

## Oxford Croquet

---

Technical

# Bayesian Updating Simplified

by Louis Nel

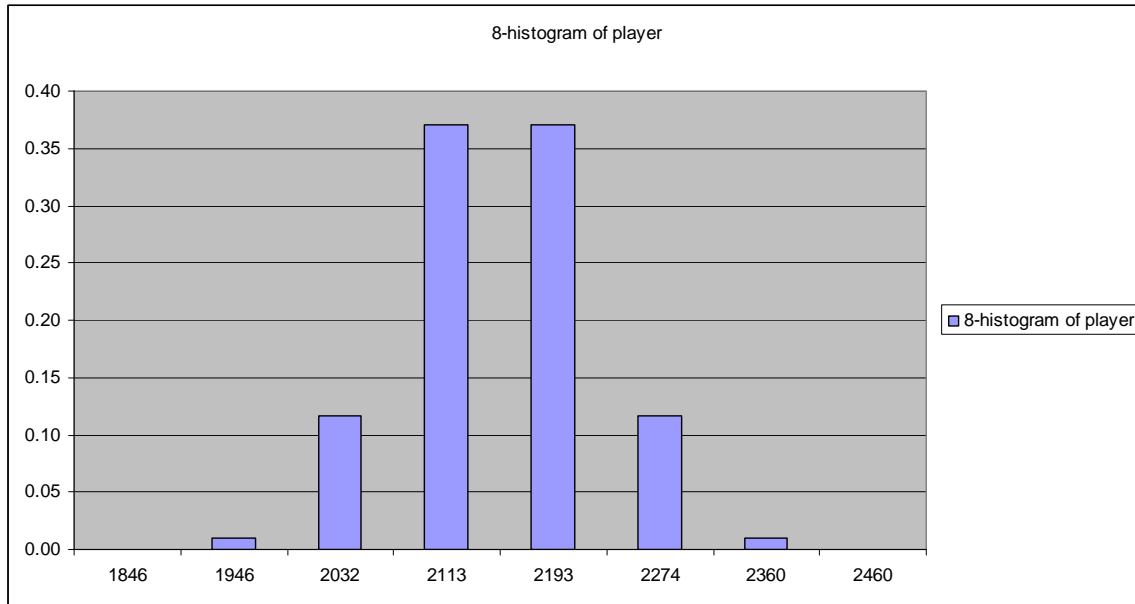
In Bayesian ranking the expected performance level of a player is represented by a Bell Curve. In my article 'Bayesian ranking for croquet' (<http://www.oxfordcroquet.com/tech/nel-by/>) the postgame updating algorithm is expressed concisely and precisely in terms of integrals that involve the Bell Curves of the players. Unfortunately, that leaves all but expert readers in the dark about how the algorithm really works.

The purpose of this article is to make this procedure understandable to a much wider audience. It begins with an overview that uses just graphics (histograms) and plain English (no formulas). This overview gives an idea of what is going on. Some of the inherent elegance of the Bayesian method already becomes visible. For many people this will likely be as much explanation as they have appetite for. They could enjoy the effects of Bayesian ranking in much the same way that most people enjoy television -- without a detailed knowledge of the underlying technology. In the further sections the calculations are explained in detail.

The introductory section on Probability Basics contains nothing new. It is there to make the updating procedure readable to people with some mathematical background but without having the specialized probability facts at their finger tips. The essential content of this article is the short piece that follows this relatively lengthy introduction. Somebody who is able to program the official Croquet Grading System should, in the light of this article, be able to program Bayesian ranking just as readily. The key formulas are explicitly stated and can be incorporated almost word for word into a computer program.

### Overview

Suppose the ranking data of player X consists of a Bayesian Grade of 2153 and a Standard Deviation of 74. For computational purposes the Bell Curve of this player can be replaced by the chart shown in the following picture.



