Chapter 6

Sampling and other Stuff

By Sariel Har-Peled, December 30, 2015[①]

6.1. Two-Point Sampling

Definition 6.1.1. A collection of random variables $X_1, \ldots, X_n$ is ***pairwise-independent***, if for any pair of variables X_i and X_j, and any pair of values α and β we have that $\mathbf{Pr}\Big[X_i = \alpha \cap X_j = \beta\Big] = \mathbf{Pr}\Big[X_i = \alpha\Big]\mathbf{Pr}\Big[X_j = \beta\Big]$.

Similarly, this collection is ***k-wise independent***, if for any $t \leq k$ variables $X_{i_1}, \ldots, X_{i_t}$ in this collection, and any set of t values, $\alpha_1, \ldots, \alpha'_t$ we have that

$$\mathbf{Pr}[(X_{i_1} = \alpha_1) \cap \ldots \cap (X_{i_t} = \alpha_t)] = \prod_{j=1}^{t} \mathbf{Pr}\Big[X_{i_j} = \alpha_j\Big].$$

Namely, pairwise independent variables behaves like independent random variables as long as you look only in pairs.

Example 6.1.2. Consider the probability space show on the right, where the triple of variables X, Y, Z can be assigned any of the rows with equal probability (i.e., 1/4).

X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	0

Clearly, for any $\alpha, \beta \in \{0, 1\}$ we have $\mathbf{Pr}[(X = \alpha) \cap (Y = \beta)] = \mathbf{Pr}[(X = \alpha)]\,\mathbf{Pr}[(Y = \beta)] = 1/4$ (this also holds for X, Z and Y, Z). Namely, X, Y, Z are all pairwise independent. However, they are not 3-wise independent (or just independent). Indeed, we have $\mathbf{Pr}[(X = 1) \cap (Y = 1) \cap (Z = 1)] = 0$, while it should have been $1/8$ if they were truly independent, or even just 3-wise independent.

6.1.1. About Modulo Rings and Pairwise Independence

Let p be a prime number, and let $\mathbb{Z}_p = \Big\{0, 1, \ldots, p-1\Big\}$ denote the ring of integers modules p. Two integers x and y are ***equivalent modulo*** p, if $x \equiv y \mod p$; namely, the reminder of dividing x and y by p is the same.

Lemma 6.1.3. *Given $y, i \in \mathbb{Z}_p$, and choosing a and b randomly, independently and uniformly from $\mathbb{Z}_p$, the probability of $y \equiv ai + b \pmod p$ is $1/p$.*

Proof: Imagine that we first choose a, then the required probability, is that we choose b such that $y - ai \equiv b \pmod p$. And the probability for that is $1/p$, as we choose b uniformly. ■

Lemma 6.1.4. *Let p be a prime, and fix $a \in \{1, \ldots, p-1\}$. Then, $\left\{ai \pmod p \;\middle|\; i = 0, \ldots, p-1\right\} = \mathbb{Z}_p$.*

Putting it differently, for any non-zero $a \in \mathbb{Z}_p$, there is a unique inverse $b \in \mathbb{Z}_p$ such that $ab \pmod p = 1$.

Proof: Assume, for the sake of contradiction, that the claim is false. Then, by the pigeon hole principle, there must exists $1 \le j < i \le p-1$ such that $ai \pmod p = aj \pmod p$. Namely, there are k', k, u such that

$$ai = u + kp \quad \text{and} \quad aj = u + k'p.$$

(Here, we know that $0 \le k < p$, $0 \le k' < p$ and $0 \le u < p$.) Since $i > j$ it must be that $k > k'$. Subtracting the two equalities, we get that $a(i-j) = (k-k')p > 0$. Now, $i - j$ must be larger than one, since if $i - j = 1$ then $a = p$, which is impossible. Similarly, $i - j < p$. Also, $i - j$ can not divide p, since p is a prime. Thus, it must be that $i - j$ must divide $k - k'$. So, let us set $\beta = (k-k')/(i-j) \ge 1$. This implies that $a = \beta p \ge p$, which is impossible. Thus, our assumption is false. ■

Lemma 6.1.5. *Given $y, z, x, w \in \mathbb{Z}_p$, such that $x \neq w$, and choosing a and b randomly and uniformly from $\mathbb{Z}_p$, the probability that $y \equiv ax + b \pmod p$ and $z = aw + b \pmod p$ is $1/p^2$.*

Proof: This equivalent to claiming that the system of equalities $y \equiv ax + b \pmod p$ and $z = aw + b$ have a unique solution in a and b.

