Acta Univ. M. Belii Math.
No. 12(2005), pp. 3–11

ON THE CONVERGENCE OF AVERAGE PRODUCTIVITY OF LABOUR AMONG ECONOMIES

Andrej Balciar

Abstract. This article deals with the various approaches to the verification of absolute convergence and conditional convergence of average productivity of labour among economies. In our contribution hypothesis of absolute and conditional convergence is originally formulated and original verification of the both convergences within the augmented Solow-Swan model is conducted.

In the paper two concepts of the convergence of average productivity of labour are distinguished: absolute and conditional. The hypothesis that poor economies tend to grow faster per capita than rich ones - without conditioning on any other characteristics of economies - is referred to as absolute convergence. If steady states differ, then we consider a concept of conditional convergence ([1]).

The concept of productivity convergence was derived from the Solow-Swan neoclassical growth model (Solow, 1956; Swan, 1956). This model examines stability of GDP per capita around the steady state assuming exogenously given constant growth rate of population. In the following years the model was augmented by the assumption of the constant rate of depreciation of capital goods within economy and exogenous technical progress.

We will consider the augmented Solow-Swan model of economic growth given by the following conditions:

$$I = i.Y, i > 0, \tag{1}$$

$$I = K' + \delta K, \delta > 0, \tag{2}$$

$$Y = F\left(K, A\left(t\right).L\right), \tag{3}$$

$$\frac{L'}{L} = n, n > 0, \tag{4}$$

where Y - production, I - investments, K - capital, L - labour, $A\left(t\right) = e^{x.t}$ - rate of technological progress, i - marginal rate to investment, δ - depreciation rate, α - elasticity of change in capital with respect to change in production, n - the growth rate of labour, $' = \frac{d}{dt}$, t - time.

2000 Mathematics Subject Classification. 91B62.

Key words and phrases. absolute convergence, average productivity of labour, capital-labour ratio, conditional convergence.

Submitted: June 13, 2005

Production function $F(.)$ is homogeneous of the first degree and satisfies the following conditions:

$$\frac{\partial F}{\partial L} > 0, \frac{\partial^2 F}{\partial L^2} < 0, \frac{\partial F}{\partial K} > 0, \frac{\partial^2 F}{\partial K^2} < 0 \cdot \tag{5}$$

It is also assumed that $F(.)$ satisfies the Inada conditions:

$$\lim_{K\to 0} F_K = \lim_{L\to 0} F_L = \infty, \lim_{K\to\infty} F_K = \lim_{L\to\infty} F_L = 0. \tag{6}$$

We will work with the variable average productivity of labour defined by $y = \frac{Y}{L}$ and capital-labour ratio defined by $k = \frac{K}{L}$. If we differentiate $K = k.L$ with respect to time and utilize (4), we obtain

$$K' = k'L + k.nL. \tag{7}$$

Putting (1) and (3) into (2) the condition for change in capital stock can be written in the formula:

$$K' = i.F(K, A(t).L) - \delta K. \tag{8}$$

If we utilize $K = k.L$ and apply the fact that $F(.)$ is a homogeneous function of the first degree, we can rewrite (8):

$$K' = i.L.F(k, A(t)) - \delta.k. \tag{9}$$

Comparing right-hand sides of (7) and (8) we obtain an expression for the growth rate of capital-labour ratio:

$$\frac{k'}{k} = \frac{i.F(k, A(t))}{k} - (n+\delta). \tag{10}$$

Because $F(.)$ is a homogeneous function of the first degree, new formula for average productivity of labour can be derived:

$$y = \frac{Y}{L} = \frac{F(K, A(t).L)}{L} = F(k, A(t)). \tag{11}$$

Effective labour is defined by $\hat{L} \equiv A(t).L$. R. J. Barro and X. X. Sala-i-Martin ([1]) work with the variable capital per effective labour defined by $\hat{k} \equiv k/A(t)$ and with the variable average productivity of effective labour defined by $\hat{y} \equiv y/A(t)$. Because $F(.)$ is a homogeneous function of the first degree, (11) can be re-written in the following formula:

$$\hat{y} = F\left(\hat{k}, 1\right) \equiv f\left(\hat{k}\right), \tag{12}$$

where $f(.)$ is the intensive production function. Putting $\hat{k} = k/A(t)$, (11) and (12) into (10) we have:

$$\frac{\hat{k}'}{\hat{k}} = \frac{i.f\left(\hat{k}\right)}{\hat{k}} - (x+n+\delta). \tag{13}$$

We will consider a Cobb-Douglas production function:

$$F(K, A(t).L) = K^{\alpha}.[A(t).L]^{1-\alpha}.$$

Utilizing Cobb-Douglas production function in (11) and (12) we obtain $f\left(\hat{k}\right) = \hat{k}^{\alpha}$, i.e. $\hat{y} = \hat{k}^{\alpha}$. Differentiating of $\hat{y} = \hat{k}^{\alpha}$ with respect to time we can write:

$$\frac{\hat{y}'}{\hat{y}} = \alpha.\frac{\hat{k}'}{\hat{k}}. \tag{14}$$

In [1] it is shown that $i.\frac{f(\hat{k})}{\hat{k}}$ is a decreasing function of time. As a consequence, according to (13) and (14) $\hat{y}'/\hat{y}$ and $\hat{k}'/\hat{k}$ are also decreasing functions of time. Utilizing these information hypotheses of absolute convergence and conditional convergence is verified on the basis of graphical analysis (see [1]).

G. Gandolfo ([2]) verified hypothesis of absolute convergence within the simple Solow-Swan model of economic growth (technological progress $A(t)$ is not considered). He proved that

$$\frac{\partial(y'/y)}{\partial y_0} < 0, \tag{15}$$

where y_0 is the initial level of average productivity of labour. The growth rate of average productivity of labour is inversely related to the initial level of average productivity of labour.

We will formulate hypothesis of absolute convergence and hypothesis of conditional convergence of average productivity of labour among economies precisely and find conditions of their validity. We will consider a Cobb-Douglas production function

$$F(K, A(t).L) = e^{\mu.t}K^{\alpha}L^{1-\alpha},$$

where $\mu \equiv x.(1-\alpha)$ is a measure of technological progress. Because $F(.)$ is a homogeneous function of the first degree, new formula for average productivity of labour can be derived:

$$y = \frac{Y}{L} = \frac{F(K, A(t).L)}{L} = F(k, A(t)) = e^{\mu.t}.k^{\alpha}. \tag{16}$$

We will distingusih 2 types of economies. We say that economy is less developed economy if it has lower initial level of capital-labour ratio. Developed economy is an economy with higher initial level of capital-labour ratio. In the paper we will denote less developed economies by subindex 1 and developed economies by subindex 2.

Hypothesis of Absolute Convergence of Average Productivity of Labour among Economies

Consider two groups of economies with the same marginal rate to investment, i, the same depreciation rate, δ, the same growth rate of labour force, n, but different initial levels of capital-labour ratio, $k_{0,1} < k_{0,2}$. Under the certain circumstances the difference between average productivity of labour of less developed economies and average productivity of labour of developed economies is a decreasing function of time and it converges to zero.

We have just stated hypothesis of absolute convergence of average productivity of labour among economies. Putting (16) into (10) and multiplying the both sides of this equation by k, we obtain a basic differential equation of the Sollow-Swan model:

$$k' + (n+\delta)\,k = ie^{\mu t}k^{\alpha}. \tag{17}$$

Equation (17) is a Bernoulli differential equation. Utilizing $q = r^{1-\alpha}$ we can transfer it into a homogeneous linear differential equation:

$$q' + (n+\delta)\,(1-\alpha)\,q = ie^{\mu t}\,(1-\alpha)\,. \tag{18}$$

We can find its particular solution in the form $q_p = ae^{\mu t}$:

$$q_p = \frac{i\,(1-\alpha)}{\mu + (n+\delta)\,(1-\alpha)}e^{\mu t}. \tag{19}$$

The solution of equation (17) given by the initial condition: $q(0) = q_0$ is:

$$q\,(t) = \left(q_0 - \frac{i\,(1-\alpha)}{\mu + (n+\delta)\,(1-\alpha)}\right)e^{-(n+\delta)(1-\alpha)t} + \frac{i\,(1-\alpha)}{\mu + (n+\delta)\,(1-\alpha)}e^{\mu t}. \tag{20}$$

Putting $k = q^{1/(1-\alpha)}$, $k\,(0) = k_0$ and $q_0 = k_0^{1-\alpha}$ into (20) we can come back to variable $k\,(t)$:

$$k\,(t) = \left\{\left(k_0^{1-\alpha} - \frac{i\,(1-\alpha)}{\mu + (n+\delta)\,(1-\alpha)}\right)e^{-(n+\delta)(1-\alpha)t} + \right.$$

$$\left. + \frac{i\,(1-\alpha)}{\mu + (n+\delta)\,(1-\alpha)}e^{\mu t}\right\}^{\frac{1}{1-\alpha}}. \tag{21}$$

We will utilize the following denotations: $A = \frac{\mu}{1-\alpha}$, $B_1 = k_{0,1}^{1-\alpha} - \frac{i(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, $B_2 = k_{0,2}^{1-\alpha} - \frac{i(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, $C = (n+\delta)\,(1-\alpha) + \mu$, $D = \frac{i(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, $E = B_1e^{-Ct} + D$ and $F = B_2e^{-Ct} + D$, where A, B_1, B_2, C, D, E and F are positive constants. Now, formula (22) can be written for less developed and for developed economies:

$$y_1\,(t) = e^{At}\left(B_1e^{-Ct} + D\right)^{\frac{\alpha}{1-\alpha}}, \tag{23}$$

$$y_2\,(t) = e^{At}\left(B_2e^{-Ct} + D\right)^{\frac{\alpha}{1-\alpha}}. \tag{24}$$

Because $k_{0,1} < k_{0,2}$, it holds: $B_1 < B_2$. Comparing (23) and (24) under assumption $B_1 < B_2$ we find out that average productivity of labour of less developed economies is in every moment lower than average productivity of labour of developed ones:

$$\forall t \geqslant 0; y_1\,(t) < y_2\,(t)\,. \tag{25}$$

Utilizing (23) and (24) we can count limits:

$$\lim_{t\to\infty}\frac{y_1\,(t)}{y_2\,(t)} = \lim_{t\to\infty}\frac{\left(B_1e^{-Ct} + D\right)^{\frac{\alpha}{1-\alpha}}}{\left(B_2e^{-Ct} + D\right)^{\frac{\alpha}{1-\alpha}}} = \frac{D^{\frac{\alpha}{1-\alpha}}}{D^{\frac{\alpha}{1-\alpha}}} = 1, \tag{26}$$

$$\lim_{t\to\infty} y_1(t) = \infty, \tag{27}$$

$$\lim_{t\to\infty} y_2(t) = \infty. \tag{28}$$

We are going to count a limit:

$$\lim_{t\to\infty} (y_1(t) - y_2(t)),$$

which is $\infty - \infty$ type. We will transform it into the limit of $\frac{0}{0}$ type:

$$\lim_{t\to\infty} (y_1(t) - y_2(t)) = \lim_{t\to\infty} \frac{\frac{y_1}{y_2} - 1}{\frac{1}{y_2}}. \tag{29}$$

