CS70: Lecture 28.

Variance; Inequalities; WLLN

1. Review: Independence
2. Variance
3. Inequalities
   - Markov
   - Chebyshev
4. Weak Law of Large Numbers

Review: Independence

Definition

X and Y are independent

$\Leftrightarrow Pr[X = x, Y = y] = Pr[X = x]Pr[Y = y], \forall x, y$

$\Leftrightarrow Pr[X \in A, Y \in B] = Pr[X \in A]Pr[Y \in B], \forall A, B.$

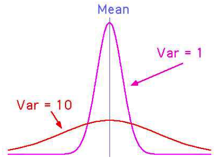
Theorem

X and Y are independent

$\Rightarrow f(X), g(Y)$ are independent $\forall f(\cdot), g(\cdot)$

$\Rightarrow E[XY] = E[X]E[Y].$

Variance



The variance measures the deviation from the mean value.

Definition: The variance of X is

$$\sigma^2(X) := var[X] = E[(X - E[X])^2].$$

$\sigma(X)$ is called the standard deviation of X.

Variance and Standard Deviation

Fact:

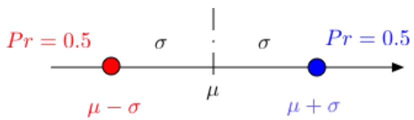
$$var[X] = E[X^2] - E[X]^2.$$

Indeed:

$$\begin{aligned} var(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2) \\ &= E[X^2] - 2E[X]E[X] + E[X]^2, \text{ by linearity} \\ &= E[X^2] - E[X]^2. \end{aligned}$$

A simple example

This example illustrates the term 'standard deviation.'



Consider the random variable X such that

$$X = \begin{cases} \mu - \sigma, & \text{w.p. } 1/2 \\ \mu + \sigma, & \text{w.p. } 1/2. \end{cases}$$

Then, $E[X] = \mu$ and $(X - E[X])^2 = \sigma^2$. Hence,

$$var(X) = \sigma^2 \text{ and } \sigma(X) = \sigma.$$

Example

Consider X with

$$X = \begin{cases} -1, & \text{w. p. } 0.99 \\ 99, & \text{w. p. } 0.01. \end{cases}$$

Then

$$\begin{aligned} E[X] &= -1 \times 0.99 + 99 \times 0.01 = 0. \\ E[X^2] &= 1 \times 0.99 + (99)^2 \times 0.01 \approx 100. \\ Var(X) &\approx 100 \implies \sigma(X) \approx 10. \end{aligned}$$

Also,

$$E(|X|) = 1 \times 0.99 + 99 \times 0.01 = 1.98.$$

Thus, $\sigma(X) \neq E[|X - E[X]|]$!

Exercise: How big can you make $\frac{\sigma(X)}{E[|X - E[X]|]}$?

Uniform

Assume that $Pr[X=i] = 1/n$ for $i \in \{1,\ldots,n\}$. Then

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} i \times Pr[X=i] = \frac{1}{n}\sum_{i=1}^{n} i \\ &= \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}. \end{aligned}$$

Also,

$$\begin{aligned} E[X^2] &= \sum_{i=1}^{n} i^2 Pr[X=i] = \frac{1}{n}\sum_{i=1}^{n} i^2 \\ &= \frac{1+3n+2n^2}{6}, \text{ as you can verify.} \end{aligned}$$

This gives

$$var(X) = \frac{1+3n+2n^2}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.$$

Variance of geometric distribution.

X is a geometrically distributed RV with parameter p.
Thus, $Pr[X=n] = (1-p)^{n-1}p$ for $n \geq 1$. Recall $E[X] = 1/p$.

$$\begin{aligned} E[X^2] &= p + 4p(1-p) + 9p(1-p)^2 + \ldots \\ -(1-p)E[X^2] &= \quad -[p(1-p) + 4p(1-p)^2 + \ldots] \\ pE[X^2] &= p + 3p(1-p) + 5p(1-p)^2 + \ldots \\ &= 2(p + 2p(1-p) + 3p(1-p)^2 + ..) \quad E[X]! \\ &\quad -(p + p(1-p) + p(1-p)^2 + \ldots) \quad \text{Distribution.} \\ pE[X^2] &= 2E[X] - 1 \\ &= 2(\frac{1}{p}) - 1 = \frac{2-p}{p} \end{aligned}$$

$\implies E[X^2] = (2-p)/p^2$ and
$var[X] = E[X^2] - E[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.
$\sigma(X) = \frac{\sqrt{1-p}}{p} \approx E[X]$ when p is small(ish).

Fixed points.

Number of fixed points in a random permutation of n items.
"Number of student that get homework back."

