Near-Optimal Target Learning With Stochastic Binary Signals

Mithun Chakraborty, Sanmay Das, Malik Magdon-Ismail,

Department of Computer Science, Rensselaer Polytechnic Institute

The Learning Problem

We wish to learn, within error tolerance ε, an unknown value $V \in \mathbb{R}$, given access only to sequential binary thresholded observations under an additive Gaussian noise model. At t = 0, 1, 2,..., we maintain a probability distribution $p_t(v)$ over V and set a threshold θ_t. The observation (stochastic binary signal) is

$$x_t = \text{sign}(V + z_t - \theta_t),$$

where $z_t \sim \mathcal{N}(0, \sigma_z)$.

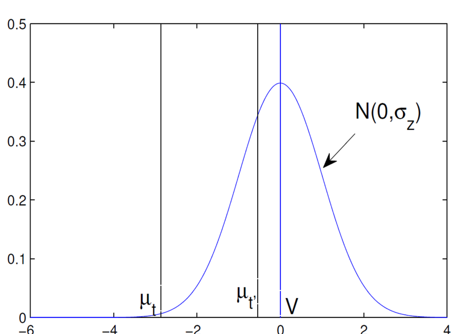

Fig 1. The effect of noise increases (*i.e.* probability of getting a noisy signal becomes greater than 1/2) as our estimate gets closer to the target

Starting prior: Gaussian $\mathcal{N}(\mu_0, \sigma_0)$. Define $\rho_t = \sigma_t/\sigma_z$, the information disadvantage (measure of the harshness of the learning environment).

Myopic threshold: $\theta_t = \mu_t$, the mean of the belief distribution.

Bayesian setting: The probability estimate is updated as:

$$p_{t+1}(v) = \Phi\left(x_t(v - \theta_t)/\sigma_z\right) p_t(v)/A_t,$$

$$A_t = \int_{-\infty}^{\infty} dv \;\; p_t(v) \;\; \Phi\left(x_t(v - \mu_t)/\sigma_z\right),$$

$$\mu_t = \mathbb{E}_{p_t}[V] = \int_{-\infty}^{\infty} dv \; v p_t(v).$$

Bayesian Inference on Exact Signals (Optimal)

If z_t are available, we maintain a Gaussian belief (simplified scalar Kalman filter) with parameters

$$\mu_{t+1} = \frac{\mu_t + \rho_t^2(V + z_t)}{1 + \rho_t^2} \quad \text{and} \quad \rho_{t+1}^2 = \frac{\rho_t^2}{1 + \rho_t^2}.$$

Assume w. l. o. g. $\mu_0 = 0$, $\sigma_z = 1$.

Theorem 1. *Fix* $\varepsilon < \frac{|V|}{2}$, $\delta \le \Phi(-1)$. *For* $t > |V|/2\varepsilon\rho_0^2$, $|V| - |\mathbb{E}[\mu_t]| < \varepsilon$. *For* $t > \max\left\{\frac{2|V|}{\varepsilon\rho_0^2}, \frac{4\zeta^2}{\varepsilon^2}\right\}$ *then* $\Pr[\mu_t > V - \varepsilon] > 1 - \delta$ *for* $V > 0$ *and* $\Pr[\mu_t < V + \varepsilon] > 1 - \delta$ *for* $V < 0$.

where $\zeta = -\Phi^{-1}(\delta)$.

Non-parametric histograms

Exact Bayesian inference on thresholded signals is analytically intractable. Alternative: Numerical Integration.

- Efficiency: $O(Nt)$ computations for posterior $p_t(v)$ with N quadrature points.
- Numerical instability

Solution: Non-parametric, discrete, finite distribution as a near-exact benchmark.

$$p_{t+1}(v_i) = \frac{1}{A_t}\Phi\left(x_t(v_i - \theta_t)\right) p_t(v_i),$$

$$A_t = \sum_{i=1}^{N} \Phi\left(x_t(v_i - \theta_t)\right) p_t(v_i) \qquad \mu_t = \sum_{i=1}^{N} v_i p_t(v_i).$$

- Computationally intense: resolution in the finite prior should be $O(\varepsilon)$ so that $N = \Omega(1/\varepsilon)$.
- Serious problems if V is outside the finite range.

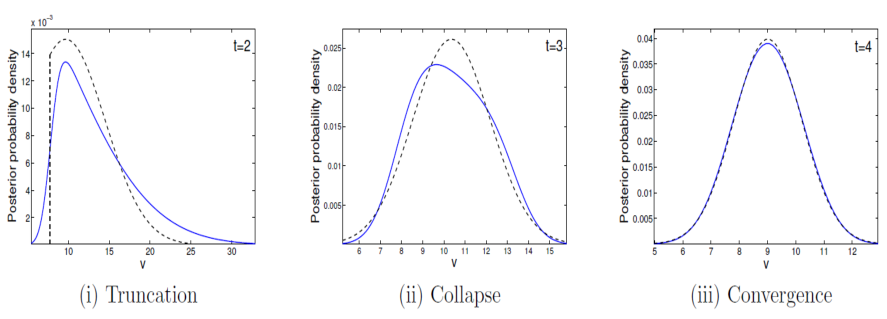

Fig 2. Evolution of $p_t(v)$ using non-parametric histogram representation and Bayesian updates. $p_0(v)$ is N(0,10), V = 9.45.

