Lecture 3

- **Read:** Chapter 2.1-2.8.
- Discrete Random Variables
- Definitions
- Probability Mass Function
- Some Useful Discrete Random Variables
- Cumulative Distribution Function (CDF)
- Averages
- Functions of a Random Variable
- Expected Value of a Derived Random Variable
- Variance and Standard Deviation

Random Variables

- **Experiment:** Procedure + Observation
- Observation is a particular outcome
- **Random variable:** Assign a real number to each outcome

  Possibilities: An RV will be denoted by X

  1. RV may be the observation
     e.g., roll of die
  2. RV is a function of the observation
     e.g., {heads, tails}
     X: heads $\to \pi$
     tails $\to 0$
  3. RV could be a function of another RV
     e.g., $Y = cos(X)$

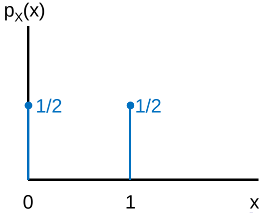
Discrete Random Variable (RV)

- S_X = range of X (set of possible values X can take)
- S_X is discrete $\Rightarrow$ S_X has a countable number of elements
  e.g., $S_X = \{1, 2, 3, 4, 5, 6\}$ ✓
  $S_X = \mathbb{Z}$ (set of integers) ✓
  $S_X = [0, 1]$ is not a countable set ✗
- A discrete RV has probability mass function (PMF)

$$\underbrace{p_X(x)}_{PMF} = P[X = x],\ x \in S_X$$

e.g., suppose $S_X = \{0, 1\}$ $\quad p_X(0) = 1/2$

$p_X(1) = 1/2$



Probability Mass Function (PMF)

Convention: uppercase characters for RVs, lowercase characters for numerical values the RV can take.

To compute PMF $p_X(x)$:

1. Pick an x, collect all samples that give rise to $X = x$
2. Add their probabilities
3. Repeat for all x

PMF and Its Properties

- $x \in S_X, \qquad p_X(x) \geq 0$
- $\sum_{x \in S_X} p_X(x) = 1$
- For an event $B \subset S_X$,

$$P[B] = P[x \in B] = \sum_{x \in B} p_X(x)$$

(e.g., $B = \{0, 1\}$)

Bernoulli RV

A RV is a Bernoulli RV if it can take only two values

$$\text{e.g., } p_X(x) = \begin{cases} 1-p & , x = 0 \\ p & , x = 1 \\ 0 & , \text{otherwise} \end{cases}$$

We say $X \sim \text{Bernoulli}(p)$, where "$\sim$" signifies X has a Bernoulli PMF with parameter p.
Example:

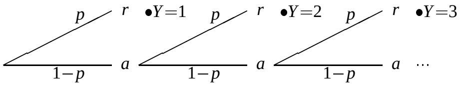
- Get the phone number of a random student
- Let $X = 0$ if the last digit is even
- Otherwise, let $X = 1$

With multiple Bernoulli trials, we can construct more complicated RVs.

Geometric RV

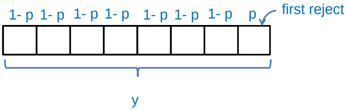
Example:

- Suppose a circuit is rejected with probability p
- Let $Y = \#$ of circuits tested up to and including the first rejection



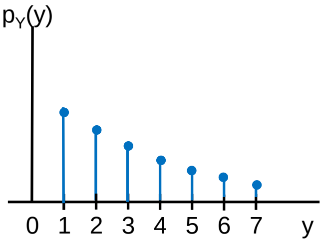
- $S_Y = \{1, 2, 3, 4, \ldots\}$
- $P[Y = 1] = p$, $P[Y = 2] = (1-p)p$, $P[Y = 3] = (1-p)^2 p$, and in general, $P[Y = y] = (1-p)^{y-1}p$

$$p_Y(y) = \begin{cases} (1-p)^{y-1}p & , \ y = 1, 2 \ldots \\ 0 & , \ \text{otherwise} \end{cases}$$



Geometric RV (cont.)

