

Hashing Problem

Probability of Collisions vs. No Collisions → C vs. no C

Birthday Problem Analogy

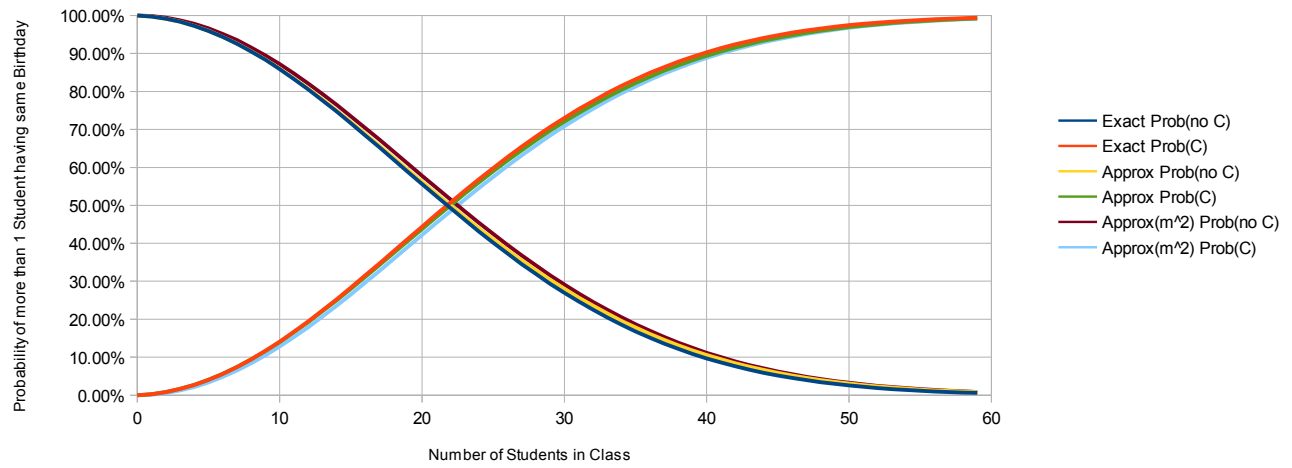
Given a room full of students, what is the probability that 2 or more have the same Birthday

