STOR608 Sprint 4: Statistical Learning for Decisions

Group 2:
Graham Burgess, Max Howell, James Neill, Adam Page
Sprint Leader: James Grant

15 December 2022

Introduction

We will be investigating three problems:

- Part A: Restaurant problem
- Part B: Two-armed bandit problem
- Part C: Expected improvement problem

Restaurant Data

We have some data on the quality of various restaurants in Lancaster, rated on a 0-10 scale.

Restaurant Data

We have some data on the quality of various restaurants in Lancaster, rated on a 0-10 scale.

Aquila	Bella Italia	Cafe Dolce	Domino's
4	2	5	3
5	6		5
			6

Restaurant Data

We have some data on the quality of various restaurants in Lancaster, rated on a 0-10 scale.

Aquila	Bella Italia	Cafe Dolce	Domino's
4	2	5	3
5	6		5
			6

Which restaurant should we go to?

Model

For each restaurant k, we assume that $r(k) \sim \mathrm{N}(\mu_k, \sigma_k^2)$, independent across restaurants.

Model

For each restaurant k, we assume that $r(k) \sim \mathsf{N}(\mu_k, \sigma_k^2)$, independent across restaurants.

We then assume standard priors

$$\begin{aligned} \sigma_k^2 &\sim \text{Inv-Gamma}(\alpha, \beta), \\ \mu_k | \sigma_k^2 &\sim \mathsf{N}(m, \sigma_k^2 \tau^2), \end{aligned}$$

for some fixed values (hyperparameters) α, β, m, τ^2.

Choice of Hyperparameters

$$\alpha = 2$$
$$\beta = 1$$
$$m = 5$$
$$\tau^2 = 1$$

Choice of Hyperparameters

$$\alpha = 2$$
$$\beta = 1$$
$$m = 5$$
$$\tau^2 = 1$$

We let $\alpha = 2$ and $\beta = 1$ so that the mean of σ_k^2 was 1, and so it was unlikely to generate rewards greater than 10.

Choice of Hyperparameters

$$\alpha = 2$$
$$\beta = 1$$
$$m = 5$$
$$\tau^2 = 1$$

We let $\alpha = 2$ and $\beta = 1$ so that the mean of σ_k^2 was 1, and so it was unlikely to generate rewards greater than 10.

Since the marks awarded are on a scale from 0-10, we let $m = 5$.

Choice of Hyperparameters

$$\alpha = 2$$
$$\beta = 1$$
$$m = 5$$
$$\tau^2 = 1$$

We let $\alpha = 2$ and $\beta = 1$ so that the mean of σ_k^2 was 1, and so it was unlikely to generate rewards greater than 10.

Since the marks awarded are on a scale from 0-10, we let $m = 5$.

We let $\tau^2 = 1$ so that variance of $\mu_k|\sigma_k^2$ was fixed as σ_k^2.

Posterior Calculation

For any given restaurant k, we derive the posterior distribution of μ_k and σ_k^2:

Posterior Calculation

For any given restaurant k, we derive the posterior distribution of μ_k and σ_k^2:

$$
\begin{aligned}
p(\mu_k, \sigma_k^2 | x_k) &\propto f(x_k | \mu_k, \sigma_k^2) p(\mu_k, \sigma_k^2) \\
&= f(x_k | \mu_k, \sigma_k^2) p(\sigma_k^2) p(\mu_k | \sigma_k^2) \\
&\vdots \\
&\propto \left(\sigma_k^2\right)^{\alpha + 3/2 + n_k/2} \\
&\quad \exp\left(-\frac{1}{\sigma_k^2}\left(\beta + \frac{1}{2\tau^2}(\mu_k - m)^2 + \frac{1}{2}\sum_{i=1}^{n_k}(x_{k,i} - \mu_k)^2\right)\right)
\end{aligned}
$$

Posterior Calculation

For any given restaurant k, we derive the posterior distribution of μ_k and σ_k^2:

$$
\begin{aligned}
p(\mu_k, \sigma_k^2 | x_k) &\propto f(x_k | \mu_k, \sigma_k^2) p(\mu_k, \sigma_k^2) \\
&= f(x_k | \mu_k, \sigma_k^2) p(\sigma_k^2) p(\mu_k | \sigma_k^2) \\
&\vdots \\
&\propto \left(\sigma_k^2\right)^{\alpha + 3/2 + n_k/2} \\
&\quad \exp\left(-\frac{1}{\sigma_k^2}\left(\beta + \frac{1}{2\tau^2}(\mu_k - m)^2 + \frac{1}{2}\sum_{i=1}^{n_k}(x_{k,i} - \mu_k)^2\right)\right)
\end{aligned}
$$

This is intractable. :(

Conditional Distributions

Considering the two parameters separately, given that we know the value of the other, we have

$$p(\sigma_k^2|\mu_k, x_k) \sim \text{Inv-Gamma}\left(A, B\right),$$

where

$$A = \alpha + \frac{1}{2} + \frac{n_k}{2},$$

$$B = \beta + \frac{1}{2\tau^2}\left(\mu_k - m\right)^2 + \frac{1}{2}\sum_{i=1}^{n_k}\left(x_{k,i} - \mu_k\right)^2,$$

Conditional Distributions

and

$$p(\mu_k|\sigma_k^2, x_k) \sim \mathrm{N}\left(\frac{m + \tau^2 \sum_{i=1}^{n_k} x_{k,i}}{1 + \tau^2 n_k}, \frac{\tau^2 \sigma_k^2}{1 + \tau^2 n_k}\right).$$

Conditional Distributions

and

$$p(\mu_k|\sigma_k^2, x_k) \sim \mathrm{N}\left(\frac{m + \tau^2 \sum_{i=1}^{n_k} x_{k,i}}{1 + \tau^2 n_k}, \frac{\tau^2 \sigma_k^2}{1 + \tau^2 n_k}\right).$$

This means we can use the Gibbs sampler. For each restaurant, we use the Gibbs sampler to get 1000 pairs (μ_k, σ_k^2). Then for each pair we sample 1000 points from $\mathrm{N}(\mu_k, \sigma_k^2)$, which gives us 10^6 points for each restaurant to analyse using Monte Carlo methods.

Results

For each restaurant k, we are interested in the average quality, the probability of having at least an okay meal, and the probability of having a great meal.

Results

For each restaurant k, we are interested in the average quality, the probability of having at least an okay meal, and the probability of having a great meal.

	Aquila	Bella Italia	Cafe Dolce	Domino's
$\mathbb{E}(r(k)\|\mathcal{D})$	4.692	4.384	5.037	4.773

Results

For each restaurant k, we are interested in the average quality, the probability of having at least an okay meal, and the probability of having a great meal.

	Aquila	Bella Italia	Cafe Dolce	Domino's
$\mathbb{E}(r(k)\|\mathcal{D})$	4.692	4.384	5.037	4.773
$\mathbb{P}(r(k) > 3\|\mathcal{D})$	0.976	0.748	0.981	0.908