- Y is referred to as a geometric RV because the probabilities in the PMF decline geometrically

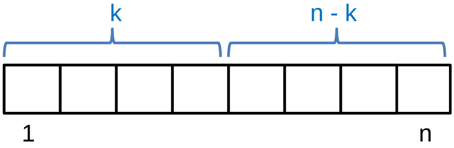


Binomial RV

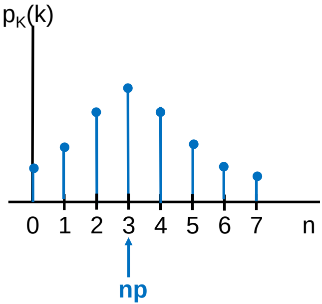
Example:

- Test n circuits, each circuit is rejected with probability p
- Let $K = \#$ of rejected circuits
  $S_K = \{0, 1, 2, ..., n-1, n\}$

$$p_K(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & , k = 0, ..., n \\ 0 & , \text{otherwise} \end{cases}$$



Binomial RV (cont.)



Binomial RV Example: Service Facility Design

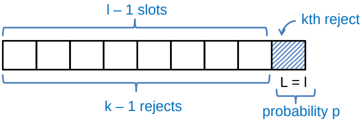
- n: customers
- p: probability customer requires service
- s: no. of service persons
- X: no. of service requests (RV)

$$
\begin{aligned}
P[X > s] &= P[X = s + 1] + \ldots + P[X = n] \\
&= p_X(s + 1) + \ldots + p_X(n) \\
&= \sum_{i=s+1}^{n} \binom{n}{i} p^i (1 - p)^{n-i}
\end{aligned}
$$

Pascal RV

Example:

- Suppose a circuit is rejected with probability p
- Let $L = \#$ of tests until we see k rejects



$$P[L = l] = P[k-1 \text{ rejects in } l-1 \text{ attempts, success on attempt } l]$$
$$= \binom{l-1}{k-1} p^{k-1} (1-p)^{l-1-(k-1)} \times p$$

- $S_L = \{k, k+1, k+2, \dots\}$

$$p_L(l) = \begin{cases} \binom{l-1}{k-1} p^k (1-p)^{l-k} & , \ l = k, k+1, k+2, \dots \\ 0 & , \text{ otherwise} \end{cases}$$

Random Variable Examples: Lottery

Example: Lottery $\rightarrow$ 2 possible outcomes: Win or Lose

1. What distribution would you use to model 1 play of the lottery?

   $\Rightarrow$ Bernoulli
   $$X = \begin{cases} 1 & \text{, with probability } p \\ 0 & \text{, with probability } (1-p) \end{cases}$$

   $1 =$ win, p is small (e.g., 10^{-12})

2. I keep playing the lottery until I win. How long until I win for the first time?
   $\Rightarrow$ Geometric

3. How long until I have won k times?
   $\Rightarrow$ Pascal

4. How many times will I win if I play n times?
   $\Rightarrow$ Binomial

Random Variable Examples: Summary

- **Bernoulli:** Number of successes in one trial
- **Geometric:** Number of trials until first success
- **Binomial:** Number of successes in n trials
- **Pascal:** Number of trials until success k

Discrete Uniform RV

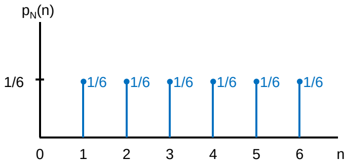
X is a discrete uniform random variable if the PMF of X has the form

$$p_X(x) = \begin{cases} \frac{1}{(l-k+1)} & , x = k, k+1, k+2, ..., l \\ 0 & , \text{otherwise} \end{cases}$$

where the parameters k and l are integers such that $k < l$

Example:

- Roll a fair die
- N is a discrete uniform random variable with $k = 1$ and $l = 6$



