4. The Binomial Model (Discrete)

Compare and Contrast: The **Geometric Model** gives probabilities for the number of trials until the first success.

The **Binomial Model** gives probabilities for the number of successes in a fixed number of trials.

Example: We would use the Geometric model to count up the number of flips until the first head appears when tossing coins. We would use the Binomial model to count up the number of heads in ten flips.

Notation: The number of trials is denoted n.

The probability of success is denoted p.

Trials are independent and the number of trials is fixed.

Each trial has 2 outcomes.

The mean is denoted:

$$E[X] = \mu = np$$

The standard deviation is denoted:

$$SD[X] = \sigma = \sqrt{np(1-p)}$$

The probability of getting $X = x$ successes in n trials is denoted:

$$P(X=x) = ({}_nC_x)(p^x)(1-p)^{n-x}$$

where $${}_nC_x = \frac{n!}{x!\,(n-x)!}$$

and $$n! = n(n-1)(n-2)\cdots(2)(1)$$

So $$5! = (5)(4)(3)(2)(1) = 120$$

Example: In 2010, Zillow reported that 23.3% of U.S. homes are underwater – in other words the homeowners owe more on the mortgage than the house is worth. Assuming the same trend holds true in Elkton, city officials take a random sample of 15 homes.

a. If we are counting up the number of underwater homes, name the probability model and its parameters.

Binomial with $n = 15$, $p = 0.233$, and X = no. of underwater homes

b. Home many homes do we expect to be underwater?

$$E[X] = np = 15(0.233) = 3.495$$

c. In the sample of 15 homes, what is the chance that not a single one is underwater?

$$P(X=0) = ({}_{15}C_0)(0.233)^0(1-0.233)^{15}$$

$$= 0.0187 \qquad 1.87\% \text{ chance}$$

d. What is the chance that at most, one home is underwater?

$$P(\text{AT MOST 1 underwater}) = P(X \leq 1) =$$

$$P(X=0) + P(X=1) = 0.0187 +$$

$$({}_{15}C_1)(0.233)^1(1-0.233)^{14} = 0.0187 + 0.0852$$

$$= 0.1039$$

10.39% chance

e. On the TI calculator, there are functions **binompdf** and **binomcdf**. Use binompdf to determine the probability of getting exactly 5 homes that are underwater.

$$P(X=5) = ({}_{15}C_5)(0.233)^5(1-0.233)^{10}$$

$$= \text{binompdf}(15, 0.233, 5)$$

$$= 0.1453$$

14.53% chance

f. Use binomcdf to determine the probability of getting 5 or fewer homes underwater.

$$P(X \le 5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$$

$$= \text{binomcdf}(15, 0.233, 5) = 0.8865$$

88.65% chance at most 5 in 15 are underwater.

g. Use binomcdf to determine the probability of getting 10 or more homes underwater.

$$P(X \ge 10) = 1 - P(X \le 9) =$$

$$1 - \text{binomcdf}(15, 0.233, 9) =$$

$$0.000433$$

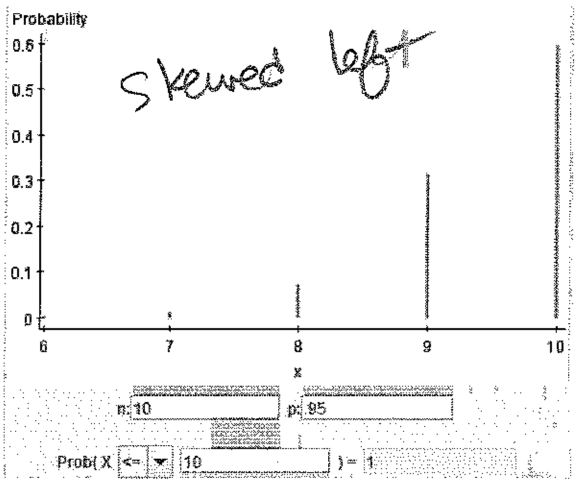
Normal Approximation to the Binomial Model

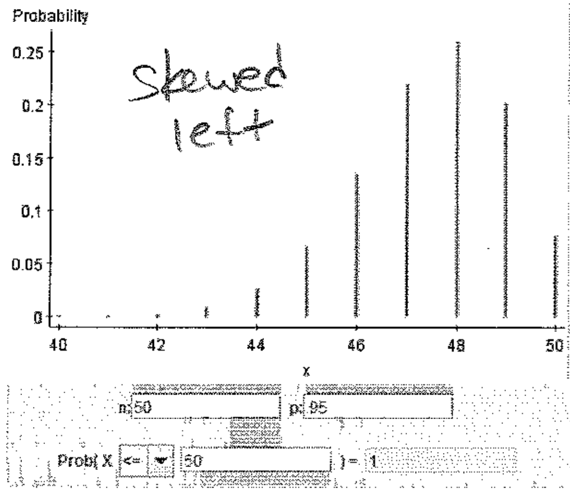
- When the number of trials gets large, the $\binom{n}{x}$ term gets very large. For instance, at Cecil College there are 2800 students. In how many ways can we form a unique committee with 10 students?

