Lecture 5: Notions of independence and applications

Rajat Mittal

IIT Kanpur

First, let's revisit some of the concepts introduced in conditional probability. There are many situations where an event affects the probability of another event. Your chances of winning a lottery might severely decrease or increase given that the winning number is even. Your chances of coming to the class might depend on whether your friend is in the class or not.

To capture these situations, we defined *conditional probability.* Given two events A, B, the conditional probability of A given B is defined by,

$$P\left(\frac{A}{B}\right) := \frac{P(A \cap B)}{P(B)}.$$

We also defined the related concept of independent events.

1 Independence

Exercise 1. What is the probability that the two throws of a dice give the same number? What if we know that the sum of numbers displayed is prime? What if we know that the number displayed on the first dice is prime? What if we know that the sum is less than 5?

We notice the two extremes; if sum of numbers is prime then they can't be equal. On the other hand, the number on the first dice being prime has no effect on them being equal. The later situation can be thought as, event A of the first number being prime is completely independent to event B of them being equal. Again, we formalize this notion with the concept of *independence* between events.

Let us take a few more examples.

1. What is the probability of obtaining a head again if we knew that the first toss gave a head, while tossing an unbiased coin twice.
2. Suppose Euler misses school on a day with probability 1/2. What is the probability that he misses school twice on two consecutive days?

Given that the coin is unbiased in the first example, the outcome of second toss is independent of the first toss. On the other hand, if Euler misses the school on first day, he might also miss it the next day with high probability (because he might be out of station or ill). In this sense, the event that Euler is absent on the first day is not independent of the event that he misses the school on the second day.

Our intuition suggests that event A and B should be independent if $P(A)$ and $P[A|B]$ are equal. We define two events A, B to be independent if,

$$P(A \cap B) = P(A)P(B) \Leftrightarrow P(A) = P[A|B].$$

Suppose there are two red and two blue balls in a bin. You pick two balls out of it sequentially, one after another. Let R_1 be the event that the first draw is a red ball. Similarly, let B_2 be the event that second ball is blue.

Exercise 2. Are these two events, R_1 and B_2, independent?

There can be two ways to perform this experiment. In the first case, you can pick a ball, replace it, and then draw the second ball.

Exercise 3. Show that R_1 and B_2 are independent in this case.

In the other case, you do not replace the ball. It seems that the probability of event B_2 should increase if the first ball is red (event R_1 occurred). Intuitively, they should be dependent. Let us check it.

The conditional probability $P\left[B_2|R_1\right]$ is 2/3 (since two blue ball and one red ball remains). What is the probability of second ball being blue?

Exercise 4. Show that $P(B_2)$ is 1/2, significantly lower than the conditional probability. You can calculate it using partition formula or by using symmetry.

The difference between *disjoint* events and *independent* events can be confusing. Remember, if A and B are independent, occurrence of A has no effect on occurrence of B. On the other hand, if A and B are disjoint, occurrence of A rules out the occurrence of B (they are *highly* dependent).

Like the case of probability function, two random variables are called independent if the product of their probability mass function gives the probability mass function of both.

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$$

Exercise 5. Let X be the random variable that assigns 1 if the number on the throw of a dice is even else it is -1. Let Y be the random variable that assigns 1 if the number on the throw of a dice is prime else it is -1. Show that X and Y are not independent.

One of the implication of two random variables being independent is $E[XY] = E[X]E[Y]$, you will prove this in the assignment.

1.1 Independent events

Two events A, B are said to be independent if,

$$P(A \cap B) = P(A)P(B).$$

Again, A and B being independent means, occurrence of A has no effect on occurrence of B. We keep this definition because it takes care of the case $P(B) = 0$, as compared to the equivalent (almost) definition, $P\left[A|B\right] = P(A)$.

Note 1. Both these definitions are pretty intuitive.

If there is no relation between A and B (independent), we expect the probability of $P(A \cap B)$ to be the product of the two individual probabilities. If the probability is lower, they are said to be *negatively correlated*,

$$P(A \cap B) < P(A)P(B).$$

Similarly, two events are said to be *positively correlated* if,

$$P(A \cap B) > P(A)P(B).$$

Intuitively, negative correlation means event A reduces the probability of event B. On the same lines, positive correlation means occurrence of A increases the chance of event B.

Exercise 6. What is the relationship between independence and conditional probability of two events?

