A. Proofs

In this appendix, we provide all proofs for Theorems and Corollaries stated in the paper. We emphasize that we are aware of existing theoretical tools provided in (van der Vaart et al., 2008) and (Knapik et al., 2011), but our approach is different and specific to the current setup.

A.1. Proofs of point-wise Bayesian update in dynamic case

Theorem 5. *Suppose $\tilde{\pi}(x) \sim \sum_{i=0}^{n} C_i^n \mathcal{B}(\alpha+i, \beta+n-i)$ with $\sum_{i=0}^{n} C_i^n = 1$, and we observe the result s of a sample from a Bernoulli random variable with parameter $A\pi(x)+B$. Then the Bayesian posterior for $\tilde{\pi}(x)$ conditioned on this observation is:*

$$\tilde{\pi}(x|s) \sim \sum_{i=0}^{n+1} C_i^{n+1} \mathcal{B}(\theta, \alpha+i, \beta+n-i) \tag{10}$$

where $\forall i = 0, ..., n+1$:

$$C_i^{n+1} = \frac{1}{E_s^n}(BC_i^n(\beta+n-i) + (A+B)C_{i-1}^n(\alpha+i-1))$$

if $s=1$ and

$$C_i^{n+1} = \frac{1}{E_f^n}((1-B)C_i^n(\beta+n-i) + (1-A-B)C_{i-1}^n(\alpha+i-1))$$

if $s=0$. E_s^n and E_f^n are normalization factors that ensure $\sum_{i=0}^{n} C_i^{n+1} = 1$. For simplicity of notation $C_{-1}^n = C_{n+1}^n = 0$ $\forall n$.

Proof. Suppose the observation is a success, i.e. $s=1$. Let $f_{\tilde{\pi}(x)} : [0,1] \to [0,1]$ be the density function of the random variable $\tilde{\pi}(x)$, and let $f_{\tilde{\pi}(x)|s=1} : [0,1] \to [0,1]$ be its the density function conditioned on this observation. Then,

$$\begin{aligned}
f_{\tilde{\pi}(x)|s=1}(\theta) &= \frac{Pr(s=1|\tilde{\pi}(x)=\theta) f_{\tilde{\pi}(x)}(\theta)}{Pr(s=1)} \\
&\propto (A\theta+B)\sum_{i=0}^{n} C_i^n \mathcal{B}(\alpha+i, \beta+n-i) \\
&= (B(1-\theta)+(A+B)\theta)\sum_{i=0}^{n} C_i^n \frac{\theta^{\alpha+i-1}(1-\theta)^{\beta+n-i-1}}{\mathrm{B}(\alpha+i, \beta+n-i)} \\
&= B\sum_{i=0}^{n} C_i^n \frac{\theta^{\alpha+i-1}(1-\theta)^{\beta+n-i}}{\mathrm{B}(\alpha+i, \beta+n-i+1)} \frac{\mathrm{B}(\alpha+i, \beta+n-i+1)}{\mathrm{B}(\alpha+i, \beta+n-i)} \\
&+ (A+B)\sum_{i=0}^{n} C_i^n \frac{\theta^{\alpha+i}(1-\theta)^{\beta+n-i-1}}{\mathrm{B}(\alpha+i+1, \beta+n-i)} \frac{\mathrm{B}(\alpha+i+1, \beta+n-i)}{\mathrm{B}(\alpha+i, \beta+n-i)} \\
&= B\sum_{i=0}^{n} C_i^n \mathcal{B}(\alpha+i, \beta+n-i+1)\frac{\beta+n-i}{\alpha+\beta+n} \\
&+ (A+B)\sum_{i=0}^{n} C_i^n \mathcal{B}(\alpha+i+1, \beta+n-i)\frac{\alpha+i}{\alpha+\beta+n} \\
&\propto \sum_{i=0}^{n+1} (BC_i^n(\beta+n-i) + (A+B)C_{i-1}^n(\alpha+i-1))\mathcal{B}(\alpha+i, \beta+n-i) \\
&\propto \sum_{i=0}^{n+1} C_i^{n+1} \mathcal{B}(\theta, \alpha+i, \beta+n-i)
\end{aligned}$$

where B is the Beta function, and satisfies $\frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)} = \frac{\alpha}{\alpha+\beta}$ and $\frac{\mathrm{B}(\alpha,\beta+1)}{\mathrm{B}(\alpha,\beta)} = \frac{\beta}{\alpha+\beta}$.

In order to ensure that this remains a probability distribution, coefficients C_i^{n+1} must satisfy $\sum_{i=0}^{n+1} C_i^{n+1} = 1$. The result for $s = 0$ can be showed similarly. □

Theorem 2 is a special case of this result, for $n = 0$. Corollary 1 directly follows from this theorem, by applying it recursively for each observations.

Corollary 2. *Suppose $\tilde{\pi}(x) \sim \mathcal{B}(\alpha, \beta)$ and we observe the outputs of experiments $\mathcal{S} = \{(s_i, x, 1-B, B)\}_{i=1,...,t}$ where $s_i \sim$ Bernoulli$((1-B)\pi(x) + B)$. Then the Bayesian posterior $\tilde{\pi}(x|\mathcal{S})$ conditioned on these observations is given by*

$$\tilde{\pi}(x|\mathcal{S}) \sim \sum_{i=0}^{S} C_i^t \mathcal{B}(\alpha + i, \beta + t - i) \tag{11}$$

where $S = \sum_{i=1}^{t} s_i$ is the total number of successes and

$$C_i^t \propto \binom{S}{i} (\alpha - 1 + i)!(\beta + t - 1 - i)! B^{S-i} \tag{12}$$

$\forall i = 0, ..., S$. Using the relation $C_{i+1}^t = \frac{(S-i)(\alpha+i)}{B(i+1)(\beta+t-1-i)} C_i^t$, we can compute all C_i^t's in time $\mathcal{O}(t)$.

Proof. We want to prove that the iterative process for computing the coefficients C_i^t's in Corollary 1 ends with coefficients C_i^t's of equation (12). We prove this by induction over t. For $t = 0$, the result is obvious, since $S = 0$, and $C_0^0 = 1$.

