Bounding changes in probability over time: It is unlikely that you will change your mind very much very often

Ernest Davis
Dept. of Computer Science
New York University
davise@cs.nyu.edu

November 10, 2013

Abstract

Under some circumstances, a probabilistic reasoner who encounters a sequence of observations will change his estimate of the likelihood of a proposition ϕ from very low to very high and back again many times. However, the prior probability of such a sequence of observations is necessarily very small. We compute an exact value of the maximal value of this probability, as a function of the amount of change and the number of changes. The calculation does not involve any independence assumptions over the observations.

1 Introduction

In some situations a probabilistic reasoner, updating her estimate of the likelihood of a proposition ϕ over time, will go many times from a very low estimate to a very high estimate and back again. For instance, consider a person executing a random walk on a straight line with absorbing boundaries at 0 and at 50. Let ϕ be the proposition, "I will reach the barrier at 0 (rather than the barrier at 50)." Then if the person is currently at 1, the likelihood of ϕ is 49/50; if the person is currently at 49, the likelihood is 1/50. If the person happens to wander back and forth between 49 and 1 many times, her evaluation of the likelihood of ϕ will likewise vary back and forth between 0.02 and 0.98.

However, it is very unlikely that this will happen to her. The point of this paper is that this observation generalizes: for any proposition ϕ in any stochastic setting, the prior probability that a sequence S of observations will occur such that $\mathrm{P}(\phi|S)$ will vary widely very many times must be very small. We compute an exact value for the maximum probability that the reasoner's estimate of the likelihood will vary up and down by a quantity α n times; as one would expect, the probability is a decreasing function of α and decreases exponentially as a function of n. The calculation does not involve any independence assumptions over the observations.

2 Basic Argument

We begin with a simple calculation. Suppose that, up to time t_1 the reasoner has made observations X, and that between times t_1 and t_2 the reasoner makes observations Y.

Let $p = \mathrm{P}(\phi|X)$ and $q = \mathrm{P}(\phi|X, Y)$.
Then $p = \mathrm{P}(\phi, Y|X) + \mathrm{P}(\phi, \neg Y|X) \geq q \cdot \mathrm{P}(Y|X)$. Thus $\mathrm{P}(Y|X) \leq p/q$.

Likewise $1 - p = \mathrm{P}(\neg\phi|X) = \mathrm{P}(\neg\phi, Y|X) + \mathrm{P}(\neg\phi, \neg Y|X) \geq (1 - q) \cdot \mathrm{P}(Y|X)$,
Thus $\mathrm{P}(Y|X) \leq (1 - p)/(1 - q)$.

Note that if p is substantially less than q, then the first bound above is significantly less than 1, and if q is substantially less than p then the second bound is significantly less than 1. That is, if there is a large change in either direction between $\mathrm{P}(\phi|X)$ and $\mathrm{P}(\phi|X, Y)$, then $\mathrm{P}(Y|X)$ must be small.

What we do in this paper is to show that this can be extended to many steps up and down; the probability that a sequence of events will occur that will cause the conditional probability of ϕ to go up and down and up and down ... is bounded tightly by the product of terms of the above form. We make a probabilistic argument to establish this bound and a calculus argument to compute a numeric value for the maximal value of this product. The proofs are all elementary.

3 Framework and Terminology

We will use the following general probabilistic framework. Let $X_1, X_2 \ldots$ be a infinite sequence of random variables corresponding to observations. There is additionally a Boolean random variable Q, which is the inference being drawn. We define the event ϕ as $Q = T$. We will be characterizing the behavior of the probability of ϕ given the values of $X_1 \ldots X_k$, and the classes of sequences of values that give rise to a specified behavior.

Let $D_0 = \{T, F\}$ and let D_i be the domain of X_i. For simplicity we will assume that each of the D_i's is finite; the extension to the case where these are measure spaces is straightforward, but complicates the exposition. The probability space $\mathcal{S}$ is thus the cross product $\mathcal{S} = D_0 \times D_1 \times D_2 \times \ldots$. An atomic event is a element in this space: an infinite tuple, assigning values to each of the random variables. For any atomic event $s \in S$, we will write s_i for the corresponding element of s.

We posit that there is an α-algebra $\mathcal{A} \subset 2^{\mathcal{P}}$, which is closed under complement and countable union, and has the following property: for any i, for any $v_i \in D_i$, the set $\{s|s_i = v_i\} \in \mathcal{A}$. We posit that there is a measure μ with the usual properties such that $\mu(\mathcal{S}) = 1$. Then for any events $E, F \in \mathcal{A}$, we define $\mathrm{P}(E) = \mu(E)$ and $\mathrm{P}(E|F) = \mu(E \cap F)/\mu(F)$, assuming $\mu(F) \neq 0$. In this paper, such a collection of $\mathcal{P}, \mathcal{A}, \mu, D_0, D_1 \ldots$ will be called an *observational stochastic framework.*

Definition 1. *An* initial sequence of observations *is either an event of the form $X_1 = v_1, \ldots, X_k = v_k$ or the null sequence, which is just the universal event $\mathcal{P}$.*

Note that, by construction of $\mathcal{A}$, each initial sequence is indeed an event in $\mathcal{A}$. Note also that since there are only countably many initial sequences, the union of any set of initial sequences is likewise an event in $\mathcal{A}$.

