Bounds on Birthday Attack Times

Michael J. Wiener

20 Hennepin St., Nepean, Ontario, Canada K2J 3Z4
michael.wiener at sympatico.ca

2005 September 8

Abstract. We analyze a generic birthday attack where distinct inputs to some function f are selected until two of the inputs map through f to the same output, meaning that the attack has succeeded. We give tight bounds on the probability of attack success after a given number of inputs are selected as well as tight bounds on the expected number of inputs that must be selected for the attack to succeed.

The types of functions considered include random functions with uniformly random outputs, random functions whose outputs have some arbitrary (biased) probability distribution, concrete functions that are balanced (all outputs have the same number of pre-images), and arbitrary concrete functions. In each case the bounds are given in terms of the probability $(1/\beta)$ that a pair of inputs give the same output, which is different for each type of function. The expected number of steps required to complete a birthday attack in all cases is between $0.7\sqrt{\beta}$ and $2\sqrt{\beta}$. In some cases tighter bounds than this are given.

Compared to previous work in this area, the analysis here gives tighter bounds and is more applicable to the most efficient practical methods used to carry out birthday attacks, such as a generalization of Pollard's rho-method and parallel collision search. However, significant challenges remain in proving bounds for these methods.

Keywords. Birthday attack, hash function.

1 Introduction

Let f be a function (of finite domain) with d inputs and n outputs, $f : D \rightarrow R$, $R = \{R_1, \ldots, R_n\}$. A collision in f is two inputs $a, b \in D$ such that $a \neq b$ and $f(a) = f(b)$. A birthday attack is the process of selecting inputs to f until some pair of them form a collision. For suitable functions f, such collisions can solve certain cryptanalytic problems such as computing discrete logarithms and finding hash function collisions to defeat digital signature schemes. The best techniques known for performing birthday attacks are Teske's modification [4] of a generalization of Pollard's rho-method [2] and parallel collision search [5].

We analyze a generic version of the birthday attack where distinct inputs to f are chosen randomly until some pair of them produce a collision. Each input to f is assumed to be chosen uniformly at random among inputs not already chosen. The known efficient methods of finding collisions choose inputs by repeated iteration of f rather than randomly, and thus significant challenges remain in proving any bounds for practical birthday attacks. However, all empirical evidence seems to show that these results hold for the practical methods.

We consider two types of functions: random and concrete. In the random function case, we treat each output as being randomly-selected according to some probability distribution. The analysis of this case is divided into the uniform probability distribution and the general case of any (finite) discrete distribution. In the concrete function case, we have some known function. The analysis of this case is divided into balanced functions where all outputs R_i have the same number of inputs that map to R_i, and the general case of any function with a finite number of inputs and outputs.

In all four cases considered, tight bounds are given on the probability of finding a collision after a given number of inputs are chosen and on the expected number of inputs that must be selected to get a collision. Approximate analyses for the uniform case appear in many places in the literature (e.g., [3, 5]) and the results here prove that these approximate analyses give accurate results.

Bellare and Kohno also studied this topic for random and concrete functions (balanced and unbalanced) [1]. A difference in their analysis is that they assumed that inputs were chosen uniformly at random (with replacement). This means that they had to factor out false collisions where the same input is chosen twice. Unfortunately, their bounds on the probability of finding a collision after k steps are not particularly tight, especially as k increases into the range where the collision is likely to occur. This made it difficult to accurately estimate the expected number of steps before a collision is found. A restriction in most of their analysis is that $d/n \geq 2$ (domain at least twice as large as the range). The reason for this is that the bounds become progressively less tight as $d \to n$. In practical attacks, the functions used have the same domain and range ($D = R$ and thus $d = n$). However, some of the range values have no pre-image and can be ignored, which reduces the effective n. Thus for practical purposes, as long as f is not a permutation, d is greater than the effective value of n. In these practical attacks, f is hoped to be random, and for a random function with domain and range equal, the number of range elements with no pre-image is about n/e, and if we reduce n by throwing these elements out of the range, then $d/n = 1/(1 - 1/e) \approx 1.58$. Thus an analysis that requires $d/n \geq 2$ does not apply particularly well to practical attacks.

A naive analysis of the probability of having no collision after k steps proceeds as follows. Compute the probability that any particular pair of inputs give a collision. Call this probability $1/\beta$. To have no collision after k steps requires that none of $k(k-1)/2$ pairs of inputs give a collision. If we treat the events that each pair of inputs do not collide as independent (which they clearly are not), then the probability of no collision is $(1 - 1/\beta)^{k(k-1)/2}$, which is close to $e^{-k(k-1)/(2\beta)}$ for large β. This probability reaches roughly 60% when k reaches $\sqrt{\beta}$, and thus the expected number of inputs required to produce a collision is close to $\sqrt{\beta}$. This paper shows that despite the obvious flaws in this reasoning, it comes close to giving the correct answer for both random and concrete functions. In fact, we show that the expected number of steps to find a collision is always between $0.7\sqrt{\beta}$ and $2\sqrt{\beta}$. In some cases we give much tighter bounds than this factor of nearly 3.

In the case of random functions, the quantity β is a function of the probability distribution of the function's outputs. If A is the random variable for this probability distribution, then we write β as $\beta(A)$. In the case of a uniform distribution across n outputs, the probability that two outputs collide is $1/n$, and thus $\beta(A) = n$. We assume that all that is known about the random variable A is the value of $\beta(A)$. The bounds computed are valid across all random variables with equal $\beta(A)$.

In the case of concrete functions, we think of the function as a hash function and call it h instead of f, and we write β as $\beta(h)$. Again, we assume that all that is known about the hash function h is the value of $\beta(h)$. The bounds computed are valid across all hash functions with equal $\beta(h)$. While it may not be possible to fully characterize the nature of some hash function in terms of how its inputs map to its outputs, there is some hope that one may be able to compute or place bounds on $1/\beta(h)$ which is the probability that two distinct inputs chosen uniformly at random produce the same output.

An interesting question is whether the random function model or concrete function model better applies to the case of finding hash function collisions. On the surface, one might think that the concrete model is better. However, real hash functions have an infinite domain. The

analysis here applies only to concrete functions with finite domain. Of course the attacker who wishes to find a hash collision must restrict the domain to some finite space to mount the attack. If we know what restriction the attacker will use, then the concrete model applies. However, it seems unlikely that we will know how the attacker restricts the domain. We are then forced to assume that the attacker chooses some random finite space within the domain. In this case, the random function model is the better model.

The four cases analyzed in detail are uniformly random functions (Section 2), general random functions with arbitrary distribution (Section 3), concrete balanced functions (Section 4), and arbitrary concrete functions (Section 5). The theorem proofs and all supporting lemmas are deferred to appendices.

2 Uniformly Random Case

This section gives bounds on the probability of collision after k steps and the expected number of steps to produce a collision for a uniformly random function.

Definition 1. *Let W_n be the random variable for the number of iterations in the following procedure. Repeatedly choose one of n values uniformly at random (with replacement) and stop when any value is selected a second time (a "collision").*

The following equivalent definition of W_n is more familiar in the context of birthday attacks. Let $F_{D,R}$ denote the set of all possible functions from D to R, where $|D| = d$, and $|R| = n$. Assume for now that $d > n$ so that there exists a collision in every function in $F_{D,R}$. Choose a particular function $f : D \to R$ uniformly at random from $F_{D,R}$ and select distinct elements of D, x_1, x_2, ..., until we have a collision among the range elements, $f(x_i) = f(x_j)$, for some $i \neq j$. W_n is the random variable for the number of elements of D selected in this process. Note that W_n cannot exceed $n + 1$ because in the worst case the first n inputs will cover all range elements, and the next one must cause a collision. We can extend this definition to the $d = n$ case if we define the number of steps to a collision to be $n + 1$ in the case where no collision exists, which happens if and only if f is a permutation. In analyzing the run-time of collision search in the random case, we assume that we do not know which function f is chosen from $F_{D,R}$. If we did know the particular f, then we would be in the concrete function case (see Section 5).

Definition 2. *Let $P_{n,k} = \prod_{j=0}^{k-1}(1 - \frac{j}{n})$. This is just the probability that no collision occurs in the first k steps in the procedure of Definition 1 (i.e., $P(W_n > k) = P_{n,k}$). This definition provides a short-hand notation, but also has the advantage that $P_{n,k}$ is defined for non-integer n, which will be useful in later sections.*

The probability that two inputs produce a collision is $P(W_n = 2) = 1/n$. The quantity β is defined to be the reciprocal of this probability, and thus $\beta = n$. Theorem 1 gives tight bounds on $P(W_n > k)$. The upper bound is valid for all n and k, while the lower bound is valid for $n \geq 1000$ and k well beyond the range where the collision is likely to occur. In fact, the lower bound becomes invalid after the probability of not having a collision has dropped to 1 in a million. Theorem 2 gives tight bounds on the expected value $E(W_n)$ that are valid for all n.

Theorem 1. Probability bounds for the uniformly random case.
For integers $n > 0$ and $k \geq 0$, $P(W_n > k) \leq e^{-\frac{k(k-1)}{2n}}$.
If $n \geq 1000$, and $0 \leq k \leq 2\sqrt{n \ln n}$, $P(W_n > k) \geq e^{-\frac{k^2}{2n} - \frac{k^3}{6n^2}}$.

If $n \geq 1000$, and $k > 2\sqrt{n \ln n}$, $P(W_n > k) < \frac{1}{n^2} \leq 10^{-6}$.
(See Appendix A for the proof.)

Theorem 2. Expectation bounds for the uniformly random case.

$$-\frac{2}{5} < E(W_n) - \sqrt{\frac{\pi n}{2}} < \frac{8}{5}.$$

(See Appendix A for the proof.)

Example 1. Consider the case of a random function with 160-bit outputs ($n = 2^{160}$). Then $-2/5 < E(W_n) - 2^{80}\sqrt{\pi/2} < 8/5$, and thus $E(W_n) \approx 1.25 \times 2^{80}$.

A cryptographer may also be concerned with the number of steps required for the probability of finding a collision to reach some threshold such as 1%. Solving $P(W_n > k) = 1 - 0.01$ for k with both bounds in Theorem 1 gives two values for k that are separated by about $\frac{1}{2}$ (far more accuracy than is required) with $k \approx 0.142 \times 2^{80}$.

A cryptanalyst who actually performs an attack in practice is more concerned with how long it will take to have a 99% chance of finding a collision. Solving $P(W_n > k) = 1 - 0.99$ for k with both bounds gives two values for k separated by about $\frac{1}{2}$ with $k \approx 3.03 \times 2^{80}$.

3 General Random Case

This section gives bounds on the probability of collision after k steps and the expected number of steps to produce a collision for a random function in the general case where the probability distribution is not necessarily uniform.

Definition 3. *Let X_A be the random variable for the number of iterations in the following procedure. Let A be a random variable for choosing among elements of a set based on some probability distribution. Let $a_i = P(A = R_i)$, $i = 1, \ldots, n$. Choose among values in this set with probability distribution determined by A (with replacement) and stop when any value is selected a second time.*

X_A is a generalization of W_n that allows for non-uniform random selection of values. If U_n is the random variable for the uniform probability distribution among n range values (i.e., $P(U_n = R_i) = 1/n$ for $i = 1, \ldots, n$), then $X_{U_n} = W_n$.

An equivalent definition of X_A is to think of a random function f whose output probabilities have bias determined by random variable A. Then repeatedly select distinct inputs to f and stop when there is a collision in the outputs.

Definition 4. *Let $\beta(A)$ be the reciprocal of the probability that two trials with random variable A give the same output. Then*

$$\beta(A) = \frac{1}{P(X_A = 2)} = \frac{1}{\sum_{i=1}^{n} a_i^2}.$$

For the uniform probability distribution, $\beta(U_n) = n$. For any other distribution, $1 \leq \beta(U_n) < n$. Theorems 3 and 4 give bounds on $P(X_A > k)$ and $E(X_A)$ in terms of $\beta(A)$.

Theorem 3. Probability bounds for the general random case.
Let A be any finite discrete random variable and k an integer. If $k \geq 1$,

$$P(X_A > k) \leq \left(1 - \frac{1}{\sqrt{\beta(A)}}\right)^{k-1} \left(1 + \frac{k-1}{\sqrt{\beta(A)}}\right) \leq e^{-\frac{k-1}{\sqrt{\beta(A)}}} \left(1 + \frac{k-1}{\sqrt{\beta(A)}}\right).$$

If $k \geq 0$,

$$P(X_A > k) \geq P_{\beta(A),k}.$$

If $\beta(A) \geq 1000$, and $0 \leq k \leq 2\sqrt{\beta(A) \ln \beta(A)}$,

$$P(X_A > k) \geq e^{-\frac{k^2}{2\beta(A)} - \frac{k^3}{6(\beta(A))^2}}.$$

(See Appendix B for the proof.)

