Sugar cane harvesting period optimization in Minecraft

Nukelawe

June 3, 2022

1 Theory

Consider a sugar cane plant that is harvested periodically, every t game ticks. The state of the sugar cane is identified by its height ($\in \{1, 2, 3\}$) and the internal age value of the top-most block ($\in \{0, \ldots, 15\}$).

When the sugar cane with height less than 3 receives a random tick its age is incremented by one. If the age is already 15 then it is reset to 0 and the height is incremented by one. On every random tick attempt (of which there are 3 per tick by default) the probability that the sugar cane receives a random tick is $q = \frac{1}{4096}$.

When a sugar cane is harvested its height is set to 1. The age value is reset to 0 unless the height of the sugar cane was already 1 before the harvest in which case the age will be unmodified. We assume that after the harvest the sugar cane is blocked by the harvesting machine for s ticks. This implies that during the first s ticks after the harvest any attempts to increase the age or height of the sugar cane will fail, even age changes that wouldn't cause the sugar cane to increase in height.

Let H_n be the height of the sugar cane plant n ticks after a harvest. Because harvesting removes all but the lowest sugar cane we have $H_0 = 1$. Let R_n be the number of random ticks received by the plant in the last n ticks. It is binomially distributed with $3n$ trials and success probability q, i.e. $R_n \sim B(3n, q)$. At the end of the harvest cycle the sugar cane height distribution is given by

$$\mathbb{P}\left(H_t = i\right) = \begin{cases} \mathbb{P}\left(R_{t-s} + X < 16\right) & \text{if } i = 1 \\ \mathbb{P}\left(16 \leq R_{t-s} + X < 32\right) & \text{if } i = 2 \\ \mathbb{P}\left(32 \leq R_{t-s} + X\right) & \text{if } i = 3 \end{cases} \tag{1}$$

where X is the internal age value of the sugar cane block immediately after the harvest. The reason why R_{t-s} appears in (1) instead of R_t is that for the first s ticks of the harvest cycle the sugar cane is assumed to be blocked by the harvesting machine.

We would now like to optimize the expected harvest rate

$$v = \frac{\langle H_t - 1 \rangle}{t} = \frac{\mathbb{P}\left(H_t = 2\right) + 2\mathbb{P}\left(H_t = 3\right)}{t} \tag{2}$$

by choosing a suitable harvest period t. However, to compute the probabilities in (2) we must know the initial age parameter X, which depends on the previous harvest. With this in mind let us define X_m to be the age of the sugar cane immediately after the mth harvest. The sequence $X_0, X_1, \ldots$ is a Markov process as the age parameter after $m+1$ harvests X_{m+1} only depends on the one after m harvests, X_m.

The relation between X_m and X_{m+1} is characterized by the transition probabilities

$$p_{i,j} := \mathbb{P}\left(X_{m+1} = j \mid X_m = i\right) = \begin{cases} 0 & \text{if } 0 < j < i \\ \mathbb{P}\left(R_{t-s} = j - i\right) & \text{if } j \geq i \\ \mathbb{P}\left(R_{t-s} \geq 16 - i\right) & \text{if } j = 0, i > 0 \\ \mathbb{P}\left(R_{t-s} \geq 16\right) + \mathbb{P}\left(R_{t-s} = 0\right) & \text{if } j = 0, i = 0 \end{cases} \tag{3}$$

and since we know that $R_{t-s} \sim B(3(t-s), q)$ the probabilities that appear have the following analytic expressions.

$$\mathbb{P}\left(R_{t-s} = j - i\right) = \binom{t-s}{j-i} q^{j-i}(1-q)^{t-s-j+i} \tag{4}$$

$$\mathbb{P}\left(R_{t-s} \geq 16 - i\right) = 1 - \sum_{k=0}^{15-i} \binom{t-s}{k-i} q^{k-i}(1-q)^{t-s-k+i} \tag{5}$$