Definition 2. *An* initial event *will be any event that is the union of initial sequences. For any initial sequences E and F, we say that E is a* prefix of *F or F* extends *E, written $E \prec F$, just if $F \subset E$ (viewed as sets of atomic events). Note that, for any two initial sequences E and F, either $E \prec F$, $F \prec E$, or E and F are disjoint (again, viewed as sets of atomic events).*

Definition 3. *Let $\mathcal{S}$ be a set of initial sequences. An initial sequence $W \in \mathcal{S}$ is* prefix-free in *$\mathcal{S}$ if W has no prefix in $\mathcal{S}$. Let E be an initial event. An initial sequence $W \subset E$ is* prefix-free in *E if W is not a proper extension of any initial sequence that is a subset of E. (Viewed as a set of atomic events, W is a maximal initial sequence).*

Lemma 1. *Any initial event E is equal to the union of its prefix-free initial sequences.*

Proof: By definition E is the union of a set $\mathcal{S}$ of initial sequences. Let $\mathcal{C}$ be the set of sequences that are prefix-free in $\mathcal{S}$. Then every sequence that is in $\mathcal{S}$ but not in $\mathcal{C}$ is a subset of some sequence in $\mathcal{C}$; hence $E = \bigcup \mathcal{S} = \bigcup \mathcal{C}$. ∎

We can now restate our initial calculation in terms of this framework.

Lemma 2. *Let S be any initial sequence of observations. Let $p = P(\phi|S)$. For some $q \in [0, 1]$, let $\mathcal{F}$ be the set of all initial sequences W such that $S \prec W$ and $P(\phi|W) \geq q$, and let F be the union of the events in $\mathcal{F}$. Then:*

a. $P(\phi|F) \geq q$.

b. $P(F|S) \leq p/q$.

Proof:

a. Let $\mathcal{C}$ be the set of all the prefix-free initial sequences in F. By lemma 1, $F = \bigcup \mathcal{C}$. However the elements of $\mathcal{C}$ are all pairwise disjoint; hence $\mathrm{P}(F) = \sum_{W \in \mathcal{C}} \mathrm{P}(W)$.

Then

$$\mathrm{P}(\phi|F) = \mathrm{P}(\phi, F)/\mathrm{P}(F) = \sum_{W \in \mathcal{C}} \mathrm{P}(\phi, W)/\mathrm{P}(F) = \sum_{W \in \mathcal{C}} \mathrm{P}(\phi|W) \cdot \mathrm{P}(W)/\mathrm{P}(F) \geq q \cdot \sum_{W \in \mathcal{C}} \mathrm{P}(W)/\mathrm{P}(F) = q$$

b. As in the calculation at the beginning of section 2. Let $G = S \setminus F$. Since $F \subset S$,
$p = \mathrm{P}(\phi|S) = \mathrm{P}(\phi|F) \cdot \mathrm{P}(F|S) + \mathrm{P}(\phi|G) \cdot \mathrm{P}(G|S) \geq q \cdot \mathrm{P}(F|S)$.

∎

Lemma 3. *Let S be any initial sequence of observations. Let $p = P(\phi|S)$. For some $q \in [0, 1]$, let $\mathcal{F}$ be the set all initial sequences W such that W extends S and $P(\phi|W) \leq q$, and let F be the union of the events in $\mathcal{F}$. Then:*

a. $P(\phi|F) \leq q$.

b. $P(F|S) \leq (1 - p)/(1 - q)$.

Proof: a. Analogous to the proof of lemma 2.a.

b. $(1 - p) = \mathrm{P}(\neg\phi|S) = \mathrm{P}(\neg\phi|F)\mathrm{P}(F|S) + \mathrm{P}(\neg\phi|G)\mathrm{P}(G|S) \geq (1 - q)\mathrm{P}(F|S)$.

∎.

Corresponding to lemmas 2 and 3, we define the function $\gamma(p, q) = \min(p/q, (1 - p)/(1 - q))$ for $p, q \in [0, 1]$. Note that if $p < q$ then $\gamma(p, q) = p/q < 1$ and if $p > q$, then $\gamma(p, q) = (1-p)/(1-q) < 1$. Note also that if $p < q$, then $\gamma(p, q)$ is an increasing function of p and a decreasing function of q and if $p > q$ then γ is a decreasing function of p and an increasing function of q. Thus, if you hold one of the arguments fixed and move the other toward it, the value of γ increases.

4 Large alternations

Definition 4 characterizes what it means for a sequence to go up and down a great deal many times.

