Scientific
Research
Publishing

Foundations for Wash Sales

Phillip G. Bradford

Department of Computer Science and Engineering, University of Connecticut, Stamford, CT, USA
Email: phillip.bradford@uconn.edu, phillip.g.bradford@gmail.com

Abstract

Consider an ephemeral sale-and-repurchase of a security resulting in the same position before the sale and after the repurchase. A sale-and-repurchase is a wash sale if these transactions result in a loss within ±30 calendar days. Since a portfolio is essentially the same after a wash sale, any tax advantage from such a loss is not allowed. That is, after a wash sale a portfolio is unchanged so any loss captured by the wash sale is deemed to be solely for tax advantage and not investment purposes. This paper starts by exploring variations of the birthday problem to model wash sales. The birthday problem is: Determine the number of independent and identically distributed random variables required so there is a probability of at least 1/2 that two or more of these random variables share the same outcome. This paper gives necessary conditions for wash sales based on variations on the birthday problem. Suitable variations of the birthday problem are new to this paper. This allows us to answer questions such as: What is the likelihood of a wash sale in an unmanaged portfolio where purchases and sales are independent, uniform, and random? Portfolios containing options may lead to wash sales resembling these characteristics. This paper ends by exploring the Littlewood-Offord problem as it relates capital gains and losses with wash sales.

Keywords

Wash Sales, Tax, Birthday Problem, Littlewood-Offord, Probability, Finance

1. Introduction

Wash sales occur when a security is sold and quickly bought back with the sole intent to capture a tax loss from the sale. Wash sales impact a portfolio's tax liabilities. Determining the likelihood of wash sales is also important for understanding investment strategies and for comparing actively and passively managed portfolios. Wash sales apply to investors, but not to market makers.

Taxes play a significant role in economics and finance. Taxes influence behavior, shape the engineering of financial transactions, and sometimes have unintended consequences. Therefore, thoughtful analysis is imperative for taxes. This paper adds firm mathematical foundations to aid the understanding of wash sale taxes.

The main goal of this paper is: To provide foundations for certain wash sales-in cases when they may occur as well as the capital gain implications. This may also help differentiate managed funds and unmanaged index funds in terms of wash sales.

Wash sales are sometimes created by the exercise of options, thus a portfolio manager may not be able to avoid a wash sale in some contexts. For example, suppose an in-the-money American-style put option is written in a portfolio. Provided this option remains in-the-money, it may be exercised by its holder[1] at anytime up to its expiry. If the exercise of this put option replaces shares sold at a loss in the prior 30 days, then this is a wash sale. This option's exercise is beyond control of the portfolio manager.

The foundations given here start with variations of the classical birthday problem from probability theory [1]-[3]. This work has implications on wash sales. Also, the Littlewood-Offord problem [4]-[6] is applied to understand capital gains for certain wash sales. The Littlewood-Offord problem is viewed from the perspective of the probabilistic method.

For convenience, let $[n] \equiv \{1, 2, \cdots, n\}$.

1.1. Wash Sales in Detail

Suppose a security is sold at a loss on day d_2. This sale is a wash sale if substantially the same security is purchased within ± 30 calendar days from d_2, see for example [7].

Definition 1 (US wash sale [7]) Consider three dates $d_1, d_2, d_3 : d_1 \leq d_2$ and $d_1 \leq d_3$ where $|d_3 - d_2| \leq 30$ calendar days. Suppose s shares of a security are purchased on date d_1 at price p_1. At some later date d_2, s shares are sold for price $p_2 < p_1$. Thus, the s shares are sold at a loss. Then within ± 30 days on date d_3, s shares are repurchased for price p_3. This is a wash sale and since $|d_3 - d_2| \leq 30$ days, then the next adjustments must be made [7]:

1) The loss $p_1 - p_2$ is not permissible for taxes. That is, this loss may not be subtracted from profits or gains and it may not be used to get a lower tax rate.

2) The cost-basis of the shares repurchased on d_3 is set to $p_3 + (p_1 - p_2)$. The shares purchased on d_3 have the start of their holding period reset to d_1.

Short positions may also be wash sales. For example, consider holding a short position of 100 shares of a security starting on date d_1 in a portfolio Π. Then suppose this short position is closed at a loss by purchasing 100 shares on day d_2. Once this position is closed on day d_2, then Π contains no shares of this security. Next re-short another 100 shares of substantially the same security on d_3 where

---

[1]Options, like shares of stock, are fungible and there are specific option exercise assignment allocation methods used to allocate exercised options [8].

$|d_3 - d_2| \leq 30$ days. These transactions leave the portfolio the same while getting a tax advantage for the loss. This tax advantage is also disallowed by the wash sale rules.

Consider a wash sale as described by Definition 1, where $(p_1 - p_2) + p_3 > p_1$ or in other words $p_3 > p_2$. Suppose the shares are sold at price $p_4 > (p_1 - p_2) + p_3 > p_1$ at the later date $d_4 \geq d_3$. In the case with the wash sale, there is a capital gain of $p_4 - [(p_1 - p_2) + p_3]$ which is smaller than the capital gain $p_4 - p_1$ if the wash sale had not occurred. Capital gains are taxable. A capital gain $p_4 - p_1$ is from the single purchase of the shares for price p_1 on d_1 and the single sale of the shares on date d_4 for price p_4, thus skipping the sale at a loss and repurchase.

This means such a wash sale gives $p_4 - p_1 - [p_4 - [(p_1 - p_2) + p_3]]$ or $p_3 - p_2$ less taxable income than a single purchase of the security at price p_1 on date d_1 and a single sale for price p_4 on d_4. Of course, a wash sale's loss is not allowed.

Wash sales may be avoided by restricting each security in a portfolio to be either purchased or sold only every 31 calendar days. This restriction may not be suitable for many portfolios. In a portfolio containing options, it may be impossible to maintain this restriction.

It has also been suggested, e.g. [9], wash sales may be avoided by purchasing or selling (moderately) correlated, but not substantially the same, securities. That is, if a security is sold at a loss then purchase a different but correlated security within 30 days maintaining some of a portfolio's characteristics while keeping the tax advantage.

Historically many securities are assumed to only trade on about $n = 252$ business days per year [10]. Although reflecting on global markets one may assume there are $n = 365$ trading days.

1.2. Background

There has not been much research on wash sales, e.g., [9]. There is important work on taxation and its investment implications. Take, for example, [11]-[13].

The birthday problem is classical.

Definition 2 (Birthday-Collision) Given two random variables X_1, X_2 mapping respectively to x_1, x_2 in the same range $[n]$, then a *birthday-collision* is when $x_1 = x_2$.

To model random wash sales, this paper assumes independent identically distributed random variables. A common statement of the birthday problem is:

Definition 3 (Birthday Problem) Consider n days in a year and k independent identically distributed (*iid*) uniform random variables whose range is $[n]$ and $n \geq k$. What is the probability $B(n, k)$ of at least one birthday-collision among these k random variables?

