Maire Keene
Bayesian Case Study, Fall 2014

Exploration vs. Exploitation: SIDS and Bayesian Bandits

When my mother was a girl, her baby brother Robert died in his sleep. His official cause of death was declared to be Sudden Infant Death Syndrome (SIDS). As the name implies, SIDS deaths are sudden, unexplained, and unpredictable, and are only declared after all other explanations have been ruled out. Decades later, SIDS is the leading cause of death of infants between 1 month and 1 year old in the US, and it is still unclear why seemingly healthy, happy, and energetic infants sometimes suddenly die, usually in their sleep as was the case with my uncle, and usually with no signs of apparent struggle or suffering.

That isn't to say that we have learned nothing at all in the last few decades, but the best that we have been able to do has been to identify sets of physical and sleep environmental factors that increase the risk of SIDS, and make recommendations to parents on how to put their baby to sleep in the safest way-- on their back, side or front, with or without a blanket, with or without mattress pads or crib bumpers, in a slightly warm or cold room, etc. Of course, when my grandmother had my uncle, she *also* received recommendations and instructions on how to keep her infant as safe as possible, but as you can imagine, they were based on far less data than what we have today.

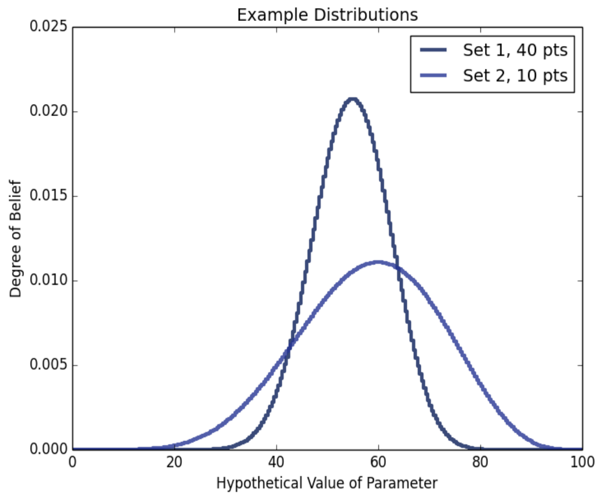
My grandmother acknowledges this and accepts it as inevitable, but what still makes her shake her head and sigh to this day is the fact that, as it so happens, the public health recommendations today are the exact opposite of what she was told was the safest course of action for her baby all those years ago. What bothers her isn't so much the changing of the recommendations, but the appearance of confidence, the implicit claim she would argue the doctors and government made, that they knew the best way to put a baby to bed when this actually wasn't the case. She instead would have wanted them to admit their ignorance in the absence of a solid body of data supporting a particular set of sleeping conditions, rather than recommending their guess with assurance while they went about gathering more data, and finally changing their recommendations when they had accumulated statistically significant proof.

In fact, it would have suited her to have their best guesses as to mortality rates associated with the sleeping conditions and their uncertainty about those guesses shared upfront with parents like her and updated as more data was gathered (Bayes' Theorem, anyone?) so parents could factor these data in when making their own personal decisions about how they could feasibly best provide care for their new infant.[1]

Pretty clearly, the entire SIDS situation is a complex problem, both medically and ethically. But suppose the public health institutions of 50 years ago were willing to admit their ignorance, and publicly update their beliefs and uncertainties about those beliefs regarding the best combination of sleeping conditions based on the incoming stream of data form the community they served. One could imagine that parents would distribute themselves pretty evenly trying out different sets of sleeping conditions at first and would begin to flock to whatever set began to emerge as having the lowest mortality rate, providing more data for that set in particular and narrowing the distribution of our belief in the mortality rate.

1 In case you're curious at this point, my grandmother was a math major, and my grandfather an actuary, which may help to explain some of their viewpoints...

However, focusing in too tightly on one set of sleeping conditions that looks promising in the beginning and gathering lots of data for that particular set may lead us to miss the *actual* set of sleeping conditions with the lowest mortality rate— for example, take a look at the figure to the left. The first set of conditions looks like it has a slightly lower most likely value (~55%), and would be selected over and over by our hypothetical parents, gathering more data and increasing our confidence in our belief, where the second distribution has a higher most likely value (~60%), but is lower and more widely spread, with less data as the parents refused to choose it over what looked to be more promising. However, the second set still has a non-trivial probability that it's true value is indeed lower than the first set's, and within two more data points could be the one that looks more appealing! We need to find a balance between exploiting what we know has a *low* mortality rate and sampling those sets that still have a decent *chance* of having a *lower* mortality rate.



