CS145: Probability & Computing

Lecture 4-5: Discrete Random Variables, Expected Values

Instructors: Eli Upfal and Alessio Mazzetto

Brown University Computer Science

Figure credits:

*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*

*Pitman, **Probability**, 1999*

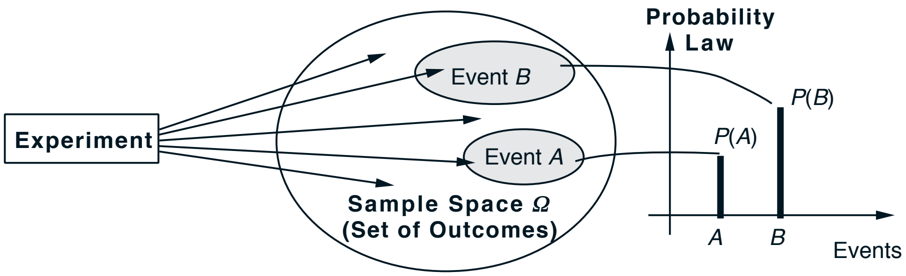
CS145: Lecture 5 Outline

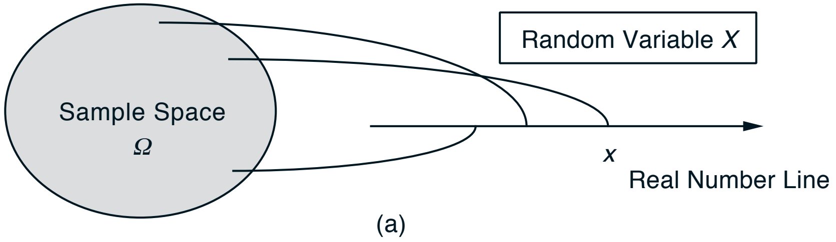
- Discrete random variables
- Expectations of discrete variables

Discrete Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$



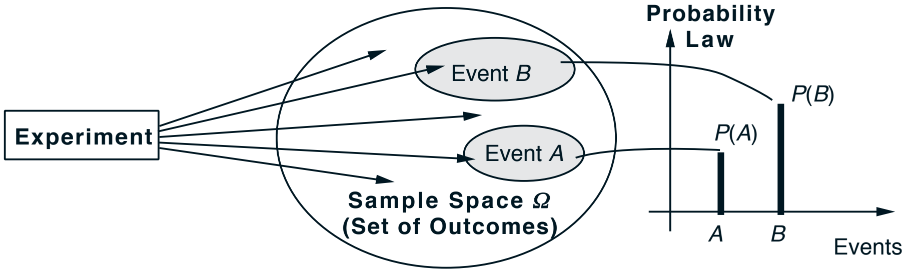


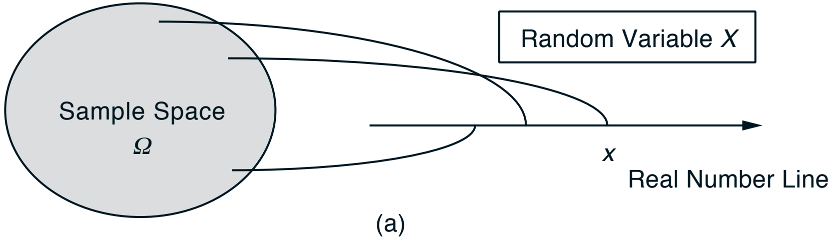
- Mathematically: A function from sample space Ω to real numbers $\mathbb{R}$
- May define several random variables on the same sample space, if there are several quantities you would like to measure
- Example:
  - Sample space: students at Brown.
  - Random variables: grade in CS 145, grade in CS 15, age,…

Discrete Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$





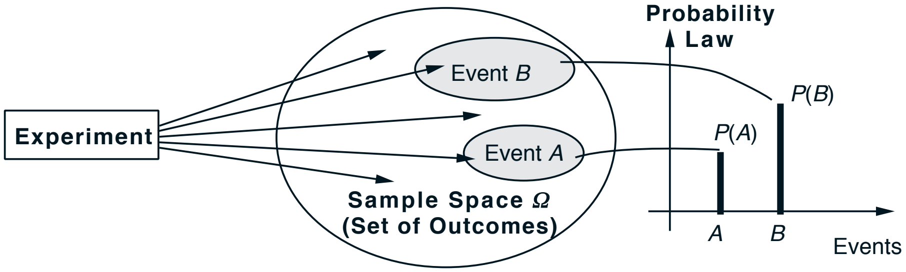
Example random variables for a day in a casino:

