STSCI 4780
Continuous parameter estimation, cont'd

Tom Loredo, CCAPS & SDS, Cornell University

2020-02-11

Recap: Inference with discrete data

- Binary data:
  - Bernoulli, binomial, negative binomial dist'ns
  - Beta posterior and prior dist'ns
- Categorical data:
  - Categorical and multinomial dist'ns
  - Dirichlet posterior and prior dist'ns
- Common ideas:
  - Sufficient statistics
  - Conjugate models (prior/likelihood)

Poisson process: A continuous analog of the Bernoulli process

Bernoulli process and binomial distribution

Bernoulli process with success probability α produces binary sequences:

$$0110010011001001101010010001000001\cdots$$

Report n, the count of 1s in a sequence of length $N \rightarrow$ **binomial distribution**:

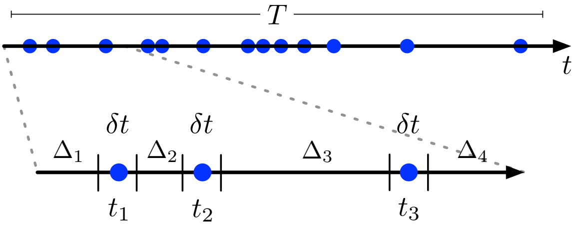
$$\begin{aligned}\mathcal{L}(\alpha) &\equiv p(n|\alpha, \mathcal{C}) \\ &= \frac{N!}{n!(N-n)!}\alpha^n(1-\alpha)^{N-n}\end{aligned}$$

Expected number of successes in N trials:

$$\mathbb{E}((n)) = \alpha N$$

Poisson point process and Poisson (counting) distribution

Poisson point process with *intensity* λ (rate per unit interval):



Report n, the number of events in an interval of size $T \rightarrow$ **Poisson distribution**:

$$\begin{aligned}\mathcal{L}(\lambda) &\equiv p(n|\lambda, \mathcal{C}) \\ &= \frac{(\lambda T)^n}{n!} e^{-\lambda T}\end{aligned}$$

Expected number of counts in T:

$$\mathbb{E}((n)) = \lambda T$$

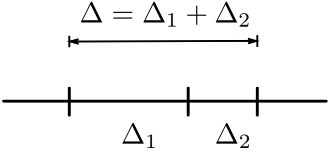
The Poisson distribution for counts from a point process

For occurence/arrival of n events in an interval Δ, let's seek

$$f_n(Delta) \equiv P(n \text{ events in } \Delta|\mathcal{P}),$$

where we'll figure out what we have to assume ($\mathcal{P}$) as we go

Partitioning an empty interval



$$f_0(\Delta) = P(\text{no events in } \Delta_1) \times P(\text{no events in } \Delta_2|\text{no events in } \Delta_1) \,||\, \mathcal{P}$$

As a simple modeling choice, let's *assume independence*:

$$P(\text{no events in } \Delta_2|\text{no events in } \Delta_1) = P(\text{no events in } \Delta_2) \quad ||\, \mathcal{P}$$

Independence implies

$$f_0(\Delta) = P(\text{no events in } \Delta_1) \times P(\text{no events in } \Delta_2) \quad || \mathcal{P}$$
$$\rightarrow \boxed{f_0(\Delta_1 + \Delta_2) = f_0(\Delta_1) \times f_0(\Delta_2)}$$

This is a **functional equation** for $f_0(\cdot)$; it has two solutions:

$$f_0(\Delta) = 0 \qquad \text{so} \quad 0 = 0 \times 0$$
$$f_0(\Delta) = e^{-\lambda\Delta} \quad \text{so} \quad e^{-\lambda(\Delta_1+\Delta_2)} = e^{-\lambda\Delta_1} \times e^{-\lambda\Delta_2}$$

Let's use the *interesting one*:

$$\boxed{f_0(\Delta) = e^{-\lambda\Delta}}$$

Note this requires that *we specify a constant, λ*

Small interval behavior

What is the meaning of λ? Note that

$$P(1 \text{ or more events in } \Delta|\mathcal{P}) \;=\; 1 - f_0(\Delta) \;=\; 1 - e^{-\lambda\Delta}$$

