

# CS70

Coupon Collecting: Fun with harmonic numbers!

Memoryless Property.

Law of the unconscious statistician. (Hmmm.)

Variance/ Covariance.

## 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.

$Pr[\text{"get second coupon"} | \text{"got milk first coupon"}] = \frac{n-1}{n}$

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

$Pr[\text{"getting } i\text{th coupon"} | \text{"got } i-1 \text{rst coupons"}] = \frac{n-(i-1)}{n} = \frac{n-i+1}{n}$

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

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

# Coupons: Poll

Collect  $n$  coupons!

What's True?

(A)  $X_1 = \frac{n}{n} = 1$ .

(B)  $X_2 = \frac{n}{n-1}$ .

(C)  $Pr[\text{getting second} | \text{got first}] = \frac{n-1}{n}$ .

(D)  $E[X_2] = \frac{n}{n-1}$ .

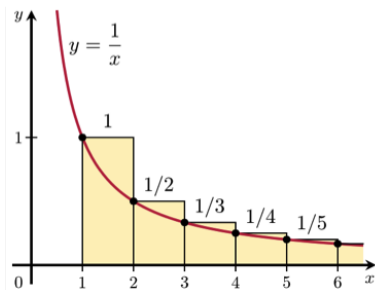
(E)  $E[X_n] = n$ .

(F)  $\sum_i E[X_i] = \sum_{i=0}^{n-1} \frac{n}{n-i}$

(G)  $\sum_i E[X_i] = \sum_{i=1}^n \frac{1}{i}$

## Review: Harmonic sum

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

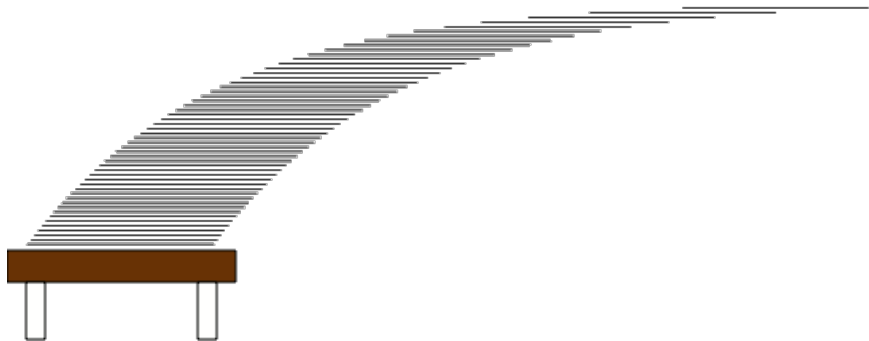


A good approximation is

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

## 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/

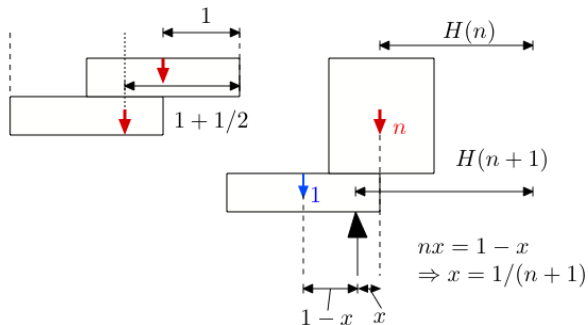
*noun*

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.

[Video.](#)

## Calculating $E[g(X)]$ : LOTUS

Let  $Y = g(X)$ . Assume that we know the distribution of  $X$ .

We want to calculate  $E[Y]$ .

**Method 1:** We calculate the distribution of  $Y$ :

$$Pr[Y = y] = Pr[X \in g^{-1}(y)] \text{ where } g^{-1}(x) = \{x \in \mathfrak{X} : g(x) = y\}.$$

This is typically rather tedious!

**Method 2:** We use the following result.

Called “Law of the unconscious statistician.”

