CS70: Lecture 20.

Coupons; Independent Random Variables; Markov; Variance

1. Time to Collect Coupons
2. Review: Independence of Events
3. Independent RVs
4. Mutually independent RVs
5. Variance
6. Inequalities
   - Markov
   - Chebyshev
7. Weak Law of Large Numbers

Coupon Collectors Problem.

Experiment: Get coupons at random from n until collect all n coupons.
Outcomes: $\{123145...,56765...\}$
Random Variable: X - length of outcome.

Before: $Pr[X \geq n \ln 2n] \leq \frac{1}{2}$.

Today: $E[X]$?

Time to collect coupons

X-time to get n coupons.

X_1 - time to get first coupon. Note: $X_1 = 1$. $E(X_1) = 1$.

X_2 - time to get second coupon after getting first.

Pr[“get second coupon”|“got ~~milk~~ first coupon”] $= \frac{n-1}{n}$

$E[X_2]$? Geometric ! ! ! $\implies E[X_2] = \frac{1}{p} = \frac{1}{\frac{n-1}{n}} = \frac{n}{n-1}$.

Pr[“getting ith coupon|“got $i-1$rst coupons”] $= \frac{n-(i-1)}{n} = \frac{n-i+1}{n}$

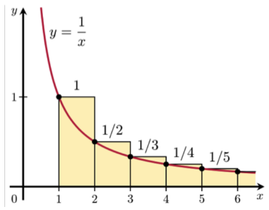
$E[X_i] = \frac{1}{p} = \frac{n}{n-i+1}, i = 1,2,\ldots,n$.

$$\begin{aligned} E[X] &= E[X_1] + \cdots + E[X_n] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} \\ &= n(1 + \frac{1}{2} + \cdots + \frac{1}{n}) =: nH(n) \approx n(\ln n + \gamma) \end{aligned}$$

Review: Harmonic sum

$$H(n) = 1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \int_1^n \frac{1}{x} dx = \ln(n).$$

.

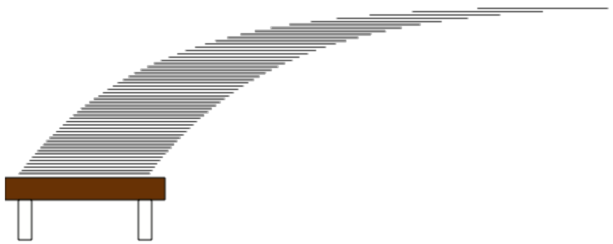


A good approximation is

$H(n) \approx \ln(n) + \gamma$ where $\gamma \approx 0.58$ (Euler-Mascheroni constant).

Harmonic sum: Paradox

Consider this stack of cards (no glue!):



If each card has length 2, the stack can extend $H(n)$ to the right of the table. As n increases, you can go as far as you want!

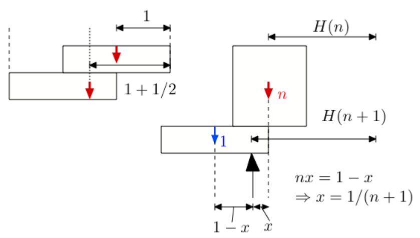
Paradox

par·a·dox
/ˈperəˌdäks/
noun

a statement or proposition that, despite sound (or apparently sound) reasoning from acceptable premises, leads to a conclusion that seems senseless, logically unacceptable, or self-contradictory.
"a potentially serious conflict between quantum mechanics and the general theory of relativity known as the information paradox"

- a seemingly absurd or self-contradictory statement or proposition that when investigated or explained may prove to be well founded or true.
  "in a paradox, he has discovered that stepping back from his job has increased the rewards he gleans from it"
  synonyms: contradiction, contradiction in terms, self-contradiction, inconsistency, incongruity; More
- a situation, person, or thing that combines contradictory features or qualities.
  "the mingling of deciduous trees with elements of desert flora forms a fascinating ecological paradox"

Stacking



The cards have width 2. Induction shows that the center of gravity after n cards is $H(n)$ away from the right-most edge.

