Lecture 5: "Two-point Sampling"

Sotiris Nikoletseas
Professor

CEID - ETY Course
2017 - 2018

Overview

A. Pairwise independence of random variables

B. The pairwise independent sampling theorem

C. Probability amplification via reduced randomness

A. On the Additivity of Variance

- In general the variance of a sum of random variables is not equal to the sum of their variances
- However, variances do add for independent variables (i.e. mutually independent variables)
- In fact, mutual independence is not necessary and pairwise independence suffices
- This is very useful, since in many situations the random variables involved are pairwise independent but not mutually independent.

Conditional distributions

- Let X, Y be discrete random variables. Their joint probability density function is
$$f(x,y) = \Pr\{(X=x) \cap (Y=y)\}$$
- Clearly $f_1(x) = \Pr\{X=x\} = \sum_y f(x,y)$
and $f_2(y) = \Pr\{Y=y\} = \sum_x f(x,y)$
- Also, the conditional probability density function is:
$$f(x|y) = \Pr\{X=x|Y=y\} = \frac{\Pr\{(X=x) \cap (Y=y)\}}{\Pr\{Y=y\}} =$$
$$= \frac{f(x,y)}{f_2(y)} = \frac{f(x,y)}{\sum_x f(x,y)}$$

Pairwise independence

- Let random variables $X_1, X_2, \ldots, X_n$. These are called pairwise independent iff for all $i \neq j$ it is
$$\Pr\{(X_i = x)|(X_j = y)\} = \Pr\{X_i = x\}, \forall x, y$$
Equivalently, $\Pr\{(X_i = x) \cap (X_j = y)\} =$
$$= \Pr\{X_i = x\} \cdot \Pr\{X_j = y\}, \forall x, y$$

- Generalizing, the collection is k-wise independent iff, for every subset $I \subseteq \{1, 2, \ldots, n\}$ with $|I| < k$ for every set of values $\{a_i\}, b$ and $j \notin I$, it is
$$\Pr\left\{ X_j = b | \bigwedge_{i \in I} X_i = a_i \right\} = \Pr\{X_j = b\}$$

Mutual (or "full") independence

- The random variables $X_1, X_2, \ldots, X_n$ are mutually independent iff for any subset $X_{i_1}, X_{i_2}, \ldots, X_{i_k}, (2 \leq k \leq n)$ of them, it is
$\Pr\{(X_{i_1} = x_1) \cap (X_{i_2} = x_2) \cap \cdots \cap (X_{i_k} = x_k)\} =$
$= \Pr\{X_{i_1} = x_1\} \cdot \Pr\{X_{i_2} = x_2\} \cdots \Pr\{X_{i_k} = x_k\}$

- Example (for $n = 3$). Let A_1, A_2, A_3 3 events. They are mutually independent iff all four equalities hold:

$$\Pr\{A_1 A_2\} = \Pr\{A_1\} \Pr\{A_2\} \quad (1)$$
$$\Pr\{A_2 A_3\} = \Pr\{A_2\} \Pr\{A_3\} \quad (2)$$
$$\Pr\{A_1 A_3\} = \Pr\{A_1\} \Pr\{A_3\} \quad (3)$$
$$\Pr\{A_1 A_2 A_3\} = \Pr\{A_1\} \Pr\{A_2\} \Pr\{A_3\} \quad (4)$$

They are called pairwise independent if (1), (2), (3) hold.

Mutual vs pairwise independence

- Important notice: Pairwise independence does not imply mutual independence in general.

- Example. Let a probability space including all permutations of a, b, c as well as aaa, bbb, ccc (all 9 points considered equiprobable). Let

$$A_k = \text{“at place } k \text{ there is an a”} \text{ (for } k = 1, 2, 3).$$

It is $\Pr\{A_1\} = \Pr\{A_2\} = \Pr\{A_3\} = \frac{2+1}{9} = \frac{1}{3}$

Also $\Pr\{A_1A_2\} = \Pr\{A_2A_3\} = \Pr\{A_1A_3\} = \frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3}$

thus A_1, A_2, A_3 are pairwise independent. But

$$\Pr\{A_1A_2A_3\} = \frac{1}{9} \neq \Pr\{A_1\}\Pr\{A_2\}\Pr\{A_3\} = \frac{1}{27}$$

thus the events are not mutually independent

Variance: key features

- Definition:
  $Var(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 \Pr\{X = x\}$
  where $\mu = E[X] = \sum_x x \Pr\{X = x\}$

- We call standard deviation of X the $\sigma = \sqrt{Var(X)}$

- Basic Properties:
  (i) $Var(X) = E[X^2] - E^2[X]$
  (ii) $Var(cX) = c^2 Var(X)$, where c constant.
  (iii) $Var(X + c) = Var(X)$, where c constant.

