Review of random variables

- A random variable \( X \) associates a number with each outcome of an experiment.
- A random variable is a fixed map from the sample space to the real line.
- Random because we do not know exactly which outcome of the experiment will be observed on the next trial, and thus which value the random variable will have.
- Observed value of \( X \) varies at random.

Discrete random variables

- A discrete random variable takes on a finite number or a countably infinite number of discrete values.
- The values taken on by a discrete random variable are discretely spaced.
- If \( u_1, u_2, \ldots \) are the values taken on by a discrete random variable, then for each choice of \( j \), \( u_j < u_{j+1} \).

Probability mass functions (pmfs)

- The probabilistic behavior of a discrete random variable \( X \) is described by its probability mass function \( p_X(u) \) or \( p(u) \).
- \( p_X(u) = p(u) = 0 \) unless \( u \) is one of the values \( u_j \) that \( X \) takes on.
- \( p_X(u_j) = p(u_j) = P(X = u_j) \).
- \( p(u) \geq 0 \) for all \( u \); \( \sum_j p(u_j) = 1 \).

Everything you always wanted to know about \( X \)

- All the probabilistic information about the discrete random variable \( X \) is summarized in its pmf.
- The pmf can be used to answer questions such as:
  - What is the probability that \( X \) has value between \( a \) and \( b \)?
  - What is the probability that \( X \) is an even number?

... but were afraid to ask!

- Example: \( X \) is a random variable taking on integer values 0 through 8.
  - \( p_X(0) = p_X(1) = p_X(7) = p_X(8) = 0.05 \)
  - \( p_X(2) = p_X(3) = p_X(5) = 0.15; p_X(4) = 0.2 \)
  - \( P(3 < X \leq 6) = p_X(4) + p_X(5) = 0.35 \)
  - \( P(3 \leq X < 6) = p_X(3) + p_X(4) + p_X(5) = 0.5 \)
  - \( P(X \text{ is odd}) = p_X(1) + p_X(3) + p_X(5) + p_X(7) = 0.4 \)
  - \( P(X = 3.13) = 0 \) because \( p_X(3.13) = 0 \).

I’m leaving on a jet plane …

- Example (continued): Suppose that \( X \) is the number of passengers (with confirmed reservations) who show up for a flight on a 5-passenger plane. Let \( Y \) denote the number of passengers who board the flight.
  - Then, \( Y = X \) if \( X \leq 5 \) and \( Y = 5 \) if \( X > 5 \).
  - The random variable \( Y \) is said to be a function of the random variable \( X \).
  - The pmf of \( Y \) can be found from \( p_X(u) \).
Repeated trials of the experiment

- Let $X$ denote a discrete random variable with pmf $p_X(u)$.
- $X$ can take on values in $\{u_1, u_2, \ldots, u_n, \ldots\}$.
- The experiment was repeated $N$ times.
- On these $N$ trials, $X$ was observed to have taken on values $x_1, x_2, \ldots, x_N$, respectively.
- Each $x_i$ is some number in the set $\{u_1, u_2, \ldots, u_n, \ldots\}$. It is possible that the same value is observed more than once.

Average of the $N$ observed values

- On $N$ trials, $X$ was observed to have taken on values $x_1, x_2, \ldots, x_N$, respectively.
- The average of these $N$ numbers is just $\frac{x_1 + x_2 + \ldots + x_N}{N}$.
- If we were to do another set of $N$ trials, the values taken on by $X$ on these trials would not be the same as $x_1, x_2, \ldots, x_N$.
- But, the average would be fairly close!

Repeated additions = multiplication!

- On $N$ trials, $X$ was observed to have taken on values $x_1, x_2, \ldots, x_N$, respectively.
- The average of these $N$ numbers is just $\frac{x_1 + x_2 + \ldots + x_N}{N}$.
- Some of the $x_i$ happen to be $u_1$, others are $u_2$, still others are $u_3$, and so on.
- If $a_i$ of the $x_i$ happen to have value $u_1$, they will contribute a total of $a_1u_1$ to the sum in the numerator.

