

CS70.

1. Random Variables: Brief Review
2. Joint Distributions.
3. Linearity of Expectation

Random Variables: Definitions

Definition

A **random variable**, X , for a random experiment with sample space Ω is a **function** $X : \Omega \rightarrow \mathfrak{R}$.

Thus, $X(\cdot)$ assigns a real number $X(\omega)$ to each $\omega \in \Omega$.

Definitions

(a) For $a \in \mathfrak{R}$, one defines

$$X^{-1}(a) := \{\omega \in \Omega \mid X(\omega) = a\}.$$

(b) For $A \subset \mathfrak{R}$, one defines

$$X^{-1}(A) := \{\omega \in \Omega \mid X(\omega) \in A\}.$$

(c) The probability that $X = a$ is defined as

$$Pr[X = a] = Pr[X^{-1}(a)].$$

(d) The probability that $X \in A$ is defined as

$$Pr[X \in A] = Pr[X^{-1}(A)].$$

(e) The **distribution** of a random variable X , is

$$\{(a, Pr[X = a]) : a \in \mathcal{A}\},$$

where \mathcal{A} is the *range* of X . That is, $\mathcal{A} = \{X(\omega), \omega \in \Omega\}$.

Some Distributions.

Binomial Distribution: $B(n, p)$, For $0 \leq i \leq n$,
 $Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}$. Geometric Distribution: $G(p)$, For $i \geq 1$,
 $Pr[X = i] = (1 - p)^{i-1} p$. Poisson: Next up.

Poisson: Motivation and derivation.

McDonalds: How many arrive at McDonalds in an hour?

Know: average is λ .

What is distribution?

Example: $Pr[2\lambda \text{ arrivals}]?$

Assumption: “arrivals are independent.”

Derivation: cut hour into n intervals of length $1/n$.

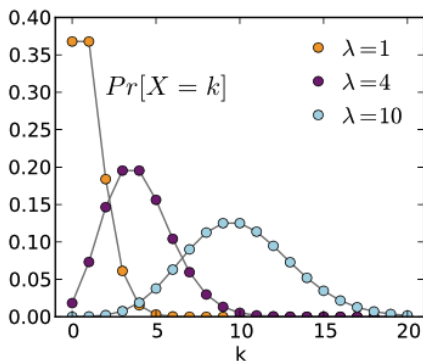
$Pr[\text{two arrivals}]$ is “ $(\lambda/n)^2$ ” or small if n is large.

Model with binomial.

Poisson

Experiment: flip a coin n times. The coin is such that $Pr[H] = \lambda/n$.
Random Variable: X - number of heads. Thus, $X = B(n, \lambda/n)$.

Poisson Distribution is distribution of X “for large n .”



Poisson

Experiment: flip a coin n times. The coin is such that $Pr[H] = \lambda/n$.
Random Variable: X - number of heads. Thus, $X = B(n, \lambda/n)$.

Poisson Distribution is distribution of X “for large n .”

We expect $X \ll n$. For $m \ll n$ one has

$$\begin{aligned}Pr[X = m] &= \binom{n}{m} p^m (1-p)^{n-m}, \text{ with } p = \lambda/n \\&= \frac{n(n-1)\cdots(n-m+1)}{m!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \\&= \frac{n(n-1)\cdots(n-m+1)}{n^m} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^{n-m} \\&\approx^{(1)} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^{n-m} \approx^{(2)} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n \approx \frac{\lambda^m}{m!} e^{-\lambda}.\end{aligned}$$

For (1) we used $m \ll n$; for (2) we used $(1 - a/n)^n \approx e^{-a}$.

Expectation - Definition

Definition: The **expected value** (or mean, or expectation) of a random variable X is

$$E[X] = \sum_a a \times Pr[X = a].$$

Theorem:

$$E[X] = \sum_{\omega} X(\omega) \times Pr[\omega].$$

Proof:

$$\begin{aligned} E[X] &= \sum_a a \times Pr[X = a] \\ &= \sum_a a \times \sum_{\omega: X(\omega)=a} Pr[\omega] \\ &= \sum_a \sum_{\omega: X(\omega)=a} X(\omega) Pr[\omega] \\ &= \sum_{\omega} X(\omega) Pr[\omega] \end{aligned}$$



Poisson Distribution: Definition and Mean

Definition Poisson Distribution with parameter $\lambda > 0$

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

Fact: $E[X] = \lambda$.