$X = X_1 + X_2 \cdots + X_n$

where X_i is indicator variable for ith student getting hw back.

$$\begin{aligned} E(X^2) &= \sum_i E(X_i^2) + \sum_{i \neq j} E(X_i X_j). \\ &= n \times \frac{1}{n} + (n)(n-1) \times \frac{1}{n(n-1)} \\ &= 1 + 1 = 2. \end{aligned}$$

$E(X_i^2) = 1 \times Pr[X_i = 1] + 0 \times Pr[X_i = 0]$
$= \frac{1}{n}$
$E(X_iX_j) = 1 \times Pr[X_i = 1 \cap X_j = 1] + 0 \times Pr[\text{“anything else”}]$
$= \frac{1 \times 1 \times (n-2)!}{n!} = \frac{1}{n(n-1)}$

$Var(X) = E(X^2) - (E(X))^2 = 2 - 1 = 1.$

Variance: binomial.

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{n} i^2 \binom{n}{i} p^i (1-p)^{n-i}. \\ &= \text{Really???!!\#\#...} \end{aligned}$$

Too hard!

Ok.. fine.
Let's do something else.
Maybe not much easier...but there is a payoff.

Properties of variance.

1. $Var(cX) = c^2 Var(X)$, where c is a constant.
   Scales by c^2.
2. $Var(X+c) = Var(X)$, where c is a constant.
   Shifts center.

Proof:

$$\begin{aligned} Var(cX) &= E((cX)^2) - (E(cX))^2 \\ &= c^2E(X^2) - c^2(E(X))^2 = c^2(E(X^2) - E(X)^2) \\ &= c^2 Var(X) \\ Var(X+c) &= E((X+c-E(X+c))^2) \\ &= E((X+c-E(X)-c)^2) \\ &= E((X-E(X))^2) = Var(X) \end{aligned}$$

□

Variance of sum of two independent random variables

Theorem:
If X and Y are independent, then

$$Var(X+Y) = Var(X) + Var(Y).$$

Proof:
Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E(X) = 0$ and $E(Y) = 0$.

Then, by independence,

$$E(XY) = E(X)E(Y) = 0.$$

Hence,

$$\begin{aligned} var(X+Y) &= E((X+Y)^2) = E(X^2 + 2XY + Y^2) \\ &= E(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2) \\ &= var(X) + var(Y). \end{aligned}$$

Variance of sum of independent random variables

Theorem:
If $X, Y, Z, \ldots$ are pairwise independent, then

$$var(X+Y+Z+\cdots) = var(X) + var(Y) + var(Z) + \cdots.$$

Proof:
Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E[X] = E[Y] = \cdots = 0$.

Then, by independence,

$$E[XY] = E[X]E[Y] = 0. \text{ Also, } E[XZ] = E[YZ] = \cdots = 0.$$

Hence,

$$\begin{aligned} var(X+Y+Z+\cdots) &= E((X+Y+Z+\cdots)^2) \\ &= E(X^2+Y^2+Z^2+\cdots+2XY+2XZ+2YZ+\cdots) \\ &= E(X^2)+E(Y^2)+E(Z^2)+\cdots+0+\cdots+0 \\ &= var(X)+var(Y)+var(Z)+\cdots. \end{aligned}$$

□

Variance of Binomial Distribution.

Flip coin with heads probability p.
X- how many heads?

$$X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$$

$E(X_i^2) = 1^2 \times p + 0^2 \times (1-p) = p.$
$Var(X_i) = p - (E(X))^2 = p - p^2 = p(1-p).$

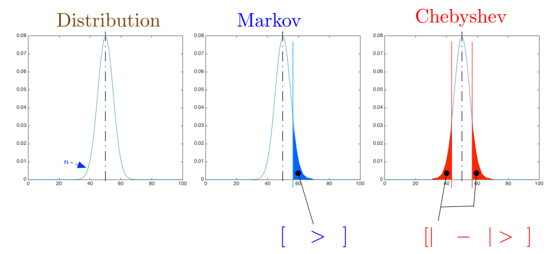
$p = 0 \implies Var(X_i) = 0$
$p = 1 \implies Var(X_i) = 0$

$X = X_1 + X_2 + \ldots X_n.$

X_i and X_j are independent: $Pr[X_i = 1 | X_j = 1] = Pr[X_i = 1].$

$$Var(X) = Var(X_1 + \cdots X_n) = np(1-p).$$

Inequalities: An Overview



Andrey Markov

Andrey (Andrei) Andreyevich Markov

Born 14 June 1856 N.S. Ryazan, Russian Empire
Died 20 July 1922 (aged 66) Petrograd, Russian SFSR

Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.

Pafnuty Chebyshev was one of his teachers.

Markov was an atheist. In 1912 he protested Leo Tolstoy's excommunication from the Russian Orthodox Church by requesting his own excommunication. The Church complied with his request.

Markov's inequality

The inequality is named after Andrey Markov, although it appeared earlier in the work of Pafnuty Chebyshev. It should be (and is sometimes) called Chebyshev's first inequality.

Theorem Markov's Inequality

Assume $f : \Re \to [0, \infty)$ is nondecreasing. Then,

$$Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.$$

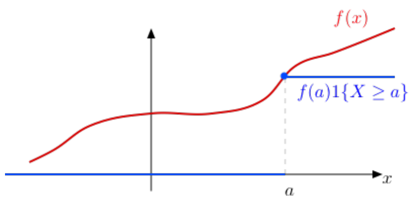
Proof:

Observe that
$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(X)/f(a)$, which holds since $f(\cdot) \geq 0$. Also, if $X \geq a$, it reads $1 \leq f(X)/f(a)$, which holds since $f(\cdot)$ is nondecreasing.

