Finding heavy hitters by chaining with few random bits

Stephen Chestnut

June 23, 2016

Joint work with
Vladimir Braverman, Nikita Ivkin, Jelani Nelson, Zhengyu Wang, David Woodruff

Outline

- What are heavy hitters and data streams?
- Well known ingredients for streaming algorithms
- Add chaining
- Limiting the number of random bits

Stream: $p_1, p_2, \ldots, p_m \in [n]$

$$f_i = \#\{j \leq m : p_j = i\}$$

$$f_i^t = \#\{j \leq t : p_j = i\}$$

Stream: $p_1, p_2, \ldots, p_m \in [n]$

$$f_i = \#\{j \leq m : p_j = i\}$$

$$f_i^t = \#\{j \leq t : p_j = i\}$$

- Read items one-at-a-time
- Approximate answer
- Succeed with probability $2/3$

Stream: $p_1, p_2, \ldots, p_m \in [n]$

$$f_i = \#\{j \le m : p_j = i\}$$

$$f_i^t = \#\{j \le t : p_j = i\}$$

- Read items one-at-a-time
- Approximate answer
- Succeed with probability $2/3$
- Goals
  - small space ($\Longrightarrow$ **few random bits**)
  - fast updates
  - fast queries

Stream: $p_1, p_2, \ldots, p_m \in [n]$

$$f_i = \#\{j \le m : p_j = i\}$$

$$f_i^t = \#\{j \le t : p_j = i\}$$

- Read items one-at-a-time
- Approximate answer
- Succeed with probability 2/3
- Goals
  - small space ($\Longrightarrow$ **few random bits**)
  - fast updates
  - fast queries
- This talk: find "frequent items" $(f_H \ge \epsilon \|f\|_2)$

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61, 389, 660, 820, 672, 271, 245, 417, 535, 2, 702, 590, 595, 671, 805, 476, 815, 6, 733, 397, 44, 917, 521, 294, 186, 727, 367, 721, 40, 666, 813, 467, 523, 261, 571, 611, 125, 870, 17, 186, 141, 250, 541, 770, 890, 832, 259, 530, 977, 970, 268, 629, 507, 830, 478, 401, 685, 988, 164, 441, 743, 905, 646, 715, 282, 774, 188, 901, 302, 802, 338, 926, 831, 627, 546, 520, 711, 231, 51, 600, 50, 852, 734, 512, 374, 942, 753, 412, 697, 776, 779, 453, 581, 312, 25, 407, 471, 92, 625, 696, 220, 24, 938, 818, 990, 927, 496, 377, 894, 575, 851, 217, 170, 59, 858, 130, 684, 732, 434, 543, 615, 972, 335, 425, 525, 79, 192, 47, 186, 588, 437, 357, 186, 954, 396, 510, 86, 189, 474, 247, 965, 69, 308, 758, 593, 532, 98, 296, 985, 256, 633, 892, 27, 26, 316, 400, 186, 257, 297, 516, 929, 908, 817, 428, 722, 826, 768, 750, 73, 888, 799, 620, 534, 522, 206, 129, 737, 526, 698, 186, 957, 238, 49, 741, 597, 177, 202, 94, 323, 999, 186, 204, 405, 778, 598, 489, 243, 937, 36, 240, 422, 232, 421, 33, 565, 804, 735, 278, 455, 759, 155, 653, 104, 352, 642, 784, 119, 209, 639, 991, 459, 163, 198, 121, 207, 274, 111, 96, 920, 184, 955, 449, 363, 195, 839, 413, 841, 582, 216, 71, 179, 107, 313, 244, 472, 186, 752, 131, 952, 95, 356, 466, 679, 18, 859, 623, 330, 365, 822, 260, 442, 775, 714, 659, 855, 789, 909, 574, 609, 448, 865, 899, 326, 275, 230, 502, 186, 406, 958, 181, 392, 394, 950, 931, 451, 488, 186, 439, 444, 355, 606, 508, 939, 554, 720, 384, 650, 183, 947, 284, 212, 375, 166, 154, 583, 771, 431, 511, 930, 362, 351, 224, 116, 12, 20, 863, 28, 506, 794, 457, 631, 161, 100, 555, 780, 236, 856, 322, 801, 108, 285, 867, 186, 447, 419, 171, 873, 113, 570, 500, 568, 884, 638, 877, 699, 210, 796, 200, 969, 572, 186, 492, 404, 145, 787, 887, 707, 80, 258, 266, 906, 66, 619, 587, 76, 346, 350, 320, 835, 227, 402, 594, 14, 382, 562, 219, 848, 962, 781, 612, 643, 87, 276, 225, 142, 676, 849, 945, 973, 742, 172, 372, 88, 381, 112, 728, 345, 481, 673, 786, 607, 90, 961, 117, 936, 933, 693, 343, 226, 557, 900, 800, 120, 823, 709, 850, 484, 935, 726, 118, 301, 688, 23, 331, 806, 993, 5, 46, 186, 808, 70, 291, 280, 292, 548, 691, 498, 527, 458, 803, 978, 561, 147, 616, 904, 912, 369, 664, 763, 123, 925, 186, 531, 333, 494, 515, 584, 746, 556, 641, 186, 767, 986, 640, 875, 658, 948, 390, 45, 635, 614, 662, 306, 215, 452, 748, 701, 139, 329, 943, 99, 205, 694, 456, 853, 395, 414, 710, 348, 829, 193, 126, 708, 749, 360, 916, 138, 152, 19, 42, 636, 897, 791, 544, 493, 3, 167, 967, 968, 103, 418, 821, 987, 307, 293, 380, 67, 11, 342, 57, 824, 60, 432, 186, 632, 628, 403, 22, 795, 55, 83, 704, 790, 483, 34, 174, 65, 599, 38, 998, 792, 39, 295, 712, 497, 857, 221, 540, 949, 995, 788, 185, 7, 461, 101, 110, 186, 435, 827, 760, 578, 248, 844, 267, 838, 689, 602, 840, 415, 186, 289, 288, 984, 895, 235, 410, 670, 370, 383, 281, 341, 358, 263, 321, 143, 487, 608, 208, 186, 992, 408, 580, 334, 386, 918, 836, 411, 433, 354, 861, 353, 15, 309, 314, 464, 589, 52, 317, 529, 186, 944, 538, 128, 975, 416, 667, 31, 907, 946, 876, 327, 644, 657, 982, 976, 963, 652, 0, 234, 964, 706, 203, 196, 159, 793, 718, 62, 158, 379, 105, 586, 53, 339, 828, 499, 373, 283, 951, 862, 686, 114, 703, 393, 246, 941, 162, 825, 315, 777, 241, 371, 463, 675, 772, 480, 690, 751, 637, 436, 385, 157, 971, 910, 603, 173, 974, 303, 89, 233, 872, 610, 186, 762, 717, 429, 450, 186, 465, 819, 64, 122, 222, 249, 340, 891, 785, 332, 186, 134, 475, 349, 723, 551, 336, 9, 700, 547, 133, 82, 84, 730, 783, 874, 560, 359, 490, 860, 814, 756, 127, 48, 656, 678, 305, 29, 810, 149, 798, 914, 424, 454, 30, 843, 878, 573, 915, 677, 996, 953, 866, 533, 4, 585, 287, 624, 409, 889, 903, 596, 344, 378, 391, 661, 869, 665, 135, 364, 387, 186, 514, 253, 738, 279, 846, 252, 242, 35, 32, 468, 680, 186, 190, 58, 552, 893, 199, 144, 630, 366, 902, 981, 924, 713, 509, 495, 754, 576, 604, 626, 186, 766, 388, 430, 669, 21, 634, 74, 959, 834, 299, 191, 115, 553, 845, 186, 591, 536, 180, 102, 194, 460, 881, 592, 186, 16, 146, 470, 265, 725, 81, 566, 868, 505, 518, 290, 254, 420, 549, 310, 37, 286, 731, 773, 911, 695, 816, 617, 519, 75, 136, 504, 85, 446, 563, 769, 692, 837, 854, 41, 361, 479, 559, 655, 153, 77, 705, 54, 739, 524, 182, 223, 928, 8, 663, 148, 186, 13, 214, 668, 517, 919, 654, 399, 63, 740, 569, 724, 879, 186, 469, 956, 262, 93, 812, 567, 201, 980, 898, 542, 151, 681, 269, 187, 745, 175, 78, 885, 298, 687, 211, 97, 747, 485, 165, 272, 807, 847, 443, 440, 558, 989, 605, 1, 237, 757, 797, 649, 477, 997, 228, 186, 736, 513, 934, 43, 68, 764, 264, 932, 719, 651, 462, 811, 503, 564, 427, 491, 445, 186, 550, 882, 137, 311, 896, 983, 765, 273, 601, 545, 10, 648