Proof:

$$\begin{aligned} E[X] &= \sum_{m=1}^{\infty} m \times \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{(m-1)!} \\ &= e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} = e^{-\lambda} \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \\ &= e^{-\lambda} \lambda e^{\lambda} = \lambda. \end{aligned}$$



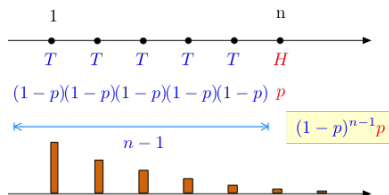
Simeon Poisson

The Poisson distribution is named after:



Equal Time: B. Geometric

The geometric distribution is named after:



I could not find a picture of D. Binomial, sorry.

Recall: An Example

Flip a fair coin three times.

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

$$X = \text{number of } H\text{'s: } \{3, 2, 2, 2, 1, 1, 1, 0\}.$$

Thus,

$$\sum_{\omega} X(\omega) Pr[\omega] = \{3 + 2 + 2 + 2 + 1 + 1 + 1 + 0\} \times \frac{1}{8}.$$

Also,

$$\sum_a a \times Pr[X = a] = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 0 \times \frac{1}{8}.$$

Win or Lose.

Expected winnings for heads/tails games, with 3 flips?

Recall the definition of the random variable X :

$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \rightarrow \{3, 1, 1, -1, 1, -1, -1, -3\}$.

$$E[X] = 3 \times \frac{1}{8} + 1 \times \frac{3}{8} - 1 \times \frac{3}{8} - 3 \times \frac{1}{8} = 0.$$

Can you ever win 0?

Apparently: expected value is not a common value, by any means.

The expected value of X is not the value that you expect!

It is the average value per experiment, if you perform the experiment many times:

$$\frac{X_1 + \cdots + X_n}{n}, \text{ when } n \gg 1.$$

The fact that this average converges to $E[X]$ is a theorem: the [Law of Large Numbers](#). (See later.)

Multiple Random Variables.

Experiment: toss two coins. $\Omega = \{HH, TH, HT, TT\}$.

$$X_1(\omega) = \begin{cases} 1, & \text{if coin 1 is heads} \\ 0, & \text{otherwise} \end{cases}$$

$$X_2(\omega) = \begin{cases} 1, & \text{if coin 2 is heads} \\ 0, & \text{otherwise} \end{cases}$$

Multiple Random Variables: setup.

Joint Distribution: $\{(a, b, Pr[X = a, Y = b]) : a \in \mathcal{A}, b \in \mathcal{B}\}$, where \mathcal{A} (\mathcal{B}) is possible values of X (Y).

$$\sum_{a \in \mathcal{A}, b \in \mathcal{B}} Pr[X = a, Y = b] = 1$$

Marginal for X : $Pr[X = a] = \sum_{b \in \mathcal{B}} Pr[X = a, Y = b]$.

Marginal for Y : $Pr[Y = b] = \sum_{a \in \mathcal{A}} Pr[X = a, Y = b]$.

X/Y	1	2	3	X
1	.2	.1	.1	.4
2	0	0	.3	.3
3	.1	0	.2	.3
Y	.3	.1	.2	.6

Conditional Probability: $Pr[X = a | Y = b] = \frac{Pr[X=a, Y=b]}{Pr[Y=b]}$.

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

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

Follows 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 H s in first three flips, Y = number of H s in last two flips. X and Y are independent.

Indeed:

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

Linearity of Expectation

Theorem:

$$E[X + Y] = E[X] + E[Y]$$

$$E[cX] = cE[X]$$

Proof: $E[X] = \sum_{\omega \in \Omega} X(\omega) \times P[\omega]$.

$$\begin{aligned} E[X + Y] &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega))Pr[\omega] \\ &= \sum_{\omega \in \Omega} X(\omega)Pr[\omega] + Y(\omega)Pr[\omega] \\ &= \sum_{\omega \in \Omega} X(\omega)Pr[\omega] + \sum_{\omega \in \Omega} Y(\omega)Pr[\omega] \\ &= E[X] + E[Y] \end{aligned}$$

Indicators

Definition

Let A be an event. The random variable X defined by

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

is called the **indicator** of the event A .

Note that $Pr[X = 1] = Pr[A]$ and $Pr[X = 0] = 1 - Pr[A]$.

Hence,

$$E[X] = 1 \times Pr[X = 1] + 0 \times Pr[X = 0] = Pr[A].$$

This random variable $X(\omega)$ is sometimes written as

$$1_{\{\omega \in A\}} \text{ or } 1_A(\omega).$$

Thus, we will write $X = 1_A$.