If Δ is small so that $\lambda\Delta \ll 1$, then $e^{-\lambda\Delta} = 1 - \lambda\Delta + O(\Delta^2)$

$$P(1 \text{ or more events in } \Delta|\mathcal{P}) = \lambda\Delta + O(\Delta^2)$$

The probability of seeing *at least* one event in a small interval is $\propto \Delta$ (and λ), and $\lambda \geq 0$

What about 2 events in a small interval?

$$\begin{aligned} P(2 \text{ or more events in } \Delta|\mathcal{P}) \quad &= [1 - f_0(\Delta)] - f_1(\Delta) \\ &= \lambda\Delta - f_1(\Delta) + O(\Delta^2) \end{aligned}$$

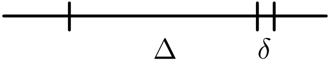
As another simplifying assumption, let's require that *events are simple*, so that this probability vanishes as $\Delta \to 0$ (no events at exactly the same instant/location):

$$\boxed{f_1(\Delta) = \lambda\Delta + O(\Delta^2)}$$

The probability of seeing *exactly* one event in a small interval is $\propto \Delta$, i.e., λ is a *rate parameter* (*intensity*)

Extending an interval

To get a handle on the exact $f_n(\Delta)$ for $n > 0$, let's look at how $f_n(\Delta)$ changes if we grow the interval by a small amount δ:



Using the LTP we can write

$$f_n(\Delta + \delta) = f_n(\Delta) f_0(\delta) + f_{n-1}(\Delta) f_1(\delta) + f_{n-2}(\Delta) f_2(\delta) + \cdots$$

Let's exploit what we know about f_0 and small-interval behavior:

$$\begin{aligned} f_n(\Delta + \delta) &= f_n(\Delta) e^{-\lambda\delta} + f_{n-1}(\Delta)\lambda\delta + O(\delta^2) \\ &= f_n(\Delta)(1 - \lambda\delta) + f_{n-1}(\Delta)\lambda\delta + O(\delta^2) \\ f_n(\Delta + \delta) - f_n(\Delta) &= -\lambda\delta f_n(\Delta) + \lambda\delta f_{n-1}(\Delta) + O(\delta^2) \end{aligned}$$

Divide by δ and take $lim_{\delta \to 0}$:

$$\boxed{f_n'(\Delta) = -\lambda f_n(\Delta) + \lambda f_{n-1}(\Delta)}$$

This is a recursive sequence of inhomogeneous differential equations—infinitely many!

$$f_n'(\Delta) = -\lambda f_n(\Delta) + \lambda f_{n-1}(\Delta)$$

Let's check $n = 0$, where there is no inhomogeneous term:

$$f_0'(\Delta) = -\lambda f_0(\Delta)$$

The solution is $Ce^{-\lambda\Delta}$, but since we know $f_0(0) = 1$, we know $C = 1$

For $n = 1$,

$$f_1'(\Delta) = -\lambda f_1(\Delta) + \lambda e^{-\lambda\Delta}$$

As an inspired guess (or using *variation of parameters*), try

$$\begin{aligned} f_1(\Delta) &= \lambda\Delta e^{-\lambda\Delta} \\ \rightarrow f_1'(\Delta) &= -\lambda^2\Delta e^{-\lambda\Delta} + \lambda e^{-\lambda\Delta} \end{aligned}$$

which satisfies the differential eq'n

Iterating, we find:

$$\boxed{f_n(\Delta) = \frac{(\lambda\Delta)^n}{n!} e^{-\lambda\Delta}}$$

The Poisson distribution

If we model events distributed in an interval Δ such that:

- A single parameter, λ, governs the process
- With λ specified, probabilities for event counts in non-overlapping intervals are independent
- The events are simple

then denoting these assumptions by $\mathcal{P} = \lambda, \mathcal{C}$ (and including the interval size in $\mathcal{C}$)

$$p(n|\lambda, \mathcal{C}) = \frac{(\lambda\Delta)^n}{n!} e^{-\lambda\Delta}$$

with λ corresponding to the event rate

We can show:

$$\mathbb{E}(n) = \lambda\Delta, \qquad \mathrm{Var}(n) = \lambda\Delta$$

"$\lambda, \mathcal{C}$" is analogous to "α, N IID trials" for binomial

Infer a Poisson rate from counts

Problem:

Observe n counts in T; infer rate (intensity), r

Likelihood

Poisson distribution:

$$\begin{aligned}\mathcal{L}(r) &\equiv p(n|r,\mathcal{C}) \\ &= \frac{(rT)^n}{n!}e^{-rT}\end{aligned}$$

Prior

Two simple "uninformative" standard choices:

- r known to be *nonzero*: it is a scale parameter; scale invariance $\rightarrow$

  $$p(r|\mathcal{C}) = \frac{1}{\ln(r_u/r_l)} \frac{1}{r}$$

  This corresponds to a flat prior on $\lambda = \log r$

- r may *vanish*; require prior predictive $p(n|\mathcal{C}) \sim$ Const:

  $$p(r|\mathcal{C}) = \frac{1}{r_u}$$

The *reference prior* ("uninformative" in an asymptotic, information-theoretic sense) is $p(r|\mathcal{C}) \propto 1/r^{1/2}$

Prior predictive

Adopting a flat (uniform) prior,

$$
\begin{aligned}
p(n|\mathcal{C}) &= \frac{1}{r_u}\frac{1}{n!}\int_0^{r_u} dr (rT)^n e^{-rT} \\
&= \frac{1}{r_u T}\frac{1}{n!}\int_0^{r_u T} d(rT)(rT)^n e^{-rT} \\
&\approx \frac{1}{r_u T} \quad \text{for} \quad r_u \gg \frac{n}{T}
\end{aligned}
$$

Posterior

A *gamma distribution*:

$$
p(r|n,\mathcal{C}) = \frac{T(rT)^n}{n!} e^{-rT}
$$

Gamma Distributions

A 2-parameter family of distributions over nonnegative x, with shape parameter α and scale parameter λ (or inverse scale ϵ):

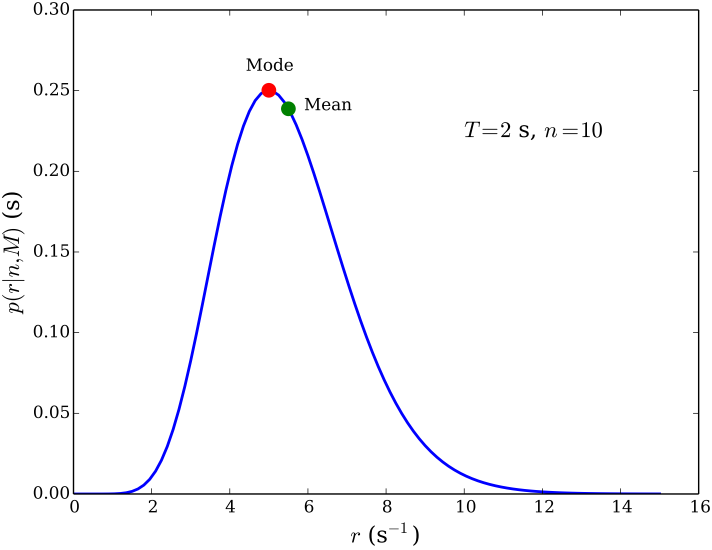
$$
\begin{aligned}
p_\Gamma(x|\alpha,\lambda) &\equiv \frac{1}{\lambda\Gamma(\alpha)}\left(\frac{x}{\lambda}\right)^{\alpha-1} e^{-x/\lambda} \\
&\equiv \frac{\epsilon}{\Gamma(\alpha)}(x\epsilon)^{\alpha-1}e^{-x\epsilon}
\end{aligned}
$$

Moments:

$$
\mathbb{E}(x) = \alpha\lambda = \frac{\alpha}{\epsilon} \qquad \mathrm{Var}(x) = \lambda^2\alpha = \frac{\alpha}{\epsilon^2}
$$

Our posterior corresponds to $\alpha = n+1$, $\lambda = 1/T$.

