Contents

1 A Model Problem **1**
- 1.1 The auto-correlation function . . . 1
- 1.2 Gradient . . . 2

2 Preturbation Method **2**
- 2.1 Computation of $\nabla_\delta I[f]$ . . . 3

3 KKT conditions **3**

4 Piecewise constant functions **5**
- 4.1 Two peaks . . . 6
- 4.2 Multiple peaks . . . 7

5 Maximum of the mutual information **8**
- 5.1 Maximum of the negative entropy . . . 8
- 5.2 Upper bound of the entropy of auto-correlation . . . 9
- 5.3 Main Theorem . . . 9

A Appendix: Compuation of the gradients **10**
- A.1 Computation of $\nabla_f I$ . . . 10
- A.2 Computation of $\nabla_\delta I$ . . . 11
- A.3 Other conclusions . . . 12

1 A Model Problem

Consider a simplified version of the continuous limit case. Assume θ is on the circle $[0,1)$ with uniform distribution $p(\theta)=1$. X, Y are conditionally independent of θ:

$$p(x,y|\theta) = p(x|\theta)p(y|\theta).$$

Also, assume that X, Y comes from the same population and satisfy translation invariance:

$$p(x|\theta) = f(x-\theta),\ p(y|\theta) = f(y-\theta)$$

and f satisfies the constraint:

$$\int_0^1 f(\theta)d\theta = 1.$$

Then the mutual information of (X,Y) and Θ is:

$$\begin{aligned} I(X,Y;\Theta) &= D_{KL}(p(x,y,\theta)\|p(x,y)p(\theta)) \\ &= \int_0^1 dx \int_0^1 dy f(x)f(y) \ln\left(\frac{f(x)f(y)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)}\right) \end{aligned} \tag{1}$$

where we have applied the prodicity of f on $[0,1)$.

1.1 The auto-correlation function

Define

$$A(u) := \int_0^1 d\theta f(\theta) f(u+\theta)$$

Properties of $A(\cdot)$ on circle $[0,1)$:

- $A(u) = A(-u)$ since $\int_0^1 d\theta f(\theta) f(u+\theta) = \int_0^1 d\theta' f(\theta'-u) f(\theta')$;
- $A(u) = A(1-u)$, i.e. $A(u)$ is symmetric w.r.t. $u = \frac{1}{2}$;
- $A(x-y) = \int_0^1 d\theta f(\theta) f(x-y+\theta) = \int_0^1 d\theta f(x-\theta) f(y-\theta)$;
- If $\bar{f} = 1$ then $\int A = 1$, $f_- \leq A \leq f_+$;
- A is continuous if f only has a finite number of discontinuities.

Thus the expression for mutual information can be simplified as:

$$\begin{aligned} I(X,Y;\Theta) &= -2H[f] - \int dx \int dy f(x) f(y) \ln (A(x-y)) \\ &= -2H[f] - \int d\theta \int dx \int dy f(x-\theta) f(y-\theta) \ln (A(x-y)) \\ &= -2H[f] + H[A] \end{aligned} \tag{2}$$

where H is the continuous entropy: $H[f] = -\int_0^1 f(x) \ln f(x) dx$.

1.2 Gradient

$$\nabla_f I(x) = 2\left(\ln f(x) - \int_0^1 dy f(y) \ln (A(y-x))\right) \tag{3}$$

Detailed computation in the Appendix A.1.

2 Preturbation Method

Let $\delta(x)$ be a smooth, 1-Lipschitz function which satisfies $\delta(0) = \delta(1) = 0$. When ϵ is small enough, $x \mapsto x + \epsilon\delta(x)$ is a one-to-one mapping. Also, we require that $\int_0^1 f(x + \epsilon\delta(x)) = 1$, which implies

$$\int_0^1 f(x)\delta'(x) = 0. \tag{4}$$

Assume f is **piecewise** C^1 on the circle $[0,1)$.

Proposition 1. *If f is piecewise C^1, for any $\delta(x) \in C^\infty(0,1)$,*

$$\int \delta'(\eta) f(\eta) d\eta = -\left[\int \delta(\eta) f_c'(\eta) + \sum_{i \in D} \triangle_i f \cdot \delta(x_i)\right] \tag{5}$$

i.e. the distributional derivative of f is

$$f'(x) = f_c'(x) + \sum_{i \in D} \triangle_i f \cdot \delta_{x_i}(x) \tag{6}$$

where $f_c'(x)$ is the point-wise derivative of f (zero at discontinuities), D is the set of discontinuities of f, $\triangle_i f := f(x_i^+) - f(x_i^-)$, where $x_i \in D$.

Proof. Assume f is discontinuous only at $a \in (0,1)$:

$$\begin{aligned} \int_0^1 \delta'(\eta) f(\eta) d\eta &= \int_0^a \delta'(\eta) f(\eta) d\eta + \int_a^1 \delta'(\eta) f(\eta) d\eta \\ &= \delta(a) f(a^-) - \delta(0) f(0) - \int_0^a \delta(\eta) f'(\eta) d\eta + \delta(1) f(1) - \delta(a) f(a^+) - \int_a^1 \delta(\eta) f'(\eta) d\eta \\ &= -\left[\int_0^1 \delta(\eta) f'(\eta) d\eta + \left(f(a^+) - f(a^-)\right) \delta(a)\right] \end{aligned}$$

Similarily, when we have more than one discontinuities $x_i \in D$,

$$\int_0^1 \delta'(\eta) f(\eta) d\eta = -\left[\int \delta(\eta) f_c'(\eta) d\eta + \sum_{i\in D} \triangle_i f \cdot \delta(x_i)\right]$$

□

2.1 Computation of $\nabla_\delta I[f]$

By writting $I[f(x+\epsilon\delta(x))] - I[f]$, we obtain

$$\begin{aligned}\nabla_\delta I[f] &= 2\left[f_c'(x)\left(\log f_c(x)+1\right) - f_c'(x)\int dy f(y)\log A(y-x) - f(x)\int dy f(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)}\right] \\ &\quad +2\sum_{i\in D}\left[\triangle_i(f\log f) - \triangle_i f\cdot\int dy f(y)\log A(y-x_i)\right]\delta_{x_i}(x) \qquad (7)\end{aligned}$$

Notice that $A(y-x) = \int_0^1 d\theta f(x-\theta)f(y-\theta)$ is continuous, and the intergral $\int_0^1 d\theta f'(x-\theta)f(y-\theta)$ is well-defined even though f' may enclude dirac delta terms. The detailed computations are in the Appendix A.2.