The second upper bound is almost as tight as the first one in the normal range of interest, and is much easier to work with when seeking the value of k that gives a particular probability of success in finding a collision.

Theorem 4. Expectation bounds for the general random case.
For any finite discrete random variable A,

$$\sqrt{\frac{\pi\beta(A)}{2}} - \frac{2}{5} < E(X_A) \leq 2\sqrt{\beta(A)}.$$

(See Appendix B for the proof.)

Theorem 4 shows that in a certain sense, random variable X_A acts like $W_{\beta(A)}$, which is the uniform case with $\beta(A)$ possible outputs. The upper and lower bounds on $E(X_A)$ differ by less than a factor of 1.6 for large $\beta(A)$.

The tightness of the bounds in Theorem 3 is not immediately obvious. This is best illustrated with an example:

Example 2. Consider the case of a random function with 160-bit outputs ($n = 2^{160}$), and $\beta(A) = 2^{128}$. Then $1.25 \times 2^{64} < E(X_A) \leq 2 \times 2^{64}$.

For the cryptographer who is concerned with the number of steps required for the probability of finding a collision to reach 1%, solving $P(X_A > k) = 1 - 0.01$ for k with both bounds in Theorem 3 gives $0.141 \times 2^{64} < k < 0.149 \times 2^{64}$.

For the cryptanalyst who is concerned with the number of steps required for the probability of finding a collision to reach 99%, solving $P(X_A > k) = 1 - 0.99$ for k with both bounds in Theorem 3 gives $3.03 \times 2^{64} < k < 6.64 \times 2^{64}$.

4 Balanced Function Case

This section gives bounds on the probability of collision after k steps and the expected number of steps to produce a collision for a concrete balanced function h.

The function h is balanced if every element R_i of its range has exactly t inputs that map to it, for some constant t. We denote such a balanced function $h_{t,n}$. We exclude $t = 1$ in this analysis because no collisions exist in h in this case.

Definition 5. *Let $Y_{t,n}$ be the random variable for the number of iterations in the following procedure. Let h be a balanced function of $d = nt$ inputs and n outputs. Choose distinct inputs to h one at a time with each input chosen uniformly at random among the inputs that have not yet been chosen. Stop when any pair of inputs map through h to the same output.*

Definition 6. *Let $\beta_{t,n}$ be the reciprocal of the probability that a uniformly randomly selected pair of distinct inputs to $h_{t,n}$ map to the same output. After the first input is selected, there are $t-1$ of the remaining $nt-1$ inputs that would cause a collision. Then*

$$\beta_{t,n} = \frac{1}{P(Y_{t,n} = 2)} = \frac{nt-1}{t-1}.$$

On the surface, it may appear that balanced functions should behave the same as uniformly random functions, but this is not the case because random functions tend to be slightly imbalanced. We see this in the definition of $\beta_{t,n}$ where once the first input has produced output R_i, the second input is less likely to produce R_i than the other possible outputs. But with uniformly random functions, all outputs are equally likely at each step. $Y_{t,n}$ tends to behave like $W_{nt/(t-1)}$.

Theorems 5 and 6 give bounds on $P(Y_{t,n} > k)$ and $E(Y_{t,n})$ in terms of $N = nt/(t-1) = \beta_{t,n} + 1/(t-1)$, a very close estimate of $\beta_{t,n}$.

Theorem 5. Probability bounds for the balanced function case.
Let t be an integer $(t \geq 2)$, n a positive integer, $N = nt/(t-1) = \beta_{t,n} + 1/(t-1)$, and k a nonnegative integer. Then $P(Y_{t,n} > k) \leq e^{-\frac{k(k-1)}{2N}}$.
If $n \geq 1000$, and $0 \leq k \leq 2\sqrt{N \ln N}$, $P(Y_{t,n} > k) \geq e^{-\frac{k^2}{2N} - \frac{2k^3}{3N^2}}$.
(See Appendix C for the proof.)

Theorem 6. Expectation bounds for the balanced function case.
Let t be an integer $(t \geq 2)$, n a positive integer, and $N = nt/(t-1) = \beta_{t,n} + 1/(t-1)$.
Then $E(Y_{t,n}) < \sqrt{\frac{\pi N}{2}} + \frac{8}{5}$. If $n \geq 1000$, $E(Y_{t,n}) > \sqrt{\frac{\pi N}{2}} - \frac{3}{2}$.
(See Appendix C for the proof.)

It is interesting that compared to uniformly random functions, balanced functions are more resistant to birthday attacks by a factor of approximately $\sqrt{t/(t-1)}$.

Example 3. Consider the case of a balanced function with 161-bit inputs and 160-bit outputs ($n = 2^{160}$ and $t = 2$). Then $-3/2 < E(Y_{t,n}) - 2^{80}\sqrt{\pi} < 8/5$, and thus $E(Y_{t,n}) \approx 1.77 \times 2^{80}$.

For the cryptographer who is concerned with the number of steps required for the probability of finding a collision to reach 1%, solving $P(Y_{t,n} > k) = 1 - 0.01$ for k with both bounds in Theorem 5 gives two values for k that are separated by about $\frac{1}{2}$ (far more accuracy than is required) with $k \approx (0.201)2^{80}$.

For the cryptanalyst who is concerned with the number of steps required for the probability of finding a collision to reach 99%, solving $P(Y_{t,n} > k) = 1 - 0.99$ for k with both bounds in Theorem 5 gives two values for k separated by about $\frac{1}{2}$ with $k \approx 4.29 \times 2^{80}$.

5 General Function Case

This section gives bounds on the probability of collision after k steps and the expected number of steps to produce a collision for a concrete function h in the general case where h is not necessarily balanced.

Definition 7. *Let Z_h be the random variable for the number of iterations in the following procedure. Let h be a function of d inputs and n outputs such that for $i = 1, \ldots, n$, the number of inputs that give output R_i is d_i. Choose distinct inputs to h one at a time with each input chosen uniformly at random among the inputs that have not yet been chosen. Stop when any pair of inputs map through h to the same output.*

Definition 8. *Let $\beta(h)$ be the reciprocal of the probability that a uniformly randomly selected pair of distinct inputs to h map to the same output. If the first input maps through h to output R_i (with probability d_i/d), the probability that the second input will map to R_i is $(d_i - 1)/(d - 1)$. Then*

$$\beta(h) = \frac{1}{P(Z_h = 2)} = \frac{d(d-1)}{\sum_{i=1}^{n} d_i(d_i - 1)}.$$

This definition of $\beta(h)$ is analogous to Bellare and Kohno's definition of a quantity ($r^{\mu(h)}$ in their parlance) that is defined to be $d^2/(\sum_{i=1}^{n} d_i^2)$ [1]. The difference between these two measures comes from the fact that they did not assume that inputs are chosen to be distinct. In practical birthday attacks [2, 5] where a function is iterated, an input can be chosen twice, but instead of leading to a false collision, this leads to finding a real collision. Thus the model where inputs are chosen to be distinct better reflects how birthday attacks are performed in practice.

Note that if the range of h include outputs that have no corresponding input ($d_i = 0$), $\beta(h)$ is not affected if we remove these range elements from R and reduce n. To guarantee that a collision exists, we require that d be greater than this reduced n. This does not affect the applicability of these results to practical birthday attacks where the domain and range are equal ($d = n$) because if any collisions exist, there must be at least one element of R that has no corresponding input, and then d is greater than the reduced n.

Theorems 7 and 8 give bounds on $P(Z_h > k)$ and $E(Z_h)$ in terms of $\beta(h)$. Throughout the proofs of these theorems, we assume that n has been reduced as described above so that $d > n$.

Theorem 7. Probability bounds for the general function case.
Let h be any function of d inputs and n outputs, $d > n \geq 2$, and k an integer. If $k \geq 1$, then

$$P(Z_h > k) \leq \left(1 - \frac{1}{\sqrt{\beta(h)}}\right)^{k-1} \left(1 + \frac{k-1}{\sqrt{\beta(h)}}\right) \leq e^{-\frac{k-1}{\sqrt{\beta(h)}}} \left(1 + \frac{k-1}{\sqrt{\beta(h)}}\right).$$

If $k < \sqrt{\beta(h)} + 1$, then

$$\begin{aligned} P(Z_h > k) &\geq \left(1 + \frac{1}{\sqrt{\beta(h)}}\right)^{k-1} \left(1 - \frac{k-1}{\sqrt{\beta(h)}}\right) \\ &\geq \left(e\left(1 - \frac{1}{2\sqrt{\beta(h)} + \frac{4}{3}}\right)\right)^{\frac{k-1}{\sqrt{\beta(h)}}} \left(1 - \frac{k-1}{\sqrt{\beta(h)}}\right). \end{aligned}$$

If $k \geq \sqrt{\beta(h)} + 1$, then $P(Z_h > k) \geq 0$.
(See Appendix D for the proof.)

The last lower bound seems vacuous, but it is not possible to do much better. Consider the case of a function h of $d = n + 1$ inputs with $d_1 = 2$, and $d_i = 1$, $i = 2, \ldots, n$. In this case, $\beta(h) = n(n+1)/2$, and $n \approx \sqrt{2\beta(h)}$. Because $P(Z_h > k) = 0$ when $k > n$, we have $P(Z_h > k) = 0$ for k greater than roughly $\sqrt{2\beta(h)}$.

The second bounds in both the upper and lower bound cases are almost as tight as the first bounds in the normal range of interest and are much easier to work with when seeking the value of k that gives a particular probability of success in finding a collision.

Theorem 8. Expectation bounds for the general function case.
For any function h with d inputs and n outputs, $d > n \geq 1$,

$$(e-2)\sqrt{\beta(h)} < E(Z_h) \leq 2\sqrt{\beta(h)}.$$

(See Appendix D for the proof.)

The tightness of the bounds in Theorem 7 is not immediately obvious. This is best illustrated with an example:

Example 4. Consider the case of a function with 160-bit outputs and $\beta(h) = 2^{128}$. Note that the actual number of inputs d is not important except to the extent that it affects $\beta(h)$. Then $0.718 \times 2^{64} < E(Z_h) \leq 2 \times 2^{64}$.

For the cryptographer who is concerned with the number of steps required for the probability of finding a collision to reach 1%, solving $P(Z_h > k) = 1 - 0.01$ for k with both bounds in Theorem 7 gives $0.135 \times 2^{64} < k < 0.149 \times 2^{64}$.

For the cryptanalyst who is concerned with the number of steps required for the probability of finding a collision to reach 99%, solving $P(Z_h > k) = 1 - 0.99$ for k with both bounds in Theorem 7 gives $0.996 \times 2^{64} < k < 6.64 \times 2^{64}$.

6 Conclusion

We studied generic birthday attacks for finding collisions in a given function f, and gave tight bounds on the probability of success of the attack after a given number of steps and tight bounds on the expected number of steps to complete the attack. These bounds are given for both random and concrete functions. In each case, we gave bounds in terms of the probability $1/\beta$ that a pair of distinct uniformly randomly selected inputs to f give the same output.

By Theorems 4 and 8, the expected number of steps required to complete a birthday attack is always between $0.718\sqrt{\beta}$ and $2\sqrt{\beta}$ for both random and concrete functions. In the case of Theorem 4, the worst case occurs for the lower bound when $\beta(A) = 1$. Even in this case, the expected number of steps is inside the range above. It may seem that this result may not hold for the uniformly random case (Theorem 2), but uniformly random functions are also part of the set of random functions with arbitrary probability distribution, and so are also governed by Theorem 4. Similarly, balanced concrete functions are governed by Theorem 8.

The range given for the expected number of steps is tight at the upper bound, but there appears to be a small amount of room to improve the lower bound. For the upper bound, consider a constant concrete function h_1 of $d \geq 2$ inputs and only one output ($n = 1$, and $d_1 = d$). Then $E(Z_{h_1}) = 2$ because the collision must occur on the second input selected. In this case $\beta(h_1) = d(d-1)/(d_1(d_1-1)) = 1$. Therefore, $E(Z_{h_1}) = 2\sqrt{\beta(h_1)}$. For the lower bound, the most extreme example we found was a concrete function h_2 of $n+1$ inputs and $n \geq 2$ outputs with $d_1 = 2$, and $d_i = 1$, $i = 2, \ldots, n$. We can show that $E(Z_{h_2}) = \frac{2}{3}(n+1)$, and

$\beta(h_2) = n(n+1)/2$. In the limit as $n \to \infty$, $E(Z_{h_2}) = \frac{2\sqrt{2}}{3}\sqrt{\beta(h_2)} \approx 0.943\sqrt{\beta(h_2)}$. This seems to leave some room to improve the lower bound. However, the existing bounds are tight enough for practical purposes. The real challenge that remains is to actually compute $\beta(h)$ or $\beta(A)$ for functions used in birthday attacks on real cryptographic schemes such as hash functions and signatures schemes based on the difficulty of computing discrete logarithms.