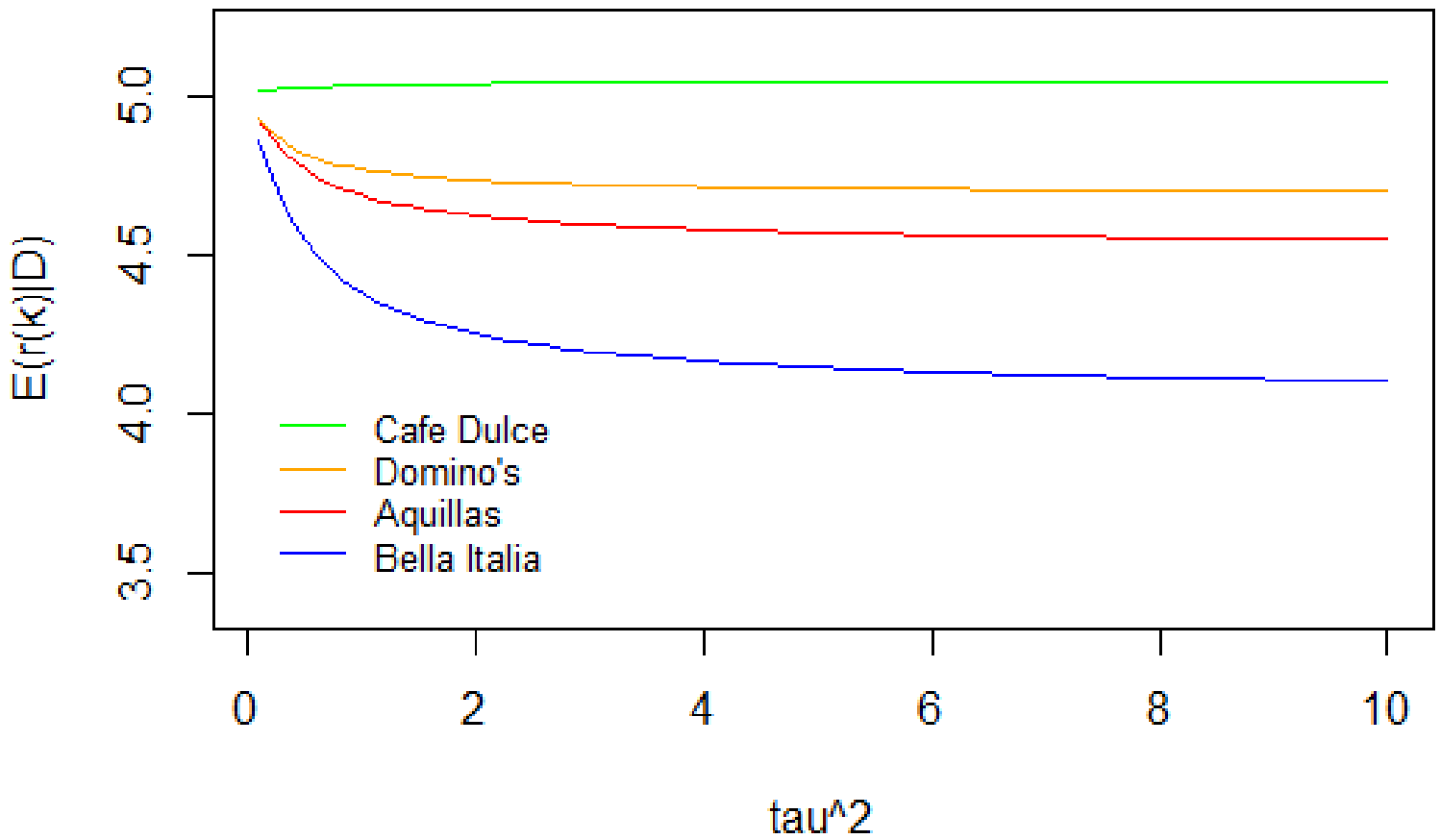
Results

For each restaurant k, we are interested in the average quality, the probability of having at least an okay meal, and the probability of having a great meal.

	Aquila	Bella Italia	Cafe Dolce	Domino's
$\mathbb{E}(r(k)\|\mathcal{D})$	4.692	4.384	5.037	4.773
$\mathbb{P}(r(k) > 3\|\mathcal{D})$	0.976	0.748	0.981	0.908
$\mathbb{P}(r(k) > 7\|\mathcal{D})$	0.011	0.128	0.023	0.061

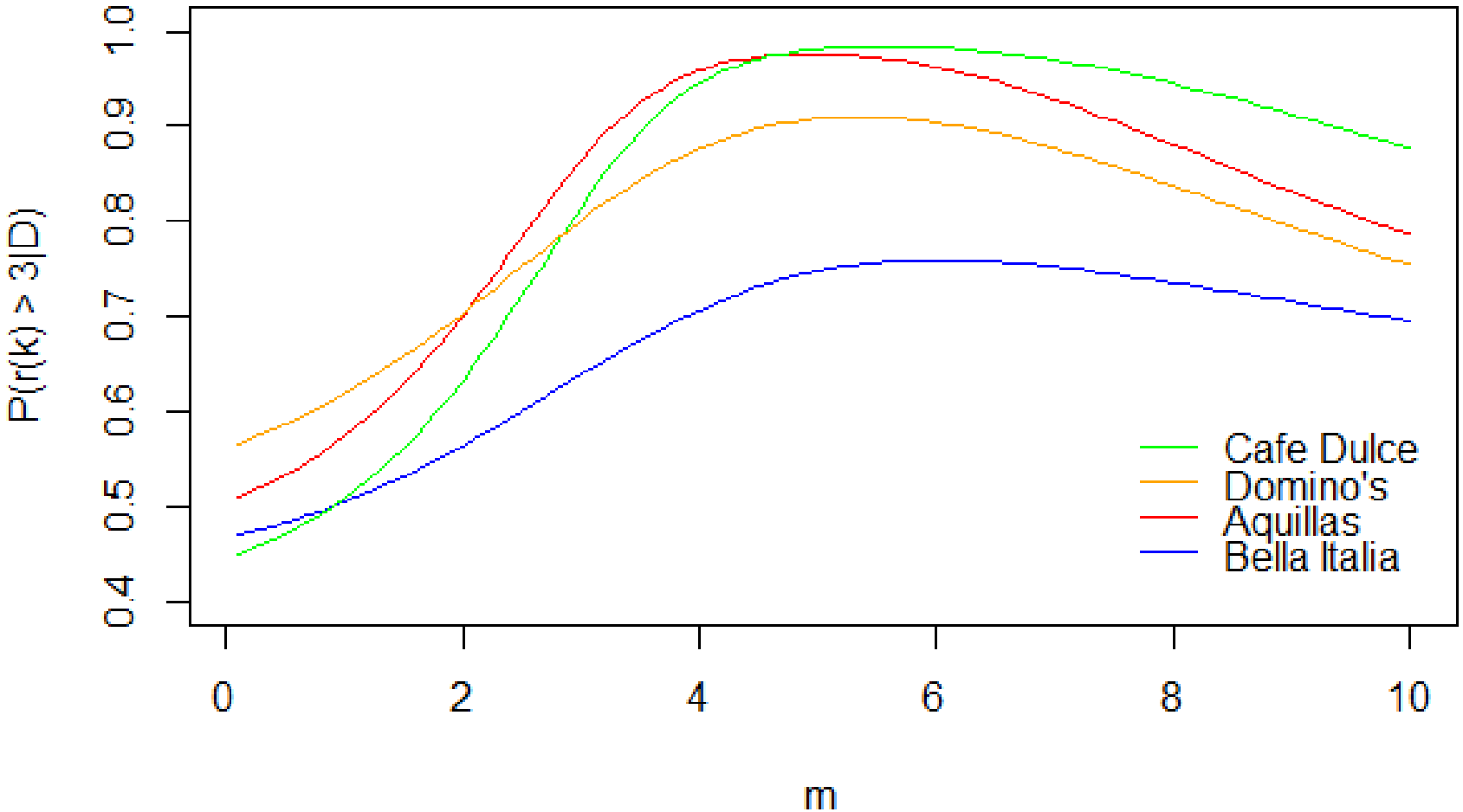
Varying τ^2

We found that varying τ^2 made no difference to our restaurant choice, regardless of our decision method. This plot illustrates this for our first decision method, $\mathbb{E}(r(k)|\mathcal{D})$