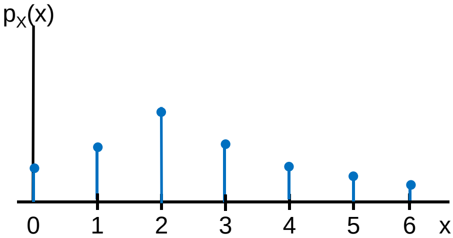
Poisson RV

- Used to count arrivals (of something)
  - Arrival of information requests at a WWW server, initiation of telephone calls, emission of particles from a radioactive source
- Arrival rate, λ arrivals/second and a time interval, T seconds
  - In this time interval, the number of arrivals X has a Poisson PMF with $\alpha = \lambda T =$ "avg # of arrivals"
- X is a Poisson random variable if its PMF is

$$p_X(x) = \begin{cases} \frac{(\lambda T)^x}{x!} e^{-\lambda T} & , x = 0, 1, 2, \ldots \\ 0 & , \text{otherwise} \end{cases}$$

- While the time of each occurrence is completely random, there is a known average number of occurrences per unit time

Poisson RV (cont.)



Binomial and Poisson Distributions

Theorem:

- Perform n Bernoulli trials
- In each trial, let the probability of success be α/n, where $\alpha > 0$ is a constant and $n > \alpha$
- Let the random variable K_n be the number of successes in the n trials
- As $n \to \infty$, $p_{K_n}(k)$ converges to the PMF of a Poisson(α) random variable

Discrete RV and PMF Summary

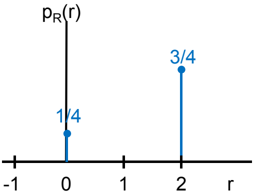
- RV $\rightarrow$ assigns a number X to each outcome
- Discrete RVs $\rightarrow$ take discrete set of values S_X
- PMF of X: $p_X(x) = P[X = x]$

Cumulative Distribution Function (CDF)

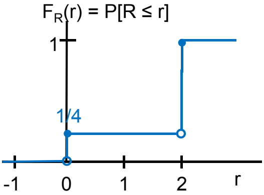
- **Definition:** The CDF of a RV X is a function

$$F_X(x) = P[X \leq x]$$

- **Example:** At the discontinuities $r = 0$ and $r = 2$, $F_R(r)$ takes on the upper values (right hand limit)



P[R = 0] = 1/4
P[R = 2] = 3/4



P[R ≤ 0] = 1/4
P[R ≤ r] = 1/4, for 0 ≤ r < 2
P[R ≤ 2] = P[R = 0] + P[R = 2] = 1

CDF Properties

For any discrete RV, with range $S_X = \{x_1, x_2, x_3, \dots\}$ satisfying $x_1 < x_2 < x_3 < \dots$

1. $F_X(-\infty) = 0, \qquad F_X(+\infty) = 1$
2. If $x' \geq x$, then $F_X(x') \geq F_X(x)$

   Proof: $F_X(x') = P[X \leq x'] = P[X \leq x] + \underbrace{P[x \leq X \leq x']}_{\geq 0}$

   $$\Rightarrow P[X \leq x'] \geq P[X \leq x] = F_X(x)$$

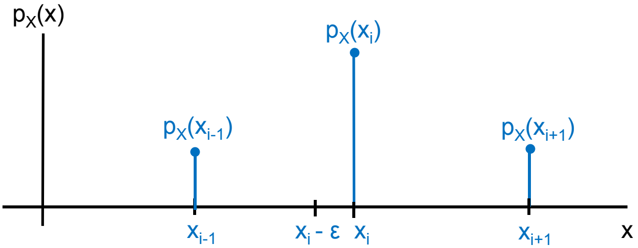
$\{X \leq x\} \cup \{x < X \leq x'\}$

disjoint

CDF Properties (cont.)

3. For $x_i \in S_X$ and an arbitrarily small positive number, $\epsilon > 0$

$$F_X(x_i) - F_X(x_i - \epsilon) = P[x_i - \epsilon < X \leq x_i]$$
$$= p_X(x_i) = P[X = x_i]$$