Because

$$\left(\frac{y_1}{y_2} - 1\right)' = \frac{\alpha}{1-\alpha} \frac{E^{\frac{2\alpha-1}{1-\alpha}} Ce^{-Ct} (B_2 E - B_1 F)}{F^{\frac{1}{1-\alpha}}}, \tag{30}$$

$$\left(\frac{1}{y_2}\right)' = -e^{-At} F^{\frac{1}{\alpha-1}} \left(AF - \frac{\alpha}{1-\alpha} B_2 Ce^{-Ct}\right), \tag{31}$$

$\lim_{t\to\infty} \frac{(y_1/y_2-1)'}{(1/y_2)'}$ exists. Utilizing l'Hospital rule, (30) and (31) we will continue counting the limit (29):

$$\lim_{t\to\infty} (y_1(t) - y_2(t)) = \lim_{t\to\infty} \frac{\left(\frac{y_1}{y_2} - 1\right)'}{\left(\frac{1}{y_2}\right)'} =$$

$$= \lim_{t\to\infty} \frac{-\alpha}{1-\alpha} \frac{\left(B_1 e^{-Ct} + D\right)^{\frac{2\alpha-1}{1-\alpha}} D (B_2 - B_1) Ce^{(A-C)t}}{\left(AF - \frac{\alpha}{1-\alpha} B_2 Ce^{-Ct}\right)}. \tag{32}$$

We can distinguish three cases:

1. If $\frac{\mu\alpha}{1-\alpha} - (n+\delta)(1-\alpha) > 0$, then

$$\lim_{t\to\infty} (y_1(t) - y_2(t)) = -\infty.$$

In this case average productivity of labour of developed economies straggles with increasing time from average productivity of labour of less developed ones (Fig. 1).

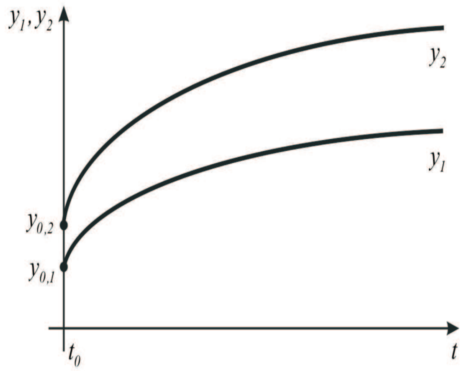

Fig. 1

2. If $\frac{\mu\alpha}{1-\alpha} - (n+\delta)(1-\alpha) = 0$ then

$$\lim_{t\to\infty}(y_1(t) - y_2(t)) = -\frac{\alpha}{\mu}\left(\frac{i(1-\alpha)}{\mu + (n+\delta)(1-\alpha)}\right)^{\frac{2\alpha-1}{1-\alpha}}\left(k_{0,2}^{1-\alpha} - k_{0,1}^{1-\alpha}\right) < 0.$$

It implies that average productivity of labour of less developed economies will not achieve average productivity of labour of developed ones over time (Fig. 2).

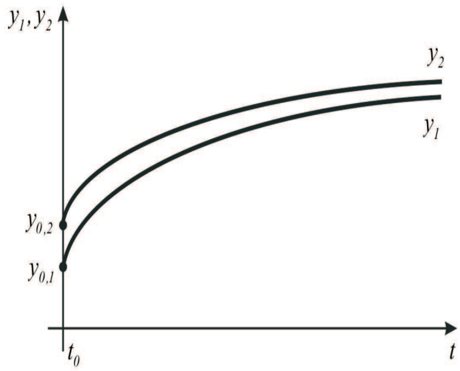

Fig. 2

3. If $\frac{\mu\alpha}{1-\alpha} - (n+\delta)(1-\alpha) < 0$ then

$$\lim_{t\to\infty}(y_1(t) - y_2(t)) = 0.$$

In this case the difference between average productivity of labour of less developed economies and average productivity of labour of developed ones converges to zero over time (Fig. 3).

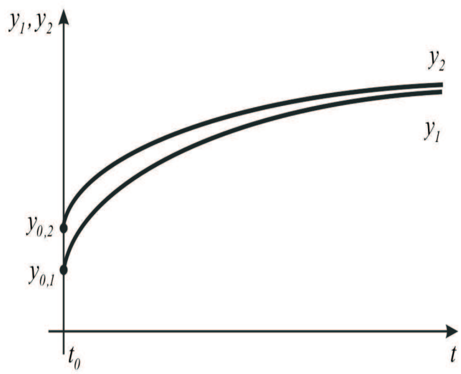

Fig. 3

On the basis of the above analysis we can formulate the following theorem.

Theorem 1. Consider two groups of economies with the same marginal rate to investment, i, the same depreciation rate, δ, the same growth rate of labour force, n, but different initial levels of capital-labour ratio, $k_{0,1} < k_{0,2}$, within the augmented Solow-Swan model given by the conditions (1) - (6), (16). Then:

1. If $\frac{\mu\alpha}{1-\alpha} - (n+\delta)(1-\alpha) = 0$, then $\lim_{t\to\infty}(y_1(t) - y_2(t)) = -\infty$.
2. If $\frac{\mu\alpha}{1-\alpha} - (n+\delta)(1-\alpha) = 0$, then $\lim_{t\to\infty}(y_1(t) - y_2(t)) < 0$.
3. If $\frac{\mu\alpha}{1-\alpha} - (n+\delta)(1-\alpha) < 0$, then $\lim_{t\to\infty}(y_1(t) - y_2(t)) = 0$.

Hypothesis of Conditional Convergence of Average Productivity of Labour among Economies

Consider two groups of economies with the same depreciation rate, δ, the same growth rate of labour, n, but different marginal rate to investment, $i_1 \neq i_2$, and the initial levels of capital-labour ratio, $k_{0,1} < k_{0,2}$. Then: if marginal rate to investment, i_1, in less developed economies exceeds marginal rate to investment, i_2, in developed ones, average productivity of labour of less developed economies will reach average productivity of labour of developed ones.

We have just stated hypothesis of conditional convergence of average productivity of labour among economies. We will utilize the following denotations: $G_1 = k_{0,1}^{1-\alpha} - \frac{i_1(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, $G_2 = k_{0,2}^{1-\alpha} - \frac{i_2(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, $H_1 = \frac{i_1(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, $H_2 = \frac{i_2(1-\alpha)}{\mu+(1-\alpha)(n+\delta)}$, where G_1, G_2, H_1, H_1 and H_2 are positive constants. Applying these denotations in (22) we obtain:

$$y_1(t) = e^{At}\left(G_1 e^{-Ct} + H_1\right)^{\frac{\alpha}{1-\alpha}}, \tag{33}$$

$$y_2(t) = e^{At}\left(G_2 e^{-Ct} + H_2\right)^{\frac{\alpha}{1-\alpha}}. \tag{34}$$

We are going to count a limit $\lim_{t\to\infty}(y_1(t) - y_2(t))$. We will distinguish two cases related to marginal rate to investment.

1. If $i_1 > i_2$, then from (33) and (34) we have:

$$\lim_{t\to\infty}(y_1(t) - y_2(t)) = \infty. \tag{35}$$

Under the assumption that less developed economies invest more in relation to production than developed ones ($i_1 > i_2$) according to (33) and (34) it holds: $y_{0,1} < y_{0,2}$ and as a consequence of (35) average productivity of labour of less developed economies must reach average productivity of labour of developed ones. We will find a time t^* when it happens. Equality $y_1(t) = y_2(t)$ is equivalent to equality $y_1^{\frac{1-\alpha}{\alpha}}(t) = y_2^{\frac{1-\alpha}{\alpha}}(t)$. Utilizing (33) and (34) in this equality we obtain:

$$t^* = \frac{1}{(n+\delta)(1-\alpha)+\mu}\ln\frac{k_{0,2}^{1-\alpha} - k_{0,1}^{1-\alpha} + (i_1 - i_2)z}{(i_1 - i_2)z}, \tag{36}$$

where z denotes $\frac{1-\alpha}{\mu+(1-\alpha)(n+\delta)}$.

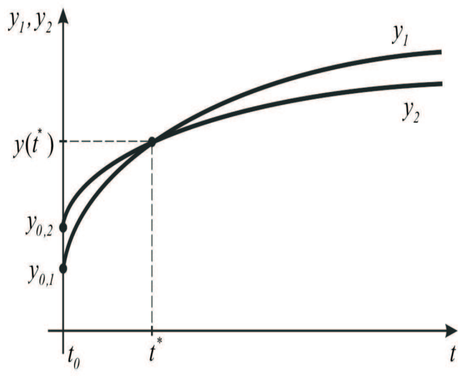

Fig. 4

Formula (36) determines a time, when average productivity of labour of less developed economies reaches average productivity of labour of developed ones (Fig. 4).

2. If $i_1 < i_2$, then from (33) and (34) we obtain:

$$\lim_{t\to\infty} (y_1(t) - y_2(t)) = -\infty. \tag{37}$$

If less developed economies invest less in relation to production than developed ones ($i_1 < i_2$), average productivity of labour of less developed economies will not reach average productivity of labour of developed ones. Under assumptions: $k_{0,1} < k_{0,2}$, $i_1 < i_2$, it holds: $G_1 < G_2, H_1 < H_2$. According to (33), (34) we find out that average productivity of labour of less developed economies is in every moment lower than average productivity of labour of developed ones:

$$\forall t \geqslant 0; y_1(t) < y_2(t). \tag{38}$$

According to (37) and (38) the difference between average productivity of labour of less developed economies and average productivity of labour of developed ones is an increasing function of time (Fig. 5).

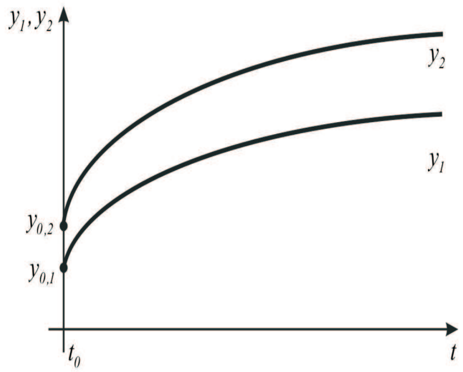

Fig. 5

On the basis of the given analysis we can formulate the following theorem.

Theorem 2. Consider two groups of economies with the same depreciation rate, δ, the same growth rate of labour, n, but different marginal rate to investment, $i_1 \neq i_2$, and different initial levels of capital-labour ratio, $k_{0,1} < k_{0,2}$, within the augmented Solow-Swan model given by the conditions (1) - (6), (16). Then:

1. If $i_1 > i_2$, then $\lim_{t\to\infty} (y_1(t) - y_2(t)) = \infty$ and there exists a time t^*, determined by (36), in which $y_1(t^*) = y_2(t^*)$.
2. If $i_1 < i_2$, then $\lim_{t\to\infty} (y_1(t) - y_2(t)) = -\infty$.