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61, 389, 660, 820, 672, 271, 245, 417, 535, 2, 702, 590, 595, 671, 805, 476, 815, 6, 733, 397, 44, 917, 521, 294, **186**, 727, 367, 721, 40, 666, 813, 467, 523, 261, 571, 611, 125, 870, 17, **186**, 141, 250, 541, 770, 890, 832, 259, 530, 977, 970, 268, 629, 507, 830, 478, 401, 685, 988, 164, 441, 743, 905, 646, 715, 282, 774, 188, 901, 302, 802, 338, 926, 831, 627, 546, 520, 711, 231, 51, 600, 50, 852, 734, 512, 374, 942, 753, 412, 697, 776, 779, 453, 581, 312, 25, 407, 471, 92, 625, 696, 220, 24, 938, 818, 990, 927, 496, 377, 894, 575, 851, 217, 170, 59, 858, 130, 684, 732, 434, 543, 615, 972, 335, 425, 525, 79, 192, 47, **186**, 588, 437, 357, **186**, 954, 396, 510, 86, 189, 474, 247, 965, 69, 308, 758, 593, 532, 98, 296, 985, 256, 633, 892, 27, 26, 316, 400, **186**, 257, 297, 516, 929, 908, 817, 428, 722, 826, 768, 750, 73, 888, 799, 620, 534, 522, 206, 129, 737, 526, 698, **186**, 957, 238, 49, 741, 597, 177, 202, 94, 323, 999, **186**, 204, 405, 778, 598, 489, 243, 937, 36, 240, 422, 232, 421, 33, 565, 804, 735, 278, 455, 759, 155, 653, 104, 352, 642, 784, 119, 209, 639, 991, 459, 163, 198, 121, 207, 274, 111, 96, 920, 184, 955, 449, 363, 195, 839, 413, 841, 582, 216, 71, 179, 107, 313, 244, 472, **186**, 752, 131, 952, 95, 356, 466, 679, 18, 859, 623, 330, 365, 822, 260, 442, 775, 714, 659, 855, 789, 909, 574, 609, 448, 865, 899, 326, 275, 230, 502, **186**, 406, 958, 181, 392, 394, 950, 931, 451, 488, **186**, 439, 444, 355, 606, 508, 939, 554, 720, 384, 650, 183, 947, 284, 212, 375, 166, 154, 583, 771, 431, 511, 930, 362, 351, 224, 116, 12, 20, 863, 28, 506, 794, 457, 631, 161, 100, 555, 780, 236, 856, 322, 801, 108, 285, 867, **186**, 447, 419, 171, 873, 113, 570, 500, 568, 884, 638, 877, 699, 210, 796, 200, 969, 572, **186**, 492, 404, 145, 787, 887, 707, 80, 258, 266, 906, 66, 619, 587, 76, 346, 350, 320, 835, 227, 402, 594, 14, 382, 562, 219, 848, 962, 781, 612, 643, 87, 276, 225, 142, 676, 849, 945, 973, 742, 172, 372, 88, 381, 112, 728, 345, 481, 673, 786, 607, 90, 961, 117, 936, 933, 693, 343, 226, 557, 900, 800, 120, 823, 709, 850, 484, 935, 726, 118, 301, 688, 23, 331, 806, 993, 5, 46, **186**, 808, 70, 291, 280, 292, 548, 691, 498, 527, 458, 803, 978, 561, 147, 616, 904, 912, 369, 664, 763, 123, 925, **186**, 531, 333, 494, 515, 584, 746, 556, 641, **186**, 767, 986, 640, 875, 658, 948, 390, 45, 635, 614, 662, 306, 215, 452, 748, 701, 139, 329, 943, 99, 205, 694, 456, 853, 395, 414, 710, 348, 829, 193, 126, 708, 749, 360, 916, 138, 152, 19, 42, 636, 897, 791, 544, 493, 3, 167, 967, 968, 103, 418, 821, 987, 307, 293, 380, 67, 11, 342, 57, 824, 60, 432, **186**, 632, 628, 403, 22, 795, 55, 83, 704, 790, 483, 34, 174, 65, 599, 38, 998, 792, 39, 295, 712, 497, 857, 221, 540, 949, 995, 788, 185, 7, 461, 101, 110, **186**, 435, 827, 760, 578, 248, 844, 267, 838, 689, 602, 840, 415, **186**, 289, 288, 984, 895, 235, 410, 670, 370, 383, 281, 341, 358, 263, 321, 143, 487, 608, 208, **186**, 992, 408, 580, 334, 386, 918, 836, 411, 433, 354, 861, 353, 15, 309, 314, 464, 589, 52, 317, 529, **186**, 944, 538, 128, 975, 416, 667, 31, 907, 946, 876, 327, 644, 657, 982, 976, 963, 652, 0, 234, 964, 706, 203, 196, 159, 793, 718, 62, 158, 379, 105, 586, 53, 339, 828, 499, 373, 283, 951, 862, 686, 114, 703, 393, 246, 941, 162, 825, 315, 777, 241, 371, 463, 675, 772, 480, 690, 751, 637, 436, 385, 157, 971, 910, 603, 173, 974, 303, 89, 233, 872, 610, **186**, 762, 717, 429, 450, **186**, 465, 819, 64, 122, 222, 249, 340, 891, 785, 332, **186**, 134, 475, 349, 723, 551, 336, 9, 700, 547, 133, 82, 84, 730, 783, 874, 560, 359, 490, 860, 814, 756, 127, 48, 656, 678, 305, 29, 810, 149, 798, 914, 424, 454, 30, 843, 878, 573, 915, 677, 996, 953, 866, 533, 4, 585, 287, 624, 409, 889, 903, 596, 344, 378, 391, 661, 869, 665, 135, 364, 387, **186**, 514, 253, 738, 279, 846, 252, 242, 35, 32, 468, 680, **186**, 190, 58, 552, 893, 199, 144, 630, 366, 902, 981, 924, 713, 509, 495, 754, 576, 604, 626, **186**, 766, 388, 430, 669, 21, 634, 74, 959, 834, 299, 191, 115, 553, 845, **186**, 591, 536, 180, 102, 194, 460, 881, 592, **186**, 16, 146, 470, 265, 725, 81, 566, 868, 505, 518, 290, 254, 420, 549, 310, 37, 286, 731, 773, 911, 695, 816, 617, 519, 75, 136, 504, 85, 446, 563, 769, 692, 837, 854, 41, 361, 479, 559, 655, 153, 77, 705, 54, 739, 524, 182, 223, 928, 8, 663, 148, **186**, 13, 214, 668, 517, 919, 654, 399, 63, 740, 569, 724, 879, **186**, 469, 956, 262, 93, 812, 567, 201, 980, 898, 542, 151, 681, 269, 187, 745, 175, 78, 885, 298, 687, 211, 97, 747, 485, 165, 272, 807, 847, 443, 440, 558, 989, 605, 1, 237, 757, 797, 649, 477, 997, 228, **186**, 736, 513, 934, 43, 68, 764, 264, 932, 719, 651, 462, 811, 503, 564, 427, 491, 445, **186**, 550, 882, 137, 311, 896, 983, 765, 273, 601, 545, 10, 648

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61, 389, 660, 820, 672, 271, 245, 417, 535, 2, 702, 590, 595, 671, 805, 476,
815, 6, 733, 397, 44, 917, 521, 294, **186**, 727, 367, 721, 40, 666, 813, 467, 523, 261, 571, 611, 125, 870, 17, **186**, 141, 250,
541, 770, 890, 832, 259, 530, 977, 970, 268, 629, 507, 830, 478, 401, 685, 988, 164, 441, 743, 905, 646, 715, 282, 774, 188,
901, 302, 802, 338, 926, 831, 627, 546, 520, 711, 231, 51, 600, 50, 852, 734, 512, 374, 942, 753, 412, 697, 776, 779, 453, 581,
312, 25, 407, 471, 92, 625, 696, 220, 24, 938, 818, 990, 927, 496, 377, 894, 575, 851, 217, 170, 59, 858, 130, 684, 732, 434,
543, 615, 972, 335, 425, 525, 79, 192, 47, **186**, 588, 437, 357, **186**, 954, 396, 510, 86, 189, 474, 247, 965, 69, 308, 758, 593,
532, 98, 296, 985, 256, 633, 892, 27, 26, 316, 400, **186**, 257, 297, 516, 929, 908, 817, 428, 722, 826, 768, 750, 73, 888, 799,
620, 534, 522, 206, 129, 737, 526, 698, **186**, 957, 238, 49, 741, 597, 177, 202, 94, 323, 999, **186**, 204, 405, 778, 598, 489, 243,
937, 36, 240, 422, 232, 421, 33, 565, 804, 735, 278, 455, 759, 155, 653, 104, 352, 642, 784, 119, 209, 639, 991, 459, 163, 198,

Decision problem:

Alon, Matias, Szegedy '96; Bar-Yossef, Jayram, Kumar, Sivakumar '02; Chakrabarti, Khot, Sun '03