To see why this is true, subtract one equation from the other. We get $y - z \equiv a(x-w) \pmod p$. Since $x - w \not\equiv 0 \pmod p$, it must be that there is a unique value of a such that the equation holds. This in turns, imply a specific value for b. The probability that a and b get those two specific values is $1/p^2$. ■

Lemma 6.1.6. *Let i and j be two distinct elements of $\mathbb{Z}_p$. And choose a and b randomly and independently from $\mathbb{Z}_p$. Then, the two random variables $Y_i = ai + b \pmod p$ and $Y_j = aj + b \pmod p$ are uniformly distributed on $\mathbb{Z}_p$, and are pairwise independent.*

Proof: The claim about the uniform distribution follows from Lemma 6.1.3, as $\mathbf{Pr}[Y_i = \alpha] = 1/p$, for any $\alpha \in \mathbb{Z}_p$. As for being pairwise independent, observe that

$$\mathbf{Pr}\Big[Y_i = \alpha \;\Big|\; Y_j = \beta\Big] = \frac{\mathbf{Pr}\Big[Y_i = \alpha \cap Y_j = \beta\Big]}{\mathbf{Pr}\Big[Y_j = \beta\Big]} = \frac{1/n^2}{1/n} = \frac{1}{n} = \mathbf{Pr}\Big[Y_i = \alpha\Big],$$

by Lemma 6.1.3 and Lemma 6.1.5. Thus, Y_i and Y_j are pairwise independent. ■

Remark 6.1.7. It is important to understand what independence between random variables mean: having information about the value of X, gives you no information about Y. But this is only pairwise independence. Indeed, consider the variables Y_1, Y_2, Y_3, Y_4 defined above. Every pair of them are pairwise independent. But, given the values of Y_1 and Y_2, one can compute the value of Y_3 and Y_4 immediately. Indeed, giving the value of Y_1 and Y_2 is enough to figure out the value of a and b. Once we know a and b, we immediately can compute all the Y_is.

Thus, the notion of independence can be extended to k-pairwise independence of n random variables, where only if you know the value of k variables, you can compute the value of all the other variables. More on that later in the course.

Lemma 6.1.8. *If X and Y are pairwise independent then $\mathbf{E}\Big[XY\Big] = \mathbf{E}\Big[X\Big]\,\mathbf{E}\Big[Y\Big]$.*

Proof: By definition, $\mathbf{E}\Big[XY\Big] = \sum_{x,y} xy\,\mathbf{Pr}\Big[(X=x)\cap(Y=y)\Big] = \sum_{x,y} xy\,\mathbf{Pr}\Big[X=x\Big]\mathbf{Pr}\Big[Y=y\Big] = \sum_x x\,\mathbf{Pr}[X=x]$ $\sum_y y\,\mathbf{Pr}[Y=y] = (\sum_x x\,\mathbf{Pr}[X=x])\Big(\sum_y y\,\mathbf{Pr}[Y=y]\Big) = \mathbf{E}\Big[X\Big]\,\mathbf{E}\Big[Y\Big]$. ■

Lemma 6.1.9. *Let $X_1, X_2, \ldots, X_n$ be pairwise independent random variables, and $X = \sum_{i=1}^n X_i$. Then $\mathbf{V}\Big[X\Big] = \sum_{i=1}^n \mathbf{V}\Big[X_i\Big]$.*