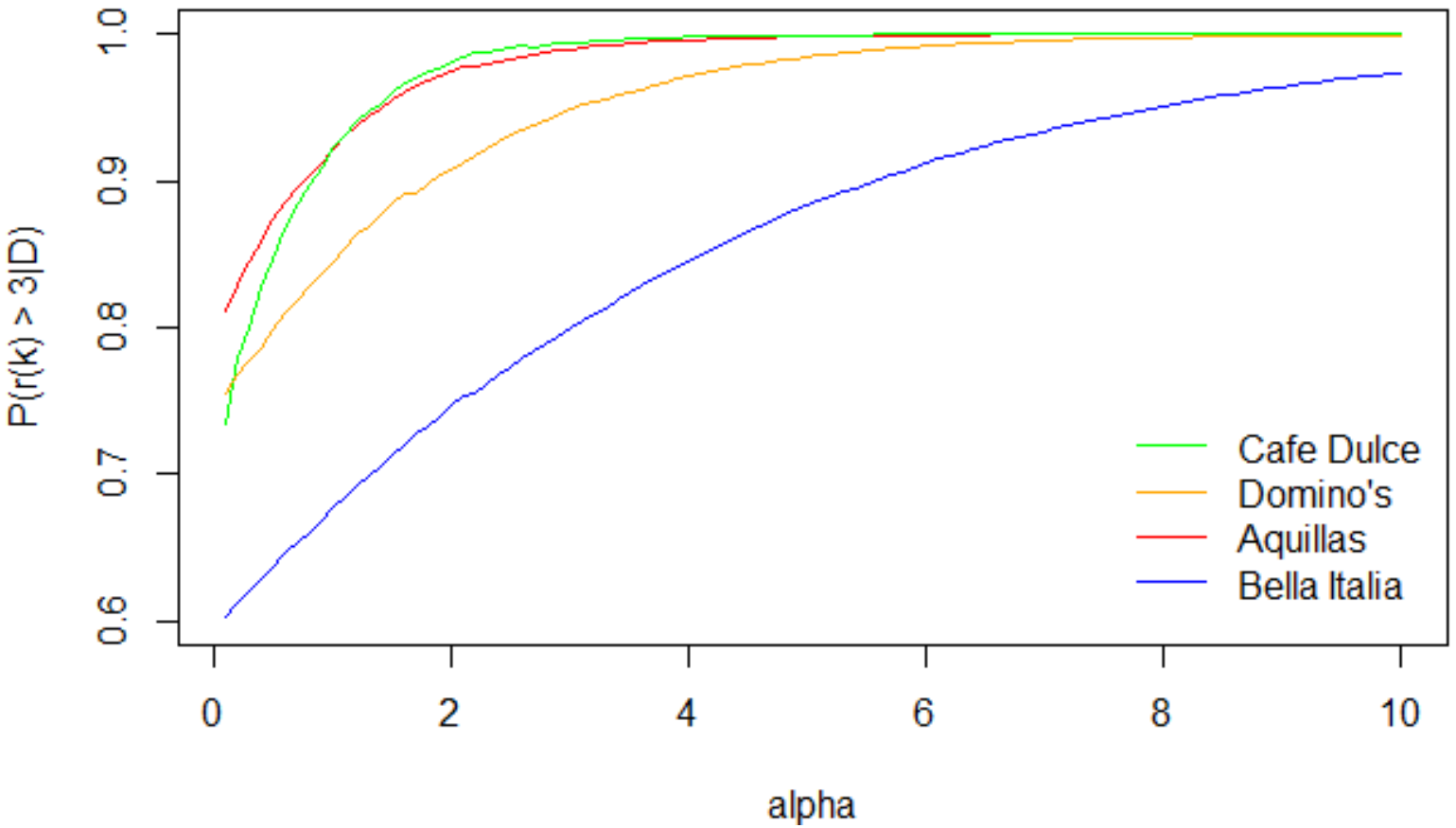
Varying m

We found that varying m did change our restaurant choice, higher values of m favoured Cafe Dulce for all decision methods. This plot illustrates this for our second decision method, $\mathbb{P}(r(k) > 3|\mathcal{D})$



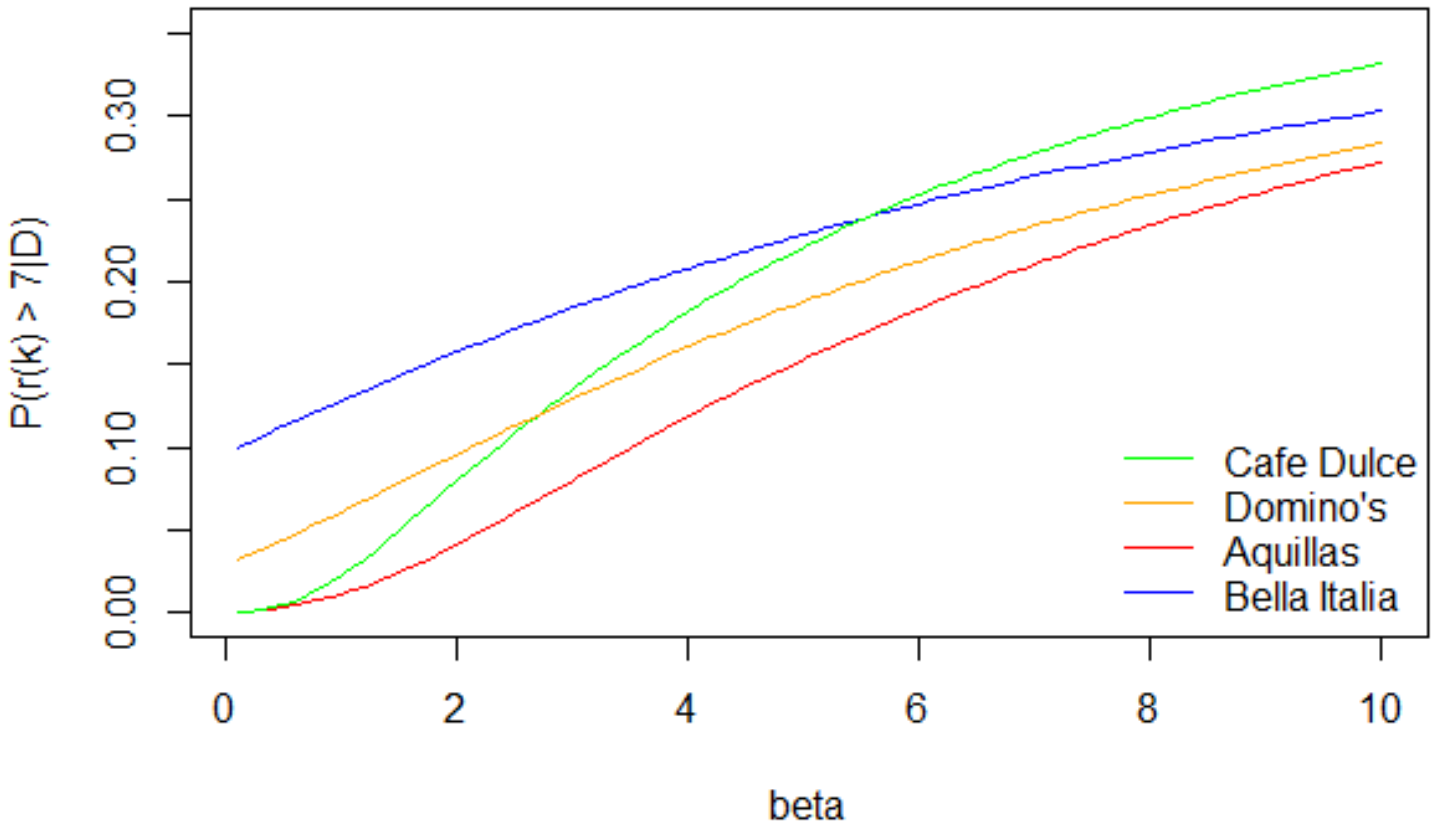
Varying α

We found that varying α did change our restaurant choice at small values of α - changing α particularly changed the behaviour of Cafe Dulce's performance for all decision methods. This plot again illustrates this for our second decision method, $\mathbb{P}(r(k) > 3|\mathcal{D})$.



Varying β

We found that varying β did change our restaurant choice at small values of β - changing β particularly changed the behaviour of Cafe Dulce's performance for our third decision method, $\mathbb{P}(r(k) > 7|\mathcal{D})$, as this plot illustrates.



Risk-Sensitivity for Gains

People are often risk-sensitive for gains – rather than choosing the outcome with the greatest expected reward, they choose an outcome with a smaller expected reward and smaller variance.

Risk-Sensitivity for Gains

People are often risk-sensitive for gains – rather than choosing the outcome with the greatest expected reward, they choose an outcome with a smaller expected reward and smaller variance.

We can quantify this by considering a new utility measure that also considers the variance, instead of just the expectation. We want a smaller variance to increase the desirability of the restaurant, so we consider

$$\frac{\mathbb{E}(r(k)|\mathcal{D})}{\sqrt{\mathrm{Var}(r(k)|\mathcal{D})}}.$$

Risk-Sensitivity in the Restaurant Example

We will now use our new utility measure on the restaurant example (using the original hyperparameters).

	Aquila	Bella Italia	Cafe Dolce	Domino's
$\mathbb{E}(r(k)\|\mathcal{D})$	4.692	4.384	5.037	4.773
$\mathrm{Var}(r(k)\|\mathcal{D})$	0.886	11.990	1.342	3.279

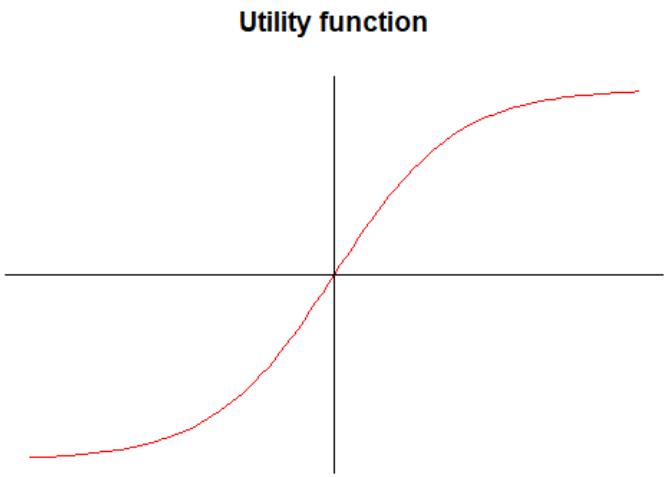
Risk-Sensitivity in the Restaurant Example

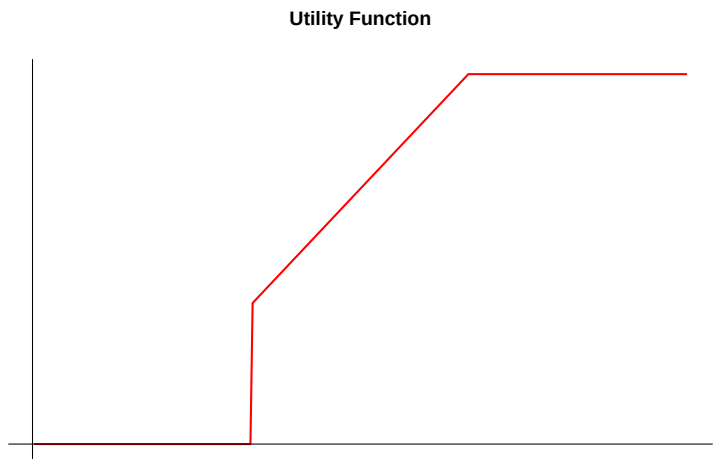
We will now use our new utility measure on the restaurant example (using the original hyperparameters).

