CS 30: Discrete Math in CS (Winter 2019): Lecture 25

Date: 22nd February, 2019 (Friday)
Topic: Probability: Independent Random Variables, Variance
Disclaimer: These notes have not gone through scrutiny and in all probability contain errors. Please discuss in Piazza/email errors to deeparnab@dartmouth.edu

---

1. **Independent Random Variables.** Two random variables X and Y are independent, if for any $x \in \mathsf{range}(X)$ and any $y \in \mathsf{range}(Y)$,

$$\mathbf{Pr}[X = x, Y = y] = \mathbf{Pr}[X = x] \cdot \mathbf{Pr}[Y = y]$$

Examples:

- If we roll two dice, and X_1 and X_2 indicate the value of the rolls, then X_1 and X_2 are independent.
- If we have two independent events $\mathcal{A}$ and $\mathcal{B}$, then their indicator random variables $\mathbf{1}_{\mathcal{A}}$ and $\mathbf{1}_{\mathcal{B}}$ are independent.
- Consider a random variable X taking value $+1$ if a toss of a coins is head, and -1 if its tails. Such random variables are called *Rademacher random variables*. Suppose we toss the coin twice and X_1 and X_2 are the corresponding random variables. Then X_1 and X_2 are independent.

A set of k random variables $X_1, \ldots, X_k$ are *mutually independent* if for any $x_1, x_2, \ldots, x_k$ with $x_i \in \mathsf{range}(X_i)$, we have

$$\mathbf{Pr}[X_i = x_i, \quad \forall 1 \le i \le k] = \prod_{i=1}^{k} \mathbf{Pr}[X_i = x_i]$$

Theorem 1. If X and Y are two independent random variables, then

$$\mathbf{Exp}[XY] = \mathbf{Exp}[X] \cdot \mathbf{Exp}[Y]$$

Proof.

$$\begin{aligned}
\mathbf{Exp}[XY] &= \sum_{x \in \mathsf{range}(x), y \in \mathsf{range}(y)} (xy) \cdot \mathbf{Pr}[X = x, Y = y] && \text{Definition of Expectation} \\
&= \sum_{x \in \mathsf{range}(x), y \in \mathsf{range}(y)} (xy) \cdot \mathbf{Pr}[X = x] \cdot \mathbf{Pr}[Y = y] && \text{Independence} \\
&= \left(\sum_{x \in \mathsf{range}(x)} x \cdot \mathbf{Pr}[X = x] \right) \cdot \left(\sum_{y \in \mathsf{range}(y)} y \cdot \mathbf{Pr}[Y = y] \right) && \text{Algebra} \\
&= \mathbf{Exp}[X] \cdot \mathbf{Exp}[Y] && \text{Definition of Expectation}
\end{aligned}$$

$\square$

Of course, there is no need to stick to two random variables. The theorem easily generalizes (do you see how?) to mutually independent random variables as follows.

Theorem 2. If $X_1, X_2, \ldots, X_k$ are mutually independent random variables, then

$$\mathbf{Exp}\left[\prod_{i=1}^{k} X_i\right] = \prod_{i=1}^{k} \mathbf{Exp}\left[X_i\right]$$

Examples.

- Let X_i and X_j be two independent Rademacher random variables. Recall, X_i takes $+1$ with probability $1/2$ and -1 with probability $1/2$. Then note (a) $\mathbf{Exp}[X_i] = \mathbf{Exp}[X_j] = 0$, (b) $\mathbf{Exp}[X_i \cdot X_i] = \mathbf{Exp}[X_j \cdot X_j] = 1$, and (c) $\mathbf{Exp}[X_i X_j] = \mathbf{Exp}[X_i] \cdot \mathbf{Exp}[X_j] = 0$. This is a very useful fact.
- Consider rolling a dice n times, independently. Let Z be the random variable indicating the *product* of all the numbers seen. What is $\mathbf{Exp}[Z]$? To solve this, let X_i be the roll of the ith dice. We know that $\mathbf{Exp}[X_i] = 3.5$ for all i. We also know $X_1, X_2, \ldots, X_n$ are all independent random variables. Thus, $\mathbf{Exp}[Z] = (3.5)^n$.