According to a blog post by Pat B [14] the birthday problem may have originally been given by Harold Davenport as cited in [15] and later published by [1]. In any case, von Mises gave the first published version to the best of our knowledge.

Bounds of day counts for the birthday problems include [16] who gives bounds for birthdays of distance d for both linear years as well as cyclic years. In a cyclic year, 1-January is a single day from 31-December of the same year. Bounds for birthdays of

distance d for cyclic years are given by [17].

The birthday problem applied to boys and girls (random variables with different labels) are discussed in [18] as well as [19]. That is, how many birthdays are shared by one or more boys and one or more girls? A comprehensive view is provided by [20] including stopping problems with the boy-girl birthday problem. Non-uniform bounds for online boy-girl birthday problems are given by [21] and [22].

Tight bounded Poisson approximations for birthday problems are given by [23]. Poisson approximations to the binomial distribution for the boy-girl birthday problem is given by [19]. A Stein-Chen Poisson approximation is used by [24] to solve variations of the standard birthday problem. Matching and birthday problems are given by [25]. Incidence variables are used to study birthday problems with Pareto-type distributions in [26].

Applications of the birthday problem include: computer security [20] [21] [26] [27], public health and epidemiology [28], psychology, DNA sequence alignment, experiments, and games [29] [30]. Summaries of work on the birthday problem are in [29]-[31].

Results on the expectation for getting j different letter k-collisions are given by [32]. Their results are expressed as truncated exponentials or gamma functions.

The Littlewood-Offord problem hails from complex analysis [6]. Erdös [5] improved Littlewood and Offord's result by an elegant application of the probabilistic method. These and related results determine the concentration of sums of random variables multiplied by integers. The Littlewood-Offord problem is applied to certain capital gains.

1.3. Structure of This Paper

Section 2 reviews variants the birthday problem applied here. First the classical birthday problem is discussed. Next this section progresses through the $\pm d$ birthday problem. After the definition and key results are given about the $\pm d$ birthday problem, the boy-girl birthday problem is explored. Finally, the $\pm d$ boy-girl birthday problem is defined and several bounds are derived as they relate to a necessary condition for wash sales.

Subsection 2.1 gives an example of wash sales based on boy-girl birthday collisions of a single day.

Section 3 generalizes results of the previous sections. In particular, it shows how to compute $B_d(n, b, g)$, the number of b boys and g girls that give a probability of 1/2 or more where a boy and a girl have birthdays within d days of each other over n days.

Subsection 3.1 gives an example of wash sales based on boy-girl birthday collisions over a range of $\pm d = 30$ days.

Finally, Section 4 explores how wash sales impact capital gains and losses. Since wash sales are capital losses, they may offset capital gains. Several results, including the Littlewood-Offord problem, are applied to capital gains and losses as they may be impacted by wash sales.

2. The Birthday Problem and Wash Sales

The birthday problem is often applied to finding the probability of coincidences. So there is a rich literature on variations of the birthday problem [29] [30]. Asset sales are often viewed as carefully selected. However, portfolios using American-style options may exhibit asset sales or purchases beyond the control of the portfolio managers.

A key question is: Over n consecutive days for what integer k does $\arg\min_k \left\{ B(n,k) \geq \frac{1}{2} \right\}$ hold for k iid uniform random variables? In other words, given n days, what is the least k iid uniform random variables so that $B(n,k) = 1/2$?

Solutions to this basic variation of the birthday problem are well known. The probability $B(n,k)$ is the compliment of the probability of k iid uniform random variables having no birthday-collisions. Therefore, if there are no birthday-collisions, then k birthdays can be in $\binom{n}{k} k!$ permutations out of all possible n^k mappings of the k random variables onto $[n]$. In other words, the $\binom{n}{k}$ subsets of k distinct elements of $[n]$ is the exact number of subsets the k variables may map to without a collision. These k variables may be ordered in $k!$ permutations. That is,

$$B(n,k) = 1 - \binom{n}{k}\frac{k!}{n^k} = 1 - \frac{n!}{(n-k)!} \cdot \frac{1}{n^k}, \tag{1}$$

for $n \geq k$ and $B(n,k) = 1$ otherwise.

Starting with n and a probability $p = B(n,k)$, then computing k is often done using the inequality $1 - x \leq e^{-x}$. In particular, the smallest k giving a probability of 1/2 that there is at least one birthday-collision requires k to be roughly $\sqrt{2(\ln 2)n}$ or about $1.18\sqrt{n}$. See for example, [1] [33] [34].

Another classical approach is to look at the random variable X as the sum of all birthday-collisions of k people over n days, see for example [19] [25] [35] [36]. A concise exposition is given in [36] which we follow. Presume the birthday of person $i \in [k]$ is given by the random variable $Y_i \in [n]$. Since a potential birthday collision is a Bernoulli trial, so X is binomially distributed. Thus, $X \in \left\{ 0, 1, 2, \cdots, \binom{k}{2} \right\}$ where $\binom{k}{2}$ is the maximum number of potential birthday-collisions. The expectation of the maximum number of birthday collisions possible is $\binom{k}{2}$ with probability $\frac{1}{n} = \mathbb{P}\left[Y_i = t \mid Y_j = t \right], t \in [n]$ where $\{i,j\} \subseteq [k]$. The expected maximum number of birthday-collisions is $\frac{1}{n}\binom{k}{2}$. If n is sufficiently larger than k, then X is approximately Poisson where $\lambda = \frac{1}{n}\binom{k}{2}$. Thus, $\mathbb{P}[X \geq 1] \approx 1 - e^{-\binom{k}{2}/n}$.

In the case of the $\pm d$ birthday problem, if two random variables X_1, X_2 map within d days of each other, then this is a $\pm d$ birthday-collision [16].

Two birthdays x_1 and x_2 of distance $|x_1 - x_2|$ demark a span of size $1 + |x_1 - x_2|$. For example, $|4\_July - 3\_July| = 1$, so these dates are in a $\pm d = \pm 2$ span, but not in a span of $\pm d = \pm 1$.

The next definition is based on [16] [17] [29].

Definition 4 ($\pm d$ Birthday Collisions) Consider n days in a year, spans of less than $\pm d$ days, and k iid uniform random variables with range $[n]$: Then $B_d(n,k)$ is the probability at least two such random variables have a $\pm d$ *birthday-collision*. That is, these two random variables have ranges in *less than d* days of each other.

In n days with a $\pm d$ span, then $\arg\min_k \left\{ B_d(n,k) \geq \dfrac{1}{2} \right\}$ gives the smallest k so there is a probability of at least 1/2 where at least two such random variables are fewer than d days from each other.