Each of n items at most once

vs

One item n^{α} times, rest $\leq$once

Requires $\Omega(n^{1-2\alpha})$ bits of space

121, 207, 274, 111 … 752, 131, 952,
95, 356, 466, 679, … 326, 275, 230,
502, **186**, 406, 95 … 183, 947, 284,
212, 375, 166, 154 … 555, 780, 236,
856, 322, 801, 10 … 969, 572, **186**,
492, 404, 145, 78 … 562, 219, 848,
962, 781, 612, 643 … 607, 90, 961,
117, 936, 933, 69 … 93, 5, 46, **186**,
808, 70, 291, 280, … 531, 333, 494,
515, 584, 746, 556 … 9, 329, 943, 99,
205, 694, 456, 85 … 4, 493, 3, 167,
967, 968, 103, 418 … 704, 790, 483,
34, 174, 65, 599, 3 … 827, 760, 578,
248, 844, 267, 83 … 321, 143, 487,
608, 208, **186**, 992 … **186**, 944, 538,
128, 975, 416, 66 … 718, 62, 158,
379, 105, 586, 53, … 463, 675, 772,
480, 690, 751, 637 … 6, 465, 819, 64,
122, 222, 249, 340, 891, 785, 332, **186**, 134, 475, 349, 723, 551, 336, 9, 700, 547, 133, 82, 84, 730, 783, 874, 560, 359, 490,
860, 814, 756, 127, 48, 656, 678, 305, 29, 810, 149, 798, 914, 424, 454, 30, 843, 878, 573, 915, 677, 996, 953, 866, 533, 4, 585,
287, 624, 409, 889, 903, 596, 344, 378, 391, 661, 869, 665, 135, 364, 387, **186**, 514, 253, 738, 279, 846, 252, 242, 35, 32, 468,
680, **186**, 190, 58, 552, 893, 199, 144, 630, 366, 902, 981, 924, 713, 509, 495, 754, 576, 604, 626, **186**, 766, 388, 430, 669, 21,
634, 74, 959, 834, 299, 191, 115, 553, 845, **186**, 591, 536, 180, 102, 194, 460, 881, 592, **186**, 16, 146, 470, 265, 725, 81, 566,
868, 505, 518, 290, 254, 420, 549, 310, 37, 286, 731, 773, 911, 695, 816, 617, 519, 75, 136, 504, 85, 446, 563, 769, 692, 837,
854, 41, 361, 479, 559, 655, 153, 77, 705, 54, 739, 524, 182, 223, 928, 8, 663, 148, **186**, 13, 214, 668, 517, 919, 654, 399, 63,
740, 569, 724, 879, **186**, 469, 956, 262, 93, 812, 567, 201, 980, 898, 542, 151, 681, 269, 187, 745, 175, 78, 885, 298, 687, 211,
97, 747, 485, 165, 272, 807, 847, 443, 440, 558, 989, 605, 1, 237, 757, 797, 649, 477, 997, 228, **186**, 736, 513, 934, 43, 68,
764, 264, 932, 719, 651, 462, 811, 503, 564, 427, 491, 445, **186**, 550, 882, 137, 311, 896, 983, 765, 273, 601, 545, 10, 648

Goal Return all H such that $f_H \geq \epsilon \|f\|_2$
Every H returned has $f_H \geq \frac{\epsilon}{2} \|f\|_2$

$[\epsilon = \Omega(1)]$	CountSketch Charikar, Chen, Farach-Colton '02	Tuesday Braverman,C, Ivkin, Woodruff '16	Today B,C,I,Nelson, Wang, W '??
space (bits)	$\log^2 n$	$\log n \log\log n$	$\log n$
update time	$\log n$	$2^{(\log\log n)^2}$	$O(1)$
query time	$O(1)$	$\tilde{O}(n)$	$O(1)$

By a standard reduction...

Goal If there exists $H \in [n]$ such that

$$f_H^2 \geq K \sum_{j \neq H} f_j^2$$

then return H

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61, 389, 660, 820, 672, 271, 245, 417, 535, 2, 702, 590, 595, 671, 805, 476, 815, 6, 733, 397, 44, 917, 521, 294, **186**, 727, 367, 721, 40, 666, 813, 467, 523, 261, 571, 611, 125, 870, 17, **186**, 141, 250, 541, 770, 890, 832, 259, 530, 977, 970, 268, 629, 507, 830, 478, 401, 685, 988, 164, 441, 743, 905, 646, 715, 282, 774, 188,

Each of n items at most once

vs

One item $K\sqrt{n}$ times, rest $\leq$once

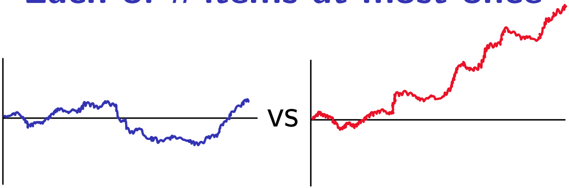
$$Z_i \overset{4w}{\sim} \{-1,+1\} \quad X = \sum_{j=1}^{m} Z_{p_j} = \langle Z, f \rangle$$

$$\mathbb{E}X^2 = \sum f_i^2 = \|f\|_2^2$$

97, 747, 485, 165, 272, 807, 847, 443, 440, 558, 989, 605, 1, 237, 757, 797, 649, 477, 997, 228, **186**, 736, 513, 934, 43, 68, 764, 264, 932, 719, 651, 462, 811, 503, 564, 427, 491, 445, **186**, 550, 882, 137, 311, 896, 983, 765, 273, 601, 545, 10, 648

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61, 389, 660, 820, 672, 271, 245, 417, 535, 2, 702, 590, 595, 671, 805, 476,
815, 6, 733, 397, 44, 917, 521, 294, **186**, 727, 367, 721, 40, 666, 813, 467, 523, 261, 571, 611, 125, 870, 17, **186**, 141, 250,
541, 770, 890, 832, 259, 530, 977, 970, 268, 629, 507, 830, 478, 401, 685, 988, 164, 441, 743, 905, 646, 715, 282, 774, 188,
901, 302, 802, 338, 926, 831, 627, 546, 520, 711, 231, 51, 600, 50, 852, 734, 512, 374, 942, 753, 412, 697, 776, 779, 453, 581,

Each of n items at most once



One item $K\sqrt{n}$ times, rest $\leq$once

$$Z_i \overset{4w}{\sim} \{-1, +1\} \quad X = \sum_{j=1}^{m} Z_{p_j} = \langle Z, f \rangle$$

$$\mathbb{E}X^2 = \sum f_i^2 = \|f\|_2^2$$

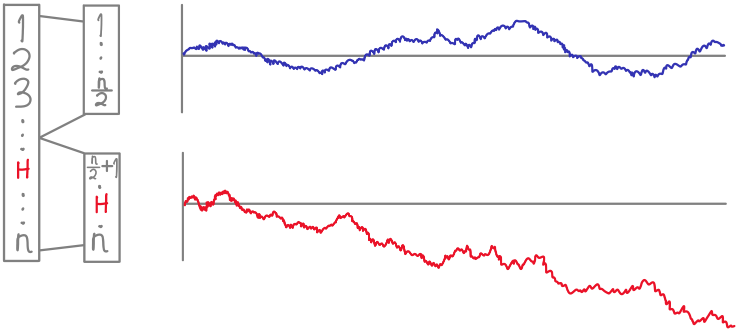
740, 569, 724, 879, **186**, 469, 956, 262, 93, 812, 567, 201, 980, 898, 542, 151, 681, 269, 187, 745, 175, 78, 885, 298, 687, 211,
97, 747, 485, 165, 272, 807, 847, 443, 440, 558, 989, 605, 1, 237, 757, 797, 649, 477, 997, 228, **186**, 736, 513, 934, 43, 68,
764, 264, 932, 719, 651, 462, 811, 503, 564, 427, 491, 445, **186**, 550, 882, 137, 311, 896, 983, 765, 273, 601, 545, 10, 648

Let's assume:

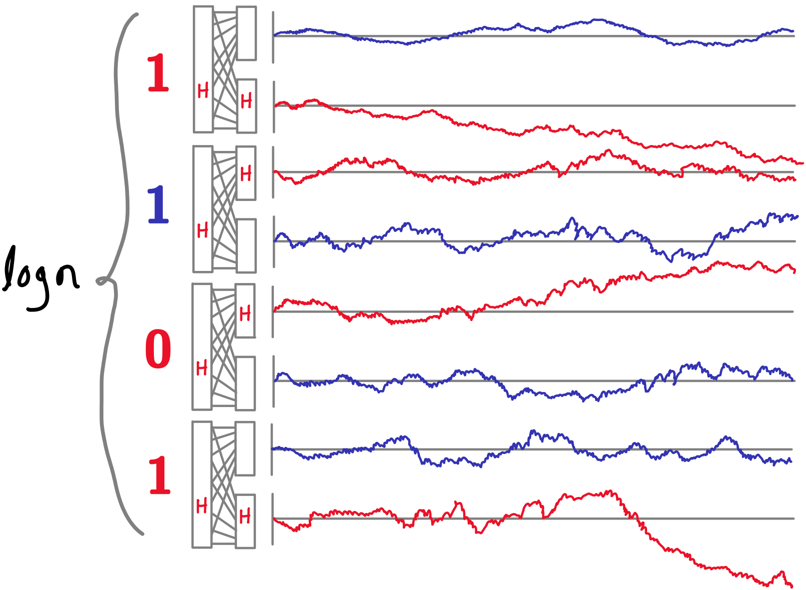
1. General frequencies
2. There is a heavy hitter $H \in [n]$
3. $f_H \geq K \left(\sum_{j \neq H} f_j^2 \right)^{1/2}$

