Recap

Counting Methods

1. With Replacement Outcomes $n = \prod_{i=1}^{I} n_i$ for I experiments

2. Without Replacement
   w/regard to order for r objects chosen from n

   $\frac{n!}{(n-r)!} = P_{n,r}$ permutation

   w/o regard to order "n choose r" $\frac{n!}{(n-r)!\,r!} = \binom{n}{r}$ binomial coefficient

Conditional Probability

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

"Probability of A given B"

A- pneumonia
B- fever $T > 101$

Intuition: Shrink the sample space from Ω to B because the event B has occurred. It has given information

~~Bayes' Rule~~

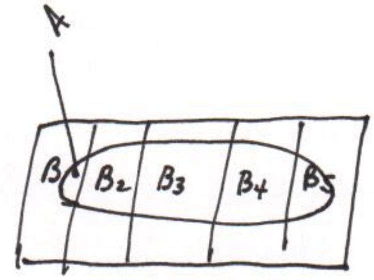
Multiplication Rule

$$Pr(A|B)\,Pr(B) = Pr(B|A)\,Pr(A) = Pr(A \cap B)$$

~~Bayes' Rule~~

Law of Total Probability

$\Omega = B = \bigcup_{i=1}^{n} B_i$ B_i disjoint



$Pr(B) = Pr(\bigcup_{i=1}^{n} B_i) = \sum_{i=1}^{n} Pr(B_i)$ 3rd Axiom

$Pr(A) = Pr(A \cap B) = \sum_{i=1}^{n} Pr(B_i \cap A)$ disjoint

$$\boxed{Pr(A) = \sum_{i=1}^{n} Pr(B_i)\,Pr(A|B_i)}$$ conditional probability

Bayes' Rule

Probability of an MI given chest pain (A)

$$Pr(B_2 \mid A) = \frac{Pr(A \mid B_2)\, Pr(B_2)}{\sum_{i=1}^{n} Pr(B_i) \cdot Pr(A \mid B_i)}$$

Bayes' Rule and Screening Test

Ex 1.E Test for Elevated Cholesterol

D^+ elevated cholesterol

D^- normal cholesterol

$Pr(D^+)$ prevalence or (incidence if per unit time)

probability that someone chosen from a specific cohort has an increased cholesterol, say African American men over 50.

How does a test perform?

D^+ Test $Pr(+ \mid D^+)$ sensitivity

$Pr(- \mid D^+)$ false negative

D^- $Pr(- \mid D^-)$ specificity

$Pr(+ \mid D^-)$ false positive

Errors

Would like to have a high sensitivity and a high specificity. Impossible in practice. A test always positive $\Rightarrow$ sensitivity of 1 but not specific.

~~Diagnosis~~ A test " negative $\Rightarrow$ specificity of 1 but not sensitive

$\Rightarrow$ Trade off

Diagnosis

$Pr(D^+ \mid +)$ predictive value positive

$Pr(D^- \mid -)$ predictive value negative

All of these terms are linked using Bayes' Rule

Predictive Value Positive 3/14

$$Pr(D^+|+) = \frac{Pr(D^+ \cap +)}{Pr(+)} = \frac{Pr(D^+)\,Pr(+|D^+)}{Pr(+)}$$

Being Pedantic

$$\{+\} = \{+ \cap D^+\} \cup \{+ \cap D^-\}$$ ← event test is positive disjoint

$$Pr(+) = Pr(+ \cap D^+) + Pr(+ \cap D^-)$$

$$= \underbrace{Pr(D^+)}_{\text{prevalence}}\,\underbrace{Pr(+|D^+)}_{\text{sensitivity}} + \overbrace{Pr(D^-)}^{\equiv 1 - Pr(D^+)}\,\underbrace{Pr(+|D^-)}_{\text{false positive}}$$

Hence,

$$Pr(D^+|+) = \frac{Pr(D^+)\,Pr(+|D^+)}{Pr(D^+)\,Pr(+|D^+) + Pr(D^-)\,Pr(+|D^-)}$$

Exercise: Work out the corresponding relationship for predictive value negative.
$Pr(D^-|-)$

Will have a problem on Homework 2.