Definition 5 (Blocks of days) Let $i : k > i > 1$. Suppose birthdays are ordered as $x_1 \leq x_2 \leq \cdots \leq x_k$, then for a birthday x_i its *nearest birthday pairs* are (x_{i-1}, x_i) and (x_i, x_{i+1}). There are no birthdays between x_{i-1} and x_i and there are no birthdays between x_i and x_{i+1}.

A *block* of days contains a single birthday on one of its end-points. The birthday x_i is associated with two blocks: $(x_{i-1}, x_i]$ and $[x_i, x_{i+1})$.

The days between x_1 and x_2 form a block of size $|x_1 - x_2|$ since there are no birthdays between x_1 and x_2. Thus, two nearest birthday pairs contained in a span of $\pm d$ are separated by a block of size $d - 1$.

Take k iid uniform random variables and consider $\pm d$ birthday-collisions over $[n]$ days. Naus [16] gives the next idea: If there are no $\pm d$ birthday-collisions, then there must be at least size $d - 1$ blocks of no birthdays between each nearest birthday pair. This gives a total of $(k-1)(d-1)$ days with no birthdays in $k - 1$ contiguous blocks of at least $d - 1$ days each. Therefore, if there are no $\pm d$ birthday-collisions, then k birthdays can be in $\dbinom{n - (k-1)(d-1)}{k} k!$ permutations out of all possible n^k mappings of the k random variables. Thus, to get the probability of at least one $\pm d$ birthday collision, take the compliment of the probability of having no $\pm d$ birthday-collisions. The next result follows.

Theorem 1 ([16]).

$$B_d(n,k) = 1 - \binom{n - (k-1)(d-1)}{k} \frac{k!}{n^k} = 1 - \frac{(n - (k-1)(d-1))!}{(n - (k-1)(d-1) - k)!} \cdot \frac{1}{n^k}, \qquad (2)$$

for $n \geq (k-1)(d-1) + k$ and $B_d(n,k) = 1$ otherwise.

Using the bound $1 - x \leq e^{-x}$ on Naus' result gives k of about $0.83\sqrt{\dfrac{n}{d-4}}$, see [16].

Also [29] approximate k to about $1.2\sqrt{\dfrac{n}{2d+1}}$ for the cyclic version.

Note, Theorem 1 with $d = 1$ gives the solution to the standard birthday problem of Definition 3. That is, a span of $d = 1$ and blocks of size $d - 1 = 0$.

The falling factorial is

$$m^{\underline{k}} = m(m-1)\cdots(m-k+1) = \binom{m}{k} k! \tag{3}$$

In these terms, Theorem 1 may be expressed as $B_d(n,k) = 1 - \dfrac{(n - (k-1)(d-1))^{\underline{k}}}{n^k}$.

The next classic result is important.

Lemma 1 (Classical) Let $m \geq k \geq 1$. The falling factorial $m^{\underline{k}}$ is the number of injective mappings of $k \geq 1$ elements to the range $[m]$.

The next definition is based on [18] [20] [23].

Definition 6 (Boy-Girl Birthdays) Consider n days in a year and two sets of distinctly labeled iid uniform random variables all with range $[n]$: g of these variables are girls and b of these variables are boys. Then $B(n, b, g)$ is the probability at least one girl and one boy have a birthday-collision.

For instance, in n days, $\underset{\substack{k=b+g \\ b=g}}{\arg\min}\left\{ B(n, b, g) \geq \dfrac{1}{2} \right\}$ gives the value $k = b + g$ and

$b = g$ so there is a probability of 1/2 where at least one girl and one boy have the same birthday.

Stirling numbers of the second kind [37] count the number of non-empty partitions of a given set. For example given the set $[m]$, the number of partitions of $[m]$ into i non-empty subsets is $\left\{ {m \atop i} \right\}$.

Due to their nature, it is common to define Stirling numbers of the second kind recursively [37]: $\left\{ {m \atop i} \right\} = i \left\{ {m-1 \atop i} \right\} + \left\{ {m-1 \atop i-1} \right\}$ with the base cases $\left\{ {m \atop 1} \right\} = 1$ and $\left\{ {m \atop m} \right\} = 1$.

Finally, $\left\{ {m \atop m+i} \right\} = 0$ for any $i > 0$. As an example,

$$\left\{ {3 \atop 2} \right\} = \left| \{ \{\{1,2\},\{3\}\}, \{\{1,3\},\{2\}\}, \{\{1\},\{2,3\}\} \} \right| = 3. \tag{4}$$

The next classical equality counts the number of functions from $[n]$ elements to $[m]$ elements, $m \geq n$,

$$m^n = \sum_{i=1}^{n} \left\{ {n \atop i} \right\} m^{\underline{i}} \tag{5}$$

expressed as the number of non-empty i partitions of the $[n]$ elements and the number of surjections from the i partitions by Lemma 1.

Theorem 2 ([18] [20]) Consider n days in a year and two sets of distinctly labeled iid uniform random variables all with range $[n]$: g random variables are girls and b random variables are boys. Then $B(n, b, g)$ is the probability at least one girl and at least one boy have a birthday-collision and

$$B(n, b, g) = 1 - \frac{1}{n^{b+g}} \sum_{i=1}^{g} (n-i)^b \begin{Bmatrix} g \\ i \end{Bmatrix} n^{\underline{i}}. \tag{6}$$

The next Lemma is from [18] [38].

Lemma 2 ([18] [38]) Consider n days in a year and two sets of distinctly labeled iid random variables all with range $[n]$: g random variables are girls and b random variables are boys. Then $B(n, b, g)$ is the probability that at least one girl and at least one boy have a birthday-collision and

$$B(n, b, g) = 1 - \frac{1}{n^{b+g}} \sum_{i=1}^{g} \sum_{j=1}^{b} \begin{Bmatrix} b \\ j \end{Bmatrix} \begin{Bmatrix} g \\ i \end{Bmatrix} n^{\underline{i+j}}. \tag{7}$$

Wash Sale Example 1: Same Day Purchase and Sale

Consider a portfolio $\Pi = \{a_1, \cdots, a_k\}$ where $a_i : k \geq i \geq 1$ is asset (security) i held in Π. At the end of business on day ℓ, consider portfolio $\Pi_\ell = \{a_{1,\ell}, \cdots, a_{k,\ell}\}$ the market value of asset i in Π_ℓ is $|a_{i,\ell}|$ and the total value of Π_ℓ is $|\Pi_\ell| = \sum_{i=1}^{k} |a_{i,\ell}|$. Just before the start of each tax year, asset i has market value $|a_{i,0}|$ and Π has total market value $|\Pi_0|$. Assume each asset is sufficiently liquid so our purchases or sales do not impact its market price.

Suppose portfolio Π has T total iid uniform and random transactions during the business days of one calendar year. Assume trades are distributed on an asset-weighted basis from the initial weight of each asset in the portfolio just before the trading year commences. Thus, just prior to the first trading day and with no other information, asset a_i is expected to have $t(i) = T \frac{|a_{i,0}|}{|\Pi_0|}$ trades in one year.