365 = Size of Bit Vector, i.e. Number of days in the year

m	PI(i=0,m-1->(1-m/n))			e(-m/n)	EXP(-m(m-1)/2n)		EXP(-m^2/2n)	
	(1-m/n)	Exact Prob(no C)	Exact Prob(C)		Approx Prob(no C)	Approx Prob(C)	Approx(m^2) Prob(no C)	Approx(m^2) Prob(C)
0	100.00%	100.00%	0.00%	100.00%	100.00%	0.00%	100.00%	0.00%
1	99.73%	99.73%	0.27%	99.73%	99.73%	0.27%	99.86%	0.14%
2	99.45%	99.18%	0.82%	99.45%	99.18%	0.82%	99.45%	0.55%
3	99.18%	98.36%	1.64%	99.18%	98.37%	1.63%	98.77%	1.23%
4	98.90%	97.29%	2.71%	98.91%	97.30%	2.70%	97.83%	2.17%
5	98.63%	95.95%	4.05%	98.64%	95.97%	4.03%	96.63%	3.37%
6	98.36%	94.38%	5.62%	98.37%	94.41%	5.59%	95.19%	4.81%
7	98.08%	92.57%	7.43%	98.10%	92.62%	7.38%	93.51%	6.49%
8	97.81%	90.54%	9.46%	97.83%	90.61%	9.39%	91.61%	8.39%
9	97.53%	88.31%	11.69%	97.56%	88.40%	11.60%	89.50%	10.50%
10	97.26%	85.89%	14.11%	97.30%	86.01%	13.99%	87.20%	12.80%
11	96.99%	83.30%	16.70%	97.03%	83.46%	16.54%	84.73%	15.27%
12	96.71%	80.56%	19.44%	96.77%	80.76%	19.24%	82.10%	17.90%
13	96.44%	77.69%	22.31%	96.50%	77.93%	22.07%	79.33%	20.67%
14	96.16%	74.71%	25.29%	96.24%	75.00%	25.00%	76.45%	23.55%
15	95.89%	71.64%	28.36%	95.97%	71.98%	28.02%	73.48%	26.52%
16	95.62%	68.50%	31.50%	95.71%	68.89%	31.11%	70.42%	29.58%
17	95.34%	65.31%	34.69%	95.45%	65.76%	34.24%	67.31%	32.69%
18	95.07%	62.09%	37.91%	95.19%	62.59%	37.41%	64.16%	35.84%
19	94.79%	58.86%	41.14%	94.93%	59.42%	40.58%	60.99%	39.01%
20	94.52%	55.63%	44.37%	94.67%	56.25%	43.75%	57.81%	42.19%
21	94.25%	52.43%	47.57%	94.41%	53.11%	46.89%	54.66%	45.34%
22	93.97%	49.27%	50.73%	94.15%	50.00%	50.00%	51.53%	48.47%
23	93.70%	46.17%	53.83%	93.89%	46.95%	53.05%	48.45%	51.55%
24	93.42%	43.13%	56.87%	93.64%	43.96%	56.04%	45.43%	54.57%
25	93.15%	40.18%	59.82%	93.38%	41.05%	58.95%	42.48%	57.52%
26	92.88%	37.31%	62.69%	93.12%	38.23%	61.77%	39.61%	60.39%
27	92.60%	34.55%	65.45%	92.87%	35.50%	64.50%	36.84%	63.16%
28	92.33%	31.90%	68.10%	92.62%	32.88%	67.12%	34.16%	65.84%
29	92.05%	29.37%	70.63%	92.36%	30.37%	69.63%	31.60%	68.40%
30	91.78%	26.95%	73.05%	92.11%	27.97%	72.03%	29.15%	70.85%
31	91.51%	24.67%	75.33%	91.86%	25.69%	74.31%	26.81%	73.19%
32	91.23%	22.50%	77.50%	91.61%	23.54%	76.46%	24.59%	75.41%
33	90.96%	20.47%	79.53%	91.36%	21.50%	78.50%	22.50%	77.50%
34	90.68%	18.56%	81.44%	91.11%	19.59%	80.41%	20.52%	79.48%
35	90.41%	16.78%	83.22%	90.86%	17.80%	82.20%	18.67%	81.33%
36	90.14%	15.13%	84.87%	90.61%	16.13%	83.87%	16.94%	83.06%
37	89.86%	13.59%	86.41%	90.36%	14.57%	85.43%	15.33%	84.67%
38	89.59%	12.18%	87.82%	90.11%	13.13%	86.87%	13.83%	86.17%
39	89.32%	10.88%	89.12%	89.87%	11.80%	88.20%	12.45%	87.55%
40	89.04%	9.68%	90.32%	89.62%	10.58%	89.42%	11.17%	88.83%
41	88.77%	8.60%	91.40%	89.38%	9.45%	90.55%	10.00%	90.00%
42	88.49%	7.61%	92.39%	89.13%	8.42%	91.58%	8.92%	91.08%
43	88.22%	6.71%	93.29%	88.89%	7.49%	92.51%	7.94%	92.06%
44	87.95%	5.90%	94.10%	88.64%	6.64%	93.36%	7.05%	92.95%
45	87.67%	5.17%	94.83%	88.40%	5.87%	94.13%	6.24%	93.76%
46	87.40%	4.52%	95.48%	88.16%	5.17%	94.83%	5.51%	94.49%
47	87.12%	3.94%	96.06%	87.92%	4.55%	95.45%	4.85%	95.15%
48	86.85%	3.42%	96.58%	87.68%	3.99%	96.01%	4.26%	95.74%
49	86.58%	2.96%	97.04%	87.44%	3.49%	96.51%	3.73%	96.27%
50	86.30%	2.56%	97.44%	87.20%	3.04%	96.96%	3.26%	96.74%
51	86.03%	2.20%	97.80%	86.96%	2.64%	97.36%	2.84%	97.16%
52	85.75%	1.89%	98.11%	86.72%	2.29%	97.71%	2.46%	97.54%
53	85.48%	1.61%	98.39%	86.48%	1.98%	98.02%	2.13%	97.87%
54	85.21%	1.37%	98.63%	86.25%	1.71%	98.29%	1.84%	98.16%
55	84.93%	1.17%	98.83%	86.01%	1.47%	98.53%	1.59%	98.41%
56	84.66%	0.99%	99.01%	85.78%	1.26%	98.74%	1.36%	98.64%
57	84.38%	0.83%	99.17%	85.54%	1.08%	98.92%	1.17%	98.83%
58	84.11%	0.70%	99.30%	85.31%	0.92%	99.08%	1.00%	99.00%
59	83.84%	0.59%	99.41%	85.07%	0.78%	99.22%	0.85%	99.15%
60	83.56%	0.49%	99.51%	84.84%	0.66%	99.34%	0.72%	99.28%