Define the 16-dimensional probability vector $\mathbf{P}_m = \big(\mathbb{P}\left(X_m = 0\right), \mathbb{P}\left(X_m = 1\right), \cdots, \mathbb{P}\left(X_m = 15\right)\big)$ to represent the age distribution after the mth harvest and the 16-by-16 transition matrix

$$\mathbf{T} = \begin{pmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,15} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,15} \\ \vdots & \vdots & \ddots & \vdots \\ p_{15,0} & p_{15,1} & \cdots & p_{15,15} \end{pmatrix} \tag{6}$$

to tabulate the transition probabilities. Now $\mathbf{P}_{m+1} = \mathbf{P}_m \mathbf{T}$ and $\mathbf{P}_m = \mathbf{P}_0 \mathbf{T}^m$. It is probably fair to assume that before the first harvest ($m = 0$) the age state of the sugar cane is 0. This could be because either the sugar can was just planted or the harvesting machine had been turned off for so long that the sugar cane had grown past height 1. Therefore, we set $\mathbf{P}_0 = \big(1, 0, 0, \ldots, 0\big)$, i.e. $X_0 = 0$ with probability 1. Over many harvest cycles the initial age state of the sugar cane will converge on a stationary distribution $\mathbf{P}_\infty = \lim_{m\to\infty} \mathbf{P}_m$, which satisfies

$$\mathbf{P}_\infty = \mathbf{P}_\infty \mathbf{T} \iff \mathbf{P}_\infty(\mathbf{T} - \mathbf{1}) = \mathbf{0}. \tag{7}$$

In other words, $\mathbf{P}_\infty$ is an eigen vector of the transition matrix.

When determining the optimal harvest period it makes sense to use the steady state distribution as the initial sugar cane age distribution as this what we would see after operating the sugar cane farm for a longer time. With this we can finally compute the sugar cane height distribution at the time of harvest starting from (1). Computing the harvest rate (2) only requires us to know the probabilities for sugar cane heights 2 and 3.

$$\mathbb{P}\left(H_t = 2\right) = \sum_{k=0}^{15} \mathbb{P}\left(X_\infty = k\right) \mathbb{P}\left(16 - k \leq R_{t-s} < 32 - k\right) \tag{8}$$

$$\mathbb{P}\left(H_t = 3\right) = \sum_{k=0}^{15} \mathbb{P}\left(X_\infty = k\right) \mathbb{P}\left(R_{t-s} \geq 32 - k\right) \tag{9}$$

where X_∞ is the random variable that represents the age state of the sugar cane at the beginning of the mth harvest for $m \to \infty$.

2 Results

One might ask how quickly the initial age distribution converges on the stationary solution, how many harvests cycles should we wait until the use of the stationary solution in the harvest rate calculations is justified? Figure 1 illustrates how the initial age distribution affects the harvest rate. We assume that there is when the harvesting machine is turned on the sugar cane is immediately broken and its age state is set to 0. Therefore the age distribution at the beginning of the first harvest cycle is assumed to be given by $\mathbf{P}_0 = (1, 0, \ldots, 0)$. Looking at the curve labeled "1. harvest" we notice that to maximize the harvest rate for the first harvest cycle, the harvest period should be about 41 minutes, which coincides with a *local* maximum of the stationary case. To maximize the rate for the *second* harvest rate however, we would have to set the harvest period to about 11 minutes. The location of the global maximum keeps increasing as more and more harvest cycles are considered until in the stationary case the maximum is obtained at $t = 0$.

This begs the question how much better it would be to use an adaptive harvest period that changes based on the number of harvest cycles since turning the harvesting machine on. In particular, we would set the delay between the $(m-1)$th and mth harvests to value at which the harvest rate for the mth harvest is maximized. However, because the global maximum of the 1st harvest rate curve occurs at the same point as that of the

