Recall the Expected Value

- The Expected Values for a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} x\, p(x)$$

Lying With Statistics

"There are three kinds of lies:
lies, damned lies, and statistics"

– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a <u>class</u> with equal probability
- X = size of chosen class
- What is E[X]?
  - $E[X] = 5\,(1/3) + 10\,(1/3) + 150\,(1/3)$
    $= 165/3 = 55$

Lying With Statistics

"There are three kinds of lies:
lies, damned lies, and statistics"

– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a <u>student</u> with equal probability
- Y = size of class that student is in
- What is E[Y]?
  - $E[Y] = 5\,(5/165) + 10\,(10/165) + 150\,(150/165)$
    $= 22635/165 \approx 137$
- Note: E[Y] is students' perception of class size
  - But E[X] is what is usually reported by schools!

Expectation of a Random Variable

- Let Y = g(X), where g is real-valued function

$$E[g(X)] = E[Y] = \sum_j y_j \, p(y_j)$$

$$= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i)$$

$$= \sum_j \sum_{i:g(x_i)=y_j} y_j \; p(x_i)$$

$$= \sum_j \sum_{i:g(x_i)=y_j} g(x_i) \, p(x_i)$$

$$= \sum_i g(x_i) \, p(x_i)$$

Other Properties of Expectations

- Linearity:

$$E[aX + b] = aE[X] + b$$

  - Consider X = 6-sided die roll, Y = 2X – 1.
  - E[X] = 3.5 E[Y] = 6

- *N*-th Moment of X:

$$E[X^n] = \sum_{x:\, p(x)>0} x^n p(x)$$

  - We'll see the 2^{nd} moment soon...

Utility

- Utility is value of some choice
  - 2 choices, each with n consequences: $c_1, c_2, \ldots, c_n$
  - One of c_i will occur with probability p_i
  - Each consequence has some value (utility): $U(c_i)$
  - Which choice do you make?
- Example: Buy a \$1 lottery ticket (for \$1M prize)?
  - Probability of winning is $1/10^7$
  - **Buy**: c_1 = win, c_2 = lose, $U(c_1) = 10^6 - 1$, $U(c_2) = -1$
  - **Don't Buy**: c_1 = lose, $U(c_1) = 0$
  - $E(\text{buy}) = 1/10^7 (10^6 - 1) + (1 - 1/10^7)(-1) \approx -0.9$
  - $E(\text{don't buy}) = 1 (0) = 0$
  - *"You can't lose if you don't play!"*

And Then There's This…

Lottery: A tax on people who are bad at math.

– Ambrose Bierce

Welcome to St. Petersburg!

- Game set-up
  - We have a fair coin (come up "heads" with p = 0.5)
  - Let *n* = number of coin flips ("heads") before first "tails"
  - You win $\$2^n$
- How much would you pay to play?
- Solution
  - Let X = your winnings
  - E[X] = $\left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \ldots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i$

    $$= \sum_{i=0}^{\infty} \frac{1}{2} = \infty$$
  - I'll let you play for $1 million... but just once! Takers?

Breaking Vegas

- Consider even money bet (e.g., bet "Red" in roulette)
  - p = 18/38 you win \$Y, otherwise (1 – p) you lose \$Y
  - Consider this algorithm for one series of bets:
    1. Y = \$1
    2. Bet Y
    3. If Win then stop
    4. If Loss then Y = 2 * Y, goto 2
  - Let Z = winnings upon stopping
  - E[Z] $= \left(\frac{18}{38}\right)1 + \left(\frac{20}{38}\right)\left(\frac{18}{38}\right)(2-1) + \left(\frac{20}{38}\right)^2\left(\frac{18}{38}\right)(4-2-1) + \ldots$

$$= \sum_{i=0}^{\infty}\left(\frac{20}{38}\right)^i\left(\frac{18}{38}\right)\left(2^i - \sum_{j=0}^{i-1} 2^j\right) = \left(\frac{18}{38}\right)\sum_{i=0}^{\infty}\left(\frac{20}{38}\right)^i = \left(\frac{18}{38}\right)\frac{1}{1-\frac{20}{38}} = 1$$