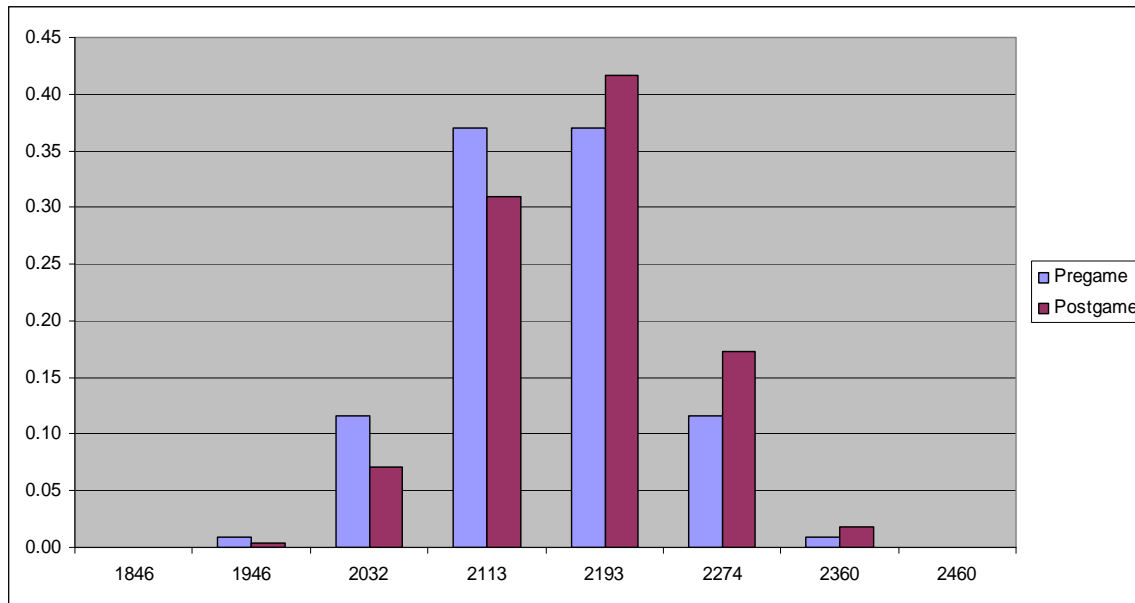
This chart shows a rectangle of height 0.37 at each of the two points 2113 and 2193. That is to signify a probability of 0.37 for the player to perform at level 2113 and also at level 2193. Similarly, the rectangles of height 0.12 at the points 2032 and 2274 and of height 0.01 at the points 1946 and 2360 signify the probabilities that the player will perform at those levels. Actually, this player also has the probability 0.0001 of performing at the levels 1846 and 2460, but this is too small to show properly in the chart. So this should simply be kept in mind. The chart implies that there is no possibility for the player to perform at any other level than the mentioned eight points. This is not to suggest that the player in real life will make such capricious leaps in performance level. It means merely that for computational purposes it is as if the player's performance will be as depicted. (In the mechanics of rigid bodies the mass of the body is often assumed similarly to be concentrated, for computational purposes, at its center of gravity).

A chart like the above one will be called the [pregame histogram](#) of the player. It can readily be constructed from the ranking data of the player. The one shown has a total of 8 nodes i.e. values like 1846, 1946, ...2460 marked on the horizontal reference line, symmetric with respect to the unmarked Mean value 2153, which is the Bayesian Grade of the player. The number N of nodes is relatively unimportant. For each chosen N the node positions and associated probabilities are quickly derived from pre-calculated nodes and weights available in the published literature on numerical analysis.

The symmetry about the Mean value implies that the player is deemed equally likely to perform above the Mean value 2153 than below it. The Bayesian Grade, as Mean value, is in fact the weighted average obtained when every node is weighted by the probability associated with it. For an intuitive idea of what this means, it is helpful to imagine the rectangles formed out of a uniform plate, so that the physical weight of a rectangle is proportional to its height. Imagine the horizontal line to be a rigid weightless beam. Then the Mean is the position on this beam for a fulcrum on which the beam will be in balance under the weights of the rectangles.