Definition 4. *Let $\alpha \in (0,1)$ and let $r_0, r_1 \dots r_n$ be a sequence of numbers in [0,1]. We say that $r_0 \dots r_n$ is an* α-alternating M-sequence *if $r_1 \geq r_0+\alpha$, $r_2 \leq r_1+\alpha$, and in general for $k = 0 \dots \lfloor n/2 \rfloor$ $r_{2k+1} \geq r_{2k}+\alpha$ and $r_{2k+2} \leq r_{2k+1}-\alpha$. It is an* α-alternating W-sequence *if $r_1 \leq r_0-\alpha$, $r_2 \geq r_1+\alpha$, and in general, $r_{2k+1} \leq r_{2k}-\alpha$ and $r_{2k+2} \geq r_{2k+1}+\alpha$.(M and W because of the shapes of the letters.) An* α-alternating sequence *is either an α-alternating M-sequence or an α-alternating W-sequence.*

For example the sequence $\langle 0.1, 0.7, 0.2, 0.9, 0.3, 0.8, 0.2 \rangle$ is a 0.5-alternating M-sequence, and the sequence $\langle 0.7, 0.2, 0.9, 0.3, 0.8, 0.2 \rangle$ is a 0.5-alternating W-sequence.

In any sequence $r_0 \dots r_n$, a *step* is a sequential pair r_i, r_{i+1}. We measure the *length* of the sequence as n, the the number of steps, rather than $n+1$, the number of elements. We will similarly refer to sequences as "even" or "odd" in terms of the number of steps.

Definition 5. *For any alternating sequence $r_0, r_1 \dots r_n$, we define $\Gamma(r_0, r_1 \dots r_n) = \Pi_{i=1}^{n} \gamma(r_{i-1}, r_i)$.*

Definition 6. *For any $\alpha \in (0,1)$ we define the function $\delta_\alpha(p,q) = \gamma(p, p+\mathrm{sign}(q-p)\cdot\alpha)$. That is, if $p < q$ then $\delta_\alpha(p,q) = p/(p+\alpha)$; if $p > q$ then $\delta_\alpha(p,q) = (1-p)/(1+\alpha-p)$;*

For any α-alternating sequence $r_0 \dots r_n$ we define $\Delta_\alpha(r_0 \dots r_n) = \Pi_{i=1}^{n} \delta_\alpha(r_{i-1}, r_i)$

Note that, if $|r_{i+1} - r_i| \geq \alpha$ then $\delta_\alpha(r_i, r_{i+1}) \geq \gamma_\alpha(r_i, r_{i+1})$. Thus if $r_0 \dots r_n$ is an α-alternating series, then $\Delta_\alpha(r_0 \dots r_n) \geq \Gamma(r_0 \dots r_n)$.

Note also that if $r_0, \dots, r_n$ is an α-alternating M-sequence then $1-r_0, \dots, 1-r_n$ is an α-alternating W-sequence, and that $\Gamma(1-r_0, \dots 1-r_n) = \Gamma(r_0, \dots r_n)$. Therefore it will suffice for most purposes to consider M-sequences; all the results transfer to W-sequences.

We define the quantity χ_α^n to be the maximum value attained by $\Gamma(r_0, r_1 \dots r_n)$ over all α-alternating sequences. We define the quantity Ψ_α^n as the maximum value attained by $\Delta_\alpha(r_0, r_1 \dots r_n)$ over all α-alternating sequences. Since $\Gamma(r_0 \dots r_n) \leq \Delta(r_0 \dots r_n)$ it follows that $\chi_\alpha^n \leq \Psi_\alpha^n$.

We will show below (theorem 10) that in fact $\chi_\alpha^n = \Delta_\alpha^n$, but until we prove that, it is convenient to treat them separately.

Lemma 4. *For any n, $\chi_\alpha^n \geq (1-\alpha)^n/(1+\alpha)^n$.*

Proof:

$\Gamma(r_0, r_1 \dots r_n)$ attains this value in the case that the sequence alternates between low values of $(1-\alpha)/2$ and high values of $(1+\alpha)/2$; that is, a range of exactly α centered at 1/2. In this case $\gamma(r_{i-1}, r_i)$ is always equal to $(1-\alpha)/(1+\alpha)$. ∎

Theorem 5. *If n is even then $\chi_\alpha^n = \Psi_\alpha^n = (1-\alpha)^n/(1+\alpha)^n$.*

Proof:

We need to show that for any α-alternating sequence $r_0, r_1 \dots r_n$, $\Delta_\alpha(r_0, r_1 \dots r_n) \leq (1-\alpha)^n/(1+\alpha)^n$. We will consider the case of M-sequences; the analysis of W-sequences is exactly analogous.