Catching a few $Z$'s

- Example (continued): For reasons of aircraft stability, control, and balance, the pilot insists that the plane must have an **odd** number of passengers. Let $Z$ denote the number of passengers left behind (to sleep in the terminal while waiting for the next flight?).
- $Z = 0$ if $X = 0, 1, 3, 5$; $Z = 1$ if $X = 2, 4, 6$; $Z = 2$ if $X = 7$; $Z = 3$ if $X = 8$.
- Sanity check: Is $\sum_p p_Z(u) = 1$? Yes.

The pmf of $Z$

- Example (continued):
  - $p_X(0) = p_X(1) = p_X(7) = p_X(8) = 0.05$
  - $p_X(2) = p_X(3) = p_X(5) = p_X(6) = 0.15$; $p_X(4) = 0.2$
- $p_Y(0) = p_Y(1) = p_Y(5) = 0.05$; $p_Y(2) = 0.15$; $p_Y(3) = 0.15$; $p_Y(4) = 0.2$
- $p_Y(5) = p_Y(6) + p_Y(7) + p_Y(8) = 0.4$
- Sanity check: Is $\sum_p p_Y(u) = 1$? Yes.

Repeated trials of the experiment

- Let $X$ denote a discrete random variable with pmf $p_X(u)$.
- $X$ can take on values in $\{u_1, u_2, \ldots, u_n, \ldots\}$.
- The experiment was repeated $N$ times.
- On these $N$ trials, $X$ was observed to have taken on values $x_1, x_2, \ldots, x_N$, respectively.
- Each $x_i$ is some number in the set $\{u_1, u_2, \ldots, u_n, \ldots\}$. It is possible that the same value is observed more than once.

Average of the $N$ observed values

- On $N$ trials, $X$ was observed to have taken on values $x_1, x_2, \ldots, x_N$, respectively.
- The average of these $N$ numbers is just $\frac{x_1 + x_2 + \ldots + x_N}{N}$.
- If we were to do another set of $N$ trials, the values taken on by $X$ on these trials would not be the same as $x_1, x_2, \ldots, x_N$.
- But, the average would be fairly close!

Repeated additions = multiplication!

- On $N$ trials, $X$ was observed to have taken on values $x_1, x_2, \ldots, x_N$, respectively.
- The average of these $N$ numbers is just $\frac{x_1 + x_2 + \ldots + x_N}{N}$.
- Some of the $x_i$ happen to be $u_1$, others are $u_2$, still others are $u_3$, and so on.
- If $a_i$ of the $x_i$ happen to have value $u_1$, they will contribute a total of $a_1u_1$ to the sum in the numerator.

Sanity check: Is $\sum p_Y(u) = 1$? Yes.
7.3 The sum of identical numbers

- On N trials, X was observed to have taken on values x_1, x_2, ..., x_N respectively
- The average of these N numbers is just x_1 + x_2 + ... + x_N
- The average can also be expressed as
  \[ \frac{a_1u_1 + a_2u_2 + \ldots + a_Nu_N}{N} \]
  where for each k, u_k occurred a_k times among the N numbers being added.

7.3 Where's the probability?

- On N trials, X was observed to have taken on values x_1, x_2, ..., x_N respectively
- What does probability mean? On N trials, an event of probability p is expected to occur roughly pN times
- The event \{X = u_k\} is expected to occur roughly \[ P(X = u_k)N = pX(u_k)N \] times
- Moral: We expect that a_k, the number of times that X had value u_k on the N trials, is approximately \[ pX(u_k)N \]

7.3 ...and the average is approximately

- On N trials, X was observed to have taken on values x_1, x_2, ..., x_N respectively
- The average of these numbers is
  \[ \sum_{k=1}^{N} a_ku_k \]
  where for each k, u_k occurs a_k times among the N numbers being added
- \[ \sum a_k = N. \] Also, for each k, \[ a_k = pX(u_k)N \]
- Moral: The average value of X over many trials is approximately \[ \sum u_kpX(u_k) \]