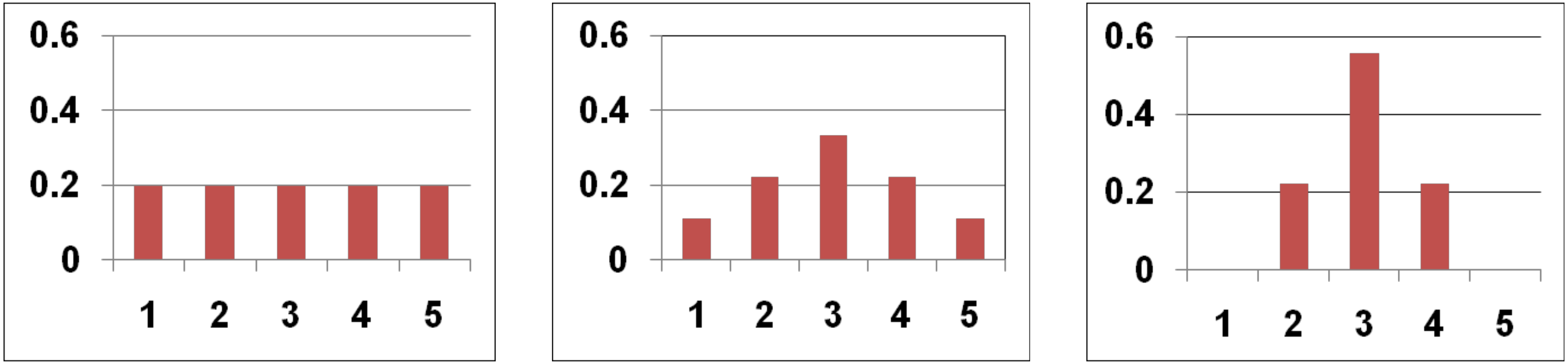
  - Expected winnings ≥ 0. Use algorithm infinitely often!

Vegas Breaks You

- Why doesn’t everyone do this?
  - Real games have maximum bet amounts
  - You have finite money
    - Not able to keep doubling bet beyond certain point
  - Casinos can kick you out
- But, if you had:
  - No betting limits, and
  - Infinite money, and
  - Could play as often as you want...
- Then, go for it!
  - And tell me which planet you are living on

Variance

- Consider the following 3 distributions (PMFs)



- All have the same expected value, E[X] = 3
- But “spread” in distributions is different
- Variance = a formal quantification of “spread”

Variance

- If X is a random variable with mean μ then the **variance** of X, denoted Var(X), is:

$$\text{Var}(X) = E[(X - \mu)^2]$$

- Note: Var(X) ≥ 0

- Also known as the 2nd Central Moment, or square of the Standard Deviation

Computing Variance

$$
\begin{aligned}
\mathrm{Var}(X) &= E[(X-\mu)^2] \\
&= \sum_x (x-\mu)^2 p(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2)\, p(x) \\
&= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\
&= \boxed{E[X^2]} - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2 \\
&\boxed{= E[X^2] - (E[X])^2}
\end{aligned}
$$

Say hello to my little friend, the 2nd moment!

Variance of 6 Sided Die

- Let X = value on roll of 6 sided die
- Recall that E[X] = 7/2
- Compute E[X^2]

$$E[X^2] = (1^2)\frac{1}{6} + (2^2)\frac{1}{6} + (3^2)\frac{1}{6} + (4^2)\frac{1}{6} + (5^2)\frac{1}{6} + (6^2)\frac{1}{6} = \frac{91}{6}$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Properties of Variance

- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
  - Proof:

    $\mathrm{Var}(aX + b) = E[(aX + b)^2] - (E[aX + b])^2$

    $= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2$

    $= a^2E[X^2] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2)$

    $= a^2E[X^2] - a^2(E[X])^2 = a^2(E[X^2] - (E[X])^2)$

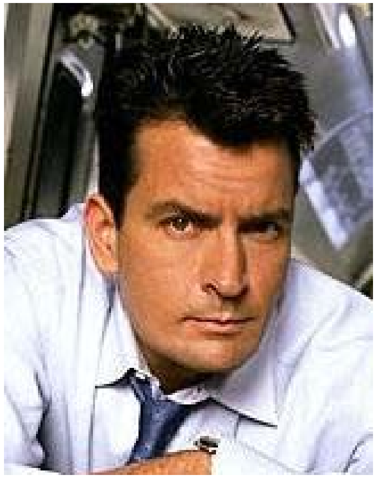
    $= a^2\mathrm{Var}(X)$
- Standard Deviation of X, denoted SD(X), is:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

  - Var(X) is in units of X^2
  - SD(X) is in same units as X

Jacob Bernoulli

- Jacob Bernoulli (1654-1705), also known as "James", was a Swiss mathematician



- One of many mathematicians in Bernoulli family
- The Bernoulli Random Variable is named for him
- He is my *academic* great^{11}-grandfather
- Resemblance to Charlie Sheen weak at best

Bernoulli Random Variable

- Experiment results in “Success” or “Failure”
  - X is random indicator variable (1 = success, 0 = failure)
  - $P(X = 1) = p(1) = p$ $P(X = 0) = p(0) = 1 – p$
  - X is a **Bernoulli** Random Variable: $X \sim Ber(p)$
  - $E[X] = p$
  - $Var(X) = p(1 – p)$
- Examples
  - coin flip
  - random binary digit
  - whether a disk drive crashed

Binomial Random Variable

- Consider *n* independent trials of Ber(p) rand. var.
  - X is number of successes in *n* trials
  - X is a **<u>Binomial</u>** Random Variable: X ~ Bin(n, p)

$$P(X = i) = p(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0,1,\ldots,n$$

  - By Binomial Theorem, we know that $\sum_{i=0}^{\infty} P(X = i) = 1$
- Examples
  - # of heads in *n* coin flips
  - # of 1's in randomly generated length *n* bit string
  - # of disk drives crashed in 1000 computer cluster
    - Assuming disks crash independently

Three Coin Flips

- Three fair (“heads” with p = 0.5) coins are flipped
  - X is number of heads
  - X ~ Bin(3, 0.5)

