PROPERTIES OF DISCRETE FISHER INFORMATION:
CRAMÉR-RAO-TYPE AND LOG-SOBOLEV-TYPE INEQUALITIES

TOMOHIRO NISHIYAMA

Abstract. The Fisher information have connections with the standard deviation and the Shannon differential entropy through the Cramér-Rao bound and the log-Sobolev inequality. These inequalities hold for continuous distributions. In this paper, we introduce the Fisher information for discrete distributions (DFI) and show that the DFI satisfies the Cramér-Rao-type bound and the log-Sobolev-type inequality.

Keywords: Fisher information, Cramér-Rao bound, log-Sobolev inequality, discrete distribution.

1. Introduction

The Fisher information [4] is defined for continuous probability density functions and plays an important role in statistics, information theory and physics. The Fisher information has connections with the standard deviation and the Shannon differential entropy [9] through the Cramér-Rao bound [3, 7] and the log-Sobolev inequality (the Stam’s inequality) [5, 10].

Recently, Moreno, Yáñez and Dehesa introduced different discrete forms of the Fisher information [8]. Furthermore, several types of discrete log-Sobolev inequalities have been studied [1, 2, 6].

In this paper, we focus on one of the discrete forms of the Fisher information shown in [8]. We aim to study properties of the Fisher information for discrete distributions (DFI) and show that DFI satisfies three new inequalities which are corresponds to the Cramér-Rao bound and the log-Sobolev inequality.

2. Fisher information

2.1. Fisher information for continuous distributions. The Fisher information for differentiable probability density function \( F(x) \) on \( \mathbb{R} \) is defined as follows.

\[
I(F) \overset{\text{def}}{=} \int_{\mathbb{R}} \frac{F'(x)^2}{F(x)} \, dx,
\]

where \( F'(x) \) denotes a derivative of \( F(x) \) with respect to \( x \). By putting \( F(x) = f(x)^2 \), we can write (1) as

\[
I(f) = 4 \int_{\mathbb{R}} f'(x)^2 \, dx.
\]

The Cramér-Rao bound is

\[
\sigma^2 I(f) \geq 1,
\]

where \( \sigma \) is the standard deviation for the probability density function \( F(x) \). The log-Sobolev inequality is

\[
\frac{1}{2} \log \left( \frac{1}{2\pi e} I(f) \right) \geq \text{Ent}(f),
\]

(4)
where $\text{Ent}(f) \overset{\text{def}}{=} \int_{\mathbb{R}} f(x)^2 \log f(x)^2 \, dx = -h(F)$ and $h(F) = -\int_{\mathbb{R}} F(x) \log F(x) \, dx$ is the Shannon differential entropy.

2.2. Fisher information for discrete distributions. First, we show some notations in this paper.

**Notations**
- Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.
- Let $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$.
- Let $\Omega$ be a set of functions $f : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ which satisfy $\sum_{i=0}^{\infty} f(i)^2 = 1$ and $\lim_{i \to \infty} f(i) = 0$. $f(i)^2$ is a probability mass function.
- Let $\mathbb{E}[A(X)] \overset{\text{def}}{=} \sum_{i=0}^{\infty} A(i)f(i)^2$ be an expected value.
- Let $\text{Ent}(f) \overset{\text{def}}{=} \sum_{i=0}^{\infty} f(i)^2 \log f(i)^2$ be an entropy.
- Let $\|f\|_{\infty} \overset{\text{def}}{=} \max_{i} f(i)$.
- Let $D$ be a difference operator defined as $Df(i) \overset{\text{def}}{=} f(i+1) - f(i)$.

Next, we introduce the discrete Fisher information.

