Supplementary Information

Pieter Libin[1,2], Timothy Verstraeten[1], Diederik M. Roijers[1], Jelena Grujic[1], Kristof Theys[2], Philippe Lemey[2], and Ann Nowé[1]

[1]Artificial Intelligence lab, Department of computer science, Vrije Universiteit Brussel, Brussels, Belgium
[2]KU Leuven - University of Leuven, Rega Institute for Medical Research, Clinical and Epidemiological Virology, Leuven, Belgium

July 6, 2018

1 Introduction

This Supplementary Information document accompanies the paper titled "Bayesian Best-Arm Identification for Selecting Influenza Mitigation Strategies".

In this document we provide a proof for BayesGap's simple regret bound (Section 2). We provide a full derivation of the integral used to compute the probability of success (Section 3). We provide some more background on the model choice (Section 4) and the seeding strategy (Section 5). In Section 6, we describe the computational resources that were used to execute the simulations. Furthermore, we provide additional figures that were omitted from the main manuscript: figures for the outcome (i.e., epidemic size) distributions (Section 7), figures for the experimental success rates (Section 8), figures for the experimental success rates with confidence intervals (Section 9), figures for the probabilities of success (i.e., P_s values) per budget (Section 10) and figures for the binned distribution over P_s values (Section 11).

2 BayesGap simple regret bound for T-distributed posteriors

Lemma 1. *Consider a Jeffrey's prior $(\mu_k, \sigma_k^2) \sim \sigma_k^{-3}$over the parameters of the Gaussian reward distributions. Then the posterior mean of arm k has the following non-standardized t-distribution at pull n_k:*

$$\mu_k \mid \bar{x}_{k,n_k}, S_{k,n_k} \sim \mathcal{T}_{n_k}(\bar{x}_{k,n_k}, n_k^{-1}\sqrt{S_{k,n_k}})$$

where n_k is the number of pulls for arm k, $\bar{x}_{k,n_k}$ is the sample mean and S_{k,n_k} is the sum of squares.

Proof. This lemma was presented and proved by Honda et al. [7]. □

Lemma 2. *Consider a random variable $X \sim \mathcal{T}_\nu(\mu, \lambda)$ with variance $\sigma^2 = \frac{\nu}{\nu-2}\lambda^2$, $\nu > 2$ and $\beta > 0$. The probability that X is within a radius $\beta\sigma$ from its mean can then*

be written as:

$$P(|X-\mu|<\beta\sigma) \geq 1 - 2\frac{\nu}{\nu-1}\frac{C(\nu)}{\beta}\left(1+\frac{\beta^2}{\nu}\right)^{-0.5(\nu-1)}$$

where

$$C(\nu) = \frac{\Gamma(0.5\nu+0.5)}{\Gamma(0.5\nu)\sqrt{\pi\nu}}$$

is the normalizing constant of a standard t-distribution.

Proof. Consider a random variable $Z \sim \mathcal{T}_\nu(0,1)$, $\nu > 2$ and $\beta > 0$. Then the probability of Z being greater than $\beta\sqrt{\frac{\nu}{\nu-2}}$ is:

$$\begin{aligned}
P(Z > \beta\sqrt{\frac{\nu}{\nu-2}}) &\overset{(1)}{=} \int_{\beta\sqrt{\frac{\nu}{\nu-2}}}^{+\infty} \mathcal{T}_\nu(z \mid 0,1)dz \\
&= C(\nu)\int_{\beta\sqrt{\frac{\nu}{\nu-2}}}^{+\infty}\left(1+\frac{z^2}{\nu}\right)^{-0.5(\nu+1)}dz \\
&\overset{(2)}{\leq} C(\nu)\int_{\beta\sqrt{\frac{\nu}{\nu-2}}}^{+\infty}\frac{z}{\beta\sqrt{\frac{\nu}{\nu-2}}}\left(1+\frac{z^2}{\nu}\right)^{-0.5(\nu+1)}dz \\
&= -\frac{\nu}{\nu-1}\frac{\sqrt{\nu-2}}{\beta\sqrt{\nu}}C(\nu)\int_{\beta\sqrt{\frac{\nu}{\nu-2}}}^{+\infty} -\frac{\nu-1}{\nu}z\left(1+\frac{z^2}{\nu}\right)^{-0.5(\nu+1)}dz \\
&\overset{(3)}{=} -\frac{\sqrt{\nu(\nu-2)}}{\nu-1}\frac{C(\nu)}{\beta}\left(1+\frac{z^2}{\nu}\right)^{-0.5(\nu-1)}\Bigg|_{\beta\sqrt{\frac{\nu}{\nu-2}}}^{+\infty} \\
&\overset{(4)}{=} \frac{\sqrt{\nu(\nu-2)}}{\nu-1}\frac{C(\nu)}{\beta}\left(1+\frac{\beta^2}{\nu-2}\right)^{-0.5(\nu-1)}
\end{aligned}$$