	Aquila	Bella Italia	Cafe Dolce	Domino's
$\mathbb{E}(r(k)\|\mathcal{D})$	4.692	4.384	5.037	4.773
$\mathrm{Var}(r(k)\|\mathcal{D})$	0.886	11.990	1.342	3.279
$\frac{\mathbb{E}(r(k)\|\mathcal{D})}{\sqrt{\mathrm{Var}(r(k)\|\mathcal{D})}}$	4.985	1.266	4.348	2.636

Risk-Sensitivity Plot

Often the utility derived is not linear, so a smaller expected reward may sometimes lead to only marginally smaller expected utility.





Restaurant Problem Conclusion

Ultimately we think we should go to **Aquila**, since it has the largest expected value when considering risk-sensitivity (variance) and one of the largest probabilities of having an at least okay quality meal.

Restaurant Problem Conclusion

Ultimately we think we should go to **Aquila**, since it has the largest expected value when considering risk-sensitivity (variance) and one of the largest probabilities of having an at least okay quality meal.

We also think **Cafe Dolce** is a good choice – it has the largest expected value and one of the largest probabilities of having an at least okay quality meal. However, we only have one datapoint and so treat these results with some skepticism (the results change the most based on the hyperparameters).

Restaurant Problem Conclusion

Ultimately we think we should go to **Aquila**, since it has the largest expected value when considering risk-sensitivity (variance) and one of the largest probabilities of having an at least okay quality meal.

We also think **Cafe Dolce** is a good choice – it has the largest expected value and one of the largest probabilities of having an at least okay quality meal. However, we only have one datapoint and so treat these results with some skepticism (the results change the most based on the hyperparameters).

The worst choice is **Bella Italia**. It has the smallest expected value, the smallest probability of having an at least okay quality meal, and by far the smallest expected value when being risk-sensitive (considering variance).

Bandits Problem

Bandits Problem: The Set Up

- Set of K actions/arms.
- There are T rounds, selecting an action a_t each round.
- Choosing action $k \in K$ gives some reward $X_{k,t}$.
- Rewards are i.i.d. across actions with $X_k \sim \nu_k$ where ν_k is unknown.
- GOAL: Identify a rule for sequentially selecting actions, which maximises expected cumulative reward over T rounds

$$\max \sum_{t=1}^{T} \mathbb{E}(X_{a_t,t})$$

UCB1

UCB1 Algorithm

- For $t = 1, \ldots K$ select arm a_t
- For $t = K + 1, \ldots, T$,
  - First calculate

$$\bar{\mu}_{k,t} = \frac{\sum_{s=1}^{t-1} X_{k,s} \mathbb{I}(a_s = k)}{\sum_{s=1}^{t-1} \mathbb{I}(a_s = k)} + \sqrt{\frac{2 \log(t)}{\sum_{s=1}^{t-1} \mathbb{I}(a_s = k)}}$$

  - Select arm with biggest $\bar{\mu}_{k,t}$

Thompson Sampling

Algorithm:

- For $t = 1, \ldots, T$:
  - For each arm k draw a sample,

$$\tilde{\mu}_{k,t} \sim p(\mu_k | X_{k,1:t-1}) \,.$$

  - Select the arm with the largest sample $\tilde{\mu}_{k,t}$ and update it's posterior.

Thompson Sampling

For our 2 armed bandit problem with $\mu_1 = 0.5\,, \mu_2 = 0.55$ we have,

$$f(X_{k,1:t-1}|\mu_k) = \mu_k^{\sum X_{k,1:t-1}}(1-\mu_k)^{t-1-\sum X_{k,1:t-1}}\,,$$
$$p(\mu_k) \propto \mu_k^{\alpha-1}(1-\mu_k)^{\beta-1}\,.$$

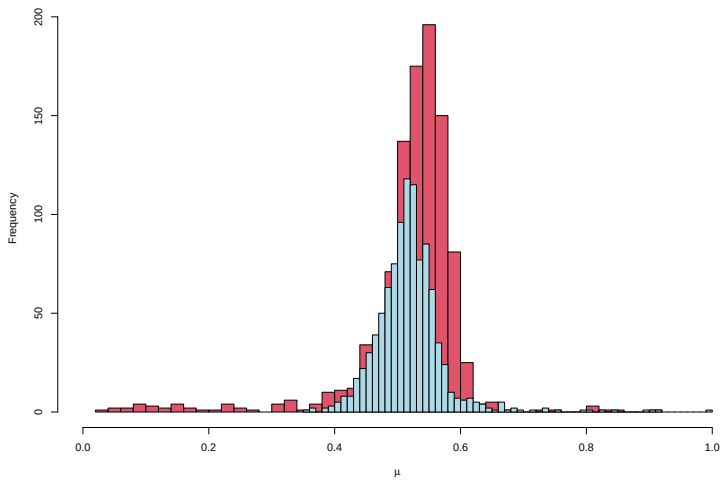
Thus when updating our Beta(α, β) will become:

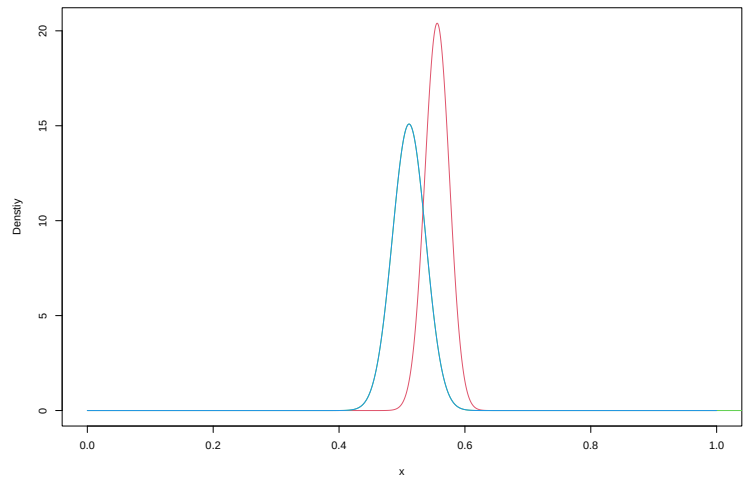
$$A \longleftarrow \alpha + \sum X_{k,1:t-1}\,,$$
$$B \longleftarrow \beta + t - 1 - \sum X_{k,1:t-1}\,,$$

and our posterior distribution for $\mu_k|X_{k,1:t-1}$ will be Beta(A, B)

Thompson Sampling

Starting with a prior Beta$(1, 1)$ our final distributions are shown:





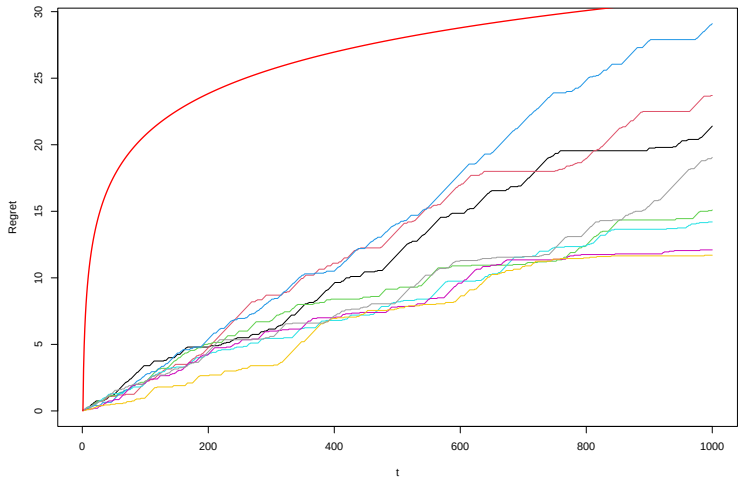
Regret

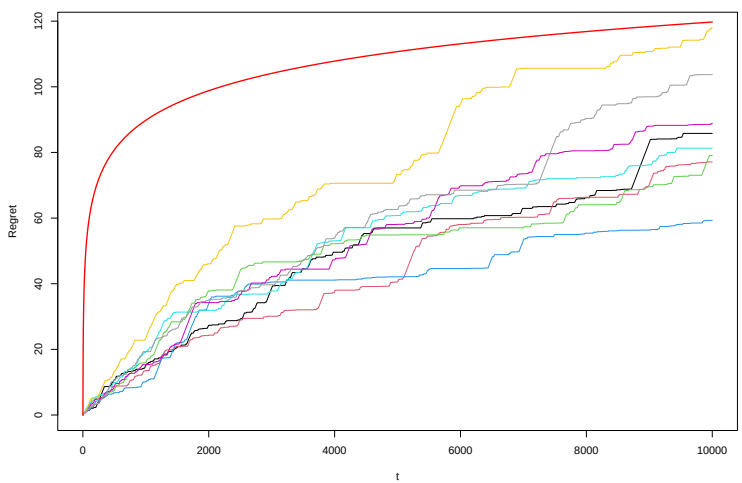
Define $\mu_k = \mathbb{E}(X_{k,t})$ for $k \in \{1, \ldots, K\}$, and $\mu^* = \max_{k \in \{1,\ldots,K\}} \mu_k$, our regret is then,