Take $t(i)$ transactions and define the independent Rademacher[2] random variables $\eta_1, \cdots, \eta_{t(i)}$ representing buys or sells of portions of asset class i in portfolio Π:

$$\eta_j = \begin{cases} +1 & \text{if transaction } j \text{ is a buy of asset } i \\ -1 & \text{if transaction } j \text{ is a sell of asset } i \end{cases} \tag{8}$$

for $j : t(i) \geq j \geq 1$. That is, the b independent Rademacher random variables where $\eta_j = +1$ represent buys (boys) and the g random variables where $\eta_j = -1$ represent sells (gals).

To apply a suitable version of Chernoff's bound ([39], Appendix A) where $\mathbb{P}[\eta_j = +1] = \mathbb{P}[\eta_j = -1] = \frac{1}{2}$, then for any $c > 0$

$$\mathbb{P}\left[\left(\eta_1 + \cdots + \eta_{t(i)} \right) > c \right] < e^{-c^2 / (2t(i))}. \tag{9}$$

So, for example, take $c = 1$, then $|b - g| \leq 1$ holds with high probability as $t(i)$ gets large. Of course, as $t(i)$ gets large, the likelihood of wash sales increases. That is, the total number of buys and sells is expected to converge to be about the same as the total number of transactions grows. However, along the way, the number of buys or sells may not be as balanced [2] [40].

---

[2]We used Bernoulli random variables for $\{0,1\}$ outcomes and we use Rademacher for $\{-1,+1\}$ outcomes.

Select the probabilities that the number of buys and sales are the same, given $t(i)$ total trades, in asset class a_i are:

$t(i)$	10	20	30	40	50
$e^{-1/(2t(i))}$	0.951	0.975	0.983	0.987	0.990

Let h be half the total trades $t(i)$. That is, $h \leftarrow t(i)/2$. Assuming $n \in \{252, 365\}$ trading days gives the probabilities of same-day girl-boy birthday collisions for a single asset-type as:

h	1	5	10	15	20	25	30	35
$B(252, h, h)$	0.0040	0.0946	0.3280	0.5909	0.7957	0.9162	0.9717	0.9921
$B(365, h, h)$	0.0027	0.0663	0.2399	0.4605	0.6660	0.8196	0.9150	0.9650

In fact, $B(252,13,13) = 0.4891$ and $B(252,14,14) = 0.5410$. So, considering only equal numbers of sales and buys over $n = 252$ days of the same asset type, 14 girls and 14 boys is the first case where there is greater than a 50% chance of a (same-day) boy-girl birthday collision.

Assuming the portfolio Π already holds this single asset type, a boy-girl collision only is a necessary condition for a wash sale. A birthday collision must be accompanied by a sale at a loss and a repurchase of substantially the same security within 30 calendar days.

3. General Wash Sales

Necessary conditions are given here for wash sales where a purchase and sale are within $\pm d$ calendar days. Since the purchase and sale are not known to be at a loss while keeping substantially the same portfolio before and after the $\pm d$ birthday collision.

Definition 7 (Boy-Girl $\pm d$ Birthdays) Consider n days in a year, spans of $\pm d$ days, and two sets of distinctly labeled iid uniform random variables all with range $[n]$: g random variables are girls and b random variables are boys. Then $B_d(n, g, b)$ is the probability at least one girl and one boy are mapped to less than d days of each other.

For example, starting with n, d and $k = g + b$ and $g = b$, then

$$\arg\min_k \left\{ B_d\left(n, \frac{k}{2}, \frac{k}{2}\right) \geq \frac{1}{2} \right\}$$ gives k so there is a probability of $\geq \frac{1}{2}$ so at least one girl and one boy have $\pm d$-birthday collisions.

The next result is based on [16] [18] [20] [38].

Theorem 3. Consider n days in a year, a span of $\pm d$ days, and two sets of distinctly labeled iid uniform random variables all with range $[n]$: g random variables are girls and b random variables are boys. Then $B_d(n, g, b)$ is the probability at least one girl and one boy have a $\pm d$ birthday-collision and:

$$B_d(n, g, b) = 1 - \frac{1}{n^{b+g}} \sum_{i=1}^{b} \sum_{j=1}^{g} \begin{Bmatrix} b \\ i \end{Bmatrix} \begin{Bmatrix} g \\ j \end{Bmatrix} (n - (i + j - 1)(d - 1))^{\underline{i+j}}. \tag{10}$$

Proof. This proof calculates the probability of not having no boy-girl $\pm d$ birthday collisions. That is, one minus the probability of no boy-girl $\pm d$ birthday collisions. This gives the probability of at least one boy-girl $\pm d$ birthday collision.

Given n days, a $\pm d$ span, and iid uniform random variables separated into g (girls) random variables and b (boys) random variables. Then the total number unconstrained mappings of the b and g variables to $[n]$ is n^{b+g} giving the denominator in front of the double sum.

The value $B_d(n,g,b)$ is not impacted if either any number of boys have the same birthday or separately any number of girls have the same birthday. Rather $B_d(n,g,b)$ is impacted by boy-girl collisions. Therefore, consider partitions of b boys and g girls. To prevent the girls' partitions and boys' partitions from colliding into $\pm d$ spans of the same range, count the number of places these i and j non-empty partitions may be mapped so there is no $\pm d > 1$ birthday-collision. By Lemma 1 there are

$$\left(n-(i+j-1)(d-1)\right)^{\underline{i+j}} = \binom{n-(i+j-1)(d-1)}{i+j}(i+j)! \tag{11}$$

injective functions to $[n]$ for sets of $i \in [b]$ boys and sets of $j \in [g]$ girls with $(i+j-1)$ blocks of $(d-1)$ contiguous days with no boy or girl in them.

Now, consider placing the i and j partitions in separate locations among the $\left(n-(i+j-1)(d-1)\right)^{\underline{i+j}}$ function mappings to $[n]$, see Naus [16]. That is, the i partitions of $[b]$ where each partition is in a different location and j partitions of $[g]$ where each partition is also in a different location by Equation (11). That is, given $i \in [b]$ and $j \in [g]$, then the product $\left\{\begin{matrix} b \\ i \end{matrix}\right\}\left\{\begin{matrix} g \\ j \end{matrix}\right\}$ is the total number of injective

mappings of boys to i non-empty partitions and independently the number of injective mappings of girls to j non-empty partitions.

This completes the proof.

Wash Sale Example 2: *d* = ±30 Calendar Days

Start with the same setup as the previous wash sale example from subsection 2.1.