The probability of Z being greater than the lower bound $\beta\sqrt{\frac{\nu}{\nu-2}}$ is the integral over its probability density function, starting from that lower bound (1). In the integral, we introduce a factor $\frac{z}{\beta\sqrt{\frac{\nu}{\nu-2}}}$, which is greater than 1 for the considered values of z (2). We then take note of the following derivative, and use this result to analytically solve the integral (3):

$$\frac{d}{dx}\left(1+\frac{x^2}{\nu}\right)^{-0.5(\nu-1)} = -\frac{\nu-1}{\nu}x\left(1+\frac{x^2}{\nu}\right)^{-0.5(\nu+1)}$$

Finally, we solve the primitive from $\frac{z}{\beta\sqrt{\frac{\nu}{\nu-2}}}$ to infinity (4).

Next, we apply a union bound to obtain a lower bound on the probability that the magnitude of Z is smaller than $\beta\sqrt{\frac{\nu}{\nu-2}}$:

$$P(|Z| < \beta\sqrt{\frac{\nu}{\nu-2}}) \geq 1 - 2\frac{\sqrt{\nu(\nu-2)}}{\nu-1}\frac{C(\nu)}{\beta}\left(1+\frac{\beta^2}{\nu-2}\right)^{-0.5(\nu-1)}$$

Finally, consider $Z = \frac{(X-\mu)}{\lambda}$:

$$P(|X-\mu| < \beta\sqrt{\frac{\nu}{\nu-2}}\lambda) \geq 1 - 2\frac{\sqrt{\nu(\nu-2)}}{\nu-1}\frac{C(\nu)}{\beta}\left(1+\frac{\beta^2}{\nu-2}\right)^{-0.5(\nu-1)}$$

□

Lemma 3. *Consider a K-armed bandit problem with budget T and K arms. Let $U_k(t)$ and $L_k(t)$ be upper and lower bounds that hold for all times $t \leq T$ and all arms $k \leq K$ with probability $1-\delta_k(t)$. Finally, let g_k be a monotonically decreasing function such that $U_k(t) - L_k(t) \leq g_k(n_k(t-1))$ and $\sum_{k=1}^{K} g_k^{-1}(H_{k,\epsilon}) \leq T-K$. We can then bound the simple regret R_T as:*

$$P(R_T < \epsilon) \geq 1 - \sum_{k=1}^{K}\sum_{t=1}^{T}\delta_k(t)$$

Proof. First, we define $\mathcal{E}$ as the event in which every mean μ_k is bounded by its associated bounds (i.e., $U_k(t)$ and $L_k(t)$) for each time step [6].

$$\mathcal{E} := \forall k \leq K, \forall t \leq T : L_k(t) \leq \mu_k \leq U_k(t)$$

The probability of μ_k deviating from a single bound at time t is by definition $\delta_k(t)$. When applying the union bound, we obtain $P(\mathcal{E}) \geq 1 - \sum_{k=1}^{K}\sum_{t=1}^{T}\delta_k(t)$. The probability of regret is equal to the probability of the event $\mathcal{E}$ occurring, as proven in [6]. □