References

1. M. Bellare and T. Kohno, Hash Function Balance and its Impact on Birthday Attacks, *Advances in Cryptology–Eurocrypt 2004* (LNCS 3027), Springer-Verlag, Berlin, 2004, pp. 401–418. The full version of this paper appears in the Cryptology ePrint Archive, http://eprint.iacr.org, number 2003/065, last revised 2004 Nov. 27.
2. J.M. Pollard, Monte Carlo Methods for Index Computation (mod p), *Mathematics of Computation*, vol. 32, no. 143 (July 1978), pp. 918–924.
3. D. Stinson, Cryptography Theory and Practice, Second edition, CRC Press, 2002.
4. E. Teske, Speeding Up Pollard's Rho Method for Computing Discrete Logarithms, *Algorithmic Number Theory Symposium III* (LNCS 1423), Springer-Verlag, Berlin, 1998, pp. 541–554.
5. P.C. van Oorschot and M.J. Wiener, Parallel Collision Search with Cryptanalytic Applications, *Journal of Cryptology*, vol. 12, no. 1 (1999), pp. 1–28.

A Proofs of Theorems 1 and 2

There are a number of supporting lemmas in addition to the theorem proofs. Undoubtedly, some of the results in this section are not new because the uniform random case has been studied extensively. We give these results to make this paper self-contained and because these results get reused in later proofs.

Lemma 1. *For $m \geq 0$, $\sum_{j=0}^{k-1} j^m \leq k^{m+1}/(m+1)$.*

Proof. Because $m \geq 0$, x^m never has negative slope for $x \geq 0$. Therefore, $j^m \leq \int_j^{j+1} x^m dx$, and applying this inequality for $j = 0, \ldots, k-1$ and combining the results gives $\sum_{j=0}^{k-1} j^m \leq \int_0^k x^m dx = k^{m+1}/(m+1)$. □

Proof of Theorem 1 (see Section 2 for statement of theorem).
Upper bound. The probability that $W_n > k$ is equal to the probability that no collision occurs after the first k iterations:

$$P(W_n > k) = \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right).$$

For the $k = 0$ case, the product is over zero values, which we define to be equal to 1, the correct value for $P(W_n > 0)$. For $k \leq n$ each of the factors in the product is positive, and we can apply the Taylor series, $-\ln(1-x) = \sum_{i=1}^{\infty} x^i/i$, to get

$$\begin{aligned} -\ln(P(W_n > k)) &= \sum_{j=0}^{k-1} \sum_{i=1}^{\infty} \frac{j^i}{in^i} = \sum_{j=0}^{k-1} \left(\frac{j}{n} + \frac{j^2}{2n^2} + \ldots\right) \\ &= \frac{k(k-1)}{2n} + \frac{k(k-1)(2k-1)}{12n^2} + \ldots . \end{aligned} \tag{1}$$

Each of the terms in the sum above is nonnegative, so that if we truncate the series at any point, we get a lower bound. Therefore,

$$-\ln(P(W_n > k)) \geq \frac{k(k-1)}{2n}, \qquad P(W_n > k) \leq e^{-\frac{k(k-1)}{2n}}.$$

Lower bound. Using $\sum_{j=0}^{k-1} j^m \le k^{m+1}/(m+1)$ from Lemma 1, we can replace all but the first term of $-\ln(P(W_n > k))$ in equation (1) as follows

$$-\ln(P(W_n > k)) \le \frac{k(k-1)}{2n} + \frac{k^3}{3 \cdot 2n^2} + \frac{k^4}{4 \cdot 3n^3} + \frac{k^5}{5 \cdot 4n^4} + \dots \ .$$

Change the constants in the denominators of all the terms after the first two to 12 (which leaves the inequality true because we have not decreased the size of the right hand side), to form a geometric series:

$$-\ln(P(W_n > k)) \le \frac{k(k-1)}{2n} + \frac{k^3}{6n^2} + \frac{k^4}{12n^2}\left(1 + \frac{k}{n} + \frac{k^2}{n^2} + \dots\right).$$

For $n \ge 1000$, and $k \le 2\sqrt{n \ln n}$, k/n will always be less than $1/6$, and thus the geometric series in brackets sums to less than $6/5$. Therefore,

$$\begin{aligned} -\ln(P(W_n > k)) &\le \frac{k(k-1)}{2n} + \frac{k^3}{6n^2} + \frac{k^4}{10n^2} \\ &\le \frac{k^2}{2n} + \frac{k^3}{6n^2} + \frac{k}{n}\left(\frac{k^3}{10n^2} - \frac{1}{2}\right). \end{aligned}$$

With the restrictions on n and k, the factor in brackets is always negative, and

$$-\ln(P(W_n > k)) \le \frac{k^2}{2n} + \frac{k^3}{6n^2}, \qquad P(W_n > k) \ge e^{-\frac{k^2}{2n} - \frac{k^3}{6n^2}}.$$

Truncating equation (1) after the second term gives

$$\begin{aligned} -\ln(P(W_n > k)) &\ge \frac{k(k-1)}{2n} + \frac{k(k-1)(2k-1)}{12n^2} \\ &\ge \frac{k}{2n}\left(k - 1 + \frac{(k-1)(2k-1)}{6n}\right). \end{aligned}$$

For $n \ge 1000$, and $k > 2\sqrt{n \ln n}$, the $(k-1)(2k-1)/(6n)$ term is always greater than 1 giving $P(W_n > k) < e^{-\frac{k^2}{2n}}$. Substituting $k = 2\sqrt{n \ln n}$ gives $P(W_n > k) < e^{-2 \ln n} = \frac{1}{n^2} \le 10^{-6}$. □

Lemma 2. *For any random variable V that takes on only positive integer values, the expected value of V is*

$$E(V) = \sum_{k=0}^{\infty} P(V > k).$$

Proof.

$$\begin{aligned} E(V) &= \sum_{k=1}^{\infty} kP(V = k) = \sum_{k=1}^{\infty} k(P(V > k-1) - P(V > k)) \\ &= \sum_{k=1}^{\infty} kP(V > k-1) - \sum_{k=1}^{\infty} kP(V > k) \\ &= \sum_{k=0}^{\infty} (k+1)P(V > k) - \sum_{k=0}^{\infty} kP(V > k) \\ &= \sum_{k=0}^{\infty} P(V > k). \end{aligned}$$

□

Lemma 3. *If a function $g(x)$ is continuous and is never increasing for $x \geq 0$, and $\int_0^\infty g(x)dx$ converges, then*

$$\sum_{k=1}^{\infty} g(k-\tfrac{1}{2}) \leq \int_0^\infty g(x)dx + \tfrac{1}{2}g(\tfrac{1}{2}), \;\; \sum_{k=0}^{m} g(k) \geq \int_0^{m+1} g(x)dx.$$

Proof. Because $g(x)$ is never increasing, for $a, b \geq 0$,

$$bg(a) \geq \int_a^{a+b} g(x)dx \geq bg(a+b). \tag{2}$$

Applying this inequality for $a = 0$, $b = 1/2$ gives $\int_0^{1/2} g(x)dx \geq (1/2)g(1/2)$, and applying it for $b = 1$, and $a = 1/2, 3/2, 5/2, \ldots$ gives

$$\int_{\frac{1}{2}}^\infty g(x)dx \geq \sum_{k=2}^{\infty} g(k-\tfrac{1}{2}).$$

Combining these two results and adding $g(\frac{1}{2})$ to the sum gives

$$\sum_{k=1}^{\infty} g(k-\tfrac{1}{2}) \leq \int_0^\infty g(x)dx + \tfrac{1}{2}g(\tfrac{1}{2}).$$

Applying inequality (2) for $b = 1$, and $a = 0, \ldots, m$ gives

$$\sum_{k=0}^{m} g(k) \geq \int_0^{m+1} g(x)dx.$$

□

Lemma 4. *For $n \geq 2$,*

$$\int_{2\sqrt{n\ln n}}^\infty e^{-\frac{x^2}{2n}}dx \leq \frac{2\sqrt{n\ln n}}{n^2-2}.$$

Proof. Because $e^{-x^2/(2n)}$ is never increasing for $x \geq 0$, if $a \geq 0$ then

$$\int_a^{2a} e^{-\frac{x^2}{2n}}dx \leq ae^{-\frac{a^2}{2n}}.$$

Substituting $a = 2^i\sqrt{n\ln n}$ and then replacing $n^{2^{2i-1}}$ with n^{2^i} (which leaves the inequality true if $i \geq 1$):

$$\int_{2^i\sqrt{n\ln n}}^{2^{i+1}\sqrt{n\ln n}} e^{-\frac{x^2}{2n}}dx \leq \frac{2^i\sqrt{n\ln n}}{n^{2^{2i-1}}} \leq \sqrt{n\ln n}\left(\frac{2}{n^2}\right)^i.$$

Applying this inequality for $i = 1, 2, \ldots$, we get a geometric progression that sums to the required expression:

$$\int_{2\sqrt{n\ln n}}^\infty e^{-\frac{x^2}{2n}}dx \leq \sqrt{n\ln n}\sum_{i=1}^{\infty}\left(\frac{2}{n^2}\right)^i = \frac{2\sqrt{n\ln n}}{n^2-2}.$$

□

Proof of Theorem 2 (see Section 2 for statement of theorem).
Upper bound. Begin with the expectation equation in Lemma 2 and use the first probability bound in Theorem 1:

$$E(W_n) = \sum_{k=0}^{\infty} P(W_n > k) = 1 + \sum_{k=1}^{\infty} P(W_n > k) \leq 1 + \sum_{k=1}^{\infty} e^{-\frac{k(k-1)}{2n}}.$$

Rewriting $k(k-1)$ as $(k-1/2)^2 - 1/4$ gives

$$E(W_n) \leq 1 + e^{\frac{1}{8n}} \sum_{k=1}^{\infty} e^{-\frac{(k-\frac{1}{2})^2}{2n}}.$$

Applying the first inequality in Lemma 3 gives

$$E(W_n) \leq 1 + e^{\frac{1}{8n}} \left(\int_0^{\infty} e^{-\frac{x^2}{2n}} dx + \frac{1}{2} e^{-\frac{1}{8n}} \right).$$

The integral is a standard definite integral equal to $\sqrt{\pi n/2}$, which gives

$$E(W_n) \leq \frac{3}{2} + e^{\frac{1}{8n}} \sqrt{\frac{\pi n}{2}}.$$

Because $e^x \leq 1 + x + x^2$ for $0 \leq x \leq 1$, we can replace $e^{1/(8n)}$ with $1 + 1/(8n) + 1/(64n^2)$:

$$E(W_n) \leq \frac{3}{2} + \sqrt{\frac{\pi n}{2}} + \left(\frac{1}{8n} + \frac{1}{64n^2} \right) \sqrt{\frac{\pi n}{2}}.$$

The final term is always less than $1/10$ for $n \geq 3$. For $n = 1$ and $n = 2$, we have $E(W_1) = 2$ and $E(W_2) = 2.5$, and the bounds in the theorem statement can be verified directly. Thus

$$E(W_n) < \sqrt{\frac{\pi n}{2}} + \frac{8}{5}.$$

Lower bound. It is easier to give a proof for $n \geq 1000$. So, we computed $E(W_n)$ for $n < 1000$ using methods that track the maximum error in each variable of a computation to verify the theorem for these cases. In fact, we showed that for $n < 1000$ and $n = 2^0, 2^1, \ldots, 2^{64}$, $0 < E(W_n) - (\sqrt{\pi n/2} + 2/3) < 1/(9\sqrt{n})$. For the remainder of the proof of the lower bound, we assume $n \geq 1000$. Begin with the expectation equation in Lemma 2 and truncate it at $m = \lfloor 2\sqrt{n \ln n} \rfloor$ so that we can use the second probability bound in Theorem 1:

$$E(W_n) = \sum_{k=0}^{\infty} P(W_n > k) \geq \sum_{k=0}^{m} P(W_n > k) \geq \sum_{k=0}^{m} e^{-\frac{k^2}{2n}} e^{-\frac{k^3}{6n^2}}.$$

Applying the second inequality in Lemma 3 gives

$$E(W_n) \geq \int_0^{m+1} e^{-\frac{x^2}{2n}} e^{-\frac{x^3}{6n^2}} dx = \int_0^{\infty} e^{-\frac{x^2}{2n}} e^{-\frac{x^3}{6n^2}} dx - \int_{m+1}^{\infty} e^{-\frac{x^2}{2n}} e^{-\frac{x^3}{6n^2}} dx.$$

In the difference of two integrals above, we can drop the $e^{-x^3/(6n^2)}$ factor in the second integral because it is less than or equal to 1 and the change increases the value of the integral, which lowers the lower bound. We can also reduce the lower limit of the second integral to $2\sqrt{n \ln n}$

because the change increases the value of the integral. The same factor in the first integral can be replaced with $1 - x^3/(6n^2)$ because for all y, $e^{-y} \geq 1 - y$.

$$E(W_n) \geq \int_0^\infty e^{-\frac{x^2}{2n}} dx - \int_0^\infty \frac{x^3}{6n^2} e^{-\frac{x^2}{2n}} dx - \int_{2\sqrt{n \ln n}}^\infty e^{-\frac{x^2}{2n}} dx.$$

The first integral is a standard definite integral equal to $\sqrt{\pi n/2}$, the second integral can be integrated by parts and is equal to $1/3$, and the last integral is less than $2\sqrt{n \ln n}/(n^2 - 2)$ by Lemma 4. For $n \geq 1000$, $1/3 + 2\sqrt{n \ln n}/(n^2 - 2) < 2/5$ which gives the lower bound

$$E(W_n) > \sqrt{\frac{\pi n}{2}} - \frac{2}{5}.$$

□

B Proofs of Theorems 3 and 4

There are a number of supporting lemmas and a definition in addition to the theorem proofs.

