LOVEGROVE MATHEMATICALS

ROUTES AND CHAINS IN LIKELINESS THEORY

RESEARCH REPORT 2014-01

Roger Lovegrove

The Chain Rule for Likelinesses contains a leading coefficient which is absent from the probabilistic equivalent. This short report gives a geometric explanation for this.

The recommended citation is

Lovegrove R., (2014), Routes and Chains in Likeliness Theory, Lovegrove Mathematicals Research Report 2014-01. March 2014.

LONDON
UNITED KINGDOM
March 2014

www.lovegrovemaths.co.uk

roger@lovegrovemaths.co.uk

Contents

1 Introduction

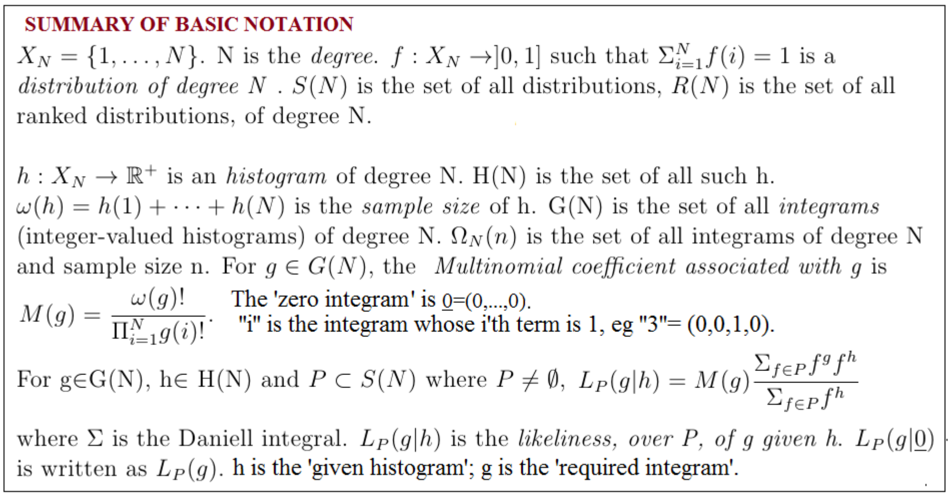
In Probability Theory, the Chain Rule states that $Pr(A \; and \; B) = Pr(A|B)Pr(B)$. By analogy, this might lead us to expect that the Likeliness Theory equivalent would say that $L_P(g_2 + g_1|h) = L_P(g_2|g_1 + h)L_P(g_1|h)$.

The Likeliness version of the Chain Rule, however, actually states[1]

$$L_P(g_2 + g_1|h) = \frac{M(g_2 + g_1)}{M(g_2)M(g_1)} L_P(g_2|g_1 + h)L_P(g_1|h).$$

Apart from the leading coefficient of $\frac{M(g_2 + g_1)}{M(g_2)M(g_1)}$, this is as expected. This report gives a geometric approach to Chains which explains why there should be that leading coefficient.

2 Notation and Terminology



SUMMARY OF BASIC NOTATION

$X_N = \{1, \ldots, N\}$. N is the *degree*. $f : X_N \to]0, 1]$ such that $\Sigma_{i=1}^N f(i) = 1$ is a *distribution of degree* N . $S(N)$ is the set of all distributions, $R(N)$ is the set of all ranked distributions, of degree N.

$h : X_N \to \mathbb{R}^+$ is an *histogram* of degree N. H(N) is the set of all such h. $\omega(h) = h(1) + \cdots + h(N)$ is the *sample size* of h. G(N) is the set of all *integrams* (integer-valued histograms) of degree N. $\Omega_N(n)$ is the set of all integrams of degree N and sample size n. For $g \in G(N)$, the *Multinomial coefficient associated with g* is $M(g) = \frac{\omega(g)!}{\Pi_{i=1}^N g(i)!}$. The 'zero integram' is $\underline{0}$=(0,...,0). "i" is the integram whose i'th term is 1, eg "3"= (0,0,1,0).

For g∈G(N), h∈ H(N) and $P \subset S(N)$ where $P \neq \emptyset$, $L_P(g|h) = M(g)\frac{\Sigma_{f \in P} f^g f^h}{\Sigma_{f \in P} f^h}$

where Σ is the Daniell integral. $L_P(g|h)$ is the *likeliness, over P, of g given h*. $L_P(g|\underline{0})$ is written as $L_P(g)$. h is the 'given histogram'; g is the 'required integram'.

Figure 1: Basic notation

Notation and terminology follow [1].

