Supplementary File for Incentivizing High Quality Crowdsourcing Information using Bayesian Inference and Reinforcement Learning

Anonymous Authors[1]

1. Basic Lemmas

Lemma 1. *If $x \sim \mathrm{Bin}(n,p)$, $\mathbb{E}t^x = (1-p+tp)^n$ holds for any $t>0$, where $\mathrm{Bin}(\cdot)$ is the binomial distribution.*

Proof.

$$t^x = e^{x\log t} = m_x(\log t) = \left(1-p+pe^{\log t}\right)^n \tag{1}$$

where $m_x(\cdot)$ denotes the moment generating function. $\square$

Lemma 2. *For given $n,m \geq 0$, if $0 \leq p \leq 1$, we can have*

$$\sum\nolimits_{x=0}^{n}\sum\nolimits_{w=0}^{m} C_n^x C_m^w p^{x+w}(1-p)^{y+z} \times B(x+z+1+t, y+w+1) = \int_0^1 [(2p-1)x+1-p]^n[(1-2p)x+p]^m x^t \mathrm{d}x$$

Proof. By the definition of the beta function (Olver, 2010),

$$B(x,y) = \int_0^{+\infty} u^{x-1}(1+u)^{-(x+y)}\mathrm{d}u \tag{2}$$

we can have

$$\begin{aligned}&\sum_{x,w} C_n^x C_m^w p^{x+w}(1-p)^{y+z} B(x+z+1+t, y+w+1)\\ &= \int_0^{+\infty} \mathbb{E}u^x \cdot \mathbb{E}u^z \cdot u^t \cdot (1+u)^{-(n+m+2+t)}\mathrm{d}u\end{aligned} \tag{3}$$

where we regard $x \sim \mathrm{Bin}(n,p)$ and $z \sim \mathrm{Bin}(m,1-p)$. Thus, according to Lemma 1, we can obtain

$$\begin{aligned}&\int_0^{+\infty} \mathbb{E}u^x \cdot \mathbb{E}u^z \cdot u^t \cdot (1+u)^{-(n+m+3)}\mathrm{d}u\\ &= \int_0^{+\infty} \frac{[1-p+up]^n \cdot [p+(1-p)u]^m \cdot u^t}{(1+u)^{n+m+2+t}}\mathrm{d}u.\end{aligned} \tag{4}$$

For the integral operation, substituting u with $v-1$ at first and then v with $(1-x)^{-1}$, we can conclude Lemma 2. $\square$

Lemma 3. $\sum_{n=0}^N C_N^n \cdot x^n = (1+x)^N$.

Lemma 4. $\sum_{n=0}^N C_N^n \cdot n \cdot x^n = N \cdot x \cdot (1+x)^{N-1}$.

Lemma 5. $\sum_{n=0}^N C_N^n \cdot n \cdot x^{N-n} = N \cdot (1+x)^{N-1}$.

Lemma 6. $\sum_{n=0}^N C_N^n \cdot n^2 \cdot x^n = Nx(1+Nx)(1+x)^{N-2}$.

Lemma 7. $\sum_{n=0}^N C_N^n \cdot n^2 \cdot x^{N-n} = N(x+N)(1+x)^{N-2}$.

Lemma 8. *If $0<x<1$, we can have*

$$\sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n \cdot x^n \geq \left(1-e^{-cN}\right)\cdot(1+x)^N$$

$$\sum_{n=\lfloor N/2 \rfloor+1}^{N} C_N^n \cdot x^{N-n} \geq \left(1-e^{-cN}\right)\cdot(1+x)^N.$$

where $c = 0.5(1-x)^2(1+x)^{-2}$.