Conclusion

First, we found out the sufficient condition for absolute convergence of the difference between average productivity of labour of less developed economies and average productivity of labour of developed ones to zero (see Theorem 1, part 3.). width Second, we found out a time, in which average productivity of labour of less developed economies reaches average productivity of labour of developed ones which proves conditional convergence of average productivity of labour among economies.

References

1. Barro, R. J. and Sala-i-Martin, X.X., *Economic Growth*, McGraw Hill, Inc., New York, 1997.
2. Gandolfo, G., *Economic Dynamics*, Springer, Berlin, 1995.

Department of Applied Informatics; Faculty of Economics; Matej Bel University; Tajovskeho 10; SK–975 90 Banska Bystrica; Slovakia

E-mail: andrej.balciar@fpv.umb.sk

Acta Univ. M. Belii Math.
No. 12(2005), pp. 13–18

ON THE STRUCTURE OF POLYHEDRAL GRAPHS WITH PRESCRIBED EDGE AND DUAL EDGE WEIGHT

Barbora Ferencová and Tomáš Madaras

Abstract. We consider families of polyhedral graphs with prescribed minimum vertex degree δ, minimum face degree ρ, minimum edge weight w and dual edge weight w^*. We determine all quadruples (δ, ρ, w, w^*) for which the asociated family is nonempty.

Introduction

Throughout this paper we consider connected plane graphs without loops or multiple edges. For a plane graph G, $V = V(G), E = E(G)$ and $F = F(G)$ denotes the set of its vertices, edges and faces, respectively. A *k-vertex* (*k-face*) will stand for a vertex (a face) of degree k, a $\geq$ *k-vertex*/$\leq$ *k-vertex* ($\geq$ *k-face*/$\leq$ *k-face*) for those of degree at least k/at most k. For an edge e being incident with an a-vertex and a b-vertex, and with a c-face and a d-face, the type of e is (a, b, c, d) where $a \leq b, c \leq d$. The weight $w(e)$ of an edge $e = uv$ is the sum $\deg_G(u) + \deg_G(v)$. The edge weight $w(G)$ of a plane graph G is equal to $\min_{uv \in E(G)} \{\deg_G(u) + \deg_G(v)\}$; the dual edge weight $w^*(G)$ of G is the edge weight of the dual of the graph G. Let $\mathcal{G}_c(\delta, \rho, w, w^*)$ be the family of all c-connected plane graphs with minimum vertex degree at least δ, minimum face degree at least ρ, edge weight at least w and dual edge weight at least w^*; for $c = 3$, we will use the notation $\mathcal{G}(\delta, \rho, w, w^*)$.

It is well known that every plane graph contains a vertex of degree at most 5. Among numerous generalizations of this result (see [6]), the fundamental role plays the Kotzig's theorem [7] stating that for each polyhedral graph G, $w(G) \leq 13$, and if G is of minimum degree at least 4, then $w(G) \leq 11$; both these bounds are sharp. Thus, the family of all polyhedral graphs with minimum edge weight at least 14 (and of the ones with minimum degree at least 4 and minimum edge weight at least 12) is empty. Considering the dual graphs, we obtain the analogical results for dual weight constraints. Note that the Kotzig's theorem provides no information about degrees of two faces incident with an edge that attains the minimum weight in a graph, but there are generalisations of theorem taking this aspect into account. For example, in [1] Borodin extended the Kotzig's result showing that each normal plane map (that is, a plane pseudograph having no vertices or faces of degree less than

2000 Mathematics Subject Classification. 05C10.
Key words and phrases. polyhedral graph, edge weight.
Supported in part by Slovak VEGA Grant 1/0424/03 and Slovak Grant APVT-20-004104.
Submitted: October 21, 2005

three) contains either a 3-face incident with an edge of weight at most 13, or a 4-face incident with an edge of weight at most 8, or else a 5-face incident with an edge of weight 6; all these bounds being sharp. Another partial results are contained in the classical paper [8]. It is also possible to consider some combinations of constraints based on minimum vertex degree, face degree, edge weight and dual edge weight; to our knowledges, the simultaneous combinations of all four mentioned constraints were not studied in deeper details.

The aim of this paper is give the complete characterization of quadruples (δ, ρ, w, w^*) for which the family $\mathcal{G}(\delta, \rho, w, w^*)$ is empty. The Euler theorem implies that $(4, 4, 8, 8)$ yields the empty family. From the Kotzig theorem, it follows that for quadruples $(3, 3, 14, 6), (4, 3, 12, 6)$, the corresponding families are empty, and from [L] follows the emptiness of families determined by quadruples $(4, 3, 8, 9)$ and $(3, 5, 7, 10)$. Using the duality, we get that the corresponding families are empty also for $(3, 3, 6, 14), (3, 4, 9, 8), (3, 4, 6, 12), (5, 3, 10, 7)$.

We prove

Theorem 1. The families $\mathcal{G}(3, 3, 7, 10), \mathcal{G}(3, 3, 8, 9), \mathcal{G}(3, 4, 7, 9)$ are empty.

In each of three cases of theorem, we proceed by contradiction, thus, assuming the non-emptiness of specified family, we consider its representant G with specified minimum vertex degree, face size, edge and dual edge weight. At this graph G, the discharging method is used. We define the charge $c:\ V \cup E \cup F \to \mathbb{Z}$ by the following assignments:

$$(\forall v \in V)\ c(v) = \deg_G(v) - 6$$
$$(\forall \alpha \in F)\ c(\alpha) = 2 \cdot \deg_G(\alpha) - 6$$
$$(\forall e \in E)\ c(e) = 0.$$

From the Euler Theorem, it follows that $\sum\limits_{x \in V \cup E \cup F} c(x) = -12.$

Next, we define the local redistribution of charges between the elements of G such that the total sum of charges remains the same. This is performed by certain rules which specify the charge transfers from elements to another elements in specific situations. After such redistribution, we obtain a new charge $\widetilde{c}:\ V \cup E \cup F \to \mathbb{Q}$. Then, we prove that for any element $x \in V \cup E \cup F$, $\widetilde{c}(x) \geq 0$ (hence, $\sum\limits_{x \in V \cup E \cup F} \widetilde{c}(x) \geq 0$). This contradiction shows that G cannot exist.

The family $\mathcal{G}(3, 3, 7, 10)$

The discharging rules are the following:

Rule 1: Each k-vertex x sends $\frac{c(x)}{k}$ to each incident edge.

Rule 2: Each k-face α, $k \notin \{7, 8, 9, 10, 11\}$ sends $\frac{c(\alpha)}{k}$ to each incident edge.

Rule 3: Each l-face β, $l \in \{7, 8, 9, 10, 11\}$ sends to each incident edge e of type (a, b, c, d) the following charge:

(a) $\frac{3}{2}$ if $a = 3, 3 \leq c \leq 4$,
(b) $\frac{4}{5}$ if $a = 3, c = 5$,
(c) $\frac{3}{4}$ if $a = 3, c \geq 6$,
(d) 1 otherwise.

For proving the nonnegativity of final charges, firstly observe that all vertices and all k-faces, $k \notin \{7,8,9,10,11\}$ are discharged to zero. Now we analyze the final charge of the remaining faces and edges.

1. Let β be an l-face, $7 \le l \le 11$. Then β is incident with at most five 3-vertices (since $w(G) \ge 7$).

   (a) If β is not incident with a 3-vertex then $\widetilde{c}(\beta) \ge 2l - 6 - l \cdot 1 = l - 6 > 0$ by Rule 3(d).

   (b) Let β be incident with exactly t 3-vertices, $1 \le t \le 5$. Then $2t$ transfers from β are by Rule 3(a), (b), or (c), and the remaining ones are by Rule 3(d). Moreover, for transfers through a pair of edges of β with common 3-vertex, Rule 3(a) may be used only with Rule 3(c) (since $w^*(G) \ge 10$). From this fact we have that the maximum charge transferred from β is in the case when each of Rules 3(a) and 3(c) is used t times; then $\widetilde{c}(\beta) \ge 2l - 6 - t\frac{3}{2} - t\frac{3}{4} - (l - 2t) \cdot 1 = l - 6 - \frac{t}{4}$. Hence, $\widetilde{c}(\beta) \ge 0$ for $t \ge 4$; for $t = 5$, we have $l \in \{10, 11\}$ and so $\widetilde{c}(\beta) > 0$.

2. Let e be an edge of G of the type (a, b, c, d); note that $a + b \ge 7$ since $w(G) \ge 7$.

   (a) If $a = 3, 3 \le c \le 4, 7 \le d \le 11$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + \frac{3}{2} = 0$ by Rules 1 and 3(a).

   (b) If $a = 3, 3 \le c \le 4, d \ge 12$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + \frac{2 \cdot d - 6}{d} \ge 0$ by Rules 1 and 2.

   (c) If $a = 3, c = 4, d = 6$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + \frac{1}{2} + \frac{2 \cdot 6 - 6}{6} = 0$ by Rules 1 and 2.

   (d) If $a = 3, c = 5, 5 \le d \le 6$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + \frac{4}{5} + \frac{2 \cdot d - 6}{d} > 0$ by Rules 1 and 2.

   (e) If $a = 3, c = 5, 7 \le d \le 11$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + \frac{4}{5} + \frac{4}{5} > 0$ by Rules 1, 2 and 3(b).

   (f) If $a = 3, c = 5, d \ge 12$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + \frac{4}{5} + \frac{2d - 6}{d} > 0$ by Rules 1 and 2.

   (g) If $a \ge 3, c \ge 6, d \ge 6$, then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + 1 + \frac{3}{4} > 0$ by Rules 1 and 2 (or 3(c) or 3(d)).

The family $\mathcal{G}(3, 3, 8, 9)$

The discharging rules are the following:

Rule 1: Each k-vertex x sends $\frac{c(x)}{k}$ to each incident edge.

Rule 2: Each k-face α, $k \notin \{6, 7\}$ sends $\frac{c(\alpha)}{k}$ to each incident edge.

Rule 3: Each k-face β, $k \in \{6, 7\}$ sends to each incident edge e of type (a, b, c, d) the following charge:

(a) $\frac{6}{5}$ if $a = 3, b \ge 5, 3 \le c \le 4$,
(b) 1 if $a \ge 4, b \ge 4, c \ge 3$,
(c) $\frac{3}{5}$ if $a = 3, b \ge 5, c \ge 5$.

Like in previous proof, all vertices and all k-faces, $k \notin \{6, 7\}$ are discharged to zero. Now we analyze the final charge of the remaining faces and edges.]

1. Let β be an l-face, $6 \le l \le 7$. Then β is incident with at most three 3-vertices (since $w(G) \ge 8$).

   (a) If β is not incident with a 3-vertex then $\widetilde{c}(\beta) \ge 2l - 6 - l \cdot 1 = l - 6 \ge 0$ by Rule 3(b).