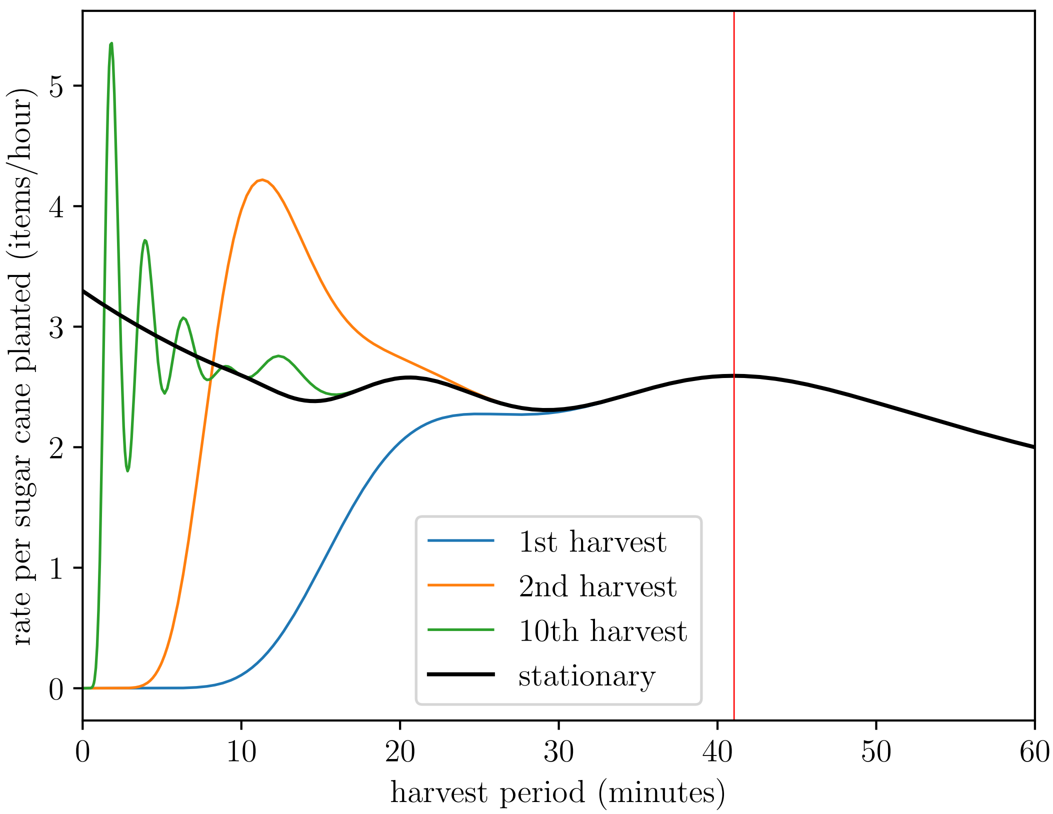

Figure 1: The harvest rate v (see (2)) graphed for different harvest periods t. Here it was assumed that the harvesting machine does *not* block plant growth at all, i.e. $s = 0$. The different curves represent harvest rates depending on the number of harvest cycles since the sugar cane was planted or the harvesting machine was turned on. The peak of the 1$^{\text{st}}$ harvest rate curve, denoted by the red vertical line, occurs at $t = 49252$ ticks or about 41 minutes.

stationary curve we conclude that such an adaptive strategy would just result in regular harvest periods of about 41 minutes, which would not be adaptive at all.

Figure 1 shows that for the first couple of harvest cycles choosing a small (< 10 minutes) harvest period would not be very productive at all, but eventually as more harvest cycles are considered a small period will actually result in higher rates than the 41 minute strategy. It would be interesting to know how many harvest cycles must be considered before using a small harvest period becomes more efficient. To figure this out we compute the expected harvest rate over the first m harvest cycles. Figure 2 shows how long a harvesting machine must be operated until it starts to perform better than the 41 minute strategy. Even though it might take hundreds of harvest cycles to reach this point, the total time is not that long since the harvest cycles also become quite short.

