An Inequality for Polynomial with nonnegative coefficients homogeneous and Rational Functions

Yueng-Tien, Lo
g96470198@csie.ntnu.edu.tw
Speech Lab, CSIE
National Taiwan Normal University

Reference:
An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology
Leonard E.Baum J.A. Eagon
An inequality for rational functions with applications to some statistical estimation problems
Gopalakrishnan, P.S. Kanevsky, D. Nadas, A. Nahamoo, D.
IBM Thomas J. Watson Res. Center, Yorktown Heights, NY;

outline

- Introduction
- Baum-Eagon inequality
- Extended Baum-Welch algorithm
- Reduction of the Case of Rational Functions to Polynomials
- An Inequality for A Polynomial
- Growth Transformations For Rational Functions
- *Gopolakrishnan's Theorem*
- *Example*

Introduction

- The well-known Baum-Eagon inequality provides an effective iterative scheme for finding a local maximum for homogeneous polynomials with positive coefficients over a domain of probability values

- However, we are interesting in maximizing a general rational function. We extend the Baum-Eagon inequality to rational function

Baum-Eagon inequality

- Baum-Eagon inequality (L. E. Baum, J. I. Eagon,1966)

Let $P(X) = P(\{X_{ij}\})$ be a polynomial with nonnegative coefficients homogeneous of degree d in its variables X_{ij} . Let $x = \{x_{ij}\}$ be any point of the domain $D:\ x_{ij} \ge 0,\ \sum_{j=1}^{q_i} x_{ij} = 1,\ i = 1,\cdots,p,\ j = 1,\cdots,q_i$,such that

$$\sum_{j=1}^{q_i} x_{ij} \frac{\partial P(x_{ij})}{\partial x_{ij}} \neq 0, \qquad \text{for all i,}$$

$$\text{where } \frac{\partial P}{\partial \mathrm{X}}(x) \textit{ denotes the value of } \frac{\partial P(X)}{\partial X} \textit{ at } x.$$

Let $\Im(x) = \Im(\{x_{ij}\})$ denote the point of D whose i, j coordinate is

$$\Im(x)_{ij} = \frac{x_{ij} \dfrac{\partial P(x_{ij})}{\partial x_{ij}}}{\sum_{j=1}^{q_i} x_{ij} \dfrac{\partial P(x_{ij})}{\partial x_{ij}}}$$

Then $P(\Im(x)) > P(x)$ unless $\Im(x) = x$

Notation

- μ will denote a doubly indexed array of nonnegative integers:
- $$\mu = \{\mu_{ij}\}, j = 1,\cdots,q_{ij}, \ i = 1,\cdots,p.$$

$$x^{\mu} \text{ then denotes } \prod_{i=1}^{p} \prod_{j=1}^{q_i} x_{ij}^{\mu_{ij}}$$

$$c_{\mu} \text{ is an abbreviation for } c_{\{\mu_{ij}\}}$$

$$\text{The polynomial } P(\{x_{ij}\}) \text{ is then written } P(x) = \sum_{\mu} c_{\mu} x^{\mu}$$

$$\Im(x)_{ij} = (\sum_{\mu} c_{\mu} \mu_{ij} x^{\mu}) \Big/ \sum_{j=1}^{q_i} \sum_{\mu} c_{\mu} \mu_{ij} x^{\mu}$$

- We wish to prove

$$P(x) = \sum_{\mu} c_{\mu} x^{\mu} \le \sum_{\mu} c_{\mu} \prod_{i=1}^{p} \prod_{j=1}^{q_i} \Im(x)_{ij}^{\mu_{ij}}.$$

Proof.(1/3)

$$P(x)=\sum_{\mu}\left\langle \boxed{\left\{c_{\mu}\prod_{i=1}^{p}\prod_{j=1}^{q_i}\Im(x)_{ij}^{\mu_{ij}}\right\}^{1/(d+1)}} \times \boxed{\left\{c_{\mu}^{d/(d+1)}x^{\mu}\prod_{i=1}^{p}\prod_{j=1}^{q_i}\left(\frac{1}{\Im(x)_{ij}}\right)^{\mu_{ij}/(d+1)}\right\}}\right\rangle$$