Note: Because of problems with embedded fonts for the larger mathematical symbols

- The greek upper-letter Σ is used instead of the summation sign
- The greek upper-case letter Π us used instead of the product sign
- Superscripts and subscripts are used instead of limits
- normal, line-height (and) have been used as brackets, wherever possible, instead of larger brackets.
- the Daniell integral is represented by the summation (Σ) sign.

There are, however, still problems with some symbols, such as large brackets around matrices.

3 Batches and Chains

Any integram other than the zero integram, $\underline{0}$, will be called a *batch*, and the term *batch size* will be used rather than *sample size.*

Given a finite sequence of batches, $g_1, \ldots, g_m$, the finite series

$$\underline{0}, \quad g_1 \ , \quad g_2 + g_1 \ , \quad \ldots \ , \quad g_m + \cdots + g_1$$

is called a *chain.* The batches are batches *of that chain* and the sequence of batches is the *generating sequence* of the chain. If $g = g_m + \cdots + g_1$ then the chain is a *chain to g*, and g is the *end-point* of the chain.

The terms of a chain apart from the initial $\underline{0}$ are called its *via-points.* Any chain which has the integram V as a via-point is said to be *via V.*

4 Routes

A chain to g whose batches all have a batch size of 1 is called a *route* to g.

4.1 Canonical Route

The canonical route to $\underline{0}$ is the route whose only term is $\underline{0}$. Otherwise, the canonical route to the integram g is that route for which the first g(1) terms of its generating sequence are $''1''$, the next g(2) are $''2''$, etc.

For example, the canonical route to (2,0,3,1) has the generating sequence
$''1'',''1'',''3'',''3'',''3'',''4''$
and therefore is the route
$\underline{0},''1'',2''1'',2''1''+''3'',2''1''+2''3'',2''1''+3''3'',2''1''+3''3''+''4''.$

4.2 Counting Routes

All routes to g have generating sequences which are permutations of one-another, and every distinct permutation gives a distinct route, so the number of distinct routes to g is the number of distinct permutations, which is M(g).

Let a chain to g have batches $g_1, \ldots, g_m$. Then there are $M(g_m) \ldots M(g_2)M(g_1)$ routes to g which are via all the via-points of that chain (Figure 2).

Of all the routes to g, the proportion which are via all of those via-points is $\dfrac{M(g_m) \ldots M(g_1)}{M(g)}$, which is $\dfrac{M(g_m) \ldots M(g_1)}{M(g_m + \cdots + g_1)}$.

Another way of putting this is that if N_V is the number of routes to g which are via all the via points of the chain then the total number of routes to g is $\dfrac{M(g_m + \cdots + g_1)}{M(g_m) \ldots M(g_1)}.N_V$

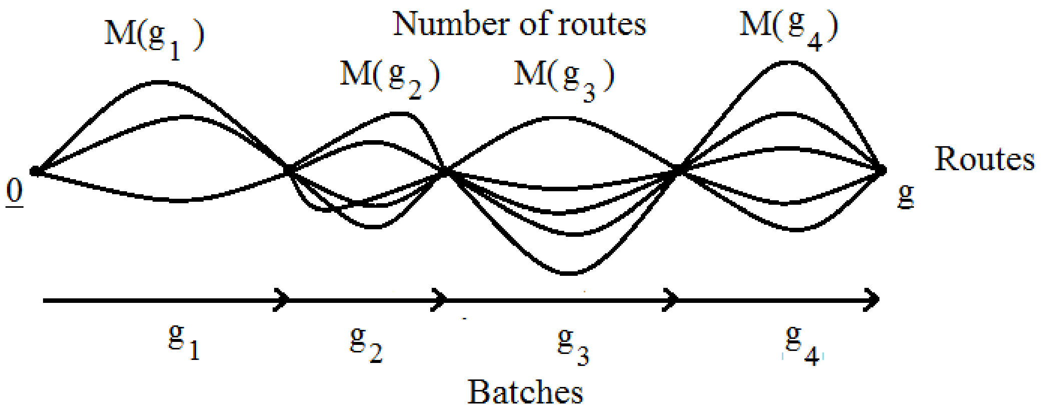


Figure 2: Routes to g via via-points

5 Likeliness of a chain

Let C be a chain to g, and let its generating sequence be $g_1, \ldots, g_m$. Then we define the Likeliness of C (over P and given h) to be $L_P(C|h)$ where

$$L_P(C|h) = L_P(g_m|g_{m-1} + \cdots + g_1 + h)L_P(g_{m-1}|g_{m-2} + \cdots + g_1 + h)\ldots L_P(g_1|h) \quad (1)$$