First, consider the case $n = 2$. Consider a sequence of the form r_0, r_1, r_2 with the constraints $r_0 \leq r_1 - \alpha$, $r_2 \leq r_1 - \alpha$. For any fixed value of r_1, it is obvious that $\gamma(r_0, r_1) = r_0/(r_0+\alpha)$ is maximized when $r_0 = r_1 - \alpha$. The value of $\gamma(r_1, r_2)$ is equal to $(1-r_1)/(1-(r_1-\alpha))$ independently of r_2. In this case $\Delta_\alpha(r_0, r_1, r_2) = (r_1-\alpha)(1-r_1)/(r_1(1-(r_1-\alpha)))$. A straightforward calculation shows this function of r_1 attains its maximum when $r_1 = (1+\alpha)/2$ and that there the value of the function is $(1-\alpha)^2/(1+\alpha)^2$.

If n is even and n is greater than 2 then $\Delta(r_0 \dots r_n) = \Pi_{k=1}^{n/2} \Delta(r_{2k-2}, r_{2k-1}, r_{2k}) \leq (1-\alpha)^n/(1+\alpha)^n$ since each of these factors corresponds to the case $n = 2$. ∎

The proof that $\chi^n_\alpha = \Psi^n_\alpha$ for odd n is harder. We first need a couple of lemmas.

Lemma 6. *For any α and n there exists a sequence $r_0, r_1 \ldots r_n$ such that $\chi^n_\alpha = \Gamma(r_0 \ldots r_n)$. and a sequence $s_0, s_1 \ldots s_n$ such that $\chi^n_\alpha = \Gamma(s_0 \ldots s_n)$.*

Proof: Γ and Δ are continuous functions, and χ and Ψ are their suprema over a compact domain. The maximum is therefore attaned. ∎

To simplify notation and terminology, we will assume for the rest of the section that we have in mind some fixed value $\alpha \in (0,1)$. Thus, all the notations below are defined relative to that value of α. We will likewise omit the subscript α on δ and Δ.

Definition 7. *An α-alternating M-sequence $r_0 \ldots r_n$ is* exact with centerpoint p *if for $k = 0 \ldots n-1$, $r_k = p + (-1)^{k+1} \cdot \alpha/2$. Likewise an α-alternating W-sequence $r_0 \ldots r_n$ is* exact with centerpoint p *if for $k = 0 \ldots n-1$, $r_k = p + (-1)^k \cdot \alpha/2$.*

That is, an exact α-alternative M-sequence with centerpoint p has the form $r_0 = p - \alpha/2$, $r_1 = p + \alpha_2$, $r_2 = p - \alpha/2$ and so on up through r_{n-1}. The final value r_n need only satisfy the constraint $r_n \leq p - \alpha/2$ if n is even and $r_n \geq p - \alpha/2$ if n is odd. An exact α-alternative W-sequence is the same thing, reversing addition and subtraction.

For any integer $n > 0$, define the function $\mu_n(p) = \Delta(r_0, r_1, \ldots, r_n)$ where $r_0 \ldots r_n$ is an exact α-alternating M-sequence centered at p; and $\omega_n(p) = \Delta(r_0, r_1, \ldots, r_n)$ where $r_0 \ldots r_n$ is an exact α-alternating W-sequence centered at p.

If $n = 2k$ is even then $\mu_n(p) = \omega_n(p) = \gamma(p - \alpha/2, p + \alpha/2)^k \cdot \gamma(p + \alpha/2, p - \alpha/2)^k$.
If $n = 2k+1$ is odd then $\mu_n(p) = \gamma(p - \alpha/2, p + \alpha/2)^{k+1} \cdot \gamma(p + \alpha/2, p - \alpha/2)^k$
and $\omega_n(p) = \gamma(p - \alpha/2, p + \alpha/2)^k \cdot \gamma(p + \alpha/2, p - \alpha/2)^{k+1}$

Lemma 7. *If n is even, then the function $\mu_n(p) = \omega_n(p)$ is an increasing function over $(0, 1/2)$, reaches a maximum at $p = 1/2$, and is a decreasing function over $(1/2, 1)$.*

Proof: $\mu_2(p) = (p - \alpha/2)(1 - p - \alpha/2)/(p + \alpha/2)(1 - p + \alpha/2)$. The numerator of $d\mu/dp$ is equal to $\alpha(1 - 2p)$, and the denominator is always positive; thus, the function is increasing over (0,1/2) and decreasing over (1/2,1). Since $\mu_{2k}(p) = (\mu_2(p))^k$ the same is true of μ_{2k}.

Lemma 8. *If n is odd, then the function $\mu_n(p)$ is an increasing function over (0,1/2) and the function $\omega_n(p)$ is an decreasing function over (1/2,1).*

Proof: We have $\mu_n(p) = \mu_{n-1}(p) \cdot (p - \alpha/2)/(p + \alpha/2)$. Since $n-1$ is even, by lemma 7 $\mu_{n-1}(p)$ is increasing over (0,1/2); and $(p - \alpha/2)/(p + \alpha/2)$ is obviously increasing over (0,1/2); so the product is likewise. Similarly $\omega_n(p) = \omega_{n-1}(p) \cdot (1 - p - \alpha/2)/(1 - p + \alpha/2)$ is the product of two expressions decreasing over (1/2,1).