Independence (Strong Condition)

Two events E_1 and E_2 are independent if

$$Pr(E_1 \cap E_2) = Pr(E_1)Pr(E_2)$$

Key idea:
Information about ~~one~~ one tells you nothing about the other

This means using the Multiplication Rule

$$Pr(E_2 | E_1) = Pr(E_1 \cap E_2)/Pr(E_1) = \frac{Pr(E_2)\cdot Pr(E_1)}{Pr(E_1)} = Pr(E_2)$$

In general, if $E_1, E_2, \ldots, E_n$ are independent events then

$$Pr(E_1 \cap E_2 \ldots \cap E_n) = \prod_{i=1}^{n} Pr(E_i)$$

Ex. 1.1 (cont'd). Three trials, performance on each trial is assume independent (Is this reasonable?). Pr(correct response) = $\frac{1}{4}$ there are 4 possible choices. Then

$$Pr(E_1 \cap E_2 \cap E_3) = \prod_{i=1}^{3} Pr(E_i) = \frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4} = \frac{1}{64}$$

We will make this assumption a lot.

Lecture 2 Discrete Probability Models

5/14

Concept of a Random Variable

Discrete Probability Models
- Bernoulli
- Binomial
- Poisson
- Discrete Counting Probability

Big Picture

Data

Discrete
1. binary response: polls, yes-no questions batting
2. counts: cases of a disease
3. neural spiking activity: action potentials

Continuous
- EEG
- local field potentials
- heights
- stock prices

Probability Models

Discrete Models
1. Bernoulli
2. Poison
3. Point processes
4. Binomial

Continuous Models
- Gaussian
- Exponential
- Gamma
- Beta
- Inverse Gaussian

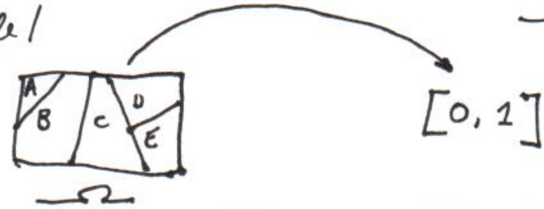
Things-to-Know
- i) Shape of the model: pdf CDF
- ii) Location: mean, mode, median
- iii) ~~Variance~~ Spread: variance, sd, range
- iV) Asymptotic properties (usually sample becomes large)
- v) WHEN DO YOU USE IT

Def'n 2.1 A random variable (RV) is a real-value function on an outcome space into the real line or R^n. The probability law on the outcome space $\Rightarrow$ probability model 6/14

$\Omega, \mathcal{F}, P$

Ex. 2.A.

Define the R.V X on Ω



$$X(\omega) = \begin{cases} 1 & \omega = A, B \\ 2 & \omega = C \\ 3 & \omega = D, E \end{cases}$$

$Pr(1) = Pr(A \cup B)$

$Pr(2) = Pr(C)$

$Pr(3) = Pr(D \cup E)$

Write Table 2.1 on B.B. before start of class

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Ex 2.0. Each outcome has Probability $\frac{1}{36}$

Define $X(\omega) = \omega_1 + \omega_2$

ω_1 – value on 1st roll

ω_2 – value on 2nd roll

Find $Pr(X(\omega))$.

- The events occur on the outcome space
- X takes on a value from 2 to 12 depending on the outcomes
- diagonals of the table define events w/ the same value of X
- This is why when playing craps in Las Vegas the casino wins if you roll a 7 (CRAPS!)