Suppose now that this player X with data (2153,74) as depicted plays a game against another player Y with data (2479, 68) and that player X wins the game. The task of the update procedure (in any ranking system) is to calculate updated ranking data for the two players in the light of the information provided by the game result. In the present case, the belief about the player's performance expressed by the pregame histogram needs to be revised. It is reasonable to conclude that the winner's probability of playing at level 2193 (which is above the Mean) must be higher than the pregame belief of 0.37 and

accordingly that his probability of playing at level 2113 (below the Mean) must be lower than 0.37. A similar consideration applies at every node. **It is remarkable that these surmised higher and lower probabilities can actually be calculated in terms of the pregame histogram data.** The calculation is done via standard probability facts (Sum Rule, Conditional Probability, Product Rule) which are introduced in detail in the introductory section. Here we just note that if these calculated revised probabilities are also charted, then we get a **postgame histogram**. The two histograms are shown together in the following picture.



It is noteworthy that to the left of the pregame Mean 2153 the revised associated probabilities (heights of the red rectangles) are all lower than the original ones (heights of the blue rectangles) while to the right of the Mean the postgame probabilities are higher than the pregame ones. One can see geometrically that the point on the horizontal reference line at which the red rectangles will be balanced is to the right of the point at which the blue rectangles will be balanced. This means **the postgame Mean will be larger than the pregame Mean**. Similar remarks, with opposite effect, apply to the loser. The role of the Standard Deviation can geometrically be described as follows. A larger SD flattens the histogram -- makes it spread out rather than having a sharp peak at the Mean. That turns out to make the weight-shift more pronounced and it results in a greater increase in the Mean for the winner than arises from a smaller SD. Similarly for the loser i.e. the larger his SD the more his reduction in Mean.

So the Bayesian postgame update procedure can be summarized in terms of the following three steps.

- Use the pregame ranking data (Mean and SD) to construct a pregame histogram for each player.
- Calculate the postgame histogram for each player.
- Calculate for each player the Mean and SD of the postgame histogram. That gives the updated ranking data.

This updating procedure involves no man-made parameter. Of course, any ranking system needs man-made initializing parameters to get going.

## 1. Probability basics

This section introduces basic probability rules for convenient reference. It contains nothing peculiar to Bayesian ranking. A reader who already knows probability theory will merely have to pick up notation.

### 1.1 Events.

For the purpose of this article an **event** is a happening which may or may not take place. When two players A and B are to play a game, then the game result 'A beat B' is an event in the above sense – it may or may not happen.

For purposes of illustration we can associate events with the roll of a die as follows. Let  $S = \{1,2,3,4,5,6\}$  denote the set of all possible outcomes. Consider the subset  $E = \{1,3,5\}$  of  $S$ . If the die is rolled, the outcome may or may not be a member of  $E$ . Thus  $E$  determines an event. In this way we can associate with every subset of  $S$  an event. The same applies to other sets  $S$  -- any subset  $E$  of  $S$  determines an event.

### 1.2 Probability of an event.

For every event  $E$  there is an associated real number  $p_E$  between 0 and 1 which expresses the degree of certainty that the event will take place. That number is called the **probability of  $E$** . If  $E$  is certain to take place, then  $p_E = 1$ . If  $E$  is certain not to take place, then  $p_E = 0$ .

For an event  $E$  associated with the roll of a die the probability  $p_E$  is given by the formula

$$p_E = m_E/m_S,$$

where  $m_E$  (the measure of  $E$ ) denotes the number of elements in  $E$ . This formula accords with observations arising from numerous experiments. In particular, when  $E = \{1,5\}$ , we have  $m_E = 2$ ,  $m_S = 6$ , so  $p_E = 2/6 = 1/3$ . In this context  $S$  is called the **sample set**. For any finite set  $S$  and any subset  $E$  of  $S$  we can similarly regard  $E$  as an event whose probability will be given by the formula  $p_E = m_E/m_S$ . The transparent probabilities of events associated with finite sets provide good illustrations for the concepts to be encountered.