Let h be half the total trades $t(i)$ in day i. That is, $h \leftarrow t(i)/2$. Assuming $n \in \{252, 365\}$ trading days and $d = \pm 30$ calendar days gives the probabilities of girl-boy ± 30-day birthday-collisions for a single asset type is:

h	1	2	3	4
$B_{30}(252,h,h)$	0.220	0.819	0.994	0.99998
$B_{30}(365,h,h)$	0.155	0.667	0.953	0.99840

Consider only a single asset type. The intuition behind these probabilities is straightforward. For instance, consider $n = 365$ days and to avoid boy-girl collisions each girl and boy must be separated by at least 30 days before and after their birthday from the other gender. So the 365 days may be broken into about six blocks of about 60 days.

4. Wash Sale and Integral Capital Gains and Losses

Capital gains or capital losses may be rounded to the nearest integer for US tax calculations. Provided all trades are rounded. Rounding drops the cents portion for gains whose cents portion is 50-cents or below. Rounding adds a dollar to the dollar portion of gains whose cents portion is greater than 50 cents while dropping the cents portion. Losses work the same way. Gains and losses must all be rounded or none must be rounded. So, from here on, let all gains or losses be integers.

Long term capital gains and losses are aggregated and at the same time short term capital gains and losses are aggregated. At the end of the tax year the long term and short term aggregates are added together to get the final capital gain or loss for taxation.

The focus here is capital gains or losses for capital assets that may have wash sales. Wash sales are losses, but losses may offset gains. The study of options and their associated premiums is classical [10] and we do not address it here. So, option premiums are ignored.

In a portfolio, individual capital gain values and individual capital loss values are usually distinct. Though rare, identical capital gains and capital losses are possible. Identical capital gains or losses are possible for portfolios built using options. We are ignoring option premiums. That is, asset purchases may be done via the exercise of cash-covered American-style put options. Also asset sales may be done via the exercise of American-style covered-call options. In these cases with options that become in-the-money, a portfolio manager has no control of the asset sales or purchases or timing of such trades. See Figure 1.

Most often, put or call option strike prices are at discrete increments. For example, many put and call equity options have strike prices in $5 or $10 increments. Suppose a portfolio is built only using the exercise of American-style options. Many asset gains and losses may be for identical amounts. Of course, this depends on the size of the underlying positions or the number of options written. Options with the same expiry on identically sized underlying assets may have very different values [10].

date d_1	date d_2	date d_3	date d_4	date d_5
Sell cash-covered put p_1 at strike price $100 on d_2	Put p_1 exercised to "purchase" the asset for $100	Covered-call c_1 sold for strike price $110 and expiry d_4	Call c_1 exercised triggering a sale giving a capital gain of $10	No wash sale
Sell cash-covered put p_2 at strike price $100 on d_2	Put p_2 exercised to "purchase" the asset for $100	Covered-call c_2 sold for strike price $90 and expiry on d_4	Call c_2 exercised triggering a sale giving a capital loss of $10	Sell cash-covered put with strike price $90 and if exercised within 30 days, then it is a wash sale

Figure 1. A potential wash sale with American-style options. Each row represents the same underlying asset type.

In such option-based portfolios assume uniform, independent, and random capital gains and capital losses. This may be modeled by the Littlewood-Offord Problem.

Definition 8 is classical and extensive discussion may be found in the likes of [6] [41].

It is based directly on [4] [6] [41].

Definition 8 (Littlewood-Offord Problem) The integer Littlewood and Offord's problem is given an integer multi-set $V = \{v_1, v_2, \cdots, v_n\}$ where $v_i \geq 1, \forall i \in [n]$ and $S_v = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_n v_n$ so each ξ_i is such that $\mathbb{P}[\xi_i = -1] = \mathbb{P}[\xi_i = +1] = \frac{1}{2}$, for $i \in [n]$, then what is $\max_{x \in \mathbb{Z}} \mathbb{P}[S_v = x]$?

Assuming equal probability of gains and losses and no drift [10]. Given an integer multi-set $V = \{v_1, v_2, \cdots, v_n\}$ so $v_i \geq 1, \forall i \in [n]$. The multi-set V represents capital gains and capital losses. Capital gains and capital losses are all from sales. The iid Rademacher random variables $\xi_i \in \{+1, -1\}$ determine if a v_i is a capital gain or loss. All v_i are positive since all the Rademacher variables have range $\{-1, +1\}$, see also [5] and [40].

Over a tax year, the total capital gain or loss is

$$S_v = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_n v_n. \tag{12}$$

In an optimal solution of this version of the Littlewood-Offord problem, [5] showed the n-element multi-set $V = \{1, 1, \cdots, 1\}$ has $\max_{x \in \mathbb{Z}} \{\mathbb{P}[S_v = x]\} = O\left(\frac{1}{\sqrt{n}}\right)$.

The next lemma's proof follows immediately from the linearity of expectation given Rademacher random variables. See, for example, [39].

Lemma 3. Consider any integer multi-set $V = \{v_1, v_2, \cdots, v_n\}$ where $v_i \geq 1, \forall i \in [n]$ and the random variable $S_v = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_n v_n$, where $\mathbb{P}[\xi_i = -1] = \mathbb{P}[\xi_i = +1] = \frac{1}{2}$, for all $i \in [n]$, then $\mathbb{E}[S_v] = 0$.

For any Rademacher random variable ξ_i, it must be $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i^2] = 1$. Since v_i is constant $\sigma_{\xi_i v_i}^2 = \mathbb{E}[\xi_i^2 v_i^2] - \mathbb{E}[\xi_i v_i]^2 = v_i^2$. Thus, a proof of the next theorem follows since the variance of a sum of independent random variables is the sum of the variances.

Theorem 4. Consider any non-negative integer vector v and the random variable $S_v = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_n v_n$, where $\mathbb{P}[\xi_i = -1] = \mathbb{P}[\xi_i = +1] = \frac{1}{2}$, for all $i \in [n]$, then $\mathbb{E}[S_v^2] = v_1^2 + v_2^2 + \cdots + v_n^2$ and $\sigma_{S_v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.

Thus, the lowest variance, $\sigma_{S_v}^2$, for the integer Littlewood-Offord problem occurs exactly when $V = \{1, 1, \cdots, 1\}$ and $|V| = n$. Assuming the $\xi_i, \forall i \in [n]$ are all Rademacher random variables, then $\mathbb{P}_{x \in \mathbb{Z}}[S_v = x]$ is maximized [6] [40] [41] as $O(1/\sqrt{n})$ and $\sigma_{S_v} = \sqrt{n}$.

Theorem 4 implies the next corollary.

Corollary 1. Assume $1 = v_1 = v_2 = \cdots = v_n$ and $S_v = \xi_1 v_1 + \xi_2 v_2 + \cdots + \xi_n v_n$ where $\mathbb{P}[\xi_i = -1] = \mathbb{P}[\xi_i = +1] = \frac{1}{2}$, for all $i \in [n]$, then the standard deviation of S_v is $\sigma_{S_v} = \sqrt{n}$.

Corollary 1 highlights an exceptional case where all capital gains and capital losses are the same. Wash sales require the loss and gain to be from essentially the same security.