The exact probability $P(X_A > k)$ is determined by summing the probabilities of all possible ways that the first k outputs can be distinct. For convenience of notation, we will sometimes think of the random variable A as being equal to the set of its probabilities, $A = \{a_1, \ldots, a_n\}$. The following definition provides some useful notation.

Definition 9. *Let $V = \{v_1, \ldots, v_n\}$, and let $S_k(V)$ be the sum of products of all subsets of size k in V:*

$$S_k(V) = \sum_{1 \leq i_1 < \ldots < i_k \leq n} \left(\prod_{j=1}^{k} v_{i_j} \right).$$

For example, $S_2(\{x, y, z\}) = xy + xz + yz$. Let $S_0(V) = 1$, and for $k > n$, $S_k(V) = 0$. When using this definition with a random variable A, one of more of the a_i can be excluded from A with the notation $S_k(A \backslash \{a_i\})$ or $S_k(A \backslash \{a_i, a_j\})$ for example.

Now we can write $P(X_A > k) = (k!)S_k(A)$. The extra factor of $k!$ comes from the fact that S was defined to include subsets without regard to order, and in the probability that k outputs are distinct, each k-subset of distinct outputs can occur in all of its $k!$ orderings. Some useful properties of S are proven in the following lemma.

Lemma 5. *For $0 < i, j \leq n$, $k > 0$,*

$$\begin{aligned}
&S_k(A) = a_i S_{k-1}(A \backslash \{a_i\}) + S_k(A \backslash \{a_i\}), \\
&S_k(A \backslash \{a_i\}) - S_k(A \backslash \{a_j\}) = (a_j - a_i) S_{k-1}(A \backslash \{a_i, a_j\}), \\
&\frac{\partial S_k(A)}{\partial a_i} = S_{k-1}(A \backslash \{a_i\}).
\end{aligned}$$

Proof. $S_k(A)$ consists of some terms involving a_i and others that do not. The sum of the terms without a_i is just $S_k(A \backslash \{a_i\})$. There is a term including a_i for each (k-1)-subset of $A \backslash \{a_i\}$. Therefore,

$$S_k(A) = a_i S_{k-1}(A \backslash \{a_i\}) + S_k(A \backslash \{a_i\}).$$

Applying this rule to separate a_j from $S_k(A\backslash\{a_i\})$, and to separate a_i from $S_k(A\backslash\{a_j\})$ gives

$$S_k(A\backslash\{a_i\}) = a_j S_{k-1}(A\backslash\{a_i, a_j\}) + S_k(A\backslash\{a_i, a_j\}),$$
$$S_k(A\backslash\{a_j\}) = a_i S_{k-1}(A\backslash\{a_i, a_j\}) + S_k(A\backslash\{a_i, a_j\}).$$

Taking the difference between these two equations gives

$$S_k(A\backslash\{a_i\}) - S_k(A\backslash\{a_j\}) = (a_j - a_i) S_{k-1}(A\backslash\{a_i, a_j\}).$$

The final identity follows directly from the first identity by taking partial derivatives with respect to a_i. □

We now seek bounds on $P(X_A > k)$ assuming that $\beta(A)$ is a known value, say β_0. This situation is a little different from the uniformly random case because the bounds need to be valid for all random variables A whose distributions are such that $\beta(A) = \beta_0$ rather than just being concerned with the single uniform distribution. A starting point is to find those distributions that minimize and maximize $P(X_A > k)$:

Lemma 6. *The minimum and maximum of $P(X_A > k)$ across fixed k and varying random variable A, subject to the constraint $\beta(A) = \beta_0$ can be found by considering only those random variables A such that the probabilities $a_1, \ldots, a_n$ take on only two distinct values other than zero.*

Proof. If one or more of the probabilities in A is zero, this is equivalent to a different random variable with only the non-zero probabilities and a smaller n. Without loss of generality, assume $a_1 \geq \ldots \geq a_n > 0$ for some $n > 0$.

The $k = 0, 1, 2$ cases are trivially true because $P(X_A > 0) = P(X_A > 1) = 1$, and $P(X_A > 2) = \beta_0$, which are all constants. For the remainder of the proof, assume $k \geq 3$.

Using Lagrange multipliers to find the stationary points (including local minima, maxima, and high-order saddle points) of $P(X_A > k)$ subject to the constraints $\sum_{i=1}^n a_i - 1 = 0$ and $\sum_{i=1}^n a_i^2 - 1/\beta_0 = 0$ gives the following equation for some constants λ_1 and λ_2:

$$\frac{\partial P(X_A > k)}{\partial a_i} = \lambda_1 \frac{\partial(\sum_{i=1}^n a_i - 1)}{\partial a_i} + \lambda_2 \frac{\partial(\sum_{i=1}^n a_i^2 - 1/\beta_0)}{\partial a_i}, \quad i = 1, \ldots, n.$$

Using $P(X_A > k) = (k!) S_k(A)$ and applying the third identity in Lemma 5 gives

$$(k!) S_{k-1}(A\backslash\{a_i\}) = \lambda_1 + 2\lambda_2 a_i, \quad i = 1, \ldots, n. \tag{3}$$

The only way for all a_i to be equal is if $\beta_0 = n$, in which case the probabilities are fully determined to be $a_1 = \ldots = a_n = 1/n$. This lemma is trivially true in this case. For the rest of this proof, assume that $a_{j-1} > a_j$ for some j, $1 < j \leq n$. Let m be such that $1 \leq m < j$. Then by the earlier assumption about the probabilities being in descending order, and $a_{j-1} > a_j$, we have $a_m > a_j$. Apply equation (3) for the $i = m$ and $i = j$ cases, and solve for λ_1:

$$\lambda_1 = (k!) S_{k-1}(A\backslash\{a_m\}) - 2\lambda_2 a_m = (k!) S_{k-1}(A\backslash\{a_j\}) - 2\lambda_2 a_j.$$

Rearranging this equation and applying the second identity in Lemma 5 gives

$$\begin{aligned} 2\lambda_2(a_m - a_j) &= (k!)(S_{k-1}(A\backslash\{a_m\}) - S_{k-1}(A\backslash\{a_j\})) \\ &= (k!)(a_j - a_m) S_{k-2}(A\backslash\{a_m, a_j\})). \end{aligned}$$

Because $a_m \neq a_j$, we can cancel out the factors of $(a_m - a_j)$ to get

$$S_{k-2}(A\backslash\{a_m, a_j\})) = -\frac{2\lambda_2}{k!}, \;\; m = 1, \ldots, j-1.$$

Using this equation for $m = u$ and $m = v$, $1 \leq u, v < j$, and taking the difference between the equations gives

$$S_{k-2}(A\backslash\{a_u, a_j\})) - S_{k-2}(A\backslash\{a_v, a_j\})) = 0.$$

Applying the second identity in Lemma 5 gives

$$(a_v - a_u)S_{k-3}(A\backslash\{a_u, a_v, a_j\})) = 0.$$

For $k \geq 3$, $S_{k-3}(A\backslash\{a_u, a_v, a_j\}) > 0$ because it is equal to 1 for $k = 3$, and is the sum of products of positive probabilities for $k > 3$. Thus, $a_u = a_v$, and because this is true for $1 \leq u, v < j$, the a_m are all equal for $1 \leq m < j$. By similar reasoning, the a_m are all equal for $j \leq m \leq n$. The stationary points occur when $a_1 = \ldots = a_{j-1} > a_j = \ldots = a_n$. Therefore, the stationary points of $P(X_A > k)$ occur when $a_1, \ldots, a_n$ take on at most two distinct non-zero values.

The maximum and minimum of $P(X_A > k)$ must occur at one of the stationary points or on the boundary of the region where $a_1, \ldots, a_n$ satisfy the constraints that their sum is 1, and $\beta(A) = \beta_0$. At all points on this boundary, one or more of the a_i is zero, or at least one pair of probabilities in A are equal. To see this, consider any A that satisfies the constraints, but has no zero probabilities and all probabilities are distinct. Because there are two constraint equations, only $n-2$ of the probabilities are independent. We can make independent infinitesimal changes to any $n-2$ of the probabilities, and still be able to adjust the final two probabilities to satisfy the constraint equations without using negative or complex values. Thus the case where there are no zero probabilities and all probabilities are distinct is not a boundary case. So, the boundary cases to consider are those with one or more zero probabilities or one or more subsets of the probabilities constrained to be equal. The case with the zeros can be eliminated because this is just the same as having no zeros with a smaller value of n. When one or more subsets of the probabilities are constrained to be equal, we can repeat the Lagrange multiplier analysis with more constraint equations. Suppose that for some constants j and m, $j < m$, the probabilities $a_j, a_{j+1}, \ldots, a_m$ are constrained to be equal. Then we have the extra constraint equations $a_j - a_{j+1} = 0$, $a_{j+1} - a_{j+2} = 0$, ..., $a_{m-1} - a_m = 0$. Let $\mu_j, \ldots, \mu_{m-1}$ be the constants associated with these equations when we use Lagrange multipliers. Then equation (3) for $i = j, \ldots, m$ becomes

$$\begin{aligned}
(k!)S_{k-1}(A\backslash\{a_j\}) &= \lambda_1 + 2\lambda_2 a_j + \mu_j, \\
(k!)S_{k-1}(A\backslash\{a_i\}) &= \lambda_1 + 2\lambda_2 a_i, + \mu_i - \mu_{i-1}, \;\; i = j+1, \ldots, m-1, \\
(k!)S_{k-1}(A\backslash\{a_m\}) &= \lambda_1 + 2\lambda_2 a_m - \mu_{m-1}.
\end{aligned} \tag{4}$$

Note that even if another subset of probabilities outside of $a_j, \ldots, a_m$ were constrained to be equal, it would not affect these equations because the partial derivatives for the additional constraint equations with respect to each of $a_j, \ldots, a_m$ would be zero. Because $a_j = \ldots = a_m$, the left hand sides of all equations in (4) are equal. The terms involving λ_1 and λ_2 are also equal in all equations. Thus $\mu_j = \mu_{j+1} - \mu_j = \mu_{j+2} - \mu_{j+1} = \ldots = \mu_{m-1} - \mu_{m-2} = -\mu_{m-1}$. The only solution to this set of equations is $\mu_j = \mu_{j+1} = \ldots = \mu_{m-1} = 0$. If another subset of probabilities were constrained to be equal, the constants associated with their constraint equations would be zero as well. Thus the equations in (3) are not affected by the additional constraints, and the final conclusion remains the same: all stationary points of $P(X_A > k)$

occur when $a_1, \ldots, a_n$ take on at most two distinct non-zero values. Therefore, the maximum and minimum of $P(X_A > k)$ must occur among the cases where $a_1, \ldots, a_n$ take on at most two distinct non-zero values. □