Learning 1 bit

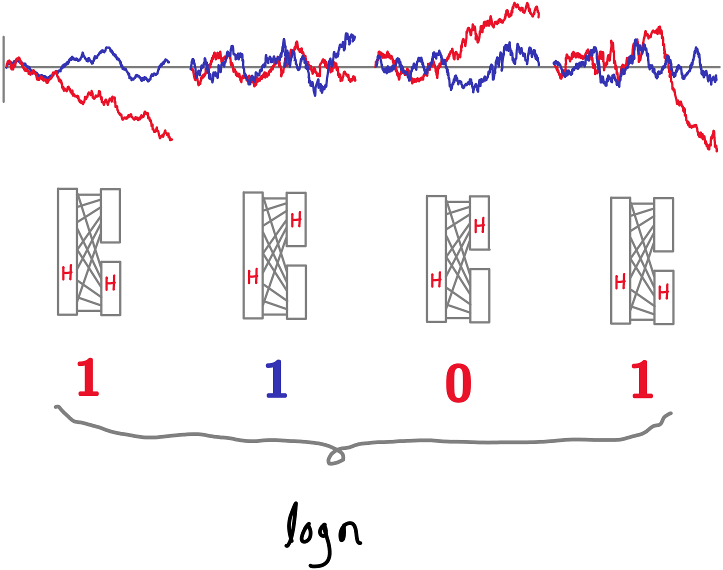
1



Learning many bits



Could it be so simple?...



Could it be so simple?... no

How do we choose the break points?

Could it be so simple?... no

How do we choose the break points?

H may not be heavy in the interval!

$$f_H \implies \frac{f_H}{\log n} \qquad \text{but} \qquad \|f\|_2 \implies \frac{\|f\|_2}{\sqrt{\log n}}$$

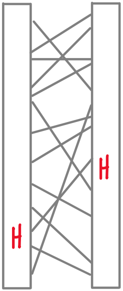
Could it be so simple?... no

How do we choose the break points?

H may not be heavy in the interval!

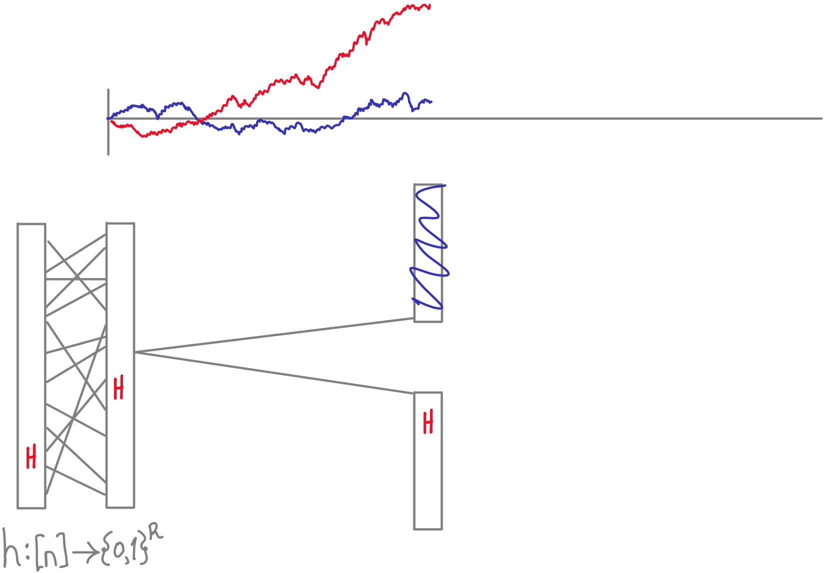
$$f_H \implies \frac{f_H}{\log n} \quad \text{but} \quad \|f\|_2 \implies \frac{\|f\|_2}{\sqrt{\log n}}$$

Can't store $O(\log n)$ independent hash functions.

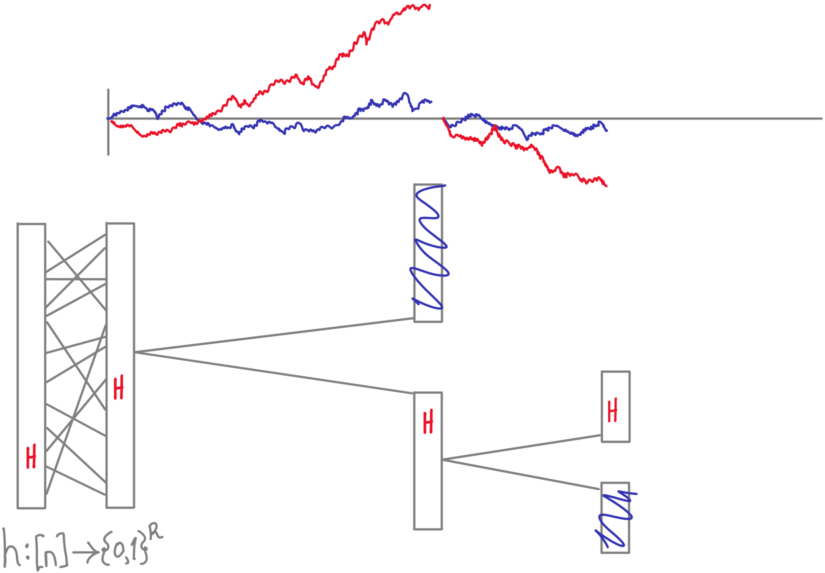


$h : [n] \to \{0,1\}^R$

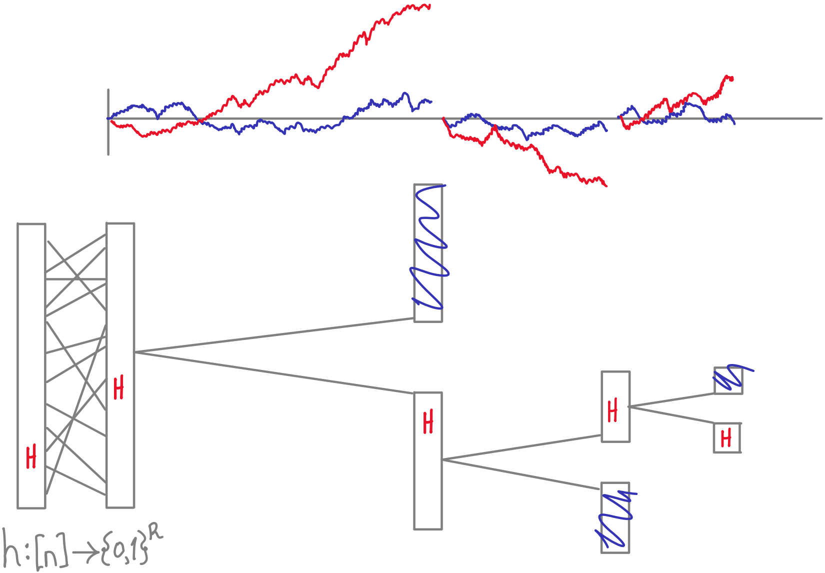
Suppose we guarantee H always “wins”. . .



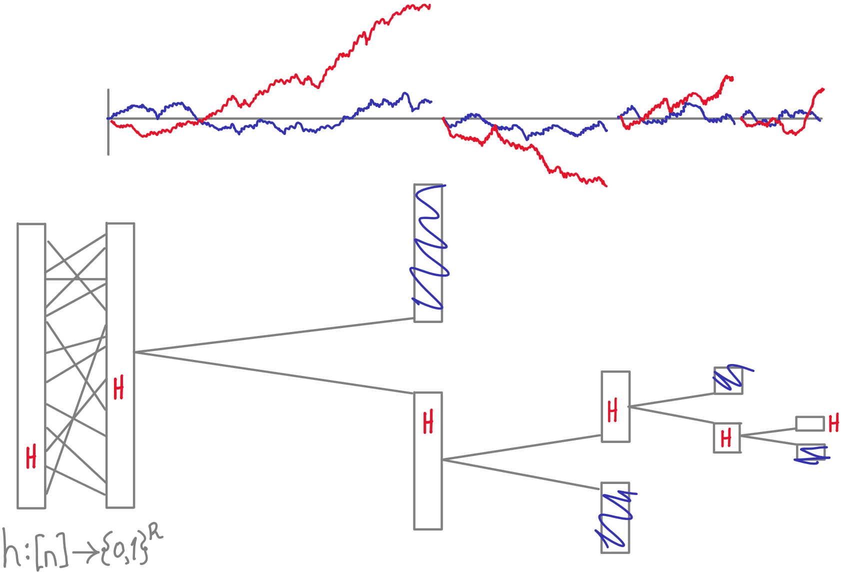
Suppose we guarantee H always “wins”. . .