Theorem 1. *Consider a K-armed Gaussian bandit problem with budget T and unknown variance. Let σ_G^2 be a generalization of that variance over all arms, and $U_k(t)$ and $L_k(t)$ respectively be the upper and lower bounds for each arm k at time t, where $U_k(t) = \hat{\mu}_k(t) + \beta\hat{\sigma}_k(t)$ and $L_k(t) = \hat{\mu}_k(t) - \beta\hat{\sigma}_k(t)$. The simple regret is then bounded as:*

$$\begin{aligned} P(R_T \leq \epsilon) &\geq 1 - 2\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\sqrt{n_k(t)(n_k(t)-2)}}{n_k(t)-1}\frac{C(n_k(t))}{\beta}\left(1+\frac{\beta^2}{n_k(t)-2}\right)^{-0.5(n_k(t)-1)} \\ &\geq 1 - O\left(KT\left(1+\frac{\beta^2}{\min\limits_{k,t} n_k(t)}\right)^{-0.5\min\limits_{k,t} n_k(t)}\right) \end{aligned}$$

where:

$$\beta = \sqrt{\frac{T-3K}{4H_\epsilon\sigma_G^2}}$$

Note that when $\min\limits_{k,t} n_k(t) \to +\infty$, the bound decreases exponentially in β, similar to the problem setting presented in [6]. Intuitively, this result makes sense, as for known variances, a Gaussian can be used to describe the posterior means, and indeed, as the number of pulls approaches infinity, our t-distributions converge to Gaussians.

Proof. According to Lemma 1, the posterior over the average reward is a t-distribution with scaling factor $\lambda_k(t) = n_k(t)^{-1}\sqrt{S_{k,n_k(t)}}$. Therefore,

$$\begin{aligned} U_k(t+1) - L_k(t+1) &= 2\beta\hat{\sigma}_k(t) \\ &\overset{(1)}{=} 2\beta\sqrt{n_k(t)(n_k(t)-2)^{-1}\lambda_k(t)^2} \\ &\overset{(2)}{=} \sqrt{n_k(t)(n_k(t)-2)^{-1}n_k(t)^{-2}S_{k,n_k(t)}} \\ &= \sqrt{(n_k(t)-2)^{-1}\frac{S_{k,n_k(t)}}{n_k(t)}} \\ &\overset{(3)}{=} \sqrt{(n_k(t)-2)^{-1}s_k^2(t)} \\ &\overset{(4)}{=} g_k(n_k(t)) \end{aligned}$$

The variance of a t-distribution equals $\frac{n_k(t)}{n_k(t)-2}\lambda_k(t)^2$ for arm k at time t, with scaling factor $\lambda_k(t)$ as described in Lemma 1 (1 + 2). We denote the variance over rewards per arm as $s_k^2(t)$ (3) and define $g_k(n_k(t))$ to be the upper bound expression as specified in Lemma 3 (4).

Next, we compute the inverse of $g_k(n)$:

$$g_k^{-1}(m) = \frac{4\beta^2 s_k^2(t)}{m^2} + 2$$

We generalize $s_k^2(t)$ to a variance σ_G^2 representative for all arms.[1] Approximating the hardness of the problem as $H_\epsilon = \sum_k H_{k,\epsilon}^{-2}$, where $H_{k,\epsilon}$ is the arm-dependent hardness defined in [6], we obtain β as follows:

$$\begin{aligned} &\sum_{k=1}^{K} g_k^{-1}(H_{k\epsilon}) \approx 4\beta^2\sigma_G^2 H_\epsilon + 2K = T - K \\ &\Leftrightarrow \beta = \sqrt{\frac{T-3K}{4H_\epsilon\sigma_G^2}} \end{aligned}$$

Finally, as the conditions in Lemma 3 on the function g_k are now satisfied, the simple regret bound can be obtained using Lemma 3 and the probability that the true mean is out of the arm-specific bounds $U_k(t)$ and $L_k(t)$, given in Lemma 2. □

3 Probability of success

As Top-two Thompson sampling recommends the arm with the highest average reward, and we assume that the arm's reward distributions are independent, the probability of success can be computed using :

[1] In the main paper, we choose $\sigma_G^2 = \bar{s}_G^2$ to be the mean over all arm-specific variances obtained after the initialization phase.

