

Birthday Problem

Assume a hash function H that pretty much randomly maps an integer input to an integer output. Suppose the number of output values for H is k . Pick n input integers randomly. How large should n be so that the probability that at least one pair of input integers map to the same output is $1/2$?

Birthday Problem

Assume a hash function H that pretty much randomly maps an integer input to an integer output. Suppose the number of output values for H is k . Pick n input integers randomly. How large should n be so that the probability that at least one pair of input integers map to the same output is $1/2$?

Let a_1, a_2, \dots, a_n be n input random numbers

Birthday Problem

Assume a hash function H that pretty much randomly maps an integer input to an integer output. Suppose the number of output values for H is k . Pick n input integers randomly. How large should n be so that the probability that at least one pair of input integers map to the same output is $1/2$?

Let a_1, a_2, \dots, a_n be n input random numbers

Define $X_{i,j} = 1$ iff $H(a_i) = H(a_j)$ $X_{i,j} = 0$ otherwise

Birthday Problem

Assume a hash function H that pretty much randomly maps an integer input to an integer output. Suppose the number of output values for H is k . Pick n input integers randomly. How large should n be so that the probability that at least one pair of input integers map to the same output is $1/2$?

Let a_1, a_2, \dots, a_n be n input random numbers

Define $X_{i,j} = 1$ iff $H(a_i) = H(a_j)$ $X_{i,j} = 0$ otherwise

Define $X = \sum_{i < j} X_{i,j}$ (X could be as high as $n(n-1)/2$)

Birthday Problem

Assume a hash function H that pretty much randomly maps an integer input to an integer output. Suppose the number of output values for H is k . Pick n input integers randomly. How large should n be so that the probability that at least one pair of input integers map to the same output is $1/2$?

Let a_1, a_2, \dots, a_n be n input random numbers

Define $X_{i,j} = 1$ iff $H(a_i) = H(a_j)$ $X_{i,j} = 0$ otherwise

Define $X = \sum_{i < j} X_{i,j}$ (X could be as high as $n(n-1)/2$)

We want to determine $Pr(X > 0)$

Birthday Problem

$$E\{X\} = Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} \quad \text{By definition}$$

Birthday Problem

$$\begin{aligned} E\{X\} &= Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} && \text{By definition} \\ &= Pr(X>0)E\{X|X>0\} \end{aligned}$$

Birthday Problem

$$\begin{aligned} E\{X\} &= Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} && \text{By definition} \\ &= Pr(X>0)E\{X|X>0\} \end{aligned}$$

$$\text{Therefore, } Pr(X>0) = E\{X\} / E\{X|X>0\}$$

Birthday Problem

$$\begin{aligned} E\{X\} &= Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} && \text{By definition} \\ &= Pr(X>0)E\{X|X>0\} \end{aligned}$$

$$\text{Therefore, } Pr(X>0) = E\{X\} / E\{X|X>0\}$$

Find $E\{X\}$ first:

$$E\{X\} = \sum E\{X_{i,j}\} \quad \text{but } E\{X_{i,j}\} = Pr(X_{i,j} = 1) = k/k^2 = 1/k$$

so

$$E\{X\} = n(n-1)/(2k)$$

Birthday Problem

$$\begin{aligned} E\{X\} &= Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} && \text{By definition} \\ &= Pr(X>0)E\{X|X>0\} \end{aligned}$$

$$\text{Therefore, } Pr(X>0) = E\{X\} / E\{X|X>0\}$$

Find $E\{X\}$ first:

$$E\{X\} = \sum E\{X_{i,j}\} \quad \text{but } E\{X_{i,j}\} = Pr(X_{i,j} = 1) = k/k^2 = 1/k$$

so

$$E\{X\} = n(n-1)/(2k)$$

Now find $E\{X|X>0\}$: (must be two inputs that map to same number)

Birthday Problem

$$\begin{aligned} E\{X\} &= Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} && \text{By definition} \\ &= Pr(X>0)E\{X|X>0\} \end{aligned}$$

$$\text{Therefore, } Pr(X>0) = E\{X\} / E\{X|X>0\}$$

Find $E\{X\}$ first:

$$E\{X\} = \sum E\{X_{i,j}\} \quad \text{but } E\{X_{i,j}\} = Pr(X_{i,j} = 1) = k/k^2 = 1/k$$

so

$$E\{X\} = n(n-1)/(2k)$$

Now find $E\{X|X>0\}$: (must be two inputs that map to same number)