### 1.3 Building new events out of given ones.

The probabilities of greatest interest to us, namely win probabilities like  $p(A \text{ bt } B)$ , are difficult if not impossible to determine from first principles. Fortunately, our main task is to derive them from known probabilities in various ways, particularly by building new events out of given ones. There are a number of standard ways in which events can be combined so as to form new events. The illustrative examples of combined events to follow are all given with reference to events determined by subsets of  $S = \{1,2,3,4,5,6\}$  via the roll of a die (see 1.1). Let  $X, Y$  be given events.

$X \cup Y$  denotes the event **at least one of  $X, Y$** . It is called the **union** of the given events. It takes place if and only if at least one of  $X, Y$  takes place.

**Example.** Let  $X = \{1,2\}$ ,  $Y = \{2,3\}$ . Then  $X \cup Y = \{1,2,3\}$ , because an outcome lies in  $X \cup Y$  precisely when it lies in at least one of the sets  $X, Y$ .

$X \cap Y$  denotes the event  **$X$  and  $Y$** . It is called the **intersection** of the given events. It takes place if and only if  $X$  takes place and  $Y$  takes place.

**Example.** Let  $X = \{1,2,3\}$ ,  $Y = \{2,3,4\}$ . Then  $X \cap Y = \{2,3\}$ .

The events  $X$  and  $Y$  are called **disjoint** if one of them precludes the other. This implies that  $X \cap Y$  cannot take place. In other words  $p(X \cap Y) = 0$ .

**Example.** Let  $X = \{1,2\}$ ,  $Y = \{3,4\}$ . Then  $X$  and  $Y$  are disjoint.

$X \oplus Y$  denotes the event **precisely one of  $X, Y$** . It is called the **disjoint union** of the given events. It takes place if and only if precisely one of  $X, Y$  takes place.

**Example.** Let  $X = \{1,2\}$ ,  $Y = \{3,4\}$ . Then  $X \oplus Y = \{1,2,3,4\}$ .

In general,  $X \oplus Y = X \cup Y$  provided that  $X$  and  $Y$  are disjoint. So not every pair of events can be combined so as to form a disjoint union.

The above constructions generalize to more than two events without complication, by just repeating the construction. For example,

$$X \cup Y \cup Z = (X \cup Y) \cup Z.$$

Since  $(X \cup Y) \cup Z = X \cup (Y \cup Z)$  there is no ambiguity. For more than three terms it is often helpful to use operator notation as follows. Given the events  $X_1, X_2, \dots, X_N$  we put

$$\cup_r X_r = X_1 \cup X_2, \cup \dots, \cup X_N$$

$$\cap_r X_r = X_1 \cap X_2, \cap \dots, \cap X_N.$$

$$\oplus_r X_r = X_1 \oplus X_2, \oplus \dots, \oplus X_N$$

$$\sum_r t_r = t_1 + t_2 + \dots + t_N$$

where  $r = 1, 2, \dots, N$  and where the  $t_r$  are real numbers. For the disjoint union to exist the events in question need to be pairwise disjoint i.e. every pair of them should be disjoint.

Corollary.  $W \cap \oplus_r X_r = \oplus_r (W \cap X_r)$

### 1.4 Sum Rule.

$$p(X_1 \oplus X_2, \oplus \dots, \oplus X_N) = pX_1 + pX_2 + \dots + pX_N.$$

**Example.** Let  $X = \{1,2\}$ ,  $Y = \{3,4\}$ ,  $Z = \{5\}$ . Then  $X \oplus Y \oplus Z = \{1,2,3,4,5\}$ . So  $p(X \oplus Y \oplus Z) = 5/6$  while  $p X + p Y + p Z = 2/6 + 2/6 + 1/6 = 5/6$ .