Now suppose the result is true for some time n and let us prove that it remains true for time $n+1$. Let S_n be the total number of successes observed up to time n, and let s_{n+1} be the new observation at time $n+1$. Suppose $s_{n+1} = 1$. Then $S_{n+1} = S_n + 1$, and $\forall i = 1, ..., S_{n+1}$:

$$\begin{aligned} C_i^{n+1} &\propto BC_i^n(\beta + n - i) + C_{i-1}^n(\alpha + i - 1) \\ &\propto \binom{S_n}{i} (\alpha - 1 + i)!(\beta + n - 1 - i)! B^{S_n+1-i}(\beta + n - i) \\ &+ \binom{S_n}{i-1} (\alpha - 1 + i - 1)!(\beta + n - i)! B^{S_n+1-i}(\alpha + i - 1) \\ &= \binom{S_{n+1}}{i} (\alpha - 1 + i)!(\beta + (n+1) - 1 - i)! B^{S_{n+1}-i} \end{aligned}$$

Similarly, if $s_{n+1} = 0$, then $S_{n+1} = S_n$, and $\forall i = 1, ..., S_{n+1}$:

$$\begin{aligned} C_i^{n+1} &\propto (1-B)C_i^n(\beta + n - i) \\ &\propto \binom{S_{n+1}}{i} (\alpha - 1 + i)!(\beta + (n+1) - 1 - i)! B^{S_{n+1}-i} \end{aligned}$$

In particular, we can see that the number of coefficients increases only when we observe a success.

□

A.2. Proof of convergence in the static case

Theorem 1. *Let $\pi : [0,1]^d \to [0,1]$ be L-Lipschitz continuous. Suppose we measure the results of experiments $\mathcal{S} = \{(x_i, s_i)\}_{i=1,...,t}$ where s_i is a sample from a Bernoulli distribution with parameter $\pi(x_i)$. Experiment points $\{x_i\}_{i=1,...,t}$ are assumed to be i.i.d. and uniformly distributed over the space. Then, starting with a uniform prior $\alpha(x) = \beta(x) = 1\ \forall x \in [0,1]^d$, the posterior $\tilde{\pi}(x|\mathcal{S})$ obtained from Algorithm 1 uniformly converges in L_2-norm to $\pi(x)$, i.e.*

$$\sup_{x\in[0,1]^d} \mathbb{E}_{\mathcal{S}} \left(\mathbb{E}\left((\tilde{\pi}(x|\mathcal{S}) - \pi(x))^2 \right) \right) = \mathcal{O}\left(t^{-\frac{2}{d+2}} \right), \tag{13}$$

where the outer expectation is performed over experiment points $\{x_i\}_{i=1,...,t}$ *and their results* $\{s_i\}_{i=1,...,t}$*. Moreover, Algorithm 1 computes the posterior in time* $\mathcal{O}(t)$*.*

Proof. For simplicity, suppose we start with a uniform prior for each x, i.e. $\tilde{\pi}(x) \sim \mathcal{B}(1,1)$. Let $x \in \mathcal{X}$, $\Delta \in [0,1]$ be arbitrary. Suppose we fix the experiment points $X = \{x_i\}_{i=1,...,t}$ and that among these t points, n of them are at most Δ far from x along all of d dimensions. We assume without loss of generality that these points are $x_1, ..., x_n$. Let D_x be the random variable denoting the number of experiments occurring at most Δ far from x along each dimension. Since we assume that experiment points $\{x_i\}_{i=1,...,t}$ are uniformly distributed over $[0,1]^d$, it follows that $D_x \sim \text{Bin}(t, \Delta^d)$.

Let S_x denote the number of successes that occurred among these n experiments. S_x can be written as a $S_x = \sum_{i=1}^n s_i$ where $\mathbf{s} = \{s_i\}_{i=1,...,n}$ are sampled independently, and $s_i \sim \text{Bernoulli}(\pi(x_i))$ denotes whether experiment on x_i was successful or not. Thus, S_x follows a Poisson-Binomial distribution, and it follows:

$$\mathbb{E}(S_x | D_x = n) = \sum_{i=1}^{n} \pi(x_i) \tag{14}$$

and

$$\mathbb{E}(S_x^2 | D_x = n) = \sum_{i=1}^{n} \pi(x_i)(1-\pi(x_i)) + \left(\sum_{i=1}^{n} \pi(x_i)\right)^2 \tag{15}$$

Note that after s successes among n experiments, the update rule 3 leads to the posterior:

$$\tilde{\pi}(x|\mathcal{S}) \sim \mathcal{B}(1+s, 1+n-s). \tag{16}$$

Using the properties of the Beta distribution, we have:

$$\mathbb{E}(\tilde{\pi}(x|\mathcal{S}) | S_x = s, D_x = n) = \frac{s+1}{n+2} \tag{17}$$

and

$$\begin{aligned}\mathbb{E}(\tilde{\pi}(x|\mathcal{S})^2 | S_x = s, D_x = n) &= \frac{(s+1)(n+1-s)}{(n+2)^2(n+3)} + \frac{(s+1)^2}{(n+2)^2} \\ &= \frac{(s+1)(s+2)}{(n+2)(n+3)} \\ &= \frac{s^2}{(n+2)^2} + \mathcal{O}\left(\frac{1}{n+1}\right)\end{aligned}$$

Therefore:

$$\begin{aligned}\mathbb{E}_{X,\mathbf{s}}\left(\mathbb{E}\left((\tilde{\pi}(x|\mathcal{S}) - \pi(x))^2\right)\right) &= \sum_{n=0}^{t} Pr(D_x = n)\mathbb{E}_{x_1,...,x_n}\left[\sum_{s=0}^{n} Pr(S_x = s|D_x = n)\left(\mathbb{E}(\tilde{\pi}(x|\mathcal{S})^2 | S_x = s, D_x = n)\right.\right. \\ &\quad \left.\left. -2\pi(x)\mathbb{E}(\tilde{\pi}(x|\mathcal{S})|S_x = s, D_x = n) + \pi(x)^2\right)\right] \\ &= \sum_{n=0}^{t} Pr(D_x = n)\mathbb{E}_{x_1,...,x_n}\left[\sum_{s=0}^{n} Pr(S_x = s|D_x = n)\left(\frac{s^2}{(n+2)^2} + \mathcal{O}\left(\frac{1}{n+1}\right) - 2\pi(x)\frac{s}{n+2} + \pi(x)^2\right)\right] \\ &= \sum_{n=0}^{t} Pr(D_x = n)\mathbb{E}_{x_1,...,x_n}\left[\frac{1}{(n+2)^2}\left(\sum_{i=0}^{n} \pi(x_i)(1-\pi(x_i)) + \left(\sum_{i=0}^{n}\pi(x_i)\right)^2\right)\right. \\ &\quad \left. -\frac{2}{n+2}\pi(x)\sum_{i=0}^{n}\pi(x_i) + \pi(x)^2 + \mathcal{O}\left(\frac{1}{n+1}\right)\right| \|x - x_i\| \leq \Delta \; \forall i = 1, ..., n\Bigg] \\ &= \sum_{n=0}^{t} Pr(D_x = n)\mathbb{E}_{x_1,...,x_n}\left[\frac{1}{(n+2)^2}\sum_{i=0}^{n}\pi(x_i)(1-\pi(x_i))\right.\end{aligned}$$