4. $F_X(x) = F_X(x_i)$ for all x such that $x_i \leq x < x_{i+1}$

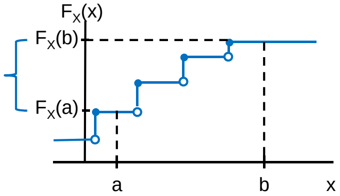
CDF Properties in Words

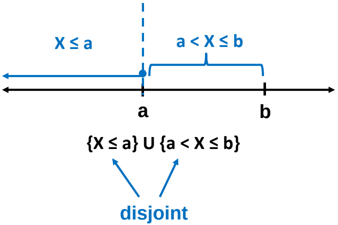
1. Going from left to right on the x-axis, $F_X(x)$ starts at 0 and ends at 1.
2. The CDF never decreases as it goes from left to right.
3. For a discrete random variable X, there is a jump (discontinuity) at each value of $x_i \in S_X$. The height of the jump at x_i is $p_X(x_i)$.
4. Between jumps, the graph of the CDF of the discrete random variable X is a horizontal line.

Another Important Consequence of Definition of CDF

The difference between the CDF evaluated at two points is the probability that the random variable takes on a value between these two points:

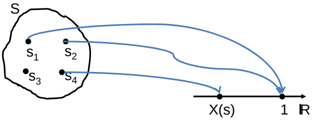
$$\text{For all } b \geq a, \qquad P[a < X \leq b] = F_X(b) - F_X(a)$$





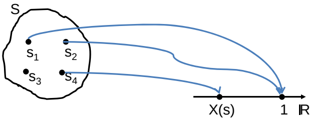
Discrete RVs

- sample space S
- an outcome $s \in S$
- an event $A \subset S$
- Probability measure assigns a number between [0,1] to each event
  - $P : A \mapsto P[A]$
  - satisfies three axioms
- random variable: X assigns a real number to each outcome
  - $X : S \mapsto \mathbb{R}$ (not necessarily a 1-1 function)
    $s \mapsto X(s)$



Discrete RVs

- discrete random variable: RV such that its range S_X is countable
- PMF: $p_X(x) = P[X = x] = P[\underbrace{\{s \in S | X(s) = x\}}_{\text{this is an event, i.e., a subset of } S}]$
  - e.g., $p_X(1) = P[X = 1] = P[\{s_1, s_2\}]$



  - geometric RV: X is such that $p_X(x) = (1 - p)^{x-1} p$, $x = 1, 2, 3, \ldots$
- CDF of a RV X:
  $F_X(x) = P[X \leq x] = P[\underbrace{\{s \in S | X(s) \leq x\}}_{\text{event}}]$

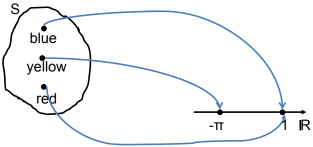
Discrete RVs: Example

Example:

- Box contains: 2 red balls, 1 yellow, 1 blue
- Experiment: select a ball out of the box at random
- $S = \{\text{red, yellow, blue}\}$
- $P[\{\text{red}\}] = \frac{1}{2}$ $\quad$ $P[\{\text{yellow}\}] = \frac{1}{4} = P[\{\text{blue}\}]$

  $P[\{\text{red,blue}\}] = \frac{3}{4}$

  X: red $\mapsto 1$
  blue $\mapsto 1$
  yellow $\mapsto -\pi$



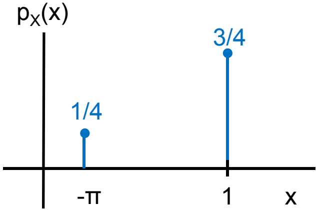
Discrete RVs: Example (cont.)

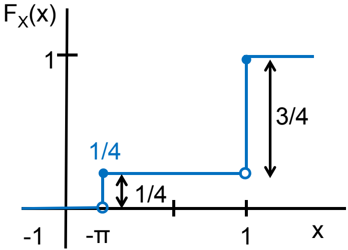
- What is PMF of X?

$$S_X = \{-\pi, 1\}$$

$$p_X(1) = P[X = 1] = P[\{s \in S | X(s) = 1\}] = P[\{\text{red,blue}\}] = \frac{3}{4}$$

$$p_X(-\pi) = P[X = -\pi] = P[\{\text{yellow}\}] = \frac{1}{4}$$





Averages: Example

- For one quiz, 10 students have the following grades (on a scale of 0 to 10):

  9,5,10,8,4,7,5,5,8,7

- Find the mean, the median, and the mode.