(b) Let β be incident with exactly t 3-vertices, $1 \le t \le 3$. Like in previous proof, $2t$ transfers from β are by Rule 3(a) or (c), and the remaining ones are by Rule 3(b). Again, for transfers through a pair of edges of β with common 3-vertex, Rule 3(a) may be used only with Rule 3(c) (since $w^*(G) \ge 9$); this yields that the maximum charge transferred from β is in the case when each of Rules 3(a) and 3(c) is used t times. Hence, $\widetilde{c}(\beta) \ge 2l - 6 - \frac{6}{5}t - \frac{3}{5}t - (l - 2t) \cdot 1 = l - 6 + \frac{t}{5} > 0$.

2. Let e be an edge of G of the type (a, b, c, d); as $w(G) \ge 8$, we have $a + b \ge 8$.
   (a) If $a = 3, b \ge 5, 4 \le c \le 5, d = 5$ then $\widetilde{c}(e) \ge -1 - \frac{1}{5} + \frac{4}{5} + \frac{1}{2} = \frac{1}{10} > 0$ by Rules 1 and 2.
   (b) If $a \ge 4, b \ge 4, c \ge 4, d \ge 5$ then $\widetilde{c}(e) \ge 2 \cdot \left(-\frac{1}{2}\right) + \frac{4}{5} + \frac{1}{2} = \frac{3}{10} > 0$ by Rules 1 and 2.
   (c) If $a = 3, b \ge 5, 3 \le c \le 4, 6 \le d \le 7$ then $\widetilde{c}(e) \ge -1 - \frac{1}{5} + \frac{6}{5} = 0$ by Rules 1 and 3(a).
   (d) If $a \ge 4, b \ge 4, c \ge 3, 6 \le d \le 7$ then $\widetilde{c}(e) \ge 2 \cdot \left(-\frac{1}{2}\right) + 1 = 0$ by Rules 1 and 3(b).
   (e) If $a = 3, b \ge 5, c = 5, 6 \le d \le 7$ then $\widetilde{c}(e) \ge -1 - \frac{1}{5} + \frac{4}{5} + \frac{3}{5} = \frac{1}{5} > 0$ by Rules 1, 2 and 3(c).
   (f) If $a = 3, b \ge 5, 6 \le c \le 7, 6 \le d \le 7$ then $\widetilde{c}(e) \ge -1 - \frac{1}{5} + 2 \cdot \frac{3}{5} = 0$ by Rules 1 and 3(c).
   (g) If $a = 3, b \ge 5, c \ge 3, d \ge 8$ then $\widetilde{c}(e) \ge -1 - \frac{1}{5} + \frac{2 \cdot 8 - 6}{8} = \frac{1}{20} > 0$ by Rules 1 and 2.
   (h) If $a \ge 4, b \ge 4, c \ge 3, d \ge 8$ then $\widetilde{c}(e) \ge 2 \cdot \frac{-1}{2} + \frac{2 \cdot 8 - 6}{8} = \frac{1}{4} > 0$ by Rules 1 and 2.

The family $\mathcal{G}(3, 4, 7, 9)$

The discharging rules are the following:

Rule 1: Each k-vertex x sends $\frac{c(x)}{k}$ to each incident edge.

Rule 2: Each k-face α, $k \ne 5$ sends $\frac{c(\alpha)}{k}$ to each incident edge.

Rule 3: Each 5-face β sends to each incident edge e of type (a, b, c, d) the following charge:
   (a) 1 if $a = 3, b \ge 4, c = 4$,
   (b) $\frac{3}{4}$ if $a = 3, b \ge 4, c \ge 5$,
   (c) $\frac{1}{2}$ if $a \ge 4, b \ge 4, c \ge 4$.

All vertices and all faces except a 5-face are discharged to zero. Consider the final charge of 5-faces and edges:

1. Let β be a 5-face. If β is not incident with a 3-vertex then $\widetilde{c}(\beta) \ge 2 \cdot 5 - 6 - 5 \cdot \frac{1}{2} = \frac{3}{2} > 0$. Otherwise, β is incident with t 3-vertices, $1 \le t \le 2$. Again, due to the fact that $w^*(G) \ge 9$, for transfers through a pair of edges of β sharing common 3-vertex, Rule 3(a) may be used only in the combination with Rule 3(c). Thus, $\widetilde{c}(\beta) \ge 2 \cdot 5 - 6 - 1 \cdot t - \frac{3}{4} \cdot t - (5 - 2t)\frac{1}{2} = \frac{3}{2} - \frac{3t}{4} \ge 0$.

2. Let e be an edge of G of the type (a, b, c, d); as $w(G) \ge 7$, $a + b \ge 7$.
   (a) If $a = 3, b \ge 4, c = 4, d \ge 5$ then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + 1 + \frac{1}{2} = 0$ by Rules 1, 2 and 3(a).
   (b) If $a = 3, b \ge 4, c = 5, d \ge 5$ then $\widetilde{c}(e) \ge -1 - \frac{1}{2} + 2 \cdot \frac{3}{4} = 0$ by Rules 1 and 3(b).

(c) If $a \geq 4, b \geq 4, c \geq 4, d \geq 5$ then $\widetilde{c}(e) \geq 2 \cdot \left(-\frac{1}{2}\right) + 2 \cdot \frac{1}{2} = 0$ by Rules 1 and 3(c).

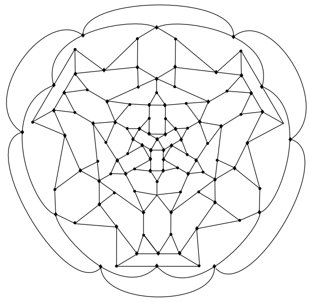

Fig. 1

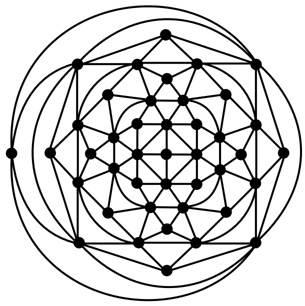

Fig. 2

Concluding remarks

1. Concerning the quadruples $(3,3,6,13)$, $(3,3,7,9)$, $(3,3,8,8)$, $(3,4,6,11)$, $(3,4,7,8)$ and $(3,5,6,11)$ (and the quadruples derived from them by swapping the first entry with the second one, and the third with the fourth one), it is easy to show that the corresponding families are nonempty (and, in fact, infinite); the examples are: the truncated dodecahedron, the graph of Fig. 1, the icosidodecahedron, the dual of the graph of Fig. 2, the rhombic dodecahedron and the truncated icosahedron (for the names of these polyhedra, see [9]). In this sense, our results are best possible. The diagram on Fig. 3 presents the hierarchy of all nonempty families generated by quadruples (δ, ρ, w, w^*) under the set inclusion partial ordering.

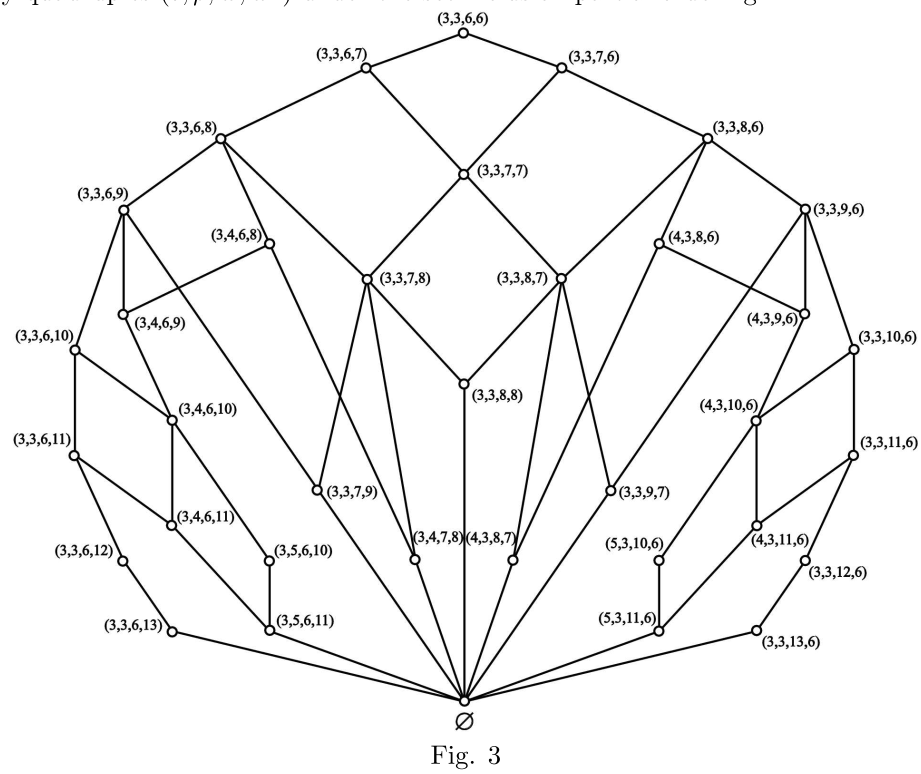

Fig. 3

2. The Kotzig theorem was further generalized in many ways, one of which considered, for the specified family of polyhedral graphs, the existence of longer paths with the degrees of their vertices bounded above by a finite constant that depends only of the specified family and of path length; such paths are called light. For the related results regarding various families, see [2], [3], [4] or [5]. In the connection with the results presented in this paper, one may consider, for given integer $k \geq 1$, the families of polyhedral graphs with prescribed weight of i-paths and dual i-paths for all $i \in \{1, \ldots, k\}$. To our knowledge, currently there are no results involving both normal and dual weights for $k \geq 3$.

References

1. O.V. Borodin, *Joint generalization of the Lebesgue and Kotzig theorems on combinatorics of plane maps*, Diskretn. Mat. **3** (1991), no. 4, 24–27. (Russian)
2. I. Fabrici, E. Hexel, S. Jendrol', H. Walther, *On vertex-degree restricted paths in polyhedral graphs*, Discrete Math. **212** (2000), 61–73.
3. I. Fabrici, S. Jendrol', *Subgraphs with restricted degrees of their vertices in planar 3-connected graphs*, Graphs and Combin. **13** (1997), 245–250.
4. J. Harant, S. Jendrol', M. Tkáč, *On 3-connected plane graphs without triangular faces*, J. Comb. Theory Ser. B **77** (1999), 150–161.
5. S. Jendrol', P. Owens, *On light graphs in* 3*-connected plane graphs without triangular or quadrangular faces*, Graphs and Combin. **17** (2001), no. 4, 659 – 680.
6. S. Jendrol', H.-J. Voss, *Light subgraphs of graphs embedded in the plane and in the projective plane – A survey*, Discrete Math (to appear).
7. A. Kotzig, *Contribution to the theory of Eulerian polyhedra*, Mat. Čas. SAV (Math. Slovaca) **5** (1955), 111–113.
8. H. Lebesgue, *Quelques consequences simples de la formule d'Euler*, J. Math. Pures Appl **19** (1940), 19–43.
9. Eric W. Weisstein., *"Uniform Polyhedron."*, From MathWorld–A Wolfram Web Resource., `http://mathworld.wolfram.com/UniformPolyhedron.html`.

