I find this paper very interesting. However, I find the main body of the paper too wordy and occasionally unclear. This summary is an attempt to express the main theoretical result in notation form so as to develop more precise understanding.

1 Proposition 1

What this proposition means is the following:

1. Let P_1 and P_2 be 2 distributions over the same set.
2. We refer to P_1 and P_2 as the training distribution and the transfer distribution respectively.
3. Let V be a random variable sampled either from P_1 or P_2.
4. Define the regret over one sample V as the negative log likelihood: $R(V) = -\mathcal{L}(V) = -\log P_\theta(V)$
5. where θ are learnt parameters, found by solving $argmax_\theta E_{V\sim P_1}[\mathcal{L}(V)]$.
6. Let $P_\theta(V)$ decomposes into product of conditional probabilities: $\prod_i P_{\theta_i}\left(V_i|\operatorname{pa}\left(i, V, B_i\right)\right)$
7. where the i component of θ parameterizes the learnt conditional probability of V_i given its causes.
8. If θ_i satisfies $(a), (b)$ in the proposition, that means that $P_{\theta_i}\left(V_i|\operatorname{pa}\left(i, V, B_i\right)\right) = P_1\left(V_i|\operatorname{pa}\left(i, V, B_i\right)\right)$.
9. If θ_i satisfies (c) in the proposition, that means $P_2\left(V_i|\operatorname{pa}\left(i, V, B_i\right)\right) = P_1\left(V_i|\operatorname{pa}\left(i, V, B_i\right)\right)$.
10. If θ_i satisfies $(a), (b), (c)$, then $E_{V\sim P_2}[\frac{\partial \mathcal{L}(V)}{\partial \theta_i}] = 0$

2 Proposition 2

What this proposition means is the following:

1. Let there be 2 random variables A and B.
2. Let $D_2 = \{(a_t, b_t)\}_t$ be the transfer dataset, sampled from the transfer distribution.
3. Let $\mathcal{L}_{A\to B} = \prod_{t=1}^{T} P\left(a_t, b_t|A \to B, \theta_t\right)$ and $\mathcal{L}_{B\to A} = \prod_{t=1}^{T} P\left(a_t, b_t|B \to A, \theta_t\right)$.
4. θ is indexed with t because after observing the t data point (a_t, b_t), θ might be updated.
5. Let $\sigma(\gamma)$ represents $P(A \to B)$, the probability that A causes B.
6. Define the log-likelihood of D_2 under a mixture of belief
$$\mathcal{L}(D_2) = \sigma(\gamma)\mathcal{L}_{A\to B} + (1 - \sigma(\gamma))\mathcal{L}_{B\to A}$$
7. Define the regret as the negative log-likelihood $\mathcal{R} = -\log \mathcal{L}(D_2) = -\log\left[\sigma(\gamma)\mathcal{L}_{A\to B} + (1 - \sigma(\gamma))\mathcal{L}_{B\to A}\right]$
8. Then:
$$\frac{\partial \mathcal{R}}{\partial \gamma} = \sigma(\gamma) - P\left(A \to B|D_2\right) = \sigma(\gamma) - \sigma(\gamma + \Delta)$$
where $\Delta = \log \mathcal{L}_{A\to B} - \log \mathcal{L}_{B\to A}$