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

Anonymous Authors[1]

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.

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

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

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),

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

we can have

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

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

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

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

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

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

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

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

Lemma 5 can be obtained as follows

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

For Lemma 6, we can have

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

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. □

References

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