Inference for non-Stationary, non-Gaussian, Irregularly-Sampled Processes

Robert Wolpert
Duke Department of Statistical Science
wolpert@stat.duke.edu

SAMSI | ASTRO Opening WS Aug 22-26
2016 Aug 24   Hamner Conference Center, RTP, NC

Outline

Basic Ideas

The entire talk in one slide:

- Basic idea 1: Even if we only observe events Y_i at discrete times $\{t_i\}$, we can *model* these as observations $Y(t_i)$ of a *process* $Y(t)$ that is well-defined (if un-observed) in continuous time. This overcomes the "regularly-spaced observations" limitation.

Basic Ideas

The entire talk in one slide:

- Basic idea 1: Even if we only observe events Y_i at discrete times $\{t_i\}$, we can *model* these as observations $Y(t_i)$ of a *process* $Y(t)$ that is well-defined (if un-observed) in continuous time. This overcomes the "regularly-spaced observations" limitation.
- Basic idea 2: Almost anything you can do with Gaussian distributions can also be one with other Infinitely-Divisible (ID) distributions, such as Poisson, Negative Binomial, Gamma, α-Stable, Cauchy, Beta Process. Some of these feature discrete (integer) values, and some feature heavy tails, offering a wider range of behavior than is possible with Gaussian time series or processes.

Basic Ideas

The entire talk in one slide:

- Basic idea 1: Even if we only observe events Y_i at discrete times $\{t_i\}$, we can *model* these as observations $Y(t_i)$ of a *process* $Y(t)$ that is well-defined (if un-observed) in continuous time. This overcomes the "regularly-spaced observations" limitation.
- Basic idea 2: Almost anything you can do with Gaussian distributions can also be one with other Infinitely-Divisible (ID) distributions, such as Poisson, Negative Binomial, Gamma, α-Stable, Cauchy, Beta Process. Some of these feature discrete (integer) values, and some feature heavy tails, offering a wider range of behavior than is possible with Gaussian time series or processes.
- Basic idea 3: Start with Stationary processes (like the Ornstein-Uhlenbeck process) or processes with Stationary increments (like Brownian Motion), then extend to *non*-stationary processes (like Diffusions).

Outline

Easy TS Example: AR(1)

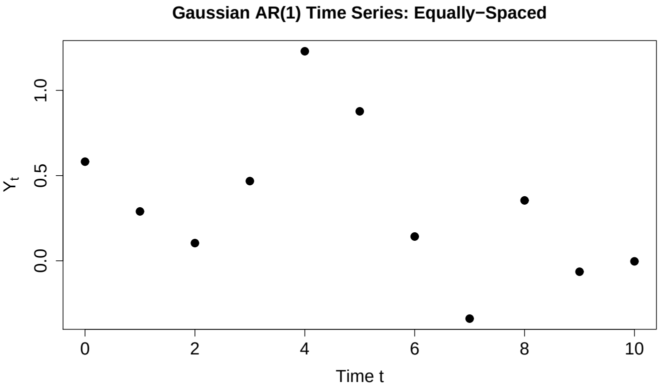
Fix $\mu \in \mathbb{R}$, $\sigma^2 > 0$, and $\rho \in (0, 1)$.
Draw $X_0 \sim \mathsf{No}(\mu, \sigma^2)$.
For times $t = 1, 2, \cdots$, set

$$X_t := \rho X_{t-1} + \zeta_t$$
$$= \rho^t X_0 + \sum_{s=0}^{t-1} \zeta_{t-s}\rho^s = \sum_{s=0}^{\infty} \zeta_{t-s}\rho^s$$

with Normal innovations $\{\zeta_t\} \overset{\text{iid}}{\sim} \mathsf{No}\big((1-\rho)\mu, (1-\rho^2)\sigma^2\big)$

Like most commonly-studied Time Series models, this features:

- Regularly-spaced observation times $t = 0, 1, 2, 3, \ldots$;
- Gaussian marginal distributions $X_t \sim \mathsf{No}(\mu, \sigma^2)$;
- Stationary distributions, with autocorrelation $\mathsf{Corr}(X_s, X_t) = \rho^{|t-s|}$.