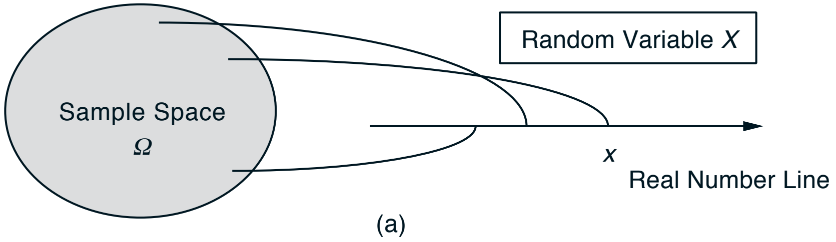
- Number of gamblers who visited
- Total money won (or probably, lost)
- Number of hands of poker played
- Total power consumed
- Number of gamblers caught cheating

Discrete Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$





- The *range* of a random variable is the set of values with positive probability

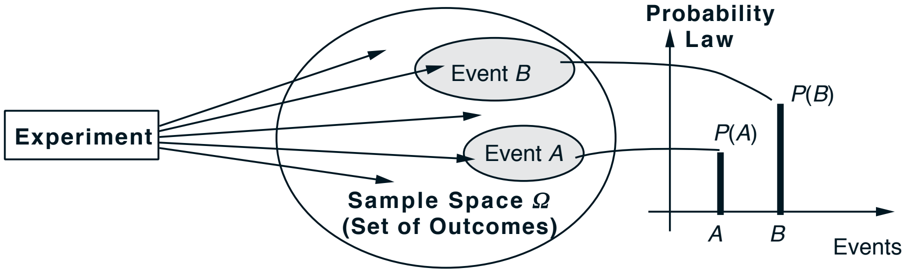
$$\mathcal{X} = \{x \in \mathbb{R} \mid X(\omega) = x \text{ for some } \omega \in \Omega, P(\omega) > 0\}$$

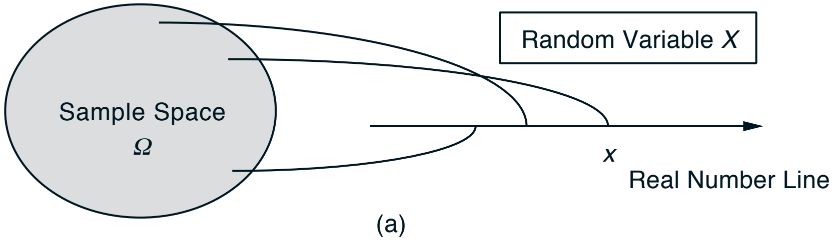
For a *discrete random variable*, the range is finite or countably infinite (we can map it to the integers). *Coming later: continuous random variables.*

Discrete Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$





- The *probability mass function (PMF)* or *probability distribution* of variable:

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

$$p_X(x) \geq 0, \qquad \sum_{x \in \mathcal{X}} p_X(x) = 1.$$

If range is finite, this is a vector of non-negative numbers that sums to one.

Computing a PMF

- Notation:

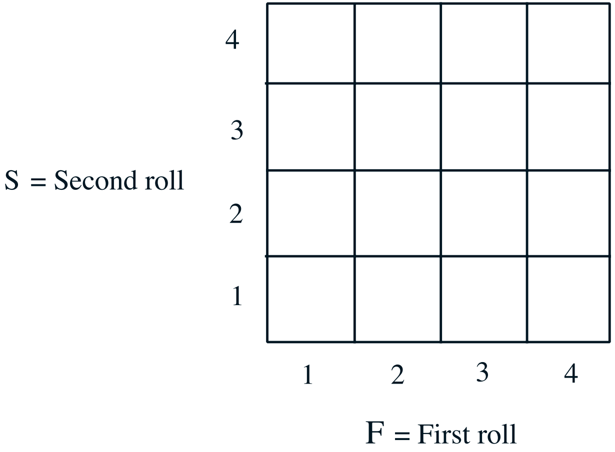
$$p_X(x) = \mathbf{P}(X = x)$$
$$= \mathbf{P}(\{\omega \in \Omega \text{ s.t. } X(\omega) = x\})$$

– collect all possible outcomes for which X is equal to x
– add their probabilities
– repeat for all x

- $p_X(x) \geq 0 \qquad \sum_x p_X(x) = 1$

- **Example:** Two independent rools of a fair tetrahedral die

F: outcome of first throw
S: outcome of second throw
$X = \min(F, S)$



Computing a PMF

Sample space $= S = \{s_1, s_2, \ldots, s_m\}$, a group of students at CIT.
Distribution: $P(s) =$ probability that I choose student s.
Random variables:

- $X = 1, 2, 3, 0$ - grade in CS 145 of the student I chose
- $Y = 1, 2, 3, 0$ - grade in CS 155 of the student I chose
- $Z = 1, 2, 3, 4$ - years in Brown of the student I chose