**Theorem:**

$$E[g(X)] = \sum_x g(x) Pr[X = x].$$

**Proof:**

$$\begin{aligned} E[g(X)] &= \sum_{\omega} g(X(\omega)) Pr[\omega] = \sum_x \sum_{\omega \in X^{-1}(x)} g(X(\omega)) Pr[\omega] \\ &= \sum_x \sum_{\omega \in X^{-1}(x)} g(x) Pr[\omega] = \sum_x g(x) \sum_{\omega \in X^{-1}(x)} Pr[\omega] \\ &= \sum_x g(x) Pr[X = x]. \end{aligned}$$





# Poll.

Which is LOTUS?

(A)  $E[X] = \sum_{x \in \text{Range}(X)} g(x) Pr[g(X) = g(x)]$

(B)  $E[X] = \sum_{x \in \text{Range}(X)} g(x) Pr[X = x]$

(C)  $E[X] = \sum_{x \in \text{Range}(g)} x Pr[g(X) = x]$

# Geometric Distribution.

Experiment: flip a coin with heads prob.  $p$ . until Heads.

Random Variable  $X$ : number of flips.

And distribution is:

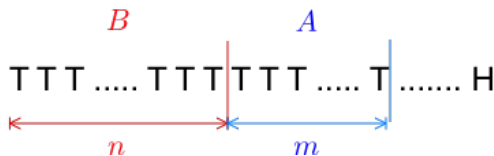
$$(A) X \sim G(p) : Pr[X = i] = (1 - p)^{i-1} p.$$

$$(B) X \sim B(p, n) : Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}.$$

$$(A) \text{ Distribution of } X \sim G(p): Pr[X = i] = (1 - p)^{i-1} p.$$

# Geometric Distribution: Memoryless - Interpretation

$$Pr[X > n + m | X > n] = Pr[X > m], m, n \geq 0.$$



$$Pr[X > n + m | X > n] = Pr[A|B] = Pr[A'] = Pr[X > m].$$

$A'$ : is  $m$  coin tosses before heads.

$A|B$ :  $m$  'more' coin tosses before heads.

The coin is memoryless, therefore, so is  $X$ .

Independent coin:  $Pr[H | \text{any previous set of coin tosses}] = p$

## Geometric Distribution: Memoryless by derivation.

Let  $X$  be  $G(p)$ . Then, for  $n \geq 0$ ,

$$\Pr[X > n] = \Pr[\text{first } n \text{ flips are } T] = (1 - p)^n.$$

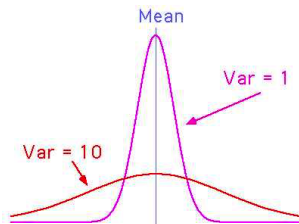
### Theorem

$$\Pr[X > n + m | X > n] = \Pr[X > m], m, n \geq 0.$$

### Proof:

$$\begin{aligned} \Pr[X > n + m | X > n] &= \frac{\Pr[X > n + m \text{ and } X > n]}{\Pr[X > n]} \\ &= \frac{\Pr[X > n + m]}{\Pr[X > n]} \\ &= \frac{(1 - p)^{n+m}}{(1 - p)^n} = (1 - p)^m \\ &= \Pr[X > m]. \end{aligned}$$

# Variance



The variance measures the deviation from the mean value.

**Definition:** The **variance** of  $X$  is

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

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

# Variance and Standard Deviation

**Fact:**

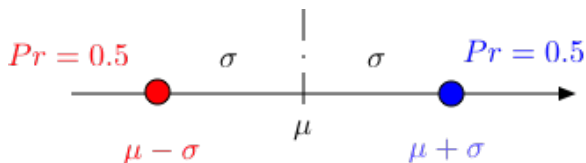
$$\text{var}[X] = E[X^2] - E[X]^2.$$

Indeed:

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

## 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,

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

## Example

Consider  $X$  with

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

Then

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

Also,

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

Thus,  $\sigma(X) = \sqrt{E[(X - E(X))^2]} \neq E[|X - E[X]|]$ !

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



## Uniform

Assume that  $Pr[X = i] = \frac{1}{n}$  for  $i \in \{1, \dots, n\}$ . 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,

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

This gives

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

(Sort of  $\int_0^{1/2} x^2 dx = \frac{x^3}{3}$ .)