Lemma 9. *For any n, the function $\Delta(r_0 \ldots r_n)$ achieves its maximum Ψ^n_α for some exact α-alternating M sequence.*

Proof by contradiction. Assume that $r_0 \ldots r_n$ is not exact and attains the maximum value of Δ; we will show that this is inconsistent.

We say that the step of the sequence r_i, r_{i+1} is a "large gap" if $|r_i - r_{i-1}| > \alpha$. Since $r_0 \ldots r_n$ is not exact, it must contains some large gap other than the last step.

Suppose that the first large gap occurs from r_0 to r_1. Since $r_1 > r_0$ we can increase the value of Δ by changing r_0 to $r_1 - \alpha$. This increase the value of $\delta(r_0, r_1)$ without changing any of the rest of the sequence so Δ increases.

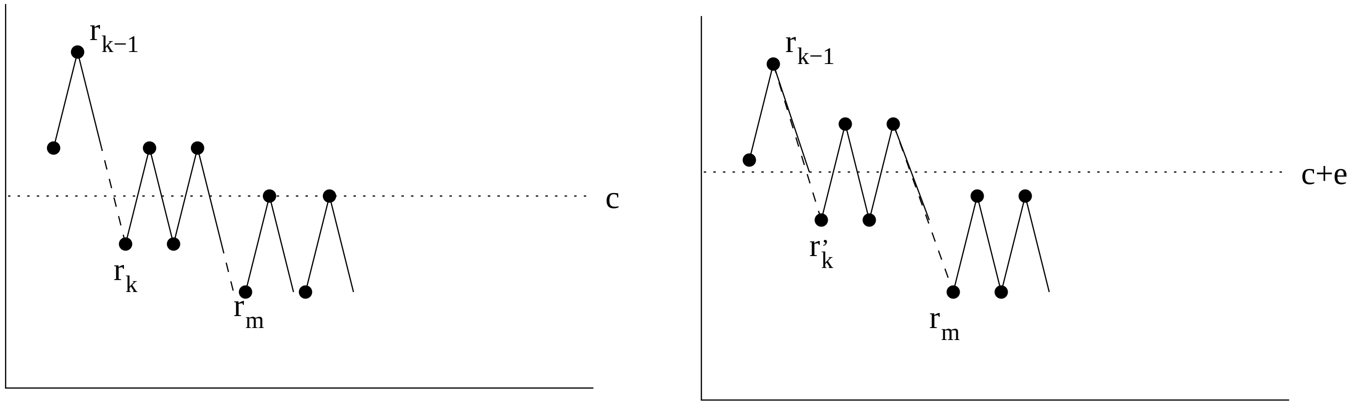

Figure 1: Moving the points in a subsequence by e

Similarly, an optimal sequence cannot have two large gaps in a row. Suppose a sequence has two large gaps in a row; that is for some j either $[r_{j-1} < r_j - \alpha$ and $r_{j+1} < r_j - \alpha]$ or $[r_{j-1} > r_j + \alpha$ and $r_{j+1} > r_j + \alpha]$. In either case, moving r_j "toward" r_{j-1} and r_{j+1} will increases $\delta(r_j, r_{j+1})$ without affecting either $\delta(r_{j-1}, r_j)$ or the other terms in the product, and so will increase Δ.

We can therefore partition the n steps of the sequence into subsequences such that every subsequence contains at least 2 edges, is exact, and ends in a large gap (the final subsequence may not end in a large gap).

Let $r_k \ldots r_m$ be any such subsequence; thus r_{m-1}, r_m is a large gap, and so is r_{k-1}, r_k if $k > 1$. Let $d = \min(|r_k - r_{k-1}| - \alpha, |r_m - r_{m-1}| - \alpha) > 0$, the smaller of the gaps on the two sides. Let c be the centerpoint of $r_k \ldots r_m$. We note now that we can move the points $r_k \ldots r_{m-1}$ in concert by any amount $-d \leq e \leq d$, and the sequence remains a valid α-alternating sequence (figure 1). That is, let $r'_i = r_i + e$ for $i = k \ldots m-1$ and $r'_i = r_i$ for $i < k$ and $i \geq m$. Then $r'_0 \ldots r'_n$ is still an α-alternating sequence, and the subsequence $r'_k \ldots r'_m$ is an exact α-alternative sequence with center $c + e$.

However, if $r_k \ldots r_m$ is an M-sequence then $\Delta(r_0 \ldots r_n) = \Delta(r_0 \ldots r_{k-1}) \cdot \mu_{m-k}(c) \cdot \Delta(r_m, \ldots r_n)$ and $\Delta(r'_0 \ldots r'_n) = \Delta(r_0 \ldots r_{k-1}) \cdot \mu_{m-k}(c+e) \cdot \Delta(r_m, \ldots r_n)$. Since $r_0 \ldots r_n$ achieves the maximal value of Δ, we must have $\mu_{m-k}(c) \geq \mu_{m-k}(c+e)$; since e is an arbitrary small value, c must be a local maximum of μ_{m-k}. By the same argument, if $r_k \ldots r_m$ is an W-sequence then c must be a local maximum of ω_{m-k}.