**Definition 1.** (Discrete Fisher information) Let $f \in \Omega$. We define the discrete Fisher information (DFI) as

$$I(f) \overset{\text{def}}{=} 4 \sum_{i=0}^{\infty} |Df(i)|^2 = 4 \sum_{i=0}^{\infty} |f(i+1) - f(i)|^2.$$  (5)

This is a natural discretization of (2) and the DFI can be also written as

$$I(f) = 4\{2 - f(0) - 2R_{ff}(1)\},$$ (6)

where $R_{ff}(t) \overset{\text{def}}{=} \sum_{i=0}^{\infty} f(i)f(i+t)$ is an autocorrelation and we use $\sum_{i=0}^{\infty} f(i)^2 = 1$.

3. Main Results

3.1. New inequalities for DFI.

**Theorem 1.** (Cramér-Rao-type bound) Let $f \in \Omega$ and let $I(f)$ be the DFI.

Then,

$$\mathbb{E}[X^2]I(f) \geq |1 - f(0)|^2,$$ (7)

with equality if and only if $f(i) = \delta_{i0}$. $\delta_{ij}$ denotes the Kronecker delta.

This is an expansion of (3) for the DFI

**Theorem 2.** (Inequality for the maximum of probability) Let $f \in \Omega$ and let $I(f)$ be the DFI.

Then,

$$I(f) > \|f\|_{\infty}^4.$$ (8)

Furthermore, this inequality is “sharp” in the sense that $\alpha = 1$ is the optimal constant for an inequality $\alpha I(f) > \|f\|_{\infty}^4$ which holds for all $f \in \Omega$.

**Proposition 1.** (Log-Sobolev-type inequality) Let $f \in \Omega$ and let $I(f)$ be the DFI.

Then,

$$\frac{1}{2} \log I(f) > \text{Ent}(f).$$ (9)

If there exists the optimal constant for an inequality $\frac{1}{2} \log(\beta I(f)) > \text{Ent}(f)$ which holds for all $f \in \Omega$, $\beta$ must be $e^{-2} \leq \beta \leq 1$.

This is an expansion of (4) for the DFI.
3.2. Proofs of main results. We show proofs of the inequalities in subsection 3.1.

Proof of Theorem 1

We consider a quantity as follows.

\[ V = -\sum_{i=0}^{\infty} i D[f(i)^2] = -\sum_{i=0}^{\infty} i \{f(i+1)^2 - f(i)^2\} \]  

(10)

From \(\lim_{i \to \infty} f(i) = 0\) and \(\sum_{i=0}^{\infty} if(i+1)^2 = \sum_{i=1}^{\infty} (i-1)f(i)^2\), we get

\[ V = \sum_{i=1}^{\infty} f(i)^2 = 1 - f(0)^2. \]  

(11)

Furthermore, we get

\[ V = -\sum_{i=0}^{\infty} i D[f(i)^2] = \sum_{i=0}^{\infty} i \{f(i+1) + f(i)\} Df(i). \]  

(12)

Applying the Cauchy-Schwarz inequality to this equation yields

\[ V^2 \leq \frac{1}{4} I(f) \sum_{i=0}^{\infty} i^2 |f(i+1) + f(i)|^2. \]  

(13)

By using \((x + y)^2 \leq 2(x^2 + y^2)\),

we get

\[ \sum_{i=0}^{\infty} i^2 |f(i+1) + f(i)|^2 \leq 2 \sum_{i=0}^{\infty} i^2 \{f(i+1)^2 + f(i)^2\} \]  

(15)

\[ \leq 2 \sum_{i=0}^{\infty} \{(i+1)^2 f(i+1)^2 + i^2 f(i)^2\} = 4E[X^2]. \]

By substituting this inequality into (13) and combining with (11), we obtain (7).

Next, we show the equality condition. If equality holds in (15), \(f(i)\) must be 0 for all \(i \geq 1\). Furthermore, since \(f(i)\) satisfies \(\sum_{i=0}^{\infty} f(i)^2 = 1\), \(f(i)\) must be \(\delta_0\). By confirming the equality holds for \(\delta_0\), the result follows.

Proof of Theorem 2

First, we prove the first half of the theorem. Let \(m\) be an index which satisfies \(f(m) = \|f\|_{\infty}\).