### 1.5 Conditional probability.

Consider the following real world problem: what is the probability of producing 2 or 3 or 4 with a roll of a die?. By putting  $E = \{2,3,4\}$ , we readily solve this problem as follows:  $pE = mE/mS=3/6$ . Let us now pose a related problem. What is the probability of producing a 2 or 3 or 4 under the assumption that the outcome is going to be an even number? This means effectively that we would ignore outcomes that are not even numbers. Or it could mean that we wish to compute retroactively what the probability would be if the additional information of an even outcome is available.

Let  $F = \{2,4,6\}$  (even numbers). So our problem becomes: what is the probability of selecting a point from  $E$  given that the point is in  $F$ . To calculate this probability we need merely put  $F$  in the role of the sample set  $S$ . That raises the issue that  $E$  is not a subset of  $F$ . However,  $E \cap F$  is a subset of  $F$  and we notice that the probability of  $E \cap F$  as event in the sample set  $F$  is the solution to the stated problem. As in 1.1 we can calculate this probability to be  $m(E \cap F)/mF = (m(E \cap F)/mS) / (mF/mS) = p(E \cap F) / pF$ .

This motivates the following general definition. If  $E$  and  $F$  are arbitrary events such that  $pF > 0$ , we define the **conditional probability of  $E$  relative to  $F$ , denoted  $p(E|F)$** , by the formula

$$(a) \quad p(E|F) = p(E \cap F)/pF.$$

The expression  $E|F$  is also known as  **$E$  given  $F$** .

### 1.6 Conditional Probability Rule.

$$pF * p(E|F) = pE * p(F|E)$$

where  $E$  and  $F$  are events such that  $pE > 0$  and  $pF > 0$  and where  $*$  denotes multiplication. To derive this rule we depart from the above equation 1.5(a) and multiply both sides by  $pF$ . We thus arrive at the expression

$$p(E \cap F) = p(E|F) * pF.$$

Since  $E$  and  $F$  are both events with positive probability, their roles could be interchanged. So we have similarly

$$p(E \cap F) = p(E|F) * pF = p(F|E) * pE.$$

### 1.7 Independent events.

It is clear from the motivating example in 1.5 that  $pE$  is in general different from  $p(E|F)$ . When  $pE$  is different from  $p(E|F)$ ,  $pE$  can be seen to be influenced by  $F$ . Indeed, in the motivating example of 1.5, where  $E = \{2,3,4\}$  and  $F = \{2,4,6\}$ , it is evident that if an outcome is known to lie in  $F$  then it is more likely to lie in  $E$  than an outcome about which nothing is known, so  $p(E|F) > pE$ . When  $pE = p(E|F)$ , and therefore (by symmetry of the Conditional Probability Rule)  $pF = p(F|E)$ , we see that  $pE$  is not influenced by  $F$  nor is  $pF$  influenced by  $E$ . This motivates the following definition applicable to events of positive probability:

$E$  and  $F$  are called **independent** if  $pE = p(E|F)$ .

In the context of example 1.1, if  $E$  and  $F$  are associated with different rolls of a die, then neither can influence the probability of the other, so they are automatically independent.

### 1.8 Product Rule.

If  $E$  and  $F$  are independent events with positive probability then

$$p(E \cap F) = pE * pF.$$

Indeed, by equation (a) of 1.5 we have  $p(E|F) = p(E \cap F)/pF$ . So it follows by independence that  $p(E \cap F) = pF * p(E|F) = pF * pE = pE * pF$ .

## 2. From ranking data to histogram.

In the Overview it was mentioned that for computational purposes the Bell Curve that represents a player's performance can be replaced by a histogram. We now describe how that histogram can be constructed from the given ranking data ( $\mu, \sigma$ ) of the player, where  $\mu$  (Greek letter mu) denotes the pregame Mean of the player's Bell Curve (i.e. the Bayesian Grade) and  $\sigma$  (sigma) the Standard Deviation.