Proof. To prove the lemmas above, we firstly define

$$F_t(x) = \sum_{n=0}^{N} C_N^n n^t x^n \tag{5}$$

Then, Lemma 3 can be obtained by expanding $(1+x)^N$. Lemma 4 can be proved as follows

$$\begin{aligned}F_1(x) &= \sum_{n=0}^{N} C_N^n (n+1)x^n - (1+x)^N\\ \sum_{n=0}^{N} C_N^n (n+1)x^n &= \frac{\mathrm{d}}{\mathrm{d}x}\left[xF_0(x)\right]\\ &= Nx(1+x)^{N-1} + (1+x)^N.\end{aligned} \tag{6}$$

Lemma 5 can be obtained as follows

$$\begin{aligned}\sum_{n=0}^{N} C_N^n n x^{N-n} &= x^N \sum_{n=0}^{N} C_N^n n \left(\frac{1}{x}\right)^n\\ &= x^N \cdot N \cdot \frac{1}{x} \cdot \left(1+\frac{1}{x}\right)^{N-1}.\end{aligned} \tag{7}$$

For Lemma 6, we can have

$$\begin{aligned}F_2(x) &= \sum_{n=0}^{N} C_N^n (n+2)(n+1)x^n - 3F_1(x) - 2F_0(x)\\ &= \left[x^2F_0(x)\right]' - 3F_1(x) - 2F_0(x)\end{aligned} \tag{8}$$

Thus, we can have

$$F_2(x) = Nx(1+Nx)(1+x)^{N-2} \tag{9}$$

which concludes Lemma 6. Then, Lemma 7 can be obtained by considering Equation 10.

$$\sum_{n=0}^{N} C_N^n n^2 x^{N-n} = x^N \sum_{n=0}^{N} C_N^n n^2 \left(\frac{1}{x}\right)^n. \tag{10}$$

For Lemma 8, we can have

$$\sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n x^n = (1+x)^N \sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n p^n (1-p)^{N-n} \tag{11}$$

where $p = x(1+x)^{-1}$. Let $X \sim \mathrm{Bin}(N,p)$, we can have

$$\sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n p^n (1-p)^{N-n} \geq 1 - P\left(X \geq N/2\right). \tag{12}$$

Since $x < 1$, $p < 0.5$ and $Np < N/2$. Considering Hoeffding's inequality, we can get

$$P\left(X \geq N/2\right) \leq \exp\left[-\frac{N(1-x)^2}{2(1+x)^2}\right] \tag{13}$$

which concludes the first inequality in Lemma 8. Similarly, for the second inequality, we can have

$$\sum_{n=K}^{N} C_N^n x^{N-n} = (1+x)^N \sum_{n=K}^{N} C_N^n (1-p)^n p^{N-n} \tag{14}$$

where $K = \lfloor N/2 \rfloor + 1$. Suppose $Y \sim \mathrm{Bin}(N, 1-p)$, we can have

$$\sum_{n=K}^{N} C_N^n (1-p)^n p^{N-n} \geq 1 - P\left(Y \leq N/2\right). \tag{15}$$

Considering Hoeffding's inequality, we can also get

$$P\left(Y \leq N/2\right) \leq \exp\left[-\frac{N(1-x)^2}{2(1+x)^2}\right] \tag{16}$$

which concludes the second inequality in Lemma 8. □

Lemma 9. *For any $x, y \geq 0$, we can have*

$$(1+x)^y \leq e^{xy}.$$

Proof. Firstly, we can know $(1+x)^y = e^{y\log(1+x)}$. Let $f(x) = x - \log(x)$. Then, we can have $f(0) = 0$ and $f'(x) \geq 0$. Thus, $x \geq \log(1+x)$ and we can conclude Lemma 9 by taking this inequality into the equality. □

Lemma 10.

$$g(x) = \frac{e^x}{e^x+1}$$

is a concave function when $x \in [0, +\infty)$.