Taking the expectation yields the inequality, because expectation is monotone. □

A picture



$$f(a)1\{X \geq a\} \leq f(x) \Rightarrow 1\{X \geq a\} \leq \frac{f(X)}{f(a)}$$

$$\Rightarrow Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}$$

Markov Inequality Example: G(p)

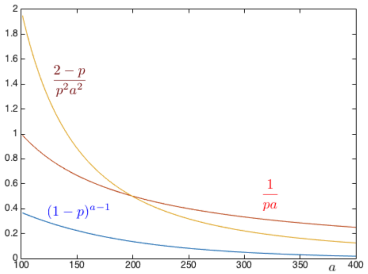
Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{2-p}{p^2a^2}.$$



Markov Inequality Example: $P(\lambda)$

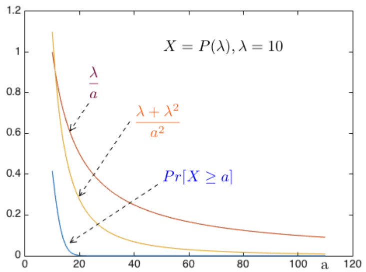
Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.$$

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda+\lambda^2}{a^2}.$$



Chebyshev's Inequality

This is Pafnuty's inequality:
Theorem:

$$Pr[|X - E[X]| > a] \leq \frac{var[X]}{a^2}, \text{ for all } a > 0.$$

Proof: Let $Y = |X - E[X]|$ and $f(y) = y^2$. Then,

$$Pr[Y \geq a] \leq \frac{E[f(Y)]}{f(a)} = \frac{var[X]}{a^2}.$$

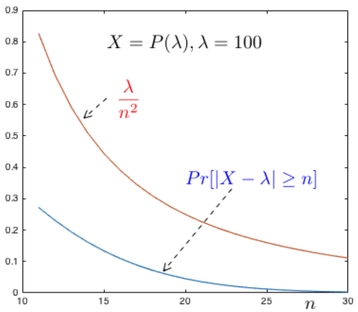
□

This result confirms that the variance measures the "deviations from the mean."

Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. Thus,

$$Pr[|X - \lambda| \geq n] \leq \frac{var[X]}{n^2} = \frac{\lambda}{n^2}.$$



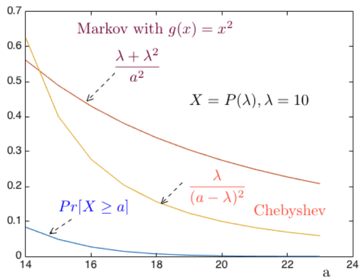
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. By Markov's inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda+\lambda^2}{a^2}.$$

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$.

Hence, for $a > \lambda$, $Pr[X \geq a] \leq Pr[|X - \lambda| \geq a - \lambda] \leq \frac{\lambda}{(a-\lambda)^2}$.



Fraction of H's

Here is a classical application of Chebyshev's inequality.

How likely is it that the fraction of H's differs from 50%?

Let $X_m = 1$ if the m-th flip of a fair coin is H and $X_m = 0$ otherwise.

Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{var[Y_n]}{(0.1)^2} = 100 var[Y_n].$$

Now,

$$var[Y_n] = \frac{1}{n^2}(var[X_1] + \cdots + var[X_n]) = \frac{1}{n} var[X_1] \leq \frac{1}{4n}.$$

$$Var(X_i) = p(1 - lp) \leq (.5)(.5) = \frac{1}{4}$$

Fraction of H's

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of Hs is within $\varepsilon > 0$ of 50% approaches 1:

$$Pr[|Y_n - 0.5| \leq \varepsilon] \to 1.$$

This is an example of the Law of Large Numbers.

We look at a general case next.

Weak Law of Large Numbers

Theorem Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean μ. Then, for all $\varepsilon > 0$,

$$Pr[|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon] \to 0, \text{ as } n \to \infty.$$

Proof:
Let $Y_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$\begin{aligned} Pr[|Y_n - \mu| \geq \varepsilon] &\leq \frac{var[Y_n]}{\varepsilon^2} = \frac{var[X_1 + \cdots + X_n]}{n^2\varepsilon^2} \\ &= \frac{nvar[X_1]}{n^2\varepsilon^2} = \frac{var[X_1]}{n\varepsilon^2} \to 0, \text{ as } n \to \infty. \end{aligned}$$

□

Summary

$\boxed{\text{Variance; Inequalities; WLLN}}$

- Variance: $var[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Fact: $var[aX + b]a^2 var[X]$
- Sum: X, Y, Z pairwise ind. $\Rightarrow var[X + Y + Z] = \cdots$
- Markov: $Pr[X \geq a] \leq E[f(X)]/f(a)$ where ...
- Chebyshev: $Pr[|X - E[X]| \geq a] \leq var[X]/a^2$
- WLLN: X_m i.i.d. $\Rightarrow \frac{X_1 + \cdots + X_n}{n} \approx E[X]$