Law of Rare Event.

Sparse Sampling: The Poisson Distribution

* Sparse Sampling: counting rare events.

* Examples.

Canonical example [
- Count the meteorites striking Houston during a time period of length t ($= 10^6$ years).
- Count the errors (typographical) in a novel of 500 pages.
- Count the accidents in a stretch of road during a fixed period.
- ....

* Counting the meteorites for a period $[0, t]$:
  - divide the period $[0, t]$ into n intervals.
  - assume that n is sufficient large so that the intervals are so small that the chance of two or more strikes in the same interval is negligible.
  - assume that strikes in different intervals are independent, and that the chance of a strike is the same for each of the n intervals, p say.

$\Rightarrow$ $\underline{X}$ = total number of strikes in the n intervals

$\sim$ the number of successes in n Bernoulli trials

with distribution

$$p(k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

$\uparrow$ binomial distribution.

* Note that if p is the chance of a strike in one minute, then the chance of a strike in two minutes should be $2p$, and so on.

$\Rightarrow$ This amounts to the assumption that np/t is a constant, which we call λ, a rate (>0) per unit time (or per unit area). Such that

$$np = \lambda t$$

* Note that p deceases when n increases or vis versa.

Proposition: Suppose that in $b(k; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that $np = \lambda t$. Then

$$\lim_{n\to\infty} b(k; n, p) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Proof:

$$P[k \text{ strikes in } [0,t]] = \binom{n}{k} p^k (1-p)^{n-k}$$

as $n \to \infty$

$\left(1 - \frac{\lambda t}{n}\right)^{-k} \to 1$

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

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

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

$$\to e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad \text{as } k \to \infty$$

* check that the Poisson distribution is a proper probability distribution:

$$e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda t} e^{\lambda t} = 0$$

Note that $e^{\alpha} = \sum_{k=0}^{\infty} \alpha^k / k!$

* Approximation: An Other Look.

Example: <u>Polling voters.</u>

Given

$v = r + g$ voters altogether

with r : reds

g : greens

Sampling without replacement.

n : sample size

A_k : event that the sample includes k greens.

$$P(A_k) = \binom{g}{k}\binom{r}{n-k} \Big/ \binom{g+r}{n} \quad : \text{ hypergeometric distribution}$$

* Assume that v, g, and r are very large compared with k and n (Typically $n \sim 1000$, while r and $g \sim 10^6$). We set

$$p = g/v, \quad q = 1-p = r/v$$

For fixed n and k, as v, g, r becomes increasingly large,

$$P(A_k) = \frac{g \cdots (g-k+1)}{k!} \cdot \frac{r \cdots (r-n+k+1)}{(n-k)!} \Big/ \frac{v \cdots (v-n+1)}{n!}$$

$$= \frac{n!}{k!(n-k)!} \left\{ \left(\frac{g}{v}\right) \cdots \left(\frac{g-k+1}{v}\right) \cdot \left(\frac{r}{v}\right) \cdots \left(\frac{r-n+k+1}{v}\right) \right\} \Big/ \left\{ \left(\frac{v}{v}\right) \cdots \left(\frac{v-n+1}{v}\right) \right\}$$

as $v \to \infty$

$$\frac{g}{v}, \ \frac{g-1}{v}, \ \ldots, \ \frac{g-k+1}{v} \to p$$

$$\frac{r}{v}, \ \frac{r-1}{v}, \ \ldots, \ \frac{r-n+k+1}{v} \to q$$

$$\frac{v}{v}, \ \frac{v-1}{v}, \ \ldots \ \frac{v-n+1}{v} \to 1$$

Then $P(A_k) \to \binom{n}{k} p^k (1-p)^{n-k}$

\# Rare greens

if g is very small, hence p, we have $P(A_1) = np(1-p)^{n-1}$. we must increase n so that $np \simeq \lambda = \text{consta}$

as $n \to \infty$

$P(A_k) \sim e^{-\lambda} \frac{\lambda^k}{k!}$

Poisson dist. $\Leftarrow$