### 2.1 Construction of the pregame histogram.

By an **N-histogram** will be meant a list of pairs  $(r_i, q_i)$  ( $i=1,2,\dots,N$ ) of real numbers such that the  $r_i$  are in ascending order, symmetric with respect to their average value, and the  $q_i$  are nonnegative numbers that sum to 1. For positive integers  $N$  up to at least 50 one can download the list of  $N$  Gauss-Hermite nodes  $g_1, g_2, \dots, g_N$  and weights  $h_1, h_2, \dots, h_N$  (just Google "Gauss-Hermite quadrature" to get a clue.) For example, when  $N = 3$  they are as follows:

<b>i</b>	<b><math>g_i</math></b>	<b><math>h_i</math></b>
<b>1</b>	-1.224744871391589	0.2954089751509193
<b>2</b>	0	1.181635900603677
<b>3</b>	1.224744871391589	0.2954089751509193

The pairs  $(g_i, h_i)$  do not form an N-histogram since the sum of the  $h_i$  does not equal 1. For a given player with ranking data  $(\mu_X, \sigma_X)$  we build an N-histogram out of them by putting

$$x_i = \mu_X + \sqrt{2}\sigma_X g_i$$
$$\xi_i = h_i / \sqrt{\pi}$$

The numbers  $x_1, x_2, \dots, x_N$  so produced are symmetric about the mean value  $\mu_X$  on the croquet ranking scale. In the case  $N = 8$  these numbers (rounded) appeared in the chart of the Overview as the values marked on the horizontal line. The numbers  $\xi_i$  (Greek letter xi) have become normalized so that they satisfy  $\xi_1 + \xi_2 + \dots + \xi_N = 1$ , so they now represent probabilities. We will refer to them as **probability weights**. The list of pairs  $(x_i, \xi_i)$  so obtained is the wanted **pregame N-histogram** of player X. In the Overview an 8-histogram was depicted. There the numbers were rounded for the sake of readability. When  $N = 3$  the normalized weights are pleasantly simple:  $\xi_1 = 1/6$ ,  $\xi_2 = 2/3$  and  $\xi_3 = 1/6$ .

### 2.2 Why histograms can replace Bell Curves.

The Bayesian postgame updating algorithm is expressed in terms of integrals that involve the representing Bell Curves of the players. In the next section the calculation of the postgame histogram from the pregame histogram is going to involve finite sums arising from the histograms. Those finite sums are nothing but approximations of the relevant integrals via the Gauss-Hermite numerical integration method. The approximations have astonishing accuracy and efficiency. To process more than 140000 games since 1985 to produce a ranking list takes about 10 seconds on my humble PC when 8 nodes are used. The accuracy (comparable to that of 50 nodes) is more than enough in view of the inherent inaccuracies that attend ranking.

## 3. From pregame to postgame histogram.

### 3.1 Classical win probability.

If player A is assumed to perform at level  $x$  (on the croquet ranking scale) and player B is assumed to play at level  $y$ , then the classical win probability of A over B is given by the formula

$$\text{cwp}(x,y) = 1/(1 + 10^{(d/500)}) \text{ where } d = y - x.$$

The factor 1/500 in the exponent arises from the specific scale used for croquet world ranking. This formula has been used for decades.

### 3.2 Calculation of Bayesian win probability.

Suppose now that player X plays against Y with pregame histograms  $(x_i, \xi_i)$  and  $(y_j, \eta_j)$  respectively ( $i, j = 1, 2, \dots, N$ ) ( $\eta$  is the Greek letter eta). The result X bt Y is here considered as an event. We need to calculate the Bayesian win probability  $p(X \text{ bt } Y)$  in terms of the probability weights  $\xi_i, \eta_j$  and the classical win probabilities  $\text{cwp}(x_i, y_j)$ .

