Augmenting Thompson Sampling with Poisson-Based Time Decay for Randomly Changing Treatments

Benjamin Klieger - August 10th, 2026

Standard Thompson sampling is optimally effective under static conditions. If the effectiveness of treatments remains fixed, it produces an optimal result. This relies upon an inherent assumption: the future effectiveness of the treatment can be determined equally from historical observations.

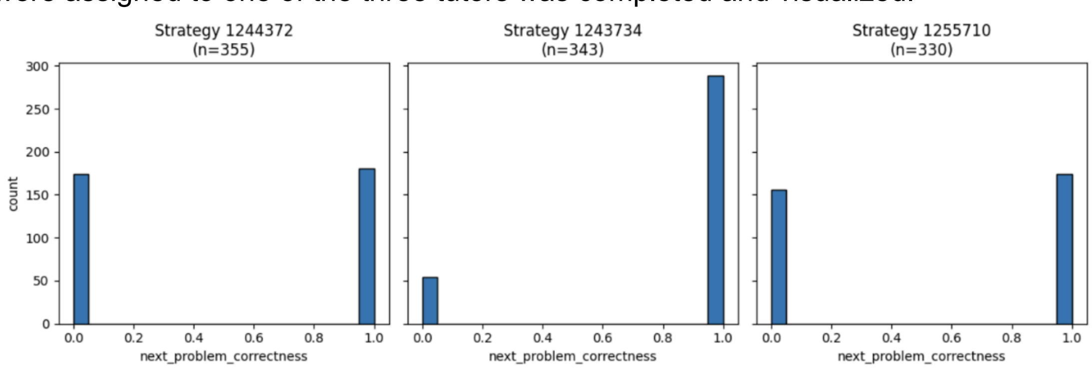
One Thompson sampling problem is a tutoring system. There are three tutors, each with their own strategy. Each strategy has a distinct effectiveness. We must route students to the most effective tutor. Thompson Sampling works well here.

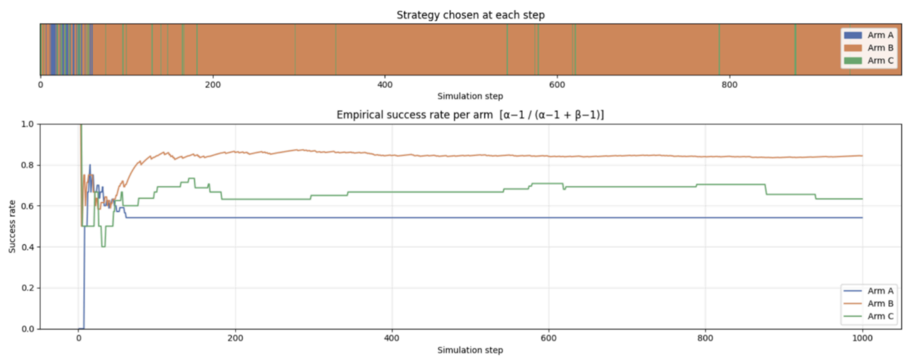
But what happens if the tutor's effectiveness changes over time? What happens if they change strategies? If we know when they changed strategies, we can reset the beta distribution, and treat it like a new treatment. Often, however, we do not know when a change has happened. In this case, Thompson sampling over-indexes on previous observations that may no longer be relevant.

More precisely, there is diminishing predictive power of observations as they age, as the chance they are no longer predictive due to a change in strategy increases. The normal beta distribution does not account for this paradigm.

In many real-world settings, effectiveness changes over time in a manner we could predict with a rate but not an exact timing. In healthcare, treatment effectiveness changes as diseases grow resistant. In education, tutoring effectiveness changes as strategies and student needs evolve. In disaster response, routing emergency paths requires adaptation when newly downed power lines or trees can inhibit path effectiveness after occurring without notification. This research aims to modify Thompson sampling to produce a more effective algorithm for these cases.