$$\text{Reg}_\pi(T) = \sum_{t=1}^{T} \mu^* - \mathbb{E}_\pi(\mu_{a_t}) \, .$$

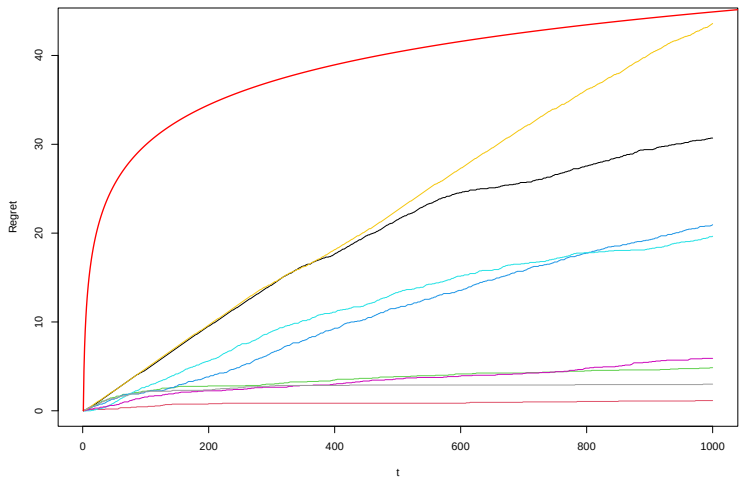
Regret

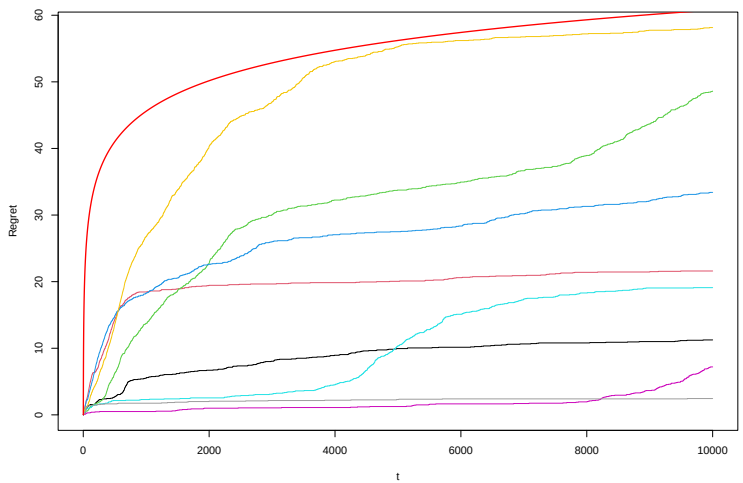
UCB1



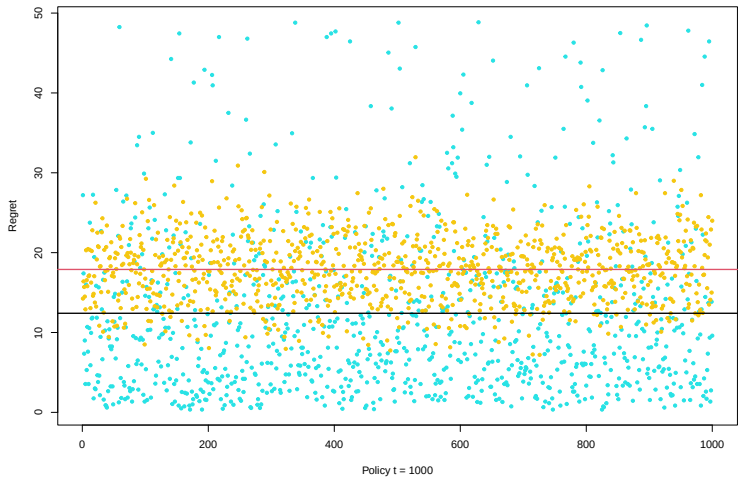


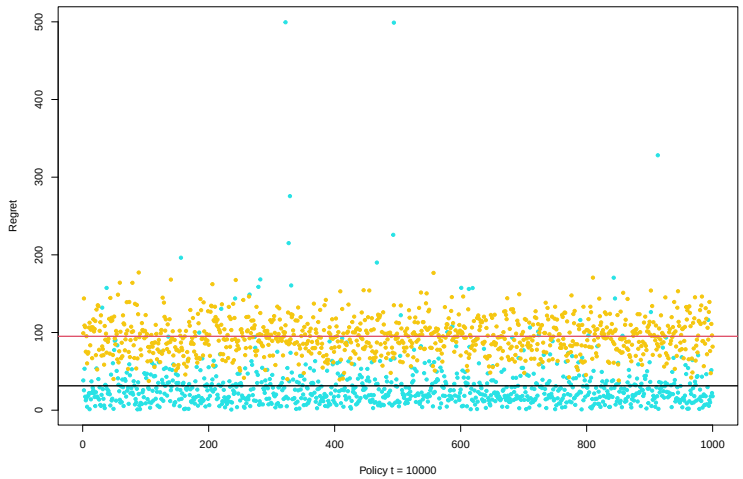
Thompson Sampling





Regret

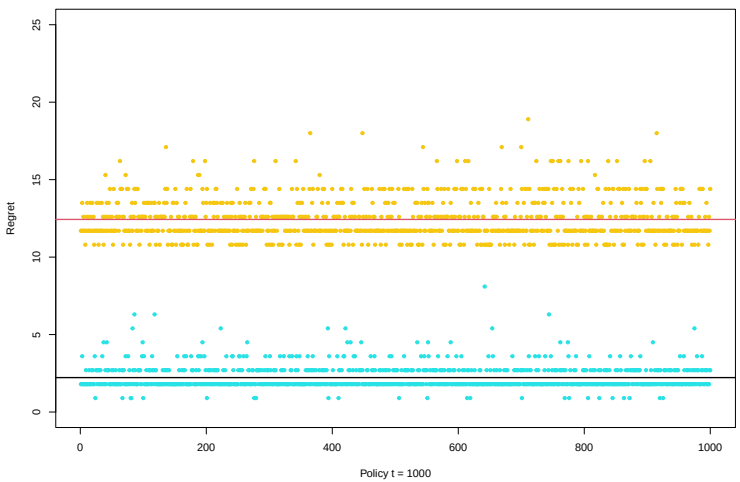


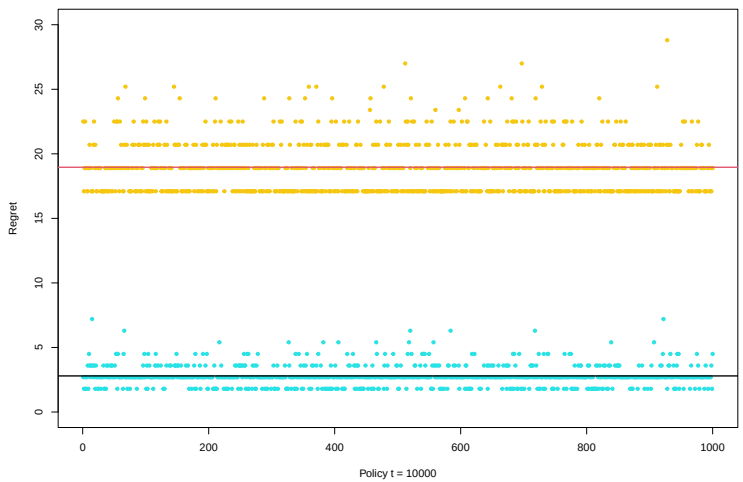


Yellow: UCB1 Blue: Thompson

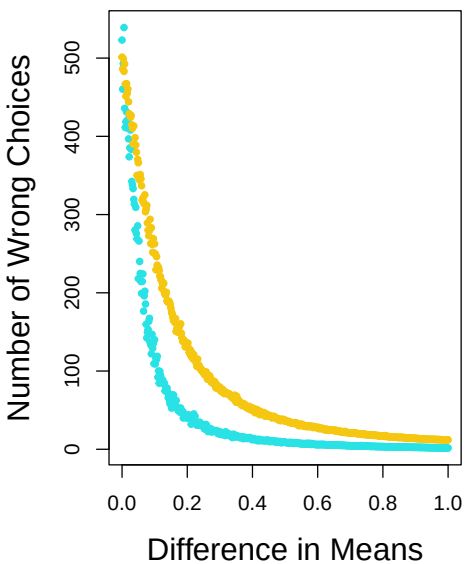
Regret for New Means

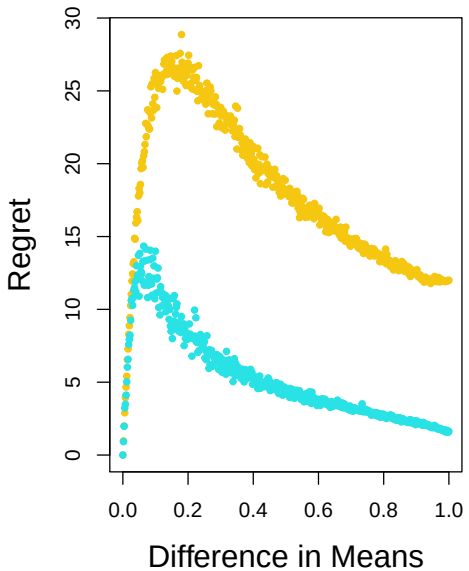
For new $(\mu_1, \mu_2) = (0.05, 0.95)$,

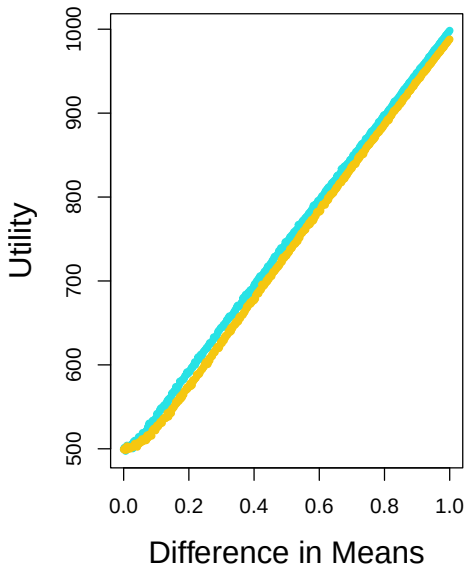




Regret for New Means







Thompson sampling: varying the hyperparameters

Our hyperparameters α and β inform our early estimates of the means of our 2 armed bandit. As time goes on, are estimates are influenced by the data we have collected from each arm.

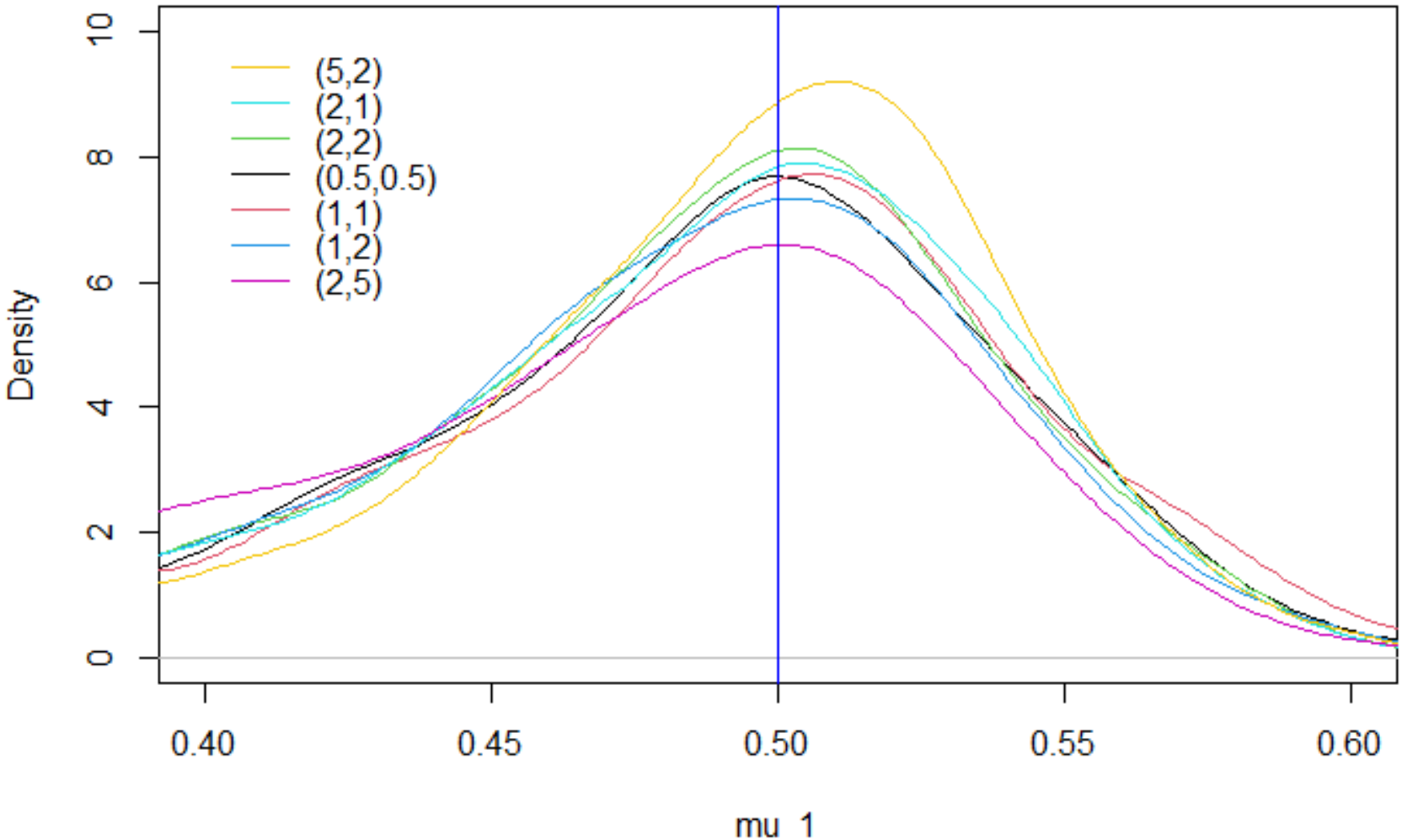
We have changed these hyperparameters and explored how our Thompson sampling performs as a result.

Our choices for (α, β) are:

$(5, 2)$, $(2, 1)$, $(2, 2)$, $(0.5, 0.5)$, $(1, 1)$, $(1, 2)$, $(2, 5)$.