. . . .

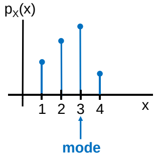
- mean$=68/10=6.8$
- median$=7$ (since there are four scores below 7 and four scores above 7)
- mode$=5$ (since that score occurs more often than any other)

Average and Expected Value

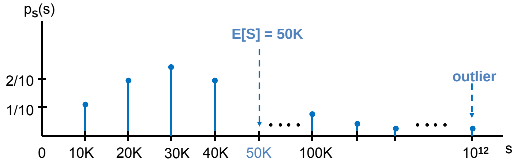
- Preceding comments on averages apply to sets of data collected by an experimenter.
- Corresponding to these averages are mathematical quantities that describe the random variables and their probability models.
  - Each average is a number that can be computed from the PMF or CDF of the RV.
- The most important of these is the **expectation**, or **expected value**, of an RV.
  - We will work with expectations throughout the course.

Mode and Median

- **Mode:** A mode, x_{mode} of RV X is a number that satisfies $p_X(x_{mode}) \geq p_X(x)$ for all x ("value most likely to occur").



- **Median:** A median, x_{median} of RV X is the number that satisfies $P[X < x_{median}] = P[X > x_{median}]$.



- E[S] = "can be large due to outlier"
- mode = 30K
- median ≈ 30K

Mode and Median (cont.)

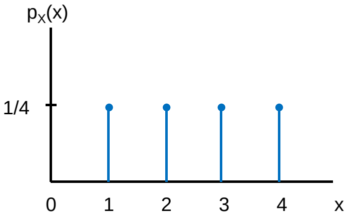
- If we read the definitions of mode and median carefully, we will observe that neither the mode nor the median of an RV need be unique.
    - A random variable can have several modes or medians.

Expected Value

- Corresponds to adding up a number of measurements and dividing by the number of terms in the sum
- Two notations: $E[X]$ and μ_x
- Synonyms: Expectation and mean value
- **Definition:(Expected Value)** The expected value of X is

$$E[X] = \mu_x = \sum_{x \in S_x} \underset{\text{possible value}}{x} \, \underbrace{p_X(x)}_{\text{likelihood or probability that that } x \text{ is taken}}$$

Expected Value: Example

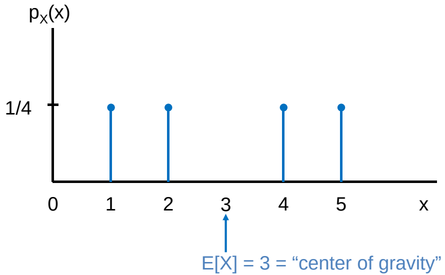


- Expected value of X is

$$E[X] = \sum_{x \in S_X} x p_X(x)$$
$$= 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 2.5$$

Expected Value: Center of Mass Interpretation

- The definition of expected value may look familiar from physics (mechanics).
- Think of point masses on a line with a mass of $p_X(x)$ kg at a distance of x meters from the origin.
- In this model, μ_x in the definition above is the **center of mass**.
  - This is why $p_X(x)$ is called probability **mass** function.



Expected Value & Sum of a Collection of Measurements

- To understand how this definition of expected value corresponds to the notion of adding up a set of measurements, suppose we have an experiment that produces an RV X and we perform n independent trials of this experiment.
- We denote the value that x takes on in the ith trial by $x(i)$.
  - We say that $x(1)$, ..., $x(n)$ is a set of n sample values of X.
  - Corresponding to the average of a set of numbers, we have after n trials of the experiment, the sample average

$$m_n = \frac{1}{n} \sum_{i=1}^{n} x(i)$$

  - Each $x(i)$ takes values from the set S_x. Out of the n trials, assume that each $x \in S_x$ occurs N_x times.
  - Then, the sum above becomes

$$m_n = \frac{1}{n} \sum_{x \in S_x} N_x x = \sum_{x \in S_x} \frac{N_x}{n} x$$

Expected Value and Relative Frequency (I)