$P(X = 2) = \sum_{s:X(s)=2} P(s)$

$P(2 \leq Z \leq 4) = \sum_{s:2\leq Z(s)\leq 4} P(z)$

Geometric Probabilities

- Repeatedly flip a coin with probability of Heads *p*, count the number of tosses *X* until the first Head is observed:

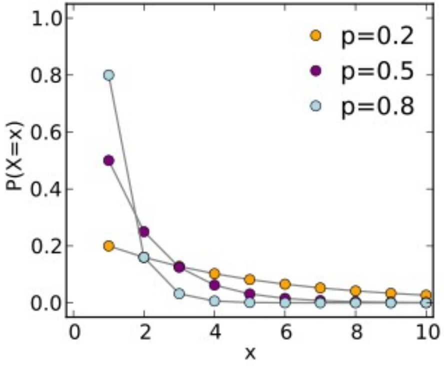
$$P(X=1)=p, P(X=2)=(1-p)p, P(X=3)=(1-p)^2p, \dots$$

$$P(X=k)=(1-p)^{k-1}p \text{ for } k=1,2,3,\dots$$

- The number of possible outcomes is *infinite*: there is no *k* after which the next toss is guaranteed to be Heads

Example:

- Your laptop hard drive independently fails on each day with (hopefully small) probability *p*. What is the distribution of the number of days until failure?



Wikipedia

Geometric Probabilities

- Repeatedly flip a coin with probability of Heads *p*, count the number of tosses *X* until the first Head is observed:

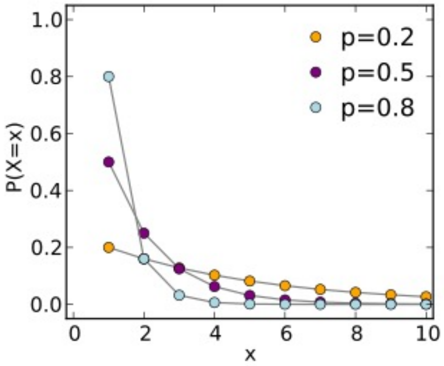
$$P(X=1)=p, P(X=2)=(1-p)p, P(X=3)=(1-p)^2p, \ldots$$

$$P(X=k)=(1-p)^{k-1}p \text{ for } k=1,2,3,\ldots$$

- Recall the *geometric series*: $$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, 0<q<1.$$

- Verify that geometric probabilities are *normalized*:

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=0}^{\infty}(1-p)^k = \frac{p}{1-(1-p)} = 1$$



Wikipedia

Geometric Probabilities

- Repeatedly flip a coin with probability of Heads *p*, count the number of tosses *X* until the first Head is observed:

$$P(X=1)=p, P(X=2)=(1-p)p, P(X=3)=(1-p)^2 p, \ldots$$

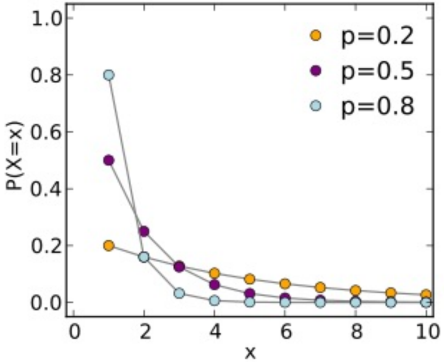
$$P(X=k)=(1-p)^{k-1}p \text{ for } k=1,2,3,\ldots$$

- What is the probability that the number of tosses *X* is odd?

$$P(X \text{ odd}) = \sum_{k=1}^{\infty}(1-p)^{2(k-1)}p = \frac{1}{2-p}$$

- For a fair coin, this equals

$$P(X \text{ odd}) = \frac{2}{3} \text{ if } p = \frac{1}{2}$$

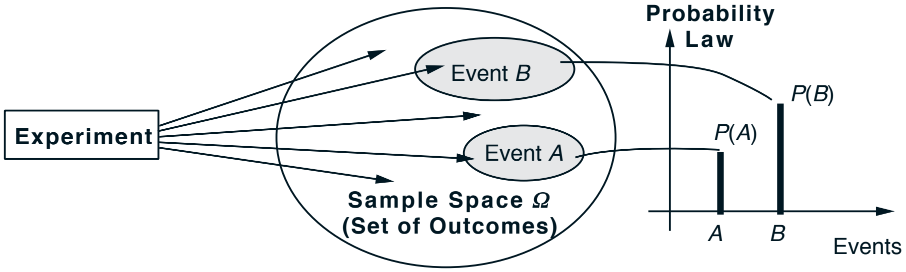


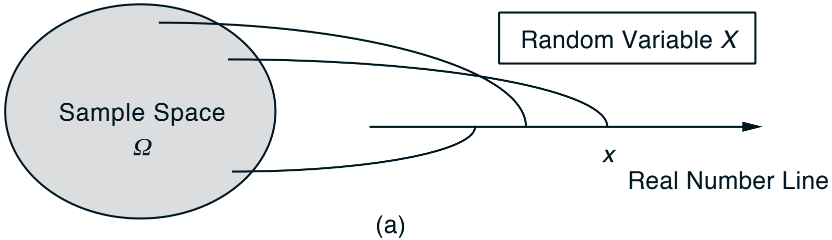
Wikipedia

Discrete Random Variables

- Computing *probabilities of sets of values*:

$$P(X \in S) = \sum_{x \in S} p_X(x) \text{ for any } S \subset \mathbb{R}.$$





- The *probability mass function* or *probability distribution* of random variable:

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

$$p_X(x) \geq 0, \qquad \sum_{x \in \mathcal{X}} p_X(x) = 1.$$

If range is finite, this is a vector of non-negative numbers that sums to one.

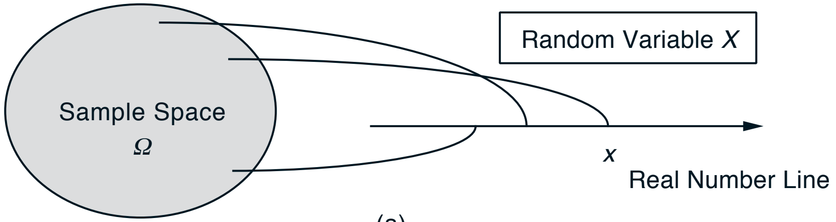
Functions of Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$

$$p_X(x) = P(X = x)$$

$$p_X(x) \geq 0, \qquad \sum_{x \in \mathcal{X}} p_X(x) = 1.$$



(a)

- If we take any *non-random (deterministic) function* of a random variable, we produce another random variable: $Y = g(X)$

$$g : \mathbb{R} \to \mathbb{R}$$
$$g \circ X : \Omega \to \mathbb{R}$$

- *Example:* Degrees Celsius *X* to degrees Fahrenheit *Y*: $Y = 1.8X + 32$
- *Example:* Current drawn *X* to power consumed *Y*: $Y = rX^2$

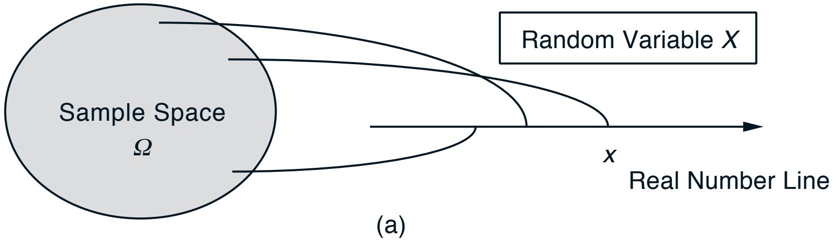
Functions of Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$

$$p_X(x) = P(X = x)$$

$$p_X(x) \geq 0, \qquad \sum_{x \in \mathcal{X}} p_X(x) = 1.$$



- If we take any *non-random (deterministic) function* of a random variable, we produce another random variable: $Y = g(X)$

$$g : \mathbb{R} \to \mathbb{R}$$
$$g \circ X : \Omega \to \mathbb{R}$$

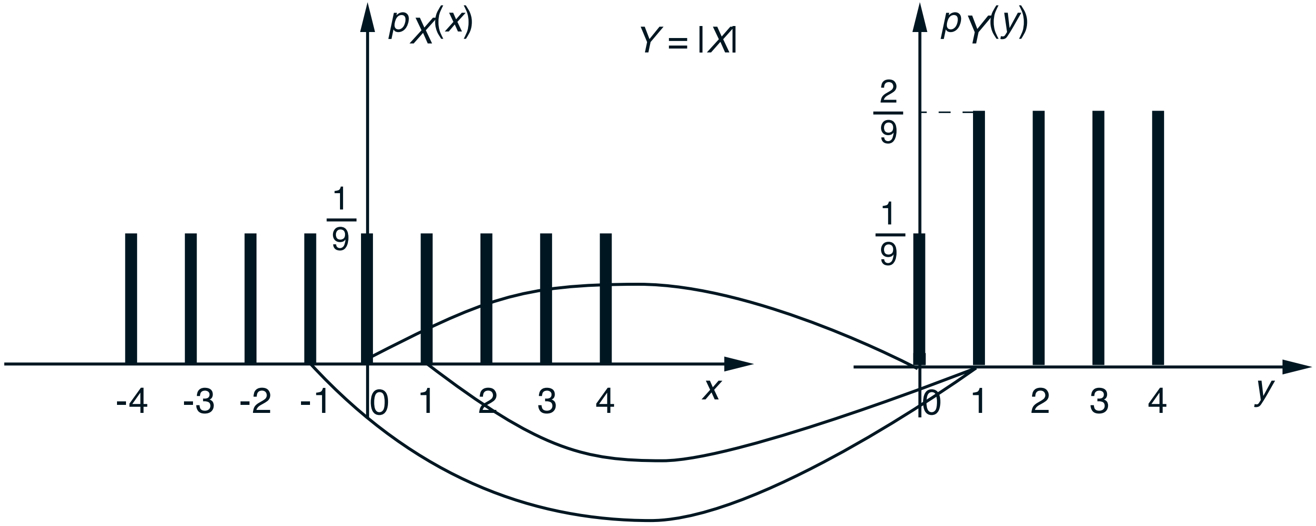
- By definition, the *probability mass function* of *Y* equals