Proposition 2.

$$\int \nabla_\delta I[f](x)dx = 0$$

Proof. From integration by parts,

$$\begin{aligned}\int dx f(x)\int dy f(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)} &= \int f(y)dy\int f(x)\frac{\partial}{\partial x}\log\left(A(y-x)\right)dx \\ &= -\int f(y)dy\int f'(x)\log\left(A(y-x)\right)dx \\ &= -\int f(y)dy\int f_c'(x)\log\left(A(y-x)\right)dx - \sum_{i\in D}\triangle_i f\int f(y)\log\left(A(y-x_i)\right)dy\end{aligned}$$

Thus

$$\int dx f(x)\int dy f(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)} + \int f(y)dy\int f_c'(x)\log\left(A(y-x)\right)dx + \sum_{i\in D}\triangle_i f\int f(y)\log\left(A(y-x_i)\right)dy = 0,$$

$$\int \nabla_\delta I[f](x)dx = \int dx\left(f_c'(x)\left(\log f_c(x)+1\right) + \sum_{i\in D}\triangle_i(f\log f)\delta_{x_i}(x)\right) = \int (f\log f)'(x)dx = 0.$$

□

3 KKT conditions

$$\nabla_f I(x) = \mu + \alpha_+(x) - \alpha_-(x) \qquad (8)$$

$$\nabla_\delta I(x) = \nu\left[f_c'(x) + \sum_{i\in D}\triangle_i f\cdot\delta_{x_i}(x)\right] \qquad (9)$$

where the 2nd one comes from minimizing $I[f(x+\epsilon\delta(x))]$ w.r.t δ, where $\int f(x)\delta'(x)dx = -\int\delta(x)dx\left(f_c'(x) + \sum_{i\in D}\triangle_i f\delta_{x_i}(x)\right) = 0$.

Proposition 3. *If $f(x)$ is continous at x, then $\nabla_f I(x)$, $\alpha_+(x)$, $\alpha_-(x)$ are continuous at x.*

Proof. Directly observe from $\nabla_f I(x) = 2\left(\log f(x) - \int dy f(y) \log A(y-x)\right)$ and $\nabla_f I(x) = \mu + \alpha_+(x) - \alpha_-(x)$. Note that $\alpha_+(x)$ and $\alpha_-(x)$ cannot be nonzero at the same time. $\square$

Take (distributional) derivative of the first condition with respect to x:

$$
\begin{aligned}
2\left(\log f(x) - \int_0^1 dy f(y) \log\left(A(y-x)\right)\right) &= \mu + \alpha_+(x) - \alpha_-(x) \\
\frac{f_c'(x)}{f_c(x)} + \sum_{i \in D} \triangle_i (\log f) \delta_{x_i}(x) - \int_0^1 dy f(y) \frac{\int_0^1 f'(x-\theta) f(y-\theta) d\theta}{\int_0^1 f(x-\theta) f(y-\theta) d\theta} &= \frac{\alpha_+'(x)}{2} - \frac{\alpha_-'(x)}{2} + \frac{\sum_{i \in D} (\triangle_i \alpha_+ - \triangle_i \alpha_-) \delta_{x_i}(x)}{2} \\
\frac{f'(x)}{f(x)} - \int_0^1 dy f(y) \frac{\int_0^1 f'(x-\theta) f(y-\theta) d\theta}{\int_0^1 f(x-\theta) f(y-\theta) d\theta} &= \frac{\alpha_+'(x)}{2} - \frac{\alpha_-'(x)}{2} \text{ for } x \notin D \qquad (10) \\
\triangle_i (\log f) &= \frac{\triangle_i \alpha_+}{2} - \frac{\triangle_i \alpha_-}{2} \text{ for } x_i \in D \qquad (11)
\end{aligned}
$$

Proposition 4. *If $f(x)$ is C^1 on $[0,1)$ and does not reach the upper/lower bounds (f_+, f_-), then $f(x)$ is constant.*

Proof. Assume $\frac{f'(x)}{f(x)}$ is not constant, set $x = \arg\max \frac{f'(x)}{f(x)}$. Then there exists an interval where $\frac{f'(x-\theta)}{f(x-\theta)} < \frac{f'(x)}{f(x)}$. Averaging $f'(x-\theta)$ by $f(y-\theta)$,

$$
\begin{aligned}
\int_0^1 f'(x-\theta) f(y-\theta) d\theta &< \int_0^1 \frac{f'(x)}{f(x)} f(x-\theta) f(y-\theta) d\theta \\
\frac{\int_0^1 f'(x-\theta) f(y-\theta) d\theta}{\int_0^1 f(x-\theta) f(y-\theta) d\theta} &< \frac{f'(x)}{f(x)} \\
\int_0^1 dy f(y) \frac{\int_0^1 f'(x-\theta) f(y-\theta) d\theta}{\int_0^1 f(x-\theta) f(y-\theta) d\theta} &< \frac{f'(x)}{f(x)}
\end{aligned}
$$

Also, since f does not take f_+ or f_-, $\alpha_+(x) = \alpha_-(x) = 0$ for all x. From equation (10),

$$
\frac{f'(x)}{f(x)} = \int_0^1 dy f(y) \frac{\int_0^1 f'(x-\theta) f(y-\theta) d\theta}{\int_0^1 f(x-\theta) f(y-\theta) d\theta}
$$

which contradicts the inequality above. Therefore $\frac{f'(x)}{f(x)}$ is constant, $\log f(x) = kx + c$, $f(x) = ce^{kx}$. However since f is continuous and periodic on [0,1), $k = 0$, $f(x)$ is constant. From $\int_0^1 f(y) dy = 1$, $f(x) \equiv 1$. $\square$

Proposition 5. *If $f(x)$ has a jump discontinuity at x, then either one of $f(x^+)$ and $f(x^-)$ must take f_+ or f_-. That is, every jump of f takes the function value either to or from the upper/lower bounds.*

Proof. Assume that at $x_i \in D$, neither $f(x_i^+)$ nor $f(x_i^-)$ reach the bounds. Then $\alpha_+(x_i^+) = \alpha_+(x_i^-) = \alpha_-(x_i^+) = \alpha_-(x_i^-) = 0$, which contradicts the equation (11):

$$
\frac{\triangle_i \alpha_+}{2} - \frac{\triangle_i \alpha_-}{2} = \log f(x_i^+) - \log f(x_i^-) \neq 0
$$

$\square$

4 Piecewise constant functions

We focus on piecewise constant f which only takes f_+ and f_-.