Lemma 7. *Given a triple of real numbers (t_1, t_2, t_3) such that $t_1 = t_2 > 0$ and $t_3 > 0$, there exists another triple (u_1, u_2, u_3) such that $u_1, u_2, u_3 > 0$, $u_1 + u_2 + u_3 = t_1 + t_2 + t_3$ and $u_1^2 + u_2^2 + u_3^2 = t_1^2 + t_2^2 + t_3^2$. If $t_3 > t_1$ there exists a triple (u_1, u_2, u_3) with the additional property that $u_1u_2u_3 < t_1t_2t_3$, and if $t_3 < t_1$ there exists a triple (u_1, u_2, u_3) with the additional property that $u_1u_2u_3 > t_1t_2t_3$.*

Proof. Write (t_1, t_2, t_3) in the form $(a, a, a(1+b))$, $a > 0$, $b > -1$. Let

$$(u_1, u_2, u_3) = (a(1 + b\epsilon), a(1 + b(1 - \epsilon - c)/2), a(1 + b(1 - \epsilon + c)/2)),$$

where $c = \sqrt{(1+\epsilon)^2 - 4\epsilon^2}$, for some ϵ, $0 < \epsilon < 1$, and $\epsilon < 1/|b|$. Note that because $0 < \epsilon < 1$, c is real, and $0 < c < 1 + \epsilon$. With simple (but messy) algebra, one can verify that $u_1 + u_2 + u_3 = t_1 + t_2 + t_3$ and $u_1^2 + u_2^2 + u_3^2 = t_1^2 + t_2^2 + t_3^2$, and show that

$$\delta = u_1u_2u_3 - t_1t_2t_3 = a^3b^3\epsilon^2(\epsilon - 1).$$

What remains to be shown is that $u_1, u_2, u_3 > 0$, and for the $t_3 > t_1$ ($b > 0$) case, $\delta < 0$, and for the $t_3 < t_1$ ($b < 0$) case, $\delta > 0$. For brevity, the rest of this proof uses the inequalities $a > 0$, $b > -1$, $0 < \epsilon < 1$, $\epsilon < 1/|b|$, and $0 < c < 1 + \epsilon$ without stating them.

Case 1: $b > 0$.

$$\begin{aligned}
&u_1 = a(1 + b\epsilon) > 0.\\
&u_2 = a(1 + b(1 - \epsilon - c)/2) > a(1 + b(1 - \epsilon - (1 + \epsilon))/2) = a(1 - b\epsilon) > 0.\\
&u_3 = a(1 + b(1 - \epsilon + c)/2) > a(1 + b(1 - \epsilon)/2) > 0.\\
&\delta = a^3b^3\epsilon^2(\epsilon - 1) < 0.
\end{aligned}$$

Case 2: $b < 0$.

$$\begin{aligned}
&u_1 = a(1 + b\epsilon) > a(1 - b/b) = 0.\\
&u_2 = a(1 + b(1 - \epsilon - c)/2) > a(1 + b(1 - \epsilon)/2) > a(1 + b) > 0.\\
&u_3 = a(1 + b(1 - \epsilon + c)/2) > a(1 + b(1 - \epsilon + (1 + \epsilon))/2) = a(1 + b) > 0.\\
&\delta = a^3b^3\epsilon^2(\epsilon - 1) > 0.
\end{aligned}$$

□

Lemma 8. *Given a triple of real numbers (t_1, t_2, t_3) such that $t_1 = 0$ and $t_3 > t_2 > 0$, there exists another triple (u_1, u_2, u_3) such that $u_1, u_2, u_3 > 0$, $u_1+u_2+u_3 = t_1+t_2+t_3$, $u_1^2+u_2^2+u_3^2 = t_1^2 + t_2^2 + t_3^2$, and $u_1u_2u_3 > t_1t_2t_3$.*

Proof. Write (t_1, t_2, t_3) in the form $(0, a, a(1+2b))$, $a, b > 0$. Let

$$(u_1, u_2, u_3) = (2a\epsilon, a(1 + b - \epsilon - c), a(1 + b - \epsilon + c)),$$

where $c = \sqrt{(b + 3\epsilon)(b - \epsilon) + 2\epsilon}$, for some ϵ, $0 < \epsilon < 1/4$, and $\epsilon < b/4$. With these restrictions on ϵ, $c > 0$, and rewriting c as $c = \sqrt{(b + \epsilon + 1/2)^2 - 4\epsilon^2 - (b - \epsilon) - 1/4}$, we see that $c < b+\epsilon+1/2$.

One can verify that $u_1 + u_2 + u_3 = t_1 + t_2 + t_3$ and $u_1^2 + u_2^2 + u_3^2 = t_1^2 + t_2^2 + t_3^2$. What remains to be shown is that $u_1, u_2, u_3 > 0$, and $\delta = u_1u_2u_3 - t_1t_2t_3 > 0$:

$$u_1 = 2a\epsilon > 0.$$
$$u_2 = a(1 + b - \epsilon - c) > a(1 + b - \epsilon - (b + \epsilon + 1/2)) = a(1 - 4\epsilon)/2 > 0.$$
$$u_3 = a(1 + b - \epsilon + c) > a(1 + b - \epsilon) > a(1 + (3/4)b) > 0.$$

Because $t_1 = 0$ and $u_1, u_2, u_3 > 0$, we have $t_1t_2t_3 = 0$, $u_1u_2u_3 > 0$, and $\delta > 0$. □

Lemma 9. *Let $n \geq k \geq 3$. Let A be a random variable containing the subset of probabilities $\{t_1, t_2, t_3\}$ and such that $A\backslash\{t_1, t_2, t_3\}$ is either empty or contains at least $k - 3$ nonzero probabilities. Let A' be a member of the set of random variables such that $\beta(A') = \beta(A)$.*
(1) If $t_1 = t_2$ and $t_3 > t_1 > 0$, A' exists such that $P(X_{A'} > k) < P(X_A > k)$.
(2) If $t_1 = t_2$ and $t_1 > t_3 > 0$, A' exists such that $P(X_{A'} > k) > P(X_A > k)$.
(3) If $t_1 = 0$ and $t_3 > t_2 > 0$, A' exists such that $P(X_{A'} > k) > P(X_A > k)$.

Proof. Permuting the probabilities in A does not affect $P(X_A > k)$, and thus without loss of generality, we can assume that if a subset $\{t_1, t_2, t_3\}$ of probabilities are in A, they occur in the first three positions (a_1, a_2, a_3). Let $B = A\backslash\{a_1, a_2, a_3\}$. Repeated use of the first identity in Lemma 5 gives

$$S_k(A) = (a_1a_2a_3)S_{k-3}(B) + (a_1a_2 + a_1a_3 + a_2a_3)S_{k-2}(B) + (a_1 + a_2 + a_3)S_{k-1}(B) + S_k(B). \quad (5)$$

Suppose that random variable A' differs from A only in the first three probabilities (a'_1, a'_2, a'_3 replace a_1, a_2, a_3). Suppose also that $a'_1 + a'_2 + a'_3 = a_1 + a_2 + a_3$ so that all the probabilities in A' sum to 1, and $a'^2_1 + a'^2_2 + a'^2_3 = a_1^2 + a_2^2 + a_3^2$ so that $\beta(A') = \beta(A)$. The identity $a_1a_2 + a_1a_3 + a_2a_3 = ((a_1 + a_2 + a_3)^2 - (a_1^2 + a_2^2 + a_3^2))/2$ means that we also have $a'_1a'_2 + a'_1a'_3 + a'_2a'_3 = a_1a_2 + a_1a_3 + a_2a_3$. Referring to equation (5), $S_k(A')$ and $S_k(A)$ differ only in the first term, which gives

$$P(X_{A'} > k) - P(X_A > k) = (k!)(a'_1a'_2a'_3 - a_1a_2a_3)S_{k-3}(B).$$

Note that because B is either empty or contains at least $k-3$ nonzero probabilities (by assumption in the lemma statement), $S_{k-3}(B) > 0$, and thus the sign of $P(X_{A'} > k) - P(X_A > k)$ is the same as the sign of $a'_1a'_2a'_3 - a_1a_2a_3$. Parts (1) and (2) of this lemma follow from Lemma 7, and part (3) follows from Lemma 8. □

Lemma 10. *Consider the set of random variables A than consist of at most n probabilities ($n > 1$), and have $\beta(A) = \beta_0$ for some constant β_0. $P(X_A > k)$ is a maximum across this set when $A = A_{\max}$, where*

$$P(A_{\max} = R_1) = (1 + \sqrt{((n/\beta_0) - 1)(n - 1)})/n,$$
$$P(A_{\max} = R_i) = (1 - \sqrt{((n/\beta_0) - 1)/(n - 1)})/n, \;\; i = 2, \ldots, n,$$

and is a minimum when $A = A_{\min}$, where $m = \lceil \beta_0 \rceil$ and

$$P(A_{\min} = R_1) = (1 - \sqrt{((m/\beta_0) - 1)(m - 1)})/m,$$
$$P(A_{\min} = R_i) = (1 + \sqrt{((m/\beta_0) - 1)/(m - 1)})/m, \;\; i = 2, \ldots, m,$$
$$P(A_{\min} = R_i) = 0, \;\; i = m + 1, \ldots, n.$$

Proof. This lemma is trivially true if $n = 2$ or $\beta_0 = 1$ or $\beta_0 = n$ because in each of these cases there is only one random variable A that satisfies $\beta(A) = \beta_0$. If $n = 2$, then a_1 and a_2 are fixed by β_0. If $\beta_0 = 1$, then one of the a_i is 1 and the rest are zero. If $\beta_0 = n$, then $a_1 = \ldots = a_n = 1/n$. This lemma also is trivially true if $k < 3$ or $k > n$ because then $P(X_A > k)$ is a fixed constant. $P(X_A > 0) = P(X_A > 1) = 1$, $P(X_A > 2) = 1/\beta_0$, and if $k > n$, then $P(X_A > k) = 0$. Assume $n \geq k \geq 3$ and $1 < \beta_0 < n$ for the rest of this proof.

By Lemma 6, the maximum and minimum values of $P(X_A > k)$ occur when A contains at most two distinct probabilities other than zero. Because $1 < \beta_0 < n$, there must be exactly two distinct nonzero probabilities. Let the two probabilities be x and y, $y > x > 0$.

Maximum case. We seek a random variable A that maximizes $P(X_A > k)$. We can eliminate any random variable that does not have at least k nonzero probabilities because in this case $P(X_A > k) = 0$ which cannot be larger than $P(X_{A_{\max}} > k)$. This handles the requirement concerning nonzero probabilities in Lemma 9. If the y probability appears more than once in A, then $\{y, y, x\}$ is a subset of A and we can increase $P(X_A > k)$ by changing A as explained in case (2) of Lemma 9. Therefore, the maximum occurs when A contains exactly one copy of y. If a zero probability appears in A, then $\{0, x, y\}$ is a subset of A and we can increase $P(X_A > k)$ by changing A as explained in case (3) of Lemma 9. Therefore, the maximum occurs when A contains one copy of y and $n - 1$ copies of x. These probabilities must sum to 1, and their squares must sum to $1/\beta_0$, which gives $y = (1 + \sqrt{((n/\beta_0) - 1)(n - 1)})/n$ and $x = (1 - \sqrt{((n/\beta_0) - 1)/(n - 1)})/n$ as required.