Our algorithm: approximate inference

Algorithm 1 The Learning Algorithm

Initialize $l_0 = -\infty, r_0 = \infty, m_0 = \mu_0, s_0 = \sigma_0$.
for $t = 0, 1, 2, \ldots$ **do**
  Set threshold at μ_t;
  Receive noisy thresholded signal x_t;
  Update l_t, r_t, m_t, s_t;
  Compute $\mu_{t+1}, \rho_{t+1} = \sigma_{t+1}/\sigma_z$;
end for

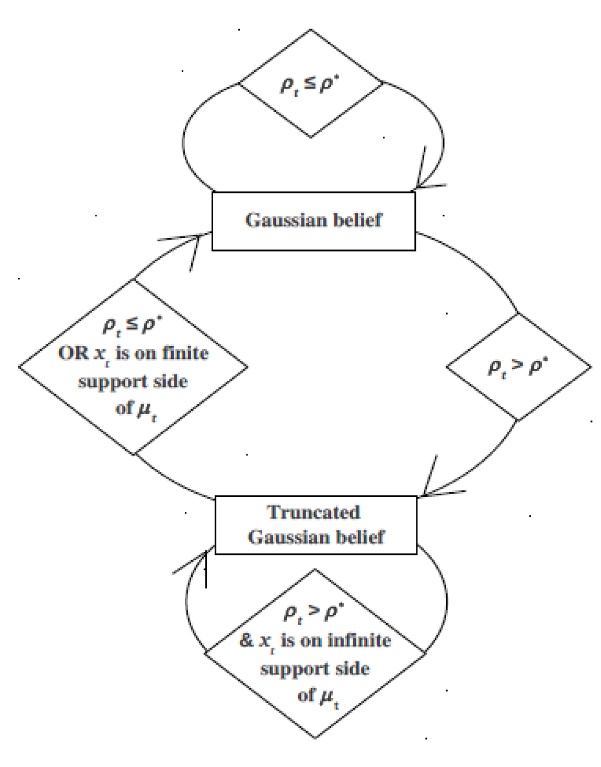

Fig 3. Learner's state transitions

Update rules

Gaussian Approximation (moment matching) to Bayesian posterior ($\rho_t <= \rho^*$):

$$\mu_{t+1} = \mu_t + x_t \frac{\left(\sigma_z\sqrt{2/\pi}\right)\rho_t^2}{\sqrt{1 + \rho_t^2}};$$

$$\rho_{t+1}^2 = \rho_t^2\left[\frac{1 + \rho_t^2(1 - 2/\pi)}{1 + \rho_t^2}\right].$$

Truncation ($\rho_t > \rho^*$):

$$(l_t, \infty, m_t, s_t) \overset{(\theta_t, x_t = +1)}{\longrightarrow} (\theta_t - 2\sigma_z, \infty, m_t, s_t);$$

$$(-\infty, r_t, m_t, s_t) \overset{(\theta_t, x_t = -1)}{\longrightarrow} (-\infty, \theta_t + 2\sigma_z, m_t, s_t).$$

Collapse (truncated Gaussian to Gaussian: entropy matching):

$$(l_t, \infty, m_t, s_t) \overset{(\theta_t, x_t = -1)}{\longrightarrow} (-\infty, \infty, \mu_{t+1}, \rho_{t+1});$$

$$(-\infty, r_t, m_t, s_t) \overset{(\theta_t, x_t = +1)}{\longrightarrow} (-\infty, \infty, \mu_{t+1}, \rho_{t+1}).$$

$$\mu_{t+1} = m_{t+1} + s_{t+1}\left[\frac{N(l') - N(r')}{\Phi(r') - \Phi(l')}\right],$$

$$\sigma_{t+1}^2 = s_{t+1}^2\left[\left(\Phi(r') - \Phi(l')\right)^2 e^{\frac{l'N(l') - r'N(r')}{\Phi(r') - \Phi(l')}}\right].$$

where $l' = (l - m_t)/s_t$, $r' = (r - m_t)/s_t$; ($l = \mu_t - 2\sigma_z$, $r = r_t$) for $x_t = +1$, ($l = l_t$, $r = \mu_t + 2\sigma_z$) for $x_t = -1$.

Convergence properties (for $\rho_0 < 1$)

Theorem 2. *There exist absolute positive constants* $C > 0$ *and* k, $1 \le k < \pi\sqrt{2} \approx 4.443$ *such that, if* $t > C/(\rho_0^2\varepsilon^k)$, *then* $|V| - |\mathbb{E}[\mu_t]| < \varepsilon$.

Thus, the waiting time for the expected mean belief to get within ε of the target V is $O(1/\rho_0^2\varepsilon^k)$ while the corresponding time bound for exact signals is $O(1/\rho_0^2\varepsilon)$. This shows that the dependence of our algorithm on ρ_0 is optimal and the algorithm is polynomial in $1/\varepsilon$. However, simulations show that k is practically equal to 1, as illustrated by an example in Fig. 4.