- We apply Hölder's inequality to obtain

$$P(x)\le\left\{\sum_{\mu}c_{\mu}\prod_{i=1}^{p}\prod_{j=1}^{q_i}\Im(x)_{ij}^{\mu_{ij}}\right\}^{1/(d+1)}\times\left\{\sum_{\mu}c_{\mu}x^{\mu}\prod_{i=1}^{p}\prod_{j=1}^{q_i}\left(\frac{x_{ij}}{\Im(x)_{ij}}\right)^{\mu_{ij}/d}\right\}^{d/(d+1)}$$

- (In the last braces we have used $\left(x^{\mu}\right)^{(d+1)/d}=x^{\mu}\prod_{i=1}^{p}\prod_{j=1}^{q_i}x_{ij}^{\mu_{ij}/d}$)
- Since $\sum_{i=1}^{p}\sum_{j=1}^{q_i}\mu_{ij}/d\equiv 1$ by homogeneity of P, we can apply the inequality of geometric and arithmetic means to the double products of the second brace of (3) to conclude:

$$\sum_{\mu}c_{\mu}x^{\mu}\prod_{i=1}^{p}\prod_{j=1}^{q_i}\left(\frac{x_{ij}}{\Im(x)_{ij}}\right)^{\mu_{ij}/d}\le\sum_{\mu}c_{\mu}x^{\mu}\sum_{i=1}^{p}\sum_{j=1}^{q_i}\frac{\mu_{ij}}{d}\frac{x_{ij}}{\Im(x)_{ij}}$$

Proof.(2/3)

- We now substitute the definition (1) of $\Im(x)_{ij}$ in the expression on the right of (4) and interchange the order of summation to obtain:

$$\sum_{\mu} c_\mu x^\mu \sum_{i=1}^{p}\sum_{j=1}^{q_i} \frac{\mu_{ij}}{d} \frac{x_{ij}}{\Im(x)_{ij}}$$

$$= \frac{1}{d}\sum_{\mu} c_\mu x^\mu \sum_{i=1}^{p}\sum_{j=1}^{q_i} \left[\mu_{ij} x_{ij} \cdot \left(\sum_{j_0=1}^{q_i}\sum_{\mu'} c_{\mu'} \mu'_{ij}\, x^{\mu} \right) \Big/ \left(\sum_{\mu'} c_{\mu'} \mu'_{ij}\, x^{\mu'} \right) \right]$$

$$= \frac{1}{d}\sum_{i=1}^{p} \left\{ \sum_{j=1}^{q_i} x_{ij} \underline{\left[\left(\sum_{\mu} \mu_{ij} c_\mu x^\mu \right) \Big/ \left(\sum_{\mu'} \mu'_{ij}\, c_{\mu'} x^{\mu'} \right) \right]} \cdot \sum_{j_0=1}^{q_i}\sum_{\mu'} c_{\mu'} \mu'_{ij_0}\, x^{\mu'} \right\}$$

- For each $\langle i,j \rangle$ the expression within the brackets is =1 and by hypothesis for each i, $\sum_{j=1}^{q_i} x_{ij} = 1$.Hence the whole last expression of (5) reduces to $(1/d)\sum_{i=1}^{p}\sum_{j_0=1}^{q_i}\sum_{\mu'} c_{\mu'} \mu'_{ij_0}\, x^{\mu'}$ .But this is just $(1/d)\sum_{ij_0} x_{ij_0} \cdot \left(\partial P / \partial x_{ij_0}\right)$. So by the Euler theorem for homogeneous functions it is equal to $\sum_{\mu} c_\mu x^\mu$

Proof.(3/3)

- Finally, if we use this upper bound $\sum_{\mu} c_{\mu} x^{\mu}$ for the expression within the second braces in (3), raise both sides of (3) to the (d+1)st power, and divide both sides of the resulting inequality by $\left(\sum_{\mu} c_{\mu} x^{\mu}\right)^{d}$ we obtain the desired inequality (2).

- That $P(\Im\{x_{ij}\}) > P\{x_{ij}\}$ if $\{x_{ij}\} \neq \{x_{ij}\}$ follows from (4) and the strictness of the inequality of geometric and arithmetic means if all summands are not equal.