## 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  $E[X] = 1/p$ .

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

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

## Fixed points.

Number of fixed points in a random permutation of  $n$  items.

“Number of student that get homework back.”

$$X = X_1 + X_2 + \dots + X_n$$

where  $X_i$  is indicator variable for  $i$ th student getting hw back.

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

$$E(X_i^2) = 1 \times Pr[X_i = 1] + 0 \times Pr[X_i = 0]$$

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

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

## Poll: fixed points.

What's true?

(A)  $X_i$  and  $X_j$  are independent.

(B)  $E[X_i X_j] = Pr[X_i X_j = 1]$

(C)  $Pr[X_i X_j] = \frac{(n-2)!}{n!}$

(D)  $X_i^2 = X_i$ .

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

### Proof:

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

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



## Independent random variables.

Independent:  $P[X = a, Y = b] = Pr[X = a]Pr[Y = b]$

Fact:  $E[XY] = E[X]E[Y]$  for independent random variables.

$$\begin{aligned} E[XY] &= \sum_a \sum_b a \times b \times Pr[X = a, Y = b] \\ &= \sum_a \sum_b a \times b \times Pr[X = a]Pr[Y = b] \\ &= \left(\sum_a aPr[X = a]\right)\left(\sum_b bPr[Y = b]\right) \\ &= E[X]E[Y] \end{aligned}$$

# Variance of sum of two independent random variables

## Theorem:

If  $X$  and  $Y$  are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

## 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,

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



## Variance of sum of independent random variables

### Theorem:

If  $X, Y, Z, \dots$  are pairwise independent, then

$$\text{var}(X + Y + Z + \dots) = \text{var}(X) + \text{var}(Y) + \text{var}(Z) + \dots .$$

### Proof:

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

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

Then, by independence,

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

Hence,

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



## Variance of Binomial Distribution.

Flip coin with heads probability  $p$ .

$X$ - how many heads?

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

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

$$\text{Var}(X_i) = p - (E(X_i))^2 = p - p^2 = p(1 - p).$$

$$p = 0 \implies \text{Var}(X_i) = 0$$

$$p = 1 \implies \text{Var}(X_i) = 0$$

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

$X_i$  and  $X_j$  are independent:  $\text{Pr}[X_i = 1 | X_j = 1] = \text{Pr}[X_i = 1]$ .

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = np(1 - p).$$

## Poisson Distribution: Variance.

**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 Poisson is the limit of the Binomial with  $p = \lambda/n$  as  $n \rightarrow \infty$ .

Mean:  $pn = \lambda$

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

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

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

# Covariance

**Definition** The covariance of  $X$  and  $Y$  is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Fact**

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

**Proof:**

Think about  $E[X] = E[Y] = 0$ . Just  $E[XY]$ .

□ish.

For the sake of completeness.

$$\begin{aligned} E[(X - E[X])(Y - E[Y])] &= E[XY - E[X]Y - XE[Y] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]. \end{aligned}$$

# Correlation

**Definition** The correlation of  $X, Y$ ,  $Cor(X, Y)$  is

$$corr(X, Y) : \frac{cov(X, Y)}{\sigma(X)\sigma(Y)}.$$

**Theorem:**  $-1 \leq corr(X, Y) \leq 1$ .

**Proof:** Idea:  $(a - b)^2 > 0 \rightarrow a^2 + b^2 \geq 2ab$ .

Simple case:  $E[X] = E[Y] = 0$  and  $E[X^2] = E[Y^2] = 1$ .

$Cor(X, Y) = E[XY]$ .

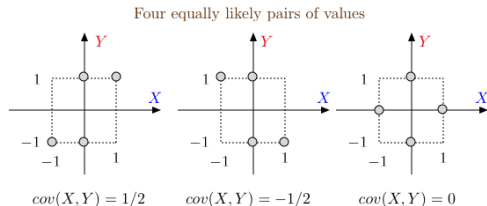
$$E[(X - Y)^2] = E[X^2] + E[Y^2] - 2E[XY] = 2(1 - E[XY]) \geq 0 \\ \rightarrow E[XY] \leq 1.$$

$$E[(X + Y)^2] = E[X^2] + E[Y^2] + 2E[XY] = 2(1 + E[XY]) \geq 0 \\ \rightarrow E[XY] \geq -1.$$



Shifting and scaling doesn't change correlation.