We consider a quantity as follows.

\[ V = -\sum_{i=m}^{\infty} D[f(i)^2] = f(m)^2 = \|f\|_{\infty}^2. \]  

(16)

By using \(D[f(i)^2] = \{f(i+1) + f(i)\} Df(i)\), we get

\[ V = -\sum_{i=m}^{\infty} \{f(i+1) + f(i)\} Df(i). \]  

(17)

Applying the Cauchy-Schwarz inequality to this equation yields

\[ V^2 \leq \sum_{i=m}^{\infty} [Df(i)]^2 \sum_{i=m}^{\infty} |f(i+1) + f(i)|^2. \]  

(18)
By using \((x + y)^2 \leq 2(x^2 + y^2)\), we get
\[
\sum_{i=m}^{\infty} |f(i + 1) + f(i)|^2 < 2 \sum_{i=m}^{\infty} \{f(i + 1)^2 + f(i)^2\}. \tag{19}
\]

Since \((x + y)^2 = 2(x^2 + y^2)\) holds if and only if \(x = y\), if \(\sum_{i=m}^{\infty} |f(i + 1) + f(i)|^2 = 2 \sum_{i=m}^{\infty} \{f(i + 1)^2 + f(i)^2\}\) holds, \(f(i)\) must be a constant for \(i \geq m\). By combining with \(\sum_{i=m}^{\infty} f(i)^2 \leq 1\), the constant must be 0 and \(\max_i f(i) = f(m) = 0\) holds. However, \(\max_i f(i) = 0\) is inconsistent with \(\sum_{i=0}^{\infty} f(i)^2 = 1\). Hence, the equality doesn’t hold in (19).

By using \(\sum_{i=0}^{\infty} f(i)^2 = 1\) for (19), we get
\[
\sum_{i=m}^{\infty} |f(i + 1) + f(i)|^2 < 4. \tag{20}
\]

Substituting this inequality into (18) and combining with (16) yields
\[
\|f\|_{\infty}^4 < 4 \sum_{i=m}^{\infty} |Df(i)|^2. \tag{21}
\]

Since \(\sum_{i=m}^{\infty} |Df(i)|^2 \leq \sum_{i=0}^{\infty} |Df(i)|^2\), the result follows.

Next, we prove the latter half of the theorem. If an inequality \(\alpha I(f) > \|f\|_{\infty}^4\) holds, \(\alpha\) must satisfy \(\alpha > \frac{\|f\|_{\infty}^4}{\|f\|_{\infty}^4}\).

For the geometric distribution \(f(i)^2 = p(1 - p)^i\), the DFI and the maximum of probability are
\[
I(f) = 4(1 - \sqrt{1 - p})^2. \tag{22}
\]
\[
\|f\|_{\infty}^2 = p. \tag{23}
\]
For \(p \sim 0\), from \(\sqrt{1 - p} = 1 - \frac{p}{2} + O(p^2)\), we get
\[
I(f) = p^2 + O(p^3). \tag{24}
\]

Hence, \(\lim_{p \to 0} \frac{\|f\|_{\infty}^4}{I(f)} = 1\) and \(\alpha\) must be \(\alpha \geq 1\). Since the inequality (23) is the case for \(\alpha = 1\), the result follows.

**Proof of Proposition 1**

First, we prove the first half of the proposition. Since \(\|f\|_{\infty} \geq f(i)\) and \(\sum_{i=0}^{\infty} f(i)^2 = 1\), we get
\[
\text{Ent}(f) = \sum_{i=0}^{\infty} f(i)^2 \log f(i)^2 \leq \log \|f\|_{\infty}^2. \tag{24}
\]

From Theorem 2, the result follows.

Next, we prove the latter half of the theorem. If an inequality \(\frac{1}{2} \log(\beta I(f)) > \text{Ent}(f)\) holds, \(\beta\) must satisfy \(\beta > \frac{\exp(2\text{Ent}(f))}{I(f)}\).