This balance is an example of an Exploration vs Exploitation problem, and our goal is to identify, with a certain amount of confidence, the best choice from among the many as fast as possible. Exploration vs Exploitation is a problem that pops up all over the place, but the canonical example deals with an array of slot machines. To borrow shamelessly from [Cam Davidson-Pilon's article on the Multi-Armed Bandit Problem], the problem can be stated:

> "Suppose you are faced with *N* slot machines (colorfully called multi-armed bandits). Each bandit has an unknown probability of distributing a prize (assume for now the prizes are the same for each bandit, only the probabilities differ). Some bandits are very generous, others not so much. Of course, you don't know what these probabilities are. By only choosing one bandit per round, our task is devise a strategy to maximize our winnings."

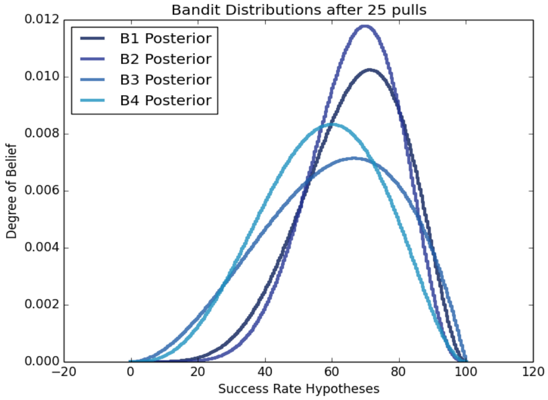
We can use a Bayesian approach to accomplish this with an algorithm referred to as Bayesian Bandits. Walking up to these "bandits," we are completely ignorant of their success rates, making for a flat prior. Bayesian Bandits, then, is an algorithm for choosing which machine to sample in a given round in a way that is *influenced*, though not entirely determined, by our current distribution of beliefs about each machine, and uses Bayes' theorem to update the prior of the relevant machine with the sampled data.

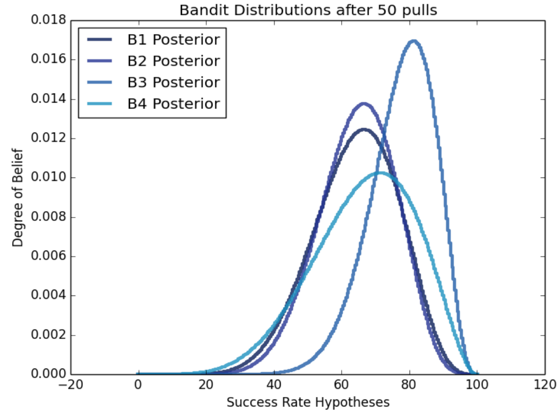
The algorithm is very simple, involving choosing a random hypothesis from the prior of each machine-- *where the likelihood of each hypothesis being chosen is our belief in that hypothesis*-- then selecting the machine with the (in this case) highest valued chosen hypothesis and sampling that machine. After a Bayesian update, the process is repeated. Choosing hypotheses with equal probability given to each would be useless to us since we don't actually care about identifying the success rates themselves to a certain confidence, we care about identifying the *best of the machines* to a certain confidence. In this way, machines that we know very little about but have most of their probability mass lower than another will be chosen very rarely, which is good, since we have very high confidence that they are not better than our other machines, even if we can't say very narrowly what their success rate is. This also means, however, that we will not deterministically choose the machine with the highest return rate to date every time, creating that exploration and exploitation balance determined by the overlaps in the machine's probability masses.

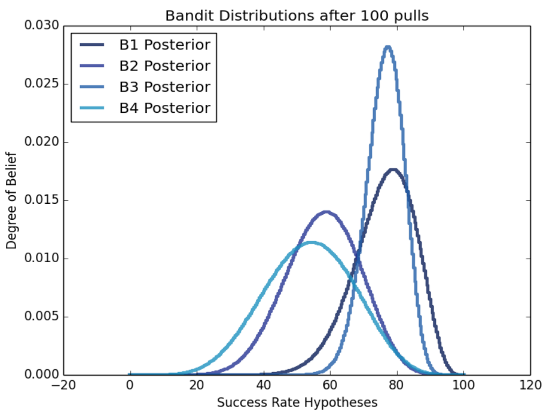
Below is my implementation of Bayesian Bandits in the thinkbayes framework, executed with four slot machines with the true success rates 85%, 60%, 75%, and 45%.