Hölder's inequality

Let $\frac{1}{p}+\frac{1}{q}=1$

- With $p,q>1$ inequality for integrals states that

$$\int_a^b \left| f(x)g(x) \right| dx \le \left[\int_a^b \left| f(x) \right|^p dx \right]^{1/p} \left[\int_a^b \left| g(x) \right|^q dx \right]^{1/q}$$

With equality when

$$\left| g(x) \right| = c \left| f(x) \right|^{p-1}$$

If $p=q=2$,this inequality becomes Schwarz's inequality.

- For sums states that

$$\sum_{k=1}^{n} \left| a_k b_k \right| \le \left(\sum_{k=1}^{n} \left| a_k \right|^p \right)^{1/p} \left(\sum_{k=1}^{n} \left| b_k \right|^q \right)^{1/q}$$

With equality when

$$\left| b_k \right| = c \left| a_k \right|^{p-1}$$

If $p=q=2$,this inequality becomes Cauchy's inequality

Euler's Homogeneous Function Theorem

- Let $f(x,y)$ be a homogeneous function of order n so that

$$f(tx,ty) = t^n f(x,y)$$

- Then define $x' \equiv xt$ and $y' \equiv yt$. Then

$$\begin{aligned} nt^{n-1} f(x,y) &= \frac{\partial f}{\partial x'}\frac{\partial x'}{\partial t} + \frac{\partial f}{\partial y'}\frac{\partial y'}{\partial t} \\ &= x\frac{\partial f}{\partial x'} + y\frac{\partial f}{\partial y'} \\ &= x\frac{\partial f}{\partial (xt)} + y\frac{\partial f}{\partial (yt)} \end{aligned}$$

- Let t = 1, then

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = nf(x,y)$$

- This can be generalized to an arbitrary number of variables

$$x_i \frac{\partial f}{\partial x_i} = nf(x)$$

Where Einstein summation has been used

Extend the Baum-Welch algorithm

- an arbitrary homogeneous polynomial $P(X) = P(\{X_{ij}\})$ with nonnegative coefficient of degree *d* in variables $X_{ij}, i = 1,..., p, \ j = 1,..., q_i$

  Assuming that this polynomial is defined over a domain of probability values, they show how to construct a transformation $T : U \rightarrow D$ for some $U \subseteq D$ such that following the property:

  $$D : x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1$$

  property A : For any $x \in U$ and $y = T(x)$, $P(y) > P(x)$ unless $y = x$

- Polynomials of this type appear in various statistical problems dealing with the estimation of probabilistic functions of Markov processes via the maximum likelihood technique, and the previous inequality provides an effective iterative scheme for find a local maximum. A general paradigm used for maximization is an E-M algorithm (Expectation-Maximization).

- In certain statistical problems it was found that estimation of parameter by some other criteria that use conditional likelihood, mutual information can give empirically better results than estimation via maximum likelihood. But these problems require finding local maxima for rational functions.

Extend the Baum-Welch algorithm

- Given the domain D (previously defined) and a rational function $R(X) = S_1(X)/S_2(X)$ (where $S_1(X), S_2(X)$ are polynomials with real coefficients in variables $X = \{X_{ij}\}_i$ and $S_2(X)$ has only positive values in D) we provide a way for constructing a large class of transformations $T : D \rightarrow D$ such that the analog of the Property A (with $R(x)$ instead of $P(x)$) holds.

property B : For any $x \in D$ and $y = T(x)$, $R(y) > R(x)$ unless $y = x$

- If a transformation $T : D \rightarrow D$ satisfies Property B we will say that T is a growth transformation of D for $R(x)$

Reduction of the Case of Rational Functions to Polynomials

- $R(X) = S_1(X)/S_2(X)$ is a ratio of two polynomials $S_1(X), S_2(X) > 0$ in variables
- $X = \{X_{ij}\}, i = 1,\ldots, p, \ j = 1,\ldots, q_i$ defined over a domain $D : x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1$

  we are looking for a growth transformation $T : D \to D$ such that
  for any $x \in D$ and $y = T(x)$, $R(y) > R(x)$ unless $y = x$

- we reduce the problem of finding a growth transformation for a rational function to one of finding that for a specially formed polynomial.