Now let us consider the first two of these subsequences: $r_0 \ldots r_k$ and $r_k \ldots r_m$. Let b be the centerpoint of $r_0 \ldots r_k$ and let c be the centerpoint of $r_k \ldots r_m$. By assumption, $r_0 \ldots r_k$ is an M-sequence so $b \geq 1/2$.

There are now two cases (figure 2):

Case 1: k is even. In that case, $b = 1/2$, since by lemma 7 that is the unique local maximum of μ_k. However, the last step r_{k-1}, r_k is downward, so $r_k < r_{k-1} - \alpha$. Therefore $c < 1/2$. But $r_k \ldots r_m$ is an M-sequence and c is not a local maximum of μ_{m-k}, which is a contradiction.

Case 2: k is odd. In that case, $b \geq 1/2$. since no value less than 1/2 can be a local maximum. However, the last step r_{k-1}, r_k is upward. so $r_k > r_{k-1} + \alpha$. Therefore $c > b \geq 1/2$. But $r_k \ldots r_m$ is an W-sequence and c is not a local maximum of ω_{m-k}, which is a contradiction.

∎.

Theorem 10. *Let n be an positive integer and let $k = \lfloor n/2 \rfloor$. If n is even, let $p_c = 1/2$; if n is odd, let $p_c = k + 1 - \sqrt{k^2 + k + (\alpha^2/4)}$. Then both $\Gamma(r_0 \ldots r_n)$ and $\Delta(r_0 \ldots r_n)$ achieve their common*

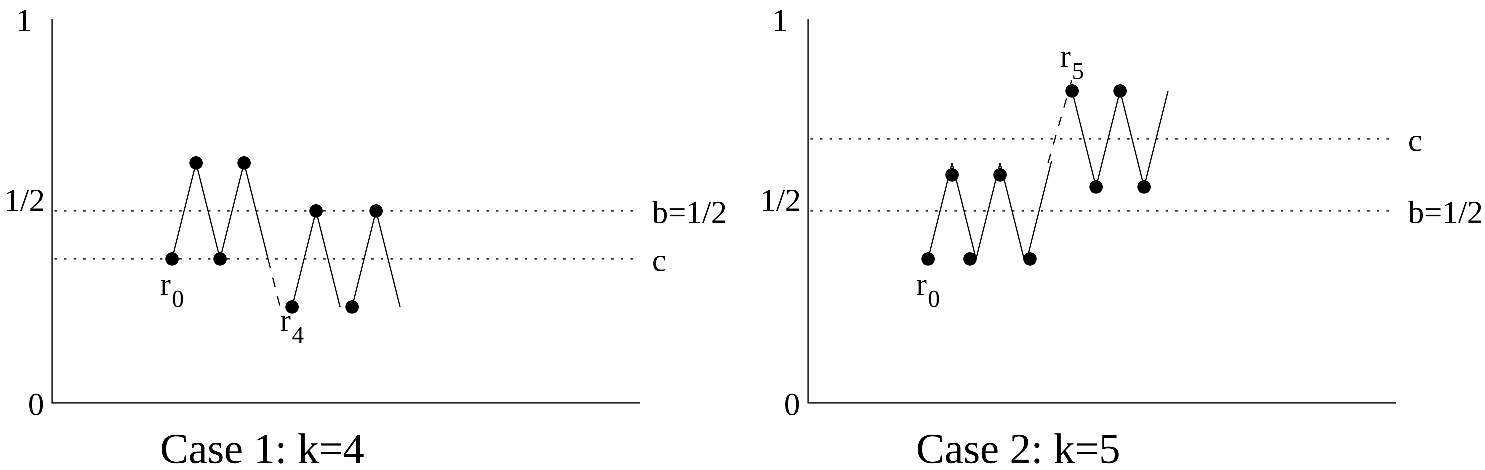

Figure 2: Two cases in lemma 9

maximum $\chi_\alpha^n = \Psi_\alpha^n$ *for the arguments* $p_c - \alpha/2, p_c + \alpha/2, p_c - \alpha/2 \ldots$

If n *is even,*

$$\chi_\alpha^n = \Psi_\alpha^n = \frac{(p_c - \alpha/2)^k \cdot (1 - p_c - \alpha/2)^k}{(p_c + \alpha/2)^k \cdot (1 - p_c + \alpha/2)^k} = \frac{(1-\alpha)^n}{(1+\alpha)^n}$$

If n *is odd,*

$$\chi_\alpha^n = \Psi_\alpha^n = \frac{(p_c - \alpha/2)^{k+1} \cdot (1 - p_c - \alpha/2)^k}{(p_c + \alpha/2)^{k+1} \cdot (1 - p_c + \alpha/2)^k}$$

Proof: Directly from combining lemmas 5, 6, and 9. The optimal value of p_c can be obtained by finding the zero of the derivative of $\mu_n(p)$; this boils down to a quadratic equation in p. (There is a second zero outside the range (0,1).)

