- We want to represent counts

Poisson R.V.'s capture the idea of uniformity

- The count in an interval
  - depends on length of interval
  - in fact, $\propto$ length
- Events are independent.

(Idea works in multiple dimensions).

If $X$ is poisson, intensity $\lambda$

- $E(X) = \lambda$
- $Var(X) = \lambda$

Notice $\lambda$ <u>has units</u> (eg $\frac{\#}{s}$, etc

and <u>depends on scale of interval</u>.

PMF:

$$P(X = n \mid \text{unit interval}, \lambda) = \frac{e^{-\lambda} \cdot \lambda^n}{n!}$$

straightforward series manipulation gives
- this is a PMF
- expectation
- variance.

Now assume whatever we're watching
is • Poisson

- types are independent

[These assumptions are a stretch; words aren't like this; neither are animals or objects; but they're simple + generic]

Notice:

1) If I observe a Poisson RV for an interval longer than 1

$$P(X = n \mid \text{Interval length } t, \lambda) = e^{-\lambda t} \frac{(\lambda t)^{x}}{x!}$$

2) If I observe Poisson RV for N unit intervals, seeing count $n_i$ in the i'th, Max likelihood est of $\lambda$ is

$$\lambda^* = \frac{1}{N} \sum_i n_i$$

3) If I observe poisson RV for N ~~unit~~ intervals, i'th has length $L_i$, and see $n_i$ in i'th, Max likelihood gives

$$\lambda^* = \frac{1}{N} \sum_i \left( \frac{n_i}{L_i} \right)$$

③

Now assume we see $\left\{ \begin{matrix} \text{words} \\ \text{animals} \\ \text{objects} \end{matrix} \right\}$ for an interval (which could be time or space). We ~~choose~~ a scale so this interval is $[-1, 0]$.

Each word type has intensity

$$\lambda_w \quad \begin{matrix} \leftarrow \text{intensity} \\ \leftarrow \text{type} \end{matrix}$$

which is unknown

Natural to try and draw conclusions from word type counts.

$$n_x = \left[ \begin{matrix} \text{number of word types} \\ \text{that appear } x \text{ times} \end{matrix} \right]$$

Natural because Max likelihood on individual words is no help.

$\rightarrow$ ~~ML~~

- Also, assume future is like the past.

  i.e. if a word has $\lambda_w$ in $[-1,0]$ it has this $\lambda$ later.

- E + T phrase this as "conditionally binomial"

- ML est. of $\lambda$ for words we <u>haven't</u> seen is $0$
- But $n_1$ large compared to $n_2$, etc suggests there are ~~words~~ types where
  - × we haven't seen them
  - × $\lambda_w > 0$
- So we should be looking at $G(\lambda)$.

$$G(\lambda) = P(\lambda_w \leq \lambda)$$

- clearly, a discrete distribution
- represents CDF (cumulative dist. function)

$$dG(\lambda) \equiv p(\lambda_w = \lambda)\, d\lambda$$

$\delta$ functions or atoms

Now

$$P(\text{a}\overset{\text{given}}{\uparrow}\text{ word type has count } x \mid \lambda_w)$$

$$= \frac{e^{-\lambda_w}\, \lambda_w^{x}}{x!}.$$

$$P(\text{a}\overset{\text{given}}{\uparrow}\text{ word type has count } x)$$

$$= \int \frac{e^{-\lambda}\, \lambda^{x}}{x!}\, dG(\lambda)$$

<u>or</u> $\int \frac{e^{-\lambda}\, \lambda^{x}}{x!}\, p(\lambda)\, d\lambda$

$E[\#\text{ of word types w/ count } x]$

$$= \sum_{i \in \text{word types}} P(\text{word type } i \text{ has count } x)$$

$$= C \int \frac{e^{-\lambda}\, \lambda^{x}}{x!}\, dG(\lambda) = \eta_x$$

total # of word types, <u>unknown</u>!

- Another way to look at this is that

$d\Gamma(\lambda) = C\, dG(\lambda)$ is a <u>measure</u>

(like a PDF, +ve, but <u>doesn't</u> $\int$ to 1)

- Notice that $C$ could be hard to get, because we could have support for $G(\lambda)$ at, say,

$$\lambda = 10^{-12}$$

$$\Updownarrow$$

- there is a word we see about once in $10^{12}$ intervals.
- affects $C$, but not a significant effect on observations.

Now

- $\eta_x$ is the observed value of an RV
- call it $r_x$
- we have
  - $E(r_x) = \eta_x$
- reasonable approx ?
  - $Var(r_x) = \eta_x$

Sum of Poisson

$$r_x \sim N(\eta_x, \sqrt{\eta_x})$$

sum of random variables

- This will come in useful.

Now consider

$$\Delta(t) = E\left[\begin{array}{l}\text{\# of types seen in } [0;t], \\ \text{but } \underline{\text{not}} \text{ in } [-1;0]\end{array}\right]$$

then

$$\Delta(t) = C\int_0^\infty \left[e^{-\lambda}\frac{\lambda^0}{0!}\right]\left[1 - e^{-\lambda t}\right] dG(\lambda)$$

seen 0 times in $-1, 0$

not seen 0 times in $0, t$.

Notice you can derive expressions for

$$E\left[\begin{array}{l}\text{\# seen } a \text{ times in } [-1,0] \text{ and} \\ b \text{ times in } [0,t]\end{array}\right]$$

in the straightforward way.

- Notice also that you <u>dont</u> need $C$ to evaluate this, just $C\,dG(\lambda)$.
- Q: <u>estimate $\Delta(t)$ given $n_x$</u>

- <u>Notice</u>

$$1-e^{-\lambda t} = \lambda t - \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^3}{3!} - \frac{(\lambda t)^4}{4!} \cdots$$

<u>So:</u>

$$\Delta(t) = \eta_1 t - \eta_2 t^2 + \eta_3 t^3 - \eta_4 t^4 \cdots$$

(assuming convergence, etc).

<u>Natural estimator:</u>

assume $\eta_1 = n_1$, $\eta_2 = n_2$, etc.

substitute