Linearity of Expectation

Theorem: Expectation is linear

$$E[a_1 X_1 + \cdots + a_n X_n] = a_1 E[X_1] + \cdots + a_n E[X_n].$$

Proof:

$$\begin{aligned} E[a_1 X_1 + \cdots + a_n X_n] &= \sum_{\omega} (a_1 X_1 + \cdots + a_n X_n)(\omega) Pr[\omega] \\ &= \sum_{\omega} (a_1 X_1(\omega) + \cdots + a_n X_n(\omega)) Pr[\omega] \\ &= a_1 \sum_{\omega} X_1(\omega) Pr[\omega] + \cdots + a_n \sum_{\omega} X_n(\omega) Pr[\omega] \\ &= a_1 E[X_1] + \cdots + a_n E[X_n]. \end{aligned}$$



Note: If we had defined $Y = a_1 X_1 + \cdots + a_n X_n$ and had tried to compute $E[Y] = \sum_y y Pr[Y = y]$, we would have been in trouble!

Using Linearity - 1: Pips (dots) on dice

Roll a die n times.

X_m = number of pips on roll m .

$X = X_1 + \cdots + X_n$ = total number of pips in n rolls.

$$\begin{aligned} E[X] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n], \text{ by linearity} \\ &= nE[X_1], \text{ because the } X_m \text{ have the same distribution} \end{aligned}$$

Now,

$$E[X_1] = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{6 \times 7}{2} \times \frac{1}{6} = \frac{7}{2}.$$

Hence,

$$E[X] = \frac{7n}{2}.$$

Note: Computing $\sum_x xPr[X = x]$ directly is not easy!

Using Linearity - 2: Fixed point.

Hand out assignments at random to n students.

X = number of students that get their own assignment back.

$X = X_1 + \dots + X_n$ where

$X_m = 1$ {student m gets his/her own assignment back}.

One has

$$\begin{aligned} E[X] &= E[X_1 + \dots + X_n] \\ &= E[X_1] + \dots + E[X_n], \text{ by linearity} \\ &= nE[X_1], \text{ because all the } X_m \text{ have the same distribution} \\ &= nPr[X_1 = 1], \text{ because } X_1 \text{ is an indicator} \\ &= n(1/n), \text{ because student 1 is equally likely} \\ &\quad \text{to get any one of the } n \text{ assignments} \\ &= 1. \end{aligned}$$

Note that linearity holds even though the X_m are not independent (whatever that means).

Note: What is $Pr[X = m]$? Tricky

Using Linearity - 3: Binomial Distribution.

Flip n coins with heads probability p . X - number of heads

Binomial Distribution: $Pr[X = i]$, for each i .

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

$$E[X] = \sum_i i \times Pr[X = i] = \sum_i i \times \binom{n}{i} p^i (1-p)^{n-i}.$$

Uh oh. ... Or... a better approach: Let

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

$$E[X_i] = 1 \times Pr[\text{"heads"}] + 0 \times Pr[\text{"tails"}] = p.$$

Moreover $X = X_1 + \dots + X_n$ and

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_n] = n \times E[X_i] = np.$$

Using Linearity - 4

Assume A and B are disjoint events. Then $1_{A \cup B}(\omega) = 1_A(\omega) + 1_B(\omega)$.

Taking expectation, we get

$$Pr[A \cup B] = E[1_{A \cup B}] = E[1_A + 1_B] = E[1_A] + E[1_B] = Pr[A] + Pr[B].$$

In general, $1_{A \cup B}(\omega) = 1_A(\omega) + 1_B(\omega) - 1_{A \cap B}(\omega)$.

Taking expectation, we get $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$.

Observe that if $Y(\omega) = b$ for all ω , then $E[Y] = b$.

Thus, $E[X + b] = E[X] + b$.

Empty Bins

Experiment: Throw m balls into n bins.

Y - number of empty bins.

Distribution is horrible.

Expectation? X_i - indicator for bin i being empty.

$$Y = X_1 + \cdots + X_n.$$

$$\Pr[X_1 = 1] = \left(1 - \frac{1}{n}\right)^m. \rightarrow E[Y] = n\left(1 - \frac{1}{n}\right)^m.$$

For $n = m$ and large n , $\left(1 - 1/n\right)^n \approx \frac{1}{e}$.

$\frac{n}{e} \approx 0.368n$ empty bins on average.

Coupon Collectors Problem.

Experiment: Get random coupon from n until get all n coupons.

Outcomes: {123145..., 56765...}

Random Variable: X - length of outcome.

Today: $E[X]$?