Notice that for given f_- , f_+ and $\int f(x)dx = 1$, the proportion of f taking f_+ and f_- are fixed. Then $H[f] = -\int f(x)\log f(x)dx$ is fixed. Since $I[f] = -2H[f] + H[A]$, it only sufficies to study the auto-correlation function $A(x) = \int f(\theta)f(x+\theta)d\theta$. First, we look at when $f(\cdot)$ and $f(\cdot + x)$ both reach f_+ or f_-:

Proposition 6. *Denote:*

$$\begin{aligned} d_+(x) &\coloneqq |\{y : f(y) = f(y+x) = f_+\}| \\ d_-(x) &\coloneqq |\{y : f(y) = f(y+x) = f_-\}| \end{aligned}$$

Then $d_+(x) - d_-(x) = d_+(0) - d_-(0) \triangleq d$.

Proof. Denote

$$\begin{aligned} l_+(x) &= |\{y : f(y) = f_+, f(y+x) = f_-\}| \\ l_-(x) &= |\{y : f(y) = f_-, f(y+x) = f_+\}| \end{aligned}$$

Then

$$\begin{aligned} d_+(0) - l_+(x) &= |\{y : f(y) = f_+\}| - |\{y : f(y) = f_+, f(y+x) = f_-\}| = |\{y : f(y) = f_+, f(y+x) = f_+\}| = d_+(x) \\ d_-(0) - l_-(x) &= |\{y : f(y) = f_-\}| - |\{y : f(y) = f_-, f(y+x) = f_+\}| = |\{y : f(y) = f_-, f(y+x) = f_-\}| = d_-(x) \end{aligned}$$

Since $f(x+\cdot)$ is a translation of f, $l_+(x) = l_-(x)$. Thus $d_+(x) - d_-(x) = d_+(0) - d_-(0)$. □

Denote $d \coloneqq d_-(0) - d_+(0)$,

$$\begin{aligned} d_+(0) + d_-(0) &= 1 \\ d_+(0)f_+ + d_-(0)f_- &= \int f dx = \bar{f} \end{aligned}$$

We get

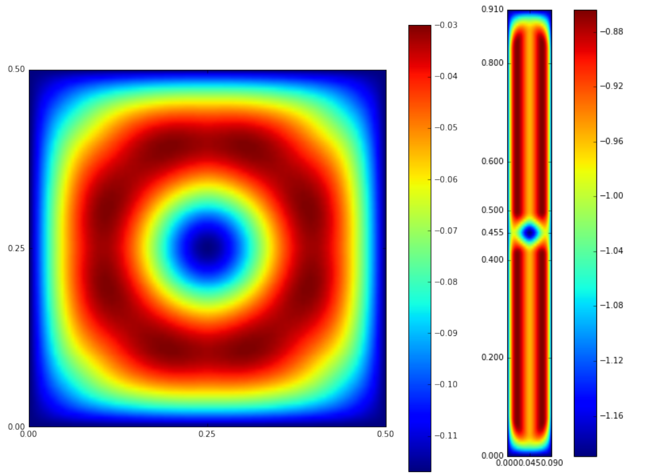
$$d_+(0) = \frac{\bar{f} - f_-}{f_+ - f_-} = \frac{1 - f_-}{f_+ - f_-} \tag{12}$$

$$d_-(0) = \frac{f_+ - \bar{f}}{f_+ - f_-} = \frac{f_+ - 1}{f_+ - f_-} \tag{13}$$

$$d = \frac{f_+ + f_- - 2\bar{f}}{f_+ - f_-} = \frac{f_+ + f_- - 2}{f_+ - f_-} \tag{14}$$

Then

$$\begin{aligned} A(x) &= \int f(\theta)f(x+\theta)d\theta \\ &= f_+^2 d_+(x) + f_-^2 d_-(x) + f_+f_-(1 - d_+(x) - d_-(x)) \\ &= f_+^2 d_+(x) + f_-^2(d_+(x) + d) + f_+f_-(1 - d_+(x) - d_+(x) - d) \\ &= (f_+ - f_-)^2 d_+(x) - f_-(f_+ - f_-)d + f_+f_- && (15) \\ &= (f_+ - f_-)^2 d_+(x) - f_-(f_+ + f_- - 2\bar{f}) + f_+f_- \\ &= (f_+ - f_-)^2 d_+(x) + f_-(2\bar{f} - f_-) && (16) \\ &= (f_+ - f_-)^2 d_-(x) + f_+(2\bar{f} - f_+) && (17) \end{aligned}$$



(a) $f_+ = 1.9$, $f_- = 0.1$, $\triangle = 0.5$ (b) $f_+ = 10.1 \gg f_- = 0.1$, $\triangle = 0.09$

Figure 1: Mutual information I with fixed f_+, f_- and varying (t, d)

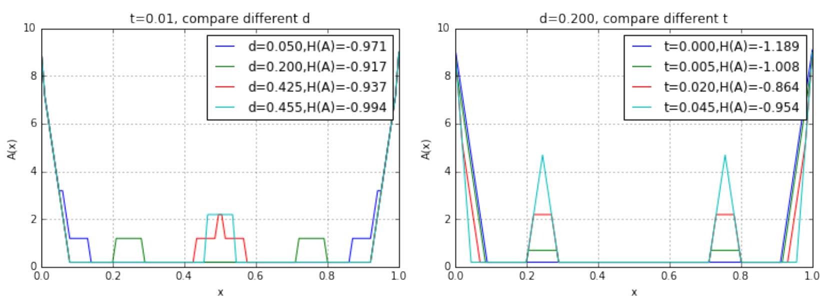


Figure 2: The shapes of $A(x)$ and $H[A]$ with varying (t, d)

4.1 Two peaks

We first focus on the case when f only has two peaks. Fix the proportion of $f = f_+$ to be $\triangle = \frac{1-f_-}{f_+-f_-}$. Express f as:

$$f = f_+ 1_{[0,t)} + f_- 1_{[t,t+d)} + f_+ 1_{[t+d,\triangle+d]} + f_- 1_{[1-\triangle-d,1)} \tag{18}$$

where $0 \le t \le \triangle$, $0 \le d \le 1 - \triangle$.