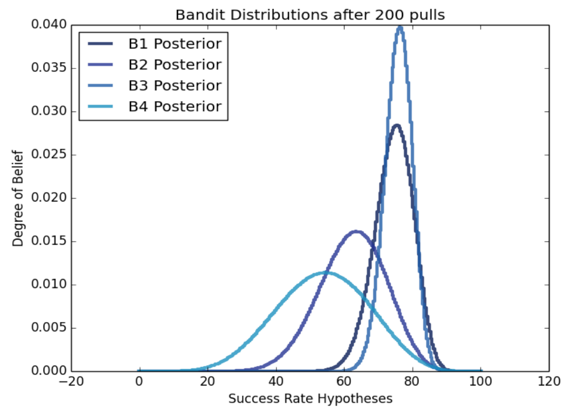
```python
import thinkbayes2
import thinkplot
from random import random
import numpy

class Bandit(thinkbayes2.Suite):
    """Suite of hypotheses of rates of success getting reward"""
    def Likelihood(self, data, hypo):
        """Computes the likelihood of the data under the hypothesis.

        hypo: value of x, the probability of success (0-100)
        data: True or False, result of a simulated pull
        """
        x = hypo / 100.0
        if data == True:
            return x
        else:
            return 1-x

def SimPull(hidden):
    return random()< hidden

def main():
    hypos=numpy.linspace(0,100,250)
    B1=Bandit(hypos)
    B2=Bandit(hypos)
    B3=Bandit(hypos)
    B4=Bandit(hypos)

    Bandits=[B1,B2,B3, B4]
    hiddenRates=[.85, .60, .75, .45]
    thinkplot.Pmf(B1, label='Uniform Prior')

    NumPulls=1000
    for i in range(0,NumPulls):
        #Sample Random Var Xb for each bandit
        RandHypo=[]
        for pmf in Bandits:
            RandHypo.append(pmf.Random())

        #Choose highest randomly sampled hyposthesis and simulate a pull on that bandit
        OptimisticHypo=max(RandHypo)
        BestBanditIndex=RandHypo.index(OptimisticHypo)
        print(BestBanditIndex)
        data=SimPull(hiddenRates[BestBanditIndex])

        #update belief about that bandit's probability distribution
        Bandits[BestBanditIndex].Update(data)


        if i%25==0:
            thinkplot.Pmf(B1, label='B1 Posterior')
            thinkplot.Pmf(B2, label='B2 Posterior')
            thinkplot.Pmf(B3, label='B3 Posterior')
            thinkplot.Pmf(B4, label='B4 Posterior')
            thinkplot.Show()

if __name__ == '__main__':
    main()
```

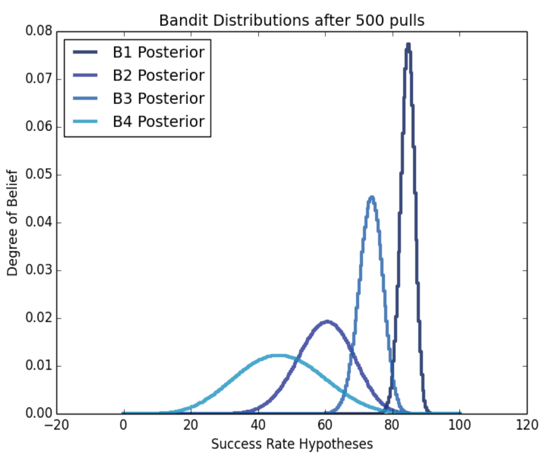
When run for 1000 rounds, we see intermediary and end distributions like we would expect, where the machines we become more confident are inferior get pulled less and less frequently as we gather more data, resulting in a progression of distributions that go from likely low-valued and spread out to likely high-valued and very confident:

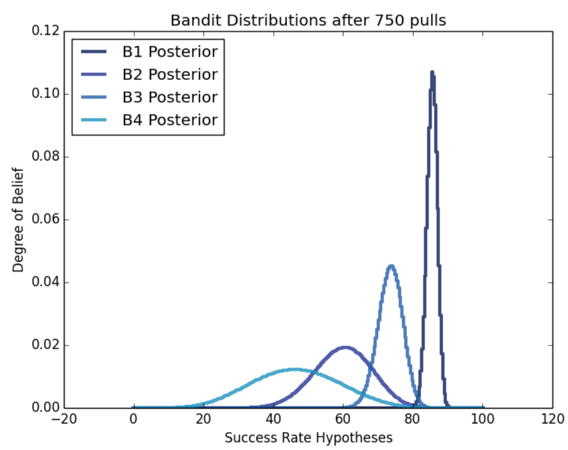


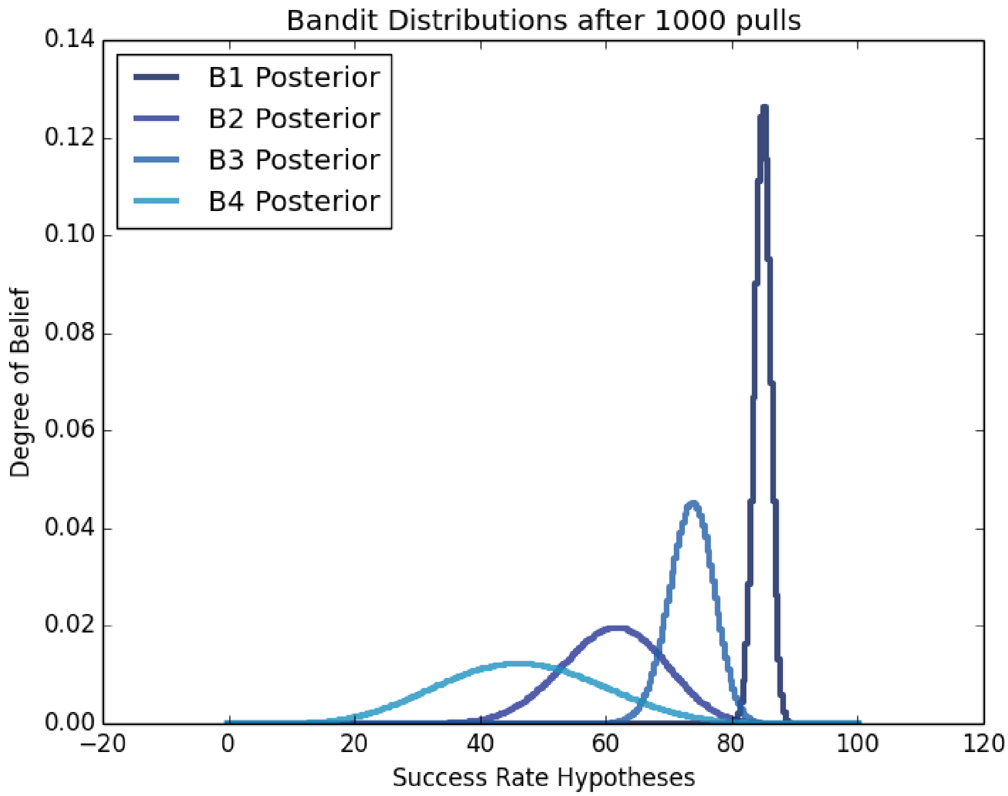












In many ways, Bayesian Bandits resembles what we might expect a group of reasonable people to do when approaching a situation like the Multi-armed Bandit problem—sampling around to gather a bit of data and settling more and more on a favorite—just as Bayes theorem can be thought of as a quantification of what most people do intuitively, increasing their belief in something when they see more or more extraordinary data in support of it. But we also see that the algorithm is a bit more stochastic and risk-taking then we might expect humans to go for—for example, many people would be skeptical after seeing just 50 pulls that they should try any machine other than B3, looking at the most likely values for each distribution and being convinced enough that they had chosen the correct one. Yet we can see that the algorithm kept testing around and in fact rooted out the correct ordering of the Bandits with great confidence by 500 pulls. (Admittedly, probably longer than most people would probably be willing to actually sit there collecting data on an array of slot machines...)

One could easily see this reluctance to keep testing apparent “losers” when something much more high stake is at risk, like in the SIDS situation, being even stronger, and causing a group of people to be more vulnerable to sticking too closely to early data 'red herrings', as it were. At 100 and 200 pulls especially, it would be difficult to convince people it was worthwhile to take the risk of trying B1 over B3, when it appeared to have a slightly greater chance of being *lower*-valued than B3.

What's more, what if the underlying rates actually changed with time? How difficult would it be to convince a group of people that they should give up on a machine or method that had worked pretty well for them in the past to try something that was historically less effective? Or, approaching it from the other direction, in contexts like medical trials, at what point does it become questionably ethical to keep sampling from something we became more and more convinced was less effective?

I won't pretend to really have answers on these fronts, but this does bring us back to a larger, more fundamental discussion of the place of intelligently applied, purposeful, trial and error in matters of public health—or any sort of larger governmental policy, really. This, of course, being extremely controversial—no one seems really comfortable with the idea that they might be being experimented on. But is simply recommending a best guess for large scale policy and sticking with it until the medical field can begin to provide statistically significant and robust evidence from smaller clinical trials any better?

(If you find this interesting, I highly recommend Tim Hartford's fantastic TED talk, Trial, error, and the God complex.)