  1.Change ratio of polynomial $R(x)$ into a polynomial $P(x)$

- reduce to Non-homogeneous polynomial with nonnegative

  2.Add a constant to $P(x)$ so that the new polynomial $P'(x) \hat{=} P(x) + C(x)$ has only nonnegative coefficients.

- Extend Baum-Eagon inequality to Non-homogeneous polynomial with nonnegative

  3.Do a variable substitution to ensure that the resulting polynomial $P''(x)$ is homogeneous.

Reduction of the Case of Rational Functions to Polynomials Step.1

- Step1:

  for any $x \in D$ there exists a polynomial $P_x(X)$ such that if $P_x(y) > P_x(x), y \in D$ then $R(y) > R(x)$

  For this is enough to set $P_x(X) = S_1(X) - R(x)S_2(X)$

  Indeed, it is easy to see that $P_x(x) = 0$ and therefore

$$if\ \ P_x(y) > P_x(x) = 0\ then\ \ R(y) > R(x)$$

$$P_x(y) = S_1(y) - R\ (x)S_2(y) > 0 = P_x(x)\ \ then\ \ R(y) = \frac{S_1(y)}{S_2(y)} > R(x)$$

- Now suppose that for polynomial $P_x(X), x \in D$, we could construct a growth transformation $T_x\ of\ D$, such that $P_x(T_x(y)) > P_x(y)$ for any $y \in D$ unless $y = T_x(y)$

  then we could define a growth transformation T of D for $R(X)$ as follows

$$T(y) = T_y(y)$$

  for any $y \in D$ (the fact the T is a growth transformation would follow from the fact that $R(T(y)) > R(y)\ if\ \ P_y(T_y(y)) > P_y(y)$)

Reduction of the Case of Rational Functions to Polynomials Step.2

- Step2:

  we show that for any polynomial $P_y(X)$ there exists a polynomial $P_y'(X)$ with nonnegative coefficients such that any growth transformation of D for the $P_y'(X)$ is also a growth transformation for $P_y(X)$.

Lemma : Let $P(X) = P(\{X_{ij}\})$ be a polynomial with real coefficients in variable $X_{ij}, i = 1 \dots p, j = 1 \dots q_i$

Let domain D be $x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1$

(a) there exist a polynomial $C(X)$ such that the polynomail $P'(X) = P(X) + C(X)$ has only nonnegative coefficients and such that the value $C(x)$ at any $x \in D$ is a constant

(b) the set of growth transformations of D for $P(X)$ coincide with the set of growth transformations for $P'(X)$

Step 2 Lemma proof

- Let d be the degree P(X) and a its minimal negative coefficient (a=0 if no negative coefficient exists)

  Then a polynomial,

$$C(X) = -a\left(\sum_{i=1,j=1}^{p,q} x_{ij} + 1\right)^d$$

  is constant in D (its value in D is equal to $-a(p+1)^d$ since $\sum_{j=1}^{q_i} X_{ij} = 1$)

  Now, every possible monomial in P(X) also occurs in C(X). Since $a \neq 0$ is the smallest negative coefficient in P(X), it is easily seen that the sum of the coefficient of corresponding monomials in P(X) and C(X) is nonnegative.

- Since P'(X) and P(X) differ only by a constant in D, it is clear that P'(y)>P'(x) for any $y, x \in D$ if and only if P(y)>P(x).

Reduction of the Case of Rational Functions to Polynomials Step.3

- Step3: finding a growth transformation for a polynomial with nonnegative coefficients can be reduce to the same problem for a homogeneous polynomial with nonnegative coefficients.

$$\textit{consider the homogeneous polynomial } P''(Y) = P''(\{Y_{lm}\}) = Y_{p+1,1}^{d} \cdot P'(\{\underbrace{Y_{ij}/Y_{p+1,1}}_{1}\})$$

$$\textit{in variables } Y_{lm}, l = 1 \ldots p+1, m = 1 \ldots q_i \textit{ where } q_{p+1} = 1$$

$$D' : \sum_{j=1}^{q_i} y_{ij} = 1,\ y_{ij} > 0,\ i = 1 \ldots p+1,\ j = 1 \ldots q_i$$

$(P''(Y), D') \equiv (P'(Y), D)$

bijection $f : D \to D'$, mapping $x = \{x_{ij}\} \in D$ into $x' = \{y_{\ln}\}$

such that $x_{ij} = y_{ij}$ for $(i, j) \neq (p+1, 1)$

and such that for any $x \in D$, $P'(x) = P''(f(x))$

Thus, if T is a growth transformation for transformation for P''
Then the composition of the maps $f^{-1} \cdot T \cdot f$ is a growth transformation For P'.