The magnitude of mutual information w.r.t. (t, d) generally follows a diamond shape. In figure 2, we plot different shapes of $A(x)$ with varying (t, d).

One **principle observation** is that large $H[A]$ favors those curves f, such that for all $x \in (0, 1)$, the number of overlappling peaks of $f(\cdot)$ and $f(x + \cdot)$ is at most one. (When $x = 0$, f and $f(x + \cdot)$ totally overlaps and there are two peaks).

In this case, $A(x) = \int f(x + \theta) f(\theta) d\theta$ puts more mass on smaller values of A, i.e. the values of A is more distributed around f_-. This phenomena might be roughly explained by the following change of variable:

$$-\int_0^1 A(x) \log(A(x)) dx = -\int_B^H \log(A) \ell(A) dA$$

where ℓ is the lebesgue measure:$\ell(A) = \lambda(\{x : A(x) < A\})$. Because $\log x$ is a concave function and $\int_B^H \ell(A) dA = \int_0^1 A(x) dx = 1$, putting more mass $\ell(A)$ on smaller values of A decreases $\int \log(A) \ell(A) dA$, thus increases the entropy.

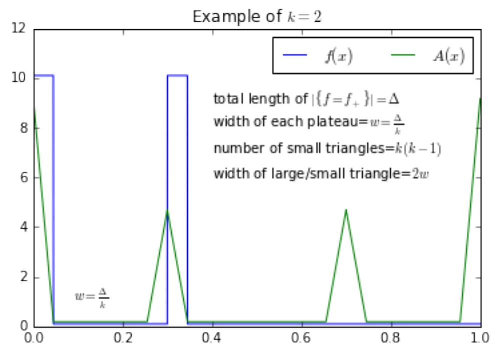


Figure 3: Illustration of a function with k peaks

4.2 Multiple peaks

We still consider $f_+ \gg f_-$, $|\{f=f_+\}| = \Delta = \frac{1-f_-}{f_+-f_-}$. Also, assume $\{f=f_+\}$ is splitted into k peaks with equal width $w = \frac{\Delta}{k}$.

Based on our observation from the 2-peaks case, we first focus on those function f such that for all $x \in (0,1)$, the number of overlapping peaks of $f(x+\cdot)$ and $f(\cdot)$ is at most one.

Since $A(x)$ is a piecewise linear function, the width and heights of each linear part can be calculated:

$$
\begin{aligned}
w &:= \frac{\Delta}{k} \\
H &:= A(0) = f_+^2\Delta + f_-^2(1-\Delta) \\
B &:= 2f_+f_-\Delta + f_-^2(1-2\Delta) \\
h(k) &:= f_+^2\frac{\Delta}{k} + 2(k-1)f_+f_-\frac{\Delta}{k} + f_-^2\left(1-(2k-1)\frac{\Delta}{k}\right)
\end{aligned}
$$

In this case, $H[A]$ can be expressed as

$$
\begin{aligned}
H[A] &= -\int_0^1 A(x)\log(A(x))dx \\
&= 2Int(H,B,w) + 2\binom{k}{2}\cdot 2Int(h(k),B,w) + \left(-B\log B\left(1-2w-2\cdot 2\binom{k}{2}w\right)\right) \\
&= 2Int(H,B,w) + 2k(k-1)\,Int(h(k),B,w) + \left(-B\log B\left(1-2w-2k(k-1)w\right)\right)
\end{aligned} \tag{19}
$$

An illustration of this setting is shown in Figure 3. Here $Int(h,b,w)$ stands for the integral of $-A(x)\log A(x)$ on each interval where $A(x)$ is linear, with upper height h, lower height b and width w:

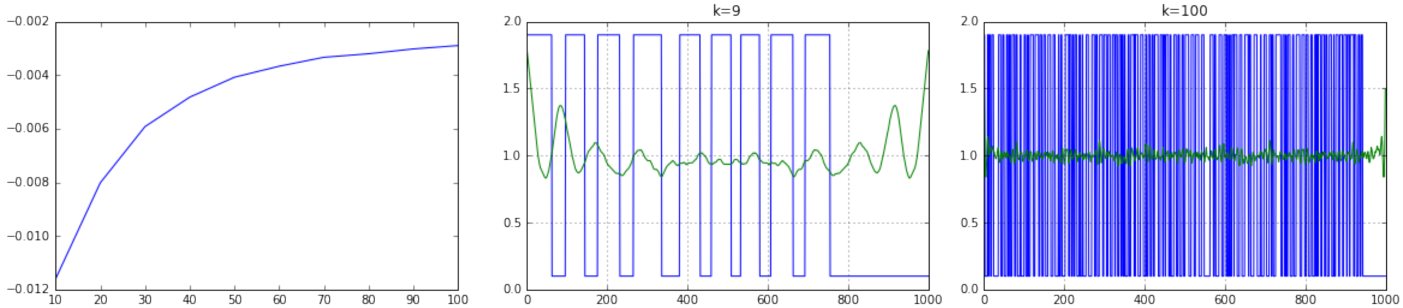
$$
Int(h,b,w) := -\int_0^w (h-\frac{h-b}{w}x)\log(h-\frac{h-b}{w}x) = \frac{w}{4(h-b)}\left[h^2-b^2-2(h^2\log h - b^2\log b)\right] \tag{20}
$$

We first evaluate equation (19) numerically for different k's. We observe that:

1. $H(A)$ increases as k is increasing, suggesting that large $H[A]$ favors the splitting of $f(x)$ into more peaks.
2. k cannot increase to infinity since $k(k-1)2\frac{\Delta}{k} + 2\frac{\Delta}{k} \le 1$.