Minimum case. We seek a random variable A that minimizes $P(X_A > k)$. If $\beta(A) = \beta_0$, then A must have at least $m = \lceil \beta_0 \rceil$ nonzero probabilities because it is not possible for fewer than m real numbers to have a sum of 1 and the sum of their squares less than $1/m$. $A_{\min}$ is an example of a random variable with m nonzero probabilities and $\beta(A_{\min}) = \beta_0$. For $k > m$, $P(X_{A_{\min}} > k) = 0$, which is the minimum, and we need only consider the case where $k \leq m$. This handles the requirement concerning nonzero probabilities in Lemma 9. If the x probability appears more than once in A, then $\{x, x, y\}$ is a subset of A and we can decrease $P(X_A > k)$ by changing A as explained in case (1) of Lemma 9. Therefore, the minimum occurs when A contains exactly one copy of x. If there are a total of m' nonzero probabilities in A (one x and $m' - 1$ copies of y), then $x + (m' - 1)y = 1$ and $x^2 + (m' - 1)y^2 = 1/\beta_0$. If $m' > m$, then $x \leq 0$, which contradicts the initial assumption that $x > 0$. Therefore, $m' \leq m$. If $m' < m$, then x is a complex number. Therefore, $m' = m$, $x = (1 - \sqrt{((m/\beta_0) - 1)(m - 1)})/m$, and $y = (1 + \sqrt{((m/\beta_0) - 1)/(m - 1)})/m$, as required. □

Lemma 11. *Let n and k be integers, $n \geq k > 1$. Let u and v be real numbers, with $v = \sqrt{1/n^2 + (1 - 1/n^2)u^2}$. Let $g(u) = (1 - u)^{k-1}(1 + (k - 1)u)$. Over the range $0 \leq u < 1$, $g(v)/g(u)$ is a minimum of $(1 - 1/n)^{k-1}(1 + (k - 1)/n)$ when $u = 0$ and $v = 1/n$.*

Proof. Because $0 \leq u < 1$, we have $1/n \leq v < 1$, and thus the derivative

$$\frac{d(g(v)/g(u))}{du} = \frac{u(n^2 - 1)(1 - v)^{k-2}k(k - 1)(k - 2)}{n^4(1 - u)^{k-2}(1 + (k - 1)u)^2(v + 1 - (1 - u)(1 - 1/n^2))}$$

is never negative, and $g(v)/g(u)$ is a minimum when u is at its minimum ($u = 0$). □

Lemma 12. *Let n and k be integers, $n > 1$, $k \geq 2$, β_0 a real number, $1 \leq \beta_0 \leq n$, and $\alpha_n = \sqrt{((n/\beta_0) - 1)/(n - 1)}$. Then $p(n) = (1 - \alpha_n)^{k-1}(1 + (k - 1)\alpha_n)P_{n,k}$ never decreases as n increases.*

Proof. If $\beta_0 = 1$ or $k > n$, then $p(n) = 0$ and this lemma is trivially true. Assume that $\beta_0 > 1$ and $n \geq k \geq 2$ for the rest of this proof. Consider the ratio $p(n+1)/p(n)$. We can apply Lemma 11 with $u = \alpha_n$, $v = \alpha_{n+1}$, $g(u) = p(n)/P_{n,k}$, and $g(v) = p(n+1)/P_{n+1,k}$, because the required relationship $v = \sqrt{1/n^2 + (1-1/n^2)u^2}$ holds. Then $p(n+1)/p(n)$ is a minimum across varying β_0 when $\alpha_n = 0$ and $\beta_0 = n$, and thus

$$\frac{p(n+1)}{p(n)} \geq \left(\frac{n-1}{n}\right)^{k-1} \left(\frac{n+(k-1)}{n}\right) \frac{P_{n+1,k}}{P_{n,k}}.$$

Because $P_{n,k} = (1-1/n)(1-2/n)\ldots(1-(k-1)/n)$, we can show that

$$P_{n+1,k} = \left(\frac{n}{n-(k-1)}\right) \left(\frac{n}{n+1}\right)^{k-1} P_{n,k}.$$

Then

$$\frac{p(n+1)}{p(n)} \geq \left(\frac{n-1}{n+1}\right)^{k-1} \left(\frac{n+(k-1)}{n-(k-1)}\right).$$

Differentiating the right hand side with respect to n gives

$$-\frac{2\left(\frac{n-1}{n+1}\right)^k k(k-1)(k-2)}{(n-1)^2(n-k+1)^2}$$

which is never positive for $n \geq k \geq 2$. Thus the lower bound on $p(n+1)/p(n)$ is minimized as $n \to \infty$.

$$\lim_{n\to\infty} \left(\frac{n-1}{n+1}\right)^{k-1} \left(\frac{n+(k-1)}{n-(k-1)}\right) = 1.$$

This means $p(n+1)/p(n) \geq 1$, and thus $p(n)$ never decreases as n increases. □

Lemma 13. *Let $\beta_0 > 1$, $m = \lceil \beta_0 \rceil$, and $\delta = m - \beta_0$, $0 \leq \delta < 1$. Let $0 \leq k \leq m$, and $\alpha = \sqrt{((m/\beta_0)-1)/(m-1)}$. Then $(1+\alpha)^{k-1}(1-(k-1)\alpha)P_{m,k} \geq P_{\beta_0,k}$.*

Proof. We proceed by induction on k. When $k = 0$ or $k = 1$, $(1+\alpha)^{k-1}(1-(k-1)\alpha) = 1$, and the inequality to be proven reduces to $P_{m,k} \geq P_{\beta_0,k}$, which is true because $m-1 < \beta_0 \leq m$, and when $k \leq n+1$, $P_{n,k}$ never decreases as n increases. The induction assumption is to assume that this lemma is true for $k = i-1$, for some i, $2 \leq i \leq m$:

$$(1+\alpha)^{i-2}(1-(i-2)\alpha) \geq P_{\beta_0,i-1}/P_{m,i-1}.$$

Then to show this lemma is true, it is sufficient to show that

$$\frac{(1+\alpha)^{i-1}(1-(i-1)\alpha)}{(1+\alpha)^{i-2}(1-(i-2)\alpha)} \geq \frac{P_{\beta_0,i}/P_{m,i}}{P_{\beta_0,i-1}/P_{m,i-1}}$$

because this result combined with the induction assumption implies that this lemma is true for $k = i$. Because $P_{n,k}/P_{n,k-1} = 1-(k-1)/n$, this simplifies to having to show that

$$\frac{(1+\alpha)(1-(i-1)\alpha)}{1-(i-2)\alpha} \geq \frac{1-(i-1)/\beta_0}{1-(i-1)/m} = 1 - \frac{(i-1)\delta}{\beta_0(m-(i-1))}.$$

Substituting $m = \beta_0 + \delta$ into the definition of α gives $\alpha^2 = \delta/(\beta_0(m-1))$. Because $\delta < 1$, we have $\beta_0 > m-1$ and $\alpha < 1/(m-1)$. Multiplying both sides of this inequality by $-(i-2)(m-1)$

and adding $m-1$ gives $(m-1)(1-(i-2)\alpha) \geq m-(i-1)$. Inverting both sides, multiplying by δ/β_0, and using $\alpha^2 = \delta/(\beta_0(m-1))$ gives

$$\frac{\alpha^2}{1-(i-2)\alpha} \leq \frac{\delta}{\beta_0(m-(i-1))}.$$

Negating both sides, multiplying by $i-1$, and adding 1 gives the required result:

$$\frac{(1+\alpha)(1-(i-1)\alpha)}{1-(i-2)\alpha} \geq 1 - \frac{(i-1)\delta}{\beta_0(m-(i-1))}.$$

□

Proof of Theorem 3 (see Section 3 for statement of theorem).
Let n be the number of possible outputs of A. If $n=1$, then $\beta(A)=1$, and this theorem is trivially true. Assume $n>1$ for the rest of this proof.

Upper bound. Let $\alpha = \sqrt{((n/\beta(A))-1)/(n-1)})$. Note that $0 \leq \alpha \leq 1$ because $1 \leq \beta(A) \leq n$. Let $x=(1-\alpha)/n$, $y=(1+(n-1)\alpha)/n$, and define random variable $A_{\max}$ as follows: $P(A_{\max}=R_1)=y$, and $P(A_{\max}=R_i)=x$ for $i=2,\ldots,n$. By Lemma 10, $P(X_A>k)$ is a maximum when $A=A_{\max}$. The event that there are no collisions after k outputs of $A_{\max}$ can be split into two disjoint cases. In the first case, there are no outputs equal to R_1, and all k outputs are different. The probability of this case is $(1-y)^k$ times the probability that all outputs are different given that no output is equal to R_1. The probabilities of the remaining $n-1$ outputs are all equal which is just the uniform case. The conditional probability is $P_{n-1,k}$, and the probability of this case is $(1-y)^k P_{n-1,k}$. In the second case, there is one output equal to R_1, and all $k-1$ remaining outputs are different. The probability of this case is $ky(1-y)^{k-1}P_{n-1,k-1}$. This gives

$$\begin{aligned} P(X_{A_{\max}}>k) &= (1-y)^k P_{n-1,k} + ky(1-y)^{k-1}P_{n-1,k-1} \\ &= (1-y)^{k-1}((1-y)P_{n-1,k} + kyP_{n-1,k-1}). \end{aligned}$$

Because $P_{n,k} = (1-1/n)(1-2/n)\ldots(1-(k-1)/n)$, we can show that

$$P_{n-1,k-1} = \left(\frac{n}{n-1}\right)^{k-1} P_{n,k}, \quad P_{n-1,k} = \left(\frac{n-k}{n-1}\right)\left(\frac{n}{n-1}\right)^{k-1} P_{n,k}.$$

These equations and $y=(1+(n-1)\alpha)/n$ lead to

$$\begin{aligned} P(X_{A_{\max}}>k) &= \left((1-y)\frac{n}{n-1}\right)^{k-1}\left((1-y)\frac{n-k}{n-1} + ky\right)P_{n,k} \\ &= (1-\alpha)^{k-1}(1+(k-1)\alpha)P_{n,k}. \end{aligned}$$

By Lemma 12, $P(X_{A_{\max}}>k)$ never decreases as n increases, and is maximized as $n\to\infty$.

$$\lim_{n\to\infty}\alpha = \frac{1}{\sqrt{\beta(A)}}, \quad \lim_{n\to\infty} P_{n,k} = 1.$$

Then the upper bound for $P(X_A>k)$ is

$$\lim_{n\to\infty} P(X_{A_{\max}}>k) = \left(1-\frac{1}{\sqrt{\beta(A)}}\right)^{k-1}\left(1+\frac{k-1}{\sqrt{\beta(A)}}\right).$$

For large $\beta(A)$, a close upper bound on $(1-1/\sqrt{\beta(A)})$ is $e^{-1/\sqrt{\beta(A)}}$:

$$P(X_A > k) \leq \left(1 - \frac{1}{\sqrt{\beta(A)}}\right)^{k-1} \left(1 + \frac{k-1}{\sqrt{\beta(A)}}\right) \leq e^{-\frac{k-1}{\sqrt{\beta(A)}}} \left(1 + \frac{k-1}{\sqrt{\beta(A)}}\right).$$

Lower bound. The lower bounds are trivially true if $k < 2$ because $P(X_A > 0) = P(X_A > 1) = 1$. For the rest of this proof, assume $k \geq 2$. Let $m = \lceil \beta(A) \rceil$, and redefine α, x, and y as follows: $\alpha = \sqrt{((m/\beta(A)) - 1)/(m-1)})$, $x = (1-(m-1)\alpha)/m$, and $y = (1+\alpha)/m$. Because $x \geq 0$, we have $\alpha \leq 1/(m-1)$. Define $A_{\min}$ as follows: $P(A_{\min} = R_1) = x$, $P(A_{\min} = R_i) = y$ for $i = 2, \ldots, m$, and $P(A_{\min} = R_i) = 0$ for $i = m+1, \ldots, n$. By Lemma 10, $P(X_A > k)$ is a minimum when $A = A_{\min}$. The event that there are no collisions after k outputs of $A_{\min}$ can be split into two disjoint cases. In the first case, there are no outputs equal to R_1, and all k outputs are different. The probability of this case is $(1-x)^k P_{m-1,k}$. In the second case, there is one output equal to R_1, and all $k-1$ remaining outputs are different. The probability of this case is $kx(1-x)^{k-1}P_{m-1,k-1}$. Proceeding similarly to the upper bound case,

$$\begin{aligned} P(X_{A_{\min}} > k) &= (1-x)^k P_{m-1,k} + kx(1-x)^{k-1} P_{m-1,k-1} \\ &= (1+\alpha)^{k-1}(1-(k-1)\alpha) P_{m,k}. \end{aligned}$$

By Lemma 13, $P(X_A > k) \geq P(X_{A_{\min}} > k) \geq P_{\beta(A),k}$. Theorem 1 gave a lower bound on $P_{n,k}$ and this proof is valid for non-integer n. Thus if $\beta(A) \geq 1000$, and $0 \leq k \leq 2\sqrt{\beta(A)\ln\beta(A)}$,

$$P(X_A > k) \geq e^{-\frac{k^2}{2\beta(A)} - \frac{k^3}{6(\beta(A))^2}}.$$

□

Proof of Theorem 4 (see Section 3 for statement of theorem).
Upper bound. Begin with the expectation equation in Lemma 2 and use the first probability bound in Theorem 3:

$$\begin{aligned} E(X_A) &= \sum_{k=0}^{\infty} P(X_A > k) = 1 + \sum_{k=1}^{\infty} P(X_A > k) \\ &\leq 1 + \sum_{k=1}^{\infty} \left(1 - \frac{1}{\sqrt{\beta(A)}}\right)^{k-1} \left(1 + \frac{k-1}{\sqrt{\beta(A)}}\right). \\ &\leq 1 + \sum_{k=0}^{\infty} \left(1 - \frac{1}{\sqrt{\beta(A)}}\right)^{k} + \frac{1}{\sqrt{\beta(A)}} \sum_{k=0}^{\infty} k \left(1 - \frac{1}{\sqrt{\beta(A)}}\right)^{k}. \end{aligned}$$