Gaussian AR(1) Time Series: Equally−Spaced
Y_t
0.0
0.5
1.0
0
2
4
6
8
10
Time t

Outline

Extension 1: Continuous Time

Set $\lambda := -\log \rho > 0$ and let $\zeta(ds)$ be a random measure assigning to disjoint intervals $(a_i, b_i]$ independent random variables

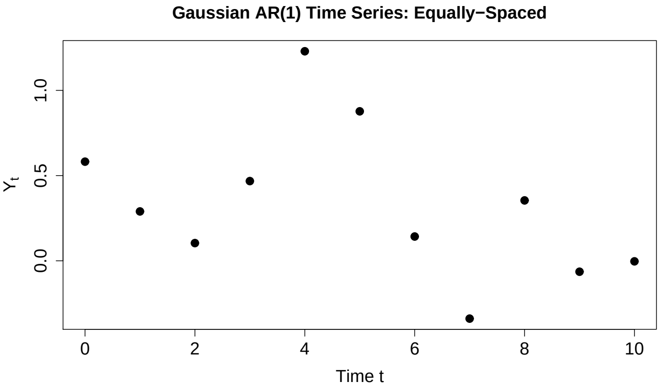
$$\zeta\big((a,b]\big) \sim \mathsf{No}\big((b-a)\lambda\mu,\ (b-a)2\lambda\sigma^2\big).$$

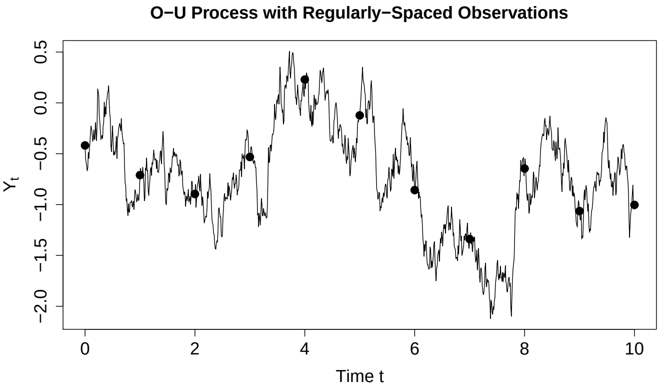
Now take $X_0 \sim \mathsf{No}(\mu, \sigma^2)$ and for $t > 0$ set

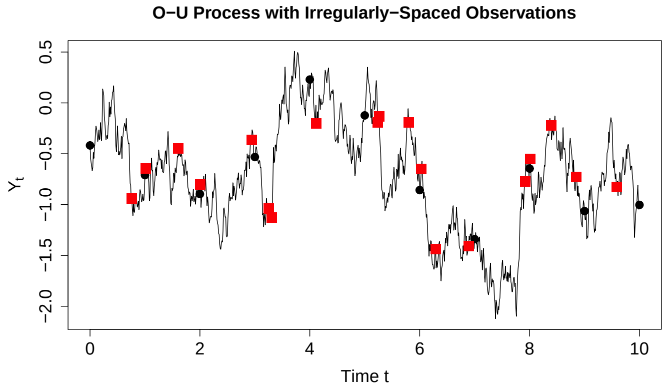
$$X_t := X_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)}\,\zeta(ds) = \int_{-\infty}^t e^{-\lambda(t-s)}\,\zeta(ds)$$

to find a process *for all* $t > 0$ (or all $t \in \mathbb{R}$) with the exact same joint distribution at $t \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$, with

- Irregularly-spaced observation times $\{t_i\} \subset \mathbb{R}_+$;
- Gaussian marginal distributions $X_t \sim \mathsf{No}(\mu, \sigma^2)$;
- Stationary distributions, with autocorrelation $\mathsf{Corr}(X_s, X_t) = \rho^{|t-s|}$.