$$
\begin{aligned}
P(\mu_J = \max_{1 \le k \le K} \mu_k) &= P(\cap_{k \neq J}^{K} (\mu_k \le \mu_J)) \\
&= \int_{x \in \mathbb{R}} P(\cap_{k \neq J}^{K} (\mu_k \le x)) P(\mu_J = x) dx \\
&= \int_{x \in \mathbb{R}} \big[\prod_{k \neq J}^{K} P(\mu_k \le x) \big] P(\mu_J = x) dx \\
&= \int_{x \in \mathbb{R}} \big[\prod_{k \neq J}^{K} F_{\mu_k}(x) \big] f_{\mu_J}(x) dx
\end{aligned}
$$

where $\mu_{_J}$ is the random variable that represents the mean of the recommended arm's reward distribution, f_{μ_J} is the recommended arm's posterior probability density function and F_{μ_k} is the other arms' cumulative density function.

4 Model choice

The epidemiological model used in the experiments is the FluTE stochastic individual-based model. In our experiment we consider the population of Seattle (United States) that includes 560,000 individuals [2]. This population is realistic both with respect to the number of individuals and its community structure [8]. In [8], it is shown that the Seattle model behaves similarly to the Los Angeles county model with respect to symptomatic attack rate. This demonstrates that the Seattle model constitutes a realistic proxy for an even larger population, and is well suited to evaluate vaccine strategies. As the computational complexity of FluTE simulations depends both on the size of the susceptible population and the proportion of the population that becomes infected, using the Seattle model allowed us to determine a ground truth, which would not have been computationally tractable using the Los Angeles county model. This ground truth enabled us to validate the performance of our framework.

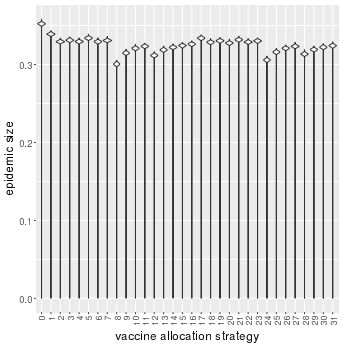
5 Seeding strategy

Following many other research efforts [3, 2, 5, 1], we use a static seeding strategy. As we show in the paper, the number of infection seeds affects the epidemic size distribution (i.e., the bimodal nature of this distribution), which we accommodate in our bandit model censoring the outcome distribution. According to [8, 4], there is no other concern with respect to the number of seeds on the epidemic size.

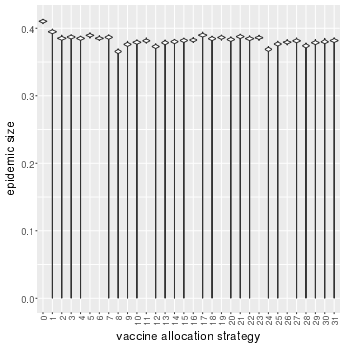
6 Computational resources

The simulations were run on a high performance cluster (HPC). On this HPC, we used 'Ivy Bridge' nodes, more specifically nodes with two 10-core "Ivy Bridge" Xeon E5-2680v2 CPUs (2.8 GHz, 25 MB level 3 cache) and 64 GB of RAM. This infrastructure allowed us to run 20 FluTE simulations per node.

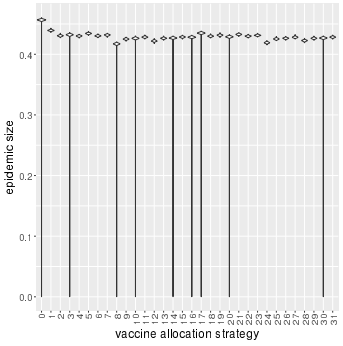
7 Outcome (i.e., epidemic size) distributions



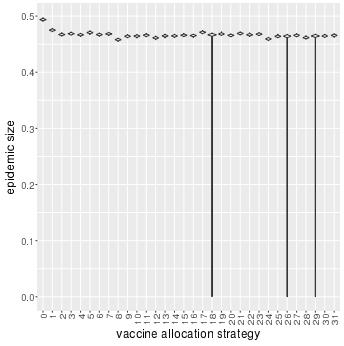
(a) Outcome distributions for $R_0 = 1.6$.



(b) Outcome distributions for $R_0 = 1.8$.