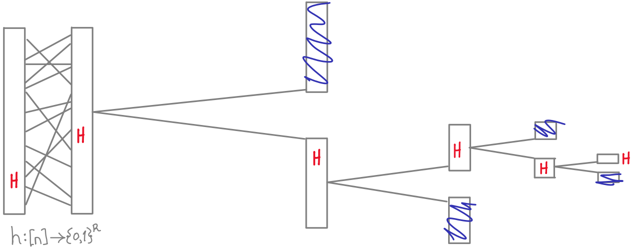


Suppose we guarantee H always "wins"...



Suppose we guarantee H always "wins"...





for p in stream **do**

  if $h(p) = b_1 b_2 \cdots b_r i *$ **then**

    $X_i \mathrel{+}= Z_p$

  if $|X_0 + X_1| \geq c\sigma\beta^r$ **then**

    $b_{r+1} \leftarrow \mathbf{1}(|X_1| > |X_0|)$

    if $r + 1 = R$ **then**

      return p where $h(p) = b_1 b_2 \ldots b_R$

    else

      restart with $b = b_1 \cdots b_r b_{r+1} 0 \cdots 0$

$R = O(\log n)$

$\sigma = \|f\|_2$

$c = 1/32$

$\beta = 3/4$

$O(1)$ update time and $O(\log n)$ bits storage

for p in stream **do**

  if $h(p) = b_1 b_2 \cdots b_r i*$ **then**

    $X_i \mathrel{+}= Z_p$

  if $|X_0 + X_1| \geq c\sigma\beta^r$ **then**

    $b_{r+1} \leftarrow \mathbf{1}(|X_1| > |X_0|)$

    if $r + 1 = R$ **then**

      return p where $h(p) = b_1 b_2 \ldots b_R$

    else

      restart with $b = b_1 \cdots b_r b_{r+1} 0 \cdots 0$

$R = O(\log n)$

$\sigma = \|f\|_2$

$c = 1/32$

$\beta = 3/4$

Key to the analysis

Theorem (Chaining Inequality)

Let $Z \in_u \{-1,+1\}^n$ and $X^t = \langle f^t, Z \rangle$.

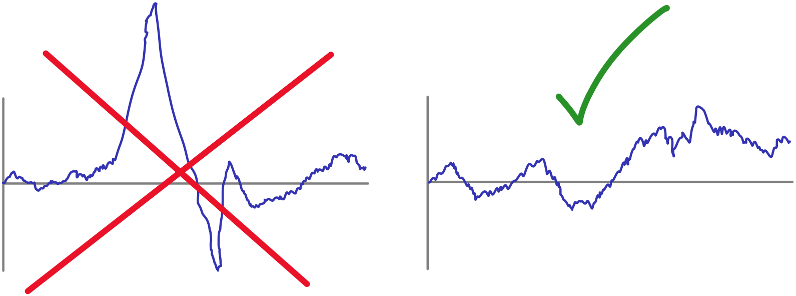
$$\mathbb{E} \sup_{1 \le t \le m} |X^t| \lesssim \sqrt{\mathrm{Var}(X^m)} = \|f\|_2$$

Key to the analysis

Theorem (Chaining Inequality)

Let $Z \in_u \{-1,+1\}^n$ and $X^t = \langle f^t, Z \rangle$.

$$\mathbb{E} \sup_{1 \le t \le m} |X^t| \lesssim \sqrt{\mathrm{Var}(X^m)} = \|f\|_2$$

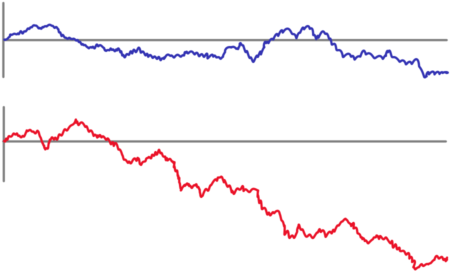


Key to the analysis

Theorem (Chaining Inequality)

Let $Z \in_u \{-1, +1\}^n$ *and* $X^t = \langle f^t, Z \rangle$.

$$\mathbb{E} \sup_{1 \le t \le m} |X^t| \lesssim \sqrt{\mathrm{Var}(X^m)} = \|f\|_2$$

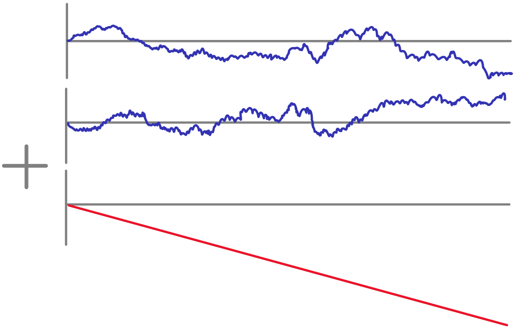


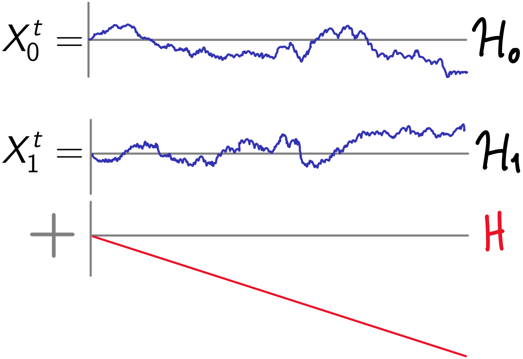
Key to the analysis

Theorem (Chaining Inequality)

Let $Z \in_u \{-1,+1\}^n$ *and* $X^t = \langle f^t, Z \rangle$.

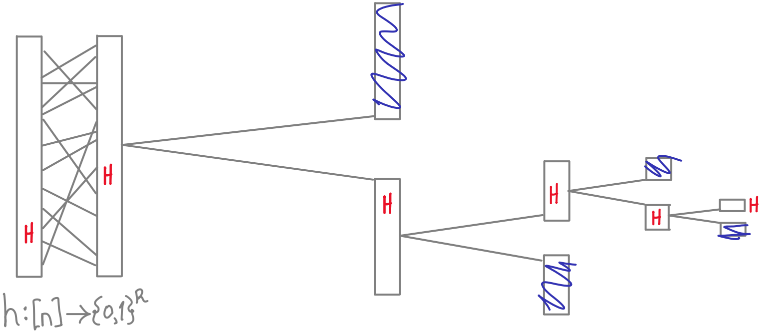
$$\mathbb{E} \sup_{1 \le t \le m} |X^t| \lesssim \sqrt{\mathrm{Var}(X^m)} = \|f\|_2$$





for p in the stream **do**
  if $h(p) = b_1 b_2 \cdots b_r i *$ **then**
    $X_i \mathrel{+}= Z_p$
  if $|X_0 + X_1| \geq c\sigma\beta^r$ **then**
    $b_{r+1} \leftarrow \mathbf{1}(|X_1| > |X_0|)$

$$\Pr\left(\sup_t |X_0^t| \geq \frac{1}{2} c\sigma\beta^r \,\middle|\, \mathcal{H}_0\right) \lesssim \frac{\left(\sum_{j\in\mathcal{H}_0} f_j^2\right)^{1/2}}{\sigma\beta^r} \overset{\mathbb{E}}{\leq} \frac{1}{K(\beta\sqrt{2})^r}$$



$$\Pr(H \text{ loses}) \le \sum_{r\ge 0} \frac{1}{cK(\beta\sqrt{2})^r} \lesssim \frac{1}{K}$$