Gaussian AR(1) Time Series: Equally−Spaced
1.0
0.5
0.0
Yt
0
2
4
6
8
10
Time t


O−U Process with Regularly−Spaced Observations
Y_t
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
2
4
6
8
10
Time t


O−U Process with Irregularly−Spaced Observations
0.5
0.0
−0.5
−1.0
−1.5
−2.0
Yt
0
2
4
6
8
10
Time t

Inference 1: Continuous Time

Inference is easy— MLEs or conjugate Bayes estimators are readily available, as is LH for arbitrary observation pairs $\{(X_{t_i}, t_i)\}$ because $s < t \Rightarrow$

$$X_t \mid \{X_u : \; u \le s\} \sim \mathsf{No}\big(\mu + (X_s - \mu)\rho, \sigma^2(1 - \rho^2)\big), \quad \rho := e^{-\lambda|t-s|}$$

so with Metropolis-Hastings we can sample posterior for any prior $\pi(\mu, \sigma^2, \lambda)$.

Easy extension to AR(p):

$$X_t + a_1 X_{t-1} + ... + a_p X_{t-p} := \zeta_t$$

For example, by expressing as a **vector** AR(1) for $\mathbf{X}_t = [X_t, X_{t-1}, \cdots, X_{t-p+1}]'$:

$$\mathbf{X}_t := R\mathbf{X}_{t-1} + \zeta_t$$

for some $a \in \mathbb{R}^p$

Extension 1: Continuous Time, More Broadly

Any stationary AR(p) has a MA(∞) representation:

$$p(L)X(t) = \zeta_t$$

for the left-shift operator $LX(t) := X(t-1)$ and the polynomial

$$p(z) = 1 + \sum_{j=1}^{p} a_j z^j = 1 + a_1 z + \cdots + a_p z^p.$$

If $p(z)$ has all its roots outside the unit circle, then

$$\frac{1}{p(z)} = 1 + \sum_{i=1}^{\infty} b_i z^i = 1 + b_1 z + b_2 z^2 + \cdots$$

$$X(t) = \frac{1}{p}(L)(\zeta_t) \qquad = \zeta_t + b_1 \zeta_{t-1} + b_2 \zeta_{z-2} + \cdots$$

Under suitable conditions, this has a continuous version

$$X(t) \qquad = \int_{-\infty}^{t} b(t-s)\zeta(ds)$$

for all $t \geq 0$, with the same distribution at times $t \in \mathbb{N}_0$.

Outline

Extension 2: Non-Gaussian

Just as AR(p) Time Series

$$X_t = \rho^t X_0 \quad + \sum_{s=0}^{t-1} \zeta_{t-s}\rho^s$$

can be constructed with *any* iid innovations $\{\zeta_s\}$, so too for the continuous-time version

$$X_t := X_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)}\, \zeta(ds).$$

The random measure $\zeta(ds)$ can be anything that is "iid" in the sense that:

- For disjoint sets $A_i \subset \mathbb{R}$, the RVs $\{\zeta(A_i)\}$ are indep;
- The $\{\zeta(a, b]\}$ distributions only depend on $(b - a)$.

For example, can have

$\zeta(a, b] \sim \mathsf{No}\big(\mu(b-a),\ \sigma^2(b-a)\big)$	Ornstein-Uhlenbeck
$\zeta(a, b] \sim \mathsf{Ga}\big(\alpha(b-a),\ \beta\big)$	Positive, Expo. Tails
$\zeta(a, b] \sim \mathsf{St}_{\mathsf{A}}\big(\alpha, \beta, \gamma(b-a), \delta(b-a)\big)$	Heavy (Pareto) Tails

Extension 2: Non-Gaussian

The two conditions

- For disjoint sets $A_i \subset \mathbb{R}$, the RVs $\{\zeta(A_i)\}$ are indep;
- The $\{\zeta(a, b]\}$ distributions only depend on $(b - a)$.

require that each $\zeta(A)$ should be Infinitely Divisible, or ID.

Examples:

ID Continuous	ID Discrete	Not ID
Normal	Poisson	Binomial
Gamma	Negative Binomial	Beta
α-Stable	$p_i \propto \frac{1}{(i+a)(i+b)}$	Uniform

Extension 2: Non-Gaussian

But what about Autocorrelated **Count** Data?

- Binned photon counts in satellite Gamma Ray detectors?
- Binned photon counts in particle accelerator detectors?
- Seismic event counts?
- Pyroclastic flow counts?
- Rockfall counts near active volcano?
- Failures of complex systems?
- Rare disease case counts?