$$+\frac{1}{(n+2)^2}\left(\sum_{i,j=0}^{n}(\pi(x)-\pi(x_i))(\pi(x)-\pi(x_j))\right)+\mathcal{O}\left(\frac{1}{n+1}\right)\Bigg|\, \|x-x_i\|\leq\Delta\ \forall i=1,...,n\Bigg]$$
$$\leq\sum_{n=0}^{t}Pr(D_x=n)\left(\frac{1}{4(n+2)}+\mathcal{O}\left(\frac{1}{n+1}\right)\right)+L^2\Delta^2$$
$$=L^2\Delta^2+\mathcal{O}\left(\frac{1}{\Delta^d(t+1)}\right)$$

Therefore, assuming $L>0$, we can choose $\Delta=\frac{1}{L^{\frac{2}{d+2}}}t^{-\frac{1}{d+2}}$, and we obtain:

$$\mathbb{E}_{X,\mathbf{s}}\left(\mathbb{E}((\tilde{\pi}(x)-\pi(x))^2\right)=\mathcal{O}\left(L^{\frac{2d}{d+2}}t^{-\frac{2}{d+2}}\right) \tag{18}$$

In particular, we observe that the smaller L, the larger Δ. Indeed, the smoother the function, the more we can share experience between points $\{x_i\}$. □

A.3. Proof of convergence in the simplified dynamic case

Theorem 3. *Let $\pi:[0,1]^d\to]0,1]$ be L-Lipschitz continuous. Suppose we observe the results of experiments $\mathcal{S}=\{(x_i,s_i,1-B,B)\}_{i=1,...,t}$ where $s_i\sim$ Bernoulli$((1-B_i)\pi(x)+B_i)$. Experiment points $\{x_i\}_{i=1,...,t}$ are assumed to be uniformly distributed over the space. Then, $\forall x\in\mathcal{X}$, the posterior $\tilde{\pi}(x|\mathcal{S})$ obtained from Algorithm 2 converges in L_2-norm to $\pi(x)$:*

$$\mathbb{E}_{\mathcal{S}}\left(\mathbb{E}\left((\tilde{\pi}(x)-\pi(x))^2\right)\right)=\mathcal{O}\left(((1-B)t)^{-\frac{2}{d+2}}\right). \tag{19}$$

Moreover, Algorithm 2 computes the posterior in time $\mathcal{O}(t)$.

Proof. Let $x\in\mathcal{X}$, $\Delta\in]0,1]$ be arbitrary. Suppose we fix the experiment points X and that among these t points, n of them are at most Δ far from x, i.e. $D_x=n$ where $D_x\sim\text{Bin}(t,\Delta^d)$ is the random variable as defined in A.2. We assume without loss of generality that these points are $x_1,...,x_n$. For simplicity, we treat the case where $\alpha=\beta=1$, i.e. the prior for $\tilde{\pi}(x)$ is uniform $\forall x\in\mathcal{X}$. Note that in this case, the coefficients C_i's in Corollary 2 can be written as:

$$C_i^n=\frac{1}{E'}\binom{n-i}{S-i}B^{S-i}, \tag{20}$$

$i=0,...,S$ where E' is the normalization factor and S is the number of observed successes.

$$\begin{aligned}
\mathbb{E}_{\mathbf{s}}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S}))|D_x=n\right]&=\sum_{s=0}^{n}Pr(S_x=s)\sum_{i=0}^{s}C_i^{n,s}(x)\frac{i+1}{n+2}\\
&=\sum_{s=0}^{n}Pr(S_x=s)\frac{\sum_{i=0}^{s}\binom{n-i}{s-i}B^{s-i}\frac{i+1}{n+2}}{\sum_{j=0}^{s}\binom{n-j}{s-j}B^{s-j}}\\
&=\sum_{s=0}^{n}Pr(S_x=s)\left(\frac{s+1}{n+2}-\frac{\sum_{i=0}^{s}\binom{n-s+i}{i}B^i\frac{i}{n+2}}{\sum_{j=0}^{s}\binom{n-s+j}{j}B^j}\right)\\
&=\sum_{s=0}^{n}Pr(S_x=s)\left(\frac{s+1}{n+2}-\frac{B}{1-B}\left(1-\frac{s+1}{n+2}\right)\left(1-\frac{\binom{n+1}{s}B^s}{\sum_{j=0}^{s}\binom{n+1}{j}B^j(1-B)^{s-j}}\right)\right)\\
&=\frac{1+\sum_{i=1}^{n}(B+(1-B)\pi(x_i))}{(n+2)(1-B)}-\frac{B}{1-B}\\
&+\frac{B}{1-B}\sum_{s=0}^{n}Pr(S_x=s)\left(1-\frac{s+1}{n+2}\right)\frac{\binom{n+1}{s}B^s(1-B)^{n-s+1}}{\sum_{j=0}^{s}\binom{n+1}{j}B^j(1-B)^{n+1-j}}
\end{aligned}$$

$$= \frac{\sum_{i=1}^{n} \pi(x_i)}{n+2} + \frac{1-2B}{(1-B)(n+2)} + \frac{B}{1-B} \sum_{s=0}^{n} Pr(S_x = s) \left(1 - \frac{s+1}{n+2}\right) \frac{\binom{n+1}{s} B^s (1-B)^{n-s+1}}{\sum_{j=0}^{s} \binom{n+1}{j} B^j (1-B)^{n+1-j}}$$

At the fourth equality, we used the fact that $\sum_{j=0}^{s} \binom{n-j}{s-j} B^{s-j} = \sum_{j=0}^{s} \binom{n+1}{j} B^j (1-B)^{s-j}$, which can be shown by induction over s. We also used the following calculations:

$$\begin{aligned}
\sum_{i=0}^{s} \binom{n-s+i}{i} B^i i &= (n-s+1) \sum_{i=1}^{s} \binom{n-s+i}{i-1} B^i \\
&= B(n-s+1) \sum_{i=0}^{s-1} \binom{n-s+1+i}{i} B^i \\
&= B(n-s+1) \left(\sum_{i=0}^{s-1} \binom{n-s+i}{i} B^i + \sum_{i=1}^{s-1} \binom{n-s+i}{i-1} B^i \right) \\
&= B(n-s+1) \left(\sum_{i=0}^{s} \binom{n-s+i}{i} B^i - \binom{n}{s} B^s + B \sum_{i=0}^{s-1} \binom{n-s+1+i}{i} B^i - \binom{n+1}{s} B^s \right)
\end{aligned}$$

