High Frequency Sampling Simulation

Updated: 2020-09-03

Simulation Goal The goal of this simulation is to demonstrate that high-frequency sampling can better detect rare species than conventional sampling when controlling for the total number of samples.

Simulation Overview

The simulations should (somewhat) mirror Fig 1 from *Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One.*

Occupancy Framework Data is simulated from an occupancy model framework where

$$Z_i \sim Bernoulli(\psi_i), \tag{1}$$

$$Y_{itj} \sim Binomial(n_j, \psi_i p_{it}), \;\; p_{it} = \frac{exp(x_{it}\beta)}{1 + \exp(x_{it}\beta)} \tag{2}$$

where Z_i is the latent occupancy at site i, ψ_i is the probability that the species is present at site i, Y_{itj} is the observed occupancy at site i, time t, and for the j^{th} replicate, and p_{it} is the probability that the species will be detected at site i and time t, given presence.

The simulations do not actually fit occupancy models, but rather, just focus on the sampling process and whether a rare species is detected, at least, once.

Simulation 1 Simulation 1 uses a fixed occupancy probability (ψ) and fixed detection probability (p). The study explores the impact of different levels of $\psi = \{.05, .10, .15\}$, $p = \{.05, .10, .15\}$, and the total number of samples. A high-frequency approach takes daily samples, whereas, the weekly sampling approach takes 7 samples once a week. With this study, we assume that the species is rare and any detection would lead to actionable interventions. Hence, we compare the total frequency of the sampling regimes, either daily samples for a certain number of weeks or 7 samples collected for a certain number of weeks, that result in at least one detection.

There are two possible ways to think about collecting a batch of seven samples on a single day. The first approach treats the batch of samples as subsamples; in other words, the organism is present, or not, and each subsample has a probability of detecting the organism, given that it is present. The second approach would treat the seven samples as independent samples, where for each sample the organism is present or not, and the sample can be detected with a given probability of the organism is present.

In practice, either scenario is plausible, but it likely depends on the underlying ecological process and how samples are collected. If the sampling process involves collecting a set of water samples at a single location and time point, then the first approach, subsampling, is most relevant. However, if samples could be spread out over time, and potentially space, then using independent samples is reasonable.

The results for the subsampling setting can be seen in Figure 1. As the species becomes easier to detect, there is separation between the two approaches, where the high frequency sampling has a larger proportion of