Using the identities

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad \sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2}, \quad 0 \leq x < 1,$$

for $x = 1 - 1/\sqrt{\beta(A)}$, we get

$$E(X_A) \leq 1 + \frac{1}{\frac{1}{\sqrt{\beta(A)}}} + \frac{1}{\sqrt{\beta(A)}} \left(\frac{1 - \frac{1}{\sqrt{\beta(A)}}}{\frac{1}{\beta(A)}}\right) = 2\sqrt{\beta(A)}.$$

Lower bound. By Theorem 3, $P(X_A > k) \geq P_{\beta(A),k}$. The proof in Theorem 2 of the lower bound case when $n \geq 1000$ works equally well for non-integer n, and thus the lower bound in the

theorem statement is true when $\beta(A) \geq 1000$. For the $\beta(A) < 1000$ case, we begin by observing that because $P_{\beta(A),k} \geq P_{\lfloor \beta(A) \rfloor,k}$, we have $E(X_A) \geq E(W_{\lfloor \beta(A) \rfloor})$. The proof of Theorem 2 states that we proved by computer computation that for integers $n < 1000$, $E(W_n) > \sqrt{\pi n/2} + 2/3$. Thus $E(X_A) > \sqrt{\pi \lfloor \beta(A) \rfloor /2} + 2/3$. The right hand side of this inequality is always greater than $\sqrt{\pi \beta(A)/2} - 2/5$ for $\beta(A) \geq 1$, and thus

$$E(X_A) > \sqrt{\frac{\pi \beta(A)}{2}} - \frac{2}{5}.$$

□

C Proofs of Theorems 5 and 6

There are a number of supporting lemmas in addition to the theorem proofs.

Lemma 14. *If $t \geq 2$, $n \geq 1000$, $N = nt/(t-1)$, and $0 \leq k \leq 2\sqrt{N \ln N}$, then $k/(tn) < 1/8$, and $k/N < 1/6$.*

Proof. The ratio $k/(tn)$ is maximized when k is at its maximum:

$$\frac{k}{tn} \leq \frac{2\sqrt{N \ln N}}{tn} = 2\sqrt{\frac{\ln(tn/(t-1))}{tn(t-1)}}.$$

The right hand side is at its maximum when t is at its minimum ($t = 2$):

$$k/(tn) \leq 2\sqrt{(\ln(2n))/(2n)}.$$

The derivative of the right hand side with respect to n is negative for $n > e/2$. Therefore, we find the maximum value of $k/(tn)$ when n is its minimum of 1000:

$$k/(tn) \leq 2\sqrt{(\ln 2000)/2000} < 0.124 < 1/8.$$

The ratio k/N is maximized when k is at its maximum:

$$k/N \leq 2\sqrt{N \ln N}/N = 2\sqrt{(\ln N)/N}.$$

The right hand side decreases for $N > e$ and thus is maximized when N is its minimum of 1000 when $n = 1000$ and $t \to \infty$:

$$k/N \leq 2\sqrt{(\ln 1000)/1000} < 0.1663 < 1/6.$$

□

Lemma 15. *If $t \geq 2$, $n > 0$, $N = nt/(t-1)$, and $0 \leq k \leq n$, then*

$$\sum_{j=0}^{k-1} \frac{j(t-1)}{tn-j} \geq \frac{k(k-1)}{2N}.$$

If $t \geq 2$, $n \geq 1000$, and $0 \leq k \leq 2\sqrt{N \ln N}$, then

$$\sum_{j=0}^{k-1} \frac{j(t-1)}{tn-j} \leq \frac{k^2}{2N} + \frac{8k^3}{21N^2}.$$

Proof. Start by using the identity $j/(tn-j) = \sum_{m=1}^{\infty}(j/(tn))^m$ and separate out the $m=1$ terms:

$$\sum_{j=0}^{k-1}\frac{j(t-1)}{tn-j} = (t-1)\sum_{j=0}^{k-1}\sum_{m=1}^{\infty}\left(\frac{j}{tn}\right)^m = (t-1)\sum_{j=0}^{k-1}\frac{j}{tn} + (t-1)\sum_{j=0}^{k-1}\sum_{m=2}^{\infty}\left(\frac{j}{tn}\right)^m.$$

Use the identity $\sum_{j=0}^{k-1} j = k(k-1)/2$ on the first sum. Because the second sum consists of nonegative terms, this establishes the lower bound in the lemma statement. Continuing with the upper bound, for the second sum we know that $j/(tn) < 1/8$ by Lemma 14 and the fact that $j < k$:

$$\sum_{j=0}^{k-1}\frac{j(t-1)}{tn-j} \le \frac{k(k-1)}{2N} + (t-1)\sum_{j=0}^{k-1}\sum_{m=2}^{\infty}\left(\frac{j}{tn}\right)^2\left(\frac{1}{8}\right)^{m-2}.$$

For the first term replace $k-1$ with k, and for the sum, the powers of $1/8$ sum to $8/7$, and $\sum_{j=0}^{k-1} j^2 = k(k-1)(2k-1)/6$:

$$\sum_{j=0}^{k-1}\frac{j(t-1)}{tn-j} \le \frac{k^2}{2N} + \frac{8}{7}\left(\frac{t-1}{(nt)^2}\right)\frac{k(k-1)(2k-1)}{6}.$$

Multiply the second term by $t-1$, and replace $k(k-1)(2k-1)$ with $2k^3$ to get

$$\sum_{j=0}^{k-1}\frac{j(t-1)}{tn-j} \le \frac{k^2}{2N} + \frac{8k^3}{21N^2}.$$

□

Lemma 16. *If $t \ge 2$, $n \ge 1000$, $N = nt/(t-1)$, and $0 \le k \le 2\sqrt{N\ln N}$, then*

$$\sum_{j=0}^{k-1}\sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i \le \frac{32k^3}{119N^2}.$$

Proof. Using Lemma 14 and the fact that $j < k$,

$$tn - j = tn\left(1-\frac{j}{tn}\right) > tn\left(1-\frac{k}{tn}\right) > \frac{7}{8}tn.$$

Using this inequality and reversing the order of summation,

$$\sum_{j=0}^{k-1}\sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i \le \sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{8(t-1)}{7tn}\right)^i\sum_{j=0}^{k-1} j^i.$$

Using Lemma 1,

$$\sum_{j=0}^{k-1}\sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i \le \sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{8}{7N}\right)^i\frac{k^{i+1}}{i+1}.$$

Pull a factor of $k/(i(i+1))$ out of the sum, and replace $i(i+1)$ with 6 in the denominator (because $i \ge 2$) to get

$$\sum_{j=0}^{k-1}\sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i \le \frac{k}{6}\sum_{i=2}^{\infty}\left(\frac{8k}{7N}\right)^2\left(\frac{8k}{7N}\right)^{i-2}.$$

By Lemma 14, $k/N < 1/6$, $8k/(7N) < 4/21$, and the sum of powers of $4/21$ is $21/17$:

$$\sum_{j=0}^{k-1}\sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i \le \frac{k}{6}\left(\frac{8k}{7N}\right)^2\frac{21}{17} = \frac{32k^3}{119N^2}.$$

□

Proof of Theorem 5 (see Section 4 for statement of theorem).
If the first j outputs have not produced a collision, there are $n-j$ remaining outputs that have not appeared yet and $t(n-j)$ possible inputs corresponding to those outputs. The number of remaining inputs is $tn-j$, and thus the probability that the next output does not produce a collision is $t(n-j)/(tn-j) = 1-j(t-1)/(tn-j)$. The probability that $Y_{t,n} > k$ is equal to the probability that no collision occurs after the first k iterations:

$$P(Y_{t,n} > k) = \prod_{j=0}^{k-1}\left(1-\frac{j(t-1)}{tn-j}\right).$$

For the $k=0$ case, the product is over zero values, which we define to be equal to 1, the correct value for $P(Y_{t,n} > 0)$. For $k \le n$ each of the factors in the product is positive, and we can apply the Taylor series, $-\ln(1-x) = \sum_{i=1}^{\infty} x^i/i$, to get

$$-\ln(P(Y_{t,n} > k)) = \sum_{j=0}^{k-1}\sum_{i=1}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i.$$

Separating the $i=1$ terms gives

$$-\ln(P(Y_{t,n} > k)) = \sum_{j=0}^{k-1}\frac{j(t-1)}{tn-j} + \sum_{j=0}^{k-1}\sum_{i=2}^{\infty}\frac{1}{i}\left(\frac{j(t-1)}{tn-j}\right)^i. \tag{6}$$

Using the fact that the second sum consists of nonnegative terms along with the first inequality in Lemma 15 gives

$$-\ln(P(Y_{t,n} > k)) \ge \frac{k(k-1)}{2N},$$

which establishes the upper bound in the theorem statement. Applying the second inequality of Lemma 15 to the first term of equation (6), and Lemma 16 to the second term gives

$$-\ln(P(Y_{t,n} > k)) \le \frac{k^2}{2N} + \frac{8k^3}{21N^2} + \frac{32k^3}{119N^2}.$$

Grouping the last two terms and observing that $8/21 + 32/119 < 2/3$ establishes the lower bound in the theorem statement. □

Proof of Theorem 6 (see Section 4 for statement of theorem).
We begin with the probability bounds in Theorem 5 and proceed almost exactly as we did in Theorem 2. For the upper bound, the logic is the same except that N is used in place of n, and this gives

$$E(Y_{t,n}) < \sqrt{\frac{\pi N}{2}} + \frac{8}{5}.$$

as required. For the lower bound, N is used in place of n, and the probability expression uses the constant 2/3 instead of 1/6: $e^{-k^2/(2N)-2k^3/(3N^2)}$ instead of $e^{-k^2/(2n)-k^3/(6n^2)}$. This gives

$$E(Y_{t,n}) > \sqrt{\frac{\pi N}{2}} - \frac{4}{3} - \frac{2\sqrt{N \ln N}}{N^2 - 2}.$$

If $n \geq 1000$, the final term never exceeds 1/6, and thus

$$E(Y_{t,n}) > \sqrt{\frac{\pi N}{2}} - \frac{3}{2}.$$

□

D Proofs of Theorems 7 and 8

There are a number of supporting lemmas in addition to the theorem proofs.

Lemma 17. *Let h be a function of d inputs and n outputs ($d > n$), such that for $i = 1, \ldots, n$, there are d_i inputs that map to the i^{th} output R_i. Let A be a random variable such that $P(A = R_i) = d_i/d$, $i = 1, \ldots, n$. Let k be an integer, $0 \leq k \leq n$. Then $P(Z_h > k) = P(X_A > k)/P_{d,k}$.*

Proof. To get $P(Z_h > k)$ we need to sum the probabilities corresponding to the possible sets of non-colliding outputs (see Definition 9):

$$P(Z_h > k) = \frac{(k!)S_k(\{d_1, \ldots, d_n\})}{d(d-1)\ldots(d-k+1)}.$$

But $S_k(\{d_1, \ldots, d_n\}) = d^k S_k(A)$, and $P(X_A > k) = (k!)S_k(A)$:

$$P(Z_h > k) = \frac{d^k P(X_A > k)}{d(d-1)\ldots(d-k+1)}.$$

But $P_{d,k} = d(d-1)\ldots(d-k+1)/d^k$, and the statement of this lemma follows. □

Lemma 18. *Consider the set of functions h of d inputs and at most n outputs ($n > 1$) such that $\beta(h) = \beta_0$ for some constant β_0. Let d_i be the number of inputs that map to R_i, the i^{th} output of h. Then $P(Z_h > k)$ is some function Q of k and $d_1, \ldots, d_n$. Let $\beta_1 = d\beta_0/(d-1+\beta_0)$. If we generalize Q to permit non-integer values for $d_1, \ldots, d_n$, then Q is a maximum across the set of functions h when $h = h_{\max}$, where*

$$d_1^{\max} = d(1 + \sqrt{((n/\beta_1) - 1)(n-1)})/n,$$
$$d_i^{\max} = d(1 - \sqrt{((n/\beta_1) - 1)/(n-1)})/n, \quad i = 2, \ldots, n,$$

and is a minimum when $h = h_{\min}$, where $m = \lceil \beta_1 \rceil$ and

$$d_1^{\min} = d(1 - \sqrt{((m/\beta_1) - 1)(m-1)})/m,$$
$$d_i^{\min} = d(1 + \sqrt{((m/\beta_1) - 1)/(m-1)})/m, \quad i = 2, \ldots, m,$$
$$d_i^{\min} = 0, \quad i = m+1, \ldots, n.$$