$Pr(X=2) = \frac{1}{36}$

$Pr(X=3) = \frac{2}{36}$

$Pr(X=4) = \frac{3}{36}$

$Pr(X=5) = \frac{4}{36}$

$Pr(X=6) = \frac{5}{36}$

$Pr(X=7) = \frac{6}{36}$

$Pr(X=8) = \frac{5}{36}$

$Pr(X=9) = \frac{4}{36}$

$Pr(X=10) = \frac{3}{36}$

$Pr(X=11) = \frac{2}{36}$

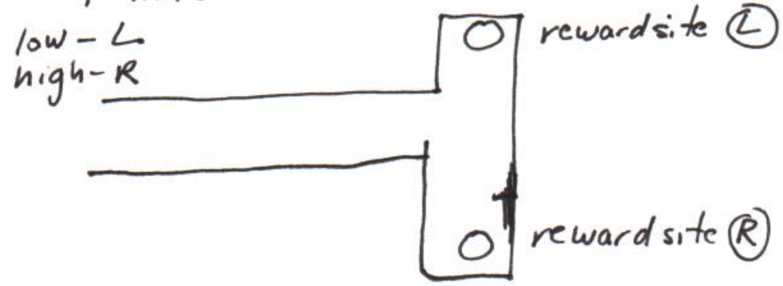
$Pr(X=12) = \frac{1}{36}$

Ex 2.1 Learning Experiment (Any binary outcome)
Graybiel Lab

hit - no hit
correct incorrect
Coakley Baker

7/14

T-maze



Previous day: 40 trials
22 correct
18 incorrect

What is the probability of a correct response today?

p - prob of a correct response

$1-p$ prob of an incorrect response

define $X = \begin{cases} 1 & \text{correct} \\ 0 & \text{if incorrect} \end{cases}$ R.V.

$Pr(X=0) = p \qquad Pr(X=1) = 1-p$

or

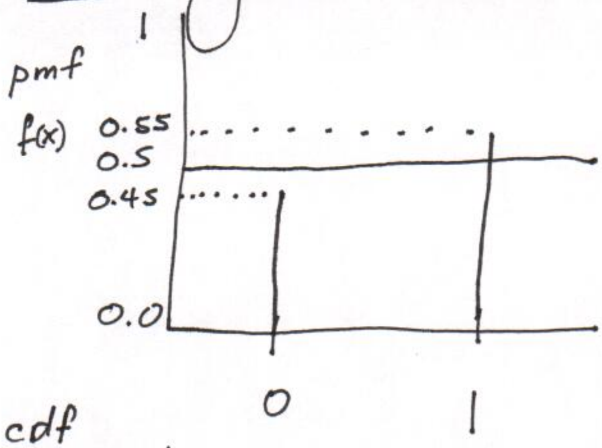
$Pr(X=x) = Pr(x=0 \text{ or } x=1) = p^x(1-p)^{1-x}$

Explain the exponents here $x=1 \Rightarrow p$, $x=0 \Rightarrow 1-p$

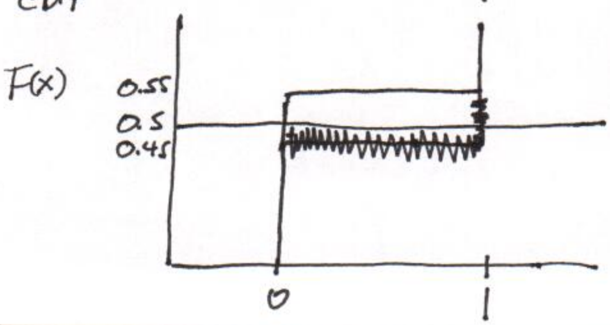
Bernoulli Probability Model (Simplest probability model)

Probability mass function (pmf): defines the probability of each value of X.

$$\sum_{x=0}^{1} Pr(X=x) = (1-p)+p = 1$$ sums to 1



cumulative distribution function (cdf)