AR(p) and its continuous versions wouldn't respect Integer Nature of data. If $X_t \in \mathbb{Z}$, and $|\rho| < 1$, then

$$X_{t+1} = \rho X_t + \zeta_{t+1} \notin \mathbb{Z}.$$

Alternatives?

Extension 2: Non-Gaussian

Here's a way to construct, and make inference in,

- stationary (for now) process X_t with
- continuous time $t \in \mathbb{R}_+$ for *any*
- non-Gaussian marginal **Infinitely Divisible (ID)** dist'ns including both continuous dist'ns (Ga(θ, β), St$_A(\alpha, \beta, \theta, \delta)$, *etc.*) and discrete **count** distributions (Po(θ), NB(θ, p), *etc.*)

Extension 2: Non-Gaussian

Here's a way to construct, and make inference in,

- stationary (for now) process X_t with
- continuous time $t \in \mathbb{R}_+$ for *any*
- non-Gaussian marginal **Infinitely Divisible (ID)** dist'ns including both continuous dist'ns ($\mathsf{Ga}(\theta, \beta)$, $\mathsf{St}_{\mathsf{A}}(\alpha, \beta, \theta, \delta)$, *etc.*) and discrete **count** distributions ($\mathsf{Po}(\theta)$, $\mathsf{NB}(\theta, p)$, *etc.*)

In fact, except for the special cases of Gaussian and Poisson, we can do this in many ways.

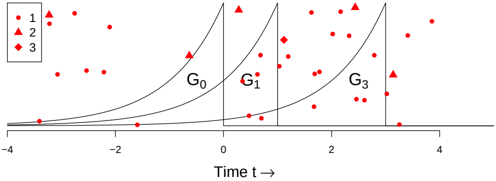
Extension 2: Non-Gaussian

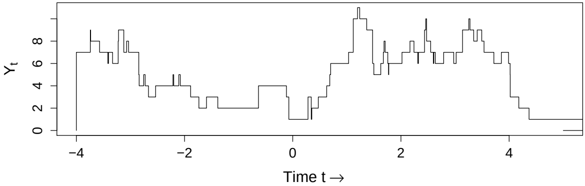
Here's a way to construct, and make inference in,

- stationary (for now) process X_t with
- continuous time $t \in \mathbb{R}_+$ for *any*
- non-Gaussian marginal **Infinitely Divisible (ID)** dist'ns including both continuous dist'ns ($\mathsf{Ga}(\theta, \beta)$, $\mathsf{St}_{\mathsf{A}}(\alpha, \beta, \theta, \delta)$, *etc.*) and discrete **count** distributions ($\mathsf{Po}(\theta)$, $\mathsf{NB}(\theta, p)$, *etc.*)

In fact, except for the special cases of Gaussian and Poisson, we can do this in many ways. Let's look at an example.

Ex 2: Negative Binomial Example





AR(1)-like Negative Binomial Process

Based on **Random Measure** $\mathcal{N}(dx\,dy) \sim \mathsf{NB}(\alpha\,dx\,dy, \beta)$ on $\mathbb{R}^2$

AR(1)-like Negative Binomial Process

Properties:

- $X_t \sim \mathsf{NB}(\alpha, \beta)$ for all t;
- $\mathsf{Corr}(X_s,\ X_t)\ = \exp\big(-\lambda|t-s|\big)$.

AR(1)-like Negative Binomial Process

Properties:

- $X_t \sim \mathsf{NB}(\alpha, \beta)$ for all t;
- $\mathsf{Corr}(X_s,\ X_t) \ = \exp\big(-\lambda|t-s|\big)$.

Does this characterize the joint distribution of $\{X_{t_i}\}$?

AR(1)-like Negative Binomial Process

Properties:

- $X_t \sim \mathsf{NB}(\alpha, \beta)$ for all t;
- $\mathsf{Corr}(X_s,\ X_t)\ = \exp\big(-\lambda|t-s|\big)$.

Does this characterize the joint distribution of $\{X_{t_i}\}$?