Institute of Mathematics, Faculty of Sciences; University of P. J. Šafárik; Jesenná 5, SK–04154 Košice; Slovak Republic

E-mail: `ferencova@szm.sk`
`madaras@science.upjs.sk`

Acta Univ. M. Belii Math.
No. 12(2005), pp. 19–29

CONDITIONING BY RARE SOURCES

M. Grendár

To Mar, in memoriam.
To George Judge, on the occasion of his eightieth birthday.

Abstract. In this paper we study the exponential decay of posterior probability of a set of sources and conditioning by rare sources for both uniform and general prior distributions of sources. The decay rate is determined by L-divergence and rare sources from a convex, closed set asymptotically conditionally concentrate on an L-projection. L-projection on a linear family of sources belongs to Λ-family of distributions. The results parallel those of Large Deviations for Empirical Measures (Sanov's Theorem and Conditional Limit Theorem).

1. Introduction

Information divergence minimization, which is also known as Relative Entropy Maximization or MaxEnt method, has – thanks to Large Deviations Theorems for Empirical Measures – gained a firm probabilistic footing, which justifies its application in the area of the convex Boltzmann Jaynes Inverse Problem (the α-problem, for short). For the β-problem – an 'antipode' of the α-problem – Large Deviations Theorems for Sources single out the L-divergence minimization method.

The paper is organized as follows: First, necessary terminology and notation are introduced. A brief survey of Large Deviations Theorems for Empirical Measures that includes Sanov's Theorem and a Conditional Limit Theorem is given next. Then, a set-up for a study of conditioning by rare sources is formulated and Sanov's Theorem and the Conditional Limit Theorem for Sources are stated; under various assumptions. Next, Theorems are proven for the continuous case and the results are applied to a criterion choice problem associated with the β-problem. An End-Notes section points to relevant literature and contains further discussion.

2000 Mathematics Subject Classification. Primary 60F10; Secondary 94A15.

Key words and phrases. Conditional Limit Theorems, Information projection, Reverse information projection, L-projection, Kerridge's inaccuracy, Watanabe's confirmability, Large Deviations for Sources, Maximum Non-parametric Likelihood, Ill-posed inverse problems, Criterion Choice Problem, Bayesian nonparametric consistency.

Supported in part by Slovak VEGA Grant 1/0424/03 and Slovak Grant APVT-20-004104.

I have greatly benefited from valuable discussions with George Judge. Without implicating him, I am indebted to Ľubomír Snoha for clarifications of a couple of important technical questions. Substantive comments and suggestions from three anonymous referees are gratefully acknowledged. Supported by VEGA 1/7295/20 and 1/3016/06 grants.

Submitted: August 31, 2005

2. Terminology and notation

Let $\mathcal{P}(\mathcal{X})$ be a set of all probability mass functions on a finite alphabet $\mathcal{X} \triangleq \{x_1, x_2, \dots, x_m\}$ of m letters. The support of $p \in \mathcal{P}(\mathcal{X})$ is a set $S(p) \triangleq \{x : p(x) > 0\}$.

A probability mass function (pmf) from $\mathcal{P}(\mathcal{X})$ is rational if it belongs to the set $\mathcal{R} \triangleq \mathcal{P}(\mathcal{X}) \cap \mathbb{Q}^m$. A rational pmf is n-rational, if denominators of all its m elements are n. The set of all n-rational pmf's will be denoted by $\mathcal{R}_n$.

Let $x_1, x_2, \dots, x_n$ be a sequence of n letters, that is identically and independently drawn from a source $q \in \mathcal{P}(\mathcal{X})$. Type and n-type are other names for empirical measures induced by a sequence of the length n. Formally, type $\nu^n \triangleq [n_1, n_2, \dots, n_m]/n$, where n_i is the number of occurrences of i-th letter of the alphabet in the sequence. Note that there are $\Gamma(\nu^n) \triangleq n!(\prod_{i=1}^m n_i!)^{-1}$ different sequences of length n, which induce the same type ν^n. $\Gamma(\nu^n)$ is called the multiplicity of type. Finally, observe that ν^n is n-rational; $\nu^n \in \mathcal{R}_n$.

Let $\Pi \subseteq \mathcal{P}(\mathcal{X})$. $\Pi_n \triangleq \Pi \cap \mathcal{R}_n$.

The information divergence ($\pm$-relative entropy, Kullback-Leibler distance etc.) $I(p||q)$ of p with respect to q (both from $\mathcal{P}(\mathcal{X})$) is $I(p||q) \triangleq \sum_{\mathcal{X}} p \log \frac{p}{q}$, with conventions that $0 \log 0 = 0$, $\log b/0 = +\infty$. The information projection $\hat{p}$ of q on Π is $\hat{p} \triangleq \arg\inf_{p \in \Pi} I(p||q)$. The value of the I-divergence at an I-projection of q on Π is denoted by $I(\Pi||q)$.

On $\mathcal{P}(\mathcal{X})$ topology induced by the standard topology on $\mathbb{R}^m$ is assumed.

The support $S(\mathcal{C})$ of a convex set $\mathcal{C} \subset \mathcal{P}(\mathcal{X})$ is just the support of the member of $\mathcal{C}$ for which $S(\cdot)$ contains the support of any other member of the set.

The following families of distributions will be needed:

1) Linear family $\mathcal{L}(u, a) \triangleq \{p : \sum_{\mathcal{X}} p(x) u_j(x) = a_j, j = 1, 2, \dots, k\}$, where u_j is a real-valued function on $\mathcal{X}$ and $a_j \in \mathbb{R}$.

2) Exponential family $\mathcal{E}(\rho, u, \theta) \triangleq \{p : p(x) = z\rho(x) \exp(\sum_{j=1}^k \theta_j u_j(x)), x \in \mathcal{X}\}$, where a normalizing factor $z \triangleq \sum_{\mathcal{X}} \rho(x) \exp(\sum_{j=1}^k \theta_j u_j(x))$ and ρ belongs to $\mathcal{P}(\mathcal{X})$; $\theta_j \in \mathbb{R}$.

3) Λ-family $\Lambda(\rho, u, \theta, a) \triangleq \{p : p(x) = \rho(x)[1 - \sum_{j=1}^k \theta_k (u_j(x) - a_j)]^{-1}, x \in \mathcal{X}\}$.

The definitions of the families can be extended to continuous $\mathcal{X}$ in a straightforward way.

In what follows, $r \in \mathcal{P}(\mathcal{X})$ will be the 'true' source of sequences and hence types.

3. Conditioning by rare types

It is convenient to begin with a brief survey of the Large Deviations Theorems for Empirical Measures (Sanov's Theorem and a Conditional Limit Theorem).

First, it is necessary to introduce the probability $\pi(\nu^n; r)$ that the source r generates an n-type ν^n. The probability that r generates a sequence of n letters $x_1, x_2, \dots, x_n$ which induces a type ν^n is $\prod_{i=1}^m (r_i)^{n\nu_i^n}$. As it was already mentioned, there is a number $\Gamma(\nu^n)$ of sequences of length n, which induce the same type ν^n. The probability $\pi(\nu^n; r)$ that r generates type ν^n is thus $\pi(\nu^n; r) \triangleq \Gamma(\nu^n) \prod_{i=1}^m (r_i)^{n\nu_i^n}$. Consequently, for $A \subseteq B \subseteq \mathcal{P}(\mathcal{X})$, $\pi(\nu^n \in A | \nu^n \in B; r) = \frac{\pi(\nu^n \in A; r)}{\pi(\nu^n \in B; r)}$; provided that $\pi(\nu^n \in B; r) \neq 0$.

Π is rare if it does not contain r. Given that the source r produced an n-type from rare Π, it is of interest to know how the conditional probability/measure spreads among the rare n-types from Π; especially as n grows beyond any limit. For the rare set of a particular form, this issue is answered by Conditional Limit Theorem (CoLT) which is also known as Conditional Weak Law of Large Numbers.

CoLT can be established by means of Sanov's Theorem (ST).

ST. *(*[6] *Thm 3) Let Π be a set such that its closure is equal to the closure of its interior. Let r be such that $S(r) = \mathcal{X}$. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \pi(\nu^n \in \Pi; r) = -I(\Pi||r).$$

Sanov's Theorem (ST) states that the probability $\pi(\nu^n \in \Pi; r)$ decays exponentially fast, with the decay rate given by the value of the information divergence at an I-projection of the source r on Π.

CoLT. *(*[8] *Thm 4.1,* [2] *Thm 12.6.2) Let Π be a convex, closed rare set. Let $B(\hat{p}, \epsilon)$ be a closed ϵ-ball defined by the total variation metric, centered at I-projection $\hat{p}$ of r on Π. Then for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \pi(\nu^n \in B(\hat{p}, \epsilon) \,|\, \nu^n \in \Pi; r) = 1.$$

Informally, CoLT states that if a dense rare set admits a unique I-projection, then asymptotically types conditionally concentrate just on it. Thus, provided that for sufficiently large n a type from rare Π occurred, with probability close to 1 it is just a type close to $\hat{p}$. Numeric examples of ST and CoLT can be found at [2].

This suggests that, conditionally upon the rare Π, it is the I-projection $\hat{p}$ rather than r, which should be considered as the true *iid* source of data. Gibbs' Conditioning Principle (GCP) - an important strengthening of CoLT - captures this 'intuition'; cf [3], [7].

If $S(\mathcal{L}) = \mathcal{X}$ then the I-projection $\hat{p}$ of r on $\Pi \equiv \mathcal{L}$ is unique and belongs to the exponential family of distributions $\mathcal{E}(r, u, \theta)$; i.e., $\mathcal{L}(u, a) \cap \mathcal{E}(r, u, \theta) = \{\hat{p}\}$.

4. Conditioning by rare sources

In the above setting there is a fixed source r and a rare set Π_n of n-types. We now consider an opposite setting where the n-type is unique, and there is a set $\mathcal{Q}_n \triangleq \mathcal{Q} \cap \mathcal{R}_n$, where $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{X})$, of rare n-sources of the type.

Furthermore, n-sources q^n are assumed to have prior distribution $\pi(q^n)$. If from $\mathcal{R}_n$ n-source q^n occurs, then the source generates n-type ν^n with the probability $\pi(\nu^n|q^n) \triangleq \Gamma(\nu^n) \prod_{i=1}^m (q_i^n)^{n\nu_i^n}$.

We are interested in the asymptotic behavior of the probability $\pi(q^n \in B \,|\, (q^n \in \mathcal{Q}) \wedge \nu^n)$ that if the n-type ν^n and an n-source q^n from a rare set $\mathcal{Q}$ occurred, then the n-source belongs to a subset B of $\mathcal{Q}$. Note that $\pi(q^n \in B \,|\, (q^n \in \mathcal{Q}) \wedge \nu^n) = \frac{\pi(q^n \in B|\nu^n)}{\pi(q^n \in \mathcal{Q}|\nu^n)}$; provided that $\pi(q^n \in \mathcal{Q}|\nu^n) > 0$. The posterior probability $\pi(q^n|\nu^n)$ is related to the defined probabilities $\pi(\nu^n|q^n)$ and $\pi(q^n)$ via Bayes's Theorem.