Many a times it might be clear from the context that two events are independent. For example,

- If we toss two coins, we get a head on first toss is independent of getting a tails on the second toss. More generally, any event *totally dependent* on first toss will be independent of any other event *totally dependent* on second toss. This is sometimes clearly stated by stating, a coin is tossed twice independently.
- If we pick two cards from a deck sequentially, with replacement, the event that we get hearts in first pick is independent to that we get ace on the second pick. If we don't put the first card back before picking the second, without replacement, they might not be independent event.

– As mentioned before, two disjoint events (non-zero probability) are always dependent.

In some cases, it might not be this clear. The best way is to check the definition of independence. Suppose you toss two coins, let A be the event that *the first toss is head* and B be the event that *the two outcomes are same*. Calculating,

$$P(A) = 1/2,\ P(B) = 1/2,\ P(A \cap B) = 1/4.$$

So, the two events are independent.

Let's look at another scenario for independence in probability. There was a survey conducted by the Health dept. in a hospital (with Asthma and Diabetes patients), it found that people who had diabetes had a lower probability of having Asthma as compared to the general population in the hospital. This suggests that people who have Asthma have less likelihood of getting Diabetes.

It turns out that even if having Asthma and Diabetes are independent of each other, there will be a negative relation between Asthma and Diabetic patients in a hospital.

In other words, suppose A, B are two independent events. They will not be mutually independent if we consider the conditional probabilities given $A \cup B$. That is, events $A|(A \cup B)$ and $B|(A \cup B)$ will be negatively related even if A, B are independent. This is known as *Berkson's Paradox*. The simple proof follows from the fact that $P((A \cap B)|(A \cup B)) \leq P(A|(A \cup B))P(B|A \cup B))$, if A, B are independent (given as an exercise in the assignment).

To make it more quantitative, consider a sample of 1000 balls. We know that 100 of them are red, 50 of them are shiny and 5 of them are red and shiny. The probability of being shiny is 1/20 and also the probability of a red ball being shiny is $5/100 = 1/20$. Hence being red and being shiny are independent.

Say, we pick only the balls which are red and/or shiny. Then the probability that a ball is shiny is close to 1/3 but a red ball being shiny remains at 1/20. This will show that a red ball is mostly not shiny.

Exercise 7. Say A be the event that the ball is shiny and B being the event that the ball is red. Prove that $P(A|(A \cup B)) > P(A)$. Convince yourself that this is the reason why events A and B seem to be negatively related.

Exercise 8. Convince yourself that the same thing happened in the above example (survey in a hospital).

1.2 Independence for family of events

The criteria for independence becomes more involved if there are more than two events. Consider a family of events, $\{A_i\}_{i \in I}$, each being indexed by i in some index set I.

The simplest notion of independence is called *pairwise independence*, implying that any two pair of events in the family are independent.

$$P(A_i \cap A_j) = P(A_i)P(A_j) \ \ \forall i \neq j \in I.$$

It can be strengthened to *k-wise independence*, it holds iff

$$P(\cap_{j \in J} A_j) = \mathit{\Pi}_{j \in J} P(A_j) \ \ \forall J \subseteq I, |J| \leq k.$$

Exercise 9. Can we restrict the previous condition to be $|J| = k$ and get the same notion of independence?

Exercise 10. Show that pairwise independence is same as 2-wise independence.

The strongest notion is of *mutual independence*, which says that the events are $|I|$-wise independent.

$$P(\cap_{j \in J} A_j) = \mathit{\Pi}_{j \in J} P(A_j) \ \ \forall J \subseteq I.$$

Exercise 11. Convince yourself that k-wise independence implies l-wise independence if $l \leq k$.

It might seem that all these notions are equivalent. Consider the following example which shows that these definitions are different. The simplest way would be to come up with 3 events which are pairwise independent but not mutually independent.

Exercise 12. Can you think of an example?

We toss an unbiased coin twice. Define A to be the event that first toss is same as the second toss. Define B, C to be events that first (second) toss is head respectively. You will show in the assignment that these three events are pairwise independent but not mutually (3-wise) independent.

Exercise 13. Is it possible that a set of events are k and k+2-wise independent but not k+1-wise independent?

1.3 Independent random variables

We can similarly define independence of random variables. Remember, a random variable X is a mapping from sample space to real numbers, $X : \Omega \to \mathbb{R}$. The conditional distribution of X given an event B was defined to be,

$$P_{X|B}(x) = P(X = x|B).$$

Using these definitions, two random variables X and Y are independent iff,

$$P(X = x \cap Y = y) = P(X = x)P(Y = y) \;\; \forall x, y.$$

Here x, y are in the range of X, Y respectively.