7.3 Definition of expectation

- The average value of the discrete random variable X, averaged over many trials, is approximately \[ \sum u_kpX(u_k) \]
- We cannot guarantee that the average value will be exactly this number, but we expect it to be fairly close
- This notion is enshrined in probability theory by defining the expectation of X as \[ E[X] = \sum u_kpX(u_k) \]
- Other names for expectation are
  - the expected value of X
  - the mean or mean value of X
  - the average or the average value of X

7.3 I say what I mean ...

- The expectation of the discrete random variable X is given by
  \[ E[X] = \sum u_kpX(u_k) \]
  where the sum is over all values of X
- Read as “E of X” or “E X”
- Other names for expectation are
  - the expected value of X
  - the mean or mean value of X
  - the average or the average value of X

7.3 ... and I mean what I say

- The expectation of the discrete random variable X is
  \[ E[X] = \sum u_kpX(u_k) \]
- The operational meaning of \[ E[X] \] is exactly what we have discussed in this lecture: \[ E[X] \] is approximately what we expect to observe as the average value of X over many trials
- But there are no guarantees
Easy examples I

- Consider an experiment of rolling a fair die
- If the outcome is {1, 4}, you win $1; if the outcome is {2, 3, 5, 6}, you lose $1
- The random variable $X$ that denotes your winnings takes on values +1 and –1
- $p_X(u) = 1/3$ if $u = +1$; $p_X(u) = 2/3$ if $u = –1$
- $E[X] = (+1)\times(1/3) + (–1)\times(2/3) = –1/3$
- Note that in this experiment, $X$ will always differ from its mean $E[X]$ on every trial

Easy examples II

- Example: $X$ is a Bernoulli random variable with parameter $p$, $0 < p < 1$
- $p_X(u) = p$ if $u = 1$; $p_X(u) = 1–p$ if $u = 0$
- $E[X] = 0\times(1–p) + 1\times(p) = p$
- Here too, $X$ can never be equal to its mean on any trial

Easy examples III

- Example: $X$ is a random variable taking on integer values 0 through 8
- $p_X(0) = p_X(1) = p_X(7) = p_X(8) = 0.05$
- $p_X(2) = p_X(3) = p_X(5) = p_X(6) = 0.15$; $p_X(4) = 0.2$
- $E[X] = 0\times(0.05) + 1\times(0.05) + \ldots + 8\times(0.05) = 4$
- In this case, $X$ can equal its mean on some trials (in fact, $X = E[X]$ on roughly 20% of the trials)

The moment about the origin

- The expectation of the discrete random variable $X$ is $E[X] = \sum u_k \times p_X(u_k)$
- The pmf defines a set of point masses
- The point mass $p_X(u_k)$ is at distance $u_k$ from the origin and has a moment of $u_k \times p_X(u_k)$ about the origin
- Total moment of all the point masses is $E[X]$, the expectation of $X$

Location of the center of mass

- Total moment of all the point masses is just $E[X]$, the expectation of $X$
- Center of mass is defined by the equation $\text{Total moment} = \text{total mass} \times \text{center of mass}$
- But total probability mass is 1
- Hence, $E[X] = \text{location of center of mass}$
- There need not be any actual mass at the center of mass
- Practical example: consider a doughnut!

Fair entry fee for playing the game

- You stake $1 on the roll of a fair die
- If the outcome is {1, 4}, you get $2 back if the outcome is {2,3,5,6}, you get $0 back
- The random variable $Z$ denotes the return on your stake or investment
- $E[Z] = (2)\times(1/3) + (0)\times(2/3) = 2/3$
- $1$ stake means a loss of $1/3 on average
- A fair entry fee for playing this game (that is, your stake) would be $2/3

Fair entry fee for playing the game

- You stake $1 on the roll of a fair die
- If the outcome is {1, 4}, you get $2 back if the outcome is {2,3,5,6}, you get $0 back
- The random variable $Z$ denotes the return on your stake or investment
- $E[Z] = (2)\times(1/3) + (0)\times(2/3) = 2/3$
- $1$ stake means a loss of $1/3 on average
- A fair entry fee for playing this game (that is, your stake) would be $2/3