Asymptotic investigations will be first carried on under the assumption of uniform prior distribution of n-sources (Sect. 4.1). The assumption will be relaxed in Section 4.2. Within each of the sections, two cases of convergence will be considered: a static and a dynamic case. For the static case asymptotic investigations are carried over a subsequence of types, which are k-equivalent to ν^{n_0}. A type

$\nu^{kn_0} \triangleq [kn_1, \ldots, kn_m]/kn_0$, $k \in \mathbb{N}$, is called k-equivalent to ν^{n_0}. The dynamic case assumes that there is a sequence of n-types which converges in the total variation to some $p \in \mathcal{P}(\mathcal{X})$. For each case what is meant by rare source will be defined separately.

For $p, q \in \mathcal{P}(\mathcal{X})$, the L-divergence $L(q||p)$ of q with respect to p is the map $L : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \to \mathbb{R} \cup \{\infty\}$, $L(q||p) \triangleq -\sum_{\mathcal{X}} p \log q$. The L-projection $\hat{q}$ of p on set of sources $\mathcal{Q}$ is: $\hat{q} \triangleq \arg\inf_{q\in\mathcal{Q}} L(q||p)$. The value of L-divergence at an L-projection of p on $\mathcal{Q}$ (i.e., $\inf_{q\in\mathcal{Q}} L(q||p)$) is denoted by $L(\mathcal{Q}||p)$.

4.1 Uniform prior.

Within this section it is assumed that n-sources have a uniform prior distribution. Since there is total $N = \binom{n+m-1}{m-1}$ n-sources (cf. [4]), the uniform prior probability $\pi(q_n) = 1/N$, for all $q^n \in \mathcal{R}_n$.

4.1.1 Static case.

Let there be an n_0-type ν^{n_0}. A set $\mathcal{Q}$ of sources is rare if it does not contain ν^{n_0}.

Sanov's Theorem for Sources (abbreviated LST) is a counterpart of the Sanov's Theorem for Types.

Static LST. *Let ν^{n_0} be a type. Let $\mathcal{Q}$ be an open set of sources. Then, for $n \to \infty$ over a subsequence $n = kn_0$, $k \in \mathbb{N}$,*

$$\frac{1}{n} \log \pi(q^n \in \mathcal{Q}|\nu^n) = -\{L(\mathcal{Q}||\nu^{n_0}) - L(\mathcal{P}||\nu^{n_0})\}.$$

Proof. Under the assumption of uniform prior distribution of of n-sources

$$\log \pi(q^n \in \mathcal{Q}|\nu^n) = \log \sum_{q^n\in\mathcal{Q}} \prod_{\mathcal{X}} (q^n)^{n\nu^n} - \log \sum_{q^n\in\mathcal{P}} \prod_{\mathcal{X}} (q^n)^{n\nu^n}.$$

Since $N < (n+1)^m$ (cf. Lemma 2.1.2 of [7]), $\frac{1}{n_0} \log \pi(q^{n_0} \in \mathcal{Q}|\nu^{n_0})$ can be bounded from above and below as:

$$-L(\mathcal{Q}_{n_0}||\nu^{n_0}) + L(\mathcal{R}_{n_0}||\nu^{n_0}) - \frac{m}{n_0}\log(n_0+1) \le \frac{1}{n_0}\log \pi(q^{n_0} \in \mathcal{Q}|\nu^{n_0}) \le$$
$$\le -L(\mathcal{Q}_{n_0}||\nu^{n_0}) + L(\mathcal{R}_{n_0}||\nu^{n_0}) + \frac{m}{n_0}\log(n_0+1).$$

Fix $p \in \mathcal{P}(\mathcal{X})$. Equip $\mathbb{R} \cup \{\infty\}$ with the standard topology (i.e., the topology induced by the total order). As for each open subset A of $\mathbb{R} \cup \{\infty\}$, $L^{-1}(A)$ is an open subset of $\mathcal{P}(\mathcal{X})$, the L-divergence is continuous in q.

$\mathcal{Q}$ is open by the assumption.

Thus, $L(\mathcal{Q}_{n_0}||\nu^{n_0})$ converges to $L(\mathcal{Q}||\nu^{n_0})$ as $n \to \infty$, $n = kn_0$, $k \in \mathbb{N}$. Also, $L(\mathcal{R}_{n_0}||\nu^{n_0})$ converges to $L(\mathcal{P}||\nu^{n_0})$ for $n \to \infty$, $n = kn_0$, $k \in \mathbb{N}$. □

The Law of Large Numbers for Sources (LLLN) is a direct consequence of LST.

Static LLLN. *Let ν^{n_0} be a type. Let $\hat{q}$ be L-projection of ν^{n_0} on $\mathcal{P}(\mathcal{X})$. And let $B(\hat{q}, \epsilon)$ be a closed ϵ-ball defined by the total variation metric, centered at $\hat{q}$. Then, for $\epsilon > 0$ and $n \to \infty$ over the types which are k-equivalent with ν^{n_0},*

$$\pi(q^n \in B(\hat{q}, \epsilon) \,|\, (q^n \in \mathcal{P}) \wedge \nu^n) = 1.$$

Proof. Let $B^C(\hat{q}, \epsilon) \triangleq \mathcal{P}(\mathcal{X}) \backslash B(\hat{q}, \epsilon)$. Since $B^C(\hat{q}, \epsilon)$ is open by the assumption, LST can be applied to it. Since $B^C \subset \mathcal{P}$, $L(B^C||\nu^{n_0}) - L(\mathcal{P}||\nu^{n_0}) > 0$. Thus, $\pi(q^n \in B^C(\hat{q}, \epsilon)|\nu^n)$ converges to 0, as $n \to \infty$ over a subsequence of $n = kn_0$, $k \in \mathbb{N}$. □

Obviously, the L-projection $\hat{q}$ of ν^{n_0} on $\mathcal{P}(\mathcal{X})$ is $\hat{q} \equiv \nu^{n_0}$.

LLLN is a special case of the Conditional Limit Theorem for Sources (LCoLT), which is a consequence of LST, as well.

Static LCoLT. *Let ν^{n_0} be a type. Let $\mathcal{Q}$ be a convex, closed rare set of sources. Let $\hat{q}$ be the L-projection of ν^{n_0} on $\mathcal{Q}$ and let $B(\hat{q}, \epsilon)$ be a closed ϵ-ball defined by the total variation metric, centered at $\hat{q}$. Then, for $\epsilon > 0$ and $n \to \infty$ over a subsequence $n = kn_0$, $k \in \mathbb{N}$,*

$$\pi(q^n \in B(\hat{q}, \epsilon) \,|\, (q^n \in \mathcal{Q}) \wedge \nu^n) = 1.$$

Proof. Let $B^C(\hat{q}, \epsilon) \triangleq \mathcal{P}(\mathcal{X}) \backslash B(\hat{q}, \epsilon)$. Clearly,

$$\log \pi(q^{n_0} \in B^C(\hat{q}, \epsilon) \,|\, (q^{n_0} \in \mathcal{Q}) \wedge \nu^{n_0}) = \log \pi(q^{n_0} \in B^C|\nu^{n_0}) - \log \pi(q^{n_0} \in \mathcal{Q}|\nu^{n_0}).$$

Since both $B^C(\hat{q}, \epsilon)$ and $\mathcal{Q}$ are open, LST can be applied. As $B^C(\hat{q}, \epsilon) \subset \mathcal{Q}$, $L(B^C||\nu^{n_0}) - L(\mathcal{Q}||\nu^{n_0}) > 0$. Hence $\pi(q^n \in B^C|(q^n \in \mathcal{Q}) \wedge \nu^n)$ converges to 0, as $n \to \infty$ over a subsequence of $n = kn_0$, $k \in \mathbb{N}$. Since under the assumptions on $\mathcal{Q}$ the L-projection of ν^{n_0} on $\mathcal{Q}$ is unique, the claim of the Theorem follows. □

Example. Let $\mathcal{X} = \{1, 2, 3, 4\}$. Let $\mathcal{Q} = \{q : \sum_{x \in \mathcal{X}} q(x)x = 1.7\}$. Let $n_0 = 10$ and $\nu^{n_0} = [1, 1, 1, 7]/10$. The L-projection of ν^{n_0} on $\mathcal{Q}$ is $\hat{q} = [0.705,\ 0.073,\ 0.039,\ 0.183]$. Let $\epsilon = 0.1$. The concentration of n-sources on the L-projection, which is captured by the Static LCoLT, is for types k-equivalent to ν^{n_0} ($k = 5, 10, 20, 30$) illustrated in Table 1.

TABLE 1. Values of $\pi(q^n \in B(\hat{q}, \epsilon)|(q^n \in \mathcal{Q}) \wedge \nu^n)$ for $n = kn_0$, $k = 5, 10, 20, 30$.

n	$\pi(\cdot\|\cdot)$
50	0.868
100	0.948
200	0.994
300	0.999

The L-projection at the above Example can be found by means of the following Proposition.

Proposition. *Let $\mathcal{Q} \equiv \mathcal{L}(u, a)$. Let $p \in \mathcal{P}(\mathcal{X})$ be such that $S(p) = S(\mathcal{L})$. Then the L-projection $\hat{q}$ of p on $\mathcal{Q}$ is unique and belongs to $\Lambda(p, u, \theta, a)$ family; i.e., $\mathcal{L}(u, a) \cap \Lambda(p, u, \theta, a) = \{\hat{q}\}$.*

Proof. In light of Theorem 9 of [6] it suffices to check that $\hat{q} = p[1 - \sum_{j=1}^{k} \theta_k(u_j(x) - a_j)]^{-1}$, with θ such that $\hat{q} \in \mathcal{L}(u, a)$, satisfies:

$$\sum_{S(p)} p \left(1 - \frac{q'}{\hat{q}}\right) = 0,$$

for all $q' \in \mathcal{Q}$, which is indeed the case. $\square$

4.1.2 Dynamic case.

Let there be a sequence of n-types which converges in the total variation to a pmf $p \in \mathcal{P}(\mathcal{X})$, denoted as $\nu^n \to p$. In this case, a set $\mathcal{Q}$ of sources is rare if it does not contain p.

Dynamic LST. *Let $\nu^n \to p$. Let $\mathcal{Q}$ be an open set of sources. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \pi(q^n \in Q|\nu^n) = -\{L(\mathcal{Q}||p) - L(\mathcal{P}||p)\}.$$

Dynamic LLLN. *Let $\nu^n \to p$. Let $\hat{q}$ be L-projection of p on $\mathcal{P}$. And let $B(\hat{q}, \epsilon)$ be a closed ϵ-ball defined by the total variation metric, centered at $\hat{q}$. Then, for $\epsilon > 0$,*

$$\lim_{n\to\infty} \pi(q^n \in B(\hat{q}, \epsilon) \,|\, (q^n \in \mathcal{P}) \wedge \nu^n) = 1.$$

Dynamic LCoLT. *Let $\nu^n \to p$. Let $\mathcal{Q}$ be a convex, closed rare set of sources. Let $\hat{q}$ be the L-projection of p on $\mathcal{Q}$ and let $B(\hat{q}, \epsilon)$ be a closed ϵ-ball defined by the total variation metric, centered at $\hat{q}$. Then, for $\epsilon > 0$,*

$$\lim_{n\to\infty} \pi(q^n \in B(\hat{q}, \epsilon) \,|\, (q^n \in \mathcal{Q}) \wedge \nu^n) = 1.$$

Proofs can be constructed along the lines for the static case.