The generality of Theorem 4 asserts large variances too. Consider the set $V = \{2^0, 2^1, \cdots, 2^{n-1}\}$, then by Theorem 4, $\sigma_{S_v}^2 = \sum_{i=0}^{n-1} 2^{2i} = \frac{2^{2n}-1}{3}$. This last equality follows since the sum is a geometric series.

Definition 9 (Distinct sums of a set or multi-set V) Consider a set or multi-set $V = \{v_1, v_2, \cdots, v_n\}$ and let each element of the lists $H_1 = \langle \hat{\xi}_{1,1}, \hat{\xi}_{2,1}, \cdots, \hat{\xi}_{n,1} \rangle$ and $H_2 = \langle \hat{\xi}_{1,2}, \hat{\xi}_{2,2}, \cdots, \hat{\xi}_{n,2} \rangle$ be fixed values from $\{-1, +1\}$. The two sums of V,

$$s_{v,1} = \hat{\xi}_{1,1} v_1 + \hat{\xi}_{2,1} v_2 + \cdots + \hat{\xi}_{n,1} v_n \tag{13}$$

$$s_{v,2} = \hat{\xi}_{1,2} v_1 + \hat{\xi}_{2,2} v_2 + \cdots + \hat{\xi}_{n,2} v_n, \tag{14}$$

are *distinct* iff there is some $\hat{\xi}_{i,1} \neq \hat{\xi}_{i,2}$, for $i \in [n]$.

Given any multi-set of positive integers $V = \{v_1, v_2, \cdots, v_n\}$, enumerate all 2^n distinct sums as $s_v[1] \geq s_v[2] \geq \cdots \geq s_v[2^n]$, for example, see Figure 2. Given any set of positive integers $V = \{v_1, v_2, \cdots, v_n\}$, where none of the 2^n distinct sums add to the same value gives $s_v[1] > s_v[2] > \cdots > s_v[2^n]$.

An important observation by [5], is that for any fixed sum s the values $s + v_i$ and $s - v_i$ differ by $2v_i$. Next, this observation is used to show the set $V = \{2^0, 2^1, \cdots, 2^{n-1}\}$ has no distinct sums that add to the same value.

In particular, take any distinct sums $s_{v,1}$ and $s_{v,2}$ with associated fixed values $\hat{\xi}_{i,1} \in \{-1, +1\}$ and $\hat{\xi}_{i,2} \in \{-1, +1\}$, respectively, for all $i \in [n]$. Suppose, for the sake of a contradiction, that $s_{v,1} = s_{v,2}$. Building on Erdös' observation, the values $s_{v,1}$ and $s_{v,2}$ may be written as $s_{v,1} = 2^n - 1 - 2m_1$ where $m_1 = \sum_{i \in I_1} 2^{i-1}$ and $I_1 = \{i : \hat{\xi}_{i,1} = -1\}$ and likewise $s_{v,2} = 2^n - 1 - 2m_2$ where $m_2 = \sum_{i \in I_2} 2^{i-1}$ and $I_2 = \{i : \hat{\xi}_{i,2} = -1\}$, for all $i \in [n]$. Finally, the uniqueness of binary-number representations means $m_1 = m_2$ which in turn means $\hat{\xi}_{i,1} = \hat{\xi}_{i,2}$, for all $i \in [n]$. So, in fact, the sums $s_{v,1}$ and $s_{v,2}$ are equal, giving a contradiction.

Thus, the set $V = \{2^0, 2^1, \cdots, 2^{n-1}\}$ satisfies the antecedent of the next theorem.

$V = \{1, 1, 1\}$	s_v
$+1 + 1 + 1$	$s_v[1] = 3$
$-1 + 1 + 1$	$s_v[2] = 1$
$+1 - 1 + 1$	$s_v[3] = 1$
$+1 + 1 - 1$	$s_v[4] = 1$
$-1 - 1 + 1$	$s_v[5] = -1$
$-1 + 1 - 1$	$s_v[6] = -1$
$+1 - 1 - 1$	$s_v[7] = -1$
$-1 - 1 - 1$	$s_v[8] = -3$

Figure 2. The case where $v_1 = v_2 = v_3 = 1$ and s_v is made of $\binom{3}{0}, \binom{3}{1}, \binom{3}{2}, \binom{3}{3}$ elements of $3, 1, -1, -3$, respectively.

Theorem 5. Among all sets of distinct positive integers where no two distinct sums add to the same value, the set $V = \{2^0, 2^1, \cdots, 2^{n-1}\}$ has a minimal sum

$s_v = v_1 + v_2 + \cdots + v_n = 2^n - 1$.

Proof. Suppose, for the sake of a contradiction, that $s_v = v_1 + v_2 + \cdots + v_n < 2^n - 1$ for some set of distinct positive integers $V = \{v_1, v_2, \cdots, v_n\}$ where no two distinct sums add to the same value.

Take the next enumeration of the 2^n distinct sums, $s_v[1] > s_v[2] > \cdots > s_v[2^n]$, and by our supposition, $2^n - 2 \geq s_v[1]$ and $s_v[2^n] \geq -2^n + 2$, so $s_v[1] - s_v[2^n] \leq 2^{n+1} - 4$.

Let $\{s_{v,1}, s_{v,2}\} \subseteq \{s_v[1], s_v[2], \cdots, s_v[2^n]\}$ where sum $s_{v,1}$ has the list of fixed values $H = \langle \hat{\xi}_{1,1}, \hat{\xi}_{2,1}, \cdots, \hat{\xi}_{n,1} \rangle$ so that $s_{v,1} = \langle v_1, v_2, \cdots, v_n \rangle \cdot H$, where $\cdot$ is the vector dot product. Likewise, the sum $s_{v,2}$ has the list of fixed values $\langle \hat{\xi}_{1,2}, \hat{\xi}_{2,2}, \cdots, \hat{\xi}_{n,2} \rangle$.

The difference of any two distinct sums $s_{v,1} - s_{v,2}$ must be even since any fixed values $\hat{\xi}_{i,1} \in \{-1, +1\}$ and $\hat{\xi}_{i,2} \in \{-1, +1\}$, for $i \in [n]$, are so that,

$$\hat{\xi}_{i,1} - \hat{\xi}_{i,2} \in \{0, -2, +2\}, \tag{15}$$

giving

$$s_{v,1} - s_{v,2} = \sum_{i=1}^{n} v_i \left(\hat{\xi}_{i,1} - \hat{\xi}_{i,2} \right) \tag{16}$$

which must be even.

Starting from $s_v[1]$ and going to $s_v[2^n]$ contains $2^n - 1$ intervals. Since all $s_v[i]$, for $i \in [2^n]$, are different and their differences must be even so $s_v[1] - s_v[2^n]$ spans at least $2(2^n - 1) = 2^{n+1} - 2$. That is, $s_v[1] - s_v[2^n] \geq 2^{n+1} - 2$. This gives a contradiction of the assumption $s_v[1] - s_v[2^n] \leq 2^{n+1} - 4$, completing the proof.