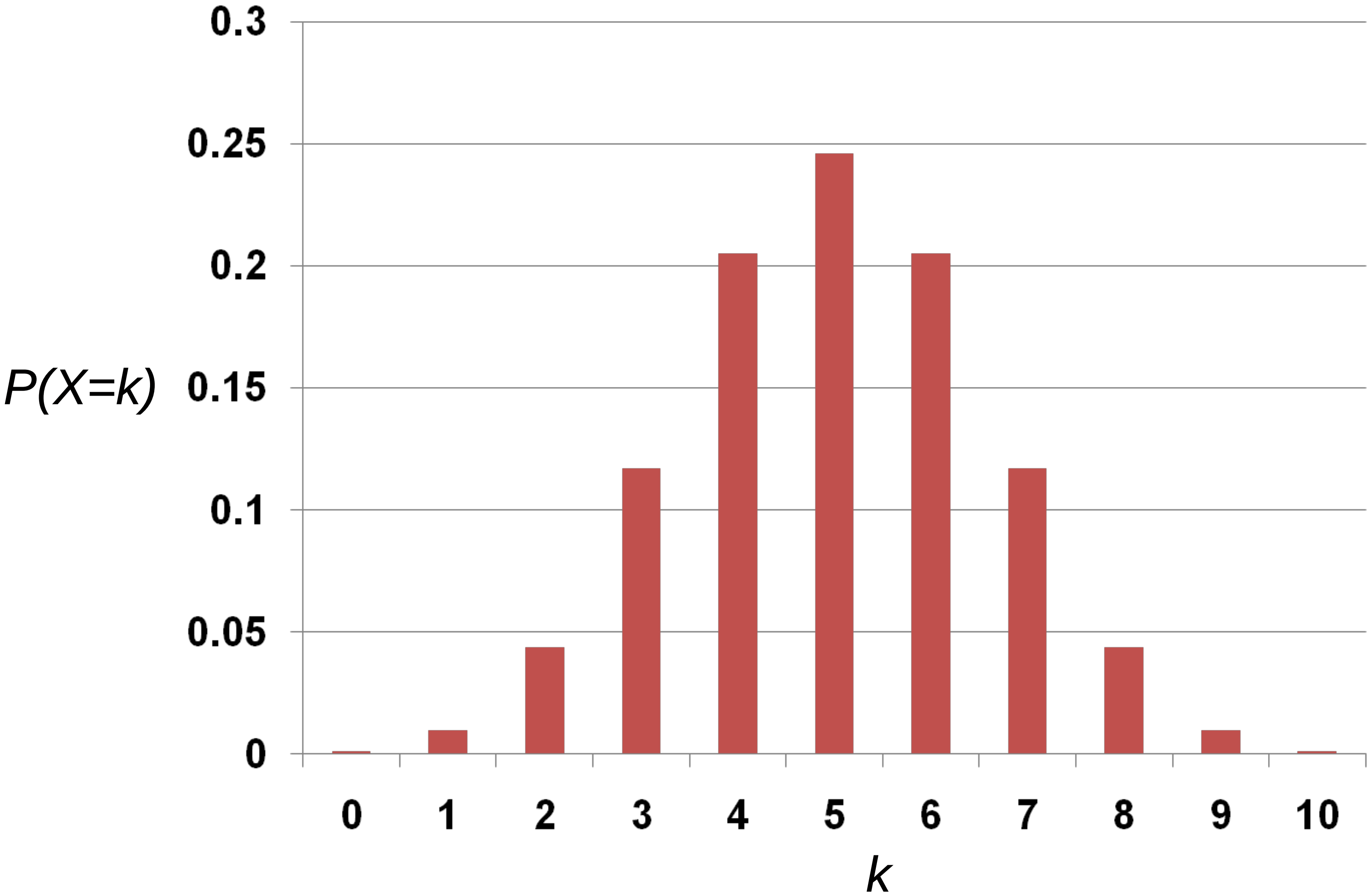
$$P(X = 0) = \binom{3}{0} p^0 (1-p)^3 = \frac{1}{8}$$

