Lecture 14: Markov and Chebyshev's Inequalities

Anup Rao

February 5, 2018

We continue our discussion of Variance, and talk about using it to prove our first concentration bound.

In the last lecture, we discussed how to calculate the variance of a random variable $\text{Var}[X] = \mathbb{E}[X^2]$. We showed that the variance of the sum of two independent random variables is just the sum of the variance. Another nice way to calculate the variance is:

Fact 1. $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Proof. Let $\mu = \mathbb{E}[X]$. Then

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}\left[(X-\mu)^2\right] \\ &= \mathbb{E}\left[X^2 + \mu^2 - 2X\mu\right] \\ &= \mathbb{E}\left[X^2\right] + \mu^2 - 2\mu\,\mathbb{E}[X] \\ &= \mathbb{E}\left[X^2\right] + \mu^2 - 2\mu^2 = \mathbb{E}\left[X^2\right] - \mu^2.\end{aligned}$$

by linearity of expectation

□

Definition 2. *The* standard deviation *of the random variable X is defined to be*

$$\sigma(X) = \sqrt{\text{Var}[X]}.$$

Example: n coin tosses

Suppose we toss a fair coin n times. What is the variance of the number of heads?

Let X denote the number of heads. As usual, it is best to represent X using the indicator random variables for each coin toss being a heads:

$$X_i = \begin{cases} 1 & \text{if } i\text{'th toss is heads,} \\ 0 & \text{otherwise.} \end{cases}$$

Then we have $X = X_1 + X_2 + \ldots + X_n$. Since $X_1, \ldots, X_n$ are independent, we have

$$\text{Var}[X] = \text{Var}[X_1] + \text{Var}[X_2] + \ldots + \text{Var}[X_n].$$

Now we can calculate:

$$\text{Var}[X_i] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}[X_i]^2.$$

First let us calculate $\mathbb{E}\left[X_i^2\right]$,

$$\mathbb{E}\left[X_i^2\right] = (1/2)\cdot 1^2 + (1/2)\cdot 0 = 1/2.$$

$$\mathbb{E}\left[X_i\right]^2 = ((1/2)\cdot 1 + (1/2)\cdot 0)^2 = (1/2)^2 = 1/4,$$

so we get

$$\mathrm{Var}\left[X_i\right] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[X_i\right]^2 = 1/2 - 1/4 = 1/4.$$

So, we conclude that

$$\mathrm{Var}\left[X\right] = \mathrm{Var}\left[X_1\right] + \mathrm{Var}\left[X_2\right] + \ldots + \mathrm{Var}\left[X_n\right] = n/4.$$

The standard deviation is

$$\sigma(X) = \sqrt{n}/2.$$

The variance gives a powerful way to measure the probability that a random variable deviates from its expectation by a lot. As we have seen, $\mathbb{E}\left[X\right]$ does not tell us anything about how far X can be from its expectation. However, we do have the following simple inequality:

Fact 3 (Markov's inequality)**.** *If X is a non-negative random variable, then*

$$p(X \geq \alpha) \leq \frac{\mathbb{E}\left[X\right]}{\alpha}.$$

Proof. We can use conditional expectation to express:

$$\begin{aligned}
\mathbb{E}\left[X\right] &= p(X \geq \alpha)\cdot \mathbb{E}\left[X | X \geq \alpha\right] + p(X < \alpha)\cdot \mathbb{E}\left[X | X < \alpha\right] \\
&\geq p(X \geq \alpha)\cdot \mathbb{E}\left[X | X \geq \alpha\right] \\
&\geq \alpha \cdot p(X \geq \alpha),
\end{aligned}$$

proving that

$$p(X \geq \alpha) \leq \frac{\mathbb{E}\left[X\right]}{\alpha}.$$

□

Applying Markov's inequality to the variance gives us Chebyshev's inequality:

Fact 4 (Chebyshev's inequality)**.**

$$p(|X - \mathbb{E}\left[X\right]| \geq \alpha) \leq \frac{\mathrm{Var}\left[X\right]}{\alpha^2}.$$

Proof.

$$\begin{aligned}
p(|X - \mathbb{E}\left[X\right]| \geq \alpha) &= p((X - \mathbb{E}\left[X\right])^2 \geq \alpha^2) \\
&\leq \frac{\mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right]}{\alpha^2} \\
&= \frac{\mathrm{Var}\left[X\right]}{\alpha^2}.
\end{aligned}$$

by Markov's inequality applied to the non-negative random variable $(X - \mathbb{E}\left[X\right])^2$.

□

Let us apply Markov and Chebyshev's inequality to some common distributions.

Example: Bernoulli Distribution

The *Bernoulli* distribution is the distribution of a coin toss that has a probability p of giving heads. Let X denote the number of heads. Then we have

$$\mathbb{E}[X] = p,$$
$$\text{Var}[X] = p - p^2.$$

Markov's inequality gives

$$p(X = 1) = p(X \geq 1) \leq \frac{\mathbb{E}[X]}{1} = p.$$

Chebyshev's inequality gives

$$p(X = 1) = p(|X - p| \geq 1 - p) \leq \frac{p - p^2}{(1-p)^2} = \frac{p}{1-p}.$$

Neither of these inequalities are very interesting, because it is much easier to see directly that $p(X = 1) = p$.

Example: Binomial Distribution

The binomial distribution is the outcome of tossing n independent coins, each of which gives heads with probability p. Let X be the number of heads.

Suppose that $p < 1/2$. We have:

$$\mathbb{E}[X] = pn,$$
$$\text{Var}[X] = (p - p^2)n.$$

Markov's inequality gives

$$p(X \geq 2pn) \leq \frac{\mathbb{E}[X]}{2pn} = \frac{pn}{2pn} = \frac{1}{2}.$$

Chebyshev's inequality gives

These are much more interesting inequalities, because it is hard to calculate $p(X \geq 2pn)$ directly.

$$\begin{aligned} p(X \geq 2pn) &= p(|X - np| \geq pn)) \\ &\leq \frac{\text{Var}[X]}{p^2n^2} = \frac{p(1-p)n}{p^2n^2} = \frac{1-p}{pn}. \end{aligned}$$

We see that n gets large, this probability goes to 0.