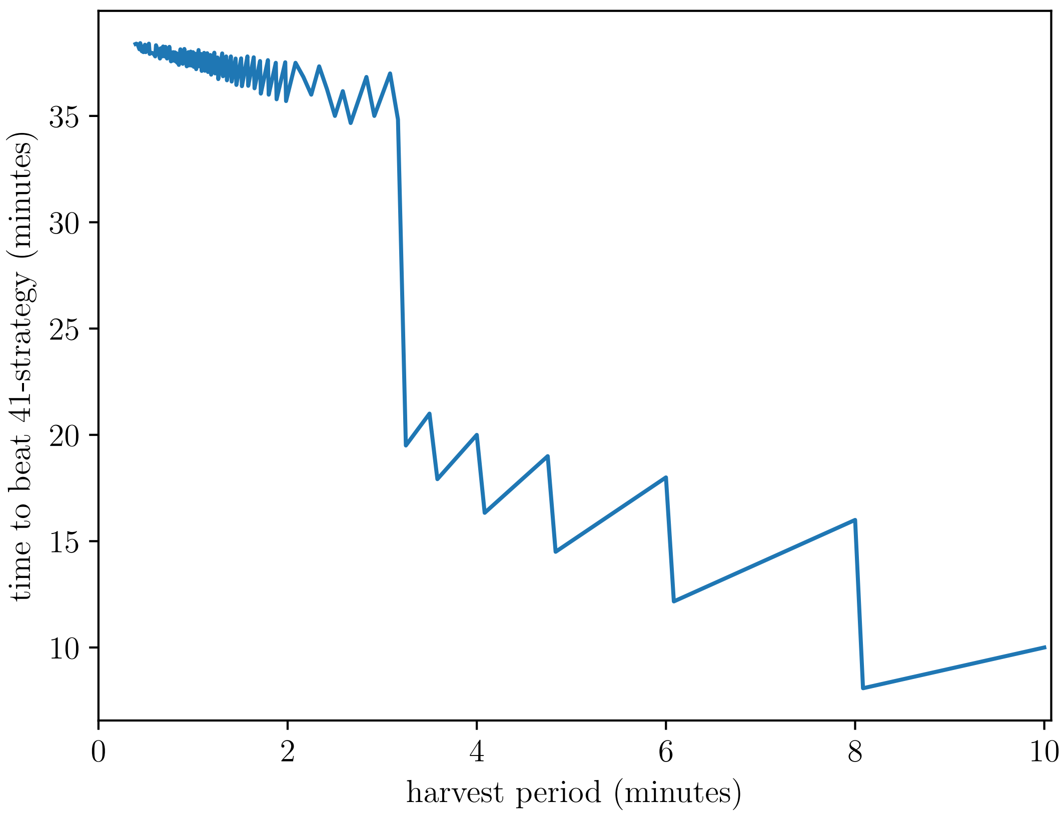

Figure 2: Graph showing how long a sugar cane farm should be run with a small clock speed to reach at least as good a rate as by harvesting every 41 ticks.

So far we have only considered no growth blocking by the harvesting machine ($s = 0$). Figure 3 shows how the harvest rates (assuming stationary age distribution) are affected by growth blocking. Unsurprisingly, this has very little effect on the long harvest periods. Now the global maximum is attained at some non-zero value of the harvest period since for $t \leq s$ growth is prevented all the time and therefore the harvest rate is 0.

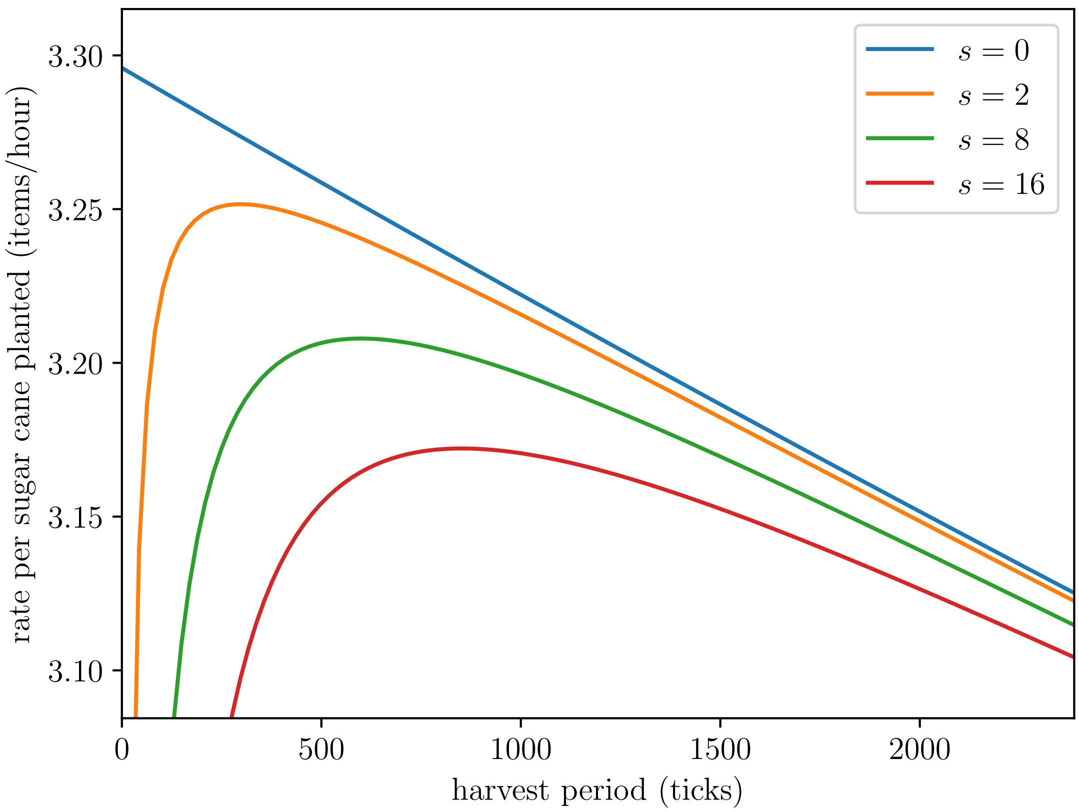

Figure 3: Effect of growth blocking on the harvest rates. Only smaller harvest periods are shown.