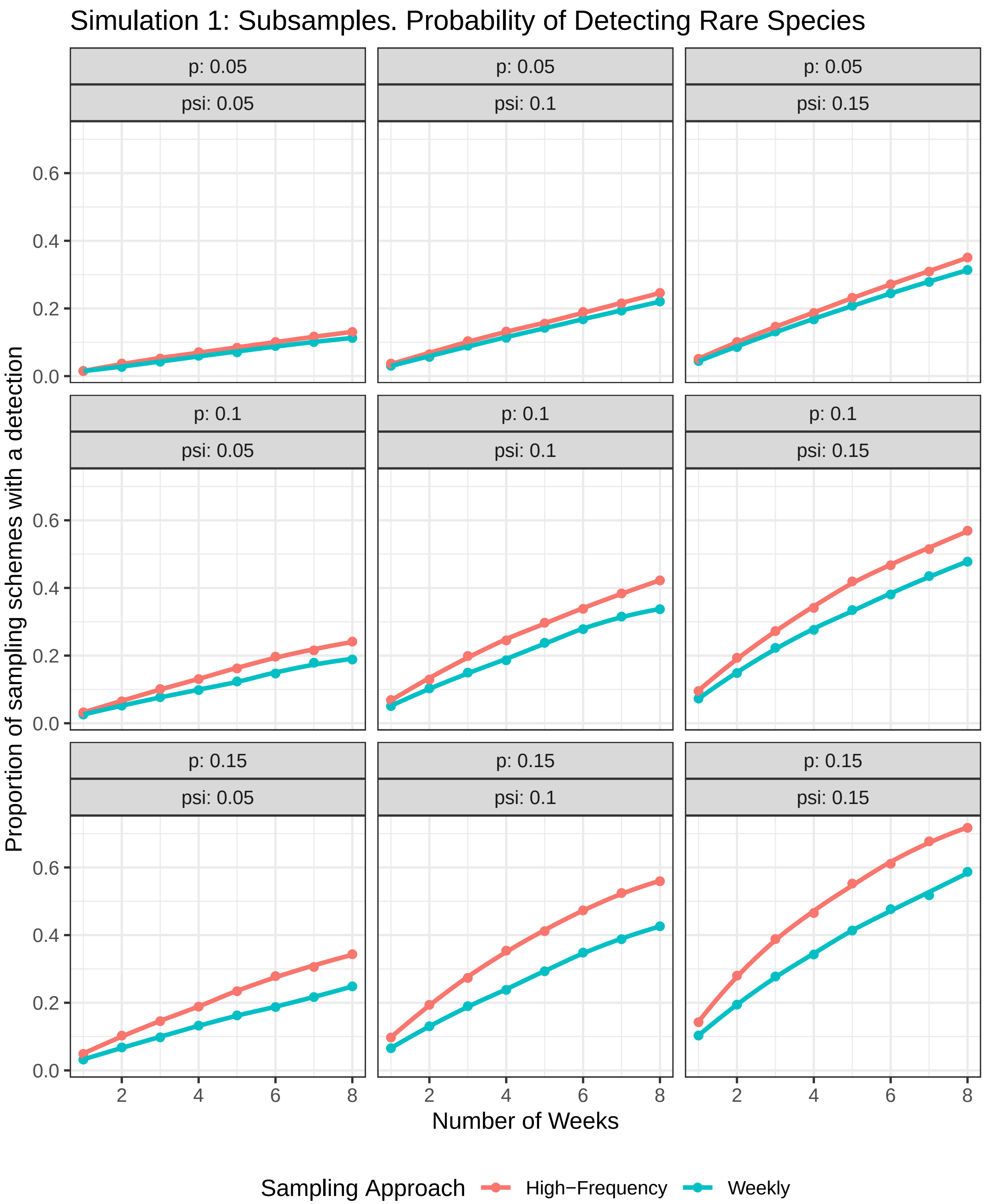

Figure 1: Simulation 1: Subsampling. Probability of detection as a function of p, psi, and number of weeks. The high-frequency approach collects samples daily and the weekly approach takes 7 samples once a week. The figure shows that, even when the total number of samples are the same, the high frequency approach detects the rare species in a higher frequency of the sampling schemes.

sampling regimes that end up with at least one positive detection. The result is not surprising as subsamples would inherently have less information than independent samples.

Figure 2 shows the independent sampling procedure, which, in contrast to Figure 1, does not show any differences with the high-frequency sampling approach.

Simulation 2 Simulation 2 once again uses a fixed occupancy probability (ψ), but now the detection probability (p) changes with time. The study explores the impact of different levels of $\psi = \{.05, .10, .15\}$ and the total number of samples. Furthermore, p will be vary stochastically on a day-to-day basis, but have a median value of either $\{.05, .10, .15\}$. Similar to simulation 1, we will compare the high-frequency sampling approach with both the sub-sampling and independent sample frameworks. For clarity the detection probability will be the same for all independent samples within a day.