Therefore, by equaling lines 2 and 4 and using $\binom{n+1}{s+1} = \binom{n+1}{s} + \binom{n}{s}$, we get:

$$\sum_{i=0}^{s-1} \binom{n-s+1+i}{i} B^i = \frac{1}{1-B} \left(\sum_{i=0}^{s} \binom{n-s+i}{i} B^i - \binom{n+1}{s+1} B^s \right) \tag{21}$$

Thus:

$$\sum_{i=0}^{s} \binom{n-s+i}{i} B^i i = \frac{B}{1-B} \left(1 - \frac{s+1}{n+2}\right) \left(\sum_{i=0}^{s} \binom{n-s+i}{i} B^i - \binom{n+1}{s+1} B^s \right) \tag{22}$$

Let $Z \sim \text{Bin}(n+1, B)$. Then:

$$\left| \sum_{s=0}^{n} Pr(S_x = s) \left(1 - \frac{s+1}{n+2}\right) \frac{\binom{n+1}{s} B^s (1-B)^{n-s+1}}{\sum_{j=0}^{s} \binom{n+1}{j} B^j (1-B)^{n+1-j}} \right| \leq \sum_{s=0}^{t} Pr(S_x = s) \frac{Pr(Z = s)}{Pr(Z \leq s)} \tag{23}$$

We know that $\mathbb{E}(Z) = (n+1)B$ and $\mathbb{E}(S_x) = nB + \sum_{i=1}^{n} (1-B)\pi(x_i)$. We then have:

$$\begin{aligned}
\sum_{s=0}^{n} Pr(S_x = s) \frac{Pr(Z = s)}{Pr(Z \leq s)} &= \sum_{s=0}^{\frac{\mathbb{E}(Z)+\mathbb{E}(S_x)}{2}} Pr(S_x = s) \frac{Pr(Z = s)}{Pr(Z \leq s)} + \sum_{s=\frac{\mathbb{E}(Z)+\mathbb{E}(S_x)}{2}+1}^{n} Pr(S_x = s) \frac{Pr(Z = s)}{Pr(Z \leq s)} \\
&\leq Pr\left(S_x \leq \frac{\mathbb{E}(Z) + \mathbb{E}(S_x)}{2}\right) + 2Pr\left(Z \geq \frac{\mathbb{E}(Z) + \mathbb{E}(S_x)}{2}\right) \\
&\leq 3e^{-\frac{(\mathbb{E}(S_x) - \mathbb{E}(Z))^2}{2n}} \\
&\leq Ce^{-\frac{(1-B)^2 \bar{\pi} n}{2}}
\end{aligned}$$

where $C \in \mathbb{R}$, $\bar{\pi} = \frac{1}{n} \sum_{i=1}^{n} \pi(x_i) > 0$. In the second step, we used $Pr(Z \leq s) \geq \frac{1}{2}$ for any $s \geq \mathbb{E}(Z)$. The last step follows from Hoeffding's inequality. So the previous upper bound decays exponentially to 0. We thus have:

$$\mathbb{E}_{\mathbf{s}} \left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S})) | D_x = n \right] = \frac{\sum_{i=1}^{n} \pi(x_i)}{n+2} + \frac{1-2B}{(1-B)(n+2)} \tag{24}$$

We now bound the second moment of $\tilde{\pi}(x|\mathcal{S})$. With the same notations as previously, we have:

$$\mathbb{E}_{\mathbf{s}}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S})^2)|D_x=n\right]=\sum_{s=0}^{n}Pr(S_x=s|D_x=n)\sum_{i=0}^{s}C_i^{n,s}\frac{(i+1)(i+2)}{(n+2)(n+3)}$$
$$=\sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\frac{(s+1)(s+2)}{(n+2)(n+3)}-2\frac{s+1}{n+3}\sum_{i=0}^{s}C_{s-i}^{n,s}\frac{i}{n+2}+\sum_{i=0}^{s}C_{s-i}^{n,s}\frac{i(i-1)}{(n+2)(n+3)}+\mathcal{O}\left(\frac{1}{n+2}\right)\right)$$
$$=\sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\frac{(s+1)(s+2)}{(n+2)(n+3)}-2\frac{B}{1-B}\frac{(s+1)(n-s+1)}{(n+2)(n+3)}\left(1-\frac{\binom{n+1}{s}B^s}{\sum_{j=0}^{s}\binom{n+1}{j}B^j(1-B)^{s-j}}\right)\right.$$
$$\left.+\frac{B^2}{1-B^2}\frac{(n-s+1)(n-s+2)}{(n+2)(n+3)}\left(\frac{1+B}{1-B}-\frac{2\binom{n+1}{s}\frac{B^{s+1}}{1-B}+\binom{n+2}{s}B^s+\binom{n+1}{s-1}B^{s-1}}{\sum_{j=0}^{s}\binom{n+1}{j}B^j(1-B)^{s-j}}\right)\right)$$
$$=\frac{1}{(n+2)(n+3)}\sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\frac{s^2}{(1-B)^2}-2sn\frac{B}{(1-B)^2}+\frac{B^2}{(1-B)^2}n^2\right)+\mathcal{O}\left(\frac{1}{(1-B)(n+2)}\right)$$
$$=\frac{1}{(1-B)^2(n+2)^2}\sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\left(\sum_{i=1}^{n}(B+(1-B)\pi(x_i))\right)^2\right.$$
$$\left.-2Bn\sum_{i=1}^{n}(B+(1-B)\pi(x_i))+B^2n^2\right)+\mathcal{O}\left(\frac{1}{(1-B)(n+2)}\right)$$
$$=\frac{1}{(n+2)^2}\sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\sum_{i=1}^{n}\pi(x_i)\right)^2+\mathcal{O}\left(\frac{1}{(1-B)(n+2)}\right)$$

where the four terms with denominator $\sum_{j=0}^{s}\binom{n+1}{j}B^j(1-B)^{s-j}$ in the third line can be shown to decay exponentially fast to 0 similarly as previously. We computed $\sum_{i=0}^{s}C_{s-i}^{n,s}\frac{i(i-1)}{(n+2)(n+3)}$ in the second line using similar calculations as were done for $\sum_{i=0}^{s}C_{s-i}^{n,s}\frac{i}{n+2}$:

$$\sum_{i=0}^{s}\binom{n-s+i}{i}B^i i(i-1)=B^2(n-s+1)(n-s+2)\sum_{i=0}^{s-2}\binom{n-s+2+i}{i}B^i \tag{25}$$