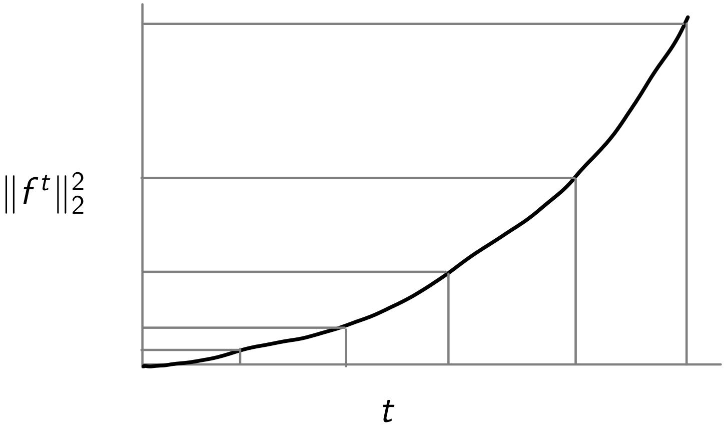
$$\#H\text{'s used} \le \sum_{r\ge 0} 2c\sigma\beta^r = \sigma/4 < f_H$$

$$(c = 2^{-5}, \beta = 3/4)$$

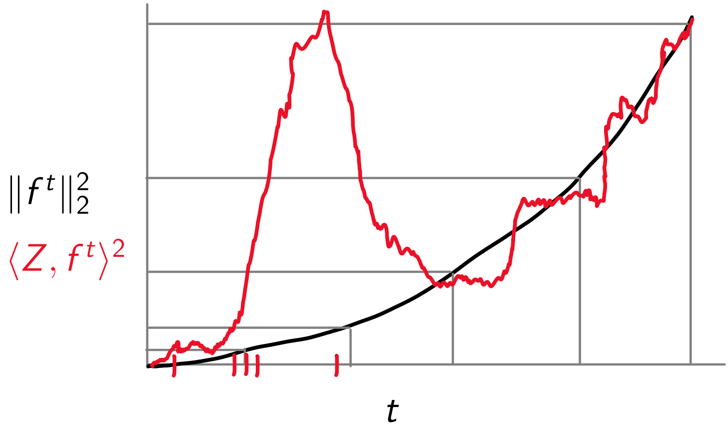
We conclude...

If there is a K-heavy hitter H then **HH** outputs H with probability at least $2/3$.

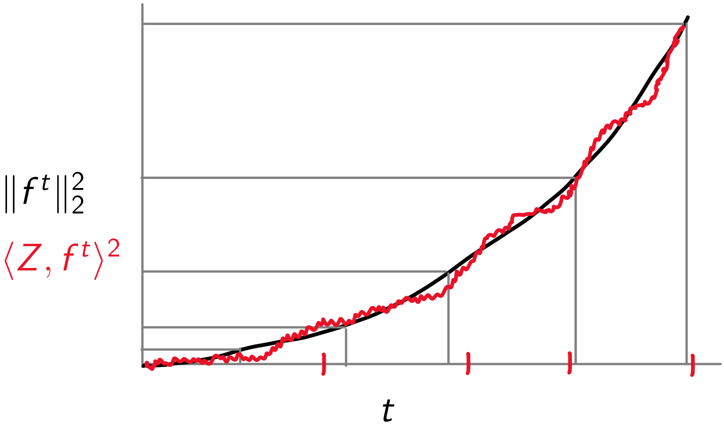
Guessing $\sigma \approx \|f\|_2$



Guessing $\sigma \approx \|f\|_2$



Guessing $\sigma \approx \|f\|_2$



Theorem ($\|\mathbf{f}\|_2$ Tracking)

There exists a streaming algorithm using $O(\frac{1}{\epsilon^2}\log n)$ bits of space that at each time t outputs Y^t where with probability at least 2/3

$$\left|Y^t - \|f^t\|_2^2\right| \leq \epsilon\|f^m\|_2^2 \text{ for all } t \leq m.$$

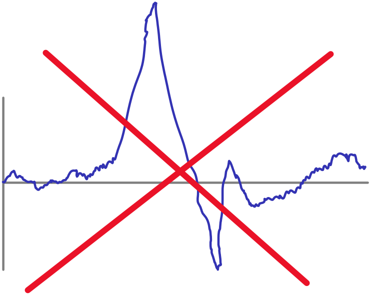
Chaining with few random bits

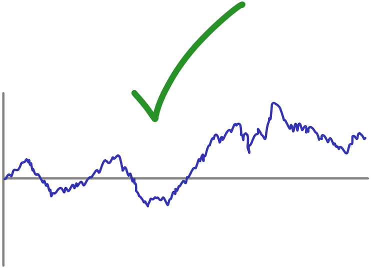
Key to the analysis

Theorem (Chaining Inequality)

Let $Z \in_u \{-1,+1\}^n$ *and* $X^t = \langle f^t, Z \rangle$.

$$\mathbb{E} \sup_{1 \le t \le m} |X^t| \lesssim \sqrt{\mathrm{Var}(X^m)} = \|f\|_2$$





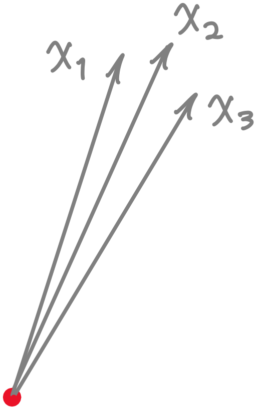
Union bound?

If $Z \in_u \{-1,+1\}^n$ and $G \sim N(0, I_{n\times n})$ then

$$\mathbb{E}\sup\langle Z, f^t\rangle \le \sqrt{\frac{\pi}{2}}\,\mathbb{E}\sup\langle G, f^t\rangle.$$

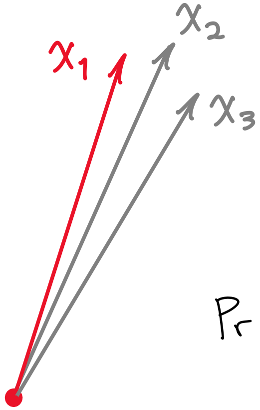
$$\Pr\left(\sup_{1\le t\le m} |\langle G, f^t\rangle| \le u\|f\|_2\right) \le \sum_{t=1}^{m} \Pr\left(|\langle G, f^t\rangle| \le u\|f\|_2\right)$$

$$\le \sum_{t\ge 1} 2\exp\left\{\frac{-u^2}{2}\cdot\frac{\|f\|_2^2}{\|f^t\|_2^2}\right\}$$



$$\mathbb{E}\langle G, x\rangle\langle G, y\rangle = \langle x, y\rangle$$

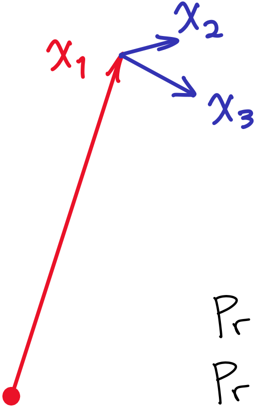
independence $\equiv$ orthogonality



$$\mathbb{E}\langle G, x\rangle\langle G, y\rangle = \langle x, y\rangle$$

independence $\equiv$ orthogonality

$$\Pr(|\langle G, x_1\rangle| \geq u) \lesssim 2\exp\left\{\frac{-u^2}{2\|x_1\|}\right\}$$



$$\mathbb{E}\langle G, x\rangle\langle G, y\rangle = \langle x, y\rangle$$

independence $\equiv$ orthogonality

$$\Pr\left(|\langle G, x_1\rangle| \geq u\right) \lesssim 2\exp\left\{\frac{-u^2}{2\|x_1\|}\right\}$$

$$\Pr\left(|\langle G, x_1 - x_2\rangle| \geq u\right) \lesssim 2\exp\left\{\frac{-u^2}{2\|x_1 - x_2\|}\right\}$$

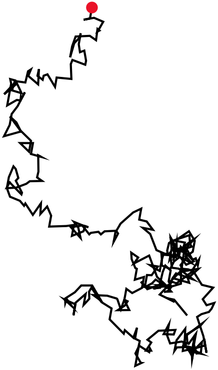
$$\Pr\left(|\langle G, x_1 - x_3\rangle| \geq u\right) \lesssim 2\exp\left\{\frac{-u^2}{2\|x_1 - x_3\|}\right\}$$

