_0\mu_0}{(\tau_L + \tau_0)}\right)^2 + \text{terms without n}}{2} \tag{18}$$

This gives us the form of a normal distribution with:

$$\tau_{post} = \tau_0 + \tau_L \tag{19}$$

$$\mu_{post} = \frac{\tau_0\mu_0 + \tau_L x}{\tau_{post}} \tag{20}$$

# B Complete Sequential Updating Calculations

**Step 1:** Starting with season prior: $\mu_0 = 26.93$, $\sigma_0 = 7.90$

$$\tau_0 = \frac{1}{\sigma_0^2} = \frac{1}{7.90^2} = 0.0160 \tag{21}$$

$$\tag{22}$$

**Step 2:** Update with home game evidence: $\mu_H = 27.20$, $\sigma_H = 12.36$

$$\tau_H = \frac{1}{\sigma_H^2} = \frac{1}{12.36^2} = 0.0065 \tag{23}$$

$$\tau_1 = \tau_0 + \tau_H = 0.0160 + 0.0065 = 0.0225 \tag{24}$$

$$\mu_1 = \frac{\mu_0\tau_0 + \mu_H\tau_H}{\tau_1} = \frac{26.93 \cdot 0.0160 + 27.20 \cdot 0.0065}{0.0225} = 27.01 \tag{25}$$

$$\sigma_1 = \sqrt{\frac{1}{\tau_1}} = \sqrt{\frac{1}{0.0225}} = 6.66 \tag{26}$$

**Step 3:** Update with Lakers matchup evidence: $\mu_L = 27.60$, $\sigma_L = 10.45$

$$\tau_L = \frac{1}{\sigma_L^2} = \frac{1}{10.45^2} = 0.0092 \tag{27}$$

$$\tau_2 = \tau_1 + \tau_L = 0.0225 + 0.0092 = 0.0317 \tag{28}$$

$$\mu_2 = \frac{\mu_1\tau_1 + \mu_L\tau_L}{\tau_2} = \frac{27.01 \cdot 0.0225 + 27.60 \cdot 0.0092}{0.0317} = 27.18 \tag{29}$$

$$\sigma_2 = \sqrt{\frac{1}{\tau_2}} = \sqrt{\frac{1}{0.0317}} = 5.62 \tag{30}$$

**Step 4:** Update with recent form evidence: $\mu_R = 28.60$, $\sigma_R = 12.22$

$$\tau_R = \frac{1}{\sigma_R^2} = \frac{1}{12.22^2} = 0.0067 \tag{31}$$

$$\tau_3 = \tau_2 + \tau_R = 0.0317 + 0.0067 = 0.0384 \tag{32}$$

$$\mu_3 = \frac{\mu_2\tau_2 + \mu_R\tau_R}{\tau_3} = \frac{27.18 \cdot 0.0317 + 28.60 \cdot 0.0067}{0.0384} = 27.43 \tag{33}$$

$$\sigma_3 = \sqrt{\frac{1}{\tau_3}} = \sqrt{\frac{1}{0.0384}} = 5.10 \tag{34}$$