Given a set of distinct positive integers V where $|V| = n$, Theorem 5 indicates that $\max_{x \in \mathbb{Z}} \{\mathbb{P}[S_v = x]\} \leq \frac{1}{2^n}$. So in the case where all distinct sums of V add to different values, erasing a wash sale loss may have a very large impact. In particular, the multi-set $V = \{1, 1, \cdots, 1\}$ has largest loss $s_v[2^n] = -n$, where Theorem 5 indicates $V = \{2^0, 2^1, \cdots, 2^{n-1}\}$ has the largest loss $s_v[2^n] = -2^n + 1$. In this case, when no distinct sums add to the same value, let $U = \{2^n - (2i - 1) : i \in [2^n]\}$ giving $\max_{x \in U} \{\mathbb{P}[S_v = x]\} = \frac{1}{2^n}$. Assuming wash sales occur with the same random and uniform probability among all losses, the expected disallowed loss is $\frac{2^n - 1}{n}$. This is because all losses are of the form $-(2^{i-1})$, for $i \in [n+1]$, and by assumption these losses all have the same probability of occurring.

Since $\mathbb{E}[S_v] = 0$ by Lemma 3, Littlewood-Offord results are useful for understanding likely values for S_v. That is, $\max_{x \in \mathbb{Z} - \{0\}} \{\mathbb{P}[S_v = x]\}$ gives most likely capital gains or losses outside of the expected value $\mathbb{E}[S_v] = 0$. None of the s_v values in Figure 2 are 0, but if V has an even number of 1s, then the most common value is 0.

The following tail bound is given by [42] where $\|v_1, v_2, \cdots, v_n\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$,

$$\mathbb{P}\left[\sum_{i=1}^{n} \xi_i v_i > t \|v_1, v_2, \cdots, v_n\|_2 \right] \leq e^{-t^2/2} \tag{17}$$

$$\mathbb{P}\left[S_v > t \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \right] \leq e^{-t^2/2} \tag{18}$$

$$\mathbb{P}\left[S_v > t\sigma_{S_v}\right] \le e^{-t^2/2} \tag{19}$$

Since by Theorem 4, $\sigma_{S_v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.

Suppose $V = \{1, 1, \cdots, 1\}$ and $|V|$ is odd. Since no sum of V is 0, there are $\dfrac{n}{2}$ capital gains and $\dfrac{n}{2}$ capital losses. This means if $S_v = t\sigma$, then there are $\dfrac{n}{2} + \dfrac{t\sigma}{2}$ capital gains and $\dfrac{n}{2} - \dfrac{t\sigma}{2}$ capital losses. Losses are necessary for wash sales. Therefore, the bound $\mathbb{P}\left[S_v > t\sigma_{S_v}\right] \le e^{-t^2/2}$ gives the probability there are at least $\dfrac{t\sigma_{S_v}}{2}$ more gains than losses. That is, there are $\dfrac{t\sigma_{S_v}}{2}$ fewer opportunities for wash sales.

Following **Figure 2**, given $|V| = n$ then $s_v[1] = n$ is the case with zero capital losses. Likewise, $s_v[2^n] = -n$ is the case with zero capital gains. By Lemma 3, since $\mathbb{E}[S_v] = 0$ and $s_v[1] + \cdots + s_v[2^n] = 0$, thus $-n = s_v[2] + \cdots + s_v[2^n]$. Also suppose a single wash sale disallows a capital loss among all identical capital gains and losses. The single wash sale disallows a single capital loss giving the expected capital gain or loss:

$$\frac{\left(s_v[2]+1\right) + \left(s_v[3]+1\right) + \cdots + \left(s_v[2^n]+1\right)}{2^n - 1}. \tag{20}$$

The term $s_v[1]$ is excluded since it has no losses, hence no wash sales.

The boy-girl ± 30 birthday problem gives a necessary condition for wash sales of substantially identical securities. Recall $B_{30}(252, g, b)$ is the probability of at least one boy-girl ± 30 birthday collision, so $1 - B_{30}(252, g, b)$ is the probability of no such birthday collision.

Given any number of boy-girl ± 30 birthday collisions of the same security and suppose these birthday collisions produce at most a single wash sale. In this case let G be a total taxable gain or loss where all gains and losses are the same. Suppose these gains and losses are all 1. This gives,

$$\mathbb{E}[G] = \left(1 - B_{30}(252, g, b)\right)\frac{s_v[1] + s_v[2] + \cdots + s_v[2^n]}{2^n} \tag{21}$$

$$+ B_{30}(252, g, b)\frac{\left(s_v[2]+1\right) + \cdots + \left(s_v[2^n]+1\right)}{2^n - 1} \tag{22}$$

$$= \frac{1 - B_{30}(252, g, b)}{2^n - 1}(0) + \frac{B_{30}(252, g, b)}{2^n - 1}\left(2^n - 1 - n\right) \tag{23}$$

$$= B_{30}(252, g, b)\left(1 - \frac{n}{2^n - 1}\right). \tag{24}$$

5. Conclusions and Further Directions

This paper shows the probabilistic method may be used to model some tax implications for wash sales. Variations of the birthday problem and the Littlewood-Offord problem are applied to certain tax implications of wash sales.

Modeling and simulating taxes are important in both public policy settings as well as in practical tax planning. In public policy settings, conflicting fiscal and social policies make tax rules contentious. In tax planning, unexpected events may have serious consequences. Thus, reducing certain taxes to mathematical terms gives an unusual level of percision. Such percision can only benefit public policy and tax planning.

Acknowledgements

References

[1]   von Mises, R. (1939) Über Aufteilungs-und Besetzungs-Wahrscheinlichkeiten. *Revue de la Faculté des Sciences de l' Université d' Istanbul*, **4**, 145-163.

[2]   Feller, W. (1968) An Introduction to Probability Theory and Its Applications. 3rd Edition, John Wiley, Hoboken.

[3]   Cormen, T., Leiserson, C., Rivest, R. and Stein, C. (2001) Introduction to Algorithms. 2nd Edition, MIT Press, Cambridge.

[4]   Littlewood, J. and Offord, A. (1943) On the Number of Real Roots of A random Algebraic Equation (III). *Matematicheskii Sbornik*, **12**, 277-286.