We finally tried to do **randomly** split f into widths $\{\alpha_i\}_{i=1,..,k}$,$\{\beta_i\}_{i=1,...,k}$, with $\sum_{i=1}^k \alpha_i = \triangle$, $\sum_{i=1}^k \beta_i = 1-\triangle$, such that $f=f_+$ on intervals $\left[\sum_{j=1}^{i-1}(\alpha_j+\beta_j), \sum_{j=1}^{i-1}(\alpha_j+\beta_j)+\alpha_i\right)$, and $f=f_-$ on intervals $\left[\sum_{j=1}^{i-1}(\alpha_j+\beta_j)+\alpha_i, \sum_{j=1}^{i}(\alpha_j+\beta_j)\right)$. For each k, we took the maximum of $H[A]$ from 1000 samples of α_i and β_i's, drawn from Multinomial distributions.

Figure 4 shows the numerical results. As k increases, $H[A]$ increases and approaches a fixed value, and as shown in Figure 4b, 4c, the function f which produces highest $H[A]$ seems to vary more and more 'wildy' between f_- and f_+. $A(x)$ approaches a near-constant function, taking $H = f_+^2\Delta + f_-^2(1-\Delta)$ at 0, and fastly decreases to 1.



(a) $H[A]$ as $k \to 100$ (b) The maximal $f(x)$(blue) and $A(x)$(green) when $k=9$ (c) The maximal $f(x)$ and $A(x)$ when $k=100$

Figure 4: Random experiments of a function with k peaks

5 Maximum of the mutual information

5.1 Maximum of the negative entropy

Lemma 7. *Assume $f(x)$ is piece-wise C^1. Under the given constraints $f_- \le f \le f_+$ and $\int_0^1 f(x)dx = 1$, $-H[f]$ is maximized when f only takes f_- and f_+ a.e.*

Proof. We combined the KKT conditions for both $\nabla_f H$ and $\nabla_\delta H$ in the Section of preturbation method. Let $G[f] = -G[f]$ be the objective function we want to maximize. It can be derived from $G[f+h] - G[f]$ that

$$\nabla_f \left(G[f]\right)(x) = \ln f(x) + 1.$$

For a local maximizer f of G, there exists functions $\alpha_+(x) \ge 0$, $\alpha_-(x) \ge 0$ and a constant μ, such that for all x where $f(x)$ (thus also $\nabla_f G(x)$) is continous, the following KKT conditions hold:

$$\nabla_f I(x) = \alpha_+(x) - \alpha_-(x) + \mu$$

and $\alpha_+(x)\left(f(x) - f_+\right) = 0$, $\alpha_-(x)\left(f_- - f(x)\right) = 0$. From $\nabla_f G(x) = \ln f(x) + 1$, $f(x)$ is constant when $f(x) \ne f_+$ or f_-, since $\alpha_+ = \alpha_- = 0$ implies $\ln f(x) + 1 = \mu$. Hence $f(x)$ is a piece-wise constant function which takes at most 3 values:

$$f(x) = \begin{cases} f_+ \\ f_- \\ e^{\mu-1} & \text{if } f \ne f_+ \text{ or } f_- \end{cases}$$

Next we show that the third option is impossible using preturbation method. It is shown in the Appendix A.2 that

$$\nabla_\delta \left(G[f]\right)(x) = 1_{\{x \notin D\}} (f \log f)'_c(x) + \sum_{x_i \in D} \triangle_i (f \log f) \delta_{x_i}(x)$$

where D is the set of discontinuities of f, $(f \log f)'_c(x)$ is the point-wise derivative of f. From maximizing the function G with respect to δ, there exists $\nu \in \mathbb{R}$ such that the KKT condition holds:

$$\nabla_\delta \left(G[f]\right)(x) = \nu \left(f'_c(x) + \sum_{x_i \in D} \triangle_i f \delta_{x_i}(x) \right)$$

where the right-hand-side comes from the constraint $\int f(x)\delta'(x)dx = -\int \delta(x)dx \left(f'_c(x) + \sum_{i \in D} \triangle_i f \delta_{x_i}(x)\right) = 0$. Now since f is piecewise constant, for all $x \notin D$, $f'_c(x) = 0$ and $(f \log f)'_c(x) = 0$. Thus

$$\begin{aligned} \sum_{x_i \in D} \triangle_i (f \log f) \delta_{x_i}(x) &= \nu \sum_{x_i \in D} \triangle_i f \delta_{x_i}(x) \\ \frac{\triangle_i (f \log f)}{\triangle_i f} &= \nu, \forall x_i \in D \end{aligned}$$

Suppose that there exists $\tilde{f}$ such that $f_+ < \tilde{f} < f_-$and that $\{f = \tilde{f}\}$ has positive measure. Then at discontinuities, f either jumps between f_+ and f_-, or between f_+(or f_-) and $\tilde{f}$. There are 3 options for $\frac{\triangle_i(f\log f)}{\triangle_i f}$, which are equal to a constant ν:

$$\frac{f_+\log(f_+)-f_-\log(f_-)}{f_+-f_-}=\frac{f_+\log(f_+)-\tilde{f}\log(\tilde{f})}{f_+-\tilde{f}}=\frac{\tilde{f}\log(\tilde{f})-f_-\log(f_-)}{\tilde{f}-f_-}=\nu.$$

However, this contradicts the convexity of the function $y = t\log t$. Therefore it is impossible to have $f_+ < \tilde{f} < f_-$ with positive measure, and $f = f_+$ or f_- a.e. □

5.2 Upper bound of the entropy of auto-correlation

Lemma 8. *Given $\int_0^1 f(x)dx = 1$, there is an upper bound for $I[f]$ such that*

$$I[f] \quad \leq \quad -2H[f]$$

Proof. It suffices to show $H(A) \leq 0$. Since $A(x)$ satisfies $\int_0^1 A(x)dx$ and $A \geq 0$, the entropy of a density function is maximized by uniform distribution on [0,1):