Nope. Here's a different process, the *Branching* NB:

- **Immigration** at rate ι;
- **Birth** at rate β;
- **Death** at rate δ;

$$X_{t+\epsilon} = \begin{cases} X_1 + 1 & \text{with probability } o(\epsilon) + \epsilon(\iota + \beta X_t) \\ X_1 & \text{with probability } o(\epsilon) + 1 - \epsilon(\iota + (\beta+\delta)X_t \\ X_1 - 1 & \text{with probability } o(\epsilon) + \epsilon\delta X_t \end{cases}$$

$$\sim \mathsf{NB}(\alpha, p),\ \text{with autocovariance}\ \exp\big(-\lambda|t-s|\big).$$

AR(1)-like Negative Binomial Process

Are these two processes the same?

AR(1)-like Negative Binomial Process

Are these two processes the same?

Nope.

- The Random Measure NB process has jumps of all possible non-zero magnitudes $\Delta X_t := [X_t - X_{t-}] \in \mathbb{Z} \backslash \{0\}$, while the Branching NB process has only jumps of $\Delta X_t = \pm 1$;
- The Branching NB process is Markov, so for $s < t$

$$\mathsf{P}[X_t \in A \mid \mathcal{F}_s] = \mathsf{P}[X_t \in A \mid X_s],$$

  while the Random Measure NB process isn't: if it has a jump $\Delta X_s = 7$, for example, then sooner or later there must follow a jump $\Delta X_t = -7$, so $\mathsf{P}[X_t = 3 \mid X_s = 10]$ depends on the history $\mathcal{F}_s$, not just the value X_s.
- The Random Measure NB process is **multivariate** ID, while Branching NB process is not.

AR(1)-like Negative Binomial Process

Are these the only two?

AR(1)-like Negative Binomial Process

Are these the only two?

Oh no.

- A discrete time AR(1)-like Markov process exists for any ID distribution and any "auto-correlation" ρ, based on Thinning:

$$
\begin{aligned}
Z &\sim f(z \mid \lambda, \phi) = X + Y, \\
X &\sim f(x \mid \rho\lambda, \phi) \perp\!\!\!\perp Y \sim f(y \mid \bar{\rho}\lambda, \phi) \quad [\bar{\rho} := (1-\rho)] \\
X \mid Z &\sim f(x \mid z, \rho, \lambda, \phi) = f(x \mid \rho\lambda, \phi) f(z - x \mid \bar{\rho}\lambda, \phi) / f(z \mid \lambda, \phi)
\end{aligned}
$$

- Given $\{X_m : \ m \le n\} \sim f(x \mid \lambda, \phi)$, draw

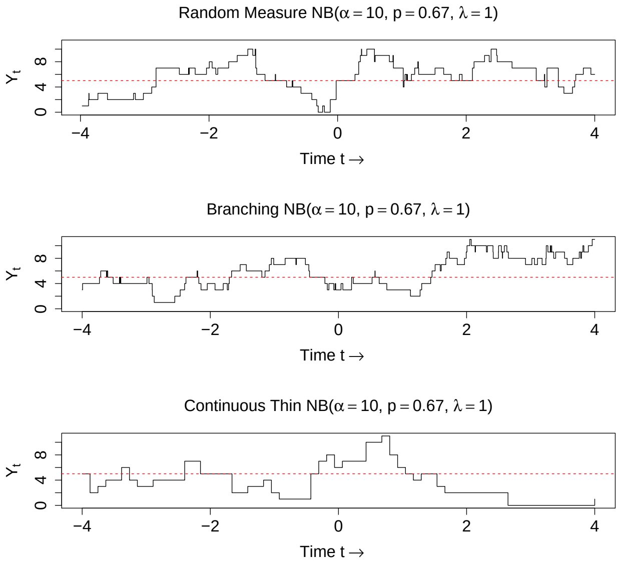
$$\xi_{n+1} \sim f(\xi \mid X_n, \rho, \lambda, \phi) \perp\!\!\!\perp \eta_{n+1} \sim f(\eta \mid \bar{\rho}\lambda, \phi)$$

  and set

$$X_{n+1} := \xi_{n+1} + \eta_{n+1} \sim f(x \mid \lambda, \phi).$$

- A continuous-time version exists too.