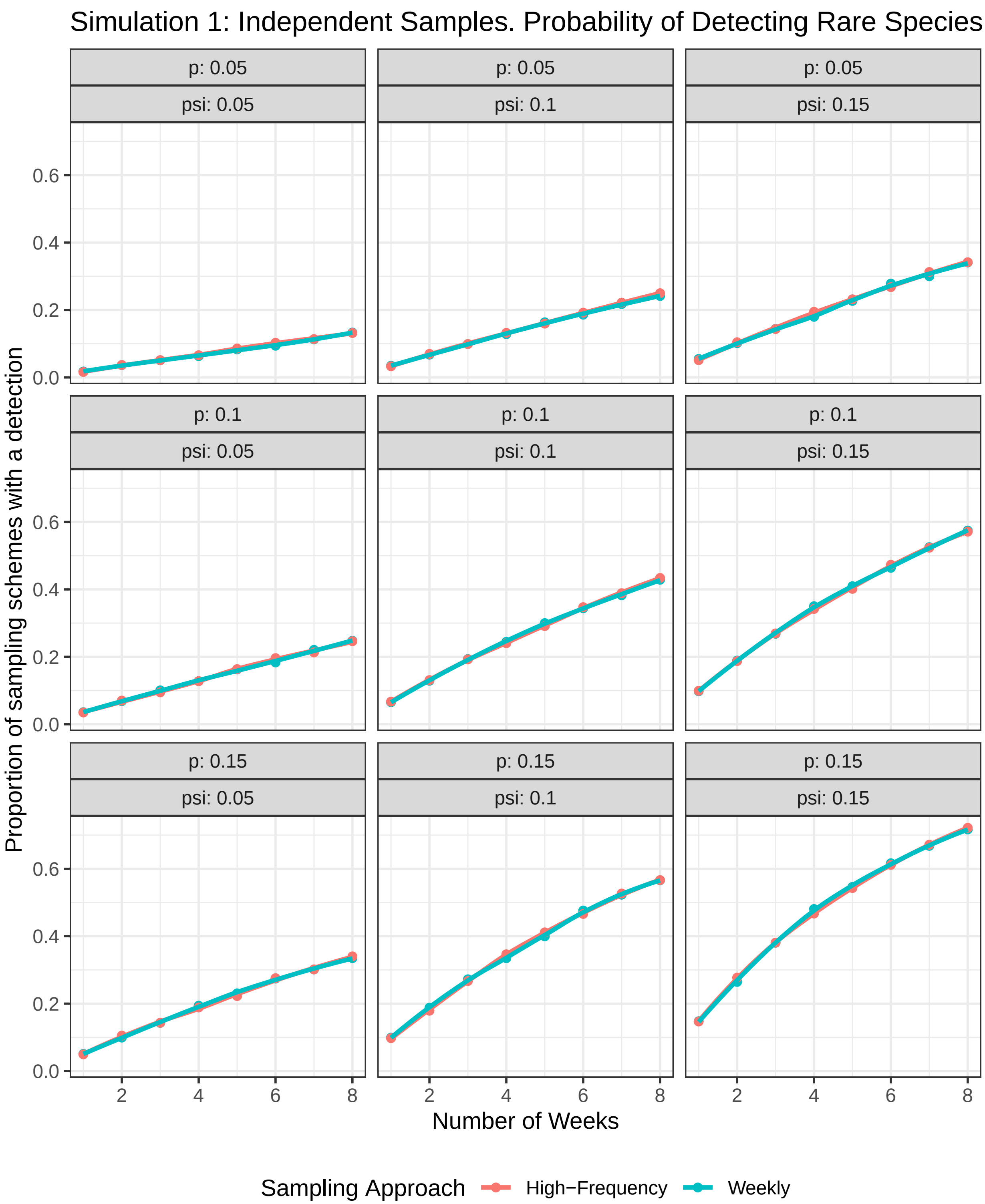

Figure 2: Simulation 1: Independent Samples. Probability of detection as a function of p, psi, and number of weeks. The high-frequency approach collects samples daily and the weekly approach takes 7 independent samples once a week. The figure shows that the detection probabilities are virtually indistinguishable.

