Review — Fundamentals

- A random variable **X** associates a number with each outcome of an experiment
- Observed value of **X** varies at random as the experiment is repeated
- Probabilistic behavior of a discrete random variable is described by its pmf
- Probabilities of events such as {a < **X** < b} can be calculated from pmf
- The pmf of **Y** = g(**X**) can be obtained from the pmf of **X**

Review — Expectation

- The average value of **X** on repeated trials
- E[**X**], the expectation of **X**, is given by

$$E[\mathbf{X}] = \sum u_k \bullet p_{\mathbf{X}}(u_k) = \mu \text{ or } \mu_{\mathbf{X}}$$

  where $p_{\mathbf{X}}(u)$ denotes the pmf of **X**
- Interpretations of E[**X**]
  - Average value of **X** over many trials
  - Moment about origin of prob. masses
  - μ is the location of the center of mass
  - Fair price to play game with winnings **X**

Review — LOTUS

- The expected value of **Y** = g(**X**) can be found by first finding the pmf of **Y** from the pmf of **X** and then using

$$E[\mathbf{Y}] = \sum v_j \bullet p_{\mathbf{Y}}(v_j)$$

  where the sum is over all values v_j of **Y**
- LOTUS: It is not necessary to first find the pmf of **Y**; E[**Y**] is also given by

$$E[\mathbf{Y}] = E[g(\mathbf{X})] = \sum g(u_k) \bullet p_{\mathbf{X}}(u_k)$$

  where the sum is over all values u_k of **X**
- Saves an extra computational step

Review — Applications of LOTUS

- **Y** = a**X** + b, where a and b are constants.
- E[**Y**] = a•E[**X**] + b • $\mu_{\mathbf{Y}} = a \bullet \mu_{\mathbf{X}} + b$
- E[a**X** + b] = a•E[**X**] + b
- Expectation is a linear operation: the expectation of a sum is the sum of the expectations
- E[a**X**] = a•E[**X**] • E[b] = b
- Masses in the pmf of **Y** = a**X** are "further away" from the origin by a factor of a, and so is the center of mass "further away"

This will hurt for just a few moments …

- E[**X** – a] is the (first) moment of **X** about a
- $E[(\mathbf{X} - a)^n]$ = n-th moment of **X** about a
- $E[\mathbf{X}^n]$ is called the n-th moment of **X**
- $E[(\mathbf{X} - \mu)^n]$ is the n-th central moment of **X**
- E[**X** – μ], the first central moment of **X** is 0
- $E[(\mathbf{X} - \mu)^2]$, the second central moment of **X**, is commonly called the variance of **X**

Review — Variance I

- $E[(\mathbf{X} - \mu)^2]$, the second central moment of **X**, is also known as the variance of **X**
- Variance is denoted by var(**X**) or σ^2 or $\sigma_{\mathbf{X}}^2$ where σ is called the standard deviation
- For a discrete random variable, LOTUS says $\sigma^2 = E[(\mathbf{X} - \mu)^2] = \sum (u_k - \mu)^2 \bullet p_{\mathbf{X}}(u_k)$
- $\sigma^2 > 0$ except for trivial random variables (a.k.a. constants) for which $\sigma^2 = 0$!
- σ^2 is the moment of inertia about μ; σ can be thought of as the radius of gyration

Review — Variance II

- $\sigma^2 = E[(\mathbf{X} - \mu)^2]$ is the variance of **X**
- $\sigma^2 = E[\mathbf{X}^2] - \mu^2$ which is also expressed sometimes as $\sigma^2 = E[\mathbf{X}^2] - (E[\mathbf{X}])^2$ or as $\sigma^2 = E[\mathbf{X}^2] - E^2[\mathbf{X}]$
- This alternative formula for σ^2 is very useful for actually calculating the variance
- Large variance means the total mass is spread widely and far away from the mean
- Small variance means mass is close to μ
- A zero variance means all the mass is at μ

