Probability Models: from Discrete to Continuous

Week 5

Prof. Brand

ADM2303

February 5, 2012

ADMINISTRATIVE ISSUES

1. Assignment-1 (grading ongoing)
2. Assignment-2 (available)
3. Quiz
   - Next week
   - Sunday, March 4th (10-11:30am)
   - Covers (first 5 weeks IN SYLLABUS)
   - Formula's provided

Last Week

1. Variance (standard deviation, coefficient of variation)
2. Rules for operating on (independent) random variables
3. Standard probability models
   - Bernoulli trial
   - Binomial
   - Geometric

This Week

1. Review: Standard probability models
   - Binomial
   - Geometric
2. Binomial examples
3. Binomial/Geometric examples
4. Poisson probability model
5. Continuous probability models
   - Uniform
   - Normal

Summary of Geometric/Binomial models

Looked at probability models deriving from simple "Bernoulli Trials." So, what are 'Bernoulli Trials?"

- Sequence of independent trials
- Outcome of trial can be classified as either "success" or "failure" (or numerically as, 1 or 0).
- The probability of success p remains constant from trial to trial

Bernoulli Trials: Probability Models

If Bernoulli trial assumptions apply then:

1. X, the trial number at which the first success occurs, has a **geometric distribution**:

$$p_X(x) = (1-p)^{(x-1)}p \text{ where } x = 1, 2, \ldots$$

EV (mean) $1/p$
variance $(1-p)/p^2$

2. X, the total number of successes in n trials has a **binomial distribution**:

$$p_X(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{(n-x)} \text{ where } x = 0, 1, 2, \ldots, n$$

EV (mean) np
variance $np(1-p)$

True/False Quiz

Consider a T/F format for a quiz with 20 questions.

- Assume answer on each T/F question selected randomly (coin toss),
- How many do you expect to get correct?
- What is the probability of passing?

True/False Quiz

Probability correct answer $p = 0.5$.
Binomial probability model.

$$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \text{ where } x = 0, 1, 2, \ldots, n$$

$X = x$	$\Pr[X = x]$
0	0.5^{20}
1	$20 \times (0.5)^1 \times 0.5^{19}$
2	$\frac{20 \times 19}{2} \times (0.5)^2 \times 0.5^{18}$
⋮	⋮
18	$\frac{20 \times 19}{2} \times (0.5)^{18} \times 0.5^2$
19	$20 \times (0.5)^{19} \times 0.5^1$
20	0.5^{20}

True/False Quiz: Solution

$X = x$	$\Pr[X = x]$	
0	0.5^{20}	9.54e-7
1	$20 \times (0.5)^1 \times 0.5^{19}$	1.91e-5
2	$\frac{20 \times 19}{2} \times (0.5)^2 \times 0.5^{18}$	1.81e-4
⋮	⋮	⋮
18	$\frac{20 \times 19}{2} \times (0.5)^{18} \times 0.5^2$	1.81e-4
19	$20 \times (0.5)^{19} \times 0.5^1$	1.91e-5
20	0.5^{20}	9.54e-7

9.54e-7	1.91e-5	1.81e-4	1.09e-3	4.62e-3	1.48e-2
3.70e-2	7.39e-2	1.20e-1	1.60e-1	**1.76e-1**	**1.60e-1**
1.20e-1	**7.39e-2**	**3.70e-2**	**1.48e-2**	**4.62e-3**	**1.09e-3**
1.81e-4	**1.91e-5**	**9.54e-7**			

Note symmetry is peculiar to setting where $p = (1 - p)$.

True/False Quiz: Solution

What is the probability that they will pass the quiz?

$$\Pr[X \geq 10] = \sum_{x=10}^{20} \frac{20!}{x!(20-x)!} p^{20} \text{ where } x = 0, 1, \ldots, 20$$

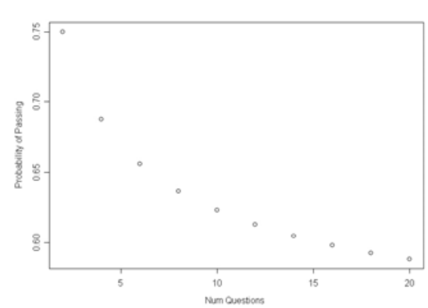
$$\Pr[X \geq 10] = 0.59$$

Why is this answer not 50%?