Using the identity $\binom{n+2}{k+2}=\binom{n}{k+2}+2\binom{n}{k+1}+\binom{n}{k}$, we have:

$$\sum_{i=0}^{s-2}\binom{n-s+2+i}{i}B^i=\sum_{i=0}^{s-2}\binom{n-s+i}{i}B^i+2\sum_{i=1}^{s-2}\binom{n-s+i}{i-1}B^i+\sum_{i=2}^{s-2}\binom{n-s+i}{i-2}B^i$$
$$=\sum_{i=0}^{s-2}\binom{n-s+i}{i}B^i+2\sum_{i=1}^{s-3}\binom{n-s+i+1}{i}B^{i+1}+\sum_{i=2}^{s-3}\binom{n-s+i+2}{i}B^{i+2}$$
$$=\sum_{i=0}^{s}\binom{n-s+i}{i}B^i-\binom{n-1}{s-1}B^{s-1}-\binom{n}{s}B^s$$
$$+2B\sum_{i=1}^{s-1}\binom{n-s+i+1}{i}B^i-2\binom{n-1}{s-2}B^{s-1}-2\binom{n}{s-1}B^s$$
$$+B^2\sum_{i=2}^{s-2}\binom{n-s+i+2}{i}B^i-\binom{n-1}{s-3}B^{s-1}-\binom{n}{s-2}B^s$$

Therefore, by isolating the term $\sum_{i=0}^{s-2}\binom{n-s+2+i}{i}B^i$, simplifying binomial coefficients and using equation (21), we get:

$$\sum_{i=0}^{s-2}\binom{n-s+2+i}{i}B^i=\frac{1}{1-B^2}\left(\frac{1+B}{1-B}\sum_{i=0}^{s}\binom{n-s+i}{i}B^i-\binom{n+2}{s}B^s-2\binom{n+1}{s}\frac{B^{s+1}}{1-B}\right.$$

$$-\binom{n+1}{s-1}B^{s-1}\Bigg)$$

Thus:

$$\begin{aligned}
\mathbb{E}_{\mathbf{s}}\left[\mathbb{E}((\tilde{\pi}(x|\mathcal{S})-\pi(x))^2)|D_x=n\right] &= \mathbb{E}_{\mathbf{s}}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S})^2)|D_x=n\right]-2\pi(x)\mathbb{E}_{\mathbf{s}}\left[\mathbb{E}\tilde{\pi}(x|\mathcal{S})|D_x=n\right]+\pi(x)^2\\
&= \sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\left(\frac{\sum_{i=1}^n\pi(x_i)}{n+1}\right)^2-2\pi(x)\frac{\sum_{i=1}^n\pi(x_i)}{n+2}+\pi(x)^2\right)+\mathcal{O}\left(\frac{1}{(1-B)(n+2)}\right)\\
&= \sum_{s=0}^{n}Pr(S_x=s|D_x=n)\left(\frac{\sum_{i,j=1}^n(\pi(x_i)-\pi(x))(\pi(x_j)-\pi(x))}{(n+1)^2}\right)+\mathcal{O}\left(\frac{1}{(1-B)(n+2)}\right)\\
&\leq L^2\Delta^2+\mathcal{O}\left(\frac{1}{(1-B)(n+2)}\right)
\end{aligned}$$

By taking the expectation over X, we finally get:

$$\begin{aligned}
\mathbb{E}_{X,\mathbf{s}}\left[\mathbb{E}((\tilde{\pi}(x|\mathcal{S})-\pi(x))^2)\right] &= \sum_{n=0}^{t}Pr(D_x=n)\mathbb{E}_{\mathbf{s}}\left[\mathbb{E}((\tilde{\pi}(x|\mathcal{S})-\pi(x))^2)|D_x=n\right]\\
&\leq L^2\Delta^2+\mathcal{O}\left(\frac{1}{(1-B)\Delta^d t}\right)
\end{aligned}$$

If we choose $\Delta=\frac{1}{L^{\frac{2}{d+2}}}((1-B)t)^{-\frac{1}{d+2}}$, we obtain the desired result.

□

A.4. Proof of convergence in the general dynamic case

Theorem 4. *Let $\pi:[0,1]^d\to]0,1]$ be L-Lipschitz continuous. Suppose we observe the results of experiments $\mathcal{S}=\{(x_i,s_i,1-B_i,B_i)\}_{i=1,\dots,t}$ where $s_i\sim$ Bernoulli($(1-(B_i+\epsilon_i))\pi(x_i)+B_i+\epsilon_i$), i.e. contextual features are noisy. We assume ϵ_i's are independent random variables with zero mean and variance σ^2. Experiment points $\{x_i\}_{i=1,\dots,t}$ are assumed to be uniformly distributed over the space. Then, $\forall x\in\mathcal{X}$, the posterior $\tilde{\pi}(x|\mathcal{S})$ obtained from Algorithm 3 converges in L_2-norm to $\pi(x)$:*

$$\mathbb{E}_{\mathcal{S}}\left(\mathbb{E}\left((\tilde{\pi}(x|\mathcal{S})-\pi(x))^2\right)\right)=\mathcal{O}\left(c(B,\sigma^2)t^{-\frac{2}{d+2}}\right), \tag{26}$$

where $c(B,\sigma^2)$ is a constant depending on $\{B_i\}_{i=1,\dots,t}$ and the noise σ^2. Moreover, Algorithm 3 computes the posterior in time $\mathcal{O}(t)$.