According to the histogram model, player X will play at precisely one of the performance levels  $x_i$  and will do so with probability  $\xi_i$ . The event X plays at level  $x_i$  will briefly be written  $X@x_i$ . Thus  $p(X@x_i) = \xi_i$  ( $i=1, 2, \dots, N$ ). Similar data and notation applies to player Y. It follows that

$$X \text{ bt } Y = \bigoplus_{ij} (X@x_i \cap Y@y_j \cap X \text{ bt } Y) \quad (i, j=1, 2, \dots, N)$$

For each  $i, j$  the events on the right are assumed independent. By applying the Sum Rule and Product Rule we obtain

$$\begin{aligned} p(X \text{ bt } Y) &= \sum_{ij} p X@x_i * p Y@y_j * \text{cwp}(x_i, y_j) \\ &= \sum_{ij} \xi_i * \eta_j * \text{cwp}(x_i, y_j) \quad (i, j=1, 2, \dots, N). \end{aligned}$$

### 3.3 The updated probability weights.

Retaining the notation of 3.2, we now proceed to revise the probability weight  $\xi_i = p(X@x_i)$  in the light of the information inherent in the game result X bt Y. To this end we do the natural thing and put

$$\upsilon_k = p(X@x_k | X \text{ bt } Y) \quad (k = 1, 2, \dots, N).$$

To express  $\upsilon_k$  (Greek letter upsilon) in terms of the pregame histogram data, we apply the definition of Conditional Probability (1.5(a)) to get

$$\upsilon_k = p(X@x_k \cap X \text{ bt } Y) / p(X \text{ bt } Y)$$

Since  $X@x_k$  is disjoint from  $X@x_i$  when  $i \neq k$  we have by 3.2 that

$$\begin{aligned} X@x_k \cap X \text{ bt } Y &= X@x_k \cap \bigoplus_{ij} (X@x_i \cap Y@y_j \cap X \text{ bt } Y) \\ &= \bigoplus_j (X@x_k \cap Y@y_j \cap X \text{ bt } Y) \quad (j=1, 2, \dots, N) \end{aligned}$$

It follows as in 3.2 that

$$\upsilon_k = \xi_k * \sum_j \eta_j * \text{cwp}(x_k, y_j) / p(X \text{ bt } Y) \quad (k, j=1, 2, \dots, N).$$

This gives the postgame histogram  $(x_k, \upsilon_k)$  for player X. By putting  $\omega_k$  (omega) =  $p(Y@y_k | X \text{ bt } Y)$  ( $k = 1, 2, \dots, N$ ) ( $\omega_k$  = Greek omega) we derive similarly

$$\omega_k = \eta_k * \sum_i \xi_i * \text{cwp}(x_i, y_k) / p(X \text{ bt } Y) \quad (i, k=1, 2, \dots, N),$$

thus to obtain the postgame histogram  $(y_k, \omega_k)$  for the player Y .  
These formulas are variants of the well known Bayes' Rule.

#### **4. From postgame histogram to updated ranking data.**

The Mean  $\mu^{\wedge}$  and Standard Deviation  $\sigma^{\wedge}$  of the postgame histogram  $(x_k, \nu_k)$  ( $k=1,2,\dots,N$ ) are respectively given by the well known formulas

$$\mu^{\wedge} = \sum_k x_k * \nu_k \text{ and } \sigma^{\wedge} = \sqrt{(\sum_k (x_k - \mu^{\wedge})^2 * \nu_k)}.$$

This gives the updated ranking data  $(\mu^{\wedge}, \sigma^{\wedge})$  for player X. The corresponding data for Y is similarly obtained. This concludes the updating procedure.

December 2008

Author: Louis Nel  
All rights reserved © 2008

---

Updated 4.xii.08

[www.oxfordcroquet.com/tech/nel-byus/index.asp](http://www.oxfordcroquet.com/tech/nel-byus/index.asp)  
on [www.oxfordcroquet.com](http://www.oxfordcroquet.com)

---