Example of variance calculations I

- Example: **X** takes on values 1, 2, …, n
- Pmf of **X** is $p(k) = 2k/[n(n+1)]$ for $1 \le k \le n$
- Find the mean and variance of **X**
- $E[\mathbf{X}] = \sum k \cdot 2k/[n(n+1)]$; sum from k = 1 to n
  $= 2/[n(n+1)] \sum k^2$; sum from k = 1 to n
  $= 2/[n(n+1)] \cdot n(n+1)(2n+1)/6$
  $= (2n+1)/3$
- $\text{var}(\mathbf{X}) = E[(\mathbf{X} - \mu)^2] = E[(\mathbf{X} - (2n+1)/3)^2]$
  $= \sum (k - (2n+1)/3)^2 \cdot 2k/[n(n+1)]$

Example of variance calculations II

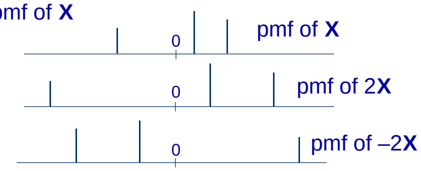
- Example: **X** takes on values 1, 2, …, n
- Pmf of **X** is $p(k) = 2k/[n(n+1)]$ for $1 \le k \le n$
- Find the mean and variance of **X**
- $E[\mathbf{X}] = (2n+1)/3$
- $E[\mathbf{X}^2] = \sum k^2 \cdot 2k/[n(n+1)]$; sum k = 1 to n
  $= 2/[n(n+1)] \sum k^3$; sum k = 1 to n
  $= 2/[n(n+1)] \cdot [n(n+1)/2]^2 = n(n+1)/2$
- $\text{var}(\mathbf{X}) = n(n+1)/2 - [(2n+1)/3]^2$
  $= (n^2 + n - 2)/18$ (Check: 0 if n = 1)

Example of variance calculations III

- If $\mathbf{Y} = g(\mathbf{X})$, what is $\text{var}(\mathbf{Y})$?
- Usually, it is easier to use the formula $\text{var}(\mathbf{Y}) = E[\mathbf{Y}^2] - (E[\mathbf{Y}])^2$ than to work directly with the definition of variance
- We already know that we can find $E[\mathbf{Y}]$ via LOTUS as $E[\mathbf{Y}] = E[g(\mathbf{X})]$
- $E[\mathbf{Y}^2] = E[(g(\mathbf{X}))^2]$ also can be found via LOTUS because $(g(\mathbf{X}))^2 = \mathbf{Z}$ (say), is just another function of **X**
- $\text{var}(\mathbf{X}^3) = E[\mathbf{X}^6] - (E[\mathbf{X}^3])^2$

Variance measures the spread I

- Let $\mathbf{Y} = a\mathbf{X}$
- What is $\text{var}(a\mathbf{X})$?
- The masses in the pmf of **Y** are "spread out" by a factor of a as compared to the pmf of **X**



Variance measures the spread II

- $E[a\mathbf{X}] = a \cdot E[\mathbf{X}] = a \cdot \mu_X$
- So the mean also moves by factor of a
- $E[(a\mathbf{X})^2] = a^2 \cdot E[\mathbf{X}^2]$
- $\text{var}(a\mathbf{X}) = E[(a\mathbf{X})^2] - (E[a\mathbf{X}])^2$
  $= a^2 \cdot E[\mathbf{X}^2] - (a \cdot \mu_X)^2$
  $= a^2 \cdot (E[\mathbf{X}^2] - (\mu_X)^2)$

  $\text{var}(a\mathbf{X}) = a^2 \cdot \text{var}(\mathbf{X})$
- If the masses are "spread out" by a factor of a, the variance increases by factor of a^2