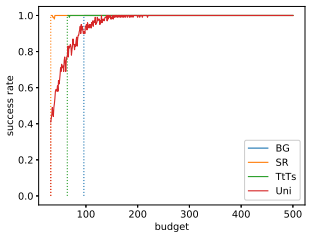


(c) Outcome distributions for $R_0 = 2.0$.

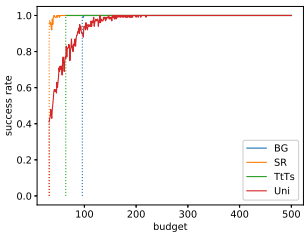


(d) Outcome distributions for $R_0 = 2.2$.

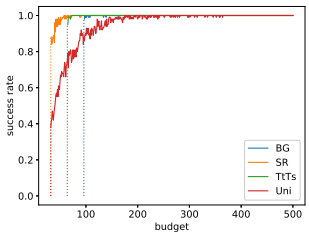
8 Bandit run success rates



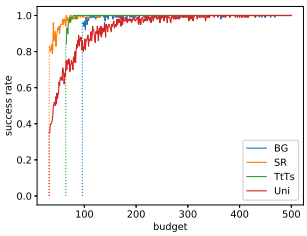
(a) Bandit run results for $R_0 = 1.6$.



(b) Bandit run results for $R_0 = 1.8$.



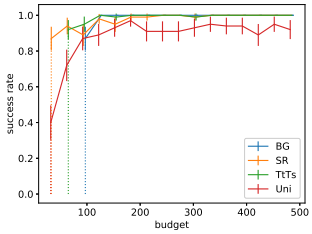
(c) Bandit run results for $R_0 = 2.0$.



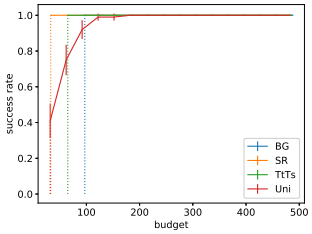
(d) Bandit run results for $R_0 = 2.2$.

9 Bandit run success rates with confidence intervals

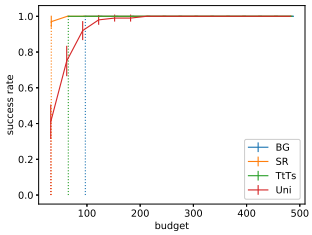
We show the bandit run success rates with a Clopper-Pearson confidence interval.



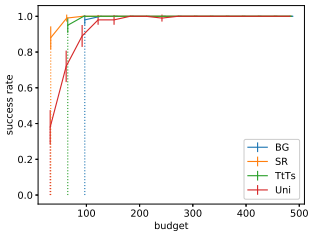
(a) Bandit run results for $R_0 = 1.4$.



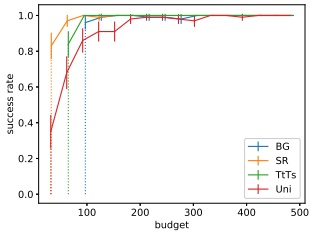
(b) Bandit run results for $R_0 = 1.6$.



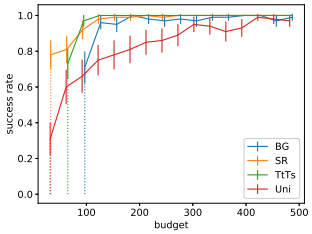
(c) Bandit run results for $R_0 = 1.8$.



(d) Bandit run results for $R_0 = 2.0$.

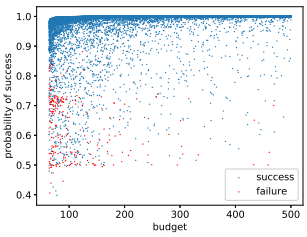


(e) Bandit run results for $R_0 = 2.2$.

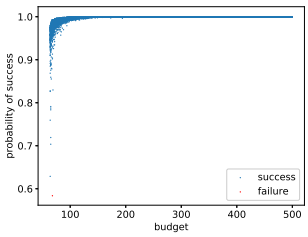


(f) Bandit run results for $R_0 = 2.4$.