Notice the difference between independence of events and random variables. For two random variables to be independent, all concerned events defined by $X = s$ *kind of events* should be independent. Intuitively, two random variables are independent iff the value of one does not give any indication about the value of other random variable.

Again, like events, independence can be generalized to family of random variables $\{X_i\}_{i \in I}$. We start with *pairwise independence*, implying that any two pair of random variables in the family are independent.

$$P((X_i = x_i) \cap (X_j = x_j)) = P(X_i = x_i)P(X_j = x_j) \;\; \forall x_i, x_j, i, j \in I.$$

Here, x_i, x_j are all possible elements from the range of X_i and X_j respectively.

It can be strengthened to *k-wise independence*, it holds iff

$$P(\cap_{j \in J}(X_j = x_j)) = \Pi_{j \in J} P(X_j = x_j) \;\; \forall \{x_j\}, J \subseteq I, |J| \leq k.$$

You will show in the assignment that $|J| \leq k$ can be changed to $|J| = k$ in the previous definition.

The last extension is to *mutual independence* and is very similar, which says that the events are $|I|$-wise independent.

$$P(\cap_{j \in J}(X_j = x_j)) = \Pi_{j \in J} P(X_j = x_j) \;\; \forall \{x_j\}, J \subseteq I.$$

Exercise 14. Convince yourself that k-wise independence implies l-wise independence for random variables if $l \leq k$.

Exercise 15. Is pairwise independence same as mutual independence, construct a counterexample.

Let us try to extend the previous example on events to random variables to set of random variables which are k-wise independent but not $k + 1$-wise independent. Pick $X_1, X_2, \cdots X_k$ uniformly independently from $\{0, 1\}$. Let X_{k+1} be the case that the sum of $X_1, X_2, \cdots, X_k$ is even. Clearly they are not $k + 1$-wise independent (the value of X_{k+1} is determined by other k random variables). On the other hand, you can show that any k are independent (the fact that $1, \cdots, k + 1 - i$ are independent is same as $1, 2, \cdots, k$ being independent).

Exercise 16. Why do we need so many notions of independence?

It seems that mutual independence is the strongest. If we need to bound the properties of $X_1, X_2, \cdots, X_n$ (like the behavior of their average), it is easiest to have mutual independence. Then, why do we need pairwise independence. The reason is, coming up with mutual independence is expensive (to create n such random variables, we need n random bits). In some case, it is enough to have pairwise (or k-wise) independence, one such example will be given in the next section. In such cases, constructing pairwise independent bits (random variables) is easier (and less expensive). We see one such construction here. There are constructions for k-wise independence too, but they will not be covered in this course.

Construction of pairwise independent bits: Given $X_1, X_2, \cdots, X_n$, n independent uniformly random bits, we will create $2^n - 1$ pairwise random bits. Each random bit, Y_S, will be indexed by a non-empty subset of $[n]$. Simply, define

$$Y_S = \sum_{i \in S} X_i \mod 2.$$

Exercise 17. What is the probability that $P(Y_S = 1)$ or $P(Y_S = 0)$?

Since S is non-empty, pick a particular index $i \in S$. Given any choice/assignment of remaining random variables in S, X_i is 0 (or 1) with probability half. That means, fixing any assignment of remaining random variables, Y_S takes value 0 or 1 with equal probability. So, $P(Y_S = 0) = P(Y_S = 1) = 1/2$.

Exercise 18. Why did we take only non-empty subsets?

That is a good start, how do we show that they are pairwise independent? A very similar strategy will work.

Theorem 1. *For $\{Y_S\}$ constructed as above, for any non-empty $U, V \subseteq [n]$,*

$$P(Y_U = a, Y_V = b) = P(Y_U = a)P(Y_V = b).$$

Proof. Without loss of generality there exist an i in U not present in V (otherwise, switch U and V). Given that Y_U is uniform, it is enough to show that $P(Y_U = a | Y_V = b) = 1/2$.

Exercise 19. Why is it enough?