Proof. Consider the random variable A such that $P(A = R_i) = d_i/d$. Then

$$\beta(A) = \frac{1}{\sum_{i=1}^{n} \left(\frac{d_i}{d}\right)^2} = \frac{d}{\sum_{i=1}^{n} \left(\frac{d_i}{d} + \frac{d_i(d_i-1)}{d}\right)} = \frac{d}{1 + (d-1)\sum_{i=1}^{n} \frac{d_i(d_i-1)}{d(d-1)}}$$
$$= \frac{d}{(d-1)\frac{1}{\beta_0} + 1} = \frac{d\beta_0}{d - 1 + \beta_0} = \beta_1.$$

Because β_0 and d are constants, β_1 is also a constant. By Lemma 17, $P(Z_h > k) = P(X_A > k)/P_{d,k}$. Because d, k, and $\beta(A) = \beta_1$ are constants, minimizing or maximizing $P(X_A > k)$ is the same as minimizing or maximizing $P(Z_h > k)$. The statement of this lemma follows from Lemma 10. □

Lemma 19. *Let d, n, and k be positive integers, $d > n \geq k \geq 2$, β_0 a real number, $\beta_1(d) = d\beta_0/(d - 1 + \beta_0)$, $1 \leq \beta_0 \leq n(d-1)/(d-n)$, and $\alpha_d = \sqrt{((n/\beta_1(d)) - 1)/(n-1)}$. Then $p(d) = (1-\alpha_d)^{k-1}(1 + (k-1)\alpha_d)/P_{d,k}$ never increases as d increases.*

Proof. If $\beta_0 = 1$, then $p(d) = 0$ and this lemma is trivially true. Assume that $\beta_0 > 1$ for the rest of this proof. Consider the ratio $p(d)/p(d+1)$. We can apply Lemma 11 with d used in place of n, $u = \alpha_{d+1}$, $v = \alpha_d$, $g(u) = p(d+1)P_{d+1,k}$, and $g(v) = p(d)P_{d,k}$ because the required relationship $\alpha_d = \sqrt{1/d^2 + (1 - 1/d^2)\alpha_{d+1}^2}$ holds. (This is a little confusing because the n in this lemma has no connection with the n in Lemma 11.) Then $p(d)/p(d+1)$ is a minimum across varying β_0 when $\alpha_{d+1} = 0$ and $\alpha_d = 1/d$ and thus

$$\frac{p(d)}{p(d+1)} \geq \left(\frac{d-1}{d}\right)^{k-1} \left(\frac{d + (k-1)}{d}\right) \frac{P_{d+1,k}}{P_{d,k}}.$$

Because $P_{d,k} = (1 - 1/d)(1 - 2/d)\ldots(1 - (k-1)/d)$, we can show that

$$P_{d+1,k} = \left(\frac{d}{d-(k-1)}\right) \left(\frac{d}{d+1}\right)^{k-1} P_{d,k}.$$

Then

$$\frac{p(d)}{p(d+1)} \geq \left(\frac{d-1}{d+1}\right)^{k-1} \left(\frac{d+(k-1)}{d-(k-1)}\right).$$

The expression on the right hand side (with d in place of n) was shown in Lemma 12 to have a lower bound of 1. This means $p(d)/p(d+1) \geq 1$, and thus $p(d)$ never increases as d increases. □

Lemma 20. $(1 + 1/x)^x > e(1 - 1/(2x + 4/3))$ *for all positive real x.*

Proof. $(1 + 1/x)^x = e^{x\ln((x+1)/x)} = e^{-x\ln(x/(x+1))} = e^{-x\ln(1-u)}$, where $u = 1/(x+1)$. From the Taylor series of $-\ln(1-u)$, if $u > 0$ then $-\ln(1-u) > u + u^2/2 + u^3/3$. Thus $(1+1/x)^x > e^{1-v}$, where $v = (3x^2 + 7x + 6)/(6(x+1)^3)$. From the Taylor series of e^{-v}, if $v > 0$, then $e^{-v} > 1 - v + v^2/2 - v^3/6$. With some algebra, this gives $(1 + 1/x)^x > e(1 - 1/(2x + 4/3) + w)$, where

$$w = \frac{486x^8 + 2889x^7 + 7587x^6 + 11889x^5 + 12831x^4 + 10114x^3 + 5508x^2 + 1728x + 216}{1296(3x+2)(x+1)^9}.$$

Because w is clearly positive for $x > 0$, the lemma statement follows. □

Proof of Theorem 7 (see Section 5 for statement of theorem).
Let $\beta_1 = d\beta(h)/(d-1+\beta(h))$. $\beta_1 = \beta(A)$ for the random variable A corresponding to function h (see proof of Lemma 18).

Upper bound. If $k > n$, the upper bound is trivially true because $P(Z_h > k) = 0$. For the rest of the upper bound proof, assume $k \leq n$. Let $\alpha = \sqrt{((n/\beta_1)-1)/(n-1)})$. Let $x = (1-\alpha)/n$, $y = (1+(n-1)\alpha)/n$, and define function $h_{\max}$ as follows: $d_1^{\max} = y$, and $d_i^{\max} = x$ for $i = 2, \ldots, n$. By Lemma 18, $P(Z_h > k)$ is a maximum when $h = h_{\max}$. Using Lemma 17 and proceeding similarly to the upper bound proof of Theorem 3,

$$P(Z_{h_{\max}} > k) = (1-\alpha)^{k-1}(1+(k-1)\alpha)P_{n,k}/P_{d,k}.$$

By Lemma 19, $P(Z_{h_{\max}} > k)$ never increases as d increases, and is maximized when d is its minimum value ($d = n+1$). With some algebra, this gives $\beta_1 = (n+1)\beta(h)/(n+\beta(h))$, $\alpha = \sqrt{((n^2/\beta(h))-1)/(n^2-1)})$, $P_{n,k}/P_{d,k} = (1+1/n)^{k-1}(1-(k-1)/n)$, and thus

$$P(Z_{h_{\max}} > k) \leq (1-\alpha)^{k-1}(1+(k-1)\alpha)\left(1+\frac{1}{n}\right)^{k-1}\left(1-\frac{k-1}{n}\right). \tag{7}$$

The derivative of the right hand side of (7) with respect to n is $k(k-1)$ times itself times

$$\frac{1}{n(n+1)(n-(k-1))} - \frac{(1-1/\beta(h))n}{(1-\alpha)(1+(k-1)\alpha)(n^2-1)^2}.$$

This derivative is minimized by minimizing k. With a little algebra, one can show that when k is at its minimum of 2, the derivative goes to zero. Thus the derivative is never negative and the right hand side of inequality (7) is a maximum as $n \to \infty$. As $n \to \infty$, the last two factors become 1, $\lim_{n\to\infty} \alpha = 1/\sqrt{\beta(h)}$, and the inequality becomes

$$P(Z_h > k) \leq \left(1-\frac{1}{\sqrt{\beta(h)}}\right)^{k-1}\left(1+\frac{k-1}{\sqrt{\beta(h)}}\right) \leq e^{-\frac{k-1}{\sqrt{\beta(h)}}}\left(1+\frac{k-1}{\sqrt{\beta(h)}}\right).$$

The second upper bound is based on $e^{-x} > 1-x$ for $x > 0$.

Lower bound. The lower bound is trivially true if $k < 2$ because $P(Z_h > 0) = P(Z_h > 1) = 1$. For the rest of this proof, assume $k \geq 2$. Let $m = \lceil \beta_1 \rceil$, and redefine α, x, and y as follows: $\alpha = \sqrt{((m/\beta_1)-1)/(m-1)})$, $x = (1-(m-1)\alpha)/m$, and $y = (1+\alpha)/m$. Because $x \geq 0$, we have $\alpha \leq 1/(m-1)$. Define $h_{\min}$ as follows: $d_1^{\min} = x$, $d_i^{\min} = y$ for $i = 2, \ldots, m$, and $d_i^{\min} = 0$ for $i = m+1, \ldots, n$. By Lemma 18, $P(Z_h > k)$ is a minimum when $h = h_{\min}$. Using Lemma 17 and proceeding similarly to the lower bound proof of Theorem 3,

$$P(Z_h > k) \geq P_{\beta_1,k}/P_{d,k}.$$

Let $g(d)$ be the right hand side of the last inequality:

$$g(d) = \frac{P_{\beta_1,k}}{P_{d,k}} = \prod_{i=0}^{k-1}\left(\frac{1-\frac{i}{\beta_1}}{1-\frac{i}{d}}\right).$$

Recalling that $\beta_1 = d\beta(h)/(d-1+\beta(h))$, and taking the derivative of $g(d)$, we find

$$\frac{\partial g(d)}{\partial d} = \frac{(d-\beta_1)g(d)}{d(d-1)}\sum_{i=0}^{k-1}\frac{i(i-1)}{(\beta_1-i)(d-i)}.$$

This derivative is nonnegative for $d > \beta_1$ and $k < \beta_1 + 1$. However, $d \geq n + 1$, $n \geq m$, $m \geq \beta_1$, and thus $d \geq \beta_1 + 1$. Therefore, $g(\beta_1 + 1)$ is a lower bound on $g(d)$:

$$P(Z_h > k) \geq \frac{P_{\beta_1,k}}{P_{\beta_1+1,k}} = \left(1 + \frac{1}{\beta_1}\right)^{k-1} \left(1 - \frac{k-1}{\beta_1}\right).$$

Recall that this inequality is only valid if $k - 1 < \beta_1$. For $k > m$, $P(Z_h > k) = 0$. This occurs because there might exist a function h with only m outputs that has the required $\beta(h)$ and thus must have a collision after selecting $m + 1$ inputs. Solving $\beta_1 = d\beta(h)/(d - 1 + \beta(h))$ with $d = \beta_1 + 1$ gives $\beta_1 = \sqrt{\beta(h)}$. Replacing β_1 with $\sqrt{\beta(h)}$ in the last inequality gives the first lower bound in the theorem statement. The second lower bound comes from applying Lemma 20 with $x = \sqrt{\beta(h)}$. □

Proof of Theorem 8 (see Section 5 for statement of theorem).
For the $n = 1$ case this theorem is true because the collision always occurs on the second input, and $\beta(h) = 1$. For the rest of this proof assume $n > 1$.

Upper bound. Use the probability upper bound in Theorem 7 and proceed in the same manner as in the proof of Theorem 4 to show that $E(Z_h) \leq 2\sqrt{\beta(h)}$.

Lower bound. Begin with the expectation equation in Lemma 2 and use the probability lower bound in Theorem 7. Let $m = \lceil \sqrt{\beta(h)} \rceil$, $b = \sqrt{\beta(h)}$, and $\delta = m - b$, $0 \leq \delta < 1$. Because the lower bound is zero for $k > m$, the sum must be stopped at m:

$$\begin{aligned} E(Z_h) &= \sum_{k=0}^{\infty} P(Z_h > k) \geq 1 + \sum_{k=1}^{m} P(Z_h > k) \\ &\geq 1 + \sum_{k=0}^{m-1} \left(1 + \frac{1}{b}\right)^{k} \left(1 - \frac{k}{b}\right). \end{aligned}$$

Using the identities

$$\sum_{k=0}^{m-1} x^k = \frac{x^m - 1}{x - 1}, \qquad \sum_{k=0}^{m-1} k x^k = \frac{x^m(m(x-1) - x) + x}{(x-1)^2}$$

with $x = 1 + 1/b$ and using some algebra, we get $E(Z_h) \geq (1 + 1/b)^{b+\delta}(b - \delta + 1) - 2b$. Because $\delta < 1$, we have

$$\left(1 + \frac{1}{b}\right)^{\delta} = \left(\frac{1}{1 - \frac{1}{b+1}}\right)^{\delta} \geq \frac{1}{1 - \frac{\delta}{b+1}} = \frac{b+1}{b - \delta + 1},$$

which gives $E(Z_h) \geq (1 + 1/b)^b(b + 1) - 2b$. By Lemma 20, $(1 + 1/b)^b > e(1 - 1/(2b + 4/3))$. With some algebra this gives

$$E(Z_h) > (e - 2)b + \frac{e(3b + 1)}{2(3b + 2)}.$$

The final term is always positive, and thus $E(Z_h) > (e - 2)\sqrt{\beta(h)}$. □