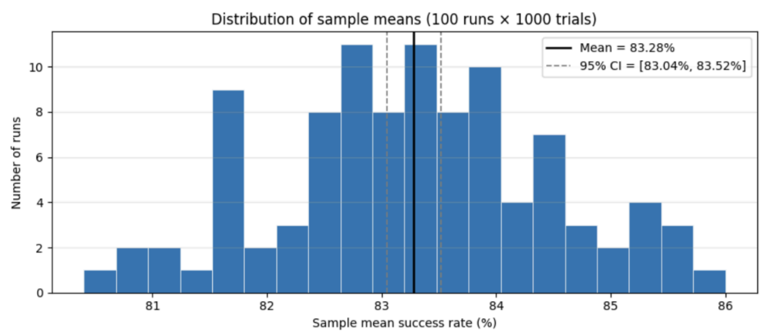
To illustrate and test the hypothesis that an alternative to Thompson sampling may be more effective, a dataset of tutoring strategies was used (**https://dl.acm.org/doi/10.1145/3636555.3636909**). First, the top three most well represented strategies were pulled. Then, a simulated treatment regime where students were assigned to one of the three tutors was completed and visualized.



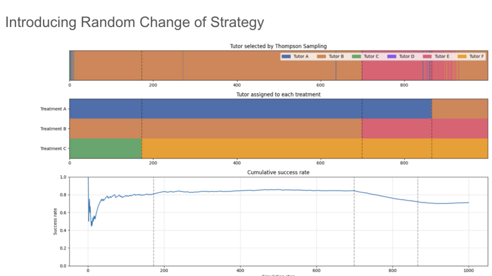


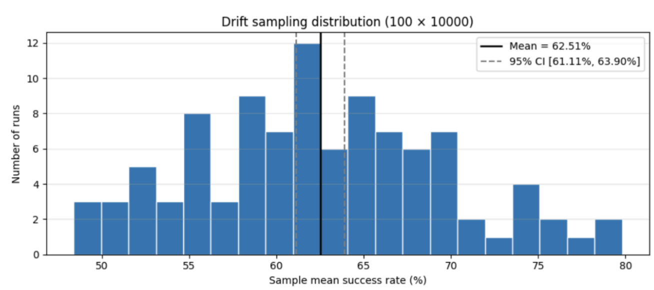
$$\theta_{\text{arm}} \sim \text{Beta}(1 + \#\text{successes},\ 1 + \#\text{failures}), \qquad \text{choose the arm with the largest } \theta_{\text{arm}}$$

The above uses unmodified Thompson sampling. As expected, the most effective strategy (B) quickly became the dominant treatment as the beta distribution updated to reflect the underlying effectiveness. The distribution of sample means showed a normal distribution around the strategy B effectiveness.



However, to make the simulation more representative of dynamic cases, we now introduce two rules: (1) every step, each tutor has a 1/1000 chance of switching the underlying strategy, (2) new strategy can be any of a pre-determined set of six, omitting samples currently in use.





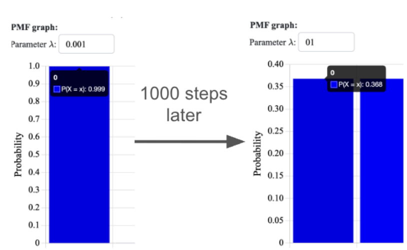
Now, the effectiveness has dropped to about 62.51% across 100 runs, each with 10,000 steps. The underlying assumption that all observations should be weighted equally is no longer valid. The probability that an observation is still relevant is not 1, as a traditional underlying alpha and beta calculation as sums of historical outcomes would assume. Rather, the weight of each observation should instead be modeled as a poisson PDF where N = 0.

$$N(k) \sim \text{Poisson}(\lambda k)$$

$$P(N = n) = \frac{(\lambda k)^n e^{-\lambda k}}{n!}$$

$$P(N = 0) = \frac{(\lambda k)^0 e^{-\lambda k}}{0!} = e^{-\lambda k}$$

$$w(k) = e^{-\lambda k}$$



The probability that an observation is still relevant is the probability that in the time since it has happened, there have been no strategy switching events. This is a poisson PDF where N = 0. For example, after 1 step, the observation is weighted as 0.999. After 1000 steps, it is weighted at 0.368.