Review: Independence of Events

- Events A, B are independent if $Pr[A \cap B] = Pr[A]Pr[B]$.
- Events A, B, C are mutually independent if

  A, B are independent, A, C are independent, B, C are independent

  and $Pr[A \cap B \cap C] = Pr[A]Pr[B]Pr[C]$.
- Events $\{A_n, n \geq 0\}$ are mutually independent if ....
- Example: $X, Y \in \{0,1\}$ two fair coin flips $\Rightarrow X, Y, X \oplus Y$ are pairwise independent but not mutually independent.
- Example: $X, Y, Z \in \{0,1\}$ three fair coin flips are mutually independent.

Independent Random Variables.

Definition: Independence

The random variables X and Y are **independent** if and only if

$$Pr[Y = b | X = a] = Pr[Y = b], \text{ for all } a \text{ and } b.$$

Fact:

X, Y are independent if and only if

$$Pr[X = a, Y = b] = Pr[X = a]Pr[Y = b], \text{ for all } a \text{ and } b.$$

Obvious from $Pr[A \cap B] = Pr[A|B]Pr[B]$ (Product rule.)

Independence: Examples

Example 1
Roll two die. X, Y = number of pips on the two dice. X, Y are independent.

Indeed: $Pr[X = a, Y = b] = \frac{1}{36}, Pr[X = a] = Pr[Y = b] = \frac{1}{6}$.

Example 2
Roll two die. X = total number of pips, Y = number of pips on die 1 minus number on die 2. X and Y are not independent.

Indeed: $Pr[X = 12, Y = 1] = 0 \neq Pr[X = 12]Pr[Y = 1] > 0$.

Example 3
Flip a fair coin five times, X = number of Hs in first three flips, Y = number of Hs in last two flips. X and Y are independent.

Indeed:

$$Pr[X = a, Y = b] = \binom{3}{a}\binom{2}{b}2^{-5} = \binom{3}{a}2^{-3} \times \binom{2}{b}2^{-2} = Pr[X = a]Pr[Y = b].$$

Mean of product of independent RV

Theorem
Let X, Y be independent RVs. Then

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

Proof:
Recall that $E[g(X,Y)] = \sum_{x,y} g(x,y)Pr[X = x, Y = y]$. Hence,

$$\begin{aligned}
E[XY] &= \sum_{x,y} xyPr[X = x, Y = y] = \sum_{x,y} xyPr[X = x]Pr[Y = y], \text{ by ind.} \\
&= \sum_x [\sum_y xyPr[X = x]Pr[Y = y]] = \sum_x [xPr[X = x](\sum_y yPr[Y = y])] \\
&= \sum_x [xPr[X = x]E[Y]] = E[X]E[Y].
\end{aligned}$$

□

Examples

(1) Assume that X, Y, Z are (pairwise) independent, with $E[X] = E[Y] = E[Z] = 0$ and $E[X^2] = E[Y^2] = E[Z^2] = 1$.

Then

$$\begin{aligned}
&E[(X + 2Y + 3Z)^2] = E[X^2 + 4Y^2 + 9Z^2 + 4XY + 12YZ + 6XZ] \\
&= 1 + 4 + 9 + 4 \times 0 + 12 \times 0 + 6 \times 0 \\
&= 14.
\end{aligned}$$

(2) Let X, Y be independent and $U[1, 2, \ldots n]$. Then

$$\begin{aligned}
E[(X - Y)^2] &= E[X^2 + Y^2 - 2XY] = 2E[X^2] - 2E[X]^2 \\
&= \frac{1 + 3n + 2n^2}{3} - \frac{(n+1)^2}{2}.
\end{aligned}$$

Mutually Independent Random Variables

Definition

X, Y, Z are mutually independent if

$$Pr[X = x, Y = y, Z = z] = Pr[X = x]Pr[Y = y]Pr[Z = z], \text{ for all } x, y, z.$$

Theorem
The events $A, B, C, \ldots$ are pairwise (resp. mutually) independent iff the random variables $1_A, 1_B, 1_C, \ldots$ are pairwise (resp. mutually) independent.
Proof:

$$Pr[1_A = 1, 1_B = 1, 1_C = 1] = Pr[A \cap B \cap C], \ldots$$

□

Functions of pairwise independent RVs

If X, Y, Z are pairwise independent, but not mutually independent, it may be that

$f(X)$ and $g(Y, Z)$ are not independent.