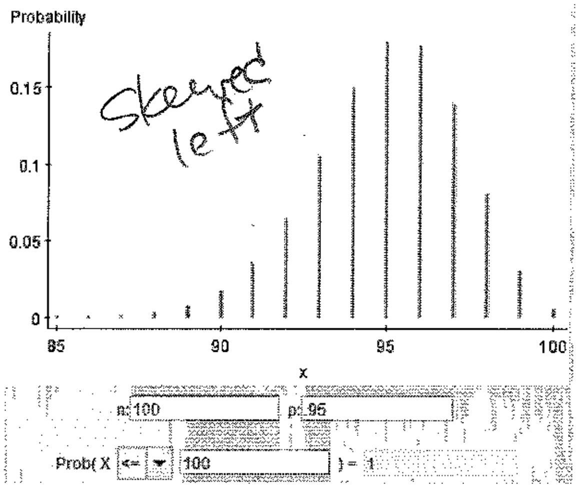
$$\binom{2800}{10} \approx 8 \times 10^{27}$$

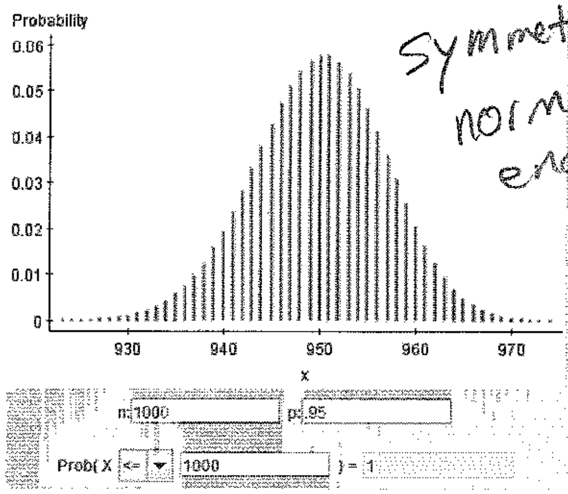
where $8 \times 10^{7} = 8{,}000{,}000$

- Suppose a Binomial Model has $p = 0.95$. We would expect a graph of the distribution to be skewed very left, with almost all the successes piled up with the successes close to the number of trials. But as the number of trials increases...









Symmetric. normal – enough.

Approximation: We can use a Normal model to approximate a binomial model, as long as we expect at least 10 successes and 10 failures. We can check:

$np \geq 10$ and nq or $n(1-p) \geq 10$

Why More Than 10 Successes / Failures (Burning Question Since Math 127)

- A Normal model extends to infinity at both ends.
- A Binomial model must have between 0 and n successes, so when we approximate with a Normal model, we must cut off the tails.
- If the center of the Normal model is far from 0 and far from *n*, no big deal (we won't chop off much of the tails).
- If the center of the Normal model is close to 0 or close to *n*, we might run into a problem, so we limit the lost area to more than 3 standard deviations.
- This means that the mean needs to be more than 3 standard deviations away from 0 and away from *n*. Check this out for the 0 end:

Require $\mu - 3\sigma > 0$

OR $\mu > 3\sigma$

For Binomial, $np > 3\sqrt{npq}$

Square $n^2p^2 > 9npq$

Simplify $np > 9q$

Since $q \leq 1$, just Require $np > 9$ and again to simplify, require $np \geq 10$

Normal Approximation

- If you are working with a Binomial model and you expect at least 10 successes and 10 failures, you can approximate the Binomial with this Normal model:

$$X \sim \text{Normal}\left(\mu = np,\ \sigma = \sqrt{np(1-p)}\right)$$

(approximately, not exactly)

Example: In an entire season of Men's League, Kupe will end up getting 1050 frames of bowling. Suppose he strikes 47% of the time. What's the probability that he gets at least 525 strikes during the course of the season?