2. **Variance and Standard Deviation.**

The expectation of a random variable is some sort of an "average behavior" of a random variable. However, the true value of a random variable may be no where close to the expectation. For instance, consider a random variable which takes the value 10000 with probability $1/2$, and -10000 with probability $1/2$. What is $\mathbf{Exp}[X]$? Yes, it is 0. Thus, there is significant *deviation* of X from its expectation.

The variance and standard deviation try to capture this deviation. In particular, the variance of a random variable is the expected value of the square of the deviation.

Theorem 3. Let X be a random variable. The variance of X is defined to be

$$\mathbf{Var}[X] := \mathbf{Exp}\left[(X - \mathbf{Exp}[X])^2\right]$$

That is, if we define another random variable $D := (X - \mathbf{Exp}[X])^2$, then $\mathbf{Var}[X]$ is the expected value of this new deviation random variable D.

The *standard deviation* $\sigma(X)$ is defined to be $\sqrt{\mathbf{Var}(X)}$.

Theorem 4. $\mathbf{Var}[X] = \mathbf{Exp}[X^2] - (\mathbf{Exp}[X])^2$.

Proof.

$$\mathbf{Var}[X] = \mathbf{Exp}[(X - \mathbf{Exp}[X])^2] = \mathbf{Exp}[X^2 - 2X\mathbf{Exp}[X] + (\mathbf{Exp}[X])^2]$$

Then, we apply linearity of expectation to get

$$\mathbf{Var}[X] = \mathbf{Exp}[X^2] - 2\mathbf{Exp}[X] \cdot \mathbf{Exp}[X] + (\mathbf{Exp}[X])^2 = \mathbf{Exp}[X^2] - (\mathbf{Exp}[X])^2$$

□

A useful corollary:

Theorem 5. For any random variable $\mathbf{Exp}[X^2] \geq (\mathbf{Exp}[X])^2$.

Examples

- *Roll of a die.* Let X be the roll of a fair 6-sided die. We know that $\mathbf{Exp}[X] = 3.5$. To calculate the variance, we can use the deviation $D := (X - \mathbf{Exp}[X])^2 = (X - 3.5)^2$. Usinhg this, we get

$$\mathbf{Var}[X] = \mathbf{Exp}[D] = \frac{1}{6}\left((2.5)^2 + (1.5)^2 + (0.5)^2 + (0.5)^2 + (1.5)^2 + (2.5)^2\right) = \frac{35}{12}$$

- *Toss of a biased coin.* Let X be a Bernoulli random variable taking value 1 if a coin tosses heads, and 0 otherwise. Suppose the probability of heads was p. Recall, $\mathbf{Exp}[X] = p$. Also note since X is a indicator random variable, $X^2 = X$. Thus, $\mathbf{Exp}[X^2] = p$ as well. We can calculate the variance as

$$\mathbf{Var}[X] = \mathbf{Exp}[X^2] - (\mathbf{Exp}[X])^2 = p - p^2 = p(1-p)$$

- *Indicator Random Variable.* Using the above toss of a biased coin example, we see that for any event $\mathcal{E}$, the variance of the indicator random variable is

$$\mathbf{Var}[\mathbf{1}_{\mathcal{E}}] = \mathbf{Pr}[\mathcal{E}] \cdot (1 - \mathbf{Pr}[\mathcal{E}])$$

Theorem 6. If X is a random variable, and c is a "scalar" (a constant), then $Z = c \cdot X$ is another random variable. $\mathbf{Var}[c \cdot X] = c^2 \cdot \mathbf{Var}[X]$.

Proof.

$$\mathbf{Var}[c \cdot X] = \mathbf{Exp}[c^2 X^2] - (\mathbf{Exp}[cX])^2 = c^2 \mathbf{Exp}[X^2] - c^2 (\mathbf{Exp}[X])^2 = c \dot{} \mathbf{Var}[X]$$

□

The next theorem is a *linearity of variance* result for ***independent*** random variables.

