Review: binomial random variables

- **X** denotes the number of occurrences of an event A of probability p on n trials
- **X** is called a binomial random variable with parameters (n, p)
- **X** takes on values 0, 1, 2, … n
- For $0 \le k \le n$,

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

- Mean $E[X] = np$, variance $np(1-p)$, and mode $\lfloor (n+1)p \rfloor$

Just so's there's no confusion …

- Binomial random variables **are** one class of the important discrete random variables that are the subject of this lecture
- Memorize the basic information shown on the previous slide
- … or at least have it on your sheet(s) of notes on the exams!
- On exams, you are expected to have this basic information at your fingertips

What if n is large and p is small?

- Suppose that n is large and p is very small
- The pmf of the binomial random variable **X** can be approximated and expressed as a function of n•p, that is, we do not need to know the values of n and p separately; their product is all that is required
- $\lambda = n \cdot p$
- It is assumed that λ is of moderate size
- What's this large, small, moderate stuff?

n is large, p is small, λ is moderate?

- And Baby Bear's chair was just right, too?
- The approximations that we will develop work reasonably well even for relatively small values of n, e.g. $n \approx 100$
- n large, p small and $\lambda = n \cdot p$ is moderate means

  $$p \ll 1 \text{ and } \lambda \ll n$$

- For $n \approx 100$, a λ of 15 or less will give reasonably accurate approximations

When all else fails, approximate …

- Remember $np = \lambda$. Then, for $k \ll n$,

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

  $$= \frac{n(n-1)(n-2)\ldots(n-k+1)}{1 \times 2 \times 3 \times \ldots \times k} \left(\frac{\lambda}{n}\right)^k \left(1-p\right)^{n-k}$$

- Consider the ratio containing k terms

  $$n(n-1)(n-2)\ldots(n-k+1)/[n(1-p)]^k$$

- The numerator terms are very slightly less than n, and so are the denominator terms
- This ratio is approximately 1

Aargh! Calculus rears its ugly head!

- Remember $np = \lambda$. Then, for $k \ll n$,

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

  $$= \frac{n(n-1)(n-2)\ldots(n-k+1)}{1 \times 2 \times 3 \times \ldots \times k} \left(\frac{\lambda}{n}\right)^k \left(1-p\right)^{n-k}$$

  $$\approx \frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n \approx \exp(-\lambda)\frac{\lambda^k}{k!}$$

- Here we have used the following result from calculus: $\lim_{n \to \infty}(1 - x/n)^n = \exp(-x)$

The Poisson approximation

- The Poisson approximation to the binomial pmf applies when n is large and p is small
- $np = \lambda$. Then, for $k \ll n$,

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

$$\approx \exp(-\lambda)\frac{\lambda^k}{k!}$$

- The Poisson approximation is sometimes called the Law of Small Numbers
- Poor results when k is comparable to n

The Poisson random variable

- **Y** is called a Poisson random variable with parameter λ if its pmf is given by

$$p_Y(k) = \exp(-\lambda)\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \ldots$$

- Note that the Poisson random variable can take on any nonnegative integer value
- Since the Taylor series for exp(x) is $\sum x^k/k!$ where k ranges from 0 to ∞, the pmf sums to $\exp(-\lambda)\bullet\exp(\lambda) = 1$

An episode of Seinfeld

- A Poisson random variable **Y** has pmf $\exp(-\lambda)\bullet\lambda^k/k!$, $k = 0, 1, 2, 3, \ldots$
- $p_Y(0) = P\{\mathbf{Y} = 0\} = \exp(-\lambda)$
- This is the probability that "nothing happens"
- $E[\mathbf{Y}] = \lambda$
- If, on average, many events are expected to occur, i.e. $\lambda \gg 0$, then $P\{\mathbf{Y} = 0\}$ is small
- $p_Y(1) = P\{\mathbf{Y} = 1\} = \lambda\bullet\exp(-\lambda)$

WCW Nitro: Poisson vs Binomial I

- A Poisson random variable **Y** can take on any nonnegative integer value: a binomial random variable **X** takes on only the integer values 0, 1, 2, … , n
- $p_X(k) \approx p_Y(k)$ under certain conditions
- Conditions:
  - n is large
  - p is very small and $p = \lambda/n$
  - $k \ll n$