AR(1)-like Negative Binomial Process



AR(1)-like Negative Binomial Process

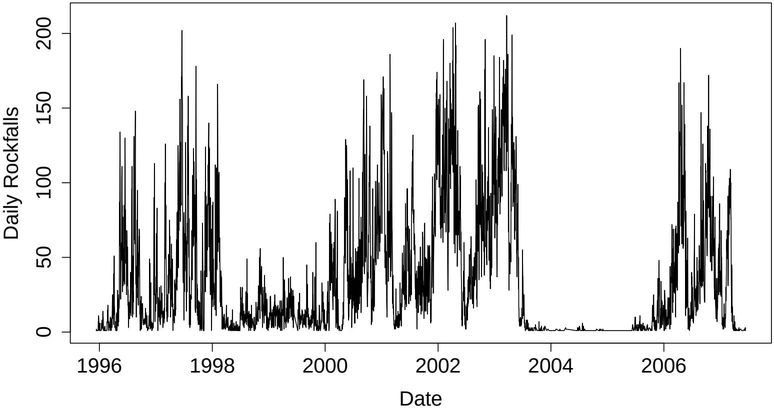
Larry Brown (U Penn) & I found six different AR(1)-like (also AR(p)-like) processes for each ID distribution, with subtly different properties:

- Markov?
- Are *all* finite-dimensional marginals ID?
- In continuous time, are paths continuous? Increments ± 1? Or bigger?
- Time-reversible?

In Wolpert & L. D. Brown (2016+) we present a complete class theorem characterizing all *Markov Infinitely-Divisible Stationary Time-Reversible Integer-Valued* ("MISTI") Processes.

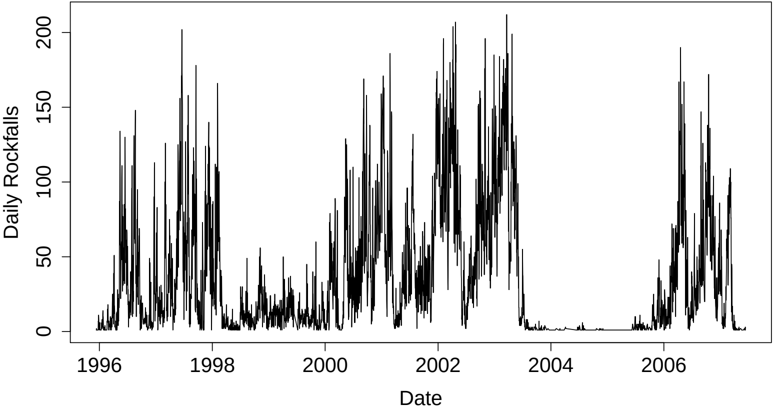
Let's look at a motivating example (not Astro, sorry!).

Motivation for a Negative Binomial series...



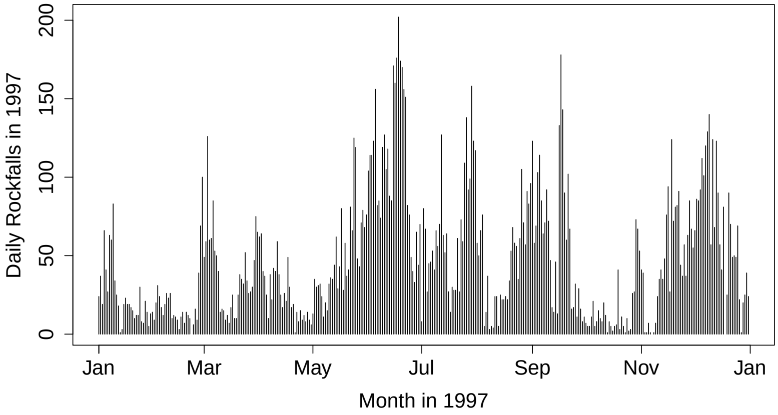
Rockfall counts at Soufriére Hills Volcano, Montserrat

Stationary?