An Inequality for A Polynomial

- Baum-Eagon inequality (L. E. Baum, J. I. Eagon)

  Let $P(X)=P(\{X_{ij}\})$ be a polynomial with nonnegative coefficients homogeneous of degree d in its variables X_{ij} . Let $x=\{x_{ij}\}$ be any point of the domain $D: x_{ij} \geq 0, \sum_{j=1}^{q_i} x_{ij} = 1,\ i=1,\cdots,p,\ j=1,\cdots,q_i$,such that

$$\sum_{j=1}^{q_i} x_{ij} \frac{\partial P(x_{ij})}{\partial x_{ij}} \neq 0, \qquad \text{for all i,}$$

$$\text{where } \frac{\partial P}{\partial \mathrm{X}}(x) \textit{ denotes the value of } \frac{\partial P(X)}{\partial X} \textit{ at } x.$$

Let $y=T(x)=T(\{x_{ij}\})$ denote the point of D whose i,j coordinate is

$$y_{ij} = \frac{x_{ij} \dfrac{\partial P(x_{ij})}{\partial x_{ij}}}{\sum_{j=1}^{q_i} x_{ij} \dfrac{\partial P(x_{ij})}{\partial x_{ij}}}$$

Then $P(T(x)) > P(x)$ unless $T(x)=x$

An Inequality for A Polynomial

- Let P be possibly nonhomogeneous:

$$Q(\{X_{ij}, Y_{p+1,1}\}) = Y_{p+1,1}^{d} + P''(\{X_{ij}, Y_{p+1,1}\})$$

and

$$P''(\{X_{ij}, Y_{p+1,1}\}) = Y_{p+1,1}^{d} \cdot P(\{X_{ij} / Y_{p+1,1}\})$$

- Let $f : D \rightarrow D'$

$$\sum_{j=1}^{q_i} x'_{ij} \frac{\partial Q}{\partial X_{ij}}(x'_{ij}) \neq 0, \qquad \textit{for all } i$$

implies that

$$\sum_{j=1}^{q_i} x_{ij} \frac{\partial P}{\partial X_{ij}}(x_{ij}) \neq 0, \qquad \textit{for all } i$$ and $$\frac{\partial Q}{\partial Y_{p+1,1}}(x'_{ij}) \neq 0$$

- Using the facts $\partial Q / \partial X_{ij}(x') = \partial P / \partial X_{ij}(x)$,Q differs from P'' in D' only by a constant(=1), and (P'',D') is the equivalent representation of (P,D)

Growth Transformations For Rational Functions

- Keep the notation and give some new definitions. Given a rational function $R(X)$ and a real number C let

$$\Gamma_{ij}(x;C)=x_{ij}\left(\frac{\partial P_x}{\partial X_{ij}}+C\right)$$

and

$$\Gamma_i(x;C)=\sum_{j=1}^{q_i}\Gamma_i(x;C).$$

- We call C admissible (for $R(X)$) if for any I, j, and $x\in D$, $\Gamma_{ij}(x;C)\geq 0$ and $\Gamma_i(x;C)>0$ .For any admissible C we consider the transformation T^C of D whose ij coordinate is defined as

$$(T^C(x))_{ij}=\frac{\Gamma_{ij}(x;C)}{\Gamma_i(x;C)}.$$

Growth Transformations For Rational Functions

- Theorem 4.1: Let $R(X)$ be a rational function in variables X_{ij} .There exists a constant N_R such that any $C \geq N_R$ is admissible for $R(X)$ and for any such C a map T^C is a growth transformation of D for $R(X)$.