Fair entry fee for playing the game

- You stake $1 on the roll of a fair die
- If the outcome is {1, 4}, you get $2 back if the outcome is {2,3,5,6}, you get $0 back
- The random variable $Z$ denotes the return on your stake or investment
- $E[Z] = (2)\times(1/3) + (0)\times(2/3) = 2/3$
- $1$ stake means a loss of $1/3 on average
- A fair entry fee for playing this game (that is, your stake) would be $2/3

Fair entry fee for playing the game

- You stake $1 on the roll of a fair die
- If the outcome is {1, 4}, you get $2 back if the outcome is {2,3,5,6}, you get $0 back
- The random variable $Z$ denotes the return on your stake or investment
- $E[Z] = (2)\times(1/3) + (0)\times(2/3) = 2/3$
- $1$ stake means a loss of $1/3 on average
- A fair entry fee for playing this game (that is, your stake) would be $2/3
How many passengers get to fly?

- Example (continued): \( Y \) denotes the number of passengers who board the flight. Then, \( Y = X \) if \( X \leq 5 \) and \( Y = 5 \) if \( X > 5 \).
- \( p_Y(0) = p_X(1) = 0.05; p_Y(2) = p_X(3) = 0.15; p_Y(4) = 0.2; p_Y(5) = 0.4 \)
- \( E[Y] = 0 \cdot p_Y(0) + 1 \cdot p_Y(1) + 2 \cdot p_Y(2) + 3 \cdot p_Y(3) + 4 \cdot p_Y(4) + 5 \cdot p_Y(5) = 3.6 \)
- On average, 4 passengers show up for the flight but only 3.6 get to fly!

Y is a function of X

- Example (continued): \( Y \) is a function of \( X \) given by \( Y = X \) if \( X \leq 5 \) and \( Y = 5 \) if \( X > 5 \).
- We can write \( E[Y] = g(X) \) where \( g(\cdot) \) is the function shown below.

It bears repeating that...

- The standard definition \( E[Y] = \sum_Y p_Y(v) \) requires finding/knowing the pmf of \( Y \).
- To find \( E[Y] \), where \( Y = g(X) \), it is not necessary to first find the pmf of \( Y \) from the pmf of \( X \). Instead, we can use \( E[Y] = E[g(X)] = \sum_X g(u_i) \cdot p_X(u_i) \) where the sum is over all values \( u_i \) of \( X \).
- The new formula is just a re-arrangement of the terms in the standard sum for \( E[Y] \).

Conclusion from the messy formula

- To find \( E[Y] \), where \( Y = g(X) \), we can find the pmf of \( Y \) from the pmf of \( X \), and then use the standard result that \( E[Y] = \sum_Y p_Y(v) \), where sum is over all values \( v_i \) of \( Y \).
- Alternatively, we can find \( E[Y] \) via the formula \( E[Y] = E[g(X)] = \sum g(u_i) \cdot p_X(u_i) \), where the sum is over all values \( u_i \) of \( X \).
- The new formula is just a re-arrangement of the terms in the standard sum for \( E[Y] \).

Expected value of a function of \( X \)

- If \( Y = g(X) \), the expected value of \( Y \), which is given by \( E[Y] = \sum_Y p_Y(v) \) can often be more easily computed as \( E[Y] = E[g(X)] = \sum g(u_i) \cdot p_X(u_i) \), where the sum is over all \( u_i \) of \( X \).
- It is not necessary to find \( p_Y(v) \) to use in the standard definition of \( E[Y] \); the pmf of \( X \) and the function \( g(\cdot) \) suffice.
- This theorem is called LOTUS.
- LOTUS makes it easy to find \( E[Z] \) (slide 8).
Political correctness and LOTUS

- Many statistics texts define $E[g(X)]$ as $E[g(X)] = \sum g(u_k) \cdot p_X(u_k)$ where the sum is over all values $u_k$ of $X$, apparently without realizing that $g(X)$ is a random variable $Y$, and hence its expected value is, by definition, $E[Y] = \sum v_j \cdot p_Y(v_j)$.
- That both computations give the same result is a theorem of probability theory.
- The first four editions of Ross call this the Law of the Unconscious Statistician.

Applications of LOTUS I

- $Y = aX + b$, where $a$ and $b$ are constants.
  