$$P(X = 1) = \binom{3}{1} p^1 (1-p)^2 = \frac{3}{8}$$

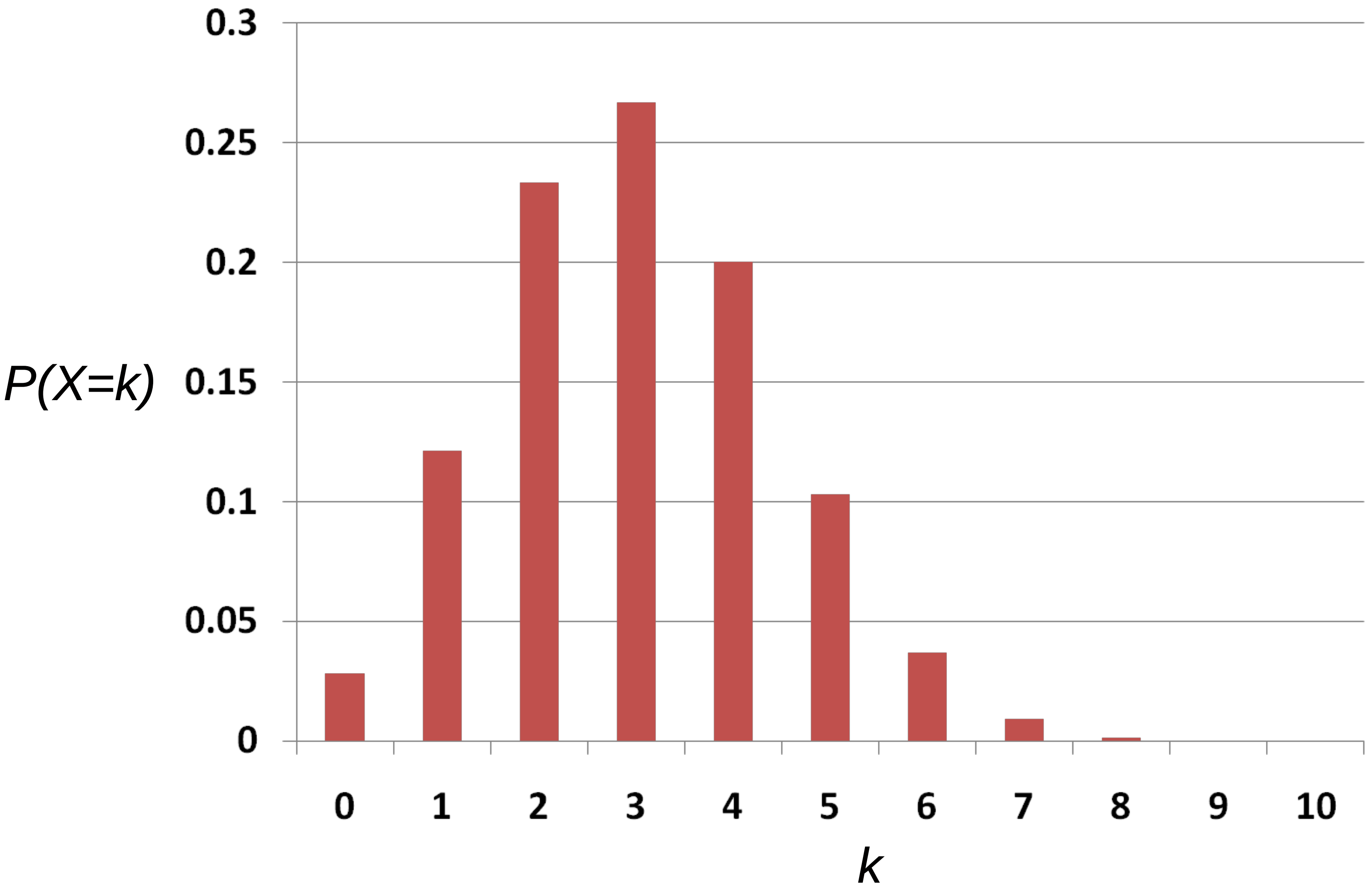
$$P(X = 2) = \binom{3}{2} p^2 (1-p)^1 = \frac{3}{8}$$

$$P(X = 3) = \binom{3}{3} p^3 (1-p)^0 = \frac{1}{8}$$

PMF for X ~ Bin(10, 0.5)



PMF for X ~ Bin(10, 0.3)



Error Correcting Codes

- Error correcting codes
  - Have original 4 bit string to send over network
  - Add 3 “parity” bits, and send 7 bits total
  - Each bit independently corrupted (flipped) in transition with probability 0.1
  - X = number of bits corrupted: $X \sim Bin(7, 0.1)$
  - But, parity bits allow us to correct at most 1 bit error
- P(a correctable message is received)?
  - $P(X = 0) + P(X = 1)$

Error Correcting Codes (cont)

- Using error correcting codes: X ~ Bin(7, 0.1)

$$P(X = 0) = \binom{7}{0}(0.1)^0(0.9)^7 \approx 0.4783$$

$$P(X = 1) = \binom{7}{1}(0.1)^1(0.9)^6 \approx 0.3720$$

  - P(X = 0) + P(X = 1) = 0.8503
- What if we didn't use error correcting codes?
  - X ~ Bin(4, 0.1)
  - P(correct message received) = P(X = 0)

$$P(X = 0) = \binom{4}{0}(0.1)^0(0.9)^4 = 0.6561$$

- Using error correction improves reliability ~30%!

Properties of Bin(n, p)

- Consider: $X \sim \text{Bin}(n, p)$
- $E[X] = np$
- $\text{Var}(X) = np(1 - p)$
- So, to compute $E[X^2]$, we have:

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

$$\begin{aligned} E[X^2] &= \text{Var}(X) + (E[X])^2 \\ &= np(1 - p) + (np)^2 \\ &= n^2p^2 - np^2 + np \end{aligned}$$

- Note: $\text{Ber}(p) = \text{Bin}(1, p)$