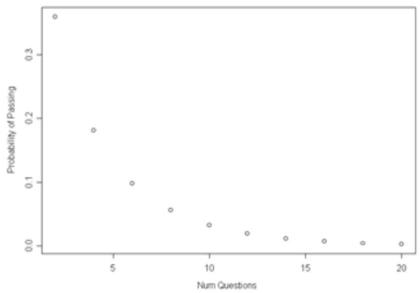
True/False Quiz: General

- How would the probability of passing a test (assuming random choice) change with the number of T/F questions (number of trials)?
- Try for $n = 2, 4, 6, \ldots 20$
- Graph probability of passing as function of n

True/False Quiz: General



Multiple-Choice quiz: $\Pr[\text{Correct}] = 0.20$



Web-browser

At the early stage of development of a web-browser one in 20 randomly chosen web-sites fails to "render" properly.

- What is the probability that the first failed rendering (a success) is the fourth web-site the browser randomly visits?
- $p = 0.95$

Consider geometric model (or use just plain old intuition)

$$\Pr[Y = 4] = (1 - 0.95) \times (1 - 0.95) \times (1 - 0.95) \times 0.95$$

Web-browser development: 6 Trials

Keep $p = 0.05$ (namely the probability failed rendering).
Question:
What is the probability that there is exactly 1 failed rendering among the first 6 trials.

- What model would you use?
- Model: Binomial.

Web-browser development: 6 Trials

$$p_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \text{ where } x = 0, 1, 2, \ldots, n$$

$$p_X(X = 1; n = 6) = \frac{6!}{1!(6-1)!} (0.05)^1 (1 - 0.05)^5$$

Universal Donor

- 20 donors come to the blood drive with two outcomes
  - Success: That is O-negative (universal donor)
  - Failure: other blood type
- $p = 0.06$
- Assume that you are sampling with replacement, or equivalently that the donors come from a large population.

Universal Donor: Analogy to Urn

- 20 draws from an urn with
  - 6 Red balls
  - 94 white balls
- Sampling with replacement
- Sampling without replacement

Universal Donor

- We are interested in the number of universal donors in our 20 donor drive.
- Denote this number X (namely the number of O-negative donors among $n = 20$).
- From last lecture we know that X follows a binomial probability model.

Questions:

1. What is the expected value for X?
2. What is the standard deviation for X?

Universal Donor

Solution (EV and Var for Binomial):

EV

$$= np = 20 \times 0.06 = 1.2$$

Var

$$= np(1-p) = 20 \times 0.06 \times (1-0.06) = 1.13$$

What about standard deviation?

Universal Donor

Question: Calculate the Probability of 2 or 3 universal donors ($n = 20$).

$$\begin{aligned}
\Pr[X = 2 \cup X = 3] &= \Pr[X = 2 \text{ OR } X = 3] \\
&= \Pr[X = 2] + \Pr[X = 3] \\
&= \frac{20!}{2!(20-2)!} p^2 (1-p)^{(20-2)} + \frac{20!}{3!(20-3)!} p^3 (1-p)^{(20-3)} \\
&= 0.2246 + 0.086 \\
&= 0.3106
\end{aligned}$$

n Choose x

How many ways can we "choose" x successes out of n tries?

$$\begin{aligned}
C_x^n &= \binom{n}{x} \\
&= \frac{n!}{x!(n-x)!}
\end{aligned}$$

Computation of Binomial can be Difficult

When n large computing the combinatoric $\frac{n!}{x!(n-x)!}$ can be intractable (*even with todays computers*).

- In some cases the Normal model can help.
- In other cases the Normal model will not work, but another approximation — **Poisson model** — will help.

What Approximation to Use for Binomial?

Normal — If expected number of successes and failures sufficiently high ($np > 10$ and $n(1-p) > 10$) then normal provides adequate approximation.

Poisson — If expected number of successes/failures is small ($np < 10$) then the Poisson model can serve as an approximation.