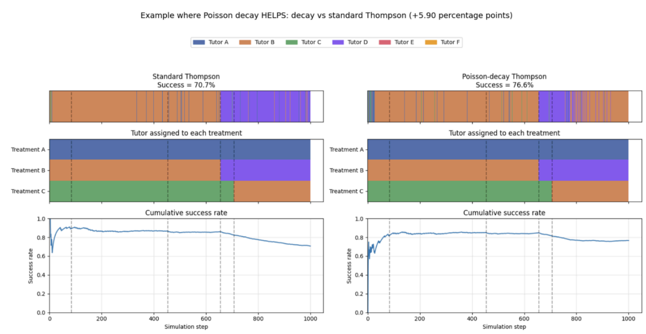
$$w(k) = e^{-\lambda k}$$

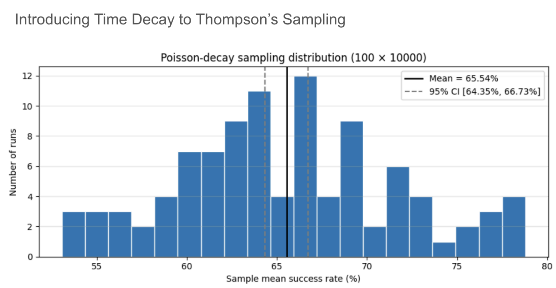
$$\alpha = 1 + \sum w(k) R \qquad R = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure} \end{cases}$$

$$\beta = 1 + \sum w(k)(1 - R) \qquad k = \text{number of steps since the observation occurred}$$

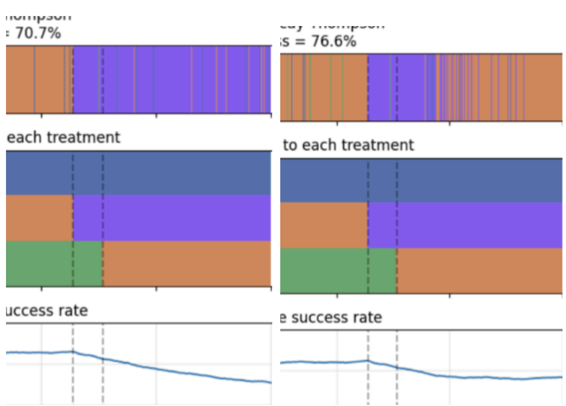
$$\theta_a \sim \mathrm{Beta}(\alpha_a, \beta_a)$$

For poisson-based time decay Thompson sampling, we weigh each previous observation with w(k). This allows us to form a new underlying Beta distribution that is used. Each observation's contribution to alpha and beta is weighted by the probability that it is still relevant.



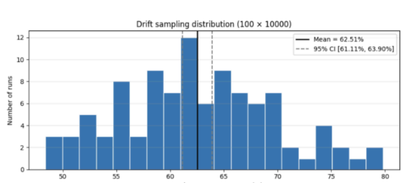


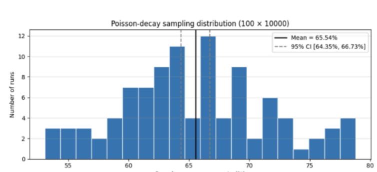
Redeploying the strategy with this poisson-based time decay, we see a higher mean outcome of 65.54%. Specifically, the impact is that the sampling algorithm is faster to react to changes in strategy effectiveness.

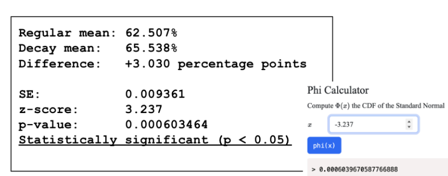


In real examples, poisson-decay Thompson is often faster to switch to the more effective strategy after the underlying strategies have changed. For instance, on the left, the regular Thompson sampling fails to adjust to favor sampling from the third arm (orange) after the change. This causes the aggregate success rate to decline over time. However, on the right, the adjusted sampling strategy switches to primarily sampling the third arm, stopping the decline in aggregate success rate.

To prove this strategy effective, a statistical p-test was used. The sample means of success from using the poisson-based time decay was found to be statistically significantly higher than regular Thompson sampling, with a p value of 0.006. Thus, the evidence suggests that the augmented version of Thompson sampling is more effective than the original version.







Appendix A: AI Disclosure and Reflection

Claude Code and ChatGPT were used as thought and implementation partners. ChatGPT was particularly effective at providing background literature review, although the suggested sources needed manual inspection and judgement as to their relevance. Claude Code was helpful for implementing some of the code, translating my research ideas into tests. Critically reviewing the code and implementation was useful to correcting mistakes in AI's interpretation of my goals and sparking improvements in the methodology.

Appendix B: Applications

Applications

- Education
  - **More effective tutoring and education strategies** that can adapt to changing effectiveness from evolving tutor strategies and student needs.

- Medicine
  - **More effective healthcare treatment allocation** and medical strategy that avoids "over-indexing" on treatment options when conditions change at known rates but unknown times.

- Disaster Response
  - **More effective disaster response** for emergency path routing as changing conditions, such as newly downed power lines or trees, can occur without notification

And more!