$$p_Y(y) = \sum_{\{x | g(x) = y\}} p_X(x) \qquad p_Y(y) \geq 0, \quad \sum_{y \in \mathcal{Y}} p_Y(y) = 1.$$

Example: Absolute Value

$$g(X) = |X|$$

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



$$p_X(x) = \begin{cases} 1/9 & \text{if } x \text{ is an integer in the range } [-4, 4], \\ 0 & \text{otherwise.} \end{cases}$$

$$p_Y(y) = \begin{cases} 2/9 & \text{if } y = 1, 2, 3, 4, \\ 1/9 & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

CS145: Lecture 5 Outline

- Discrete random variables
- Expectations of discrete variables

Expectation

- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

- The expectation is a single number, not a random variable.
  It encodes the "center of mass" of the probability distribution:

- The random variable has an expectation iff $E[|X|] < \infty$

- We may also use the terms mean, average, first moment

- Median is a different concept. It's the value M such that

$$P(X \leq M) \geq 1/2 \quad \text{and} \quad P(X \geq M) \geq 1/2$$

Bernoulli Probability Distribution

- A *Bernoulli* or *indicator* random variable *X* has one parameter *p*:

$$p_X(1) = p, \qquad p_X(0) = 1 - p, \qquad \mathcal{X} = \{0, 1\}$$

The expectation of an indicator random variable *is its probability:*

$$E[X] = \sum_{x \in \mathcal{X}} x \cdot P(X = x) = 1 \cdot p + 0 \cdot (1 - p) = p$$

Jakob Bernoulli

Examples:

- Flip a possibly biased coin with probability of coming up heads *p*
- A user answers a true/false question in an online survey
- Does it snow or not on some day

Expectation – First Moment

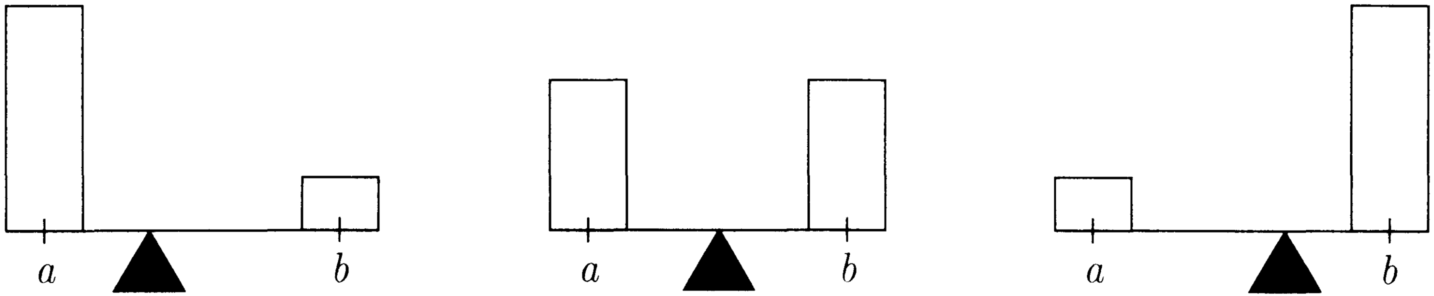
- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

- The expectation is a single number, not a random variable.
  It encodes the "center of mass" of the probability distribution:

If X takes two possible values, say a and b, with probabilities $P(a)$ and $P(b)$, then

$$E(X) = aP(a) + bP(b) \qquad P(a) + P(b) = 1$$

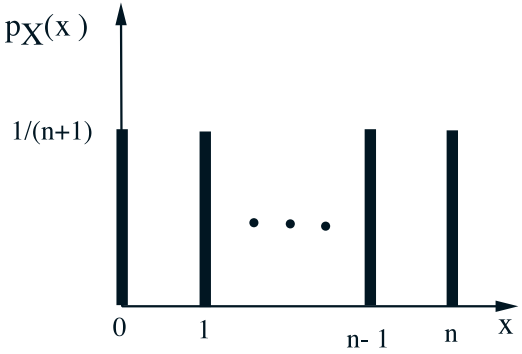


Expectation

- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

- The expectation is a number, not a random variable.
  It encodes the "center of mass" of the probability distribution
- Example: Uniform distribution on *{0,1,…,n}*



$$\mathbf{E}[X] = 0 \times \frac{1}{n+1} + 1 \times \frac{1}{n+1} + \cdots + n \times \frac{1}{n+1} = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}$$

Expectation

- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

- Example: Uniform distribution on *{0,1,…,n}*, $P(X=i) = \frac{1}{n+1}$, $E[X] = \frac{n}{2}$

- *Example:* $P(X = i) = \begin{cases} \frac{i}{\frac{1}{2}n(n+1)} & \text{for } 0 \leq i \leq n \\ 0 & \text{otherwise} \end{cases}$

$$\text{Proper distribution:} \quad \sum_{i=0}^{n} P(X = i) = \sum_{i=0}^{n} \frac{i}{\frac{1}{2}n(n+1)} = 1$$

$$E[X] = \sum_{i=0}^{n} i \frac{i}{\frac{1}{2}n(n+1)} = \frac{\frac{1}{6}n(n+1)(2n+1)}{\frac{1}{2}n(n+1)} = \frac{2}{3}n + \frac{1}{3}$$

Expectation

- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

$$x_{\min} \leq E[x] \leq x_{\max}$$

$$x_{\min} = \min\{x \mid x \in \mathcal{X}\}$$

$$x_{\max} = \max\{x \mid x \in \mathcal{X}\}$$

- The expectation is an average or interpolation. It is possible that

$$p_X(E[x]) = 0 \text{ for some random variables } X.$$

Example: $p_X(1) = p,$ $p_X(0) = 1 - p,$ $\mathcal{X} = \{0, 1\}$ $E[X] = p$

Geometric Probability Distribution

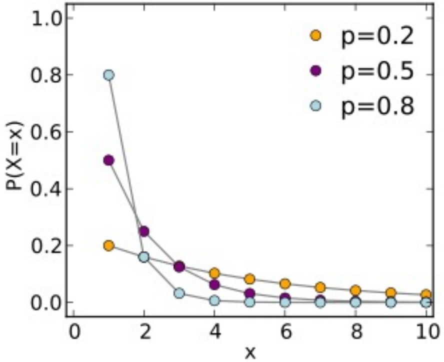
- Recall the geometric series: $$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}, 0 < q < 1.$$

- A *geometric* random variable X has parameter p, countably infinite range:

$$p_X(k) = (1-p)^{k-1}p \qquad \mathcal{X} = \{1, 2, 3, \ldots\}$$

Examples:

- Flip a coin with bias p, count number of tosses until first heads (success)
- Your laptop hard drive independently fails on each day with (hopefully small) probability p. What is the distribution of the number of days until failure?



Wikipedia

Geometric Probability Distribution

- A *geometric* random variable X has parameter p, countably infinite range:

$$p_X(k) = (1-p)^{k-1} p \qquad \mathcal{X} = \{1, 2, 3, \ldots\}$$

- The expected value equals:

$$E[X] = \sum_{k \geq 1} k(1-p)^{k-1} p = \sum_{k \geq 1} (1-p)^{k-1} = \frac{1}{p}$$

- In general, for any non-negative random variable X:

$$E[X] = \sum_{k \geq 1} k P(X = k) = \sum_{k \geq 1} \sum_{j \geq k} P(X = j) = \sum_{k \geq 1} P(X \geq k)$$

Expectation of Non-Negative R. V.

$$\begin{array}{rcccccccccccc}
E[X] & = & P(X=1) & + & 2P(X=2) & + & 3P(X=3) & + & 4P(X=4) & + & 5P(X=5) & + & \dots \\
 & = & P(X=1) & + & P(X=2) & + & P(X=3) & + & P(X=4) & + & P(X=5) & + & \dots \\
 & & & + & P(X=2) & + & P(X=3) & + & P(X=4) & + & P(X=5) & + & \dots \\
 & & & & & + & P(X=3) & + & P(X=4) & + & P(X=5) & + & \dots \\
 & & & & & & & + & P(X=4) & + & P(X=5) & + & \dots \\
 & & & & & & & & & + & P(X=5) & + & \dots \\
 & & \vdots & & & \vdots & & & & \vdots & & \vdots & \dots \\
 & = & P(X \geq 1) & + & P(X \geq 2) & + & P(X \geq 3) & + & P(X \geq 4) & + & P(X \geq 5) & + & \dots \\
 & = & \sum_{k \geq 1} P(X \geq k) & & & & & & & & & &
\end{array}$$

$$E[X] = \sum_{k \geq 1} kP(X = k) = \sum_{k \geq 1} \sum_{j \geq k} P(X = j) = \sum_{k \geq 1} P(X \geq k)$$