First, the K-L divergence $D_{KL}(f\|g) \geq$ for any two distributions f, g. Let U be the uniform distribution on $[0, 1)$, then

$$\begin{aligned}
D_{KL}(A\|U) &= \int_0^1 A(x)\log\frac{A(x)}{U(x)}dx \\
&= \int_0^1 A(x)\log\frac{A(x)}{1}dx \\
&= -H[A] \\
&\geq 0
\end{aligned}$$

Therefore $H[A] \leq 0$, $I[f] = -2H[f] + H[A] \leq -2H[f]$. □

As shown in the proof the above lemma, this maximum reachable only when A is the uniform distribution on $[0, 1)$. Ideally, this happens when $f(x)$ are i.i.d. random variables with $P(f = f_+) = \triangle$, $P(f = f_-) = 1 - \triangle$, where $\triangle = \frac{1-f_-}{f_+-f_-}$:

$$\begin{aligned}
A(x) &= \int f(\theta)f(x+\theta)d\theta \\
&= E\left[f(\theta)f(x+\theta)\right] \\
&= E\left[f(\theta)\right]E\left[f(x+\theta)\right] \\
&= (\triangle f_+ + (1-\triangle)f_-)^2 \\
&= 1
\end{aligned}$$

Questions remained:

- Integral well-defined?

Note that this auto-correlation function $A(x)$ may not be continuous, which is different from the property stated in Section 1, since here f does not satisfy the condition that it has only a finite number of discontinuities.

5.3 Main Theorem

The conclusion above can be generalized to our Poissonian model:

Theorem 9. $I[f] = -E_N\left[\int_0^1 \prod_{i=1}^n \frac{f(s_i-\theta)}{f(s_i)}d\theta\right] \leq -\bar{f}H[h]$, *where h is a piece-wise constant function only taking two values $\left\{\frac{f_-}{\bar{f}}, \frac{f_+}{\bar{f}}\right\}$ on $[0, 1)$ with $\int_0^1 h(x)dx = 1$. For such an h, $H[h]$ has fixed value: $H[h] = -\left(\frac{f_+}{\bar{f}}\ln\left(\frac{f_+}{\bar{f}}\right)\triangle + \frac{f_-}{\bar{f}}\ln\left(\frac{f_-}{\bar{f}}\right)(1-\triangle)\right)$, where $\triangle = \frac{\bar{f}-f_-}{f_+-f_-}$.*

Proof. Let $\tilde{f} = \frac{f}{\bar{f}}$, then $\frac{f_-}{\bar{f}} \leq \tilde{f} \leq \frac{f+}{\bar{f}}$,

$$
\begin{aligned}
I[f] &= -E_N\left[\int_0^1 \prod_{i=1}^n \frac{f(s_i-\theta)}{f(s_i)} d\theta\right] \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^1 \cdots \int_0^1 \prod_{i=1}^n ds_i f(s_i) e^{-\int_0^1 f(s)ds} \ln\left(\frac{\prod_{i=1}^n f(s_i)}{\int_0^1 \prod_{i=1}^n f(s_i-\theta)d\theta}\right) \\
&= \sum_n \frac{1}{n!} e^{-\bar{f}} \left(\bar{f}\right)^n \int_0^1 \cdots \int_0^1 \prod_{i=1}^n ds_i \tilde{f}(s_i) \ln\left(\frac{\prod_{i=1}^n \tilde{f}(s_i)}{\int_0^1 \prod_{i=1}^n \tilde{f}(s_i-\theta)d\theta}\right) \\
&= \sum_n \frac{1}{n!} e^{-\bar{f}} \left(\bar{f}\right)^n \left[\int_0^1 \cdots \int_0^1 \prod_{i=1}^n ds_i \tilde{f}(s_i) \ln\left(\prod_{i=1}^n \tilde{f}(s_i)\right) - \int_0^1 \cdots \int_0^1 \prod_{i=1}^n ds_i \tilde{f}(s_i) \ln\left(A(s_1,\ldots,s_n)\right)\right] \\
&\leq \sum_n \frac{1}{n!} e^{-\bar{f}} \left(\bar{f}\right)^n \int_0^1 \cdots \int_0^1 \prod_{i=1}^n ds_i \tilde{f}(s_i) \ln\left(\prod_{i=1}^n \tilde{f}(s_i)\right) \\
&= \sum_n \frac{1}{n!} e^{-\bar{f}} \left(\bar{f}\right)^n \left(-nH[\tilde{f}]\right) \\
&= \left(-H[\tilde{f}]\right) \sum_{n=1}^{\infty} \frac{1}{(n-1)!} e^{-\bar{f}} \left(\bar{f}\right)^n \\
&= -\bar{f} H[\tilde{f}] \\
&\leq -\bar{f} H[h]
\end{aligned} \tag{21}
$$

Where the last step is followed from Lemma 7. □

Note that the maximum is attained when $A(s_1,\ldots,s_n)$ is constant (equal to one) for all n.

A Appendix: Compuation of the gradients

A.1 Computation of $\nabla_f I$

First we show the following two conclusions:

$$\nabla_f|_t \left(H[f]\right) = -\left(\ln f(t) + 1\right) \tag{22}$$

$$\nabla_f|_t \left(A[f](u)\right) = f(t+u) + f(t-u) \tag{23}$$

where the 2nd one can be derived from

$$
\begin{aligned}
&\int_0^1 d\theta \left(f(u+\theta) + h(u+\theta)\right)\left(f(\theta) + h(\theta)\right) \\
=\ & A(u) + \int d\theta f(u+\theta)h(\theta) + \int d\theta f(\theta)h(u+\theta) + O(h^2) \\
=\ & A(u) + \int d\theta f(u+\theta)h(\theta) + \int d\theta' f(\theta'-u)h(\theta') + O(h^2).
\end{aligned}
$$

Now since $I = -2H(f) + H(A)$, using (22)(23), we obtain

$$
\begin{aligned}
\nabla_f I(t) &= -2\nabla_f|_t\left(H[f]\right)+\int_0^1 du\left(-\left(\ln A(u)+1\right)\right)\nabla_f|_t A(u)\\
&= 2\left(\ln f(t)+1\right)-\int_0^1 du\left(\ln A(u)+1\right)\left(f(t+u)+f(t-u)\right)\\
&= 2\left(\ln f(t)+1\right)-\int_0^1 f(t+u)\ln A(u)du-\int_0^1 f(t-u)\ln A(u)du-2\\
&= 2\left(\ln f(t)-\int_0^1 f(y)\ln A(y-t)dy\right)
\end{aligned}
$$

*Note that previously we derived $\nabla_f I(t)=2\left(\ln f(t)-H[f]-\int_0^1 f(y)\ln A(y-t)dy\right)$ from the expression of I: $\int_0^1 dx\int_0^1 dyf(x)f($ before simplification. Since the simplification envolves plugging in the constraint $\int f(x)dx=1$, which also produces some derivative, the expression of $\nabla_f I$ is differed by a constant. Either including the constant $H[f]$ or not does not affect our conclusions since it can be absorbed into μ in the KKT condition.

