Lectures 20/21 Poisson distribution

- As a limit to binomial when n is large and p is small.
- A theorem by Simeon Denis Poisson(1781-1840). Parameter λ= np= expected value
- As n is large and p is small, the binomial probability can be approximated by the Poisson probability function
- $P(X=x)= e^{-\lambda} \lambda^x / x!$, where e =2.71828
- Ion channel modeling : n=number of channels in cells and p is probability of opening for each channel;

Binomial and Poisson approximation

x	n=100, p=.01	Poisson
0	.366032	.367879
1	.36973	.367879
2	.184865	.183940
3	.06099	.061313
4	.014942	.015328
5	.002898	.003066
6	.0000463	.000511
7		

Advantage: No need to know n and p; estimate the parameter λ from data

X= Number of deaths	frequencies
0	109
1	65
2	22
3	3
4	1
total	200

200 yearly reports of death by horse-kick from10 cavalry corps over a period of 20 years in 19th century by Prussian officials.

x	Data frequencies	Poisson probability	Expected frequencies
0	109	.5435	108.7
1	65	.3315	66.3
2	22	.101	20.2
3	3	.0205	4.1
4	1	.003	0.6
	200		

Pool the last two cells and conduct a chi-square test to see if Poisson model is compatible with data or not. Degree of freedom is 4-1-1 = 2. Pearson's statistic = .304; P-value is .859 (you can only tell it is between .95 and .2 from table in the book); accept null hypothesis, data compatible with model

Rutherfold and Geiger (1910)

- Polonium source placed a short distance from a small screen. For each of 2608 eighth-minute intervals, they recorded the number of alpha particles impinging on the screen

Other related application in

Medical Imaging : X-ray, PET scan (positron emission tomography), MRI

# of □ particles	Observed frequency	Expected freq.
0	57	
1	203	211
2	383	407
3	525	
4	532	508
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	10	11
11+	6	6

Pearson's chi-squared statistics = 12.955; d.f.=12-1-1=10

Poisson parameter = 3.87, P-value between .95 and .975. Accept null hypothesis : data are compatible with Poisson model

Poisson process for modeling number of event occurrences in a spatial or temporal domain

- Homogeneity : rate of occurrence is uniform
- Independent occurrence in non-overlapping areas
- Non-clumping