$F(x) = Pr(X \le x)$

$F(0) = Pr(X \le 0) = Pr(X=0) = 1-p$

$F(1) = Pr(X \le 1) = Pr(X=0) + Pr(X=1) = 1-p+p = 1$

$F(x) \equiv$ area under pmf upto and including x

Ex. 2.1 $\hat{p}$ (p hat) $= \frac{22}{40} = 0.55$ an estimate of p previous day 8/14

or

$p = 0.5$ likely to perform at chance

Def'ns

mean — average

$$\mu = E(X) = \sum_{x=-\infty}^{\infty} x\, f(x)$$

variance — defines the spread

$$\sigma^2 = \sum_{x=-\infty}^{\infty} (x - E(x))^2 f(x)$$

$$= E[x - E(x)]^2 = E[x-\mu]^2 = EX^2 - \mu^2$$

Bernoulli

$$\mu = E(X) = \sum_{x=0}^{1} 0 \cdot (1-p) + 1 \cdot p = p$$

$$\sigma^2 = E(X^2) - \mu^2 = \sum_{x=0}^{1} x^2 - p^2 = 0^2(1-p) + 1^2 p - p^2$$

$$= p - p^2 = p(1-p)$$

Return to discuss interpretation after we do binomial model

Binomial Model (40 at bats for David Ortiz)

Ex. 2.1 Assume n-trials $n = 40$ p probability of correct ~~success~~

trials are independent

k- correct

Recall: $E_1, \ldots, E_n$ are independent events

$$Pr\left(\bigcap_{i=1}^{n} E_i\right) = \prod_{i=1}^{n} Pr(E_i)$$

$E_1, \ldots, E_n$ disjoint events

$$Pr\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} Pr(E_i)$$

Let's analyze 1st a learning experiment with 3 trials

Events C - correct $\qquad Pr(C) = Pr(X=1) = p$

I - incorrect $\qquad Pr(I) = Pr(X=0) = 1-p$

Suppose we get C_1, C_2, I_3 ← trial

$Pr(C_1 \cap C_2 \cap I_3) =$ [redacted] $Pr(C_1)Pr(C_2)Pr(I_3) = p^2(1-p)$

Make a Table on Blackboard (Tell Students)

(number of correct response)

RV X	Outcomes (independent)	No# Outcomes	Probability
0	$I_1\ I_2\ I_3$	$1 \quad \binom{3}{0}$ 3 attempts 0 correct	$(1-p)^3$
1	$C_1\ I_2\ I_3$ $I_1\ C_2\ I_3$ $I_1\ I_2\ C_3$	$3 \quad \binom{3}{1}$	$\binom{3}{1} p(1-p)^2$
2	$C_1\ C_2\ I_3$ $C_1\ I_2\ C_3$ $I_1\ C_2\ C_2$	$3 \quad \binom{3}{2}$	$\binom{3}{2} p^2(1-p)$
3	$C_1\ C_2\ C_3$	$1 \quad \binom{3}{3}$	$\binom{3}{3} p^3$

$$Pr(X=x) = \binom{3}{x} p^x (1-p)^{3-x}$$

$x = 0, 1, 2, 3$

In general, Binomial Probability Models

$B(n, p)$

$$f(x) = Pr(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$x = 0, 1, 2, \ldots, n$