4.2 General prior.

Let $\pi(q)$ be a prior pmf on $\mathcal{R}$. From this pmf, a prior distribution $\pi^{\mathcal{A}}(q^n)$ on $\mathcal{R}_n$ is constructed by a quantization $\mathcal{A} \triangleq \{A_1, A_2, \dots, A_N\}$ of $\mathcal{R}$ into disjoint sets, such that each $A \in \mathcal{A}$ contains just one q_n from $\mathcal{R}_n$. Then $\pi^{\mathcal{A}}(q^n) \triangleq \pi(\{A_j : q^n \in A_j, j = 1, 2, \dots, N\})$.

Let $\mathcal{S} \triangleq S(\pi(\cdot))$. Let $\mathcal{Q}^\pi \triangleq \mathcal{Q} \cap \mathcal{S}$, $\mathcal{P}^\pi \triangleq \mathcal{P} \cap \mathcal{S}$.

As the static case is subsumed under the dynamic one, only the latter limit theorems will be presented.

General prior LST. *Let $\nu^n \to p$. Let $\mathcal{Q}^\pi$ be an open set of sources. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \pi^{\mathcal{A}}(q^n \in \mathcal{Q}^\pi|\nu^n) = -\{L(\mathcal{Q}^\pi||p) - L(\mathcal{P}^\pi||p)\}.$$

Proof. For a zero-prior-probability n-source, the posterior probability is zero as well; so such sources can be excluded from considerations. Let $\mathcal{S}_n \triangleq S(\pi^{\mathcal{A}}(q_n))$, $\mathcal{Q}_n^\pi \triangleq \mathcal{Q} \cap \mathcal{S}_n$, $\mathcal{P}_n^\pi \triangleq \mathcal{P} \cap \mathcal{S}_n$.

$$\log \pi^{\mathcal{A}}(q^n \in \mathcal{Q}|\nu^n) = \log \sum_{q^n \in \mathcal{Q}_n^\pi} \pi^{\mathcal{A}}(q^n) \prod_{\mathcal{X}} (q^n)^{n\nu^n} - \log \sum_{q^n \in \mathcal{P}_n^\pi} \pi^{\mathcal{A}}(q^n) \prod_{\mathcal{X}} (q^n)^{n\nu^n}.$$

Denote by $\lambda(\mathcal{Q}_n^\pi||\nu^n) \triangleq \inf_{q^n \in \mathcal{Q}_n^\pi} \lambda(q^n||\nu^n)$, where $\lambda(q^n||\nu^n) \triangleq L(q^n||\nu^n) - \frac{1}{n} \log \pi^{\mathcal{A}}(q^n)$. Using this notation and invoking the same argument as in the proof of LST for uniform prior, $\frac{1}{n}\log \pi^{\mathcal{A}}(q^n \in \mathcal{Q}|\nu^n)$ can be bounded from above and below as:

$$- \lambda(\mathcal{Q}_n^\pi||\nu^n) + \lambda(\mathcal{P}_n^\pi||\nu^n) - \frac{m}{n}\log(n+1) \leq \frac{1}{n}\log \pi^{\mathcal{A}}(q^n \in \mathcal{Q}|\nu^n) \leq$$
$$\leq -\lambda(\mathcal{Q}_n^\pi||\nu^n) + \lambda(\mathcal{P}_n^\pi||\nu^n) + \frac{m}{n}\log(n+1).$$

Since for $n \to \infty$, $\mathcal{S}_n = \mathcal{S}$, and $\nu^n \to p$, and $\mathcal{Q}^\pi$ is open, it taken together, implies that $\lambda(\mathcal{Q}_n^\pi||\nu^n)$ converges to $L(\mathcal{Q}^\pi||p)$. Similarly, $\lambda(\mathcal{P}_n^\pi||\nu^n)$ converges to $L(\mathcal{P}^\pi||p)$. □

Let $\nu^n \to p$. A set of sources is rare if it does not contain p. Then, from the General prior LST, follows

General prior LCoLT. *Let $\nu^n \to p$. Let $\mathcal{Q}^\pi$ be a convex, closed rare set of sources. Let $\hat{q}^\pi$ be the L-projection of p on $\mathcal{Q}^\pi$. Let $B(\hat{q}^\pi, \epsilon)$ be a closed ϵ-ball defined by the total variation metric, centered at $\hat{q}^\pi$. Then, for $\epsilon > 0$,*

$$\lim_{n\to\infty} \pi^{\mathcal{A}}(q^n \in B(\hat{q}^\pi, \epsilon) \,|\, (q^n \in \mathcal{Q}^\pi) \wedge \nu^n) = 1.$$

4.3 Conditioning by rare sources: continuous alphabet

Sanov's Theorem for continuous alphabet can be established either via 'the method of types + discrete approximation' approach (cf. [4]) or by means of the large deviations theory (cf. [7]). The former approach will be used here to formulate continuous alphabet version of LST.

Let $(\mathcal{Y}, \mathcal{F})$ be a measurable space. Let $\mathcal{T}^m$ be a partition of the alphabet $\mathcal{Y}$ into finite number m of sets $\mathcal{T}^m \triangleq (T_1, T_2, \ldots, T_m)$; $T_i \in \mathcal{F}$. The $\mathcal{T}^m$-quantized P, denoted by $P^{\mathcal{T}}$, is defined as the distribution $P(T_1), P(T_2), \ldots, P(T_m)$ on the finite set $\mathcal{X} \triangleq \{1, 2, \ldots, m\}$.

Let $\mathcal{P}(\mathcal{Y})$ be the set of all probability measures on $(\mathcal{Y}, \mathcal{F})$. Let $\mathcal{Q} \subseteq \mathcal{P}$. For probability measures (pm's) $P, Q \in \mathcal{P}(\mathcal{Y})$, the L^m-divergence $L^m(Q||P)$ of Q with respect to P is defined as

$$L^m(Q||P) \triangleq \sup_{\mathcal{T}^m} L(Q^{\mathcal{T}}||P^{\mathcal{T}}),$$

where the supremum is taken over all m-element partitions. $L^m(\mathcal{Q}||P)$ denotes $\sup_{Q\in\mathcal{Q}} L^m(Q||P)$. Let $\mathcal{Q}^{\mathcal{T}} \triangleq \{Q : Q^{\mathcal{T}} \in \mathcal{Q}\}$, $L^m(\mathcal{Q}^{\mathcal{T}}||P^{\mathcal{T}}) \triangleq \sup_{\mathcal{Q}} L(Q^{\mathcal{T}}||P^{\mathcal{T}})$.

The empirical distribution $\nu^{n,m}$ of an n-sequence of $\mathcal{Y}$-valued random variables Y with respect to a partition $\mathcal{T}^m$ is defined as

$$\nu_j^{n,m} = \frac{1}{n}\,\mathrm{Card}\{Y_i : Y_i \in T_j;\, 1 \leq i \leq n\}, \qquad 1 \leq j \leq m.$$

The τ^m-topology of pm's on $(\mathcal{Y}, \mathcal{F})$ is the topology in which a pm belongs to the interior of a set $\mathcal{Q}$ of pm's iff for some partition $\mathcal{T}^m$ and $\epsilon > 0$

$$\{Q' : |Q'(T_j) - Q(T_j)| < \epsilon, j = 1, 2, \ldots, m\} \subset \mathcal{Q}.$$

Thus, an n-source $q^n \in \mathcal{R}_n(\mathcal{X})$ belongs to the interior of $\mathcal{Q}$ if there exists $\mathcal{T}^m$ of $\mathcal{Y}$ and $\epsilon > 0$ such that the set $\{Q' : |Q'(T_j) - q_j^n| < \epsilon, j = 1, 2, \ldots, m\}$ is a subset of $\mathcal{Q}$.

Under the assumption of uniform prior distribution of n-sources, a continuous analogue to the Dynamic LST is:

Continuous L**ST.** *Let, as* $n \to \infty$, $\nu^{n,m} \to R$, $R \in \mathcal{R}(\mathcal{X})$. *Let* $\mathcal{Q}$ *be a rare (i.e.,* $R \notin \mathcal{Q}$*) open subset of* $\mathcal{P}(\mathcal{Y})$. *Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \pi(q^n \in \mathcal{Q}|\nu^{n,m}) = -\{L^m(\mathcal{Q}||R) - L^m(\mathcal{P}||R)\}.$$

Proof. First, an asymptotic lower bound to $\frac{1}{n}\log\pi(q^n \in \mathcal{Q}|\nu^n)$ will be established. Pick up a Q such that for a $\mathcal{T}^m$, and an $\epsilon > 0$, $Q \in \mathcal{Q}$. Let $\mathcal{M}^{\mathcal{T}}(Q) \triangleq \{q^n : |q^n_j - Q(T_j)| < \epsilon, j = 1, 2, \dots, m\}$. By the Dynamic LST for uniform prior, $\lim_{n\to\infty} \frac{1}{n}\log\pi(q^n \in \mathcal{M}^{\mathcal{T}}(Q)|\nu^n) = -\{L(\mathcal{M}^{\mathcal{T}}(Q)||R^{\mathcal{T}}) - L(R^{\mathcal{T}}||R^{\mathcal{T}})\}$ which is greater or equal to $-\{L(Q^{\mathcal{T}}||R^{\mathcal{T}}) - L(R^{\mathcal{T}}||R^{\mathcal{T}})\}$, since $Q^{\mathcal{T}} \in \mathcal{M}^{\mathcal{T}}(Q)$. Let $\mathcal{M}(Q) \triangleq \cup_{\mathcal{T}^m}\mathcal{M}^{\mathcal{T}}(Q)$. Then, for $n \to \infty$, $\frac{1}{n}\log\pi(q^n \in \mathcal{M}(Q)|\nu^n) \geq \sup_{\mathcal{T}^m} -\{L(Q^{\mathcal{T}}||R^{\mathcal{T}}) - L(R^{\mathcal{T}}||R^{\mathcal{T}})\} \equiv -\{L^m(Q||R) - L^m(R||R)\}$. Since $\pi(q^n \in \mathcal{Q}|\nu^n) \geq \sup_{Q\in\mathcal{Q}} \pi(q^n \in \mathcal{M}(Q)|\nu^n)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \pi(q^n \in \mathcal{Q}|\nu^n) \geq \sup_{Q\in\mathcal{Q}} -\{L^m(Q||R) - L^m(R||R)\}$$
$$\equiv -\{L^m(\mathcal{Q}||R) - L^m(\mathcal{P}||R)\}.$$

Asymptotic upper bound: for $\mathcal{T}^m$ as above, by the Dynamic LST with a uniform prior,

$$\lim_{n\to\infty} \frac{1}{n} \log \pi(q^n \in \mathcal{Q}^{\mathcal{T}}|\nu^n) = -\{L^m(\mathcal{Q}^{\mathcal{T}}||R^{\mathcal{T}}) - L^m(\mathcal{P}^{\mathcal{T}}||R^{\mathcal{T}})\}$$
$$\equiv \sup_{\mathcal{Q}} -\{L(Q^{\mathcal{T}}||\mathcal{P}^{\mathcal{T}}) - L(R^{\mathcal{T}}||R^{\mathcal{T}})\}.$$

Since $\pi(q^n \in \mathcal{Q}|\nu^n) \leq \sup_{\mathcal{T}^m} \pi(q^n \in \mathcal{Q}^{\mathcal{T}}|\nu^n)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \pi(q^n \in \mathcal{Q}|\nu^n) \leq -\{L^m(\mathcal{Q}||R) - L^m(\mathcal{P}||R)\}.$$

As the asymptotic lower and upper bounds coincide, the claim follows. □

5. Application to Criterion Choice Problem

1. Let there be an alphabet $\mathcal{X}$ (finite, for simplicity) and prior distribution $\pi(q)$ of sources. From the prior $\pi(q)$ a source is drawn, and the source then generates an n-type ν^n. We are not given the actual source, but rather a set $\mathcal{Q}$ to which the source belongs. Given the alphabet $\mathcal{X}$, the n-type ν^n, the prior distribution of sources $\pi(\cdot)$ and the set $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{X})$ the objective is to select a source $q \in \mathcal{Q}$. This constitutes the β-problem. Since $\mathcal{Q}$ in general contains more than one source the problem is under-determined and in this sense ill-posed. This paper is concerned with the special case of the β-problem where rational sources (i.e., n-sources) are considered.