- Mode $\hat{r} = \frac{n}{T}$; mean $\langle r \rangle = \frac{n+1}{T}$ (shift down 1 with $1/r$ prior)
- Std. dev'n $\sigma_r = \frac{\sqrt{n+1}}{T}$; credible regions found by integrating (can use incomplete gamma function)



Mode
Mean
T=2 s, n=10
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$p(r|n,M)$ (s)
0
2
4
6
8
10
12
14
16
r (s^{-1})

Conjugate prior

Note that a gamma distribution prior is the conjugate prior for the Poisson sampling distribution:

$$
\begin{aligned}
p(r|n, M') &\propto \mathrm{Gamma}(r|\alpha, \epsilon) \times \mathrm{Pois}(n|rT) \\
&\propto r^{\alpha-1} e^{-r\epsilon} \times r^n e^{-rT} \\
&\propto r^{\alpha+n-1} \exp[-r(T+\epsilon)]
\end{aligned}
$$

Useful conventions

- Use a flat prior for a rate that may be zero
- Use a log-flat prior ($\propto 1/r$) for a nonzero scale parameter
- Use proper (normalized, bounded) priors
- Plot posterior with abscissa that makes prior flat (use $\log r$ abscissa for scale parameter case)

Supplementary material

The On/Off Problem: Handling a nuisance parameter

Basic problem

- Look off-source; unknown background rate b
  Count $N_{\rm off}$ photons in interval $T_{\rm off}$
- Look on-source; rate is $r = s + b$ with unknown signal s
  Count $N_{\rm on}$ photons in interval $T_{\rm on}$
- Infer s

Conventional solution

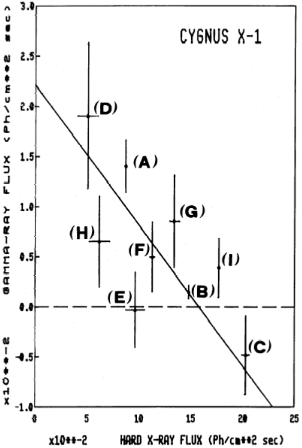
$$
\begin{aligned}
\hat{b} &= N_{\rm off}/T_{\rm off}; & \sigma_b &= \sqrt{N_{\rm off}}/T_{\rm off} \\
\hat{r} &= N_{\rm on}/T_{\rm on}; & \sigma_r &= \sqrt{N_{\rm on}}/T_{\rm on} \\
\hat{s} &= \hat{r} - \hat{b}; & \sigma_s &= \sqrt{\sigma_r^2 + \sigma_b^2}
\end{aligned}
$$

But $\hat{s}$ can be negative!