We have $n = 1050$, $p = 0.47$

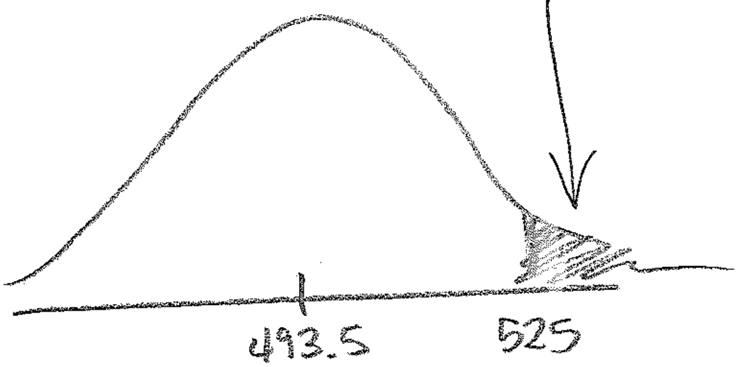
$$\underbrace{np = 493.5 \quad n(1-p) = 556.5}_{\text{OK to use Normal approximation}}$$

$\mu = np = 493.5$

$\sigma = \sqrt{np(1-p)} \approx 16.17$

$$P(X \geq 525) \approx \text{ncdf}(525, E99, 493.5, 16.17)$$

$$= 0.0257$$



$\approx$ 2.6% chance he makes at least 1/2 his strikes all season long

5. The Poisson Model (Discrete)

- The Poisson model is used to model the probability of rare events.

Example: The Ohio State University has about 50,000 students. Suppose during the fall semester, 3 students contract meningitis. If the disease is typically contracted by 1 in 100,000 people each year, is there cause for concern?

What is the expected number of cases per year at OSU?

$n = 50{,}000$

$p = \frac{1}{100{,}000}$

So $\mu = np = 50000 \times \frac{1}{100000} = \frac{1}{2}$ case

If we wanted to use a binomial model, what would the formula look like? We'd like the probability of witnessing at least 3 cases.

$$P(X \geq 3) = 1 - P(X \leq 2)$$

$$= 1 - \left[P(X=0) + P(X=1) + P(X=2)\right]$$

$$= 1 - \left[{}_{50000}C_0 \left(\frac{1}{100{,}000}\right)^0 \left(\frac{99{,}999}{100000}\right)^{50000} + {}_{50000}C_1 \left(\frac{1}{100{,}000}\right)^1 \left(\frac{99999}{100000}\right)^{49999} + {}_{50000}C_2 \left(\frac{1}{100000}\right)^2 \left(\frac{99999}{100{,}000}\right)^{49998} \right]$$

Why can't we use the Normal approximation to the Binomial model?

$np = \frac{1}{2}$ is less than 10.

Model is way too skewed to use normal approximation.

Poisson Model for Successes (A Stand Alone Model or an Approximation to the Binomial)

Require $n \geq 20$ with $p \leq 0.05$ or
$n \geq 100$ with $p \leq 0.10$

λ = mean number of successes

X = # of successes

$$\mu = E[X] = \lambda$$

$$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \sigma = \sqrt{\lambda}$$

$\lambda = 1/2$

Use Poisson to give the probability of at least three cases of meningitis.

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - P\{X=0 \text{ or } 1 \text{ or } 2\}$$

$$1 - \left[\frac{e^{-1/2}(1/2)^0}{0!} + \frac{e^{-1/2}(1/2)^1}{1!} + \frac{e^{-1/2}(1/2)^2}{2!}\right] =$$

$$1 - 0.6065 - 0.3033 - 0.0758 = 0.0144$$

~ 1.44% chance at least 3 cases in one year.

Example: In the testing of circuit boards, a certain board contains 200 diodes, and each diode has a 0.01 probability that it will fail.

a. How many diodes would we expect to fail, and what is the standard deviation?
b. Give the approximate probability that at least 4 diodes will fail on any given board.
c. A shipment of five boards are sent to a customer. How likely is it that at least four of the five boards work properly (a board works only if all diodes work).

a.) $\lambda = np = 200(0.01) = 2 = \mu$

$\sigma = \sqrt{\lambda} = \sqrt{2} = 1.41$

b.) Poisson with $\lambda = 2$

$P(X \geq 4) = 1 - P(X \leq 3) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=4)]$

$= 1 - \left[\frac{e^{-2} \cdot 2^0}{0!} + \frac{e^{-2} \cdot 2^1}{1!} + \frac{e^{-2} \cdot 2^2}{2!} + \frac{e^{-2} \cdot 2^3}{3!}\right]$

$= 1 - \underset{\text{no failures}}{0.135} - \underset{\text{1 failure}}{0.271} - \underset{\text{2 failures}}{0.271} - \underset{\text{3 failures}}{0.180} =$

$= 0.143$ or 14.3% chance at least 4 failures.

c.) $P(\text{at least 4 of 5 work}) = P(X=4) + P(X=5)$

where $X \sim \text{Binomial}(n=5, p=0.135)$

$= ({}_5C_4)(0.135)^4(1-0.135)^1 + ({}_5C_5)(0.135)^5(1-0.135)^0$

$= 0.0014 + 0.000045 = 0.0014$

Roughly 1.5 in 1000!