Expected Values of Functions

- Consider a *non-random (deterministic) function* of a random variable:

$$Y = g(X)$$

$$p_X(x) = P(X = x) \quad \Rightarrow \quad p_Y(y) = \sum_{\{x | g(x) = y\}} p_X(x)$$

- What is the expected value of random variable *Y*? $E[Y] = E[g(X)]$

- Correct approach #1: $E[Y] = \sum_y y p_Y(y)$

- Correct approach #2: $E[Y] = E[g(X)] = \sum_x g(x) p_X(x)$

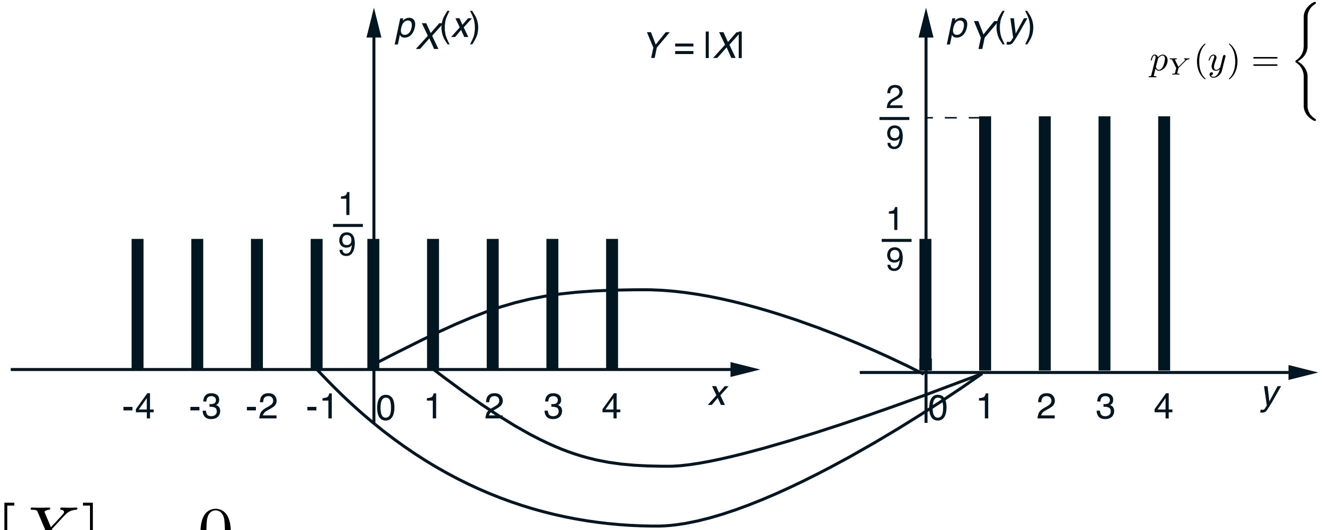
- Incorrect approach: $g(E[X]) \neq E[g(X)]$ *(except in special cases)*

Example: Absolute Value

$$g(X) = |X|$$

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

$$p_Y(y) = \begin{cases} 2/9 & \text{if } y = 1, 2, 3, 4, \\ 1/9 & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$



$$E[X] = 0$$

$$g(E[X]) = g(0) = 0$$

$$E[Y] = \frac{1}{9}(0) + \frac{2}{9}(1+2+3+4) = \frac{20}{9} \approx 2.22$$

Linearity of Expectation

- Consider a *linear function*: $Y = g(X) = aX + b$
- *Example:* Change of units (temperature, length, mass, currency, …)
- In this special case, mean of *Y* is the linear function applied to *E[X]*:

$$E[Y] = g(E[X]) = aE[X] + b$$

$$\mathbf{E}[Y] = \sum_x (ax+b)p_X(x) = a\sum_x xp_X(x) + b\sum_x p_X(x) = a\mathbf{E}[X] + b.$$

Example: You went on vacation to Europe, and want to find the average amount you spent on lodging per day. The following are equivalent (assuming a fixed exchange rate from Euros to US dollars):

- *E[g(X)]* = convert each receipt from Euros to US dollars, average result
- *g(E[X])* = average receipts in Euros, convert result to US dollars

Linearity of Expectation

- Consider a *linear function*: $Y = g(X) = aX + b$
- *Example:* Change of units (temperature, length, mass, currency, …)
- In this special case, mean of *Y* is the linear function applied to *E[X]*:

$$E[Y] = g(E[X]) = aE[X] + b$$

Example: I offer you to let you play a game where you pay a $20 entrance fee, and then I let you roll a fair 6-sided die, and pay you the rolled value times $5. What is your expected change in money?