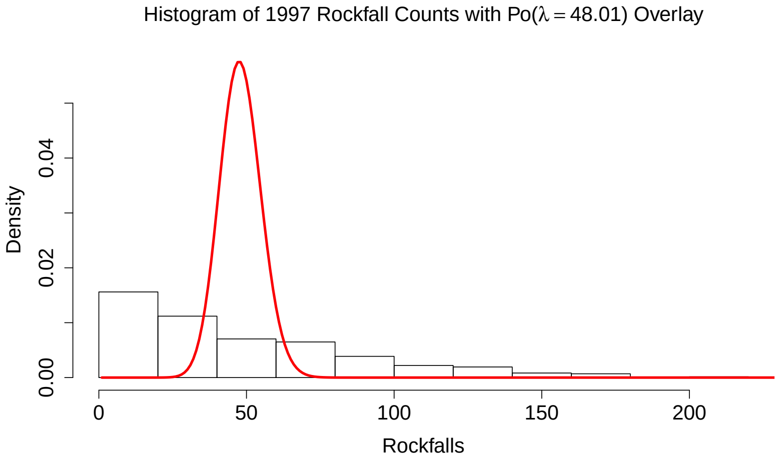


Looks a little patchy; maybe we can find a homogeneous subset. Looks like 1997 might be a less patchy year...

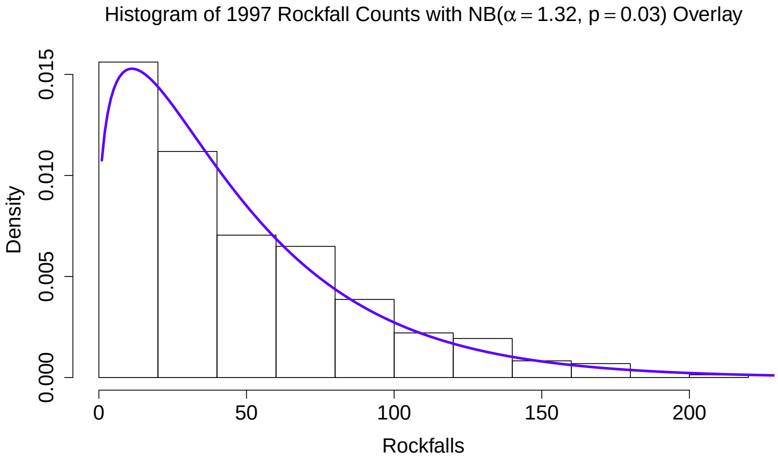
Subset, just the 1997 Rockfalls



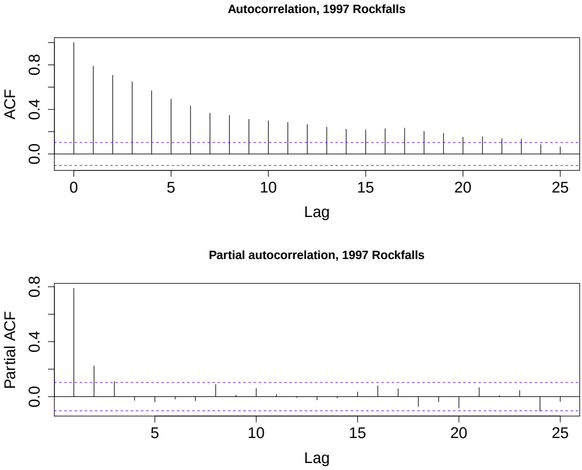
Poisson fit is horrible...



Negative Binomial fit is great.



But not IID, need autocorrelation



SO,

We (esp. **Jianyu Wang** '13) built a regression model based on one of the AR(1)-like NB models above, to relate

→ (easily counted) rockfall counts   to

→ (hard to measure) subsurface volcanic magma flows,

to improve volcanic hazard forecasting.

Related issues were addressed by **Mary Beth Broadbent** ('14) in constructing NPB light curve models for Gamma Ray Bursts, with the help of astronomers **Tom Loredo** and **Jon Hakkila** (Thanks!).