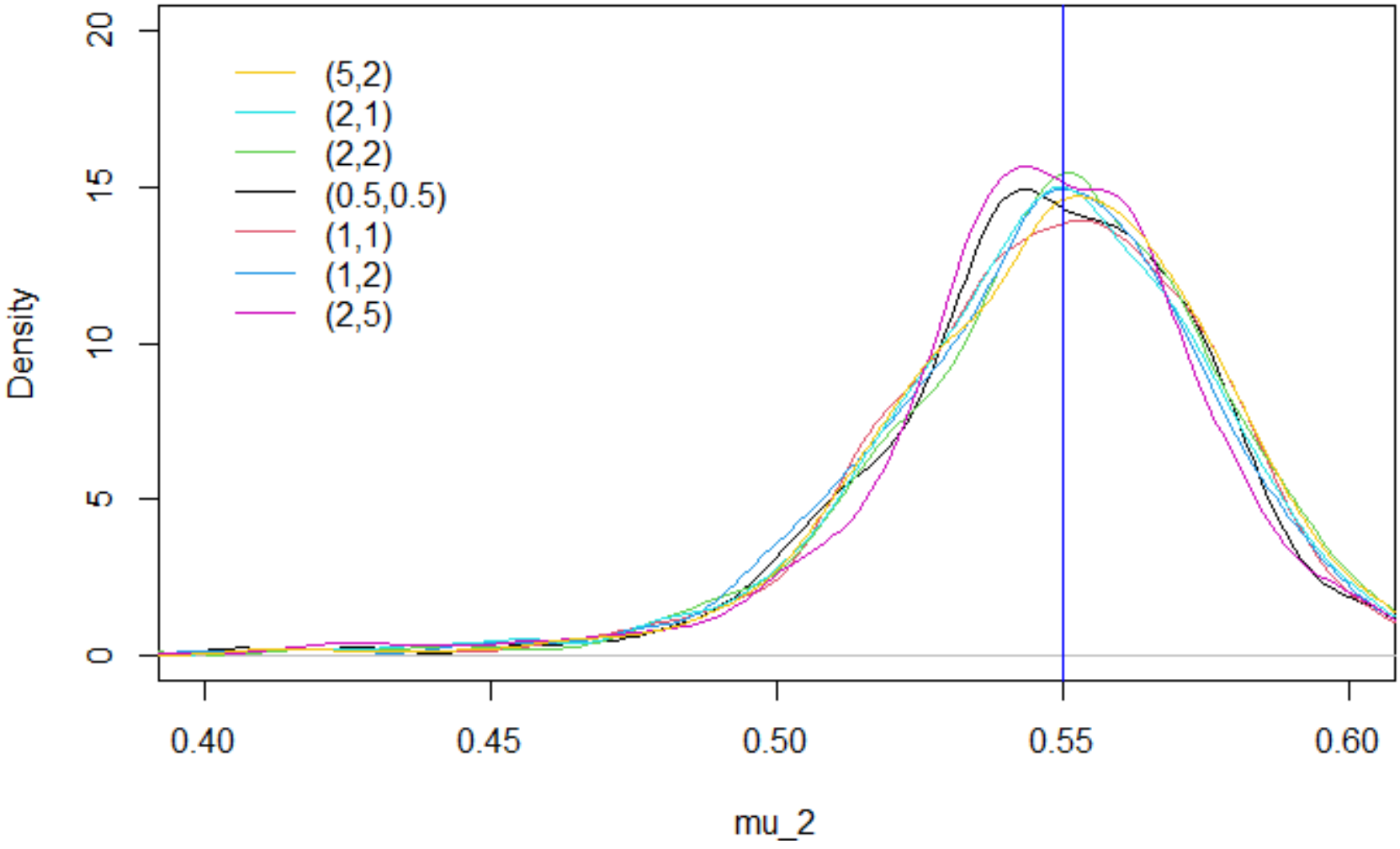
Thompson sampling: varying the hyperparameters

For each hyperparameter pair, we ran our algorithm 1000 times for T=1000, below we illustrate the spread of our final estimates of μ_1



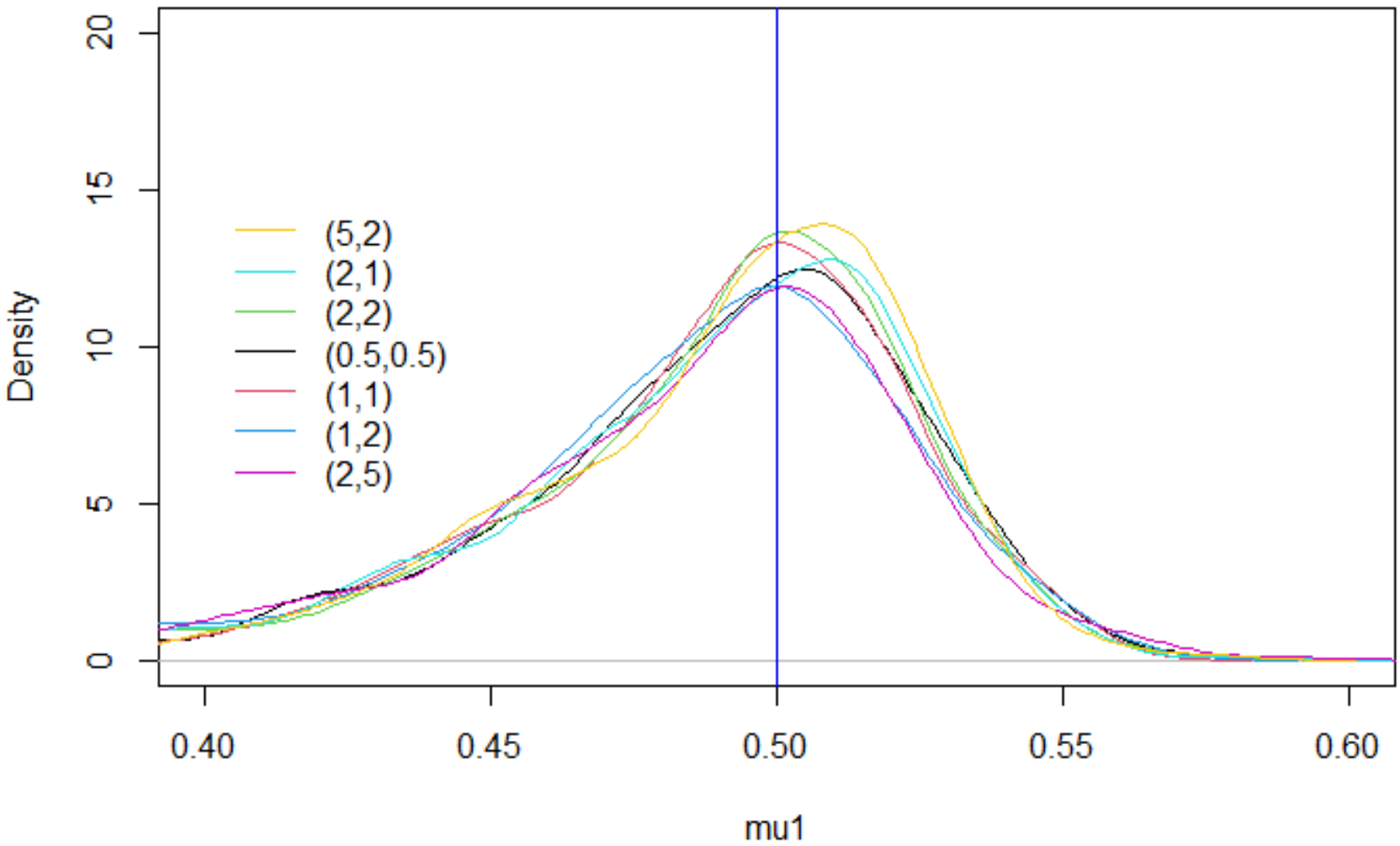
Thompson sampling: varying the hyperparameters

Below we illustrate the spread of our final estimates of μ_2



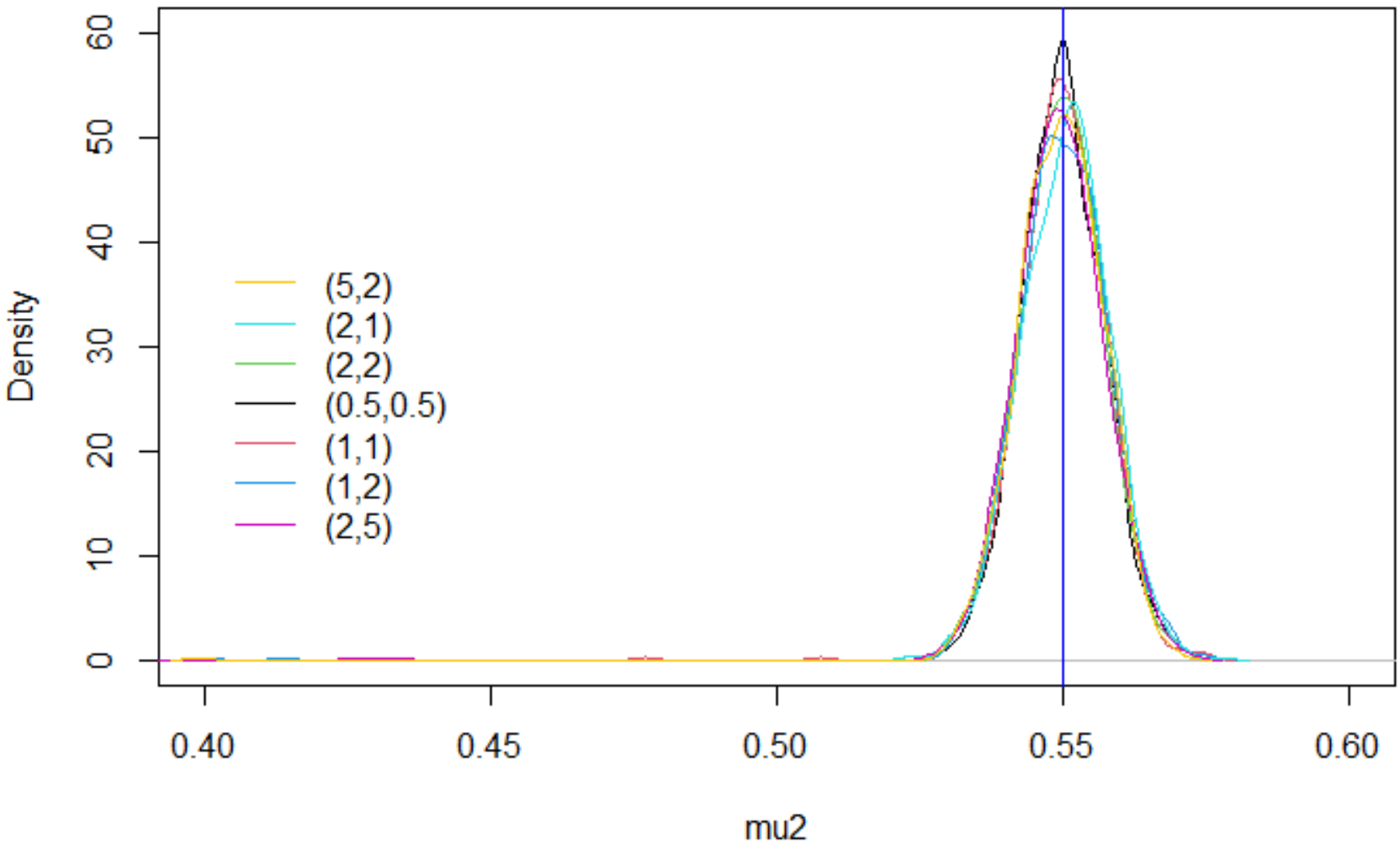
Thompson sampling: varying the hyperparameters