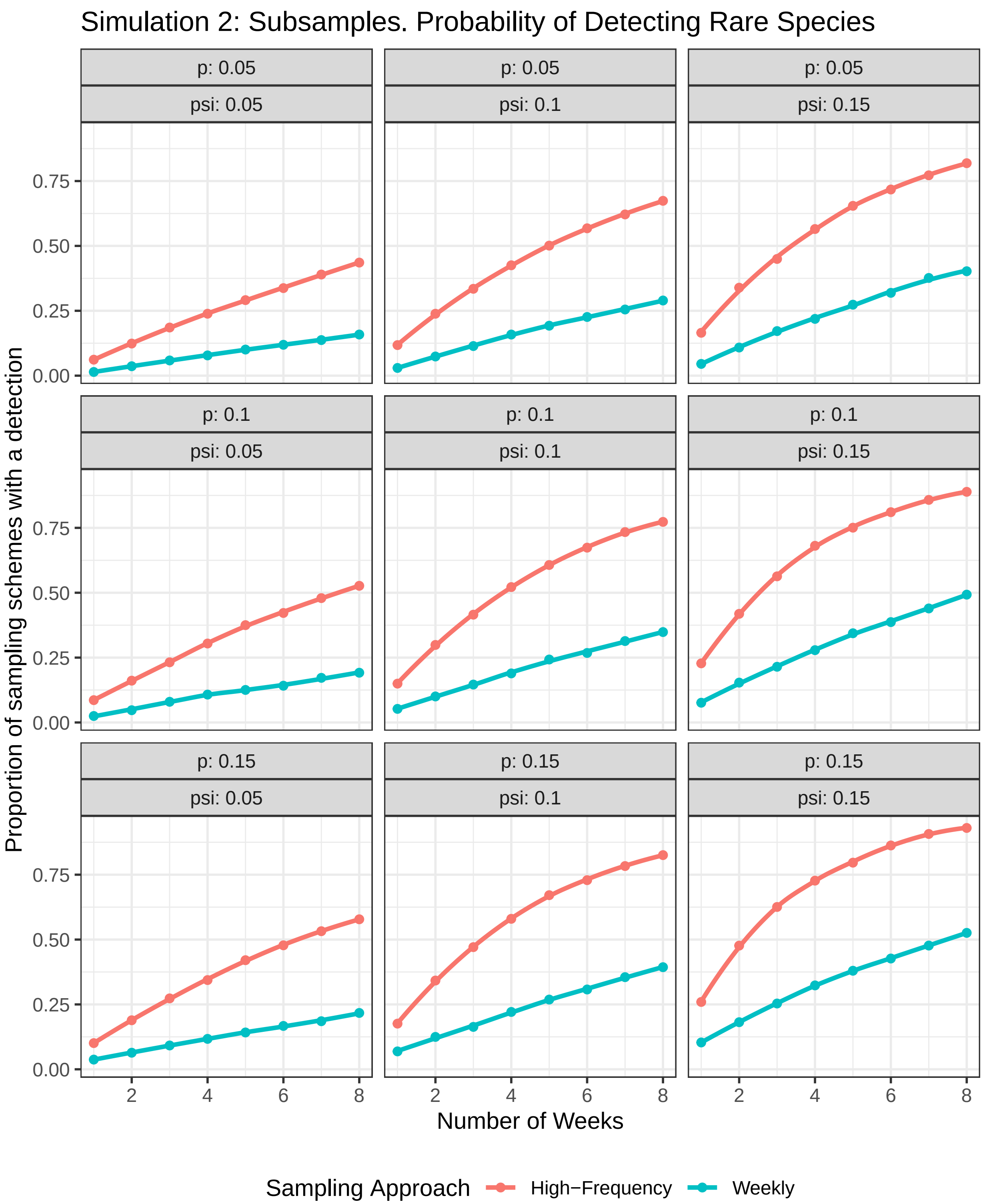

Figure 3: Simulation 2: Subsampling. Comparison of frequency of detection for seven subsamples collected on a single day and daily high-frequency samples with the detection probability changes daily. Similar to the results from simulation 1, the high-frequency approach has a higher detection rate than the weekly sampling.