Proof. The proof of theorem 3 can be completely adapted to this new setting. Let $x\in\mathcal{X}$, $\Delta\in[0,1]$ be arbitrary. Suppose we fix the experiment points X and that among these t points, n of them are at most Δ far from x. We assume without loss of generality that these points are $x_1,...,x_n$. We then define $B_X=\frac{1}{n}\sum_{i=1}^n B_i$.

$$\begin{aligned}
\mathbb{E}_{\mathbf{s}}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S}))|D_x=n\right] &= \sum_{s=0}^{n}Pr(S_x=s)\sum_{i=0}^{s}C_i^{n,s}(x)\frac{i+1}{n+2}\\
&= \sum_{s=0}^{n}Pr(S_x=s)\frac{\sum_{i=0}^{s}\binom{n-i}{s-i}B_X^{s-i}\frac{i+1}{n+2}}{\sum_{j=0}^{s}\binom{n-j}{s-j}B_X^{s-j}}\\
&= \sum_{s=0}^{n}Pr(S_x=s)\left(\frac{s+1}{n+2}-\frac{\sum_{i=0}^{s}\binom{n-s+i}{i}B_X^i\frac{i}{n+2}}{\sum_{j=0}^{s}\binom{n-s+j}{j}B_X^j}\right)\\
&= \sum_{s=0}^{n}Pr(S_x=s)\left(\frac{s+1}{n+2}-\frac{B_X}{1-B_X}\left(1-\frac{s+1}{n+2}\right)\left(1-\frac{\binom{n+1}{s}B_X^s}{\sum_{j=0}^{s}\binom{n+1}{j}B_X^j(1-B_X)^{s-j}}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1+\sum_{i=1}^n (B_i+\epsilon_i+(1-B_i-\epsilon_i)\pi(x_i))}{(n+2)(1-B_X)} - \frac{B_X}{1-B_X} \\
&+ \frac{B_X}{1-B_X}\sum_{s=0}^n Pr(S_x=s)\left(1-\frac{s+1}{n+2}\right)\frac{\binom{n+1}{s}B_X^s(1-B_X)^{n-s+1}}{\sum_{j=0}^s\binom{n+1}{j}B_X^j(1-B_X)^{n+1-j}} \\
&= \frac{\sum_{i=1}^n(1-B_i)\pi(x_i)}{(1-B_X)(n+2)} + \frac{1-2B+\sum_{i=1}^n\epsilon_i(1-\pi(x_i))}{(1-B_X)(t+2)} \\
&+ \frac{B_X}{1-B_X}\sum_{s=0}^n Pr(S_x=s)\left(1-\frac{s+1}{n+2}\right)\frac{\binom{n+1}{s}B_X^s(1-B_X)^{n-s+1}}{\sum_{j=0}^s\binom{n+1}{j}B_X^j(1-B_X)^{n+1-j}}
\end{aligned}
$$

Let $Z \sim \text{Bin}(n+1, B_X)$. Then:

$$
\left|\sum_{s=0}^n Pr(S_x=s)\left(1-\frac{s+1}{n+2}\right)\frac{\binom{n+1}{s}B_X^s(1-B_X)^{n-s+1}}{\sum_{j=0}^s\binom{n+1}{j}B_X^j(1-B_X)^{n+1-j}}\right| \leq \sum_{s=0}^t Pr(S_x=s)\frac{Pr(Z=s)}{Pr(Z\leq s)} \tag{27}
$$

We know that $\mathbb{E}(Z)=(n+1)B_X$ and $\mathbb{E}(S_x)=nB_X+\sum_{i=1}^n(1-B_i)\pi(x_i)+\sum_{i=1}^n\epsilon_i(1-\pi(x_i))$. Since $\mathbb{E}(\epsilon_i)=0$, then $\mathbb{E}(S_x)-\mathbb{E}(Z)$ will also increase linearly with n and thus the previous upper bound also decreases exponentially with n to 0 with very high probability. We thus have:

$$
\begin{aligned}
\mathbb{E}_{\mathbf{s},\epsilon}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S}))|D_x=n\right] &= \frac{\sum_{i=1}^n(1-B_i)\pi(x_i)}{(1-B_X)(n+2)} + \mathbb{E}_\epsilon\left[\frac{\sum_{i=1}^n\epsilon_i(1-\pi(x_i))}{(1-B_X)(t+2)}\right] + \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)}\right) \\
&= \frac{\sum_{i=1}^n(1-B_i)\pi(x_i)}{(1-B_X)(n+2)} + \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)}\right)
\end{aligned}
$$

We now bound the second moment of $\tilde{\pi}(x|\mathcal{S})$. With the same notations as previously, we have:

$$
\begin{aligned}
&\mathbb{E}_S\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S})^2)|D_x=n\right] = \sum_{s=0}^n Pr(S_x=s|D_x=n)\sum_{i=0}^s C_i^{n,s}\frac{(i+1)(i+2)}{(n+2)(n+3)} \\
&= \sum_{s=0}^n Pr(S_x=s|D_x=n)\left(\frac{(s+1)(s+2)}{(n+2)(n+3)} - 2\frac{s+1}{n+3}\sum_{i=0}^s C_{s-i}^{n,s}\frac{i}{n+2} + \sum_{i=0}^s C_{s-i}^{n,s}\frac{i(i-1)}{(n+2)(n+3)} + \mathcal{O}\left(\frac{1}{n+2}\right)\right) \\
&= \sum_{s=0}^n Pr(S_x=s|D_x=n)\left(\frac{(s+1)(s+2)}{(n+2)(n+3)} - 2\frac{B_X}{1-B_X}\frac{(s+1)(n-s+1)}{(n+2)(n+3)}\left(1-\frac{\binom{n+1}{s}B_X^s}{\sum_{j=0}^s\binom{n+1}{j}B_X^j(1-B_X)^{s-j}}\right)\right. \\
&\left.+\frac{B_X^2}{1-B_X^2}\frac{(n-s+1)(n-s+2)}{(n+2)(n+3)}\left(\frac{1+B_X}{1-B_X} - \frac{2\binom{n+1}{s}\frac{B_X^{s+1}}{1-B_X}+\binom{n+2}{s}B_X^s+\binom{n+1}{s-1}B_X^{s-1}}{\sum_{j=0}^s\binom{n+1}{j}B_X^j(1-B_X)^{s-j}}\right)\right) \\
&= \frac{1}{(n+2)(n+3)}\sum_{s=0}^n Pr(S_x=s|D_x=n)\left(\frac{s^2}{(1-B_X)^2} - 2sn\frac{B_X}{(1-B_X)^2} + \frac{B_X^2}{(1-B_X)^2}n^2\right) + \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)}\right) \\
&= \frac{1}{(1-B_X)^2(n+2)^2}\sum_{s=0}^n Pr(S_x=s|D_x=n)\left(\sum_{i=1}^n(B_i+\epsilon_i+(1-B_i-\epsilon_i)\pi(x_i))^2\right. \\
&\left.-2B_Xn\sum_{i=1}^n(B_i+\epsilon_i+(1-B_i-\epsilon_i)\pi(x_i)) + B_X^2n^2\right) + \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)}\right) \\
&= \frac{1}{(1-B_X)^2(n+2)^2}\sum_{s=0}^n Pr(S_x=s|D_x=n)\left(2\sum_{i,j=1}^n\epsilon_i(1-\pi(x_i))\pi(x_j) + \sum_{i,j=1}^n\epsilon_i\epsilon_j(1-\pi(x_i))(1-\pi(x_j))\right.
\end{aligned}
$$