A.2 Computation of $\nabla_\delta I$

To compute $\nabla_\delta I[f]$, first we compute $\nabla_\delta\left(\int f\log f\right)$: applying change of variable $\eta=x+\epsilon\delta(x)$, $d\eta=(1+\epsilon\delta'(x))dx$,

$$
dx=\frac{1}{1+\epsilon\delta'(x)}d\eta=(1-\epsilon\delta'(x)+O(\epsilon^2))d\eta=(1-\epsilon\delta'(\eta)+O(\epsilon^2))d\eta
$$

$$
\begin{aligned}
-H[f(x+\epsilon\delta(x))] &= \int dxf(x+\epsilon\delta(x))\log(f(x+\epsilon\delta(x)))\\
&= \int d\eta\left(1-\epsilon\delta'(\eta)+O(\epsilon^2)\right)f(\eta)\log(f(\eta))\\
&= -H[f]+\epsilon\left[\int\delta(\eta)(f\log f)_c'(\eta)d\eta+\sum_{i\in D}\triangle_i(f\log f)\delta(x_i)\right]+O(\epsilon^2)
\end{aligned}
$$

Similarily, define $F(x,y)=f(x)f(y)\ln\left(\int_0^1 d\theta f(x-\theta)f(y-\theta)\right)$, then

$$
\begin{aligned}
&\int dx\int dyF(x+\epsilon\delta(x),y+\epsilon\delta(y))\\
=&\int d\eta\int d\gamma F(\eta,\gamma)-2\epsilon\int\left(\int d\gamma F(\eta,\gamma)\right)\delta'(\eta)d\eta+O(\epsilon^2)\\
=&\int d\eta\int d\gamma F(\eta,\gamma)+2\epsilon\left[\int\left(\int d\gamma\frac{\partial F_c(\eta,\gamma)}{\partial\eta}\right)\delta(\eta)d\eta+\sum_{i\in D}\triangle_i\left(\int d\gamma F(\eta,\gamma)\right)\delta_{x_i}(\eta)\right]+O(\epsilon^2)\\
=&\int d\eta\int d\gamma F(\eta,\gamma)+O(\epsilon^2)+2\epsilon\int\delta(\eta)d\eta\int d\gamma f_c'(\eta)f(\gamma)\ln\left(A(\gamma-\eta)\right)\\
&+2\epsilon\left[\int\delta(\eta)d\eta\int d\gamma f(\gamma)f(\eta)\frac{\int_0^1 d\theta f'(\eta-\theta)f(\gamma-\theta)}{A(\gamma-\eta)}+\sum_{i\in D}\triangle_i\left(f(\eta)\int d\gamma f(\gamma)\ln\left(A(\gamma-\eta)\right)\right)\delta_{x_i}(\eta)\right]
\end{aligned}
$$

Thus

$$
\nabla_\delta\left(-H[f]\right)=(f\log f)_c'(x)+\sum_{i\in D}\triangle_i(f\log f)\delta_{x_i}(x)\tag{24}
$$

$$
\begin{aligned}
\nabla_\delta\left(\int dx\int dyF(x,y)\right) &= 2\left[f_c'(x)\int dyf(y)\log\left(A(y-x)\right)+f(x)\int dyf(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)}\right]\\
&\quad+2\sum_{i\in D}\triangle_i f\cdot\left(\int dyf(y)\log A(y-x_i)\right)\cdot\delta_{x_i}(x)
\end{aligned}
$$

Therefore from $I[f]=-2H(f)-\int dx\int dyF(x,y)$ we obtain

$$\begin{aligned}\nabla_\delta I[f] &= 2\left[f_c'(x)\left(\log f(x)+1\right)-f_c'(x)\int dyf(y)\log A(y-x)-f(x)\int dyf(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)}\right]\\ &\quad +2\sum_{i\in D}\left[\triangle_i(f\log f)-\triangle_i f\cdot\int dyf(y)\log A(y-x_i)\right]\delta_{x_i}(x)\end{aligned}$$

A.3 Other conclusions

*The following conclusions are derived from 'integration by parts' below, which we were not sure about their correctness (and they contradict the numerical simulations):

$$\begin{aligned}\int_0^1 f'(x-\theta)f(y-\theta)d\theta &= \int_0^1\left(f_c'(x-\theta)+\sum_{i\in D}\triangle_i f\delta_{x_i}(x-\theta)\right)f(y-\theta)d\theta\\ &= -\int_0^1 f_c(x-\theta)f'(y-\theta)d\theta+\sum_{i\in D}\triangle_i f\cdot f(y-x+x_i)\\ &= -\int_0^1 f(x-\theta)f(y-\theta)d\theta+\sum_{i\in D}\triangle_i f\cdot f(y-x+x_i)\end{aligned}$$

$$\begin{aligned}\int dyf(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)} &= -\int dyf(y)\frac{\int_0^1 d\theta f(x-\theta)f'(y-\theta)}{A(y-x)}+\sum_{i\in D}\triangle_i f\cdot\int\frac{f(y)f(y-x+x_i)}{A(y-x)}dy\\ &= -\int dyf(y)\frac{\partial\log A(y-x)}{\partial y}+\sum_{i\in D}\triangle_i f\cdot\int\frac{f(y)f(y-x+x_i)}{A(y-x)}dy\\ &= \int dyf'(y)\log A(y-x)+\sum_{i\in D}\triangle_i f\cdot\int\frac{f(y)f(y-x+x_i)}{A(y-x)}dy\end{aligned}$$

$$\begin{aligned}\nabla_\delta I[f] &= 2\left[f_c'(x)(\log f_c(x)+1)-\int dy\left(f_c'(x)f(y)+f(x)f'(y)\right)\log A(y-x)-f(x)\sum_{i\in D}\triangle_i f\int\frac{f(y)f(y-x+x_i)}{A(y-x)}dy\right]\\ &\quad +2\sum_{i\in D}\left(\triangle_i(f\log f)-\triangle_i f\cdot\int dyf(y)\log A(y-x_i)\right)\cdot\delta_{x_i}(x)\end{aligned}\tag{25}$$