WCW Nitro: Poisson vs Binomial II

- The Poisson pmf is the limiting form of the binomial pmf as $n \to \infty$ and $p \to 0$ such that np has the constant value λ
- In view of this, it is not too surprising that
  - $E[\mathbf{Y}] = \lambda \quad \Leftrightarrow \quad E[\mathbf{X}] = np$
  - $\mathrm{var}(\mathbf{Y}) = \lambda \quad \Leftrightarrow \quad \mathrm{var}(\mathbf{X}) = np(1-p)$
  - $\mathrm{Mode}(\mathbf{Y}) = \lfloor\lambda\rfloor \quad \Leftrightarrow \quad \mathrm{mode}(\mathbf{X}) = \lfloor(n+1)p\rfloor$
  - ML estimate of λ given $\{\mathbf{Y} = k\}$ is k

    ML estimate of p given $\{\mathbf{X} = k\}$ is k/n

WCW Nitro: Poisson vs Binomial III

- Poisson random variables are counting random variables; just like the binomial
- A Poisson random variable counts the number of occurrences of a very rare event on a very large number of trials
- These trials are imagined to have been conducted simultaneously
- In the next minute, 10000 web-surfers connected to www.yahoo.com will (or will not) decide to click on a particular link

Where do we see Poisson RVs?

- Poisson random variables model many phenomena that we measure per unit time
  - number of packet transmission requests
  - number of processes initiated
  - number of phone calls arriving at exchange
  - number of α-particles emitted by source
  - number of electrons emitted by cathode
  - number of defective widgets in production run
  - number of cars not exceeding the speed limit
  - number of RVs passing a checkpoint

Remember this one too!

- Poisson random variables are yet another class of random variables that arise so often in applications that we expect you to know (or have on your sheet of notes) all the relevant information about them
- We shall discuss Poisson random variables later in the context of what is called a Poisson process

Waiting for Godot

- We return to the consideration of repeated independent trials of a simple experiment
- We number the trials as #1, #2, etc.
- On each trial, an event A of probability $P(A) = p$ may occur ("a success") or it may not ("a failure")
- Let **X** denote the number of the trial on which the first success occurred
- Then, **X** is a counting random variable that takes on values 1, 2, 3, …

Independent trials, remember?

- Let **X** denotes the number of the trial on which event A occurred for the first time
- **X** is a counting random variable that takes on values 1, 2, 3, …
- $P\{\mathbf{X} = 1\} = ?$
- $\{\mathbf{X} = 1\}$ means A occurred on the first trial, and this has probability p. $P\{\mathbf{X} = 1\} = p$
- $P\{\mathbf{X} = 2\} = (1-p)\bullet p$ since A^c must have occurred on the 1st trial, and A on the 2nd

Geometric random variable

- More generally, $P\{\mathbf{X} = k\} = (1-p)^{k-1}\bullet p$ since A^c must have occurred on the first k–1 trials, and A on the k-th
- A random variable taking on integer values 1, 2, 3, … with probabilities

$$p_X(k) = (1-p)^{k-1}\bullet p, \quad k \geq 1$$

  is called a geometric random variable with parameter p
- Geometric because the pmf values are a geometric series

Basic properties

- For a geometric random variable with parameter p, $E[\mathbf{X}] = 1/p$
- If the chances for the Fighting Illini football team to make it to the Rose Bowl are 1/30, then on average, the team will go once every 30 years or so
- $\text{var}(\mathbf{X}) = (1-p)/p^2$
- See Ross, Chapter 4 for details of the computation

Some associated probabilities I

- A geometric random variable has pmf
  $p_X(k) = (1–p)^{k–1}•p, \quad k ≥ 1$
- $P\{X ≤ n\} = p_X(1) + p_X(2) + … + p_X(n)$
  $= p + (1–p)•p + … + (1–p)^{n–1}•p$
  $= p•[1 + (1–p) + … + (1–p)^{n–1}]$
  $= p•[1 – (1–p)^n]/[1 – (1–p)]$
  $= 1 – (1–p)^n$
- $P\{X > n\} = (1–p)^n$