$$+\left(\sum_{i=1}^{n}(1-B_i)\pi(x_i)\right)^2\Bigg) + \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)}\right)$$

where the four terms with denominator $\sum_{j=0}^{s}\binom{n+1}{j}B_X^j(1-B_X)^{s-j}$ in the third line can be shown to decay exponentially fast to 0 similarly as previously. Taking the expectation over ϵ, we then get:

$$\mathbb{E}_{S,\epsilon}\left[\mathbb{E}(\tilde{\pi}^t(x)^2)|D_x=n\right] = \sum_{s=0}^{n} Pr(S_x=s|D_x=n)\left(\frac{\sum_{i=1}^{n}(1-B_i)\pi(x_i)}{(1-B_X)(n+1)}\right)^2 + \mathcal{O}\left(\frac{1}{(1-B_X)(n+1)} + \frac{\sigma^2}{(1-B_X)^2(n+1)}\right)$$

Thus:

$$\begin{aligned}
\mathbb{E}_{\mathbf{s},\epsilon}\left[\mathbb{E}((\tilde{\pi}(x|\mathcal{S})-\pi(x))^2)|D_x=n\right] &= \mathbb{E}_{\mathbf{s},\epsilon}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S})^2)|D_x=n\right] - 2\pi(x)\mathbb{E}_{S,\epsilon}\left[\mathbb{E}(\tilde{\pi}(x|\mathcal{S}))|D_x=n\right] + \pi(x)^2\\
&= \sum_{s=0}^{n} Pr(S_x=s|D_x=n)\left(\left(\frac{\sum_{i=1}^{n}(1-B_i)\pi(x_i)}{(1-B_X)(n+1)}\right)^2 - 2\pi(x)\frac{\sum_{i=1}^{n}(1-B_i)\pi(x_i)}{(1-B_X)(n+2)} + \pi(x)^2\right)\\
&+ \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)} + \frac{\sigma^2}{(1-B_X)^2(n+2)}\right)\\
&= \sum_{s=0}^{n} Pr(S_x=s|D_x=n)\left(\frac{\sum_{i,j=1}^{n}(1-B_i)(1-B_j)(\pi(x_i)-\pi(x))(\pi(x_j)-\pi(x))}{(1-B_X)^2(n+2)^2}\right)\\
&+ \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)} + \frac{\sigma^2}{(1-B_X)^2(n+2)}\right)\\
&\leq L^2\Delta^2 + \mathcal{O}\left(\frac{1}{(1-B_X)(n+2)} + \frac{\sigma^2}{(1-B_X)^2(n+2)}\right)
\end{aligned}$$

Finally, by taking the expectation over experiment points X, we get:

$$\begin{aligned}
\mathbb{E}_{X,\mathbf{s},\epsilon}\left[\mathbb{E}((\tilde{\pi}(x|\mathcal{S})-\pi(x))^2)\right] &= \sum_{n=0}^{t} Pr(D_x=n)\mathbb{E}_{S,\epsilon}\left[\mathbb{E}((\tilde{\pi}^t(x)-\pi(x))^2)|D_x=n\right]\\
&\leq L^2\Delta^2 + \mathcal{O}\left(\frac{C^{(1)}}{\Delta^d t} + \frac{C^{(2)}\sigma^2}{\Delta^d t}\right)
\end{aligned}$$

where $C^{(i)} = \mathbb{E}_X\left[\frac{1}{(1-B_X)^i}\right]$. Therefore, if we choose $\Delta = \frac{1}{L^{\frac{2}{d+2}}}t^{-\frac{1}{d+2}}$, then we obtain the desired result.

□

B. Smooth Beta processes for classification

In this appendix, we extend the convergence rates in L_2 function approximation to L_1 and Bayes risk (misclassification error). These are to be understood as corollaries to the proofs presented in Sec. A. Furthermore, we establish the connection between SBPs in the static setting and nearest neighbor techniques. However, our method allows for precise prior knowledge injection, whose efficiency is empirically demonstrated on a synthetic classification experiment.

B.1. Convergence in L_1 norm

Leaving out constants, Theorems 1, 3, and 4 provide convergence rates of the type $\mathcal{O}\left(t^{-\frac{2}{d+2}}\right)$. In all three settings, we obtain the following corollary for the error in L_1 norm:

Corollary 3 (Convergence in L_1)**.** *Under the assumptions of Theorems 1, 3, and 4, the corresponding Algorithms 1, 2, and 3 converge in L_1 norm to $\pi(x)$:*

$$\sup_{x\in[0,1]^d} \mathbb{E}_{\mathcal{S}}\left(\mathbb{E}\left|\tilde{\pi}(x|\mathcal{S})-\pi(x)\right|\right)=\mathcal{O}\left(t^{-\frac{1}{d+2}}\right),$$

where we leave out the constants of the respective theorems.

Proof. For all three cases, the statement follows from the application of Jensen's inequality. We have

$$\mathbb{E}_{\mathcal{S}}\left(\mathbb{E}\left|\tilde{\pi}(x|\mathcal{S})-\pi(x)\right|\right)=\mathbb{E}_{\mathcal{S}}\left(\mathbb{E}\left(\sqrt{\left(\tilde{\pi}(x|\mathcal{S})-\pi(x)\right)^2}\right)\right)\leq\sqrt{\mathbb{E}_{\mathcal{S}}\left(\mathbb{E}\left(\left(\tilde{\pi}(x|\mathcal{S})-\pi(x)\right)^2\right)\right)}, \tag{28}$$

which yields the presented convergence rates by taking the square root of the rates of the respective Theorems for L_2 convergence.