Fix all the variables except X_i (call it a partial assignment A), such that, $Y_V = b$ (notice that $i \notin V$). Depending on the value of X_i, Y_U will switch from 0 to 1 with equal probability.

$$\begin{aligned} P[Y_U = a | Y_V = b] &= \frac{P[Y_U = a \cap Y_V = b]}{P[Y_V = b]} \\ &= \frac{1}{P[Y_V = b]} \left(\sum_{A \text{ s.t. } Y_V = b} P[Y_U = a|A]P[A] \right) \\ &= \frac{1}{P[Y_V = b]} \left(\sum_{A \text{ s.t. } Y_V = b} \frac{1}{2} P[A] \right) \\ &= \frac{1}{2} \frac{1}{P[Y_V = b]} \left(\sum_{A \text{ s.t. } Y_V = b} P[A] \right) \\ &= \frac{1}{2} \end{aligned}$$

Notice, it is the same argument as the one used to prove that Y_S's are uniform.

□

Exercise 20. Is this construction 3-wise independent?

2 Frequency moments

This example is based on the paper by Alon, Matias and Szegedy on *The Space Complexity of Approximating the Frequency Moments* [1]. I would like to thank Piysuh Srivastava (TIFR) for his exposition of this example.

Suppose you maintain a server for an application and keep getting requests from different IP addresses (as a data stream). You would like to detect if there are a *few* IP addresses who might be making a lot of requests. This would give evidence that there might be an attack on the server.

In general, the task is to find if there are a large number of requests from a small number of clients (heavy hitters). The easiest way is to keep a *frequency* table, where there will an entry for each IP address, and it will contain the total number of requests which have come from this IP address. This table needs to be update as more requests arrive. You might already notice that the size of the table will be huge and updating it would be a nightmare.

Approximate solution: Is it possible to find an *approximate* solution to this which requires less complexity? Let us define one such approximate solution in terms of frequencies. Let $f_1, f_2, \cdots, f_n$ be the frequencies of requests from clients. Intuitively, there are heavy hitters iff $F_1 := \sum_{i=1}^{n} f_i$ is much less than $F_2 := \sum_{i=1}^{n} f_i^2$.

We can change our problem to ask, given that requests come one by one, is there a way to find if $F_2 >> F_1$? Notice that keeping the entire frequency table is very expensive. Also, it is easy to keep a single counter and update the value of F_1.

We can ask: is there an *efficient* way (in terms of space) to find F_2? We can even allow some approximation here, i.e., with high probability our estimation of F_2 should be close to the actual value of F_2 (not away by more than a constant factor). In general, estimating the frequency moments (and specifically F_2) of a data stream is an important and well studied problem. Let us see how we can do it (and show that our computed value is close to actual value) using the concepts learned so far.

AMS algorithm The central idea (from Alon, Matias and Szegedy) is to define a new random variable $Y = \sum_i a_i f_i$, where a_i's are random variables which take value $\{-1, 1\}$ with equal probability. We will see that $Z = Y^2$ will be a good approximation of F_2 with high probability. Notice that maintaining Y is same as keeping a counter (given the knowledge of a_i).

Exercise 21. What is the expected value of a_i; what is the expected value of Y.

Using linearity of expectation, $E[Y] = 0$. Though, we are interested in the expected value of Z.

$$\begin{aligned} E[Z] &= E[(\sum_i a_i f_i)^2] \\ &= E[\sum_i f_i^2 + \sum_{i \neq j} a_i a_j f_i f_j] \\ &= \sum_i f_i^2 + \sum_{i \neq j} f_i f_j E[a_i a_j] \end{aligned}$$

We used linearity of expectation and took out constants from expectation. If we assume a_i's to be pairwise independent (do we need mutual independence?), $E[a_i a_j] = E[a_i]E[a_j] = 0$ Hence,

$$E[Z] = \sum_i f_i^2 = F_2.$$

At least, in expectation, we will get the value of F_2. But from past discussion it is quite clear that the individual values of random variable Z might be far away from the expectation.

Exercise 22. Can you construct a random variable which is far away from expectation all the time. How can we show that we are not in that case?

Intuitively, the trick should be to show that the variance of Z is pretty small. That would make sure that, with high probability, the value of Z will be close to its expectation (and hence F_2). Let us calculate the variance. (Remember that we need to calculate $E[Z^2]$ to compute the variance, we already have $E[Z]$.)

$$\begin{aligned}
E[Z^2] &= E[Y^4] \\
&= E[(\sum_i a_i f_i)^4] \\
&= \sum_i f_i^4 + \sum_{i \neq j} E[4 f_i^3 f_j a_i^3 a_j + 6 f_i^2 f_j^2 a_i^2 a_j^2] + c_1 \sum_{i,j,k} E[f_i^2 f_j f_k a_i^2 a_j a_k] + c_2 \sum_{i,j,k,l} E[f_i f_j f_k f_l a_i a_j a_k a_l] \\
&= \sum_i f_i^4 + 4 \sum_{i \neq j} f_i^3 f_j E[a_i a_j] + 6 \sum_{i<j} f_i^2 f_j^2 + c_1 \sum_{i,j,k} f_i^2 f_j f_k E[a_j a_k] + c_2 \sum_{i,j,k,l} f_i f_j f_k f_l E[a_i a_j a_k a_l]
\end{aligned}$$

The last term will be 0 assuming a_i's are 4-wise independent.