- Proof: From step 2) it follows that for any $x \in D$ there exists a constant N_x such that the following polynomial has only nonnegative coefficients- $P_x(X)+C_x(X)$ where

$$C_x(X) = N_x \left(\sum_{i=1,j=1}^{p,q_i} X_{ij} + 1 \atop x \in D \right)^d$$

  and d is the degree of P_x .Since by assumption $R(X)$ has no poles in the compact set D one can find $N \geq N_x$ for all $x \in D$ such that for any $C \geq N$ multiplied by a factor (that will be derived below), the transformation T_x^C of D constructed from the polynomial $P_x(X)+C_x(X)$ via (6) is a growth transformation for it. For this C a family of growth transformation for $R(X)$ can be constructed following (2):

$$T^C(x) = T_x^C(x)$$

Growth Transformations For Rational Functions

- In order to compute explicitly these growth transformation let us note that the polynomial

$$\frac{\partial\left(\sum_{i=1,j=1}^{p,q_i} X_{ij}+1\right)^{d}}{\partial X_{ij}} = d\left(\sum_{i=1,j=1}^{p,q_i} X_{ij}+1\right)^{d-1}$$

is equal to $d(p+1)^{d-1}$ at any point in D. Combining this remark with the formulas (6) and (16),we have

$$\left(T^{C}(x)\right)_{ij} = (T_x^{\,C}(x))_{ij}$$

$$= \frac{x_{ij}\left(\frac{\partial P_x}{\partial X_{ij}}(x)+C\right)}{\sum_{j=1}^{q_i} x_{ij}\left(\frac{\partial P_x}{\partial X_{ij}}(x)+C\right)}$$

where $C \geq N_R = Nd(p+1)^{d-1}$ and $N = max_x\, N_x$ .This completes the proof of the theorem.

Gopolakrishnan's Theorem (1991)

- $R(\Lambda)$ is a rational function of polynomials in Λ_{ij}.Then there exists a a_R such that for $C > a_R$,the following function $T^C()$ is a growth transformation in D for R:

$$(T^C(\lambda))_{ij} = \frac{\lambda_{ij}\left(\frac{\partial P(\lambda_{ij})}{\partial \lambda_{ij}} + C\right)}{\sum_{j=1}^{q_i} \lambda_{ij}\left(\frac{\partial P(\lambda_{ij})}{\partial \lambda_{ij}} + C\right)}$$

Here $a_R = ad(p+1)^{d\text{-}1}$, $a = max_\lambda \, a_\lambda$, and a_λ is the minimal negative coefficient for all polynomials over all λ .

Example

- consider

$$R(x,y,z)=\frac{x^2}{x^2+y^2+z^2}\qquad x,y,z\geq 0\quad x+y+z=1$$

1. start from some x_0, y_0, z_0 such that $x_0>0, y_0>0, z_0>0,\quad x_0+y_0+z_0=1$
   iteration index $\mathrm{i}=0$
2. $k \leftarrow R(x_i, y_i, z_i)$
3. obtain a polynomial with nonegative coefficient
   $P(x,y,z)=x^2-k(x^2+y^2+z^2)+k(\underline{x+y+z})^2$ C
4. using update formula
   let $D \leftarrow x\partial P(x,y,z)/\partial x + y\partial P(x,y,z)/\partial y + z\partial P(x,y,z)/\partial z$
   $= 2x^2+4kxy+4kxz+4kyz$

$$x_{i+1}\leftarrow\left[\frac{x\dfrac{\partial P(x,y,z)}{\partial x}}{D}\right]_{x_i,y_i z_i}\qquad y_{i+1}\leftarrow\left[\frac{y\dfrac{\partial P(x,y,z)}{\partial y}}{D}\right]_{x_i,y_i z_i}\qquad z_{i+1}\leftarrow\left[\frac{z\dfrac{\partial P(x,y,z)}{\partial z}}{D}\right]_{x_i,y_i z_i}$$

5. increment iteration index i by 1 and go to 2.

Simple Example

- About step.3

$$P'(x,y) = x^2 + xy + x + y + 1$$

$$let\ \ x' = x/z, y' = y/z$$

$$then\ \ P'' = z^2 P'(x', y') = x^2 + xy + zx + zy + z^2$$