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1^n = 1$$ by Binomial Thm

10/14

cdf

$$F(x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y}$$

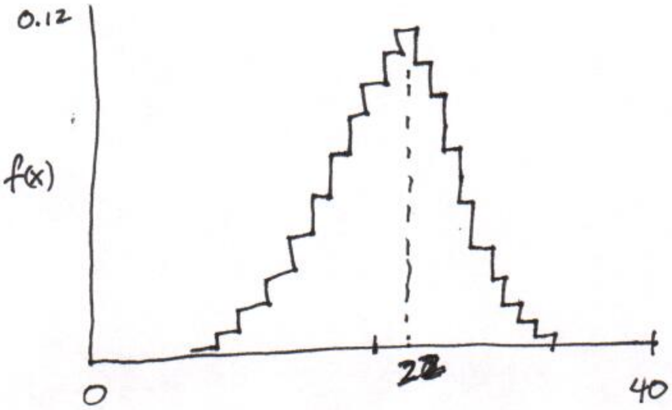
$0 \le F(x) \le 1$

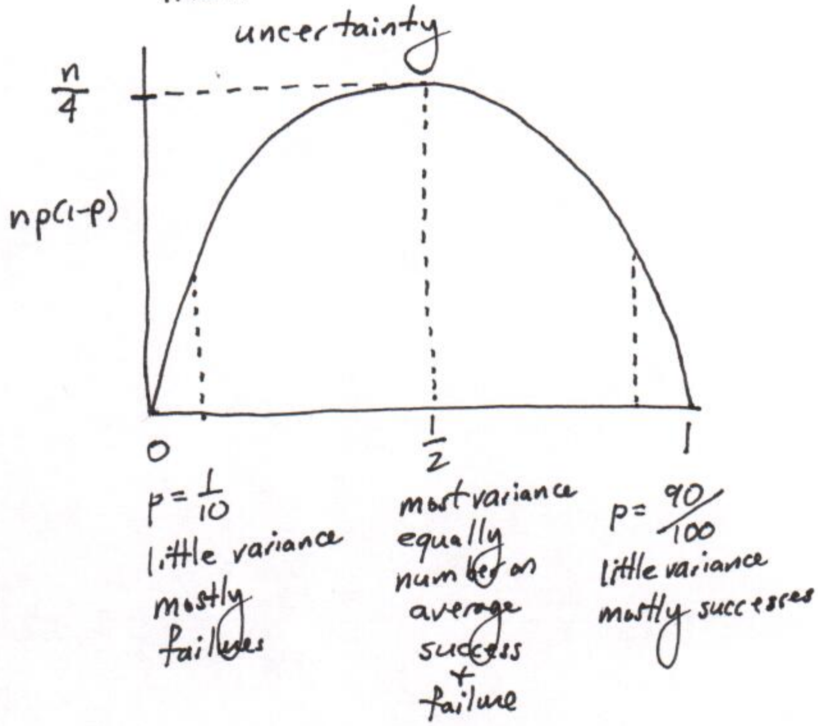
$y = 0, 1, \ldots, n$

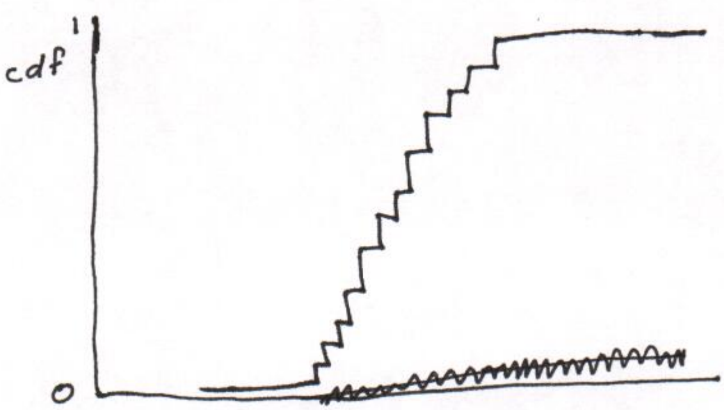
$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = np$$

number of trials $\times$ p success per trial.

$$Var(X) = np(1-p)$$







Ex. 2.1 Day 1 $\frac{22}{40}$ Day 2 $\frac{36}{40}$

Is there learning? Is p on day 2 $> \hat{p}$

Proof by Falsification (Hypothesis testing)

H_0: There is no learning

How likely are the data if there is no learning?

$$Pr(X = 36 \mid \hat{p} = 0.55, n = 40) = \binom{40}{36} \hat{p}^{36} (1-\hat{p})^4 = \binom{40}{36} (0.55)^{36} (0.45)^4$$

$$= 1.69 \times 10^{-6}$$

11/14

Conclusion : Improbable

If H_0: true $\Rightarrow$ low probability of learning $\Rightarrow$ reject H_0 (falsify H_0)

Conclude learning has happened.