Proof: Observe, that $\mathbf{V}\Big[X\Big] = \mathbf{E}\Big[(X-\mathbf{E}[X])^2\Big] = \mathbf{E}\Big[X^2\Big] - \Big(\mathbf{E}[X]\Big)^2$. Let X and Y be pairwise independent variables. Observe that $\mathbf{E}\Big[XY\Big] = \mathbf{E}[X]\,\mathbf{E}[Y]$, as can be easily verified. Thus,

$$\begin{aligned}
\mathbf{V}\Big[X+Y\Big] &= \mathbf{E}\Big[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2\Big] \\
&= \mathbf{E}\Big[\Big(X+Y\Big)^2 - 2\Big(X+Y\Big)\Big(\mathbf{E}[X]+\mathbf{E}[Y]\Big) + \Big(\mathbf{E}[X]+\mathbf{E}[Y]\Big)^2\Big] \\
&= \mathbf{E}\Big[(X+Y)^2\Big] - \Big(\mathbf{E}[X]+\mathbf{E}[Y]\Big)^2 \\
&= \mathbf{E}\Big[X^2+2XY+Y^2\Big] - \Big(\mathbf{E}[X]\Big)^2 - 2\,\mathbf{E}\Big[X\Big]\,\mathbf{E}\Big[Y\Big] - (\mathbf{E}[Y])^2 \\
&= \Big(\mathbf{E}\Big[X^2\Big] - \Big(\mathbf{E}[X]\Big)^2\Big) + \Big(\mathbf{E}\Big[Y^2\Big] - \Big(\mathbf{E}[Y]\Big)^2\Big) + 2\,\mathbf{E}\Big[XY\Big] - 2\,\mathbf{E}\Big[X\Big]\,\mathbf{E}\Big[Y\Big] \\
&= \mathbf{V}\Big[X\Big] + \mathbf{V}\Big[Y\Big] + 2\,\mathbf{E}\Big[X\Big]\,\mathbf{E}\Big[Y\Big] - 2\,\mathbf{E}\Big[X\Big]\,\mathbf{E}\Big[Y\Big] \\
&= \mathbf{V}\Big[X\Big] + \mathbf{V}\Big[Y\Big],
\end{aligned}$$

by Lemma 6.1.8. Using the above argumentation for several variables, instead of just two, implies the lemma. ■

6.1.1.1. Generating k-wise independent variable

Consider the polynomial $f(x) = \sum_{i=0}^{k-1} \alpha_i x^i$ evaluated modulo p, where the coefficients $\alpha_0, \ldots, \alpha_{k-1}$ are taken from $\mathbb{Z}_p$. We claim that $f(0), f(1), \ldots, f(p-1)$ are k-wise independent. Indeed, for any k indices $i_1, \ldots, i_k \in \mathbb{Z}_p$, and k values $v_1, \ldots, v_k \in \mathbb{Z}_p$, we have that $\beta = \mathbf{Pr}[f(i_1) = v_1 \text{ and } \ldots \text{ and } f(i_k) = v_k]$ happens only for one specific choice of the αs, which implies that this probability is $1/p^k$, which is what we need.

6.1.2. Application: Using less randomization for a randomized algorithm

We can consider a randomized algorithm, to be a deterministic algorithm $\mathbf{Alg}(x, r)$ that receives together with the input x, a random string r of bits, that it uses to read random bits from. Let us redefine **RP**:

Definition 6.1.10. The class **RP** (for Randomized Polynomial time) consists of all languages L that have a deterministic algorithm $\mathbf{Alg}(x, r)$ with worst case polynomial running time such that for any input $x \in \Sigma^*$,

- $x \in L \implies \mathbf{Alg}(x, r) = 1$ for half the possible values of r.
- $x \notin L \implies \mathbf{Alg}(x, r) = 0$ for all values of r.

Let assume that we now want to minimize the number of random bits we use in the execution of the algorithm (Why?). If we run the algorithm t times, we have confidence 2^{-t} in our result, while using $t \log n$ random bits (assuming our random algorithm needs only $\log n$ bits in each execution). Similarly, let us choose two random numbers from $\mathbb{Z}_n$, and run $\mathbf{Alg}(x, a)$ and $\mathbf{Alg}(x, b)$, gaining us only confidence $1/4$ in the correctness of our results, while requiring $2 \log n$ bits.