6 Chain Rule

The Likeliness Chain Rule states

$$L_P(g_2 + g_1|h) = \frac{M(g_2 + g_1)}{M(g_2)M(g_1)}L_P(g_2|g_1 + h)L_P(g_1|h) \quad (2)$$

This does generalise in the expected way:

$$L_P(g_m + \cdots + g_1|h) = \frac{M(g_m + \cdots + g_1)}{M(g_m)\ldots M(g_1)}L_P(g_m|g_{m-1} + \cdots + g_1 + h)\ldots L_P(g_1|h). \quad (3)$$

For example,

$$L_P(g_4{+}g_3{+}g_2{+}g_1|h) = \frac{M(g_4 + g_3 + g_2 + g_1)}{M(g_4)M(g_3)M(g_2)M(g_1)}L_P(g_4|g_3{+}g_2{+}g_1{+}h)L_P(g_3|g_2{+}g_1{+}h)L_P(g_2|g_1{+}h)L_P(g_1|h). \quad (4)$$

The LHS of this is the likeliness of $g = g_4 + g_3 + g_2 + g_1$.
On the RHS, the product of likelinesses is the likeliness of a chain to g. This chain has the via-points g_1, $g_2 + g_1$, $g_3 + g_2 + g_1$,$g_4 + g_3 + g_2 + g_1$. In the leading coefficient, the numerator is the number of routes to g and the denominator is the number of routes to g which are via all of the chain's via-points

.All of this generalises so, for clarity, we shall continue working with this specific case.

7 Likelinesses and routes

We rewrite (4) as

$$L_P(g_4|g_3+g_2+g_1+h)L_P(g_3|g_2+g_1+h)L_P(g_2|g_1+h)L_P(g_1|h) = M(g_4)M(g_3)M(g_2)M(g_1)\frac{L_P(g|h)}{M(g)} \tag{5}$$

and note that the fraction at the end of the RHS is independent of the chain.

When that chain is a route each of the $M(g_i)s$ on the RHS is 1, and we have

$$L_P(g_4|g_3+g_2+g_1+h)L_P(g_3|g_2+g_1+h)L_P(g_2|g_1+h)L_P(g_1|h) = \frac{L_P(g|h)}{M(g)} \tag{6}$$

so the RHS becomes independent of the route and we conclude that for any given (g,h,P), all routes to g have the same likeliness (over P and given h). To be precise,

Theorem 1. *Each route to g has a likeliness of* $\dfrac{L_P(g|h)}{M(g)} = \dfrac{\Sigma f^g f^h}{\Sigma f^h}$.

The LHS of (5) is the likeliness of the chain with batches g_1 , g_2 , g_3 , g_4 and so with via-points g_1 , g_2+g_1 , $g_3+g_2+g_1$ *and* $g_4+g_3+g_2+g_1$.
On the RHS, the $M(g_4)M(g_3)M(g_2)M(g_1)$ is the number of routes that are via all of those via-points, and the final $\dfrac{L_P(g|h)}{M(g)}$ can now be seen to be the likeliness per route to g.

We can thus conclude

Theorem 2.

(a) The likeliness of a chain is the number of routes via all of its via-points multiplied by the likeliness per route.

Or, alternatively,

(b)The likeliness of a chain is the sum of the likelinesses of the routes which are via all of its via-points.

Since each side of (6) is the likeliness per route, by taking the M(g) across to the other side we see that

Theorem 3.

(a) the likeliness of an integram is the number of routes to that integram multiplied by the likeliness per route.

Or, alternatively,

(b) The likeliness of an integram is the sum of the likelinesses of the routes to that integram.

8 Chain Rule revisited

If we look at the Chain Rule again, we can now view it in terms of routes (and chains).

$$L_P(g_2 + g_1|h) = \frac{M(g_2 + g_1)}{M(g_2)M(g_1)} L_P(g_2|g_1 + h) L_P(g_1|h)$$

This is about a chain to the integram $g = g_2 + g_1$.

The LHS is the likeliness of that integram.
The product of likelinesses on the RHS is the likeliness of the chain to g with via-points g_1 and g. The leading coefficient is pro-ratering up from the number of routes via those via-points to the number of routes to g:

- The likeliness of the chain is $L_P(g_2|g_1 + h)L_P(g_1|h)$
- This is the sum of the likelinesses of the $M(g_2)M(g_1)$ routes which are via its via-points
- So each route to $g_2 + g_1$ has likeliness $\frac{1}{M(g_2)M(g_1)} L_P(g_2|g_1 + h)L_P(g_1|h)$
- There are $M(g_2 + g_1)$ routes to $g_2 + g_1$, so the likeliness of $g_2 + g_1$ is $M(g_2 + g_1).\frac{1}{M(g_2)M(g_1)} L_P(g_2|g_1 + h)L_P(g_1|h)$

9 Examples

Example 1

Take $N = 3, g = (3, 1, 2), h = (2, 3, 5), P = S(3)$, and use successive applications of the Law of Succession.