5 Alternating likelihood

We now combine the probabilistic analysis of section 3 with the algebra of section 4.

Definition 8. *Let W be an initial sequence of observations. We say that* the likelihood of ϕ M-alternates by α n times in W *if there exist* $W_0 \prec W_1 \prec \ldots \prec W_{n-1} \prec W_n = W$ *such that the sequence* $P(\phi|W_0), P(\phi|W_1), \ldots P(\phi|W_n)$ *is an α-alternating M-sequence.*

Theorem 11. *Let $\mathcal{Q}$ be the set of all initial sequences W such that the likelihood of ϕ M-alternates by α n times in W. Let $Q = \bigcup \mathcal{Q}$. Then $P(Q) \leq \Psi_\alpha^n$.*

Proof: Let $\mathcal{Q}$ and Q be as in the statement of the theorem. Let $\mathcal{Q}'$ be the set of all prefix-free sequences in $\mathcal{Q}$.

For any element W of $\mathcal{Q}'$, let $W_0 \prec W_1 \prec \ldots \prec W_n$ be a sequence satisfying the condition of definition 8. For $k = 0 \ldots n$ let $q_k = \mathrm{P}(\phi|W_k)$; then the sequence $q_0 \ldots q_n$ is an α-alternating M-sequence.

For $k = 1 \ldots n$ let $\mathcal{G}_k$ be the set of all extensions E of W_{k-1} such that the likelihood of ϕ M-alternates by α k-times in E. Let $\mathcal{F}_k$ be the set of sequences that are prefix-free in $\mathcal{G}_k$. Let $F_k = \bigcup \mathcal{F}_k$. Clearly, since each such E is an extension of W_{k-1}, they are each subsets of W_{k-1}; therefore F_k is a subset of W_{k-1}. Since W_k is an element of $\mathcal{F}_k$, it follows that $W_k \subset F_k$. Therefore we have $W_0 \supset F_1 \supset W_1 \supset F_2 \ldots \supset F_n \supset W_n$. Therefore $\mathrm{P}(W_n) = \mathrm{P}(W_0) \cdot \mathrm{P}(F_1|W_0) \cdot \mathrm{P}(W_1|F_1) \cdot \mathrm{P}(F_2|W_1) \cdot \ldots \cdot \mathrm{P}(W_n|F_n)$.

Note also that, since F_k is the disjoint union of the elements in $\mathcal{F}_k$, we have $\sum_{E \in \mathcal{F}_k} \mathrm{P}(E|F_k) = 1$.

We above defined $q_k = \mathrm{P}(\phi|W_k)$. Suppose that k is even. By definition of $\mathcal{F}_k$ we know that, for every $E \in \mathcal{F}_{k+1}$, $\mathrm{P}(\phi|E) \geq q_k + \alpha$. Since F_{k+1} is the disjoint union of the elements of $\mathcal{F}_k$, by lemma 2.a $\mathrm{P}(\phi|F_{k+1}) \geq q_k + \alpha$. It follows from lemma 2.b that $\mathrm{P}(F_{k+1}|W_k) \leq \gamma(q_k, q_k + \alpha)$. By the corresponding argument, if k is odd, then $\mathrm{P}(F_{k+1}|W_k) \leq \gamma(q_k, q_k - \alpha)$.
Thus the product $\mathrm{P}(F_1|W_0) \cdot \mathrm{P}(F_2|W_1) \cdot \ldots \cdot \mathrm{P}(F_n|W_{n-1}) \leq \Delta_\alpha(q_0, q_1 \ldots q_n) \leq \Psi_\alpha^n$.
Therefore, we have $\mathrm{P}(W_n) \leq \Psi_\alpha^n \cdot \mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_n|F_n)$.

Consider now the entire tree of W's and F's as a whole. For $k = 0 \ldots n$, let $\mathcal{H}_k$ be the prefix-free collection of sequences in which the likelihood of ϕ α-alternates at least k times (thus $\mathcal{H}_n = \mathcal{Q}'$). For any $W \in \mathcal{H}_k$ define $F_1, W_1 \ldots F_k, W_k = W$ as before. We claim that

$$\sum_{W \in \mathcal{H}_k} \mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_k|F_k) = 1$$

Proof of claim by induction: For k=1, since $F_1 = \bigcup \mathcal{H}_1$ the statement is immediate. Suppose it is true for $k-1$. For any $W_{k-1} \in \mathcal{H}_{k-1}$ let $\Theta_k(W_{k-1}) = \{W_k \in \mathcal{H}_k | W_{k-1} \prec W_k\}$, the set of extensions of W_{k-1} with k alternations. We have observed above that $\sum_{W_k \in \Theta_k(W_{k-1})} \mathrm{P}(W_k|F_k) = 1$. Therefore

$$\sum_{W_k \in \mathcal{H}_k} \mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_k|F_k) =$$

$$\sum_{W_{k-1} \in \mathcal{H}_{k-1}} \left[\sum_{W_k \in \Theta_k(W_{k-1})} \mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_{k-1}|F_{k-1}) \cdot \mathrm{P}(W_k|F_k) \right] =$$

$$\sum_{W_{k-1} \in \mathcal{H}_{k-1}} \left[\mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_{k-1}|F_{k-1}) \cdot \left[\sum_{W_k \in \Theta_k(W_{k-1})} \mathrm{P}(W_k|F_k) \right] \right] =$$

$$\sum_{W_{k-1} \in \mathcal{H}_{k-1}} \mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_{k-1}|F_{k-1}) \cdot 1 = 1$$