- proof of (i):
  $Var(X) = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] + E[-2\mu X] + E[\mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$

The additivity of variance

Theorem: if $X_1, X_2, \ldots, X_n$ are pairwise independent random variables, then: $$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Proof:

$$Var(X_1 + \cdots + X_n) = E[(X_1 + \cdots + X_n)^2] - E^2[X_1 + \cdots + X_n] =$$

$$= E\left[\sum_{i=1}^{n} X_i^2 + \sum_{1 \le i \ne j \le n} X_i X_j\right] - \left(\sum_{i=1}^{n} \mu_i^2 + \sum_{1 \le i \ne j \le n} \mu_i \mu_j\right) =$$

$$= \sum_{i=1}^{n} (E[X_i^2] - \mu_i^2) + \sum_{1 \le i \ne j \le n} (E[X_i X_j] - \mu_i \mu_j) = \sum_{i=1}^{n} Var(X_i)$$

(since X_i pairwise independent, so $\forall 1 \le i \ne j \le n$
$E(X_i X_j) = E(X_i) E(X_j = \mu_i \mu_j)$ $\square$

Note: As we see in the proof, the pairwise independence suffices, and mutual (full) independence is not needed.

B. The pairwise independent sampling theorem

Another Example. Birthday matching: Let us try to estimate the number of pairs of people in a room having birthday on the same day.

- Note 1: Matching birthdays for different pairs of students are pairwise independent, since knowing that (George, Takis) have a match tell us nothing about whether (George, Petros) match.

- Note 2: However, the events are not mutually independent. Indeed they are not even 3-wise independent since if (George,Takis) match and (Takis, Petros) match then (George, Petros) match!

Birthday matching

- Let us calculate the probability of having a certain number of birthday matches.
- Let $B_1, B_2, \ldots, B_n$ the birthdays of n independently chosen people and let $\mathcal{E}_{i,j}$ be the indicator variable for the event of a (i,j) match (i.e. $B_i = B_j$). As said, the events $\mathcal{E}_{i,j}$ are pairwise independent but not mutually independent.
- Clearly, $\Pr\{\mathcal{E}_{i,j}\} = \Pr\{B_i = B_j\} = 365 \frac{1}{365} \frac{1}{365} = \frac{1}{365}$ (for $i \neq j$).
  Let D the number of matching pairs. Then
  $$D = \sum_{1 \leq i < j \leq n} \mathcal{E}_{i,j}$$
  By linearity of expectation we have
  $$E[D] = E\left[\sum_{1 \leq i < j \leq n} \mathcal{E}_{i,j}\right] = \sum_{1 \leq i < j \leq n} E[\mathcal{E}_{i,j}] = \binom{n}{2} \frac{1}{365}$$

Birthday matching

- Since the variances of pairwise independent variables $\mathcal{E}_{i,j}$ add up, it is:

$$Var[D] = Var\left[\sum_{1\le i<j\le n} \mathcal{E}_{i,j}\right] = \sum_{1\le i<j\le n} Var[\mathcal{E}_{i,j}] =$$

$$= \binom{n}{2}\frac{1}{365}(1-\frac{1}{365})$$

- As an example, for a class of $n = 100$ students, it is $E[D] \simeq 14$ and $Var[D] < 14(1-\frac{1}{365}) < 14$. So by Chebyshev's inequality we have

$$\Pr\{|D-14| \ge x\} \le \frac{14}{x^2}$$

Letting $x = 6$, we conclude that with more than 50% chance the number of matching birthdays will be between 8 and 20.

The Pairwise Independent Sampling Theorem (I)

- We can actually generalize and not restrict to sums of zero-one (indicator) valued variables neither to variables with the same distribution. We below state the theorem for possibly different distributions with same mean and variance (but this is done for simplicity, and the result holds for distributions with different means and/or variances as well).