The Model

If $np < 10$, or if $n(1-p) < 10$, and n unwieldy in size then Poisson can be used.
Notation

$\lambda =$ mean number of successes (np)

$X =$ the number of successes

Then the Poisson probability model is defined as:

$$P(X = x) = \frac{\exp(-\lambda)\lambda^x}{x!}$$

EV (mean) λ

variance λ

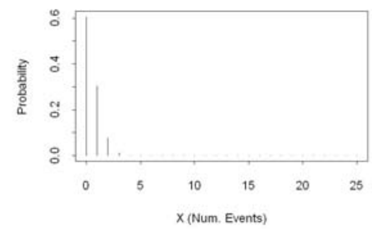
sd $\sqrt{\lambda}$

Prove that coefficient of variation $\nu_X = \frac{1}{\sqrt{\lambda}}$

The Model

For $\lambda = 0.5$

$X = x$	$\Pr[X \mid \textbf{Poisson}]$
0	$\frac{\exp(-\lambda)\lambda^0}{0!}$
1	$\frac{\exp(-\lambda)\lambda^1}{1!}$
2	$\frac{\exp(-\lambda)\lambda^2}{2!}$
3	$\frac{\exp(-\lambda)\lambda^3}{3!}$
4	$\frac{\exp(-\lambda)\lambda^4}{4!}$
$\vdots$	$\vdots$
n	$\frac{\exp(-\lambda)\lambda^n}{n!}$



Example: Blues Fest

You run a Food-booth (the Chicken-Shack) at the Blues festival. You sell chicken sandwiches, exclusively. Assume that the booth has 20000 sales and that 1 in every 10000 chicken sandwiches is contaminated with salmonella and will cause intestinal illness in the person consuming it.

- What is the mean number of cases of intestinal illness to expect?
- Solution,

$$\lambda = 20000 \times 1/10000 = 2 \text{ cases}$$

Therefore,

$$E[X] = 2 \text{ cases}$$

Example: Blues Fest. Again

- What is the probability of 3 or more cases?
- Simpler question: What is the probability of 3 cases
- Recall (from last time), $np = 2$. Thus $\lambda = 2$.

$$\begin{aligned} P(X = 3) &= \frac{\exp(-2)2^x}{x!} \\ &= \frac{\exp(-2)2^3}{3!} \\ &= 0.180 \end{aligned}$$

Example: Blues Fest. Again

- What is the probability of 3 or more cases?
- As a first step consider the complement event (probability of 0, 1, or 2 cases),

$X = x$	$\Pr[X \mid \textbf{Poisson}]$	
0	$\frac{\exp(-2)2^0}{0!} =$	0.135
1	$\frac{\exp(-2)2^1}{1!} =$	0.271
2	$\frac{\exp(-2)2^2}{2!} =$	0.271
$0 \cup 1 \cup 2$	$\sum_{x=0}^{2} \frac{\exp(-\lambda)\lambda^x}{x!} =$	0.677

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0 \cup X = 1 \cup X = 2) \\ &= 1 - 0.677 \\ &= 0.323 \end{aligned}$$

Example: Defects

Consider case involving

- 1000 trials
- $\Pr[\text{defect}] = 0.001$
- Thus $np = 1000 \times 1/1000 = 1$

$\lambda =$ mean number of successes($np = 1$)

$X = x$	$\Pr[X \mid \textbf{Poisson}]$	
0	$\frac{\exp(-1)1^0}{0!} =$	0.3679
1	$\frac{\exp(-1)1^1}{1!} =$	0.3679
2	$\frac{\exp(-1)1^2}{2!} =$	0.1839
3	$\frac{\exp(-1)1^3}{3!} =$	0.0613
4	$\frac{\exp(-1)1^4}{4!} =$	0.0153
$\vdots$	$\vdots$	$\vdots$
1000	$\frac{\exp(-1)1^{1000}}{1000!} =$	0

λ as a Rate

In a Poisson Probability model λ is often thought of as the average number of "successes" per unit time (success/time) with units such as (successes/day).

More generally λ can represent the average number of successes per unit:

- area,
- mass,
- volume,
- population size, and etc ....

Let us use time as the example but keep in mind that the same concepts could be applied to the area, mass, or volume units.