Finally, we have

$$\mathrm{P}(Q) = \sum_{W \in \mathcal{H}_n} \mathrm{P}(W) \leq \Psi_\alpha^n \sum_{W \in \mathcal{H}_n} \mathrm{P}(W_1|F_1)\mathrm{P}(W_2|F_2) \ldots \mathrm{P}(W_n|F_n) = \Psi_\alpha^n$$

▮

It is easily shown that the upper bound in theorem 11 is tight. In fact the following general result holds:

Theorem 12. *Let $r_0 \ldots r_n$ be any alternating sequence. Then there exists a observational stochastic framework such that, for $k = 1 \ldots n$,*

a. $P(\phi) = r_0$

b. $P(\phi|X_1 = v_1 \ldots X_k = v_k) = r_k$

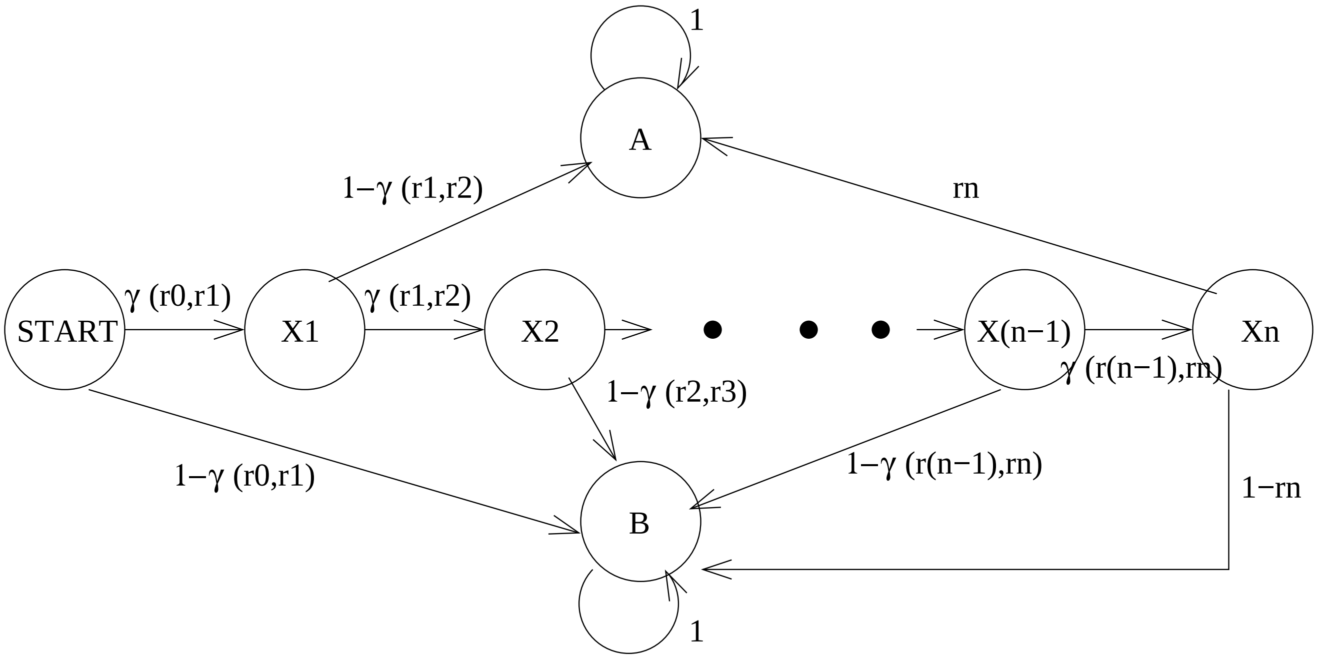

Figure 3: Markov process for theorem 12

c. $P(X_1 = v_1 \dots X_k = v_k) = \Gamma(r_0 \dots r_k)$.

(With some slight changes of definitions and of the proof below, the result holds for non-alternating sequences as well.)

Proof: Figure 3 below shows a Markov process satisfying the theorem for the case of an M-sequence of odd length; the other cases are analogous. ϕ is the proposition, "The process will end in state *A*." It is easily verified that the conditions of the theorem are satisfied. ∎

Corollary 13. *For any n, there exists a observational stochastic framework in which the set of M-sequences in which the likelihood of ϕ α-alternates n times has a total probability of Ψ^n_α.*

Proof: Immediate from theorem 10 and theorem 12. ∎