- Recall our discussion of the "relative frequency" interpretation of probability.
- If in n observations of an experiment, the event A occurs N_A times, we can interpret the probability of A as:

$$P[A] = \lim_{n \to \infty} \frac{N_A}{n}$$

- This is the relative frequency of A. In the notation of RVs, we have the corresponding observation that

$$p_X(x) = \lim_{n \to \infty} \frac{N_x}{n}$$

Expected Value and Relative Frequency (II)

- This suggests that

$$\lim_{n \to \infty} m_n = \sum_{x \in S_x} x p_X(x) = E[X]$$

- The equation above says that the definition of $E[X]$ corresponds to a model of doing the same experiment repeatedly.
- After each trial, we add up all the observations up to date and divide by the number of trials.
- The result approaches the expected value as the number of trials increases without limit.

Expected Values of RVs

- We can use the definition of expected value to derive the expected value of each family of random variables.

- **Theorem:** The expected value of Bernoulli RV X is $E[X] = p$.
  - **Proof:** $E[X] = 0 \cdot p_X(0) + 1 \cdot p_X(1) = 0(1-p) + 1(p) = p$.

- **Theorem:** The expected value of geometric RV X is $E[X] = 1/p$.
  - **Proof:** $X \sim \text{geometric}(p)$, $p_X(x) = (1-p)^{x-1}p$, x= 1, 2, ...

$$E[X] = \sum_{x=1}^{\infty} x \cdot p_X(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p = p\left[\sum_{x=1}^{\infty} x(1-p)^{x-1}\right]$$

Let $q = 1-p$.

$$E[X] = p\left[\sum_{x=1}^{\infty} xq^{x-1}\right]$$

Expected Values of RVs (cont.)

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q}$$

Taking the partial derivative of both sides,

$$\frac{\partial}{\partial q}\left(\sum_{x=0}^{\infty} q^x\right) = \frac{\partial}{\partial q}\left(\frac{1}{1-q}\right)$$
$$\left(\sum_{x=0}^{\infty} x q^{x-1}\right) = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

Thus,

$$E[X] = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

If the probability of rejecting a circuit is $p = 1/5$, to observe the first reject, we have to conduct on average $E[Y] = 1/p = 5$ tests (if $p = 1/10$, we must conduct 10 tests).

Expected Values of RVs (cont.)

- **Theorem:** The expected value of Poisson RV X is $E[X] = \alpha$.
- **Theorem:** (a) For binomial RV X, $E[X] = np$.

  (b) For Pascal RV X, $E[X] = k/p$.

  (c) For discrete uniform RV X, $E[X] = (k + l)/2$.

Derived Random Variables

- **Idea:** An RV that is a function of another RV.
- **Definition: (Derived RV)** Each sample value y of a derived RV Y is a mathematical function $g(x)$ of a sample value x of another RV X. We adopt the notation $Y = g(X)$ to describe the relationship of the two RVs.

Example: Signal-noise ratio of a radio receiver, x,

we observe as the ratio of two strengths. We convert to decibel using $y = 10log_{10}x$.

$$X \longrightarrow \boxed{g(\cdot)} \longrightarrow Y = g(X)$$

- Suppose we know the PMF of X. What is the PMF of $Y = g(X)$?

$$p_Y(y) = P[Y = y] = P[\{x \in S_x \overset{\text{such that}}{\mid} g(x) = y\}]$$

$$= \sum_{x:g(x)=y} p_X(x)$$

Sum for all outcomes $X = x$ for which $Y = y$

PMF of Derived Random Variable: Example

Example:

$$X = \begin{cases} -1 & \text{, with probability } 1/3 \\ 0 & \text{, with probability } 1/3 \\ 1 & \text{, with probability } 1/3 \end{cases}$$

- p_X(-1)$= 1/3$
- Let g(x)$= x^2$ and Y$=$ g(X)$= X^2$.
- **Note:** $S_Y = \{0{,}1\}$.
- $p_Y(0) =$ P(Y$=$ 0)$=$ P(X$=$ 0)$= 1/3$
- $p_Y(1) =$ P(Y$=$ 1)$=$ P(X$=$ -1) $+$ P(X$=$ 1)$= 2/3$

