1 Lottery

Suppose each week, 2000000 lottery tickets are sold for \$1 apiece. If 4000 of these tickets pay off \$30 each, 500 pay off \$800 each, one ticket pays off \$1200000, and no ticket pays off more than one prize?

1.1 Expected Value of one ticket

Since no ticket pays off more than one prize, and each ticket is only one option, the expected value of a ticket is:

$$
\begin{aligned}
E\left(\text{Ticket}\right) &= \frac{N_{\text{no-prize tickets}}0 + N_{\$30\text{-prize tickets}}30 + N_{\$800\text{-prize tickets}}800 + N_{\$1200000\text{-prize tickets}}1200000}{N_{\text{tickets}}} \\
&= \frac{0 + 120000 + 400000 + 1200000}{1995499} \\
E\left(\text{Ticket}\right) &= 0.86
\end{aligned}
$$

This lottery seems like a losing proposition.

1.2 Expected Value of Five Tickets

One might naively conclude that the expected value of five tickets is $5E\left(\text{Ticket}\right)$. However, that is not correct. The \$1,200,000 ticket goes into the expected value, but it can only be one of the tickets. Therefore, the expected value will be less than $5E\left(\text{Ticket}\right)$.

1.2.1 Expected Value of Two Tickets

To prepare for this calculation, I will first calculate the expected value of two tickets.

$$\begin{aligned}
E\left(\text{Two Tickets}\right) &= P\left(\text{Ticket 1 is worth \$ 0}\right)\left(0 + E\left(\text{Ticket 2} \mid \text{Ticket 1 is worth \$0}\right)\right) \\
&+ P\left(\text{Ticket 1 is worth \$ 30}\right)\left(\$30 + E\left(\text{Ticket 2} \mid \text{Ticket 1 is worth \$30}\right)\right) \\
&+ P\left(\text{Ticket 1 is worth \$ 800}\right)\left(\$800 + E\left(\text{Ticket 2} \mid \text{Ticket 1 is worth \$800}\right)\right) \\
&+ P\left(\text{Ticket 1 is worth \$ 1,200,000}\right)\left(\$1,200,000 + E\left(\text{Ticket 2} \mid \text{Ticket 1 is worth \$1,200,000}\right)\right) \\
&= \frac{1995499}{2000000}\left(\frac{120000 + 400000 + 1200000}{1999999}\right) \\
&+ \frac{4000}{2000000}\left(30 + \frac{110070 + 400000 + 1200000}{1999999}\right) \\
&+ \frac{500}{2000000}\left(800 + \frac{120000 + 399500 + 1200000}{1999999}\right) \\
&+ \frac{1}{2000000}\left(1200000 + \frac{120000 + 400000}{1999999}\right) \\
&= 0.85806 + 0.061710 + 0.20021 + 0.6 \\
&= 1.71998
\end{aligned}$$

This is very slightly smaller than $2 * E\left(\text{Ticket}\right) = 1.72$, as predicted.

2 Aircraft Departure

2.1 Problem Statement

To prepare an aircraft for departure, it takes a random amount of time between 20 and 27 minutes. If the flight is scheduled to depart at 9:00 am and preparation begins at 8:37 am, find the probability that the plane is prepared for departure on schedule.

2.2 Solution

Since time is a continuous variable, it makes sense to model the random amount of time between 20 and 27 minutes as a probability distribution on the interval $(20, 27)$. This does not preclude the possibility that the time is discreet, this can be modeled by a series of dirac delta distributions at the integer points. A distribution $\mu(t)$ is therefore a valid probability distribution for flight times if it is supported only on $(20, 27)$ and is normalized.

Given $\mu(t)$, the expected time that the loading takes is:

$$\langle t\rangle = \int_{20}^{27} t\mu(t)dt \tag{1}$$

If a flight is scheduled to depart at 9:00 am and preparation begins at 8:37 am, then the probability that the filght will be prepared for departure in time is the probability that the preparations take fewer than 23 minutes. This is given by:

$$P\left(\text{Prepared for Departure}\right) = \int_{20}^{23} \mu(t)dt \tag{2}$$

So far I have made no assumptions.