Applications of the Binomial Model

1. Learning
2. Hits in a game, a season
3. Voters for a candidate
4. k heads in n tosses of a coin with probability of heads p

Proof by Falsification

~~Is there learning if~~

Is there learning ?

H_0: There is no learning

How likely are the data if no learning ?

Ex 2.1 Day 1 $\frac{22}{40}$ Day 2 $\frac{36}{40}$

$$Pr(X=36 \mid \hat{p}=0.55,\ n=40) = \binom{40}{36} \hat{p}^{36}(1-\hat{p})^4$$

$$= \binom{40}{36}(.55)^{36}(0.45)^4$$

$$= 1.69 \times 10^{-6}$$

Conclusion Improbable

If H_0: true then low prob of no learning $\Rightarrow$ reject H_0) + conclude (falsify H_0)

~~C. Poisson~~

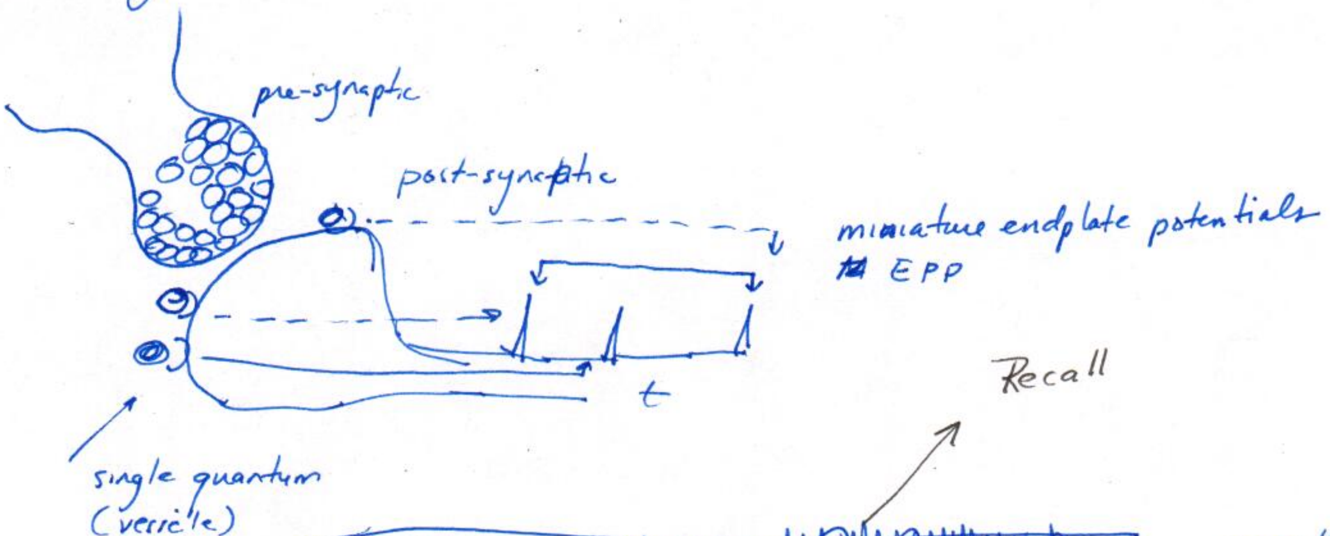
Applications

1. Learning Experiment
2. ~~Batting~~ Hits in a game (season, series) ~~N spikes~~
3. N people will vote for candidate A
4. k-heads in N tosses

H.W. 4 No independence

Poisson

Ex. 2.2 Quantal Response Hypothesis Neurotransmitter Release @ the Frog NMJ. Ach is released from the NMJ in discrete packets or quanta