Proof. $g'(x) = (2+t(x))^{-1}$, where $t(x) = e^x + e^{-x}$. $t'(x) = e^x - e^{-x} \geq 0$ when $x \in [0, +\infty)$. Thus, $g'(x)$ is monotonically decreasing when $x \in [0, +\infty)$, which concludes Lemma 10. □

Lemma 11. *For $x \in (-\infty, +\infty)$,*

$$h(x) = \frac{1}{e^{|x|}+1}$$

satisfies

$$h(x) < e^x \ \text{ and } \ h(x) < e^{-x}.$$

Proof. When $x \geq 0$, we can have

$$h(x) < \frac{1}{e^x} = e^{-x} \leq e^x. \tag{17}$$

When $x \leq 0$, we can have

$$h(x) = \frac{e^x}{e^x+1} < e^x \leq e^{-x}. \tag{18}$$

□

Lemma 12. *If $\lambda = p/(1-p)$ and $0.5 < p < 1$, then*

$$\sum\nolimits_{n=\lfloor N/2 \rfloor}^{N} C_N^n \lambda^{m-n} p^n (1-p)^m \leq [4p(1-p)]^{N/2}$$

$$\sum\nolimits_{n=0}^{\lfloor N/2 \rfloor} C_N^n \lambda^{n-m} p^n (1-p)^m \leq [4p(1-p)]^{N/2}$$

where $m = N - n$.

Proof. For the first inequality, we can have

$$\begin{aligned} &\sum_{n=\lfloor N/2 \rfloor}^{N} C_N^n \lambda^{m-n} p^n (1-p)^m \qquad (19) \\ &= \sum_{n=\lfloor N/2 \rfloor}^{N} C_N^n p^m (1-p)^n \leq \sum_{m=0}^{\lfloor N/2 \rfloor} C_N^m p^m (1-p)^n \end{aligned}$$

According to the inequality in (Arratia and Gordon, 1989), we can have

$$\sum_{m=0}^{\lfloor N/2 \rfloor} C_N^m p^m (1-p)^n \leq \exp(-ND) \tag{20}$$

where $D = -0.5\log(2p) - 0.5\log(2(p-1))$, which concludes the first inequality in Lemma 12.

For the second inequality, we can have

$$\begin{aligned} &\sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n \lambda^{n-m} p^n (1-p)^m \\ &= \frac{1}{[p(1-p)]^N} \sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n [p^3]^n [(1-p)^3]^m \qquad (21) \\ &= \frac{[p^3+(1-p)^3]^N}{[p(1-p)]^N} \sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n x^n (1-x)^m \end{aligned}$$

where $x = p^3/[p^3 + (1-p)^3]$. By using Equation 20, we can have

$$\begin{aligned}
&\sum_{n=0}^{\lfloor N/2 \rfloor} C_N^n \lambda^{n-m} p^n (1-p)^m \\
&\leq \frac{[p^3 + (1-p)^3]^N}{[p(1-p)]^N} [x(1-x)]^{N/2} \\
&= [4p(1-p)]^{N/2}
\end{aligned} \tag{22}$$

which concludes the second inequality of Lemma 12. □

2. H function analysis

Here, we present our analysis on the H function defined in the proof of Proposition 1. Firstly, we can have:

Lemma 13. $H(m, 0.5; M, t) = 2(t+1)^{-1}$.

Lemma 14. $H(m, p; M, t) = H(n, \bar{p}; M, t)$.

Lemma 15. *As a function of m, $H(m, p; M, t)$ is logarithmically convex.*

Proof. Lemma 13 can be proved by integrating $2x^t$ on $[0, 1]$. Lemma 14 can be proved by showing that $H(n, \bar{p}; M, t)$ has the same expression as $H(m, p; M, t)$. Thus, in the following proof, we focus on Lemma 15. Fixing p, M and t, we denote $\log(H)$ by $f(m)$. Then, we compute the first-order derivative as

$$H(m)f'(m) = 2^{M+1} \int_0^1 \lambda u^n (1-u)^m x^t \mathrm{d}x \tag{23}$$

where $u = (2p-1)x + 1 - p$ and $\lambda = \log(1-u) - \log(u)$. Furthermore, we can solve the second-order derivative as

$$\begin{aligned}
&2^{-2(M+1)} H^2(m) f''(m) = \\
&\int_0^1 g^2(x)\mathrm{d}x \int_0^1 h^2(x)\mathrm{d}x - \left(\int_0^1 g(x)h(x)\mathrm{d}x\right)^2
\end{aligned} \tag{24}$$