2.3 Example

There is no way of determining what the probability of being prepared for depature is without knowing the probability distribution $\mu(t)$. Let's evaluate the probability with the constant distribution:

$$\mu(t) = \begin{cases} \frac{1}{7} & 20 < t < 27 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$\begin{aligned} P\left(\text{Prepared for Departure}\right) &= \int_{20}^{23} \frac{1}{7} dt \\ &= \frac{1}{7}\left(t\right)_{t=20}^{t=23} \\ &= \frac{1}{7}\left(23 - 20\right) \\ &= \frac{3}{7} \end{aligned} \tag{4}$$

3 Nigerian Con

- The spammer sends the initial solicitation to $N = 10000000$ recipients.
- $x = 5\%$ are Very Gullible (VG) and the remaining $1 - x = 95\%$ are Reasonably Skeptical (RS). The spammer does not know which one is which.
- The spammer's cost of follow-up email exchange is $K = \$20$.
- If the follow-up email exchange is a success, the spammer gets a reward of $L = \$2000$.
- The spammer can send either Well-Written Solicitations (WWS) or Badly Written Solicitations (BWS).
- The probability of initial response from each VG individual is $p_g = 0.01$ regardless of the type of solicitation.
- The probability of initial response from each RS individual is $q_w - 0.002$ if the spammer sends WWS and $q_b = 0.001$ if the spammer sends BWS.

- Once the recipient responds, the probability of successfully fooling him is $r_1 = 0.1$ for VG responders and $r_2 = 0.005$ for RS.

3.1 Expected Cost from Follow-Up Emails

Suppose we send well-written emails.

$$\begin{aligned}
E\left(\text{Responses} \mid \text{WWS}\right) &= xNp_g + (1-x)Nq_w \\
&= 5000 + 19000 \\
&= 24000 \\
E\left(\text{Follow-Up Cost} \mid \text{WWS}\right) &= K\left(xNp_g + (1-x)Nq_w\right) \\
&= K * 5000 + K * 19000 \\
&= 100000 + 380000 \\
&= 480000 \\
E\left(\text{Gross Revenue} \mid \text{WWS}\right) &= E\left(\text{Gross Revenue from VG} \mid \text{WWS}\right) + E\left(\text{Gross Revenue from RS} \mid \text{WWS}\right) \\
&= L\left(E\left(\text{Follow-Ups from RS} \mid \text{WWS}\right) + E\left(\text{Follow-Ups from VG} \mid \text{WWS}\right)\right) \\
&= L\left(r_2 E\left(\text{Responses from RS} \mid \text{WWS}\right) + r_1 E\left(\text{Responses from VG} \mid \text{WWS}\right)\right) \\
&= L\left(r_2 p_g xN + r_1 q_w (1-x)N\right) \\
E\left(\text{Gross Revenue} \mid \text{BWS}\right) &= LN\left(r_1 p_g x + r_2 q_w (1-x)\right) \qquad (5) \\
&= LN\left(5 \times 10^{-5} + 9.5 \times 10^{-6}\right) \\
E\left(\text{Profit} \mid \text{BWS}\right) &= E\left(\text{Gross Revenue} \mid \text{WWS}\right) - E\left(\text{Follow-Up Cost} \mid \text{WWS}\right) \\
&= N\left(L\left(r_2 p_g x + r_1 q_w (1-x)\right) - K\left(xp_g + (1-x)q_w\right)\right) \qquad = \\
&= 710000
\end{aligned} \qquad (6)$$

Most of the expense comes from sending emails to the RS individuals, although they contribute only 15 % of the revenue.