Theorem 7. For any two ***independent*** random variables X and Y, let $Z := X + Y$. Then,

$$\mathbf{Var}[Z] = \mathbf{Var}[X] + \mathbf{Var}[Y]$$

Proof.

$$\begin{aligned}
\mathbf{Var}[X+Y] &= \mathbf{Exp}[(X+Y)^2] - (\mathbf{Exp}[X] + \mathbf{Exp}[Y])^2 \\
&= \mathbf{Exp}[X^2 + 2XY + Y^2] - \left(\mathbf{Exp}^2[X] - 2\mathbf{Exp}[X]\mathbf{Exp}[Y] + \mathbf{Exp}^2[Y]\right) \\
&= \left(\mathbf{Exp}[X^2] - \mathbf{Exp}^2[X]\right) + \left(\mathbf{Exp}[Y^2] - \mathbf{Exp}^2[Y]\right) + 2\left(\mathbf{Exp}[XY] - \mathbf{Exp}[X]\mathbf{Exp}[Y]\right) \\
&= \mathbf{Var}[X] + \mathbf{Var}[Y]
\end{aligned}$$

In the last equality, due to independence, we get that $2\left(\mathbf{Exp}[XY] - \mathbf{Exp}[X]\mathbf{Exp}[Y]\right) = 0$. □

We can generalize the above proof to many random variables. In particular, we can say that if $X_1, X_2, \ldots, X_k$ are mutually independent random variables, then the variance of the sum is the sum of the variances. However, we *don't need mutual independence*. Pairwise independence suffices. This is very important to note. The proof is given as a solution to the UGP; perhaps you can try it. There is nothing more than the algebra above except there are k things adding up.

Theorem 8. For any k ***pairwise independent*** random variables $X_1, X_2, \ldots, X_k$,

$$\mathbf{Var}\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} \mathbf{Var}[X_i]$$

3. **Deviation Inequalities**

We have seen an example that $\mathbf{Exp}[X]$ may not be anywhere close to what values X can take (recall the $X = 10000$ with 0.5 probability and -10000 with 0.5 probability). Deviation inequalities try to put an *upper bound* on the probability that a random walk deviates too far from the expectation.

The mother of all deviation inequalities is the following:

Theorem 9. (Markov's Inequality)

Let X be a random variable whose range is *non-negative reals*. Then for any $t > 0$, we have

$$\mathbf{Pr}[X \geq t] \leq \frac{\mathbf{Exp}[X]}{t}$$

Proof. By definition of expectation, we have

$$\mathbf{Exp}[X] = \sum_{x \in \mathsf{range}(X)} x \cdot \mathbf{Pr}[X = x] = \sum_{0 \leq x < t} x \cdot \mathbf{Pr}[X = x] + \sum_{x \geq t} x \cdot \mathbf{Pr}[X = x]$$

The first summation $\sum_{0 \leq x < t} x \cdot \mathbf{Pr}[X = x] \geq 0$. All terms are non-negative. The second summation is $\sum_{x \geq t} x \cdot \mathbf{Pr}[X = x] \geq t \cdot \sum_{x \geq t} \mathbf{Pr}[X = x] = t \cdot \mathbf{Pr}[X \geq t]$.

Putting it all together, we get

$$\mathbf{Exp}[X] \geq t \cdot \mathbf{Pr}[X \geq t]$$

which gives what we want by rearrangement. □

Markov's inequality only talks about non-negative random variables. Indeed, the example in the beginning of this bullet point shows that it cannot be true for general random variables. This is where *variance* comes to play. The following is one of the most general forms of deviation inequalities.

Theorem 10. (Chebyshev's Inequality)
Let X be a random variable. Then for any $t > 0$, we have

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \ \geq \ t] \leq \frac{\mathbf{Var}[X]}{t^2}$$

Proof. We first note that

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \geq t] = \mathbf{Pr}[(X - \mathbf{Exp}[X])^2 \geq t^2]$$

Then we notice that $D := (X - \mathbf{Exp}[X])^2$ is a non-negative random variable, and therefore we can apply Markov's inequality on it to get

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \geq t] = \mathbf{Pr}[D \geq t^2] \leq \frac{\mathbf{Exp}[D]}{t^2} = \frac{\mathbf{Var}[X]}{t^2}$$

□