where the functions $g, h : (0, 1) \to \mathbb{R}$ are defined by

$$g = \lambda\sqrt{u^n(1-u)^m}\ ,\ h = \sqrt{u^n(1-u)^m}. \tag{25}$$

By the Cauchy-Schwarz inequality,

$$\int_0^1 g^2(x)\mathrm{d}x \int_0^1 h^2(x)\mathrm{d}x \geq \left(\int_0^1 g(x)h(x)\mathrm{d}x\right)^2 \tag{26}$$

we can know that $f''(m) \geq 0$ always holds, which concludes that f is convex and H is logarithmically convex. □

Then, for the case that $t = 1$ and $M \gg 1$, we can further derive the following three lemmas for $H(m, p; M, 1)$:

Lemma 16. *The ratio between two ends satisfies*

$$\log \frac{H(0, p; M, 1)}{H(M, p; M, 1)} \approx \begin{cases} \log(M) + \epsilon(p) & p > 0.5 \\ 0 & p = 0.5 \\ -\log(M) - \epsilon(\bar{p}) & p < 0.5 \end{cases}$$

where $\epsilon(p) = \log(2p-1) - \log(p)$ and $\epsilon(p) = 0$ if $p = 0.5$.

Lemma 17. *The lower bound can be calculated as*

$$\log H(m, p) \gtrsim \begin{cases} H(0, p) - \underline{k}m & 2m \leq M \\ H(M, p) - \underline{k}n & 2m > M \end{cases}$$

where $\underline{k} = \log\left(\max\left\{p/(\bar{p} + M^{-1}), \bar{p}/(p + M^{-1})\right\}\right)$.

Lemma 18. *The upper bound can be calculated as*

$$\log H(m, p) \lesssim \begin{cases} H(0, p) - \overline{k}m & 2m \leq M \\ H(M, p) - \overline{k}n & 2m > M \end{cases}$$

where $n = M - m$ and $\overline{k} = 2\log(2 \cdot \max\{p, \bar{p}\})$.

Proof. By Lemma 13, $\log H(m, 0.5; M, 1) \equiv 0$, which proves the above three lemmas for the case that $p = 0.5$. Considering the symmetry ensured by Lemma 14, we thus focus on the case that $p > 0.5$ in the following proof and transform $H(m, p)$ into the following formulation

$$H(m, p) = \omega(p) \cdot \int_{\bar{p}}^{p} x^n (1-x)^m (x - 1 + p)\mathrm{d}x \tag{27}$$

where $\omega(p) = 2^{M+1}/(2p-1)^2$. Then, we can solve $H(0, p)$ and $H(M, p)$ as

$$\begin{aligned}
H(0, p) &= \omega(p) \int_{\bar{p}}^{p} x^M (x - \bar{p})\mathrm{d}x \\
&= \frac{(2p)^{M+1}}{(2p-1)(M+1)} - O\left(\frac{(2p)^{M+1}}{M^2}\right)
\end{aligned} \tag{28}$$

$$\begin{aligned}
&H(M, p) = \omega(p) \int_{\bar{p}}^{p} (1-x)^M (x - \bar{p})\mathrm{d}x \\
&= \frac{p(2p)^{M+1}}{(2p-1)^2(M+1)(M+2)} - O\left(\frac{(2\bar{p})^{M+1}}{M+2}\right).
\end{aligned} \tag{29}$$

Using the Taylor expansion of function $\log(x)$, we can calculate the ratio in Lemma 16 as

$$\log \frac{H(0, p)}{H(M, p)} = \log(M) + \log \frac{2p-1}{p} + O\left(\frac{1}{M}\right) \tag{30}$$

which concludes Lemma 16 when $M \gg 1$.