Exercise 23. The terms with coefficients 4 and c_1 will become 0 under what condition?

Since 4-wise independence implies 2-wise independence, we get that $E[Z^2] = \sum_i f_i^4 + 6 \sum_{i<j} f_i^2 f_j^2$ when a_i's are 4-wise independent. Calculating the variance,

$$Var[Z] = E[Z^2] - E[Z]^2 = 4 \sum_{i<j} f_i^2 f_j^2 = 2 \sum_{i \neq j} f_i^2 f_j^2.$$

Exercise 24. Calculate and confirm the variance of Z.

Since f_i^2's are positive, we get $Var[Z] \leq 2 \sum_i \sum_j f_i^2 f_j^2 = 2F_2^2$. So, we are able to bound the variance and expectation of Z. This gives us intuition that Z is close to F_2 with high probability.

Exercise 25. How do we show the previous statement formally?

This will follow from *Chebyshev inequality*, it states that given a random variable X and $a > 0$,

$$P[|X - E[X]| \geq a] \leq \frac{Var[X]}{a^2}.$$

Applying this for our case (you will show in the assignment),

$$P[|Z - F_2| \geq \alpha F_2] \leq \frac{2}{\alpha^2}.$$

The proof of Chebyshev inequality will be covered in the next lecture note.

Exercise 26. Is there a way to improve the variance, possibly by increasing the memory usage slightly?

The variance could further be decreased by using the following result (you will prove it in assignment). Let $X_1, X_2, \cdots, X_m$ be independent copies of X, then

$$Var[\frac{1}{m} \sum_i X_i] = \frac{1}{m} Var[X].$$

Taking cue from this result, we take s independent copies of Z and define our estimate to be,

$$G = \frac{1}{s} \sum_{i=1}^{s} Z_i.$$

Exercise 27. What is the mean and variance of G? What can you say after applying Chebyshev's inequality on G?

This shows that we can estimate F_2 with any constant probability (say 7/8) with error ϵF_2 using only constant number of copies of Z ($16/\epsilon^2$).

Note, we only wanted a_i's to be 4-wise independent and not mutually independent. There are techniques to use small random seeds and come up with such 4-wise independent distributions (like we constructed pairwise independent bits in the previous section), but they fall out of the scope of this course .

3 Extra reading: hash functions

We will see an application of independence in a concept called *hashing*. It is taken from Victor Shoup's book on Computational Number Theory [2].

Hashing is the practice of mapping a set bigger set S to a potentially smaller set T using random keys R. The objective is that an element $s \in S$ goes to an element in T uniformly at random and the number of collisions are small.

One application could be to store lots of elements, when it is know that all elements belong to some big set S. To reduce the size of the storage, we can store the hashed image (in T) instead of the actual element. This will require much less storage size.

Exercise 28. Read about hashing from the web.

Let us define it formally.

Given three sets R, S, T, a *hashing* from S to T is a collection of functions $\Phi_r : S \to T$, for each $r \in R$. We could think of elements of R as keys which allow us to know the mapping from S to T. We can think of $\Phi_R(s)$ for an $s \in S$ as a random variable where r is picked uniformly. Again, we would want $\Phi_R(s)$ to be distributed uniformly in T. Ideally, knowledge of $\Phi_R(s)$ should not give any information about $\Phi_R(s')$ for any $s' \neq s$.

To make these requirements concrete, a family of hash functions (hashing) is called *pairwise independent* iff

- Random variables $\{\Phi_R(s)\}_{s \in S}$ are pairwise independent.
- $\Phi_R(s)$ is uniformly distributed over T for every s.

You will show in the assignment that a hashing is pairwise independent iff

$$P(\Phi_R(s) = t \cap \Phi_R(s') = t') = \frac{1}{|T|^2} \quad \forall s, s' \in S (s \neq s'); t, t' \in T.$$

As expected, pairwise independent hashing implies that no s favors any particular area of T, and given knowledge of one hash (value of $\Phi_r(s)$) we have no clue about any other $\Phi_r(s')$ (assuming, we don't know r). There are many ways known to construct such hash functions. We will now take a look at an application of such possible constructions in cryptography.