Re-expressing the Poisson Model

One can alter Poisson model to allow an arbitrary time interval: replacing λ with λt.
This allows the Poisson model to be computed given any fixed time interval t as:

$$P(X = x) = \frac{\exp(-\lambda t)(\lambda t)^x}{x!}$$

In this case the mean and variance will be a function of the fixed unit of time, namely

EV (mean) λt

variance λt

sd $\sqrt{\lambda t}$

The above reflects what is known as a *Poisson Process*.

Conditions for Poisson Process

Assumptions:

1. We know the average rate, λ, which denotes the average number of successes per unit (length, time, volume, or etc) of interest;
2. The probability of success is the same for each **equally sized unit**;
3. What happens in one unit has no influence upon any other nonoverlapping unit;
4. We can imagine dividing our unit into tiny subunits;
   - The probability of more than one success occurring in a tiny subunit is negligible; and
   - the probability of an event in a short interval (Δt) is approximately $\lambda \Delta t$ for any t.

Left-hand Turn Lane

Consider the number of cars (X) arriving in a left-hand turn lane. Let X be characterized by Poisson PMF, with $\lambda = 1$ arrival/30 sec.
Consider at traffic-light with a 2 minute cycle.
(a) What is the mean number of cars that will arrive in a cycle?

Poisson Example: Traffic light (2)

Solution Define 30 seconds as a unit time interval and note that there are 4 such intervals per traffic light cycle.
The expected value of arrivals is:

$$\lambda t = \frac{1}{30 \text{ sec}} \times \frac{\overbrace{(4 \times 30 \text{ sec})}^{2 \text{ minutes}}}{\text{traffic light cycle}} = \frac{4 \text{ cars}}{\text{traffic light cycle}}$$

(b) What is the probability that less than 6 cars will arrive in one "traffic light cycle" (2 minutes) ?

$$P[X < 6] = \sum_{x=0}^{5} \frac{\exp(-4)(4)^x}{x!}$$

(c) If the capacity of the left hand turn lane is 5 cars, what is the probability that it will be exceeded ?

Example: Airport Taxi

- Demand (X) for Taxis at pick-up zone within local airport;
  - X is random variable described by average rate of $\lambda = 6$/hour
  - What if they allocate 2 cabs in a 30 minute period.
  - Assume a 30 minute period.
- What is probability of unmet demand?

Airport Taxi: Solution

- $\lambda = 3$ customers per half-hour (30 minutes)
- Let X = number of customers in 30 minutes (i.e., Demand)
- Interested in $\Pr[\text{unmet}]$ where "unmet" refers to unmet demand

The following tabulates all outcomes (X) consistent with demand being **met**, & calculates their respective probabilities, & sums them to get $\Pr[\text{met}]$.

$X = x$	$\Pr[X \mid \textbf{Poisson}] =$	
0	$\frac{\exp(-3)3^0}{0!} =$	0.0498
1	$\frac{\exp(-3)3^1}{1!} =$	0.1494
2	$\frac{\exp(-3)3^2}{2!} =$	0.2240
$\Pr[\text{met}]$	$\sum_{i=0}^{2} \frac{\exp(-3)3^x}{x!} =$	0.4232

$$\begin{aligned}\Pr[\text{unmet}] &= 1 - \Pr[\text{met}] \\ &= 1 - 0.4232 = 0.5768\end{aligned}$$

Other Poisson Type Problems

- Defects per 30 yard bolt of fabric;
- Leaks per square meter of foundation;
- Cancer cases per 10,000 people; and
- Product returns ($\lambda = 2/\text{month}$) .

Poisson Model

- Poisson often deals with rare events.
- So that while one can imagine counting the number of successes from a Poisson process, it is often difficult to imagine counting the number of failures.
- In one class of application (the general model), the Poisson describes a process that extends over time, over space, over volume, over mass, or over items.

You Should Be Able to . . .

- . . . recognize the need for . . .
  - Geometric or
  - Binomial or
  - Poisson (as approximation or as general model)
- . . . apply the above listed models, whether it be to . . .
  - calculate the probability of an event or
  - summarize the model (determining its expected value, variance, or . . .).