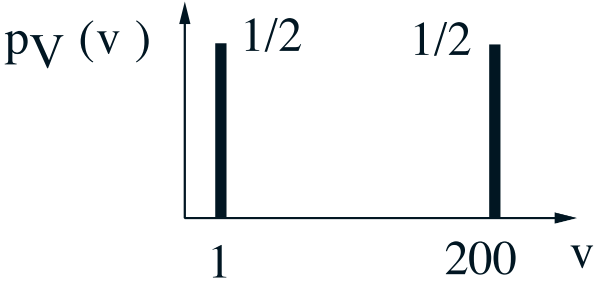
$$Y = 5X - 20$$ *(change in money Y for dice outcome X)*

$$E[X] = 3.5$$

$$E[Y] = 5E[X] - 20 = -2.5$$

Travel at a Random Speed

- You want to travel 200 miles to New York
- With 50% probability, the new high-speed train runs at a constant velocity of 200 mph
- With 50% probability, the train engine overheats and it runs at a constant velocity of 1 mph



$$E[V] = \frac{201}{2} = 100.5$$

- time in hours $= T = t(V) = \dfrac{200}{V}$
- $\mathbf{E}[T] = \mathbf{E}[t(V)] = \sum_v t(v) p_V(v) =$
- $\mathbf{E}[TV] = 200 \neq \mathbf{E}[T] \cdot \mathbf{E}[V]$
- $\mathbf{E}[200/V] = \mathbf{E}[T] \neq 200/\mathbf{E}[V]$.

$$E[T] = 1 * 1/2 + 200 * 1/2 = 100.5$$

Expectation of Multiple Variables

- The *expectation* or *expected value* of a function of two discrete variables:

$$E[g(X,Y)] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} g(x,y) p_{XY}(x,y)$$

- A similar formula applies to functions of 3 or more variables

- Expectations of sums of functions are sums of expectations:

$$E[g(X) + h(Y)] = E[g(X)] + E[h(Y)] = \left[\sum_{x \in \mathcal{X}} g(x) p_X(x) \right] + \left[\sum_{y \in \mathcal{Y}} h(y) p_Y(y) \right]$$

- This is always true, *whether or not X and Y are independent*
- Specializing to *linear functions*, this implies that:

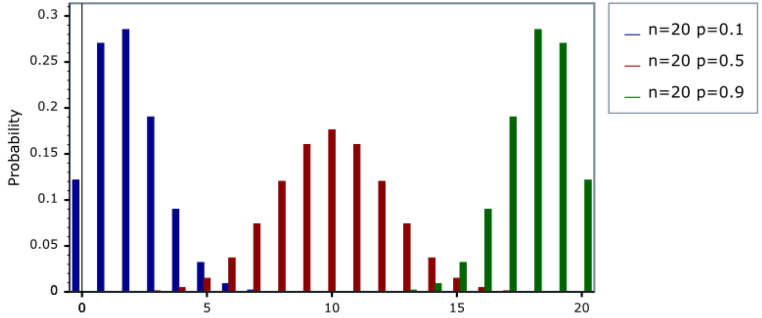
$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

Mean of Binomial Probability Distribution

- Suppose you flip *n* coins with bias *p*, count number of heads
- A *binomial* random variable *X* has parameters *n, p*:

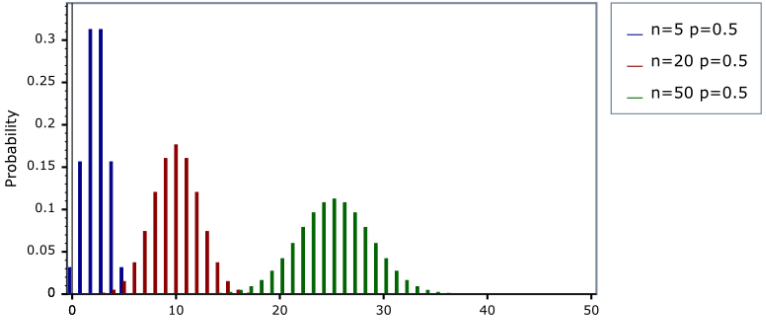
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathcal{X} = \{0, 1, 2, \ldots, n\}$$



- For binomial, *expected values are expected counts of events:*

$$E[X] = pn$$



- Simple proof uses indicator variables X_i for whether each of n tosses is heads:

$$\mathbf{E}[X_i] = p \cdot 1 + (1-p) \cdot 0 = p = \Pr(X_i = 1).$$

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np.$$

Binomial Mean: The Hard Way

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np.
\end{aligned}
$$