[5]   Erdös, P. (1945) On a Lemma of Littlewood and Offord. *Bulletin of the AMS—American Mathematical Society*, **51**, 898-902. http://dx.doi.org/10.1090/S0002-9904-1945-08454-7

[6]   Tao, T. and Vu, V. (2010) A Sharp Inverse Littlewood-Offord Theorem. *Random Structures & Algorithms*, **37**, 525-539. http://dx.doi.org/10.1002/rsa.20327

[7]   US Internal Revenue Service, Investment Income and Expenses (including Capital Gains and Losses) (2015) IRS Publication 550, Cat. No. 15093R for 2014 Tax Returns. See Pages 59+. https://www.irs.gov/pub/irs-pdf/p550.pdf

[8]   FINRA Regulatory Notice 11-35, Effective 8-Aug-2011. https://www.finra.org/sites/default/files/NoticeDocument/p124062.pdf

[9]   Jensen, B. and Marekwica, M. (2011) Optimal Portfolio Choice with Wash Sale Constraints. *Journal of Economic Dynamics and Control*, **35**, 1916-1937. http://dx.doi.org/10.1016/j.jedc.2011.06.007

[10] Hull, J. (2015) Options, Futures, and other Derivatives. 9th Edition, Pearson, New York.

[11] Constantinides, G.M. (1983) Capital Market Equilibrium with Personal Tax. *Econometrica*, **51**, 611-636.

[12] Constantinides, G.M. (1984) Optimal Stock Trading with Personal Taxes: Implications for Prices and the Abnormal January Returns. *Journal of Financial Economics*, **13**, 65-89. http://dx.doi.org/10.1016/0304-405X(84)90032-1

[13] Dammon, R., Spatt, C. and Zhang, H. (2001) Optimal Consumption and Investment with Capital Gains Taxes. *Review of Financial Studies*, **14**, 583-616. http://dx.doi.org/10.1093/rfs/14.3.583

[14] Pat's World Math Blog (2012). http://pballew.blogspot.com/2011/01/who-created-birthday-problem-and-even.html

[15] Ball, W. and Coxeter, H. (1987) Mathematical Recreations and Essays. 13th Edition, Dover, New York, 45-46.

[16] Naus, J. (1968) An Extension of the Birthday Problem. *The American Statistician*, **22**, 27-

29. http://dx.doi.org/10.2307/2681879

[17] Abramson, M. and Moser, W. (1970) More Birthday Surprises. *American Mathematical Monthly*, **77**, 856-858. http://dx.doi.org/10.2307/2317022

[18] Crilly, T. and Nandy, S. (1987) The Birthday Problem for Boys and Girls. *The Mathematical Gazette*, **71**, 19-22. http://dx.doi.org/10.2307/3616281

[19] Pinkham, R. (1988) A Convenient Solution to the Birthday Problem for Girls and Boys. *The Mathematical Gazette*, **72**, 129-130. http://dx.doi.org/10.2307/3618930

[20] Nishimura, K. and Sibuya, M. (1988) Occupancy with Two Types of Balls. *Annals of the Institute of Statistical Mathematics*, **40**, 77-91. http://dx.doi.org/10.1007/BF00053956

[21] Galbarith, S. and Holmes, M. (2012) A Non-Uniform Birthday Problem with Applications to Discrete Logarithms. *Discrete Applied Mathematics*, **160**, 1547-1560. http://dx.doi.org/10.1016/j.dam.2012.02.019

[22] Selivanov, B. (1995) On the Waiting Time in a Scheme for the Random Allocation of Colored Particles. *Discrete Mathematics and Applications*, **5**, 73-82. http://dx.doi.org/10.1515/dma.1995.5.1.73

[23] Chatterjee, S., Diaconis, P. and Meckes, E. (2005) Exchangeable Pairs and Poisson Approximation. *Probability Surveys*, **2**, 64-106. http://dx.doi.org/10.1214/154957805100000096

[24] Arratia, R., Goldstein, L. and Gordon, L. (1990) Poisson Approximation and the Chen-Stein Method. *Statistical Science*, **5**, 403-424.

[25] DasGupta, A. (2005) The Matching, Birthday and the Strong Birthday Problem: A Contemporary Review. *Journal of Statistical Planning and Inference*, **130**, 377-389. http://dx.doi.org/10.1016/j.jspi.2003.11.015

[26] Bradford, P., Perevalova, I., Smid, M. and Ward, C. (2006) Indicator Random Variables in Traffic Analysis and the Birthday Problem. 31*st Annual IEEE Conference on Local Computer Networks* (*LCN* 2006), Tampa, 14-16 November 2006, 1016-1023. http://dx.doi.org/10.1109/lcn.2006.322217

[27] Stinson, D. (2005) Cryptography Theory and Practice. 3rd Edition, CRC Press, Boca Raton.

[28] Song, R., Green, T., McKenna, M. and Glynn M. (2007) Using Occupancy Models to Estimate the Number of Duplicate Cases in a Data System without Unique Identifiers. *Journal of Data Science*, **5**, 53-66.

[29] Diaconis, P. and Mosteller, F. (1989) Methods for Studying Coincidences. *Journal of the American Statistical Association*, **84**, 853-861. http://dx.doi.org/10.1080/01621459.1989.10478847

[30] DasGupta, A. (2004) Sequences, Patterns, and Coincidences. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.8367

[31] Apoport, A. (1998) Birthday Problems: A Search for Elementary Solutions. *The Mathematical Gazette*, **82**, 111-114. http://dx.doi.org/10.2307/3620175

[32] Flajolet, P., Gardy, D. and Thimonier, L. (1992) Birthday Paradox, Coupon Collectors, Caching Algorithms and Self-Organizing Search. *Discrete Applied Mathematics*, **39**, 207-229. http://dx.doi.org/10.1016/0166-218X(92)90177-C

[33] Motwani, R. and Raghavan, P. (1995) Randomized Algorithms. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511814075

[34] Mitzenmacher, M. and Upfal, E. (2005) Probability and Computing. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511813603

[35] Blom, G., Holst, L. and Sandell, D. (1994) Problems and Snapshots from the World of

Probability. Springer, New York. http://dx.doi.org/10.1007/978-1-4612-4304-5

[36] Pinkham, R. (1985) 69.32 The Birthday Problem, Pairs and Triples. *The Mathematical Gazette*, **69**, 279. http://dx.doi.org/10.2307/3617572

[37] Graham, R., Knuth, D. and Patashnik, O. (1994) Concrete Mathematics: A Foundation for Computer Science. Addison Wesley, Boston. http://dx.doi.org/10.2307/3617572

[38] Wendl, M. (2003) Collision Probability between Sets of Random Variables. *Statistics & Probability Letters*, **64**, 249-254. http://dx.doi.org/10.1016/S0167-7152(03)00168-8

[39] Alon, N. and Spencer, J.H. (2008) The Probabilistic Method. 3rd Edition, Wiley, Hoboken. http://dx.doi.org/10.1002/9780470277331

[40] Nguyen, H. and Vu, V. (2012) Small Ball Probability, Inverse Theorems, and Applications. http://arxiv.org/abs/1301.0019

[41] Tao, T. and Vu, V. (2006) Additive Combinatorics. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511755149

[42] Montgomery-Smith, S. (1990) The Distribution of Rademacher Sums. *Proceedings of the American Mathematical Society*, **109**, 517-522. http://dx.doi.org/10.1090/S0002-9939-1990-1013975-0