In the first instance, use the canonical route, which has

generating sequence: $"1","1","1","2","3","3"$
via-points : (1,0,0),(2,0,0),(3,0,0),(3,1,0),(3,1,1),(3,1,2)

This route has the likeliness (remember to read from right to left)

$$\frac{(1+6)}{(3+15)}.\frac{(1+5)}{(3+14)}.\frac{(1+3)}{(3+13)}.\frac{(1+4)}{(3+12)}.\frac{(1+3)}{(3+11)}.\frac{(1+2)}{(3+10)} =$$
$$\frac{7}{18}.\frac{6}{17}.\frac{4}{16}.\frac{5}{15}.\frac{4}{14}.\frac{3}{13}$$

As an alternative, take the route with:

generating sequence $''3'',''1'',''1'',''2'',''3'',''1''$
via-points (0,0,1),(1,0,1),(2,0,1),(2,1,1),(2,1,2),(3,1,2)

This route has the likeliness

$$\frac{(1+4)}{(3+15)}.\frac{(1+6)}{(3+14)}.\frac{(1+3)}{(3+13)}.\frac{(1+3)}{(3+12)}.\frac{(1+2)}{(3+11)}.\frac{(1+5)}{(3+10)} =$$
$$\frac{5}{18}.\frac{7}{17}.\frac{4}{16}.\frac{4}{15}.\frac{3}{14}.\frac{6}{13}$$

It is a property of S(N) that there are no interactions across X_N, so the numerators in the two expressions are permutations of one-another -matching the permutation betwixt the generating sequences. This makes it easy to see that the two products are equal even though their individual terms are not.

Example 2

This is a more typical example.

As before, take $N = 3, g = (3, 1, 2), h = (2, 3, 5)$, but take $P = R(3)$ and use 'Great Likelinesses' to find the individual likelinesses. We shall also use the same routes as in Example 1.

In the first instance, use the canonical route, which has
generating sequence: $''1'',''1'',''1'',''2'',''3'',''3''$
via-points : (1,0,0),(2,0,0),(3,0,0),(3,1,0),(3,1,1),(3,1,2)

Using a sample of approximately 40 million distributions per calculation, this route has the likeliness

0.2377798 . 0.2253311 . 0.3078201 . 0.4561664 . 0.4452978 . 0.4355025
= 0.0014590(1)

As an alternative, again take the route with:

generating sequence $''3'',''1'',''1'',''2'',''3'',''1''$
via-points (0,0,1),(1,0,1),(2,0,1),(2,1,1),(2,1,2),(3,1,2)

This route has the likeliness

0.4268333 . 0.2431723 . 0.314448 . 0.4337043 . 0.4251456 . 0.2424419
= 0.0014590(2)

10 Discussion

The crux of the analysis is Theorem (1): that (within the context of any given problem) all routes to the same end-point have the same likeliness.

In a probabilistic context, this would be almost trivial. Essentially, it is just exchangability. It would be saying that the probability of tossing 2 Heads followed by a Tail is the same as that of throwing a Head, followed by a Tail, followed by another Head.

The reason why changing the order in this way has no effect when dealing with probabilities is because the numbers do not alter -and the commutivity of multiplication then takes care of the rest.

With likelinesses, however, the numbers *do* alter. Likeliness Theory is the theory of small samples, which means that every observation can have a profound effect: the likeliness of tossing the second Head in a sequence is not the same as that of tossing the first because of the effect of having tossed that first Head (this observation is extremely important, since it explains why the Multinomial Theorem does not apply to Likelinesses generally). Within such a context, the fact that changing the order of observations has no overall effect is actually a highly significant result. This independence of order is demonstrated by the examples.

The model that the theorems brings to mind is that of routes being fibres which are conducting likelinesses in the same way that wires conduct electricity. Bundle routes together to form a chain and the likeliness conducted by the chain is the sum of the likelinesses of the individual routes. The amount of likeliness being conducted to an end-point is the total amount being conducted by all the routes leading there.

This is, of course, only an informal way of thinking. But it does explain the leading coefficient in the Chain Rule: it's just counting fibres.

References

[1] Lovegrove,R,2013, 'The Fundamentals of Likelinesses', Lovegrove Mathematicals Research Report 2013-01,December 2013