Hashing Collision Probabilites

i.e. Birthday Problem Analogy



Hashing Problem

Probability of Collisions vs. No Collisions → C vs. no C

Birthday Problem Analogy

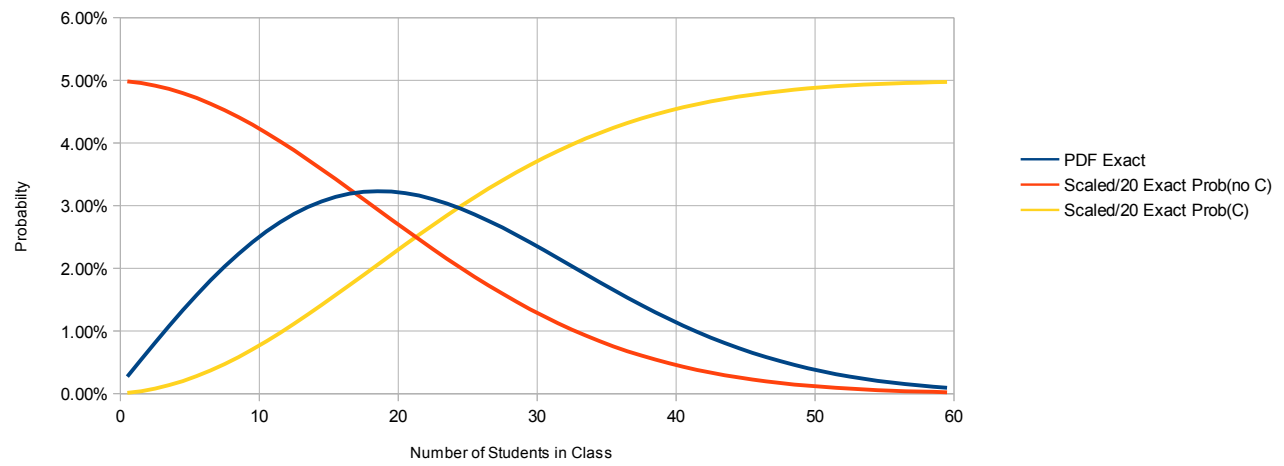
Given a room full of students, what is the probability that 2 or more have the same Birthday