For the geometric distribution \(f(i)^2 = p(1 - p)^i\), the entropy is \(\text{Ent}(f) = p \log p + (1 - p) \log(1 - p)\). Then, we get
\[
\exp(2\text{Ent}(f)) = p^2(1 - p)^{\frac{1}{1-p}}. \tag{25}
\]
Combining with \(\lim_{p \to 0} (1 - p)^{\frac{1}{1-p}} = e^{-2}\) and (23) yields
\[
\lim_{p \to 0} \frac{\exp(2\text{Ent}(f))}{I(f)} = e^{-2}. \tag{26}
\]

Hence, \(\beta\) must be \(\beta \geq e^{-2}\). Since the inequality (20) is the case for \(\beta = 1\), the result follows.
4. Examples

We show some examples of the DFI and other quantities for discrete distributions.

1. Discrete uniform distribution

\[ f(i) = \begin{cases} \frac{1}{\sqrt{N}} & (0 \leq i \leq N - 1) \\ 0 & (N \leq i) \end{cases} \]

- DFI: \( I(f) = \frac{4}{N} \).
- \( \mathbb{E}[X^2] = \frac{(N-1)(2N-1)}{6} \).
- Maximum of probability: \( \|f\|_\infty^2 = \frac{1}{N} \).
- Entropy: \( \text{Ent}(f) = -\log N \).

When \( N = 1 \), the equality holds for (7) as mentioned in Section 3.

2. Bernoulli distribution

For \( 0 \leq p \leq 1 \),

\[ f(i) = \begin{cases} \sqrt{1-p} & (i = 0) \\ \sqrt{p} & (i = 1) \\ 0 & (\text{otherwise}) \end{cases} \]

- DFI: \( I(f) = 4\{1 + p - 2\sqrt{p(1-p)}\} \).
- \( \mathbb{E}[X^2] = p \).
- Maximum of probability: \( \|f\|_\infty = \max(p, 1-p) \).
- Entropy: \( \text{Ent}(f) = p\log p + (1-p)\log(1-p) \).

3. Geometric distribution

For \( 0 < p \leq 1 \),

\[ f(i) = \sqrt{p(1-p)^i} \]

- DFI: \( I(f) = 4(1 - \sqrt{1-p})^2 \).
- \( \mathbb{E}[X^2] = \frac{(2-p)(1-p)}{p^2} \).
- Maximum of probability: \( \|f\|_\infty^2 = p \).
- Entropy: \( \text{Ent}(f) = \frac{p\log p + (1-p)\log(1-p)}{p} \).

5. Conclusion

We have introduced the discrete Fisher information (DFI) and have shown three new inequalities for the DFI. They are the Cramér-Rao-type bound, the inequality for the maximum of probability and the log-Sobolev-type inequality. We have also shown two inequalities other than the log-Sobolev-type inequality are “sharp”.

We hope we will find tighter bound for the log-Sobolev-type inequality in the future.

References

[1] Sergey G Bobkov and Michel Ledoux. On modified logarithmic sobolev inequalities for bernoulli and poisson measures. *Journal of functional analysis*, 156(2):347–365, 1998.
[2] Sergey G Bobkov and Prasad Tetali. Modified logarithmic sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006.
[3] Harald Cramér. *Mathematical methods of statistics (PMS-9)*, volume 9. Princeton university press, 2016.
[4] Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922.
[5] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.
[6] Oliver Johnson et al. A discrete log-sobolev inequality under a bakry–émery type condition. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 53, pages 1952–1970. Institut Henri Poincaré, 2017.

[7] C Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in statistics*, pages 235–247. Springer, 1992.

[8] P Sánchez-Moreno, RJ Yáñez, and JS Dehesa. Discrete densities and fisher information. In *Proceedings of the 14th International Conference on Difference Equations and Applications. Difference Equations and Applications*. Istanbul, Turkey: Bahçeşehir University Press, pages 291–298, 2009.

[9] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[10] Aart J Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2(2):101–112, 1959.