$$
\begin{aligned}
E\left(\text{Responses} \mid \text{BWS}\right) &= xNp_g + (1-x)Nq_b \\
&= 5000 + 9500 \\
&= 14500 \\
E\left(\text{Follow-Up Cost} \mid \text{BWS}\right) &= K\left(xNp_g + (1-x)Nq_b\right) \\
&= K * 5000 + K * 9500 \\
&= 100000 + 190000 \\
&= 290000 \\
E\left(\text{Gross Revenue} \mid \text{BWS}\right) &= E\left(\text{Gross Revenue from VG} \mid \text{BWS}\right) + E\left(\text{Gross Revenue from RS} \mid \text{BWS}\right) \\
&= L\left(E\left(\text{Follow-Ups from RS} \mid \text{BWS}\right) + E\left(\text{Follow-Ups from VG} \mid \text{BWS}\right)\right) \\
&= L\left(r_2 E\left(\text{Responses from RS} \mid \text{BWS}\right) + r_1 E\left(\text{Responses from VG} \mid \text{BWS}\right)\right) \\
&= L\left(r_2 p_g x N + r_1 q_w (1-x) N\right) \\
E\left(\text{Gross Revenue} \mid \text{BWS}\right) &= LN\left(r_1 p_g x + r_2 q_b (1-x)\right) \\
&= LN\left(5 \times 10^{-5} + 4.75 \times 10^{-6}\right) \\
&= \$1095000 \\
E\left(\text{Profit} \mid \text{BWS}\right) &= E\left(\text{Gross Revenue} \mid \text{BWS}\right) - E\left(\text{Follow-Up Cost} \mid \text{BWS}\right) \\
&= N\left(L\left(r_2 p_g x + r_1 q_b (1-x)\right) - K\left(x p_g + (1-x) q_b\right)\right) \\
&= 1095000 - 290000 \\
&= 805000
\end{aligned}
$$

Although the expected revenue doesn't change significantly, the cost of sending follow-ups decreases, making this strategy more profitable.

3.2 How does email cost affect strategy?

We have determined:

$$
\begin{aligned}
E\left(\text{Profit} \mid \text{WWS}\right) &= N\left(L\left(r_2 p_g x + r_1 q_w (1-x)\right) - K\left(x p_g + (1-x) q_w\right)\right) \\
E\left(\text{Profit} \mid \text{BWS}\right) &= N\left(L\left(r_2 p_g x + r_1 q_b (1-x)\right) - K\left(x p_g + (1-x) q_b\right)\right)
\end{aligned}
$$

Therefore, the two strategies are equivalent if $E\left(\text{Profit} \mid \text{WWS}\right) = E\left(\text{Profit} \mid \text{BWS}\right)$ The cost of sending an email therefore doesn't affect the strategy choice at all.

4 Speech Problem

4.1 Problem Statement

Suppose a politician is giving a speech, and wants to know how much emphasis to put on one of her policy positions. We assume that the voters are split into two groups, Pro and Con, on that position. We further assume that the

Candidate has supporters in both groups. The number of people in each group is

Pro?	Support Candidate?	# Individuals
Yes	Yes	r
Yes	No	$n-r$
No	Yes	s
No	No	$m-s$

For each mention of the issue:

- Each Non-Supporter who supports the issue becomes a supporter with probability p.
- Each Supporter who doesn't support the issue becomes a non-supporter with probability q.

4.2 How Many Times to Mention the Issue?

Suppose the candidate mentions the issue once. Then, she can expect to gain $(n-r)p$ supporters, and lose qs supporters.