365 = Size of Bit Vector, i.e. Number of days in the year

m	(1-m/n)	P(i=0,m-1→(1-m/n))		diff m	PDF Exact Prob(no C)	PDF Exact Prob(C)	Scaled/20 Exact Prob(no C)	Scaled/20 Exact Prob(C)
		Exact Prob(no C)	Exact Prob(C)					
0	100.00%	100.00%	0.00%				5.00%	0.00%
1	99.73%	99.73%	0.27%	0.5	0.27%	0.27%	4.99%	0.01%
2	99.45%	99.18%	0.82%	1.5	0.55%	0.55%	4.96%	0.04%
3	99.18%	98.36%	1.64%	2.5	0.82%	0.82%	4.92%	0.08%
4	98.90%	97.29%	2.71%	3.5	1.08%	1.08%	4.86%	0.14%
5	98.63%	95.95%	4.05%	4.5	1.33%	1.33%	4.80%	0.20%
6	98.36%	94.38%	5.62%	5.5	1.58%	1.58%	4.72%	0.28%
7	98.08%	92.57%	7.43%	6.5	1.81%	1.81%	4.63%	0.37%
8	97.81%	90.54%	9.46%	7.5	2.03%	2.03%	4.53%	0.47%
9	97.53%	88.31%	11.69%	8.5	2.23%	2.23%	4.42%	0.58%
10	97.26%	85.89%	14.11%	9.5	2.42%	2.42%	4.29%	0.71%
11	96.99%	83.30%	16.70%	10.5	2.59%	2.59%	4.16%	0.84%
12	96.71%	80.56%	19.44%	11.5	2.74%	2.74%	4.03%	0.97%
13	96.44%	77.69%	22.31%	12.5	2.87%	2.87%	3.88%	1.12%
14	96.16%	74.71%	25.29%	13.5	2.98%	2.98%	3.74%	1.26%
15	95.89%	71.64%	28.36%	14.5	3.07%	3.07%	3.58%	1.42%
16	95.62%	68.50%	31.50%	15.5	3.14%	3.14%	3.42%	1.58%
17	95.34%	65.31%	34.69%	16.5	3.19%	3.19%	3.27%	1.73%
18	95.07%	62.09%	37.91%	17.5	3.22%	3.22%	3.10%	1.90%
19	94.79%	58.86%	41.14%	18.5	3.23%	3.23%	2.94%	2.06%
20	94.52%	55.63%	44.37%	19.5	3.22%	3.22%	2.78%	2.22%
21	94.25%	52.43%	47.57%	20.5	3.20%	3.20%	2.62%	2.38%
22	93.97%	49.27%	50.73%	21.5	3.16%	3.16%	2.46%	2.54%
23	93.70%	46.17%	53.83%	22.5	3.10%	3.10%	2.31%	2.69%
24	93.42%	43.13%	56.87%	23.5	3.04%	3.04%	2.16%	2.84%
25	93.15%	40.18%	59.82%	24.5	2.95%	2.95%	2.01%	2.99%
26	92.88%	37.31%	62.69%	25.5	2.86%	2.86%	1.87%	3.13%
27	92.60%	34.55%	65.45%	26.5	2.76%	2.76%	1.73%	3.27%
28	92.33%	31.90%	68.10%	27.5	2.65%	2.65%	1.60%	3.40%
29	92.05%	29.37%	70.63%	28.5	2.53%	2.53%	1.47%	3.53%
30	91.78%	26.95%	73.05%	29.5	2.41%	2.41%	1.35%	3.65%
31	91.51%	24.67%	75.33%	30.5	2.29%	2.29%	1.23%	3.77%
32	91.23%	22.50%	77.50%	31.5	2.16%	2.16%	1.13%	3.87%
33	90.96%	20.47%	79.53%	32.5	2.03%	2.03%	1.02%	3.98%
34	90.68%	18.56%	81.44%	33.5	1.91%	1.91%	0.93%	4.07%
35	90.41%	16.78%	83.22%	34.5	1.78%	1.78%	0.84%	4.16%
36	90.14%	15.13%	84.87%	35.5	1.66%	1.66%	0.76%	4.24%
37	89.86%	13.59%	86.41%	36.5	1.53%	1.53%	0.68%	4.32%
38	89.59%	12.18%	87.82%	37.5	1.42%	1.42%	0.61%	4.39%
39	89.32%	10.88%	89.12%	38.5	1.30%	1.30%	0.54%	4.46%
40	89.04%	9.68%	90.32%	39.5	1.19%	1.19%	0.48%	4.52%
41	88.77%	8.60%	91.40%	40.5	1.09%	1.09%	0.43%	4.57%
42	88.49%	7.61%	92.39%	41.5	0.99%	0.99%	0.38%	4.62%
43	88.22%	6.71%	93.29%	42.5	0.90%	0.90%	0.34%	4.66%
44	87.95%	5.90%	94.10%	43.5	0.81%	0.81%	0.30%	4.70%
45	87.67%	5.17%	94.83%	44.5	0.73%	0.73%	0.26%	4.74%
46	87.40%	4.52%	95.48%	45.5	0.65%	0.65%	0.23%	4.77%
47	87.12%	3.94%	96.06%	46.5	0.58%	0.58%	0.20%	4.80%
48	86.85%	3.42%	96.58%	47.5	0.52%	0.52%	0.17%	4.83%
49	86.58%	2.96%	97.04%	48.5	0.46%	0.46%	0.15%	4.85%
50	86.30%	2.56%	97.44%	49.5	0.41%	0.41%	0.13%	4.87%
51	86.03%	2.20%	97.80%	50.5	0.36%	0.36%	0.11%	4.89%
52	85.75%	1.89%	98.11%	51.5	0.31%	0.31%	0.09%	4.91%
53	85.48%	1.61%	98.39%	52.5	0.27%	0.27%	0.08%	4.92%
54	85.21%	1.37%	98.63%	53.5	0.24%	0.24%	0.07%	4.93%
55	84.93%	1.17%	98.83%	54.5	0.21%	0.21%	0.06%	4.94%
56	84.66%	0.99%	99.01%	55.5	0.18%	0.18%	0.05%	4.95%
57	84.38%	0.83%	99.17%	56.5	0.15%	0.15%	0.04%	4.96%
58	84.11%	0.70%	99.30%	57.5	0.13%	0.13%	0.04%	4.96%
59	83.84%	0.59%	99.41%	58.5	0.11%	0.11%	0.03%	4.97%
60	83.56%	0.49%	99.51%	59.5	0.10%	0.10%	0.02%	4.98%