Variance measures the spread III

- $var(a\mathbf{X}) = a^2 \cdot var(\mathbf{X})$
- Note that the sign of a does not matter
- $var(a\mathbf{X}) > 0$ even if a is negative
- Exercise: Show that

  $$var(\mathbf{X} + b) = var(\mathbf{X})$$

  that is, moving all the masses by a fixed distance b does not change the variance

  $$var(a\mathbf{X} + b) = var(a\mathbf{X}) = a^2 \cdot var(\mathbf{X})$$
- Variance is the second central moment

Reminiscences about expectations

- The expectation of a random variable tells roughly what the average value of the variable can be expected to be over a large number of trials
- We will now discuss some details about these large numbers of trials
  - How are the trials conducted?
  - If the results of N trials are viewed as a gigantic single experiment, how do we assign probabilities to this experiment?

Physical Independence of Trials

- The fundamental notion underlying the idea of expectation is that the repeated trials are conducted under absolutely identical conditions
- This is called physical independence in the sense that the outcomes of the various trials do not **physically** influence or affect one another in any way
- Note that this is primarily a belief: we cannot prove this lack of influence

Independent and Nonindependent

- Sampling without replacement: successive draws cannot be viewed as independent trials of the experiment
- Outcome of the first draw does influence the second draw — the same object cannot be drawn again
- Sampling with replacement: successive draws are independent trials of the experiment — the same object can be drawn again

Sampling without replacement

- Example: An urn contains 6 red balls and 3 green balls. Two balls are drawn at random without replacement from the urn
- 72 different samples without replacement can be drawn from the urn
- P(A) = P{first ball is green} = 3/9 = 1/3
- P(B) = P{second ball is green} = 3/9 = 1/3
- P(AB) = P{both balls are green} = 3/36
- $P(AB) \neq P(A)P(B)$

Sampling with replacement

- Example: An urn contains 6 red balls and 3 green balls. Two balls are drawn at random with replacement from the urn
- Here, the ball drawn first is replaced in the urn before the second ball is drawn
- Assumption: (not explicitly stated) The urn is shaken well before the second draw
- 81 different samples with replacement can be drawn from the urn

Sampling with replacement

- Example: An urn contains 6 red balls and 3 green balls. Two balls are drawn at random with replacement from the urn
- P(A) = P{first ball is green} = 3/9 = 1/3
- P(B) = P{second ball is green} = 3/9 = 1/3
- P(AB) = P{both balls are green}
  = P{(G_1,G_1), (G_1,G_2), (G_1,G_3),…,(G_3,G_3)}
  = 9/81 = 1/9 ≠ 3/36 for sampling without replacement. Note also: P(AB) = P(A)P(B)

Independent Trials

- Notion of independent trials: the outcomes of the various trials do not influence or affect one another in any way
- Knowing that the first ball was green does not affect our beliefs about the chances of the second ball being green
- Contrast this with sampling without replacement — if the first ball is green, the second ball has only 2/8 chance of being green (instead of 3/9)

Beliefs

- It is not possible to determine whether repeated trials are independent or not
- Independence of trials is a belief and cannot be proved mathematically
- We assume that the trials are independent (and do our best (by vigorous shaking of urns) to ensure this as much as possible)
- Be careful: Eight perfect riffle shuffles of a deck of cards restore it to the original state

Simple vs Compound Experiments

- Consider an experiment with sample space Ω = {a_1, a_2, … }
- We shall call this a simple experiment
- The result of repeated independent trials of this experiment is a sequence or vector of outcomes, say, (a_5, a_2, a_7, a_9, a_1, …)
- This vector is regarded as the outcome of a compound experiment with sample space Ω × Ω × Ω × …
- Simple experiments are subexperiments

Events on Compound Experiments

- The outcome of a compound experiment is a sequence or vector of outcomes of the form

  (a_5, a_2, a_7, a_9, a_1, …)
- The simple event A occurred on i-th subexperiment if the i-th outcome in this sequence is a member of the event A   Ω
- The compound event (A, B, C, A^c, …) occurred if a_5 A, a_2 B, a_7 C, a_9 A^c, …