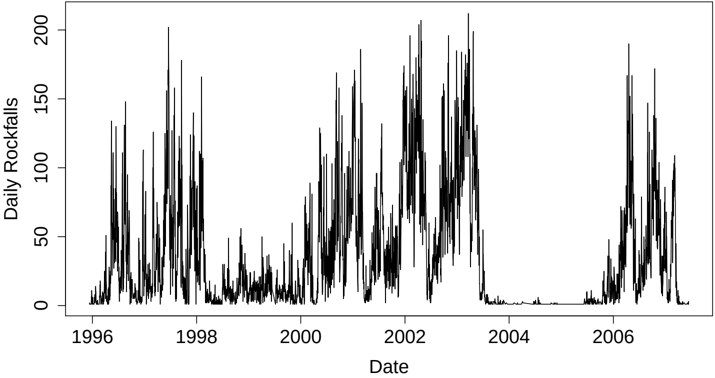
A shout out about **Another cool modeling idea:**

Robert Lund (Clemson) and his student Yunwei Cui (2009) found an interesting new way to model stationary integer-valued time series and processes using Renewal Theory. In discrete time their method (unlike ours) is able to model negative autocorrelation and cyclic behavior. If you have negatively-autocorrelated integer data, look into it.

Outline

Extension 3: Non-Stationary

You might have noticed that the rockfall counts don't look stationary



SO, we are implementing **random time-change**.

Extension 3: Non-Stationary

Begin with stationary process $X(t)\ \cdots$
Then, construct random time-change model $t \to R_t$ and set

$$Y_t := X(R_t)$$

We used:

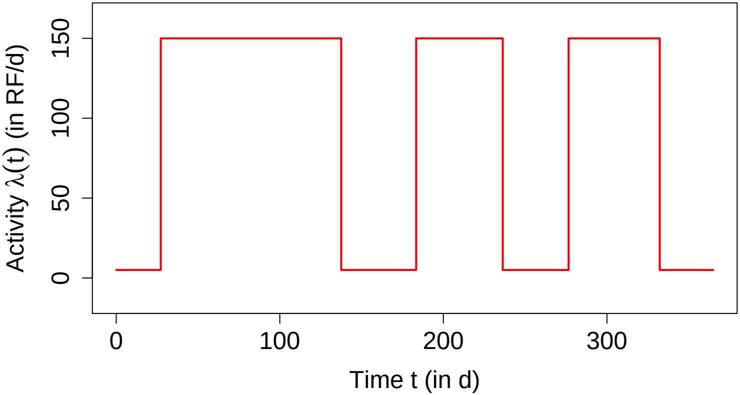
$$R_t = \int_0^t \lambda(s)\,ds,$$

with two-level

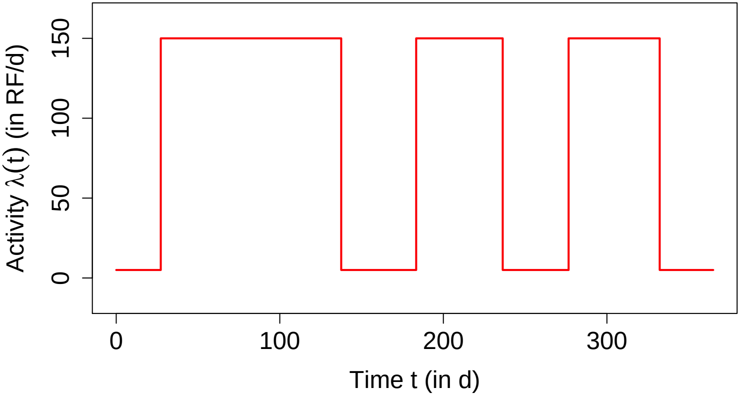
$$\lambda(s) = \begin{cases} \lambda_+ & s_i < s \le t_i \\ \lambda_- & t_i < s \le s_{i+1} \end{cases}$$

for uncertain levels $0 < \lambda_- < \lambda_+ < \infty$ and transition times $s_1 < t_1 < s_2 < t_2 < \ldots$.

Extension 3: Non-Stationary



Extension 3: Non-Stationary



A work in progress, wish us luck.

Outline

Wrap-up

Conclusions:

- Statistical methods are available for data that are non-Stationary, or non-Gaussian, or irregularly sampled, or all three.
- They're not built into SAS.
- BUT, together, Astronomers and Statisticians can build problem-specific tools to support *estimation* and *prediction* and (especially Bayesian) *inference*,
- using routine simulation-based computational methods.

Thanks!

Thanks to the **Organizers**, to the **National Science Foundation** (MPS, PHY, GEO), and to **SAMSI**'s 2016-17 Program on

Statistical, Mathematical and Computational Methods for Astronomy (ASTRO)

Glad to see you here!