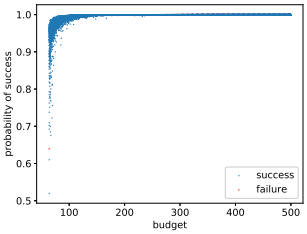
10 P_s values for Top-two Thompson sampling



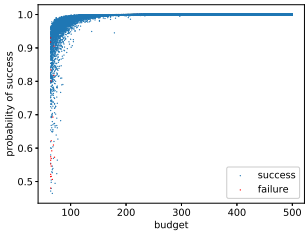
(a) P_s values for $R_0 = 1.4$.



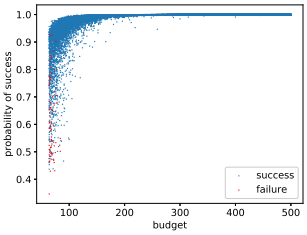
(b) P_s values for $R_0 = 1.6$.



(c) P_s values for $R_0 = 1.8$.

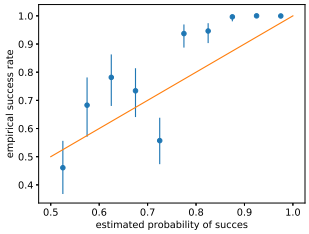


(d) P_s values for $R_0 = 2.0$.

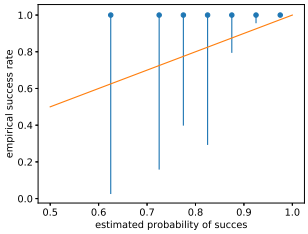


(e) P_s values for $R_0 = 2.2$.

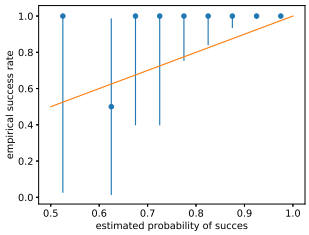
11 Binned distribution of P_s values for Top-two Thompson sampling



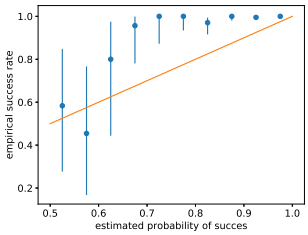
(a) Binned distribution for $R_0 = 1.4$.



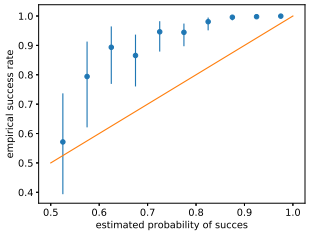
(b) Binned distribution for $R_0 = 1.6$.



(c) Binned distribution for $R_0 = 1.8$.



(d) Binned distribution for $R_0 = 2.0$.



(e) Binned distribution for $R_0 = 2.2$.

References

[1] S. Andradóttir, W. Chiu, D. Goldsman, M. L. Lee, K.-L. Tsui, B. Sander, D. N. Fisman, and A. Nizam. Reactive strategies for containing developing outbreaks of pandemic influenza. *BMC public health*, 11(1):S1, 2011.

[2] D. L. Chao, M. E. Halloran, V. J. Obenchain, and I. M. Longini Jr. FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, 6(1):e1000656, 2010.

[3] N. M. Ferguson, D. A. T. Cummings, S. Cauchemez, C. Fraser, and Others. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209, 2005.

[4] T. C. Germann, K. Kadau, I. M. Longini, and C. A. Macken. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.

[5] M. E. Halloran, N. M. Ferguson, S. Eubank, I. M. Longini, D. A. T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. C. Germann, and Others. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644, 2008.

[6] M. Hoffman, B. Shahriari, and N. Freitas. On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning. In *Artificial Intelligence and Statistics*, pages 365–374, 2014.

[7] J. Honda and A. Takemura. Optimality of Thompson Sampling for Gaussian Bandits Depends on Priors. In *AISTATS*, pages 375–383, 2014.

[8] L. Willem, S. Stijven, E. Vladislavleva, J. Broeckhove, P. Beutels, and N. Hens. Active Learning to Understand Infectious Disease Models and Improve Policy Making. *PLoS Comput Biol*, 10(4):e1003563, 2014.