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61, **186**, 389, 660, 820, 672, 271, **186**, 245, 417, 535, 2, 702, 590, 595, 671, 805, 476, 815, 6, 733, 397, 44, 917, 521, 294, 727, 367, 721, 40, 666, 813, 467, 523, 261, 571, 611, 125, **186**, 870, **186**, 17, 141, 250, 541, 770, 890, 832, 259, **186**, 530, 977, 970, 268, 629, 507, 830, 478, **186**, 401, 685, 988, **186**, 164, 441, 743, 905, 646, 715, 282, 774, 188, 901, 302, 802, 338, 926, 831, 627, 546, 520, 711, **186**, 231, 51, 600, 50, 852, 734, 512, 374, 942, 753, **186**, 412, 697, 776, **186**, 779, 453, 581, 312, 25, 407, 471, 92, 625, 696, 220, 24, 938, 818, 990, 927, **186**, 496, 377, 894, 575, 851, 217, **186**, 170, 59, 858, 130, 684, 732, 434, 543, 615, 972, 335, 425, 525, 79, 192, 47, 588, 437, 357, 954, 396, 510, 86, 189, 474, 247, **186**, 965, 69, 308, 758, 593, 532, 98, **186**, 296, 985, 256, **186**, 633, 892, 27, 26, 316, 400, 257, 297, 516, 929, 908, 817, 428, 722, 826, 768, 750, 73, 888, 799, 620, **186**, 534, 522, 206, 129, 737, 526, 698, 957, 238, 49, 741, **186**, 597, 177, 202, 94, 323, **186**, 999, 204, 405, 778, 598, 489, **186**, 243, 937, 36, 240, 422, 232, 421, **186**, 33, 565, 804, 735, 278, 455, 759, 155, 653, 104, 352, 642, 784, 119, 209, 639, 991, 459, 163, 198, 121, 207, 274, 111, **186**, 96, 920, **186**, 184, 955, 449, **186**, 363, 195, 839, 413, 841, 582, 216, 71, 179, 107, 313, 244, 472, 752, 131, 952, 95, 356, 466, 679, 18, **186**, 859, 623, 330, 365, **186**, 822, 260, 442, 775, 714, 659, **186**, 855, 789, 909, 574, 609, **186**, 448, 865, 899, **186**, 326, 275, 230, 502, 406, 958, 181, 392, 394, 950, 931, 451, 488, 439, 444, 355, 606, **186**, 508, 939, 554, 720, 384, **186**, 650, 183, 947, 284, 212, 375, 166, 154, 583, 771, 431, 511, 930, 362, **186**, 351, 224, 116, 12, 20, 863, **186**, 28, 506, 794, 457, 631, 161, 100, 555, 780, 236, 856, 322, 801, 108

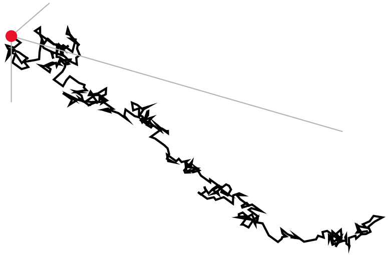
$J : \mathbb{R}^{331} \to \mathbb{R}^{d}$, i.i.d. $N(0, 1/d)$ entries

160, 755, 842, 300, 140, 324, 218, 880, 109, 486, 61,
805, 476, 815, 6, 733, 397, 44, 917, 521, 294, 727, 367
250, 541, 770, 890, 832, 259, **186**, 530, 977, 970, 268
715, 282, 774, 188, 901, 302, 802, 338, 926, 831, 627,
412, 697, 776, **186**, 779, 453, 581, 312, 25, 407, 471,
217, **186**, 170, 59, 858, 130, 684, 732, 434, 543, 615, 972, 335, 425, 525, 79, 192, 47, 588, 437, 357, 954, 396, 510, 86, 189,
474, 247, **186**, 965, 69, 308, 758, 593, 532, 98, **186**, 296, 985, 256, **186**, 633, 892, 27, 26, 316, 400, 257, 297, 516, 929, 908,
817, 428, 722, 826, 768, 750, 73, 888, 799, 620, **186**, 534, 522, 206, 129, 737, 526, 698, 957, 238, 49, 741, **186**, 597, 177, 202,
94, 323, **186**, 999, 204, 405, 778, 598, 489, **186**, 243, 937, 36, 240, 422, 232, 421, **186**, 33, 565, 804, 735, 278, 455, 759, 155,
653, 104, 352, 642, 784, 119, 209, 639, 991, 459, 163, 198, 121, 207, 274, 111, **186**, 96, 920, **186**, 184, 955, 449, **186**, 363, 195,
839, 413, 841, 582, 216, 71, 179, 107, 313, 244, 472, 752, 131, 952, 95, 356, 466, 679, 18, **186**, 859, 623, 330, 365, **186**, 822,
260, 442, 775, 714, 659, **186**, 855, 789, 909, 574, 609, **186**, 448, 865, 899, **186**, 326, 275, 230, 502, 406, 958, 181, 392, 394,
950, 931, 451, 488, 439, 444, 355, 606, **186**, 508, 939, 554, 720, 384, **186**, 650, 183, 947, 284, 212, 375, 166, 154, 583, 771,
431, 511, 930, 362, **186**, 351, 224, 116, 12, 20, 863, **186**, 28, 506, 794, 457, 631, 161, 100, 555, 780, 236, 856, 322, 801, 108

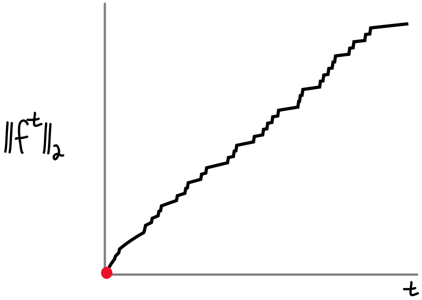
$\underline{Jf^t}$

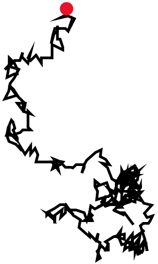


$\underline{Jf^t}$



$$X^t = \langle G, f^t \rangle, \quad t = 1, 2, \ldots, m$$



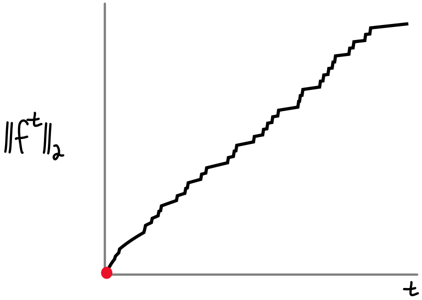


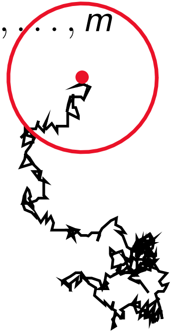
$$d(s,t) = \sqrt{\mathbb{E}(X^t - X^s)^2} = \|f^t - f^s\|_2$$

orthogonal $\equiv$ independent

far from orthogonal $\Longrightarrow$ card($\epsilon\|f\|_2$-net) $\leq 1/\epsilon^2$

$$X^t = \langle G, f^t \rangle, \quad t = 1, 2, \ldots, m$$



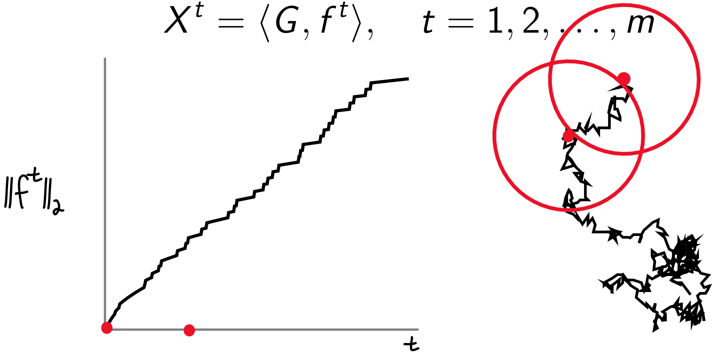


$$d(s,t) = \sqrt{\mathbb{E}(X^t - X^s)^2} = \|f^t - f^s\|_2$$

$$\text{orthogonal} \quad \equiv \quad \text{independent}$$

$$\text{far from orthogonal} \implies \text{card}(\epsilon \|f\|_2\text{-net}) \le 1/\epsilon^2$$

$$X^t = \langle G, f^t \rangle, \quad t = 1, 2, \dots, m$$

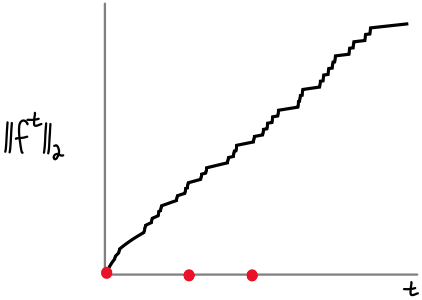


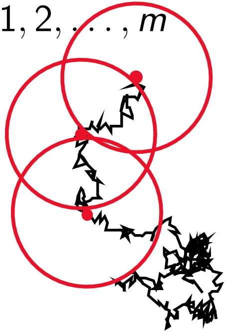
$$d(s,t) = \sqrt{\mathbb{E}(X^t - X^s)^2} = \|f^t - f^s\|_2$$