If $\mathcal{Q} \equiv \mathcal{P}(\mathcal{X})$, then under the assumption of uniform prior distribution of n-sources, Static LLLN shows that asymptotically (along the types k-equivalent with ν^n) it is just $\hat{q} \equiv \nu^n$ which is the 'only-possible' source of ν^n (i.e., of itself)[1].

[1]Note that in the case of unrestricted $\mathcal{Q}$, ν^n is known to be Non-parametric Maximum Likelihood Estimator of the source. Here, ν^n is the Maximum A-posteriori Probability source.

Dynamic LLLN, assuming that $\nu^n \to r$, implies that the n-sources concentrate on the true source. However, they do not, if a general prior is assumed, such that it puts zero probability on the true source. In the dynamic case ($\nu^n \to r$) with general prior, n-sources concentrate on the L-projection of r on $\mathcal{P}^\pi$.

What if $\mathcal{Q}$ does not contain ν^n? How should an n-source be selected in this case of static rare $\mathcal{Q}$? One possibility is to select q^n from $\mathcal{Q}$ by minimization of a distance or a convex statistical distance measure [16] between ν^n and $\mathcal{Q}_n$. In this way, the original β-problem of selecting $q^n \in \mathcal{Q}$ is transformed into an associated Criterion Choice Problem (CCP).

If the rare $\mathcal{Q}$ is convex and closed, Static LCoLT shows that - at least for n sufficiently large - the CCP associated with this instance of the β-problem should be solved by minimization of the L-divergence over $\mathcal{Q}$. A major qualifier has to be added to this statement: it holds provided that uniform prior distribution of n-sources is assumed. If a general prior, strictly positive on the entire set of rational sources is assumed, then the statement still holds. Prior matters only if it is not strictly positive on the entire $\mathcal{R}$. Then, it is the L-projection of ν^n on $\mathcal{Q}^\pi$ that should be selected (recall the General prior LCoLT).

2. Confront the β-problem with the following α-problem (also known as Boltzmann Jaynes Inverse Problem): let there be a source q that emits letters from an alphabet $\mathcal{X}$. From the source q an n-type was drawn. We are not given the actual n-type, but rather a set Π to which the n-type belongs. Given the alphabet $\mathcal{X}$, the source q and the set Π the objective is to select an n-type $\nu^n \in \Pi$.

The CCP associated with the α-problem is solved by CoLT and GCP provided that Π is a convex, closed rare set. The Theorems imply that at least for sufficiently large n, the I-projection of q on Π should be selected.

6. EndNotes

0) While preparing the final form of the paper, the author learned about the work [10] by Ayalvadi Ganesh and Neil O'Connell, where an inverse of Sanov's Theorem has already been studied and established. The authors also discuss relevance of the Theorem for Bayesian nonparametric consistency. Some differences: the authors work with prior distribution on $\mathcal{P}(\mathcal{X})$ where the alphabet $\mathcal{X}$ is finite. Here the concept of n-source is used and continuous $\mathcal{X}$ is also considered. The rate function of General prior LST of the present work appears to be more general than that of Theorem 1 of [10], as the latter was established for the case of $\pi(p) > 0$. The Conditional Limit Theorem (LCoLT) was not explicitly considered at [10].

1) The terminology and notation of this paper follow more or less closely [2], [4], [6], [7]. The brief survey of Large Deviations Theorems for Empirical Measures (Sect. 3) draws from the same sources. Reader interested in tracking evolution of the Theorems is directed to [1], [3], [7], [8], [12], [16], [18], [19], [22], [23], [24], [25], [27], among others; see [13] for new developments. In relation to the Proposition of Sect. 4.1 see also [9]. The continuous case of conditioning by rare sources (Sect. 5) is built parallel with [12] and [4].

2) This work is motivated by [11], where a problem of selecting between Empirical Likelihood and Maximum Entropy Empirical Likelihood (cf. [20], [21]) has been addressed on probabilistic, rather than statistical, grounds. Further discussion, relevant also to the CCP associated with the α and β-problems, can be found there.

3) Any of the results presented here may be stated in terms of reverse I-projections [5]. For instance the right-hand side of the General prior LST could be equivalently expressed as $-(I(p||\mathcal{Q}^\pi) - I(p||\mathcal{P}^\pi))$, where $I(p||C) \triangleq \inf_{q \in C} I(p||q)$ is the value of the I-divergence at a reverse I-projection of p on C. The above mentioned statistical considerations (and 4) below) served as a motivation for stating the results in terms of the newly introduced L-divergence, though the L-projection is formally identical with the reverse I-projection, which is already in use in a parametric context, cf. [5]. The present work leaves open the issue whether it is more advantageous to state the Theorems of conditioning by rare sources in terms of the L-projection or in terms of the reverse I-projection.

4) If p is an n-type then the L-divergence is known as Kerridge's inaccuracy; cf. [14], [15]. Watanabe in a fundamental work [26] which also addresses questions related to that of the present paper, calls negative of Kerridge's inaccuracy confirmability. A reviewer pointed out that the L-divergence can be identified with mean code length.

5) For any prior $\pi(\cdot)$, the L-projection $\hat{q}^\pi$ of p on $\mathcal{Q}^\pi$ is the same as the source which has asymptotically supremal over $\mathcal{Q}^\pi$ value of the posterior probability $\pi(q^n|\nu^n)$. In the case of uniform prior the correspondence holds for any n.

References

1. Bártfai, P., *On a conditional limit theorem*, Progress in Statistics **1** (1974), 85–91.
2. Cover, T. and Thomas, J., *Elements of Information Theory*, John Wiley and Sons, NY, 1991.
3. Csiszár. I., *Sanov Property, Generalized I-projection and a Conditional Limit Theorem*, Ann. Probab. **12** (1984), 768–793.
4. ———, *The Method of Types*, IEEE Trans. IT **44** (1998), 2505–2523.
5. Csiszár, I. and Matúš, F., *Information projections revisited*, IEEE Trans. IT **49** (2004), 1474–1490.
6. Csiszár, I. and Shields, P., *Notes on Information Theory and Statistics: A tutorial*, Foundations and Trends in Communications and Information Theory **1** (2004), 1–111.
7. Dembo, A and Zeitouni, O., *Large Deviations Techniques and Applications*, 2-nd ed., Springer, Application of Mathematics, vol. 38, NY, 1998.
8. Ellis, R. S., *The theory of large devitions: from Boltzmann's 1877 calculation to equilibrium macrostates in 2D turbulence*, Physica D (1999), 106–113.
9. Friedlander M. P. and Gupta M. R., *On minimizing distortion and relative entropy*, IEEE Trans. IT (to appear).
10. Ganesh, A. and O'Connell, N., *An inverse of Sanov's theorem*, Stat. & Prob. Letters **42** (1999), 201–206.
11. Grendár, M. and Judge, G., *Probabilistic approach to Criterion Choice Problem: Estimating Equations case* (2005) (working paper).
12. Groeneboom, P., Oosterhoff, J. and Ruymgaart, F. H., *Large deviation theorems for empirical probability measures*, Ann. Probab. **7** (1979), 553–586.
13. Harremoës, P., *Information topologies with applications* (to appear at *Bolyai Studies.*).
14. Kerridge, D. F., *Inaccuracy and inference*, J. Roy. Statist. Soc. Ser. B **23** (1961), 284–294.
15. Kulhavý, R., *Recursive Nonlinear Estimation: A Geometric Approach*, vol. 216, Springer-Verlag, London, 1996 (Lecture Notes in Control and Information Sciences).
16. La Cour, B. R. and Schieve, W. C., *Macroscopic determinism in interacting systems using Large Deviations Theory*, Jour. Stat. Phys. **107 3/4** (2002), 729-755.
17. Liese, F. and Vajda, I., *Convex Statistical Distances*, Teubner, Leipzig, 1987.
18. Leonard, Ch. and Najim, J., *An extension of Sanov's Theorem: Application to the Gibbs Conditioning Principle*, Bernoulli **8 (6)** (2002), 721–743.
19. Lewis, J. T., Pfister, C.-E. and Sullivan, W. G., *Entropy, concentration of probability and conditional theorems*, Markov Proc. Rel. Field. **1** (1995), 319–386.
20. Mittelhammer, R., Judge, G. and Miller, D., *Econometric Foundations*, CUP, Cambridge, 2000.

21. Owen, A., *Empirical Likelihood*, Chapman & Hall/CRC, NY, 2001.
22. Sanov, I. N., *On the probability of large deviations of random variables*, Mat. Sbornik **42** (1957), 11–44 (In Russian).
23. van Campenhout, J. M. and Cover, T. M., *Maximum entropy and conditional probability*, IEEE Trans. IT **27** (1981), 483–489.
24. Vasicek, O. A., *A conditional law of large numbers*, Ann. Probab. 8 (1980), 142–147.
25. Vincze, I., *On the maximum probability principle in statistical physics*, Coll. Math. Soc. J. Bolyai **9** (1972), 869–893,.
26. Watanabe, S., *Information-theoretic aspects of inductive and deductive inference*, IBM Journal (1960), 208–231.
27. Zabel, S., *Rates of convergence for conditional expectations*, Ann. Probab. **8** (1980), 928–941.

Department of Mathematics; FPV UMB; Tajovskeho 40; SK-974 01 Banska Bystrica; Slovakia.

Institute of Mathematics and Computer Science; Banska Bystrica; Slovakia.

Institute of Measurement Science; Bratislava; Slovakia

E-mail: `marian.grendar@savba.sk`