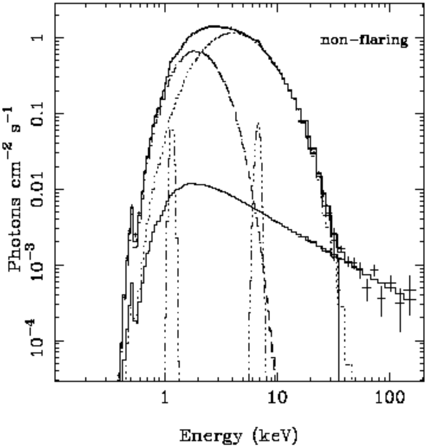
Examples

Spectra of X-Ray Sources

Bassani et al. 1989

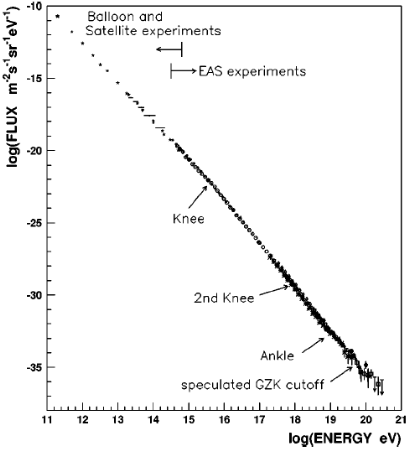


Di Salvo et al. 2001

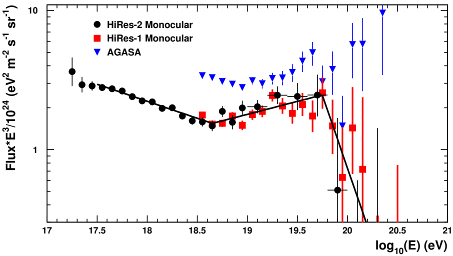


Spectrum of Ultrahigh-Energy Cosmic Rays

Nagano & Watson 2000



HiRes Team 2007



Bayesian Solution to On/Off Problem

First consider off-source data; use it to estimate b:

$$p(b|N_{\rm off}, I_{\rm off}) = \frac{T_{\rm off}(bT_{\rm off})^{N_{\rm off}} e^{-bT_{\rm off}}}{N_{\rm off}!}$$

Use this as a prior for b to analyze on-source data

For on-source analysis $I_{\rm all} = (I_{\rm on}, N_{\rm off}, I_{\rm off})$:

$$p(s, b|N_{\rm on}) \propto p(s)p(b)[(s+b)T_{\rm on}]^{N_{\rm on}} e^{-(s+b)T_{\rm on}} \qquad || \; I_{\rm all}$$

$p(s|I_{\rm all})$ is flat, but $p(b|I_{\rm all}) = p(b|N_{\rm off}, I_{\rm off})$, so

$$p(s, b|N_{\rm on}, I_{\rm all}) \propto (s+b)^{N_{\rm on}} b^{N_{\rm off}} e^{-sT_{\rm on}} e^{-b(T_{\rm on}+T_{\rm off})}$$

Now marginalize over b;

$$\begin{aligned} p(s|N_{\rm on}, I_{\rm all}) &= \int db\; p(s, b \mid N_{\rm on}, I_{\rm all}) \\ &\propto \int db\; (s+b)^{N_{\rm on}} b^{N_{\rm off}} e^{-sT_{\rm on}} e^{-b(T_{\rm on}+T_{\rm off})} \end{aligned}$$

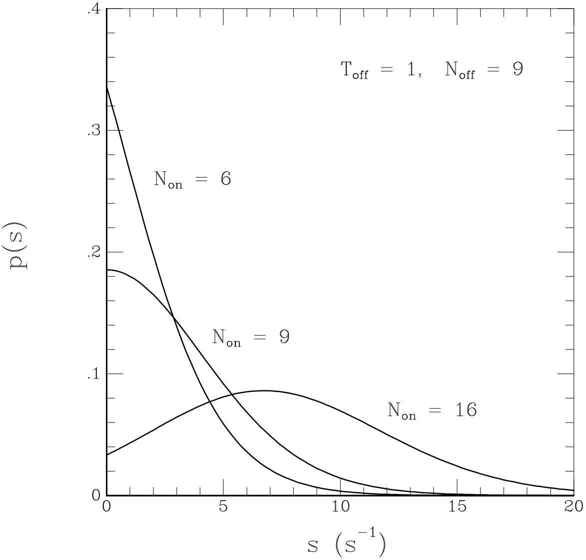
Expand $(s+b)^{N_{\rm on}}$ and do the resulting Γ integrals:

$$\begin{aligned} p(s|N_{\rm on}, I_{\rm all}) &= \sum_{i=0}^{N_{\rm on}} C_i \frac{T_{\rm on}(sT_{\rm on})^i e^{-sT_{\rm on}}}{i!} \\ C_i &\propto \left(1 + \frac{T_{\rm off}}{T_{\rm on}}\right)^i \frac{(N_{\rm on} + N_{\rm off} - i)!}{(N_{\rm on} - i)!} \end{aligned}$$

Posterior is a weighted sum of Gamma distributions, each assigning a different number of on-source counts to the source (evaluate via recursive algorithm or confluent hypergeometric function)

Example On/Off Posteriors—Short Integrations

$T_{\mathrm{on}} = 1$



Example On/Off Posteriors—Long Background Integrations

$T_{\mathrm{on}} = 1$