Below we illustrate the spread of our final estimates of μ_1, here letting $T = 10,000$



Thompson sampling: varying the hyperparameters

Below we illustrate the spread of our final estimates of μ_2, here letting $T = 10,000$



The Set Up

Let $f : \mathcal{X} \to \mathbb{R}$ be a black box function. We want to find

$$x^* = \underset{x \in \mathcal{X}}{\operatorname{argmin}} f(x).$$

The Set Up

Let $f : \mathcal{X} \to \mathbb{R}$ be a black box function. We want to find

$$x^* = \underset{x \in \mathcal{X}}{\operatorname{argmin}} f(x).$$

We sample from $f(x)$ n times and so have $\mathcal{D}_n = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$ where $y_i = f(x_i)$.

Our current best guess of f is $f_n^* = \min_{i=1,\ldots,n} y_i$.

The Set Up

Conditional on the observations, a Gaussian model predicts the value of f for any $x \in \mathcal{X}$ as $f(x) \mid \mathcal{D}_n \sim N(\mu(x), \sigma(x))$.

The Set Up

Conditional on the observations, a Gaussian model predicts the value of f for any $x \in \mathcal{X}$ as $f(x) \mid \mathcal{D}_n \sim N(\mu(x), \sigma(x))$.

To choose where next to evaluate f, we maximize an acquisition function $\alpha_n : \mathcal{X} \to \mathbb{R}$ that estimates the benefit provided by an evaluation with respect to solving our minimisation.

The Set Up

Conditional on the observations, a Gaussian model predicts the value of f for any $x \in \mathcal{X}$ as $f(x) \mid \mathcal{D}_n \sim N(\mu(x), \sigma(x))$.

To choose where next to evaluate f, we maximize an acquisition function $\alpha_n : \mathcal{X} \to \mathbb{R}$ that estimates the benefit provided by an evaluation with respect to solving our minimisation.

Now consider an acquisition function of the form $\alpha_n(x) = \mathbb{E}(u(x) \mid \mathcal{D}_n)$ where

$$u(x) = \max(0, f_n^* - f(x)).$$

Calculating $\alpha_n(x)$

$$\alpha_n(x) = \mathbb{E}\left(u(x)|\mathcal{D}_n\right)$$

Calculating $\alpha_n(x)$

$$\begin{aligned}\alpha_n(x) &= \mathbb{E}\left(u(x)|\mathcal{D}_n\right) \\ &= \mathbb{E}\left(\max(0, f_n^* - f(x))|\mathcal{D}_n\right)\end{aligned}$$

Calculating $\alpha_n(x)$

$$\begin{aligned}\alpha_n(x) &= \mathbb{E}\left(u(x)|\mathcal{D}_n\right) \\ &= \mathbb{E}\left(\max(0, f_n^* - f(x))|\mathcal{D}_n\right) \\ &= \int_{-\infty}^{\infty} \max(0, f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f\end{aligned}$$

Calculating $\alpha_n(x)$

$$
\begin{aligned}
\alpha_n(x) &= \mathbb{E}\left(u(x)|\mathcal{D}_n\right) \\
&= \mathbb{E}\left(\max(0, f_n^* - f(x))|\mathcal{D}_n\right) \\
&= \int_{-\infty}^{\infty} \max(0, f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f \\
&= \int_{-\infty}^{f_n^*} (f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f
\end{aligned}
$$

Calculating $\alpha_n(x)$

$$
\begin{aligned}
\alpha_n(x) &= \mathbb{E}\left(u(x)|\mathcal{D}_n\right) \\
&= \mathbb{E}\left(\max(0, f_n^* - f(x))|\mathcal{D}_n\right) \\
&= \int_{-\infty}^{\infty} \max(0, f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f \\
&= \int_{-\infty}^{f_n^*} (f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f \\
&= f_n^* \int_{-\infty}^{f_n^*} \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f - \int_{-\infty}^{f_n^*} f \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f
\end{aligned}
$$

Calculating $\alpha_n(x)$

$$
\begin{aligned}
\alpha_n(x) &= \mathbb{E}\left(u(x)|\mathcal{D}_n\right) \\
&= \mathbb{E}\left(\max(0, f_n^* - f(x))|\mathcal{D}_n\right) \\
&= \int_{-\infty}^{\infty} \max(0, f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f \\
&= \int_{-\infty}^{f_n^*} (f_n^* - f) \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f \\
&= f_n^* \int_{-\infty}^{f_n^*} \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f - \int_{-\infty}^{f_n^*} f \cdot \mathsf{N}(\mu_n(x), \sigma_n^2(x))\,\mathrm{d}f \\
&= f_n^* \cdot \mathbb{P}(f(x) \leqslant f_n^*) - \mathbb{E}(f(x)|f(x) \leqslant f_n^*)
\end{aligned}
$$

Calculating $\alpha_n(x)$

From the inverse Mills ratio (see Greene (2012)) we have

$$\mathbb{E}(f \mid f \leqslant f_n^*) = \mu_n(x) - \sigma_n(x)\frac{\phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right)}{1 - \Phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right)}.$$

Calculating $\alpha_n(x)$

From the inverse Mills ratio (see Greene (2012)) we have

$$\mathbb{E}(f \mid f \leqslant f_n^*) = \mu_n(x) - \sigma_n(x)\frac{\phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right)}{1 - \Phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right)}.$$

Therefore the closed form for $\alpha_n(x)$ is

$$f_n^* \cdot \Phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right) - \mu_n(x) + \sigma_n(x)\frac{\phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right)}{1 - \Phi\left(\frac{f_n^* - \mu_n(x)}{\sigma_n(x)}\right)}.$$

Interpreting $\alpha_n(x)$

$$\alpha_n(x) = f_n^* \cdot \mathbb{P}(f(x) \leqslant f_n^*) - \mathbb{E}(f(x) \mid f(x) \leqslant f_n^*)$$

Interpreting $\alpha_n(x)$

$$\alpha_n(x) = f_n^* \cdot \mathbb{P}(f(x) \leqslant f_n^*) - \mathbb{E}(f(x) \mid f(x) \leqslant f_n^*)$$

$$\bar{\mu}_{k,t} = \frac{\sum_{s=1}^{t-1} X_{k,s} \mathbb{I}(a_s = k)}{\sum_{s=1}^{t-1} \mathbb{I}(a_s = k)} + \sqrt{\frac{2\log(t)}{\sum_{s=1}^{t-1} \mathbb{I}(a_s = k)}}$$

References

Agarwal, D., Basu, K., Ghosh, S., Xuan, Y., Yang, Y., and Zhang, L. (2018). Online parameter selection for web-based ranking problems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 23–32.

Greene, W. H. (2012). *Econometric analysis*. Pearson, Harlow, 7th ed., international ed. edition.

Thank you for listening

Any questions?