(skip Fisher model)

Notice that, for $t > 1$

$$n_1 t - n_2 t^2 + n_3 t^3 - \cdots$$

is a series that oscillates savagely.

You could interpret this several ways

- it doesn't converge. — panic
- the oscillations "cancel", and we need some way to accelerate this cancellation

↑ quite plausible, as $n_x \to 0$ as $x \to \infty$.

this gives §4 — Euler's transform.

Now skip to §7.

Recall that $\Delta(t)$ may be hard to estimate for large t (because there may be low frequency words).

Instead, try for a lower bound on $\Delta(t)$ — call this bound $b(t)$

Problem becomes:

$$b(t) = \inf_{C\,dG(\lambda)} \left[\int_0^\infty e^{-\lambda} \left[1 - e^{-\lambda t} \right] \left[C\,dG(\lambda) \right] \right]$$

Subject to

$$\eta_x = \epsilon \int_0^\infty \left[e^{-\lambda} \frac{\lambda^x}{x!} \right] \left[C\,dG(\lambda) \right]$$

This would be an LP in $C\,dG(\lambda)$, IF we knew η_x.

Strategy 1:

* assume $\eta_x = n_x$.
* discretize η_x

→ then we have an honest LP and can solve.

- This works for ~~th~~ +T, but failed on object data (infeasible - ?!?)

Strategy 2:

* assume ~~$\eta_x \leq$~~ $\eta_x \leq n_x + \gamma\sqrt{n_x}$

$$\eta_x \geq n_x - \gamma\sqrt{n_x}$$

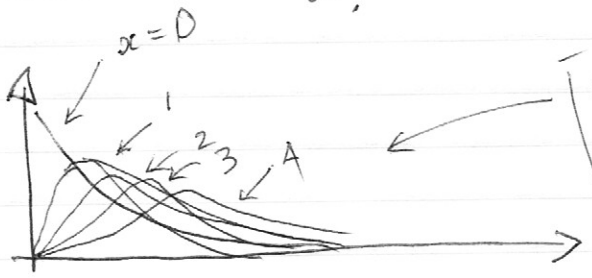
(i.e. γ stds away from mean)

- This might slacken the bound, but is probably better practice

But S2 isn't all that reliable either (on objects, get feasibility only if you apply big γ AND use only n_x for $x \in [1 \cdots 5]$ which is worrying.)

What is going on here?

Consider $\frac{e^{-\lambda} \lambda^x}{x!}$ as a function of λ

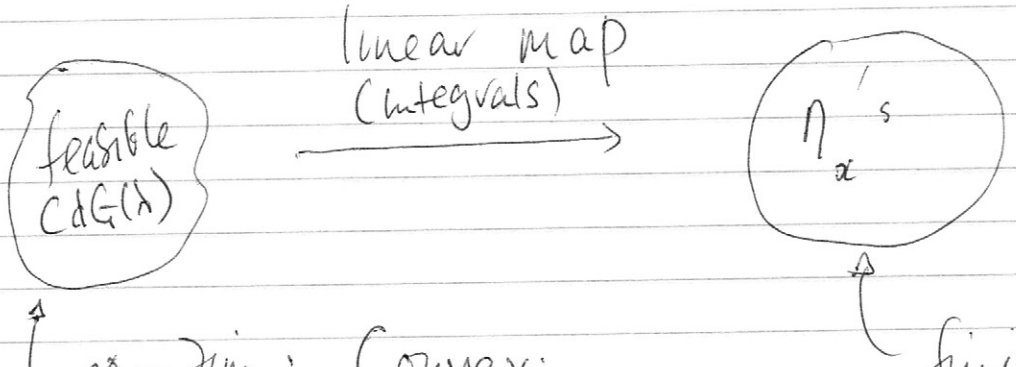


This means we can't have

$$\eta_1 = 100 \; ; \; \eta_2 = 0 \; ; \; \eta_3 = 100 \; ; \text{ etc.}$$

because these functions overlap so strongly — $\int \frac{e^{-\lambda} \lambda^1}{1!} C \, dG(\lambda)$ is similar to $\int \frac{e^{-\lambda} \lambda^2}{2!} C \, dG(\lambda)$

Alternative view:



∞-dim; Convex;

finite dim;

Since $CdG(\lambda) \geq 0$, this isn't the whole space

-ve η_x's can't occur.

~~This picture strongly implies that~~

Use this picture with plots;

there are vectors of η_x

that are (a) non-negative

(b) infeasible

and if you use $n_x = \eta_x$,

this gets infeasibility

Also explains why large r is required, AND why large x creates problems (the n_x estimates are poor).

What to do?

we actually know quite a lot about r_x, which is what we should be working with

- approximately Gaussian
- var approx = mean.

Strategy 3

assume $r_x \sim N(n_x, \sqrt{n_x})$

and $\eta_x = r_x$.

Now we must (a) estimate the counts and (b) estimate $b(t)$

$$\inf_{C\,dG(\lambda)} \int e^{-\lambda}\left[1-e^{-\lambda t}\right]\left[C\,dG(\lambda)\right] + \mu \sum_{i=1}^{K} s_i^2$$

(μ: weight)

$$s_i^2 = \frac{(r_i - n_i)^2}{2\,n_i^2}$$ ← Standard error of ith count,

$$r_i = \int \frac{e^{-\lambda}\lambda^i}{i!}\, C\,dG(\lambda)$$ ← Count estimates

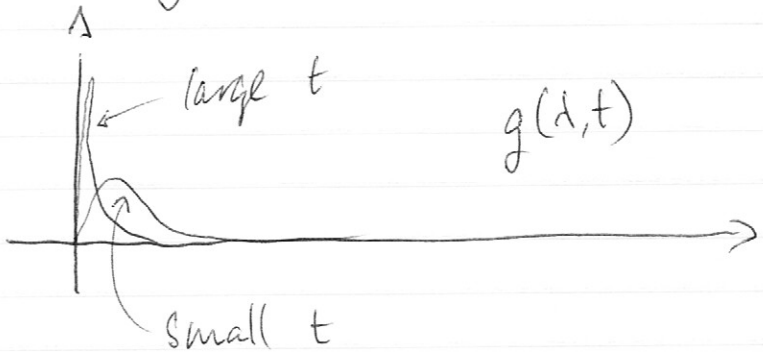
Q: how to choose μ?

A: cross validation

Now a QP, but it's convex, so no worries

Notice also that

$g(\lambda, t) = e^{-\lambda}(1 - e^{-\lambda t})$ is important



- This (basically) looks for weight in small λ's (of CdG(λ)) which makes sense.

Issue: . lower bounds are helpful, but we want more (estimates, etc).

- Options

  . work w/ continuous CdG(λ) models

  . make models explicitly discrete.

- Notice ~~probe~~ one attractive feature of this formulation
  - γ_x, $b(t)$, $\Delta(t)$ are quite insensitive to "small" changes in $Cal(G(t))$
  - because
    $$\frac{e^{-\lambda}\lambda^x}{x!}$$ is pretty smooth.
- This means that big shifts in where the weight is in a prob model are required to change counts
- In turn, we could use quite rough discretizations (but finer near 0).