Let P be the event that a_n and a_{n-1} map to the same number

Birthday Problem

$$\begin{aligned} E\{X\} &= Pr(X=0)E\{X|X=0\} + Pr(X>0)E\{X|X>0\} \quad \text{By definition} \\ &= Pr(X>0)E\{X|X>0\} \end{aligned}$$

$$\text{Therefore, } Pr(X>0) = E\{X\} / E\{X|X>0\}$$

Find $E\{X\}$ first:

$$E\{X\} = \sum E\{X_{i,j}\} \quad \text{but } E\{X_{i,j}\} = Pr(X_{i,j} = 1) = k/k^2 = 1/k$$

so

$$E\{X\} = n(n-1)/(2k)$$

Now find $E\{X|X>0\}$: (must be two inputs that map to same number)

Let P be the event that a_n and a_{n-1} map to the same number

$$\text{Then } Pr(X>0) = E\{X\} / E\{X|P\}$$

Birthday Problem

$$E\{X|P\} = \sum_{i < j < n-1} E\{X_{i,j}|P\} + \sum_{j=n-1 \text{ or } j=n} E\{X_{i,j}|P\}$$

Birthday Problem

$$E\{X|P\} = \sum_{i < j < n-1} E\{X_{i,j}|P\} + \sum_{j=n-1 \text{ or } j=n} E\{X_{i,j}|P\}$$

Observe that mapping of a_i and a_j are independent given P

Birthday Problem

$$E\{X|P\} = \sum_{i < j < n-1} E\{X_{i,j}|P\} + \sum_{j=n-1 \text{ or } j=n} E\{X_{i,j}|P\}$$

Observe that mapping of a_i and a_j are independent given P

Hence, for $i < j < n-1$,

$$E\{X_{i,j}|P\} = 1/k \text{ and } \sum E\{X_{i,j}|P\} = (n-2)(n-3)/(2k)$$

Birthday Problem

$$E\{X|P\} = \sum_{i < j < n-1} E\{X_{i,j}|P\} + \sum_{j=n-1 \text{ or } j=n} E\{X_{i,j}|P\}$$

Observe that mapping of a_i and a_j are independent given P

Hence, for $i < j < n-1$,

$$E\{X_{i,j}|P\} = 1/k \text{ and } \sum E\{X_{i,j}|P\} = (n-2)(n-3)/(2k)$$

And for $j=n-1$ or $j=n$, except $i=n-1$ and $j=n$

$$E\{X_{i,j}|P\} = 2/(k-2) \text{ and } \sum E\{X_{i,j}|P\} = 2(n-2)/(k-2)$$

Birthday Problem

$$E\{X|P\} = \sum_{i < j < n-1} E\{X_{i,j}|P\} + \sum_{j=n-1 \text{ or } j=n} E\{X_{i,j}|P\}$$

Observe that mapping of a_i and a_j are independent given P

Hence, for $i < j < n-1$,

$$E\{X_{i,j}|P\} = 1/k \text{ and } \sum E\{X_{i,j}|P\} = (n-2)(n-3)/(2k)$$

And for $j=n-1$ or $j=n$, except $i=n-1$ and $j=n$

$$E\{X_{i,j}|P\} = 2/(k-2) \text{ and } \sum E\{X_{i,j}|P\} = 2(n-2)/(k-2)$$

Therefore,

$$\begin{aligned} Pr(X > 0) &> n(n-1)/(2k) / ((n-2)(n-3)/(2k) + 2(n-2)/(k-2) + 1) \\ &\approx n^2/(n^2 + 2k) \end{aligned}$$

Birthday Problem

$$E\{X|P\} = \sum_{i < j < n-1} E\{X_{i,j}|P\} + \sum_{j=n-1 \text{ or } j=n} E\{X_{i,j}|P\}$$

Observe that mapping of a_i and a_j are independent given P

Hence, for $i < j < n-1$,

$$E\{X_{i,j}|P\} = 1/k \text{ and } \sum E\{X_{i,j}|P\} = (n-2)(n-3)/(2k)$$

And for $j=n-1$ or $j=n$, except $i=n-1$ and $j=n$

$$E\{X_{i,j}|P\} = 2/(k-2) \text{ and } \sum E\{X_{i,j}|P\} = 2(n-2)/(k-2)$$

Therefore,

$$\begin{aligned} Pr(X > 0) &> n(n-1)/(2k) / ((n-2)(n-3)/(2k) + 2(n-2)/(k-2) + 1) \\ &\approx n^2/(n^2 + 2k) \end{aligned}$$

which is $1/2$ when $2k = n^2$