# Examples of Covariance



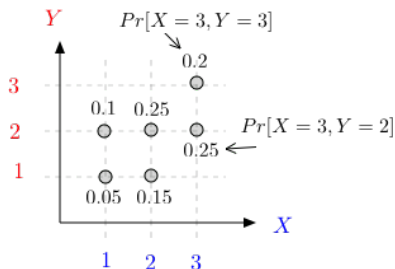
Note that  $E[X] = 0$  and  $E[Y] = 0$  in these examples. Then  $cov(X, Y) = E[XY]$ .

When  $cov(X, Y) > 0$ , the RVs  $X$  and  $Y$  tend to be large or small together.  $X$  and  $Y$  are said to be **positively correlated**.

When  $cov(X, Y) < 0$ , when  $X$  is larger,  $Y$  tends to be smaller.  $X$  and  $Y$  are said to be **negatively correlated**.

When  $cov(X, Y) = 0$ , we say that  $X$  and  $Y$  are **uncorrelated**.

## Examples of Covariance



$$E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3$$

$$E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8$$

$$E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2$$

$$E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4$$

$$E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \dots + 3 \times 3 \times 0.2 = 4.85$$

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y] = .25$$

$$\text{var}[X] = E[X^2] - E[X]^2 = .51$$

$$\text{var}[Y] = E[Y^2] - E[Y]^2 = .4$$

$$\text{corr}(X, Y) \approx 0.55$$

# Properties of Covariance

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

## Fact

(a)  $\text{var}[X] = \text{cov}(X, X)$

(b)  $X, Y$  independent  $\Rightarrow \text{cov}(X, Y) = 0$

(c)  $\text{cov}(a + X, b + Y) = \text{cov}(X, Y)$

(d)  $\text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V).$

## Proof:

(a)-(b)-(c) are obvious.

(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,

$$\begin{aligned}\text{cov}(aX + bY, cU + dV) &= E[(aX + bY)(cU + dV)] \\ &= ac \cdot E[XU] + ad \cdot E[XV] + bc \cdot E[YU] + bd \cdot E[YV] \\ &= ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V).\end{aligned}$$





# Summary

## Variance

- ▶ **Variance:**  $\text{var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$
- ▶ **Fact:**  $\text{var}[aX + b] = a^2 \text{var}[X]$
- ▶ **Sum:**  $X, Y, Z$  pairwise ind.  $\Rightarrow \text{var}[X + Y + Z] = \dots$

## Random Variables so far.

Probability Space:  $\Omega$ ,  $Pr : \Omega \rightarrow [0, 1]$ ,  $\sum_{\omega \in \Omega} Pr(\omega) = 1$ .

Random Variables:  $X : \Omega \rightarrow R$ .

Associated event:  $Pr[X = a] = \sum_{\omega: X(\omega)=a} Pr(\omega)$

$X$  and  $Y$  independent  $\iff$  all associated events are independent.

Expectation:  $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$ .

Linearity:  $E[X + Y] = E[X] + E[Y]$ .

Variance:  $Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$

For independent  $X, Y$ ,  $Var(X + Y) = Var(X) + Var(Y)$ .

Also:  $Var(cX) = c^2 Var(X)$  and  $Var(X + b) = Var(X)$ .

Poisson:  $X \sim P(\lambda)$   $E(X) = \lambda$ ,  $Var(X) = \lambda$ .

Binomial:  $X \sim B(n, p)$   $E(X) = np$ ,  $Var(X) = np(1 - p)$

Uniform:  $X \sim U\{1, \dots, n\}$   $E[X] = \frac{n+1}{2}$ ,  $Var(X) = \frac{n^2-1}{12}$ .

Geometric:  $X \sim G(p)$   $E(X) = \frac{1}{p}$ ,  $Var(X) = \frac{1-p}{p^2}$