Expected Value of a Derived Random Variable: Example

- Suppose the PMF of X is given. What is the expected value of $Y = g(X)$?
- **Theorem:** $E[Y] = \sum_{x \in S_x} g(x) p_X(x)$
- **Example:** Using the theorem,

$$E[Y] = g(-1) \cdot \frac{1}{3} + g(0) \cdot \frac{1}{3} + g(1) \cdot \frac{1}{3} = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$$

OR

$$E[Y] = 1 \cdot P[Y = 1] + 0 \cdot P[Y = 0] = 1 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3} = \frac{2}{3}$$

$$= \sum_{y \in S_y} y p_Y(y)$$

Derived Random Variables

$$X \longrightarrow \boxed{g(\cdot)} \longrightarrow Y = g(X)$$

Two types of problems:

1. Given $p_X(\mathrm{x})$ and $g()$, find $p_Y(y)$.
2. Given $p_X(\mathrm{x})$ and $g()$, find $E[Y]$.

Derived Random Variables: Problem Examples

$$X = \begin{cases} -1 & \text{, with probability } 1/6 \\ 0 & \text{, with probability } 1/3 \\ 1 & \text{, with probability } 1/2 \end{cases}$$

$g(x) = |x|$ (absolute value of x)

- Problem 1: $Y = |X|$, so $S_Y = \{0, 1\}$.

$$p_Y(0) = P[Y = 0] = P[|X| = 0] = P[X = 0] = p_X(0) = 1/3$$

$$\begin{aligned} p_Y(1) &= P[Y = 1] = P[\{X = -1\} \cup \{X = 1\}] \\ &= p_X(1) + p_X(-1) = 2/3 \end{aligned}$$

- Problem 2:

$$E[Y] = \sum_{y \in S_Y} y p_Y(y) \overset{\text{using result of Problem 1}}{\overset{\downarrow}{=}} \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$$

Alternatively,

$$\begin{aligned} E[Y] &= \sum_{x \in S_x} g(x) p_X(x) = g(-1) \cdot \frac{1}{6} + g(0) \cdot \frac{1}{3} + g(1) \cdot \frac{1}{2} \\ &= 1 \cdot \frac{1}{6} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{2} = \frac{2}{3} \end{aligned}$$

Derived Random Variable Examples: Expectation & Variance

- **Examples:**
  1. $Y = aX + b$, also given $p_X(x)$

$$
\begin{aligned}
E[Y] &= \sum_{x \in S_X} (ax + b) p_X(x) \\
&= \underbrace{a \sum_{x \in S_X} x p_X(x)}_{E[X]} + \underbrace{b \sum_{x \in S_X} p_X(x)}_{1} \\
&= aE[X] + b
\end{aligned}
$$

  2. Variance
     $Y = (X - \mu_x)^2$, **Note:** $\mu_x = E[X]$

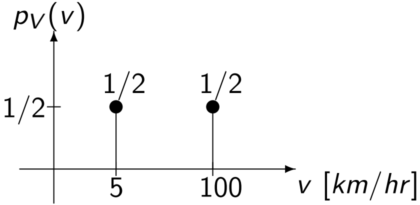
$$
E[Y] = \sum_{x \in S_x} (x - \mu_x)^2 p_X(x) = Var[X] = {\sigma_x}^2
$$

Expectation: Classic Mistake

- Caution: In general, $E[g(X)] \neq g(E[X])$.
- Example: It is NOT true, in general, that $E[X^2] = (E[X])^2$, since this means that $Var[X] = 0$.

Expectation: Classic Mistake Example

- Let us drive a distance of 200 km, at a constant but random speed V.