Now, if she mentions the issue again, the number of supporters has changed:

Pro?	Support Candidate?	# Individuals
Yes	Yes	$r+(n-r)p$
Yes	No	$(n-r)(1-p)$
No	Yes	$s(1-q)$
No	No	$m-s+qs$

After mentioning the issue a second time, she can expect the numbers to be:

Pro?	Support Candidate?	# Individuals
Yes	Yes	$r+(n-r)p+(n-r)(1-p)p$
Yes	No	$(n-r)(1-p)-(n-r)(1-p)p$
No	Yes	$s(1-q)-s(1-q)q$
No	No	$m-s+qs+s(1-q)q$

4.3 Generalization

4.3.1 Determining Net Change in supporters from k mentions

The probability of an individual being convinced after the n-th mentioning of an issue is the probability that they are not convinced the previous $n-1$ times times the probability that they are convinced the n-th time. If the probabilities are the same each time, then if the probability of being convinced each time is p, the probability of being convinced the n-th time is $(1-p)^{n-1}p$.

Therefore, if there are $(n-r)$ individuals who support the candidate, and the candidate mentions the issue k times, the expected number of people who are convinced is the sum of the expected number of people who are convinced each time. Therefore, the expected number of people to be convinced is:

$$
\begin{aligned}
E\left(\#\text{ New Supporters} \mid \text{k mentions}\right) &= (n-r)\left(p+(1-p)p+(1-p)^2p+\cdots+(1-p)^{k-1}p\right)\\
&= (n-r)p\sum_{i=0}^{k-1}(1-p)^i\\
&= (n-r)p\left(\frac{1-(1-p)^k}{1-(1-p)}\right)\\
E\left(\#\text{ New Supporters} \mid \text{k mentions}\right) &= (n-r)\left(1-(1-p)^k\right)\\
E\left(\#\text{ Lost Supporters} \mid \text{k mentions}\right) &= s\left(1-(1-q)^k\right)
\end{aligned}
$$

Therefore, if the candidate mentions the issue k times, the expected change in number of supporters is:

$$
E\left(\#\text{ Net Change in Supporters} \mid \text{k mentions}\right) = (n-r)\left(1-(1-p)^k\right) - s\left(1-(1-q)^k\right)
$$

This makes sense: if the candidate mentions the issue 0 times, then she will expect no change in supporters under this model. If the candidate mentions the issue an infinite number of times, everybody who can be convinced will be, and therefore her support will change by $(n-r)-s$.

4.3.2 Determining the Optimal Continuous Value for k

Let's assume that we can mention the issues a fractional number of times, so that we can use calculus to maximize yield. We will revert to integers later.

$$
\frac{\partial E\left(\#\text{ Net Change in Supporters} \mid \text{k mentions}\right)}{\partial k} = (n-r)\log\left(\frac{1}{1-p}\right)(1-p)^k - s\log\left(\frac{1}{1-q}\right)(1-q)^k
$$

If we set $\frac{\partial E(\#\text{ Net Change in Supporters}|\text{k mentions})}{\partial k} = 0$, then we get:

$$
\begin{aligned}
(n-r)\log\left(\frac{1}{1-p}\right)(1-p)^k &= s\log\left(\frac{1}{1-q}\right)(1-q)^k\\
\log\left((n-r)\log\left(\frac{1}{1-p}\right)\right) + k\log(1-p) &= \log\left(s\log\left(\frac{1}{1-q}\right)\right) + k\log(1-q)\\
k &= \frac{\log\left(\frac{s}{n-r}\frac{\log(1-q)}{\log(1-p)}\right)}{\log\left(\frac{1-p}{1-q}\right)}
\end{aligned}
$$

, assuming that $p \neq q$.

Note that k has to be a positive value, so therefore, for this unique local extremum to be useful, either

$$\frac{s}{n-r}\frac{\log(1-q)}{\log(1-p)} > 1$$
$$\frac{1-p}{1-q} > 1$$

or

$$\frac{s}{n-r}\frac{\log(1-q)}{\log(1-p)} < 1$$
$$\frac{1-p}{1-q} < 1$$

If the unique local extremum exists, then if it is greater than the boundary values, it is the optimum point within the interval. I could check the second derivative, but I think that it will be more tractable to just compare the expected change in support at the extremum of k to the values at 0 and ∞.