*Especially, if f is piece-wise constant, since its point-wise derivative $f_c'(x)=0$,

$$\begin{aligned}\nabla_\delta I[f] &= -2f(x)\int dyf'(y)\log A(y-x)-2f(x)\sum_{i\in D}\triangle_i f\cdot\int\frac{f(y)f(y-x+x_i)}{A(y-x)}dy\\ &\quad +2\sum_{i\in D}\left(\triangle_i(f\log f)-\triangle_i f\cdot\int dyf(y)\log A(y-x_i)\right)\cdot\delta_{x_i}(x)\\ &= -2f(x)\sum_{i\in D}\triangle_i f\left(\log A(x_i-x)+\int\frac{f(y)f(y-x+x_i)}{A(y-x)}dy\right)\\ &\quad +2\sum_{i\in D}\left(\triangle_i(f\log f)-\triangle_i f\cdot\int dyf(y)\log A(y-x_i)\right)\cdot\delta_{x_i}(x)\end{aligned}$$

By plugging in equation (10) into $\nabla_\delta I$ in (25) and (9), we get

$$\begin{aligned}
\nabla_\delta I[f] &= 2\left[f_c'(x)(\log f_c(x)+1) - f_c'(x)\int dy f(y)\log A(y-x) - f(x)\int dy f(y)\frac{\int_0^1 d\theta f'(x-\theta)f(y-\theta)}{\int_0^1 d\theta f(x-\theta)f(y-\theta)}\right] \\
&\quad +2\sum_{i\in D}\left(\triangle_i(f\log f) - \triangle_i f\cdot\int dy f(y)\log A(y-x_i)\right)\cdot\delta_{x_i}(x) \\
&= 2f'(x)(\log f(x)+1) - 2f'(x)\int dy f(y)\log\left(A(y-x)\right) - 2f(x)\left(\frac{f'(x)}{f(x)} - \frac{\alpha_+'(x)}{2} + \frac{\alpha_-'(x)}{2}\right) \\
&\quad +2\left(\triangle_i(f\log f) - \triangle_i f\cdot\int dy f(y)\log A(y-x_i)\right)\cdot\delta_{x_i}(x) \\
&= \nu\left[f'(x) + \sum_{i\in D}\triangle_i f\cdot\delta_{x_i}(x)\right]
\end{aligned}$$

$$f'(x)\log f(x) = f'(x)\left(\int dy f(y)\log A(y-x) + \frac{\nu}{2}\right) - f(x)\left(\alpha_+'(x) - \alpha_-'(x)\right) \text{ for } x\notin D \tag{26}$$

$$\frac{\triangle_i(f\log f)}{\triangle_i f} = \int dy f(y)\log A(y-x_i) + \frac{\nu}{2} \text{ for } x_i\in D \tag{27}$$

Proposition 10. *$\alpha_+(x)$, $\alpha_-(x)$,$\nabla_f I(x)$ are piecewise constant functions.*

Proof. From equation (8) we only need to show $\alpha_+(x)$ is piecewise const.

If $f(x) < f_+$, $\alpha_+(x) = 0$.

If $f(x) = f_+$, $\alpha_+(x) \geq 0$, $\alpha_-(x) = 0$. There exists an interval $\mathcal{I}$ around x such that $f = f_+$ on $\mathcal{I}$. Thus on inverval $\mathcal{I}$, $f'(x) = 0$, by equation (26),

$$\begin{aligned}
f(x)\left(\alpha_+'(x) - \alpha_-'(x)\right) &= 0 \\
\alpha_+'(x) &= 0
\end{aligned}$$

Thus $\alpha_+(x)$ is constant on inverval $\mathcal{I}$ around x, which gives us the conclusion that α_+ is piecewise const. □

Therefore, f **cannot vary continuously between** f_+ **and** f_-, otherwise ∇f is continuous but α_+ or α_- is discontinous.

We can also further simplify equation (10) as

$$\frac{f'(x)}{f(x)} = \int_0^1 dy f(y)\frac{\int_0^1 f'(x-\theta)f(y-\theta)d\theta}{\int_0^1 f(x-\theta)f(y-\theta)d\theta} \text{ for } x\notin D \tag{28}$$

Proposition 11. *If $f(x)$ has a jump discontinuity at x_0, then there exists an open interval (β_1, β_2) such that $\beta_1 < x_0 < \beta_2$ and $f(x)$ is constant on both (β_1, x_0) and (x_0, β_2).*

Proof. Suppose there's no such interval on the left of x_0. Then there exists a sequence $x_k \to x_0^-$ such that $f'(x_k) \neq 0$. Applying equations (26), (27),

$$\begin{aligned}
\log f(x_0^-) &= \int dy f(y)\log A(y-x_0) + \frac{\nu}{2} \\
\frac{\triangle_{x_0}(f\log f)}{\triangle_{x_0} f} &= \int dy f(y)\log A(y-x_0) + \frac{\nu}{2}
\end{aligned}$$

Therefore

$$\log f(x_0^-) = \frac{\triangle_{x_0}(f\log f)}{\triangle_{x_0} f} = \frac{f(x_0^+)\log f(x_0^+) - f(x_0^-)\log f(x_0^-)}{f(x_0^+) - f(x_0^-)}$$

which implies $f(x_0^+) = f(x_0^-)$. This contradicts with the discontinuity at x_0. Similarly, there's an interval to the right of x_0 such that f is constant. □