Can we do better? Let us define $r_i = ai + b \mod n$, where a, b are random values as above (note, that we assume that n is prime), for $i = 1, \ldots, t$. Thus $Y = \sum_{i=1}^{t} \mathbf{Alg}(x, r_i)$ is a sum of random variables which are pairwise independent, as the r_i are pairwise independent. Assume, that $x \in L$, then we have $\mathbf{E}[Y] = t/2$, and $\sigma_Y^2 = \mathbf{V}[Y] = \sum_{i=1}^{t} \mathbf{V}[\mathbf{Alg}(x, r_i)] \leq t/4$, and $\sigma_Y \leq \sqrt{t}/2$. The probability that all those executions failed, corresponds to the event that $Y = 0$, and

$$\mathbf{Pr}\Big[Y = 0\Big] \leq \mathbf{Pr}\Big[\big|Y - \mathbf{E}[Y]\big| \geq \frac{t}{2}\Big] = \mathbf{Pr}\Big[\big|Y - \mathbf{E}[Y]\big| \geq \frac{\sqrt{t}}{2} \cdot \sqrt{t}\Big] \leq \frac{1}{t},$$

by the Chebyshev inequality. Thus we were able to "extract" from our random bits, much more than one would naturally suspect is possible. We thus get the following result.

Lemma 6.1.11. *Given an algorithm* **Alg** *in* **RP** *that uses* $\lg n$ *random bits, one can run it* t *times, such that the runs results in a new algorithm that fails with probability at most* $1/t$.

6.2. QuickSort is quick via direct argumentation

Consider a specific element α in the input array of n elements that is being sorted by **QuickSort**, and let X_i be the size of the recursive subproblem in the ith level of the recursion that contains x. If x thus not participate in such a subproblem in this level, that $X_i = 0$. It is easy to verify that

$$X_0 = n \quad \text{and} \quad \mathbf{E}\Big[X_i \,\Big|\, X_{i-1}\Big] \leq \frac{1}{2} \cdot \frac{3}{4} X_{i-1} + \frac{1}{2} X_{i-1} \leq \frac{7}{8} X_{i-1}.$$

As such, $\mathbf{E}[X_i] = \mathbf{E}\Big[\mathbf{E}[X_i]\Big] = (7/8)^i n$. In particular, we have by Markov's inequality that

$$\mathbf{Pr}\left[\begin{array}{c}\alpha \text{ participates in more than} \\ c \ln n \text{ levels of the recursion}\end{array}\right] = \mathbf{Pr}[X_{c \ln n} \geq 1] \leq \frac{\mathbf{E}[X_{c \ln n}]}{1} \leq (7/8)^{c \ln n} n \leq \frac{1}{n^{\beta+1}},$$

if $\Big(c \ln(8/7)\Big) \ln n \geq \beta \ln n \iff c \geq \beta / \ln(8/7)$. We conclude the following.

Theorem 6.2.1. *For any* $\beta \geq 1$*, we have that the running time of* **QuickSort** *sorting* n *elements is* $O(\beta n \log n)$*, with probability* $\geq 1 - 1/n^{\beta}$.

Proof: For $c = \beta / \ln(8/7)$, the probability that an element participates in at most $c \ln n$ levels of the recursion is at most $1/n^{\beta+1}$. Since there are n elements, by the union bound, this bounds the probability that any input number would participate in more than $c \ln n$ recursive calls. But that implies that the recursion depth of **QuickSort** is $\leq c \ln n$, which immediately implies the claim. ∎

What the above proof shows is that an element can not be too unlucky – if it participates in enough rounds, then, with high probability, the subproblem containing it would shrink significantly. This fairness of luck is one of the most important principles in randomized algorithms, and we next formalize it by proving a rather general theorem on the "concentration" of luck.