Expectation: Classic Mistake Example

- What is the expected time $E[T]$ to get from A to B?

$$T = \frac{200}{V}$$

$$E[T] = \sum_{v \in S_V} \frac{200}{v} p_V(v) = \frac{200}{5} \cdot \frac{1}{2} + \frac{200}{100} \cdot \frac{1}{2} = \underline{\underline{21 \text{ hours}}}$$

Mistake:

$$\cancel{\begin{aligned} E[V] &= \frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 100 = 52.5 \text{ km/hr} \\ \implies E[T] &\neq \frac{200}{E[V]} = \frac{200}{52.5} = 3.8 \text{ hours} \end{aligned}}$$

Variance and Standard Deviation

- Variance measures the spread of an RV.
- Variance of RV X describes the difference between the random variable X and its expected value.
- **Definition:(Variance)** Variance of RV X is
$$Var[X] = E[(X - \mu_X)^2]$$
- **Important note:** Variance which is the expected value of the sum of squares can never be a negative number!
- **Definition:(Standard Deviation)** Standard deviation of RV X is
$$\sigma_x = \sqrt{Var[X]}$$
- Units of σ_x same as those of X.

Variance

- **Theorem:** The variance of RV X is given by
$$Var[X] = E[X^2] - {\mu_X}^2 = E[X^2] - (E[X])^2$$
  - **Proof:** The variance of RV X is a derived RV $Y = (X - \mu_X)^2$.

$$\begin{aligned}
Var[X] &= E[Y] = E[(X - \mu_X)^2] \\
&= E[X^2 - 2X\mu_X + {\mu_X}^2] \\
&= \sum_{x \in S_x} (x^2 - 2x\mu_X + {\mu_X}^2) p_X(x) \\
&= \sum_{x \in S_x} x^2 p_X(x) - \sum_{x \in S_x} 2x\mu_X p_X(x) + \sum_{x \in S_x} {\mu_X}^2 p_X(x) \\
&= E[X^2] - 2\mu_X E[X] + {\mu_X}^2 \\
&= E[X^2] - 2{\mu_X}^2 + {\mu_X}^2 \\
&= E[X^2] - {\mu_X}^2
\end{aligned}$$

Properties of Variance (I)

1. If $Y = X + b$, then $Var[Y] = Var[X]$. (A shift does not change the variance)

Proof: $E[Y] = E[X] + b$

$$
\begin{aligned}
Var[Y] &= \sum_{y} (y - E[X] - b)^2 p_Y(y) \\
&= \sum_{x \in S_x} (x + b - E[X] - b)^2 p_X(x) \\
&= \sum_{x \in S_x} (x - E[X])^2 p_X(x) \\
&= Var[X]
\end{aligned}
$$

Properties of Variance (II)

2. If $Y = aX$, then $Var[Y] = a^2 Var[X]$. (A scaling changes the variance by the square of the scaling)

   Proof: $E[Y] = aE[X] = a\mu_x$

$$
\begin{aligned}
Var[Y] &= E[(Y - \mu_Y)^2] \\
&= E[(aX - a\mu_X)^2] \\
&= E[a^2(X - \mu_X)^2] \\
&= a^2 E[(X - \mu_X)^2] \\
&= a^2 Var[X]
\end{aligned}
$$

3. If $Y = aX + b$, then $Var[Y] = a^2 Var[X]$.

Variance: Example

- Let X be the outcome of the roll of a die.
- Find its mean and variance.

. . . .

- We can use the definitions of expectation and variance.
- $E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 21/6 = 3.5$
- $E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$
- $Var[X] = E[X^2] - (E[X])^2 = \frac{91}{6} - (\frac{21}{6})^2 = \frac{105}{36} = \frac{35}{12} \approx 2.9$

Variance: Example (cont.)

- What if $Y = 2X - 6$?
- How does Y look? (PMF)
- What is its mean and the variance?

. . . .

- $E[Y] = 2E[X] - 6 = 2 \cdot 3.5 - 6 = 1$
- $Var[X] = 2^2 Var[X] = 4 \cdot \frac{35}{12} = \frac{35}{3} \approx 11.66$

Variances of Various Random Variables

- Bernoulli: $Var[X] = p(1-p)$
- Geometric: $Var[X] = (1-p)/p^2$
- Binomial: $Var[X] = np(1-p)$
- Pascal: $Var[X] = k(1-p)/p^2$
- Poisson: $Var[X] = \alpha$
- Discrete Uniform: $Var[X] = (l-k)(l-k+2)/12$