orthogonal $\equiv$ independent

far from orthogonal $\implies$ card($\epsilon\|f\|_2$-net) $\leq 1/\epsilon^2$

$$X^t = \langle G, f^t \rangle, \quad t = 1, 2, \ldots, m$$



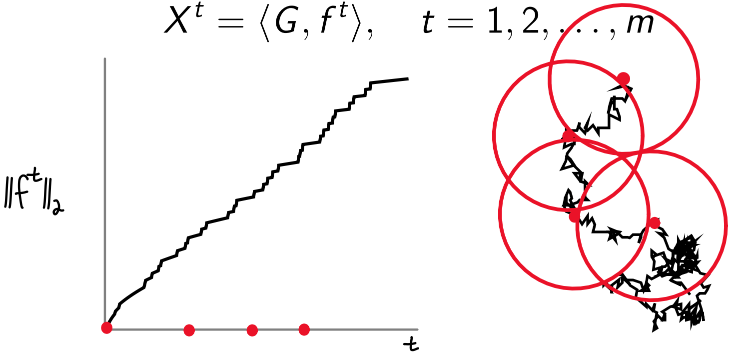


$$d(s,t) = \sqrt{\mathbb{E}(X^t - X^s)^2} = \|f^t - f^s\|_2$$

orthogonal $\equiv$ independent

far from orthogonal $\implies$ card($\epsilon\|f\|_2$-net) $\leq 1/\epsilon^2$

$$X^t = \langle G, f^t \rangle, \quad t = 1, 2, \ldots, m$$

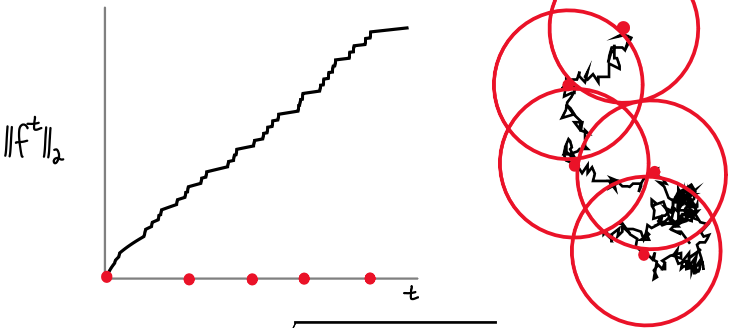


$$d(s,t) = \sqrt{\mathbb{E}(X^t - X^s)^2} = \|f^t - f^s\|_2$$

orthogonal $\equiv$ independent

far from orthogonal $\Longrightarrow$ card($\epsilon\|f\|_2$-net) $\leq 1/\epsilon^2$

$$X^t = \langle G, f^t \rangle, \quad t = 1, 2, \ldots, m$$



$$d(s,t) = \sqrt{\mathbb{E}(X^t - X^s)^2} = \|f^t - f^s\|_2$$

orthogonal $\equiv$ independent

far from orthogonal $\Longrightarrow$ card($\epsilon\|f\|_2$-net) $\leq 1/\epsilon^2$

T_i a $\frac{1}{2^{2^{i-1}}}\|f\|_2$ net of size $|T_i| \leq 2^{2^i}$, then

$$\begin{aligned}\mathbb{E}\sup_t \langle G, f^t\rangle &\lesssim \sup_t \sum_{i\geq 0} 2^{i/2} d(t, T_i) \\ &\leq \sum_{i\geq 0} 2^{i/2} \cdot 2^{-2^{i-1}} \|f\|_2 \\ &\lesssim \|f\|_2\end{aligned}$$

Fewer bits by embedding [Meka '12]

- $J : \{f^t\}_{t\in[m]} \to \mathbb{R}^{O(\log m)}$
- $O(1)$-distortion embedding (e.g. [Kane, Meka, Nelson '11])
- Replace $X^t = \langle f^t, Z\rangle$ by $\hat{X}^t = \langle Jf^t, Z\rangle$.

Lemma (Slepian)

Let X^t and $\hat{X}^t$ be Gaussian processes where

$$\mathbb{E}(\hat{X}^t - \hat{X}^s)^2 \leq \mathbb{E}(X^t - X^s)^2, \textit{ for all } t.$$

Then $\mathbb{E}\sup \hat{X}^t \leq \mathbb{E}\sup X^t$.

$$\mathbb{E}\sup_t \langle Z, Jf^t \rangle$$

$$\mathbb{E}\sup_t \langle Z, Jf^t \rangle \lesssim \mathbb{E}\sup_t \langle G, Jf^t \rangle$$

$$\mathbb{E}\,\sup_t\langle Z, Jf^t\rangle \lesssim \mathbb{E}\,\sup_t\langle G, Jf^t\rangle \lesssim \mathbb{E}\,\sup_t\langle G, f^t\rangle$$

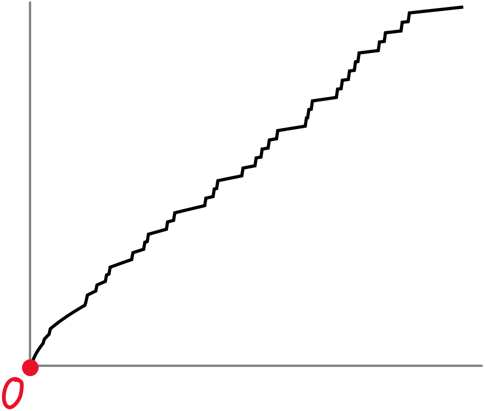
$$\mathbb{E}\,\sup_t \langle Z, Jf^t \rangle \lesssim \mathbb{E}\,\sup_t \langle G, Jf^t \rangle \lesssim \mathbb{E}\,\sup_t \langle G, f^t \rangle \lesssim \|f\|_2$$

There is another way...

Theorem (4-wise Chaining Inequality Nelson & Wang)

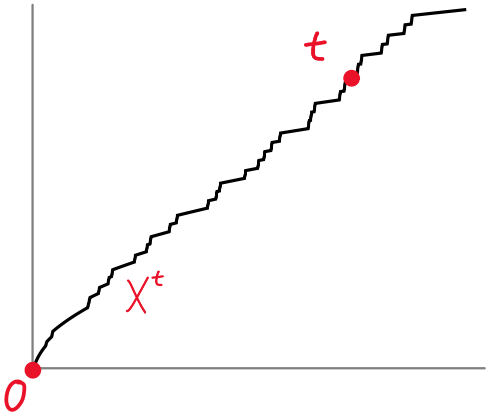
Let $\tilde{Z} \in \{-1, +1\}^n$ *have* 4-*wise independent entries. Then*

$$\mathbb{E} \sup_{1 \le t \le m} |\langle \tilde{Z}, f^t \rangle| \lesssim \|f\|_2.$$



Khintchine:

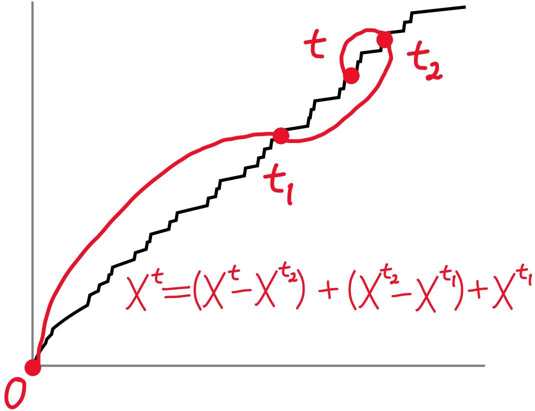
$$\mathbb{E}\langle Z, x\rangle^p \leq \sqrt{p}^p \|x\|_2^p$$
$$\mathbb{E}\langle \tilde{Z}, x\rangle^4 \leq 16\|x\|_2^4$$



Khintchine:

$$\mathbb{E}\langle Z, x\rangle^p \leq \sqrt{p}^p \|x\|_2^p$$

$$\mathbb{E}\langle \tilde{Z}, x\rangle^4 \leq 16\|x\|_2^4$$



Khintchine:

$$\mathbb{E}\langle Z, x\rangle^p \leq \sqrt{p}^p \|x\|_2^p$$

$$\mathbb{E}\langle \tilde{Z}, x\rangle^4 \leq 16\|x\|_2^4$$

T_i a $\frac{1}{2^i}\|f\|_2$ net of size $|T_i| \leq 2^{2i}$, then

$$\mathbb{E}\sup_t \langle \tilde{Z}, f^t \rangle \lesssim \sum_{i\geq 1} (|T_{i-1}| \cdot |T_i|)^{1/4} \sup_t d(t, T_i)$$
$$\leq \sum_{i\geq 1} 2^{4i/4} \cdot \frac{1}{2^i}\|f\|_2$$

T_i a $\frac{1}{2^i}\|f\|_2$ net of size $|T_i| \leq 2^{2i}$, then

$$\mathbb{E}\sup_t \langle \tilde{Z}, f^t \rangle \lesssim \sum_{i \geq 1} (\qquad 2 \cdot |T_i|)^{1/4} \sup_t d(t, T_i)$$

$$\lesssim \sum_{i \geq 1} 2^{2i/4} \cdot \frac{1}{2^i} \|f\|_2$$

T_i a $\frac{1}{2^i}\|f\|_2$ net of size $|T_i| \leq 2^{2^i}$, then

$$\begin{aligned}
\mathbb{E}\sup_t \langle \tilde{Z}, f^t\rangle &\lesssim \sum_{i\geq 1} (\qquad 2 \cdot |T_i|)^{1/4} \sup_t d(t, T_i) \\
&\lesssim \sum_{i\geq 1} 2^{2i/4} \cdot \frac{1}{2^i}\|f\|_2 \\
&\lesssim \|f\|_2
\end{aligned}$$

Summary

HH in $O(\log n)$ bits and $O(1)$ update time