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

\[ \lambda \]

\[ E(X) = \lambda \]

\[ \text{Var}(X) = \lambda \]

Notice \( \lambda \) has units (e.g. \( \frac{\#}{s} \), etc. and depends on scale of interval.)
PMF:
\[ P(X = n \mid \text{unit}) = \frac{e^{-\lambda} \lambda^n}{n!} \]

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

\[ P(X = n \mid \text{interval length } t \mid \lambda) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \]

2) If I observe Poisson RV for \( N \) unit intervals, seeing count \( n_i \) in the \( i^{th} \) interval, Max likelihood est of \( \lambda \) is

\[ \lambda^* = \frac{1}{N} \sum n_i \]

3) If I observe poisson RV for \( N \) unit intervals, \( i^{th} \) has length \( \theta_i \) and see \( n_i \) in \( i^{th} \), Max likelihood gives

\[ \lambda^* = \frac{1}{N} \sum \left( \frac{n_i}{\theta_i} \right) \]
Now assume we see \{ words \} for objects on an interval (which could be time or space), we choose a scale so this interval is \([ -1, \infty ]\).

Each word type has intensity

\[
\lambda_{\text{type}} \quad \text{intensity}
\]

which is unknown.

Natural to try and draw conclusions from word type counts

\[
N_{\text{oc}} = \left[ \frac{\text{number of word types}}{\text{that appear } x \text{ times}} \right]
\]

Natural because Max likelihood on individual words is no help.
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"
The estimated value \( \lambda \) for words we haven't seen is 0.

But if \( n_1 \) is large compared to \( n_2 \), etc., suggests there are word 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 distribution function).

\[
dG(\lambda) = p(\lambda_w = \lambda) d\lambda
\]

It's functions or atoms.
Now
\[ P(\text{a word type has count } x \mid \lambda) \]
\[ = e^{-\lambda x} \frac{\lambda^x}{x!} \]
\[ P(\text{a word type has count } x) \]
\[ = \int e^{-\lambda} \frac{x^x}{x!} \, dG(\lambda) \]
\[ \text{or } \int e^{-\lambda} \frac{x^x}{x!} \, p(\lambda) \, d\lambda \]

\[ E[\# \text{ of word types w/ count } x] \]
\[ = \sum_{i \in \text{word types}} P(\text{word type } i \text{ has count } x) \]
\[ = C \int e^{-\lambda} \frac{x^x}{x!} \, pG(\lambda) = \eta_x \]
\[ \text{total # of word types, unknown!} \]
Another way to look at this is that
\[ d\Gamma(\lambda) = C dG(\lambda) \]
is a measure
(like a PDF, the, but doesn't \( \int 1 \))

Notice that \( C \) could be hard to get, because we could have support for \( G(\lambda) \) at, say,
\[ \lambda = 10^{-12} \]

\[ \sqrt{1} \]

There is a word we see about once in \( 10^{12} \) intervals.

Affects \( C \), but not a significant effect on observations.
Now

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

\[ \sim \overset{\text{sum of Poisson}}{\sim} N(n_x, \sqrt{n_x}) \]

\[ \overset{\text{sum of random variables}}{\sim} \]

This will come in useful.
Now consider

\[ \Delta(t) = E \left[ \text{\# of types seen in } [0; t], \right. \]
\[ \left. \quad \text{but not in } [-1; 0] \right] \]

Then

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

seen 0 times
\[ \text{in } -1, 0 \]

not seen 0 times in
\[ \text{in } 0, t \]

Notice you can derive expressions for

\[ E \left[ \text{\# seen a times in } [-1, 0] \text{ and } \right. \]
\[ \left. \quad \text{6 times in } [0, t] \right] \]

in the straightforward way.
Notice also that you don't need $C$ to evaluate this, just $CdG(\lambda)$.

Q: estimate $\Delta(t)$ given $\eta_\infty$

Notice

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

So:

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

(assuming convergence, etc.).

Natural estimator:

assume $\eta_1 = \eta_1, \eta_2 = \eta_2, \ldots$

substitute