Theorem 3. *Fix* $0 < \delta < 1$, $0 < \varepsilon < V$, $0 < \rho_0 \le 1$, *and define* $\Delta = V - \varepsilon$. *There is an absolute constant* $C > 0$ *such that if* $t > T = e^{C(\ln(1/\delta) + \Delta)/\varepsilon}/\rho_0^2$, *then with probability at least* $1 - \delta$, $\max_{i \le t} \mu_i > V - \varepsilon$.

Comparing with Theorem 1, our asymptotic dependence on ρ_0 is optimal *i.e.* $O(1/\rho_0^2)$ for small ρ_0 (harsh learning environment).

Experimental results

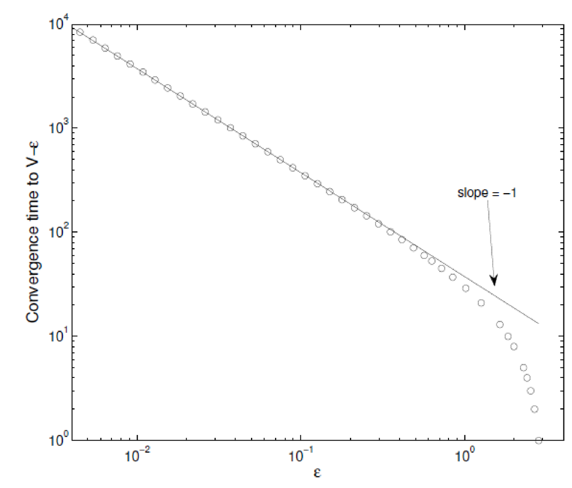

Fig 4. Convergence time approaches $O(1/\varepsilon)$ as ε becomes small

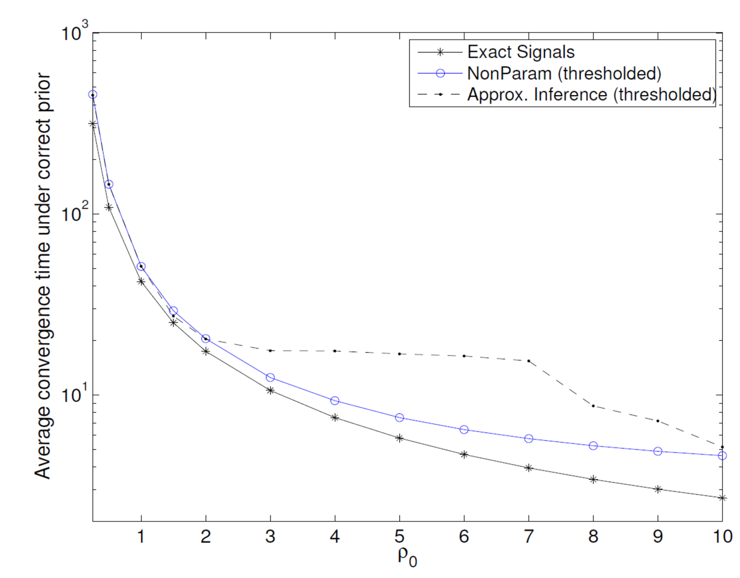

Fig 5. Plot of average correct-prior convergence time *vs* ρ_0, logarithmic along the vertical axis

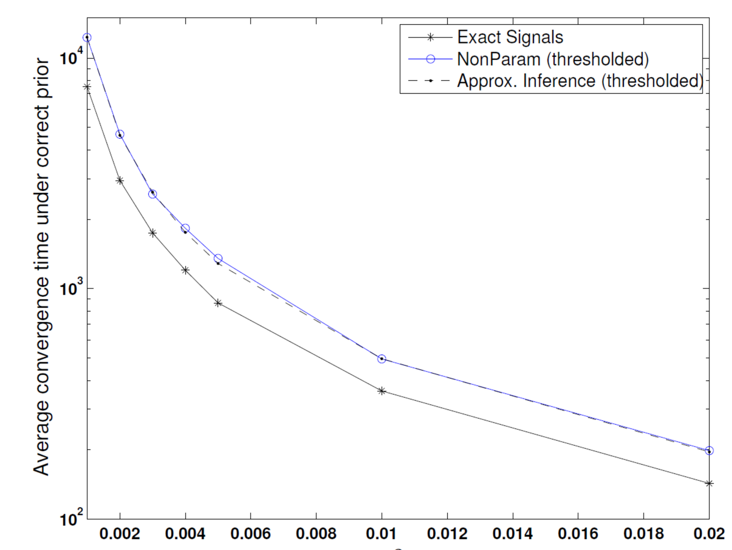

Fig 6. Plot of average correct-prior convergence time *vs* ε, logarithmic along the vertical axis

References

[1] S. Das and M. Magdon-Ismail. Adapting to a market shock: Optimal sequential market-making. *Proc. NIPS*, pages 361-368, 2008.

[2] R Waeber, P.I. Frazier, and S.G. Henderson. A Bayesian Approach to Stochastic Root Finding. *Proc. Winter Simulation Conf.* IEEE, 2011.