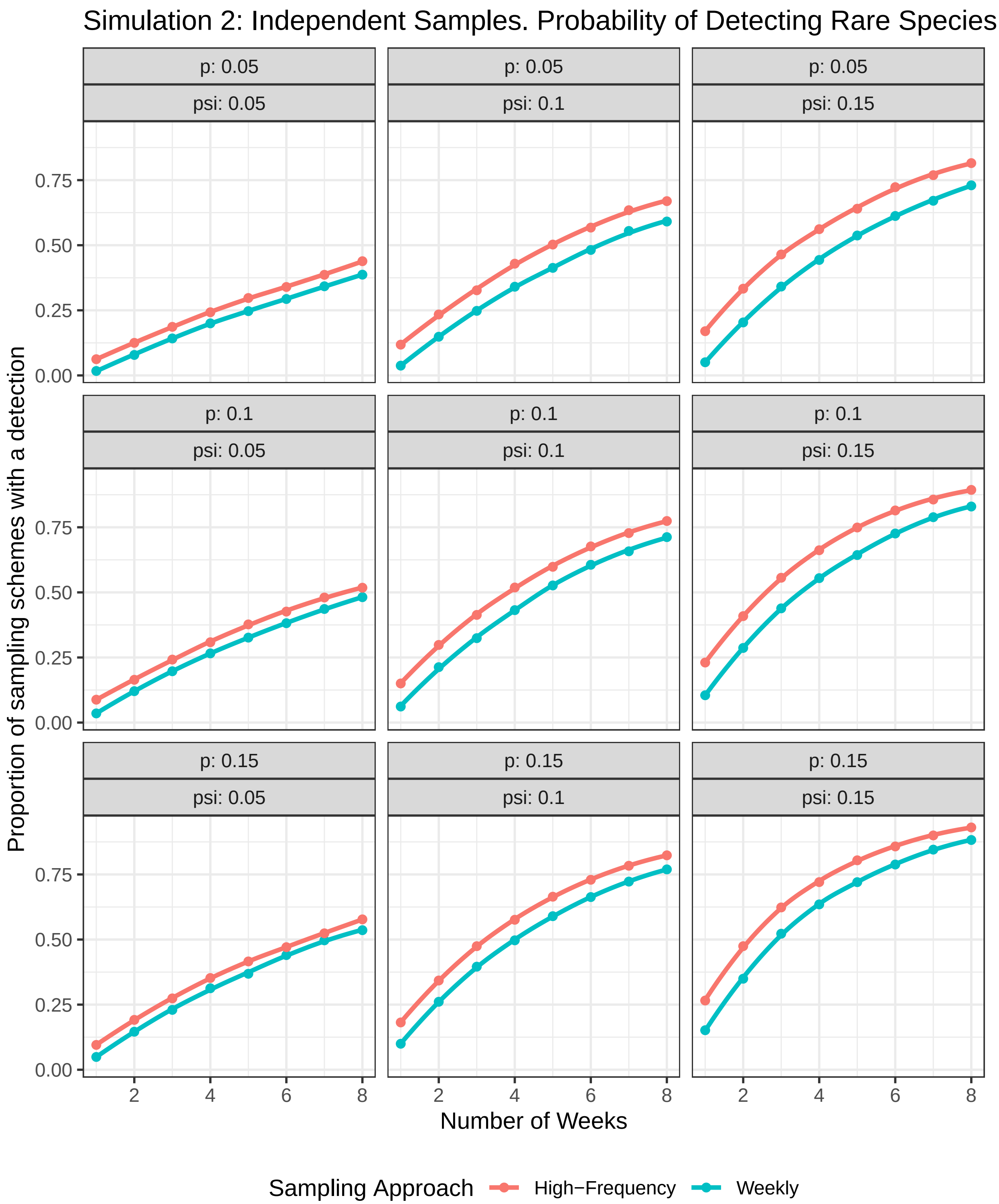

Figure 4: Simulation 2: Independent Samples. Comparison of frequency of detection for seven independent samples collected on a single day and daily high-frequency samples with the detection probability changes daily. In contrast to the results from simulation 1, the high-frequency approach has a higher detection rate than the weekly sampling.