Some associated probabilities II

- A geometric random variable has pmf
  $p_X(k) = (1–p)^{k–1}•p, \quad k ≥ 1$
- $P\{X > n\} = (1–p)^n$ can be obtained directly (and more easily) as follows
- $\{X > n\}$ if and only if the first n trials resulted in failures, that is, A^c occurred on the first n trials. The probability of this is just $(1–p)^n$

Crossing the i's and dotting the t's

- A geometric random variable has pmf
  $p_X(k) = (1–p)^{k–1}•p, \quad k ≥ 1$
- $P\{X > n\} = p_X(n+1) + p_X(n+2) + …$
  $= (1–p)^n•p + (1–p)^{n+1}•p + …$
  $= (1–p)^n•p •[1 + (1–p) + (1–p)^2 …]$
  $= (1–p)^n•p /[1 – (1–p)]$
  $= (1–p)^n$

Some associated probabilities III

- Knowing how to sum geometric series is the key
- $P\{X$ is a multiple of $k\}$
  $= p_X(k) + p_X(2k) + p_X(3k) + …$
  $= (1–p)^{k–1}•p •[1 + (1–p)^k + (1–p)^{2k} …]$
  $= (1–p)^{k–1}•p /[1 – (1–p)^k]$
- $P\{X$ is even$\} = [1–p]/[2–p]$
- Similarly, $P\{X$ is odd$\} = 1/[2–p]$

A simple game

- Fred and Wilma take turns tossing a coin with P(Head) = p. The first to toss a Head wins the game. Fred begins the game
- **X**, the number of tosses is a geometric random variable with parameter p
- P{Fred wins} = $P\{X$ is odd$\} = 1/[2–p]$
  P{Wilma wins} = $P\{X$ is even$\} = [1–p]/[2–p]$
- P{Fred wins} > P{Wilma wins}
- P{game goes on forever} = 0

Many players?

- If there are k players (Barney? Betty?…), the win probabilities are in the ratio
  $1:q:q^2:q^3:……:q^{k–1}$
  where q = 1–p (a commonly used notation)
- Notice that win probabilities are smaller for players whose turns come later
- The player whose turn is the i-th wins with probability $q^{i–1}/[1 + q + q^2 + q^3 + … + q^{k–1}]$
- P{game goes on forever} = 0

Any engineering applications?

- Data transmission systems occasionally change 0's to 1's or vice versa
- To protect against transmission errors, most systems employ an error-detecting code known as a CRC (cyclic redundancy check) code
- A data packet of N bits contains k data bits (the payload) and N–k overhead bits (including CRC, header, trailer, address, time-stamp, etc)

ARQ communication systems I

- The receiver checks if the received packet is a valid codeword in the CRC code
- A valid codeword is accepted and the payload sent to its destination
- If the received packet is not a valid codeword, the receiver asks that the complete packet be re-transmitted (in other words, repeated)
- ARQ = automatic repeat request

ARQ communication systems II

- Let p denote the probability of a bit error on the transmission channel
- Successive bit transmissions used to send the packet are N independent trials
- The number of bit errors in the packet is a binomial (N,p) random variable
- P{no errors} = $(1 - p)^N$
- If there are no transmission errors, the received packet is a valid codeword

ARQ communication systems III

- Long messages are divided into packets
- How long should each packet be?
- The overhead N–k = r bits is fixed: the question is "What should k be?"
- If k >> r, we save on the overhead
- But increasing k increases N = k+r
- $Q = (1 - p)^N$, the probability of the packet being received without any errors in it, goes to 0 as N increases

ARQ communication systems IV

- Make k very small so that Q = P{successful transmission} is high
- But, each packet contains a large number of overhead bits. A lot of the transmission capacity is wasted on the overhead
- Q is the probability that the packet is transmitted successfully
- If the transmission is unsuccessful, that is, the receiver requests a repeat, the transmitter sends the packet again

If at first you don't succeed, try, try, again…

- Q is the probability that the packet is transmitted successfully
- If the transmission is unsuccessful, that is, the receiver requests a repeat, the transmitter sends the packet again
- The total number of times that a packet is transmitted is a geometric random variable with parameter Q
- On average, a packet is transmitted 1/Q times over the channel

ARQ communication systems VI

- $Q = (1 - p)^N$ is the probability that the packet is transmitted successfully
- On average, a packet is transmitted 1/Q times over the channel
- On average, N/Q bits have been transmitted to send k bits of payload
- Effective average data transmission rate is $kQ/N = k(1 - p)^N/N = k(1-p)^{k+r}/[k+r]$
- What value of k maximizes the data rate?

