Law of Rare Events.

Sparse Sampling: The Poisson Distribution

* Sparse Sampling: counting rare events.

* Examples:
  - Count the meteorites striking Houston during a time period of length $\tau$ (= 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.

2 Counting the meteorites for a period $[0, t]$:
  - Divide the period $[0, t]$ into $n$ intervals.
  - Assume that $n$ is sufficiently 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
\( X = \text{total number of strikes in the } n \text{ intervals} \)

\( \text{the number of successes in a Bernoulli trials with distribution} \)

\[ \varphi(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n. \]

\[ \text{Binomial distribution.} \]

- 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.

- This amounts to the assumption that \( np/\tau \) is a constant, which we call \( \lambda \), a rate (per unit time) or (per unit area). Such that \( np = \lambda \tau \)

- Note that \( p \) decreases when \( n \) increases or vice versa.

**Proposition:** Suppose that in \( \varphi(k; n, p) \), we let \( n \to \infty \)
and \( p \to \lambda \) in such a way that \( np = \lambda \tau \). Then

**Proof:**
\[
\lim_{n \to \infty} \binom{n}{k} p^k (1-p)^{n-k} = \frac{(\lambda \tau)^k e^{-\lambda \tau}}{k!}
\]
as \( n \to \infty \)

\[
(1 - \frac{\lambda \tau}{n})^{-k} \to 1
\]

\[
(1 - \frac{\lambda \tau}{n})^n \to e^{-\lambda \tau}
\]

\[
P(k \text{ strikes in } \tau + 1) = \binom{n}{k} p^k (1-p)^{n-k}
\]

\[
(1 - \frac{\lambda \tau}{n})^{-k} \to 1
\]

\[
(1 - \frac{\lambda \tau}{n})^n \to e^{-\lambda \tau}
\]

\[
\to e^{-\lambda \tau} \frac{(\lambda \tau)^k}{k!}, \quad \text{as } k \to \infty
\]
a) Check that the Poisson distribution is a proper probability distribution.

\[ e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda t} e^{\lambda t} = 1 \]

Note that \( e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \).

b) Approximation: An Other Look.

Example: Polling voters.

Given \( n = r + q \) voters altogether

with \( n : \text{aeds} \)

\( q : \text{queens} \)

Sampling without replacement.

\( n : \text{sample size} \)

\( \mathbb{A}_k : \text{event that the sample includes} \)

\( k : \text{queens} \)

\[ P(\mathbb{A}_k) = \frac{\binom{q}{k} \binom{n-k}{r-k}}{\binom{n}{r}} : \text{hypergeometric distribution} \]

Assume that \( n, q, \) and \( r \) are very large compared with \( k \) and \( n \) (typically \( n \geq 1000 \), while \( n \) and \( q \geq 100 \)). We set

\( p = \frac{q}{n} \), \( q = 1 - p = \frac{r}{n} \)

For fixed \( n \) and \( k \), as \( n, q, r \) becomes increasingly large,
\[ P(A_k) = \frac{\left( \begin{array}{c} g-k+1 \\ h \end{array} \right)}{h!} \cdot \frac{\left( \begin{array}{c} v-n+k+1 \\ n-h \end{array} \right)}{(n-h)!} \]

\[ \sqrt{\left( \frac{v}{v} \right) \cdots \left( \frac{v-u+1}{v} \right)} \]

\[ = \frac{n!}{h! (n-h)!} \left( \frac{g}{v} \right) \cdots \left( \frac{g-k+1}{v} \right) \cdot \left( \frac{v}{v} \right) \cdots \left( \frac{v-u+k+1}{v} \right) \]

As \( u \to 0 \)

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

\[ \frac{v}{v}, \frac{v-1}{v}, \ldots, \frac{v-u+k+1}{v} \to g \]

\[ \frac{v}{v}, \frac{v-1}{v}, \ldots, \frac{v-u}{v} \to 1 \]

Then

\[ P(A_k) \to \left( \begin{array}{c} n \end{array} \right) \cdot p^h \cdot (1-p)^{n-h} \]

As \( n \to \infty \), if \( g \) is very small, hence \( p \), we have \( P(A_k) \ll \frac{v}{v} \).

If \( g \) is very small, hence \( p \), we must increase \( n \).