- Theorem. Let $X_1, \ldots, X_n$ pairwise independent variables with the same mean μ and variance σ^2. Let

$$S_n = \sum_{i=1}^{n} X_i$$

Then $\Pr\left\{\left|\frac{S_n}{n} - \mu\right| \geq x\right\} \leq \frac{1}{n}\left(\frac{\sigma}{x}\right)^2$

Proof. Note that $E[\frac{S_n}{n}] = \frac{n\mu}{n} = \mu$,
$Var[\frac{S_n}{n}] = \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$ and apply Chebyshev's inequality. □

The Pairwise Independent Sampling Theorem (II)

Note: This Theorem actually provides a precise general evaluation of how the average of pairwise independent random samples approaches their mean. If the number n of samples becomes large enough we can arbitrarily close approach the mean with confidence arbitrarily close to 100% ($n > \frac{\sigma^2}{x^2}$) i.e. a large number of samples is needed for distributions of large variance and when we want to assure high concentration around the mean).

C. Reduced randomness at probability amplification

Motivation:

- Randomized Algorithms, for a given input x, actually choose n random numbers ("witnesses") and run a deterministic algorithm on the input, using each of these random numbers.
- intuitively, if the deterministic algorithm has a probability of error ϵ (e.g. $\frac{1}{2}$), t independent runs reduce the error probability to ϵ^t(e.g. $\frac{1}{2^t}$) and amplify the correctness probability from $\frac{1}{2}$ to $1 - \frac{1}{2^t}$.
- however, true randomness is quite expensive! What happens if we are constrained to use no more than a constant c random numbers? The simplest case is when $c = 2$ e.g. we choose just 2 random numbers (thus the name two-point sampling)

C. Reduced randomness at probability amplification

Problem definition:

If our randomized algorithm reduced the error probability to ϵ^t with t random numbers, what can we expect about the error probability with 2 random numbers only?

- an obvious bound is ϵ^2(e.g. when $\epsilon = \frac{1}{2}$ reducing the error probability from $\frac{1}{2}$ to $\frac{1}{4}$)
- can we do any better?
- it turns out that we can indeed do much better and reduce the error probability to $\frac{1}{t}$ (which is much smaller than the constant ϵ^2)

C. Reduced randomness at probability amplification

High level idea:

- generate t ("pseudo-random") numbers based on the chosen 2 truly random numbers and use them in lieu of t truly independent random numbers.
- these generated numbers are dependent on the 2 chosen numbers. Hence they are not independent, but are pairwise independent.
- This loss of full independence reduces the accuracy of the algorithmic process but is still quite usable in reducing the error probability.

C. Reduced randomness at probability amplification

The process (high level description):

1. Choose a large prime number p
   (e.g. Mersenne prime, $2^{31} - 1$).
2. Define Z_p as the ring of integers modulo p
   (e.g $0, 1, 2, \ldots, 2^{31} - 2$)
3. Choose 2 truly random numbers, a and b from Z_p
   (e.g. $a = 2^{20} + 781$ and $b = 2^{27} - 44$).
4. Generate t "pseudo-random" numbers $y_i = (ai + b) \bmod p$
   e.g. $y_0 = 2^{27} - 44$
   $y_1 = (2^{20} + 781) \cdot 1 + 2^{27} - 44$
   $y_2 = (2^{20} + 781) \cdot 2 + 2^{27} - 44$ and so on.
5. Use each of the y_i's as t witnesses (in lieu of t purely independent random witnesses)

C. Reduced randomness at probability amplification

Performance (high level discussion)

As we will see:

- for any given error bound ϵ, the error probability is reduced from ϵ^2 to $\frac{1}{t}$.
- however, instead of requiring $O(\log \frac{1}{\epsilon})$ runs on the deterministic algorithm (in the case of t independent random witnesses) we will require $O(\frac{1}{\epsilon})$ runs in the case of 2 independent random witnesses.
- thus, we gain in the probability amplification but loose on some efficiency.
- we need significantly less true randomness.

The class RP (I)

Definition. The class RP (Random Polynomial time) consists of all languages L admitting a randomized algorithm A running in worst case polynomial time such that for any input x:

- $x \in L \Rightarrow \Pr\{A(x) \text{ accepts}\} \geq \frac{1}{2}$
- $x \notin L \Rightarrow \Pr\{A(x) \text{ accepts}\} = 0$

Notes:

- language recognition $\longleftrightarrow$ computational decision problems
- the $\frac{1}{2}$ value is arbitrary. The success probability needs just to be lower-bounded by an inverse polynomial function of the input size (a polynomial number of algorithm repetitions would boost it to constant, in polynomial time).
- RP actually includes Monte Carlo algorithms than can err only when $x \in L$ (one-sided error)