Declaration of Independence

- The assumption of independence of trials is declared in the axiomatic theory by the way probabilities are assigned to the events of the compound experiment
- If the trials are (assumed to be) independent, then we set

  P(A, B, C, A^c, …) = P(A)P(B)P(C)P(A^c)…
- Both A and A^c cannot occur on the same trial of the simple experiment: here they are occurring on different subexperiments

Random Variables

- $\mathbf{X}$ is a random variable defined on the simple experiment. It maps a_5 to $\mathbf{X}(a_5)$
- $\mathbf{X}_i$ denotes the number observed on the i-th subexperiment of the compound experiment
- $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots)$ is called a random vector
- If the outcome is $(a_5, a_2, a_7, a_9, a_1, \ldots)$, then $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5, \ldots$ have values $\mathbf{X}(a_5), \mathbf{X}(a_2), \mathbf{X}(a_7), \mathbf{X}(a_9), \mathbf{X}(a_1), \ldots$ etc

Independent random variables

- For repeated independent trials, the random variables $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \ldots$ are said to be independent random variables
- If $\mathbf{X}$ is a discrete random variable, then for repeated independent trials, we have

$P(\mathbf{X}_1 = a_5, \mathbf{X}_2 = a_2, \mathbf{X}_3 = a_7, \mathbf{X}_4 = a_9, \ldots) =$

$= P(\mathbf{X}_1{=}a_5)P(\mathbf{X}_2{=}a_2)P(\mathbf{X}_3{=}a_7)P(\mathbf{X}_4{=}a_9)\ldots$

- This is just $P(A, B, C, A^c, \ldots) = P(A)P(B)P(C)P(A^c)\ldots$

Counting random variables

- $\mathbf{X}$, the indicator function of an event A of probability p is a Bernoulli random variable with parameter p
- $\mathbf{X}$ equals 1 if the event A occurred
- $\mathbf{X}_i = 1$ if event A occurred on the i-th trial
- Let $\mathbf{Y} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_4 + \ldots$
- $\mathbf{Y}$ = number of times that A occurred on n trials, is called a counting random variable
- $\mathbf{X}_i$ is also a counting random variable

Not defined on a subexperiment!

- $\mathbf{Y} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 + \mathbf{X}_4 + \ldots$ is a random variable that is defined on the compound experiment — its value is not determined completely by what happened on any specific subexperiment
- $\mathbf{Y}$ = # of times that A occurred on n trials
- $\mathbf{Y}$ counts the occurrences of A
- $\mathbf{Y}$ takes on values 0, 1, 2, … n
- What is the pmf of $\mathbf{Y}$?

Binomial random variable

- $\mathbf{Y}$ is called a binomial random variable with parameters (n, p)
- $\mathbf{Y}$ = # of times that an event A of probability p occurred on n trials
- $\mathbf{Y}$ takes on values 0, 1, 2, … n
- $p(k) = P(\mathbf{Y} = k) = ?$
- What is P(A occurred on k specific trials out of the n (and A^c occurred on the rest))
- $P(\mathbf{X}_1 = 1, \mathbf{X}_2 = 0, \mathbf{X}_3 = 0, \mathbf{X}_4 = 1, \ldots)$

Probability of a specific outcome

- $P(\mathbf{X}_1 = 1, \mathbf{X}_2 = 0, \mathbf{X}_3 = 0, \mathbf{X}_4 = 1, \ldots)$?
- Here, there are k 1's and n–k 0's indicating the occurrence and non-occurrence of A on specific trials
- $P(\mathbf{X}_1 = 1, \mathbf{X}_2 = 0, \mathbf{X}_3 = 0, \mathbf{X}_4 = 1, \ldots)$
  $= P(\mathbf{X}_1 = 1)P(\mathbf{X}_2 = 0)P(\mathbf{X}_3 = 0)P(\mathbf{X}_4 = 1)\ldots$
  $= p(1-p)(1-p)p\ldots = p^k(1-p)^{n-k}$
- If we had chosen a different selection of k trials, we would have got the same result