Example: A customer service center receives 5 calls per every 3 minutes, on average. Use the Poisson model to answer the following questions.

a. What is the expected number of calls during the next minute?
b. What is the probability there will be no calls during the next minute?
c. What is the probability there will be at least two calls?

a.) If X = number of calls in a minute,
$E[X] = \frac{5}{3}$ and $X \sim$ Poisson model, $\lambda = \frac{5}{3}$

b.) $P(X=0) = \frac{e^{-5/3}(5/3)^0}{0!} = e^{-5/3} = 0.189$

18.9% chance of no calls in next minute.

c.) $P(X \geq 2) = P(X=2) + P(X=3) + \cdots$
$= 1 - [P(X=0) + P(X=1)] =$
$1 - 0.189 - \left[\frac{e^{-5/3}(5/3)^1}{1!}\right] =$
$1 - 0.189 - 0.315 \approx 0.496$

49.6% chance at least two calls in the next minute.

Example: A typesetter for a magazine makes one error per every 500 words of typeset. A typical page has 300 pages. What is the probability that there will be no more than two errors in a five-page article?

a. Set this up as a binomial problem, but don't bother to solve it because of the tediousness.
b. Now set this up as a Poisson problem and do, in fact, solve it.

$P(\text{Error}) = \frac{1}{500}$ and $P(\text{not error}) = \frac{499}{500}$ for any word.

5 pages are 1500 words so . . .

a.) X = no. of errors is Binomial $n = 1500$, $p = \frac{1}{500}$, $1-p = \frac{499}{500}$

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= ({}_{1500}C_0)\left(\frac{1}{500}\right)^0\left(\frac{499}{500}\right)^{1500} +$$
$$({}_{1500}C_1)\left(\frac{1}{500}\right)^1\left(\frac{499}{500}\right)^{1499} +$$
$$({}_{1500}C_2)\left(\frac{1}{500}\right)^2\left(\frac{499}{500}\right)^{1498}$$

---

b.) $X \sim$ Poisson with $\lambda = E[X] = np = 1500\left(\frac{1}{500}\right) = 3$

$$P(X \le 2) = \left[\frac{e^{-3}(3)^0}{0!}\right] + \left[\frac{e^{-3}(3)^1}{1!}\right] + \left[\frac{e^{-3}(3)^2}{2!}\right]$$

$$= 0.0498 + 0.1494 + 0.2240$$

$$\approx 0.4232$$

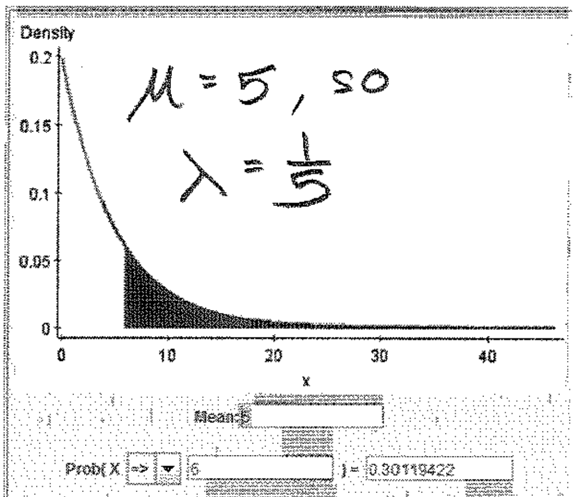
6. The Exponential Model (Continuous)