ARQ communication systems VII

- Look at the ratio

$$\frac{(k+1)(1-p)^{k+r+1}/[k+r+1]}{k(1-p)^{k+r}/[k+r]}$$

- The ratio equals 1 if and only if
  $k(k+r+1) = (1-p)r/p$
- $k(k+r+1) \approx (k+r/2)^2 = k^2 + kr + r^2/4$
- $k \approx [(1-p)r/p]^{1/2} - r/2 \approx [r/p]^{1/2} - r/2$
- Example: $p = 10^{-4}$ and $r = 30$: $k \approx 533$
- Rounding off to nearest multiple of 8 is OK

And now for something completely different…

- The geometric random variable is the simplest of a class of random variables called "waiting time variables"
- Let $\mathbf{X}_r$ denote the independent trial on which an event A occurred for the r-th time
- Then, $\mathbf{X}_r$ takes on values r, r+1, r+2, …
- For $n \geq r$, the event $\{\mathbf{X}_r = n\}$ occurs if and only if A occurred on the n-th trial and it occurred r–1 times on previous n–1 trials

The waiting time for the r-th success

- The outcomes of trials are independent of one another
- P{A occurred for r-th time on the n-th trial}
  $= P\{\mathbf{X}_r = n\}$
  = P{A on n-th}P{A on r–1 of previous n–1}
  $= p \bullet \binom{n-1}{r-1} \bullet p^{r-1}(1-p)^{n-r}$
  $= \binom{n-1}{r-1} \bullet p^r(1-p)^{n-r}$

Negative binomial random variables

- $\mathbf{X}_r$ is called a negative binomial random variable with parameters r and p
- It counts the number of trials required to accumulate r successes
- It measures the waiting time till the occurrence of the r-th success
- The pmf of $\mathbf{X}_r$ is given by

$$p_{\mathbf{X}_r}(n) = \binom{n-1}{r-1} \bullet p^r(1-p)^{n-r}, \quad r \leq n < \infty$$

Negative binomial random variables

- $\mathbf{X}_1$, the waiting time for the first success is just a geometric random variable with parameter p
- To accumulate r successes, we wait for the first success. The waiting time is a geometric random variable
- Once the first success has been achieved, we then wait for the second success, and again our waiting time is a geometric random variable

Negative binomial random variables

- $\mathbf{X}_r$, the waiting time for the r-th success is actually the sum of the r successive waiting times for a single success
- $\mathbf{X}_r$ is the sum of r independent geometric random variables with parameter p
- Independent because the geometric RVs are defined on independent trials
- $E[\mathbf{X}_r] = r \bullet E[\mathbf{X}_1] = r/p$ (expectation of a sum is the sum of the expectations)

Why so negative? or down with PC

- Why are negative binomial random variables given such a pejorative name?
- The pmf values for a binomial (n, p) RV are the ones obtained in the expansion of $[p + (1–p)]^n$ via the binomial theorem
- For a negative binomial (r, p) RV, the pmf values are the ones obtained in the expansion of $[p + (1–p)]^{–r}$ via the binomial theorem

Binomial theorem for neg. exponent

- Expanding $[1 + x]^{–r}$ via the binomial theorem gives an infinite series for which the k-th term (coefficient of x^k) is

  $$r(r+1)(r+2)\ldots(r+k–1)/k!$$
- If n = r+k,

  $$\binom{n-1}{r-1} = \binom{r+k-1}{r-1} = \binom{r+k-1}{k}$$

  $= (r+k–1)(r+k–2)\ldots(r+k–1 – k + 1)/k!$

  $= r(r+1)(r+2)\ldots(r+k–1)/k!$

Memory…

- What must you remember (or have on your sheet of notes) from this lecture
- Definitions, pmfs, means, variances, etc of the following classes of random variables
  - Binomial random variables
  - Poisson random variables
  - Geometric random variables
  - Negative binomial random variables
- Knowing how to use this is more important