  $E[Y] = E[g(X)] = \sum g(u_k) \cdot p_X(u_k) = \sum (a \cdot u_k + b) \cdot p_X(u_k) = a \cdot E[X] + b$

- $E[aX + b] = a \cdot E[X] + b$

- Expectation is a linear operation: the expectation of a sum is the sum of the expectations
  
  $E[b] = b$

Applications of LOTUS II

- Suppose that $Y = X - a$, where $a$ is some constant. Then,
  
  $E[Y] = E[X] - a$

- The mass $p_X(u_k)$ is at distance $u_k - a$ from the point $a$ on the real line

- $(u_k - a) \cdot p_X(u_k)$ is the moment of the mass $p_X(u_k)$ about the point $a$ on the real line

- $E[X - a] = \text{total moment about } a$

Applications of LOTUS III

- $E[X - a] = \text{total moment about the point } a$
- If we choose $a$ to be $E[X]$, then we get that $E[X - a] = E[X] - a = E[X] - E[X] = 0$
- $E[X]$ is the center of mass
- The moment about the center of mass is 0
- A body is "perfectly balanced" about its center of mass
- The quantity $X - E[X]$ is the deviation of $X$ from its mean
- The average deviation from the mean is 0

Applications of LOTUS IV

- The average deviation from the mean is 0
- We shall often use $\mu$ or $\mu_X$ to denote the mean, that is, expected value of $X$
- In probability theory as well as in many other areas, the average value of the squared deviation from $\mu$, that is, the expected value of $(X - \mu)^2$ is important

- $E[(X - \mu)^2]$ is called the variance of $X$

- Variance is denoted by $\text{var}(X)$ or $\sigma^2$ or $\sigma_X^2$ where $\sigma$ is called the standard deviation

The variance of a random variable

- $\sigma^2 = E[(X - \mu)^2]$ is called the variance of $X$
- LOTUS tells us that $\sigma^2 = E[(X - \mu)^2] = \sum (u_k - \mu)^2 \cdot p_X(u_k)$
- $\sigma^2 \geq 0$ since all the terms in the sum are nonnegative
- In fact, $\sigma^2 > 0$ except when the random variable happens to be a constant!
- If your computations give you negative variance or zero variance, you are likely to have made a mistake! Check your work!
More on the variance

- $\sigma^2 = E[(X - \mu)^2]$ is called the variance of $X$
- LOTUS tells us that
  
  $\sigma^2 = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2]$
  
  $= E[X^2] - 2\mu E[X] + \mu^2$
  
  $\therefore \sigma^2 = E[X^2] - \mu^2$

- Alternatively, expand out $\sum (u_k - \mu)^2 p_X(u_k)$ and verify that you get the same result
  
  $\sigma^2 = E[(X - \mu)^2] = \sum (u_k - \mu)^2 p_X(u_k)$
  
  $= E[X^2] - (E[X])^2$ or $E[X^2] - E^2[X]$

What does the variance represent?

- $\sigma^2 = E[(X - \mu)^2]$ is called the variance of $X$

- $\sigma^2 = \sum (u_k - \mu)^2 p_X(u_k)$

- The point mass $p_X(u_k)$ is at distance $u_k - \mu$ from the center of mass

- Hence, this point mass has moment of inertia $(u_k - \mu)^2 p_X(u_k)$ about the center of mass

- $\sigma^2$ denotes the moment of inertia about the center of mass

- $\sigma$ is also the radius of gyration

What does large variance mean?

- $\sigma^2$ denotes the moment of inertia about the center of mass

- $\sigma$ is also the radius of gyration: the system of point masses can be represented by a two half-unit masses at locations $\mu \pm \sigma$

- Since the total mass is always 1, a large variance means the total mass is spread widely and far away from the mean

- Small variance means mass is close to $\mu$

- A zero variance means all the mass is at $\mu$

Summary

- How to calculate probabilities of various interesting events from the pmf
- How to find the pmf of a function $g(X)$
- The average value of $X$ on repeated trials
- $E[X]$, the expectation of $X$
- Interpretations of $E[X]$
- Finding $E[g(X)]$ via LOTUS
- Variance of a random variable
- Properties of the variance