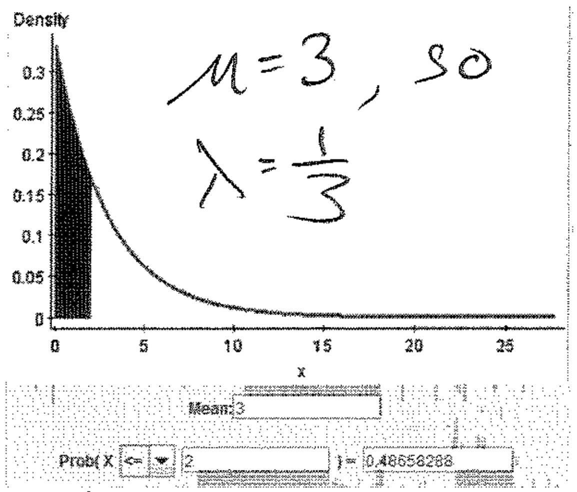
- The Poisson model counts up the number of "rare" events in a certain time frame.
- The exponential model, with same parameter λ, is used to model the time between those "rare" events.
- The exponential model is continuous because the time between events have an infinite number of possibilities.
- If λ increases in the Poisson model, we expect more events to occur in a given interval.
- Thus, if λ increases in the exponential model, we expect the time between events to decrease.
- The mean or expected value of an exponential model is: $E[X] = \mu = \frac{1}{\lambda}$.

The probability model / function is given by:

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0 \text{ and } \lambda > 0$$

To find probabilities for the time between events, we would need **calculus** to integrate and find the area under the exponential model (just like Uniform and Normal).





$f(x) = \frac{1}{5} e^{-\frac{1}{5}x}$, $x \geq 0$

$f(x) = \frac{1}{3} e^{-1/3 x}$, $x \geq 0$

There is a function to determine the probability that x lies between any two values, say s and t:

$$P(s \le X \le t) = e^{-\lambda s} - e^{-\lambda t}$$

There is also a function (by setting $s = 0$), to find the probability that the waiting time for the next event will be less than t:

$$P(X \le t) = P(0 \le X \le t) = e^{-\lambda(0)} - e^{-\lambda t}$$
$$= e^{0} - e^{-\lambda t}$$
$$= 1 - e^{-\lambda t}$$

Example: A customer service center receives 5 calls per every 3 minutes, on average. Recall that per minute, the mean for the Poisson model was $\lambda = 5/3$.

a. Give the probability model for the time between calls at the call center.

$$f(x) = \frac{5}{3} e^{-5/3 x}, \quad x \ge 0$$

b. Starting now, what is the probability next call arrives in under two minutes?

$$P(X \le 2) = 1 - e^{-\lambda x}$$
$$= 1 - e^{-5/3(2)} = 1 - e^{-10/3}$$
$$\approx 0.9643$$

96.43% chance

c. Give the probability that the next call arrives between 30 seconds and one minute.

$$P(0.5 \le X \le 1) =$$

$$e^{-5/3(0.5)} - e^{-5/3(1)} =$$

$$e^{-5/6} - e^{-5/3} \approx 0.2457$$

d. Give the probability that the next call isn't for at least five minutes.

$$P(X \ge 5) = 1 - P(X \le 5) =$$

$$1 - \left(1 - e^{-5/3(5)}\right) =$$

$$1 - 1 + e^{-25/3} = e^{-25/3}$$

$$\approx 0.00024$$

e. On average, how long do we wait in between calls?

$$E[X] = \mu = \frac{1}{\lambda}$$

$$= \frac{1}{5/3} = \frac{3}{5} \text{ of a minute}$$

$\sim$ 36 seconds

Example: An Amazon.com sales manager notes that during the late evening hours, about 10 people per minute actually buys something from the website. Purchases appear to be independent of each other.

a. What model could we use to model the time between purchases? What is the parameter?
b. What is the mean of the model? Interpret its value.
c. What is the probability function to model time between purchases?
d. What is the probability that the time until the next purchase happens in under 5 seconds?
e. What is the probability that the time until the next purchase exceeds 30 seconds?

a) Exponential model. $\lambda = 10$.

b) $\mu = \frac{1}{\lambda} = \frac{1}{10}$. We expect a purchase every 1/10 of a minute (6 seconds)

c) $f(x) = 10e^{-10x}, \quad x \geq 0$

d) $P(X \leq 5 \text{ seconds}) = P(X \leq \frac{5}{60} \text{ or } \frac{1}{12}) =$

$1 - e^{-10(\frac{1}{12})} = 1 - e^{-10/12} = 0.565$

56.5% chance.

e) $P(X \geq 30 \text{ seconds}) = P(X \geq \frac{1}{2} \text{ min})$

$= 1 - P(X \leq \frac{1}{2} \text{ minute}) =$

$1 - (1 - e^{-10(\frac{1}{2})}) =$

$1 - 1 + e^{-5} = e^{-5} = 0.0067$

0.67% chance