Geometric Distribution: Expectation

$$X =_D G(p), \text{ i.e., } Pr[X = n] = (1 - p)^{n-1} p, n \geq 1.$$

One has

$$E[X] = \sum_{n=1}^{\infty} n Pr[X = n] = \sum_{n=1}^{\infty} n(1 - p)^{n-1} p.$$

Thus,

$$\begin{aligned} E[X] &= p + 2(1 - p)p + 3(1 - p)^2 p + 4(1 - p)^3 p + \dots \\ (1 - p)E[X] &= (1 - p)p + 2(1 - p)^2 p + 3(1 - p)^3 p + \dots \\ pE[X] &= p + (1 - p)p + (1 - p)^2 p + (1 - p)^3 p + \dots \\ &\quad \text{by subtracting the previous two identities} \\ &= \sum_{n=1}^{\infty} Pr[X = n] = 1. \end{aligned}$$

Hence,

$$E[X] = \frac{1}{p}.$$

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

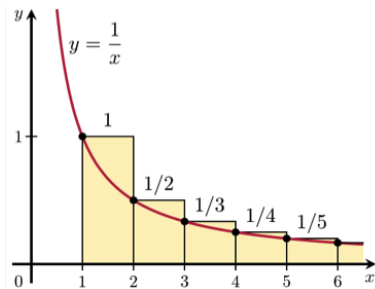
$Pr[\text{"getting } i\text{th coupon"} | \text{"got } i-1 \text{rst coupons"}] = \frac{n-(i-1)}{n} = \frac{n-i+1}{n}$

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

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

Review: Harmonic sum

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

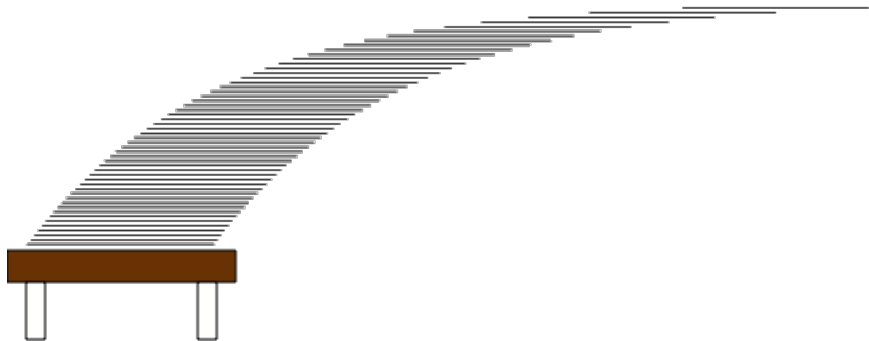


A good approximation is

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

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/

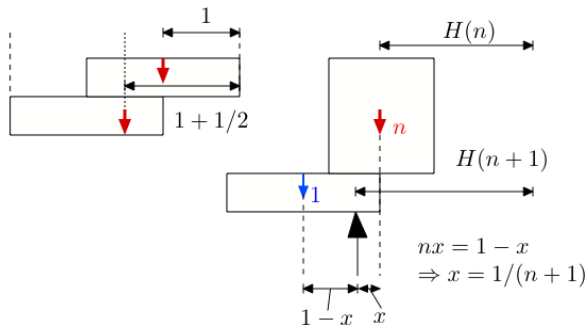
noun

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 $E[g(X)]$

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.

Theorem:

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

Proof:

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



An Example

Let X be uniform in $\{-2, -1, 0, 1, 2, 3\}$.

Let also $g(X) = X^2$. Then (method 2)

$$\begin{aligned} E[g(X)] &= \sum_{x=-2}^3 x^2 \frac{1}{6} \\ &= \{4 + 1 + 0 + 1 + 4 + 9\} \frac{1}{6} = \frac{19}{6}. \end{aligned}$$

Method 1 - We find the distribution of $Y = X^2$:

$$Y = \begin{cases} 4, & \text{w.p. } \frac{2}{6} \\ 1, & \text{w.p. } \frac{2}{6} \\ 0, & \text{w.p. } \frac{1}{6} \\ 9, & \text{w.p. } \frac{1}{6}. \end{cases}$$

Thus,

$$E[Y] = 4 \frac{2}{6} + 1 \frac{2}{6} + 0 \frac{1}{6} + 9 \frac{1}{6} = \frac{19}{6}.$$

Summary

Probability Space: Ω , $Pr[\omega] \geq 0$, $\sum_{\omega} Pr[\omega] = 1$.

Random Variable: Function on Sample Space.

Distribution: Function $Pr[X = a] \geq 0$. $\sum_a Pr[X = a] = 1$.

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

Many Random Variables: each one function on a sample space.

Joint Distributions: Function $Pr[X = a, Y = b] \geq 0$.

$\sum_{a,b} Pr[X = a, Y = b] = 1$.

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

Applications: compute expectations by decomposing.

Indicators: Empty bins, Fixed points.

Time to Coupon: Sum times to “next” coupon.

$Y = f(X)$ is Random Variable.

Distribution of Y from distribution of X .