

Birthday bounds

Theorem 1. (classical birthday bound).

Let $y_1, \dots, y_q$ be uniformly and independently drawn from a size- N set. Then

$$1 - e^{-q(q-1)/2N} \leq \Pr[\exists i \neq j, y_i = y_j] \leq \frac{q(q-1)}{2N}.$$

If furthermore $q \leq \sqrt{2N}$, the lower bound is $\geq \frac{q(q-1)}{4N}$.

Proof of the upper bound. The probability that two independent elements drawn from a size- N set are equal is at most $1/N$. Therefore, using the union bound, $\Pr[\exists i \neq j, y_i = y_j] \leq \sum_{i \neq j} 1/N = q(q-1)/2N$. $\square$

Proof of the lower bound. Let N_i be the event “there is no collision amongst $y_1, \dots, y_i$.” We want to lower bound $\Pr[\neg N_q]$, that is to upper bound $\Pr[N_q]$. Using the law of total probability, $\Pr[N_q] = \Pr[N_q|N_{q-1}]\Pr[N_{q-1}] + \Pr[N_q|\neg N_{q-1}]\Pr[\neg N_{q-1}]$. But $\Pr[N_q|N_{q-1}] = 0$. Therefore, $\Pr[N_q] = \Pr[N_q|\neg N_{q-1}]\Pr[\neg N_{q-1}]$. Hence, by induction

$$\Pr[N_q] = \Pr[N_1] \cdot \Pr[N_2|N_1] \cdots \Pr[N_q|N_{q-1}].$$

Now, $\Pr[N_1] = 1$ (since there cannot be a collision between one element). And for $i > 1$, $\Pr[N_i|N_{i-1}]$ is the probability that y_i collides with a y_j , $j < i$, assuming they are all different. Since y_i is drawn independently of the previous y_j 's, $\Pr[N_i|N_{i-1}] = 1 - (i-1)/N$.

Together, we obtain that $\Pr[N_q] = \prod_{i=1}^{q-1} 1 - i/N$. Since $1 + x \leq e^x$ for all x ,

$$\Pr[N_q] \leq \prod_{i=1}^{q-1} e^{-i/N} = e^{-\frac{1}{N} \sum_{i=1}^{q-1} i} = e^{-q(q-1)/2N}.$$

Hence $\Pr[\exists i \neq j, y_i = y_j] \geq 1 - e^{-q(q-1)/2N}$. Finally, if $q \leq \sqrt{2N}$, $q(q-1)/2N \leq 1$ and since $e^{-x} \leq 1 - x/2$ for $0 \leq x \leq 1$, the probability is at least $q(q-1)/4N$. $\square$

Theorem 2. (variant of the birthday bound).

Let $y_1, \dots, y_q, z_1, \dots, z_q$ be uniformly and independently drawn from a size- N set. Then

$$1 - e^{-q^2/N} \leq \Pr[\exists i, j, y_i = z_j] \leq \frac{q^2}{N}.$$

If furthermore $q \leq \sqrt{N}$, the lower bound is $\geq \frac{q^2}{2N}$.

Proof of the upper bound. The probability that two independent elements drawn from a size- N set are equal is at most $1/N$. Therefore, using the union bound, $\Pr[\exists i, j, y_i = z_j] \leq \sum_{i,j} 1/N = q^2/N$. $\square$

Proof of the lower bound. Let N_i be the event “for all j , $y_i \neq z_j$ ”. We want to lower bound $\Pr[\bigvee_i \neg N_i]$, that is to upper bound $\Pr[\bigwedge_i N_i]$. Assume that we first draw all the values z_j , $1 \leq j \leq q$. Then N_i is the event “ $y_i \notin Z$ ” where $Z = \{z_1, \dots, z_q\}$. Since the y_i ’s are drawn independently, the N_i ’s are independent events. Therefore, $\Pr[\bigwedge_i N_i] = \prod_i \Pr[N_i]$. Now, $N_i = \bigvee_j [y_i = z_j]$. Assume now on the contrary that we first draw y_i , then the z_j ’s independently: We see that the events $[y_i = z_j]$ are independent. That is $\Pr[N_i] = \prod_j \Pr[y_i = z_j] = (1 - 1/N)^q$. Now using $1 + x \leq e^x$, $\Pr[N_i] \leq e^{-q/N}$ and $\Pr[\bigwedge_i N_i] \leq e^{-q^2/N}$. Therefore, $\Pr[\exists i, j, y_i = z_j] \geq 1 - e^{-q^2/N}$. If $q \leq \sqrt{N}$, $q^2/N \leq 1$ and since $e^{-x} \leq 1 - x/2$ for $0 \leq x \leq 1$, the probability is at least $q^2/2N$. $\square$