Recall

Assume ~~[scribbled out]~~ in a small time window t

i) N vesicles p probability of release (and binding)

ii) release events are independent $\Rightarrow$ $Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ $np \approx \lambda$ or

iii) N - large p small $np \to \lambda$ as $n \to \infty$. $p \approx \frac{\lambda}{n}$

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

$$= \frac{n!}{(n-k)!\,k!} \cdot p^k (1-p)^{n-k}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \quad \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$n \to \infty$

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \longrightarrow 1$$

$$\left[\lim_{x \to \infty} \left(1+\frac{x}{n}\right)^n = e^x\right]$$

$$\left(1-\frac{\lambda}{n}\right)^{-k} = \frac{1}{\left(1-\frac{\lambda}{n}\right)^k} \longrightarrow 1$$

Def'n of $e = 2.718..$

$$\left(1-\frac{\lambda}{n}\right)^n \longrightarrow e^{-\lambda}$$

$\Rightarrow$

$$Pr(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Poisson Approximation to the Binomial $(N > 10^2 \quad p < 10^{-2})$

$k = 0, 1, 2, \ldots, \ldots$

Recall: $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^x$

The Poisson P.m.f.

λ - rate no# of events per unit (time or space)

$X \sim \mathcal{P}(\lambda)$

X is distributed Poisson w/ parameter λ

$$f(k) = Pr(X=k)$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

$$F(j \mid \lambda) = \sum_{k=0}^{j} \frac{\lambda^k}{k!} e^{-\lambda} = Pr(X \le j)$$

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

Intuition $np(1-p)$ for n large p small

$np(1-p) \approx np \Rightarrow E(X) \approx Var(X)$

Applications (No# of Events per unit Time)

~~No# of Events per~~

1. Number of Customers per hour at Home Depot.
2. Emergency Rm visits 12 AM - 6 AM
3. No# Miniitrature Excitating Post-Synaptic Currents
4. No# of Cases of H1N1, or Pb Poisoning, per unit area
5. No# of Soldier in the Prussian Army killed by being kicked in the head by horses in a 20 year period Bortkiewicz 1848-1868
6. Lord Rutherford + Geiger 1910
   Radioactive decay of Polonium in 125 msec

Discrete Probability Model

Any discrete-valued function can be converted in to a p.m.f. if it is finitely summable

$$\sum_{x=-\infty}^{\infty} g(x) = c < \infty$$

$$\frac{1}{1-\frac{1}{2}} = \frac{1}{\frac{1}{2}} = 2$$

$$\frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{1}{2}\right)^n = 1$$

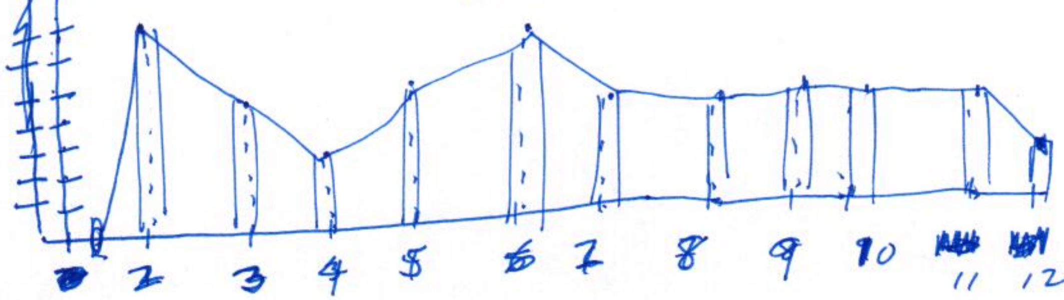
$$\sum_{n=0}^{\infty}\frac{1}{n} = \infty$$

then

$$f(x) = c^{-1} g(x) \quad \text{is a p.m.f}$$

$$Pr(X = x) = f(x) = c^{-1} g(x)$$

$c = 46$



/46