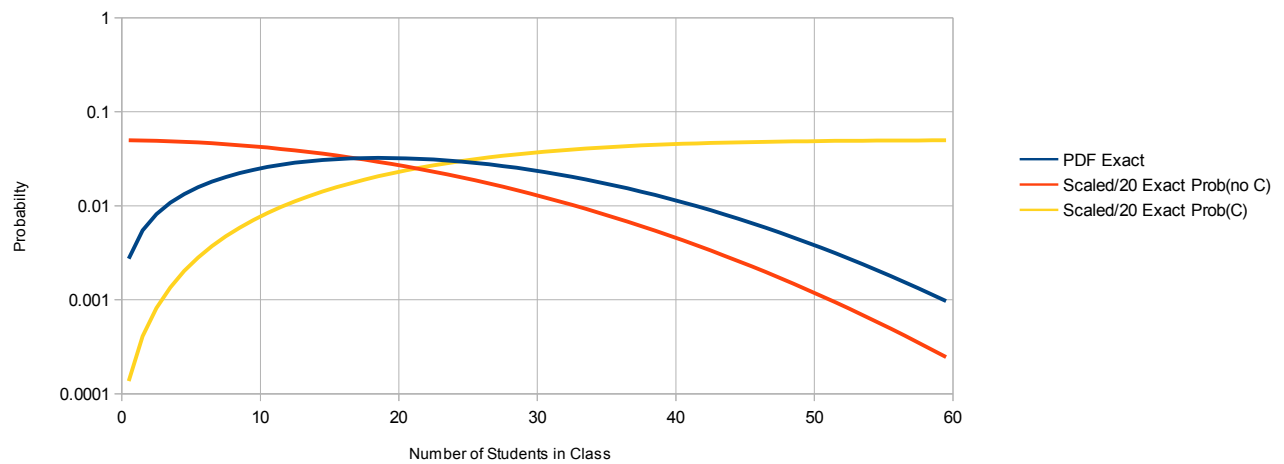
Hashing Collision Problem

Probability Density Function vs. Cumulative



Hashing Collision Problem

Probability Density Function vs. Cumulative



Hashing Problem

Probability of Collisions vs. No Collisions → C vs. no C

Birthday Problem Analogy

Given a room full of students, what is the probability that 2 or more have the same Birthday

365 = Size of Bit Vector, i.e. Number of days in the year

m	Exact (1-1/n)^m	Expected Value	Number of Exp C
0	1.0000	0.0000	0.0000
1	0.9973	1.0000	0.0000
2	0.9945	1.9973	0.0027
3	0.9918	2.9918	0.0082
4	0.9891	3.9836	0.0164
5	0.9864	4.9727	0.0273
6	0.9837	5.9591	0.0409
7	0.9810	6.9427	0.0573
8	0.9783	7.9237	0.0763
9	0.9756	8.9020	0.0980
10	0.9729	9.8776	0.1224
11	0.9703	10.8505	0.1495
12	0.9676	11.8208	0.1792
13	0.9650	12.7884	0.2116
14	0.9623	13.7534	0.2466
15	0.9597	14.7157	0.2843
16	0.9571	15.6754	0.3246
17	0.9544	16.6325	0.3675
18	0.9518	17.5869	0.4131
19	0.9492	18.5387	0.4613
20	0.9466	19.4879	0.5121
21	0.9440	20.4345	0.5655
22	0.9414	21.3785	0.6215
23	0.9388	22.3200	0.6800
24	0.9363	23.2588	0.7412
25	0.9337	24.1951	0.8049
26	0.9312	25.1288	0.8712
27	0.9286	26.0600	0.9400
28	0.9261	26.9886	1.0114
29	0.9235	27.9146	1.0854
30	0.9210	28.8381	1.1619
31	0.9185	29.7591	1.2409
32	0.9160	30.6776	1.3224
33	0.9134	31.5935	1.4065
34	0.9109	32.5070	1.4930
35	0.9084	33.4179	1.5821
36	0.9060	34.3264	1.6736
37	0.9035	35.2323	1.7677
38	0.9010	36.1358	1.8642
39	0.8985	37.0368	1.9632
40	0.8961	37.9353	2.0647
41	0.8936	38.8314	2.1686
42	0.8912	39.7250	2.2750
43	0.8887	40.6162	2.3838
44	0.8863	41.5049	2.4951
45	0.8839	42.3912	2.6088
46	0.8814	43.2750	2.7250
47	0.8790	44.1565	2.8435
48	0.8766	45.0355	2.9645
49	0.8742	45.9121	3.0879
50	0.8718	46.7863	3.2137
51	0.8694	47.6582	3.3418
52	0.8670	48.5276	3.4724
53	0.8647	49.3946	3.6054
54	0.8623	50.2593	3.7407
55	0.8599	51.1216	3.8784
56	0.8576	51.9815	4.0185
57	0.8552	52.8391	4.1609
58	0.8529	53.6944	4.3056
59	0.8506	54.5473	4.4527
60	0.8482	55.3978	4.6022