□

B.2. Convergence in Bayes risk

In the classification setting, it is natural to use the posterior predictive of the Beta-Bernoulli model. Therefore, we have the classifier $\tilde{s}(x|\mathcal{S})$ based on the posterior parameters $\tilde{\alpha}(x),\tilde{\beta}(x)$:

$$\tilde{s}(x|\mathcal{S})=\begin{cases}1 & \text{if } \frac{\tilde{\alpha}(x)}{\tilde{\alpha}(x)+\tilde{\beta}(x)}\geq 0.5,\\ 0 & \text{otherwise.}\end{cases} \tag{29}$$

To estimate the performance of a classifier, the agreement with the Bayes optimal classifier is used. The Bayes risk of a classification problem is minimized by the omniscient Bayes classifier:

Definition 1 (Bayes risk and optimal classifier)**.** *For any $x\in\mathcal{X}$, the Bayes risk of a classifier $\tilde{s}:\mathcal{X}\to\{0,1\}$ is given by*

$$R(\tilde{s},x)=\mathbb{P}_{s\sim B(\pi(x))}\left[s\neq\tilde{s}(x)\right]. \tag{30}$$

The Bayes optimal classifier is given based on the underlying probability function $\pi(x)$. The corresponding decision rule is

$$s^*(x)=\mathbb{1}_{\pi(x)\geq 0.5}, \tag{31}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. This decision rule incurs the following optimal Bayes risk:

$$R^*(x)=R(s^*,x)=\min\{\pi(x),1-\pi(x)\}. \tag{32}$$

To relate the convergence in L_1 to Bayes risk, the following simple Lemma is useful and allows to establish convergence in Bayes risk in Thm. 6.

Lemma 1. *Suppose $B(\cdot)$ denotes a Bernoulli distribution, $p,q\in[0,1]$ and $s'\in\{0,1\}$. Then we have*

$$\mathbb{P}_{s\sim B(p)}\left[s\neq s'\right]\leq\mathbb{P}_{s\sim B(q)}\left[s\neq s'\right]+|p-q|, \tag{33}$$

which relates the misclassification directly to ℓ_1 loss.

Proof. Suppose $s'=1$. Then the left-hand side is p and the right-hand side gives $q+|p-q|$. If $p>=q$, we have for the right-hand side $q+p-q=p$ and equality holds. If $p<q$, we have for the right-hand side $q+q-p$ and $2p\leq 2q$ by the assumption $p<q$. The same argument works for $s'=0$ by symmetry. □

Theorem 6 (Convergence in Bayes risk)**.** *Under the assumptions of Theorems 1, 3, and 4, the classifier in Eq. (29) based on the posterior parameters obtained by the corresponding Algorithms 1, 2, and 3 uniformly converges to the risk of the Bayes optimal classifier s^*, i.e. for any $x\in\mathcal{X}$:*

$$\mathbb{E}_{\mathcal{S}}\left[R(\tilde{s},x)\right]\leq R^*(x)+\mathcal{O}\left(t^{-\frac{1}{d+2}}\right), \tag{34}$$

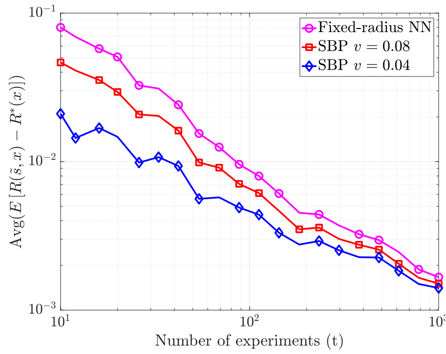

Figure 4. Bayes risk of SBP with specified informative prior, which is identical to the underlying function $\pi(x)$, compared to fixed-radius NN which can not specify a prior in its standard framework.

where constants of the respective theorems are left out (see Sec. B.1).

Proof. Using Lemma 1, we have the following for any $x \in \mathcal{X}$:

$$\begin{aligned} R(\tilde{s}, x) = \mathbb{P}_{s\sim \mathrm{B}(\pi(x))}\left[s \neq \tilde{s}(x|\mathcal{S})\right] &\leq \mathbb{P}_{s\sim \mathrm{B}(\tilde{\pi}(x|\mathcal{S}))}\left[s \neq \tilde{s}(x|\mathcal{S})\right] + |\tilde{\pi}(x|\mathcal{S}) - \pi(x)| \\ &= \min\{\tilde{\pi}(x|\mathcal{S}), 1 - \tilde{\pi}(x|\mathcal{S})\} + |\tilde{\pi}(x|\mathcal{S}) - \pi(x)| \\ &\leq \min\{\pi(x), 1 - \pi(x)\} + 2\,|\tilde{\pi}(x|\mathcal{S}) - \pi(x)| \\ &= R^*(x) + 2\,|\tilde{\pi}(x|\mathcal{S}) - \pi(x)| \end{aligned} \tag{35}$$

Now, we can apply the convergence in L_1 of Corollary 3 and get the desired result:

$$\mathbb{E}_{\mathcal{S}}\left(\mathbb{P}_{s\sim \mathrm{B}(\pi(x))}\left(s \neq \tilde{s}(x|\mathcal{S})\right)\right) \leq R^*(x) + \mathcal{O}\left(t^{-\frac{1}{d+2}}\right). \tag{36}$$

□

B.3. Related methods and practical considerations

Smooth Beta processes are designed for probability function approximation, in which case the estimation of the standard deviation on top of the function approximation is useful. In the particular static classification setting, SBPs are tightly connected to the fixed-radius nearest neighbors (*NN*) classifier. SBPs have the advantage to specify a prior, which is useful to incorporate knowledge or combat biased data. In contrast to fixed-radius NN, SBPs perform additive smoothing like the famous Krichevsky-Trofimov estimator (Krichevsky & Trofimov, 1981) by adding *pseudo-counts*. Despite the introduced bias, SBPs converge optimally to the Bayes classifier: the rate proven in Thm. 6 matches the lower-bound established by Audibert et al. (2007) for classification.

On a practical side, faster inference methods are available due to the algorithmic fixed-radius nearest neighbors problem. Both exact (e.g. k-d and ball trees) and approximate (e.g. hashing-based) methods can be used for faster inference schemes. For further practical considerations and background on the fixed-radius NN algorithm, we refer to Chen et al. (2018).

We conduct a synthetic experiment in order to show how the specification of a prior can help in the low data regime. We compare the convergence of SBP with various priors and the standard fixed-radius NN algorithm. For an informative prior, we set the prior $\tilde{\pi}(x) \sim \mathrm{Beta}(\alpha(x), \beta(x))$ such that $\mathbb{E}[\tilde{\pi}(x)] = \pi(x)$ and $\mathbb{V}[\tilde{\pi}(x)] = v$. In Fig. 4, we compare the convergence for different values of v: in the low data regime, SBPs can profit strongly from an informative prior. With increasing number of observations, the approximation quality varies less as we expect it to happen for a Bayesian method. Asymptotically, the convergence rate is the same.