Assume that two parties, let us call them Raj and Simran, want to communicate a message. Though, Simran wants to make sure that the message is actually from Raj and not transmitted by Amrish. Hashing provides a solution to this age old problem, though with two assumptions.

- Raj and Simran share a secret key.
- Simran is ready to accept a small probability of error.

The solution is quite natural given a family of hash functions.

Exercise 29. Why don't you think about a possible protocol for this problem using hashing.

So, as hinted above, assume that Raj and Simran know of a hashing scheme $R, S, T, \Phi_R(s)$. The set S will become the possible set of messages and R is the set of random keys. Raj and Simran will decide upon the key in advance, some $r \in R$. At the time of transmission, Raj will send $(s, \Phi_r(s))$ as a message to Simran.

We can model it as random variables $(X, Y = \Phi_R(X))$, where X is the random message and $\Phi_R(X)$ is the corresponding authentication tag. The message X is uniformly distributed in S and $r \in R$ is picked uniformly.

After receiving the message, Simran checks if the authentication tag Y is indeed the hash of X under the secret key R. Amrish (our adversary) fools Simran if he can find another pair (X', Y'), such that, $\Phi_r(X') = Y'$.

Ideally, ever after seeing the original message (X, Y), Amrish should not have any knowledge of key r. So, the probability that he can find X', Y' such that $\Phi_r(X') = Y'$ should be really small. We will explicitly calculate the probability of this event (say E).

$$\begin{aligned}
P(E) &= \sum_{s \in S} \sum_{t \in T} P(X = s \cap Y = t \cap E) \\
&= \sum_{s \in S} \sum_{t \in T} P((X = s) \cap (\Phi_R(s) = t) \cap (\Phi_R(X') = Y') \cap (X' \neq s)) \\
&= \sum_{s \in S} \sum_{t \in T} P(X = s) P((\Phi_R(s) = t) \cap (\Phi_R(X') = Y') \cap (X' \neq s)) \\
&\leq \sum_{s \in S} \sum_{t \in T} P(X = s) \frac{1}{|T|^2} \\
&= \frac{1}{|T|}
\end{aligned}$$

Here, third equality follows from the fact that the message and the key are picked independently.

Exercise 30. Why did the fourth inequality follow?

So, the probability of Amrish being successful is inversely proportional to the size of the image T. By picking a big enough T, Simran can ensure her failure probability to be small.

4 Assignment

Exercise 31. Prove $Var[\frac{1}{m}\sum_i X_i] = \frac{1}{m} Var[X]$, where X_i's are independent copies of X. What kind of independence do we need for this result?

Exercise 32. Given that events A, B are independent, prove that A^c, B^c are also independent. Hint: first prove that A, B^c are independent.

Exercise 33. Let X and Y be two independent random variables. Prove that $E[XY] = E[X]E[Y]$.

Exercise 34. We toss an unbiased coin twice. Define A to be the event that first toss is same as the second toss. Define B, C to be events that first (second) toss is head respectively. Show that these three events are pairwise independent but not mutually (3-wise) independent.

Exercise 35. Show that a hashing is pairwise independent iff

$$P(\Phi_R(s) = t \cap \Phi_R(s') = t') = \frac{1}{|T|^2} \quad \forall s, s' \in S (s \neq s'); t, t' \in T.$$

Exercise 36. Suppose A, B are independent events, show that $A|(A \cup B)$ and $B|(A \cup B)$ are negatively correlated.

Exercise 37. For the estimation of F_2 by AMS algorithm, show that $P[|Z - F_2| \geq \alpha F_2] \leq \frac{2}{\alpha^2}$.

Exercise 38. Show that random variables $\{X_i\}_{i \in I}$ are k-wise independent iff

$$P(\cap_{j \in J}(X_j = x_j)) = \Pi_{j \in J} P(X_j = x_j) \quad \forall \{x_j\}, J \subseteq I, |J| = k.$$

Notice that condition needs to be true only when $|J| = k$. Can we make the same change when we talk about k-wise independence of events?

References

1. N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of computer and system sciences*, pages Volume 58, Issue 1, 137–147, 1999.
2. V. Shoup. *A Computational Introduction to Number Theory and Algebra.* Cambridge University Press, 2008.