Furthermore, we can solve $H(1, p)$ as

$$\begin{aligned}
H(1, p) &= \omega(p) \int_{\bar{p}}^{p} x^{M-1} (x - \bar{p})\mathrm{d}x - H(0, p) \\
&= \frac{(2\bar{p} + M^{-1})(2p)^M}{(2p-1)(M+1)} - O\left(\frac{(2p)^{M+1}}{M^2}\right)
\end{aligned} \tag{31}$$

The value ratio between $m = 0$ and $m = 1$ then satisfies

$$\log \frac{H(1, p)}{H(0, p)} = \log \frac{p}{\bar{p} + M^{-1}} + O\left(\frac{1}{M}\right). \tag{32}$$

By Rolle's theorem, there exists a $c \in [m, m+1]$ satisfying

$$\log H(1,p) - \log H(0,p) = f'(c) \tag{33}$$

where $f(m) = \log H(m,p)$. Meanwhile, Lemma 15 ensures that $f''(m) \geq 0$ always holds. Thus, we can have

$$\log H(m+1,p) - \log H(m,p) \geq \log \frac{H(1,0)}{H(0,p)} \tag{34}$$

which concludes the first case of Lemma 17. Similarly, we compute the ratio between $m = M-1$ and M as

$$\log \frac{H(M,p)}{H(M-1,p)} = \log \frac{p}{\bar{p} + M^{-1}} + O\left(\frac{1}{M}\right). \tag{35}$$

Meanwhile, Rolle's theorem and Lemma 15 ensure that

$$\log H(m,p) - \log H(m-1,p) \leq \log \frac{H(M,0)}{H(M-1,p)} \tag{36}$$

which concludes the second case of Lemma 17.

Lastly, we focus on the upper bound described by Lemma 18. According to the inequality of arithmetic and geometric means, $x(1-x) \leq 2^{-2}$ holds for any $x \in [0,1]$. Thus, when $2m \leq M$ (i.e. $n \geq m$), we can have

$$H(m,p) \leq 2^{-2m}\omega(p) \cdot \int_{\bar{p}}^{p} x^{n-m}(x-1+p)\mathrm{d}x \tag{37}$$

where the equality only holds when $m = 0$.

$$\int_{\bar{p}}^{p} x^{n-m}(x-1+p)\mathrm{d}x = \frac{(2p-1)p^{\delta}}{\delta} + \frac{\Delta}{\delta(\delta+1)} \tag{38}$$

where $\delta = n - m + 1$ and $\Delta = \bar{p}^{\delta+1} - p^{\delta+1} < 0$. Hence,

$$\log \frac{H(m,p)}{H(0,p)} \leq -2m[\log(2p) - \varepsilon(m)] + O\left(\frac{1}{M}\right) \tag{39}$$

where $\varepsilon(m) = -(2m)^{-1}[\log(n-m+1) - \log(M+1)]$. Since $\log(x)$ is a concave function, we can know that

$$\varepsilon(m) \leq (M)^{-1}\log(M+1) = O\left(M^{-1}\right) \tag{40}$$

which concludes the first case in Lemma 18. Similarly, for $2m > M$ (i.e. $n < m$), we can have

$$\log \frac{H(m,p)}{H(M,p)} \leq -2n[\log(2p) - \hat{\varepsilon}(n)] + O\left(\frac{1}{M}\right) \tag{41}$$

where $\hat{\varepsilon}(n) \leq O(M^{-1})$. Thereby, we can conclude the second case of Lemma 18. Note that the case where $p < 0.5$ can be derived by using Lemma 14. $\square$

For the case that $t = 0$ and $M \gg 1$, using the same method as the above proof, we can derive the same lower and upper bounds as Lemmas 18 and 17. On the other hand, for $t = 0$, Lemma 16 does not hold and we can have

Lemma 19. *$H(m,p;M,0) = H(n,p;M,0)$*

Proof. When $t = 0$,

$$H(m,p) = 2^{M+1}(2p-1)^{-1}\int_{\bar{p}}^{p} x^{n}(1-x)^{m}\mathrm{d}x. \tag{42}$$

Then, substituting x as $1 - v$ concludes Lemma 19. $\square$

References

R. Arratia and L. Gordon. Tutorial on large deviations for the binomial distribution. *Bulletin of Mathematical Biology*, 51(1):125–131, Jan 1989.

Frank W. J. Olver. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.