An example

- Let n = 4 and k = 2
- $P(A, A, A^c, A^c) = p^2(1-p)^2$
  $P(A, A^c, A, A^c) = p^2(1-p)^2$
  $P(A, A^c, A^c, A) = p^2(1-p)^2$
  $P(A^c, A, A, A^c) = p^2(1-p)^2$
  $P(A^c, A, A^c, A) = p^2(1-p)^2$
  $P(A^c, A^c, A, A) = p^2(1-p)^2$
- Each of these events is disjoint from the others (because A and A^c cannot both occur on the same trial)

How many selections?

- The probability that A occurs on k specific trials and does not occur on the remaining n–k trials is $p^k(1-p)^{n-k}$
- If we had chosen a different selection of k trials, we would have got the same result
- The k trials on which A occurs can be specified by stating the subset (of size k) of {1, 2, 3, … , n} on which A occurred
- How many such subsets are there? $\binom{n}{k}$

Binomial Probabilities

- The probability that A occurs on k specific trials and does not occur on the other n–k trials is $p^k(1-p)^{n-k}$
- P{A occurs on k trials out of n} = ?
- The difference between the two questions is that we don't care which k trials A occurs on, as long as A occurs exactly k times on the n trials
- P{A occurs on k of n trials} = $\binom{n}{k} p^k(1-p)^{n-k}$

Binomial Random Variables

- **Y** is called a binomial random variable with parameters (n, p)
- **Y** counts the number of times that an event A of probability p occurs on n independent trials
- **Y** takes on values {0, 1, 2, … ,n}
- For $0 \le k \le n$,
  $p_{\mathbf{Y}}(k) = P\{\mathbf{Y} = k\} = \binom{n}{k} p^k(1-p)^{n-k}$

Is this a valid pmf?

- For $0 \le k \le n$,
  $p_{\mathbf{Y}}(k) = P\{\mathbf{Y} = k\} = \binom{n}{k} p^k(1-p)^{n-k}$
- The binomial theorem asserts that

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

- Putting x = p and y = 1 – p shows that the sum of the probability masses (the RHS) is 1, the value of the LHS

Probabilities from the binomial pmf

- Probabilities such as $P\{a < \mathbf{Y} < b\}$ are found by summing up the appropriate terms in the pmf
- There are no closed-form expressions for such probabilities: numerical evaluation is required
- $P\{a < \mathbf{Y} < b\}$ is the sum of L (say) of the n+1 probabilities in the pmf
- If L > (n+1)/2, find 1 – P{complement}

Example

- **Y** is a binomial random variable with parameters (70, 0.1). What is P{**Y** > 5}?
- P{**Y** > 5} = P{**Y**=6} + P{**Y**=7} +…+ P{**Y**=70}

$$= \sum_{k=6}^{70} \binom{70}{k} 0.1^k 0.9^{70-k}$$

- It is a lot easier to compute P{**Y** ≤ 5} (which has only 6 terms) and subtract from 1 to find P{**Y** > 5}

Mean and variance

- The mean and variance of the binomial random variable with parameters (n, p) are calculated in the textbook
- E[**Y**] = np
- var(**Y**) = np(1–p)
- In calculating the variance, it is convenient to use var(**Y**) = E[**Y**(**Y**–1)] + E[**Y**] – $(E[\mathbf{Y}])^2$
- Exercise: Prove this identity
- Exercise: Find P{**Y** is even}

Summary

- We discussed some properties of variance
- We discussed the concept of independent trials and how probabilities are assigned
- We discussed simple versus compound experiments
- We defined a binomial random variable with parameters (n, p) as a counting variable — it counts the number of occurrences of an event A of probability p on n independent trials