The RP algorithm

- An RP algorithm, for given input x, actually picks a random number r from Z_p and computes $A(x, r)$ with the following properties:
  - $x \in L \Rightarrow A(x, r) = 1$, for half of the possible values of r
  - $x \notin L \Rightarrow A(x, r) = 0$, for all possible choices of r

- As said,
  - if we run algorithm A t times on the same input x, the error probability is $\leq \frac{1}{2^t}$ but this requires t truly random numbers.
  - if we restrict ourselves to $t = 2$ true random numbers a, b from Z_p and run $A(x, a)$ and $A(x, b)$, the error probability can be as high as $\frac{1}{4}$
  - but we can do better with t pseudo-random numbers: Let $r_i = a \cdot i + b \pmod{p}$, where a, b are truly random, as above, for $i = 1, 2, \ldots, t$.

Modulo rings pairwise independence (I)

Let p a prime number and $Z_p = \{0, 1, 2, \ldots, p-1\}$ denote the ring of the integers modulo p.

Lemma 1. Given $y, i \in Z_p$ and choosing a, b randomly uniformly from Z_p, the probability of $y \equiv a \cdot i + b \pmod{p}$ is $\frac{1}{p}$

Proof. Imagine that we first choose a. Then when choosing b, it must be $y - a \cdot i \equiv b \pmod{p}$. Since we choose b uniformly and modulo p can take p values, the probability is clearly $\frac{1}{p}$ indeed.

□

Modulo rings pairwise independence (II)

Lemma 2. Given $y, z, x, w \in Z_p$ such that $x \neq w$, and choosing a, b randomly uniformly from Z_p, the probability of $y \equiv a \cdot x + b \pmod{p}$ and $z \equiv a \cdot w + b \pmod{p}$ is $\frac{1}{p^2}$

Proof. It is $y - z \equiv a \cdot (x - w) \pmod{p}$. Since $x - w \neq 0$, the equation holds for a unique value of a. This in turn implies a specific value for b. The probability that a, b get those two specific values is clearly $\frac{1}{p} \cdot \frac{1}{p} = \frac{1}{p^2}$. □

Modulo rings pairwise independence (III)

Lemma 3. Let i, j two distinct elements of Z_p,and choose a, b randomly uniformly from Z_p. Then the two variables $Y_i = a \cdot i + b \pmod{p}$ and $Y_j = a \cdot j + b \pmod{p}$ are uniformly distributed on Z_p and are pairwise independent.

Proof. From lemma 1, it is clearly $\Pr\{Y_i = \alpha\} = \frac{1}{p}$, for any $\alpha \in Z_p$, so Y_i, Y_j indeed have uniform distribution. As for pairwise independence, note that:

$$\Pr\{Y_i = \alpha | Y_j = \beta\} = \frac{\Pr\{Y_i = \alpha \cap Y_j = \beta\}}{\Pr\{Y_j = \beta\}} = \frac{\frac{1}{p^2}}{\frac{1}{p}} = \Pr\{Y_i = \alpha\}$$

Thus, Y_i, Y_j are pairwise independent. □

Modulo rings pairwise independence (IV)

Note: This is only pairwise independence. Indeed, consider the variables Y_1, Y_2 Y_3, Y_4 as defined above. Every pair of them are pairwise independent. But, if you give the value of Y_1, Y_2 then we know the values of Y_3, Y_4 immediately, since the values of Y_1 and Y_2 uniquely determine a and b, and thus we can compute all Y_i variables.

Modulo rings pairwise independence (V)

- Thus $Y = \sum_{i=1}^{t} A(x, r_i)$ is a sum of random variables which are pairwise independent, since the r_i values are pairwise independent. Assume that $x \in L$, then $E[Y] = \frac{t}{2}$ and $Var[Y] = \sum_{i=1}^{t} Var[A(x, r_i)] = t\frac{1}{2}\frac{1}{2} = \frac{t}{4}$
- The probability that all those t executions failed corresponds to the event $Y = 0$, and $\Pr\{Y = 0\} \leq \Pr\{|Y - E[Y]| \geq E[Y]\} = \Pr\{|Y - \frac{t}{2}| \geq \frac{t}{2}\} \leq$ $\leq \frac{Var[Y]}{\left(\frac{t}{2}\right)^2} = \frac{\frac{t}{4}}{\left(\frac{t}{2}\right)^2} = \frac{1}{t}$ from Chebyshev's inequality.
- Thus the use of pseudo-randomness (t pseudo-random numbers) allows to "exploit" our 2 random bits to reduce the error probability from $\frac{1}{4}$ to $\frac{1}{t}$.