Example 1: Flip two fair coins,
$X = 1\{\text{coin 1 is } H\}, Y = 1\{\text{coin 2 is } H\}, Z = X \oplus Y$. Then, X, Y, Z are pairwise independent. Let $g(Y, Z) = Y \oplus Z$.
Then $g(Y, Z) = Y \oplus X \oplus Y = X$ is not independent of X.

Example 2: Let A, B, C be pairwise but not mutually independent in a way that A and $B \cap C$ are not independent. Let
$X = 1_A, Y = 1_B, Z = 1_C$. Choose $f(X) = X, g(Y, Z) = YZ$.

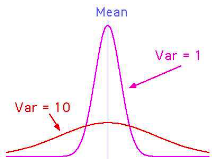
Quick Review.

Coupons; Independent Random Variables

- Expected time to collect n coupons is $nH(n) \approx n(\ln n + \gamma)$
- X, Y independent $\Leftrightarrow Pr[X \in A, Y \in B] = Pr[X \in A]Pr[Y \in B]$
- Then, $f(X), g(Y)$ are independent

  and $E[XY] = E[X]E[Y]$
- Mutual independence ....

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:

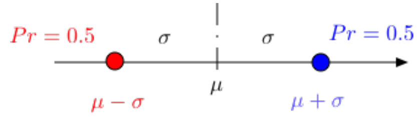
$$var[X] = E[X^2] - E[X]^2.$$

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 $E[X] = 1/p$.

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

$\implies E[X^2] = (2-p)/p^2$ 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.

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

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

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

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:

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

$\square$

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,

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

$\square$

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: $Pr[X_i=1|X_j=1] = Pr[X_i=1].$

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

Poisson Distribution: Definition

Definition Poisson Distribution with parameter $\lambda > 0$

$$X = P(\lambda) \Leftrightarrow Pr[X=m] = \frac{\lambda^m}{m!}e^{-\lambda}, m \geq 0.$$

Mean, Variance?

Ugh.

Recall that Poission is the limit of the Binomial with $p=\lambda/n$ as $n \to \infty$.

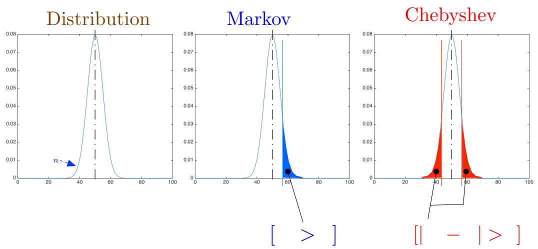
Mean: $pn = \lambda$

Variance: $p(1-p)n = \lambda - \lambda^2/n \to \lambda$.

$E(X^2)$? $Var(X) = E(X^2)-(E(X))^2$ or $E(X^2) = Var(X)+E(X)^2$.

$E(X^2) = \lambda + \lambda^2$.

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:

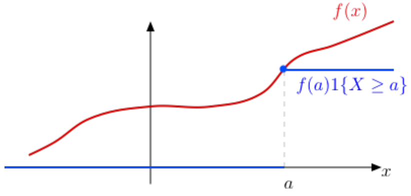
Observe that

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

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)

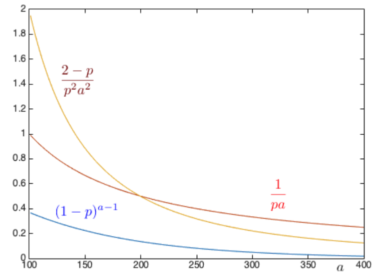
Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

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: $P(\lambda)$

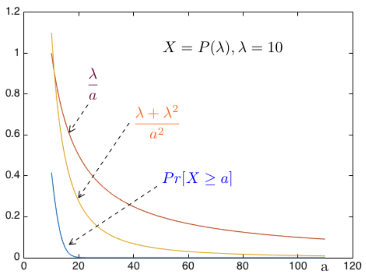
Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

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

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



Summary

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 ...