Bayesian Statistics

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Alfonso Corti 12, I-20133, Milano, Italy, European Union
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/fabrizio/

BASICS ON STOCHASTIC PROCESSES

- A stochastic process $\{X_t, t \in T\}$ is a collection of random variables X_t, indexed by a set T, defined on a sample space Ω, endowed with a σ-algebra $\mathcal{F}$ and a base probability measure $\mathbf{P}$, and taking values in a common measurable space S endowed with an appropriate σ-algebra

- T set of
  - times $\Rightarrow$ temporal stochastic process
  - spatial coordinates $\Rightarrow$ spatial process
  - both time and spatial coordinates $\Rightarrow$ spatio-temporal process

- T discrete $\Rightarrow$ process *in discrete time*, represented through $\{X_n,\ n = 0, 1, 2, ..\}$

- T continuous $\Rightarrow$ process in *continuous time*, e.g. $T = [0, \infty)$

- Values taken by process $\Rightarrow$ *states* of the process, belonging to the *state space* S, which may be either discrete or continuous

BASICS ON STOCHASTIC PROCESSES

$\{X_t, t \in T\}$ stochastic process

- Mean function: $\mu_X(t) = E[X_t]$
- Autocorrelation function: $R_X(t_1, t_2) = E[X_{t_1} X_{t_2}]$
- $\{X_t, t \in T\}$ strictly stationary if $(X_{t_1}, \ldots, X_{t_n})$ and $(X_{t_1+\tau}, \ldots, X_{t_n+\tau})$ have the same distribution for any n, $t_1, t_2, \ldots, t_n$ and τ
  - $n = 1 \Rightarrow X_t$'s have the same distribution
  - $n = 2 \Rightarrow$ joint distribution depends on difference between times and not the times themselves, i.e. $F_{X_{t_1}, X_{t_2}}(x_1, x_2) = F_{X_0, X_{t_2 - t_1}}(x_1, x_2)$
- $\{X_t, t \in T\}$ weakly stationary if
  - Constant mean function: $\mu_X(t) = \mu_X, \forall t$
  - Autocorrelation function depends on time differences: $R_X(t_1, t_2) = R(t_2 - t_1)$

DISCRETE TIME MARKOV CHAINS

- A stochastic process $\{X_n\}$, discrete in time and with a countable/finite state space, is a Markov chain if, for any $n > n_1 > \cdots > n_k$, and $j, i_1, \ldots, i_{n_k}$, we have

$$\begin{aligned} P\left(X_n = j \mid X_{n_1} = i_1, X_{n_2} = i_2, ..., X_{n_k} = i_{n_k}\right) &= \\ P\left(X_n = j \mid X_{n_1} = i_1\right) &= p_{i_1 j}^{(n_1,n)}. \end{aligned}$$

- One step *transition probability*: $p_{ij}^{(m,m+1)} = P\left(X_{m+1} = j \mid X_m = i\right)$

- $p_{ij}^{(m,m+1)}$ independent of $m \Rightarrow$ stationary process and *time homogeneous* chain

- n*-step transition probability matrix* defined as $\mathbf{P}^{(n)}$, with elements p_{ij}^n

- Matrices $\mathbf{P}^{(n)}$ characterize the transition behavior of an homogeneous Markov chain

- When $n = 1$, we refer to the *transition matrix* instead of the one step transition matrix and write $\mathbf{P}$ instead of $\mathbf{P}^{(1)}$

INFERENCE FOR FINITE, TIME HOMOGENEOUS MARKOV CHAINS

- Transition matrix $\mathbf{P} = (p_{ij})$ where $p_{ij} = P(X_n = j | X_{n-1} = i)$, for states $i, j \in \{1, \dots, K\}$

- If it exists, stationary distribution $\boldsymbol{\pi}$ unique solution of $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, $\pi_i \geq 0, \sum \pi_i = 1$

- We consider the simple experiment of observing m successive transitions of the Markov chain, say $X_1 = x_1, \dots, X_m = x_m$, given a known initial state $X_0 = x_0$

- Likelihood function $l(\mathbf{P}|\mathbf{x}) = \prod_{i=1}^{K} \prod_{j=1}^{K} p_{ij}^{n_{ij}}$ with $n_{ij} \geq 0$ number of observed transitions from state i to state j and $\sum_{i=1}^{K} \sum_{j=1}^{K} n_{ij} = m$

- $\widehat{\mathbf{P}}$ MLE for $\mathbf{P}$, with $\widehat{p}_{ij} = \dfrac{n_{ij}}{n_{i\cdot}}$, where $\displaystyle n_{i\cdot} = \sum_{j=1}^{K} n_{ij}$

- However, especially in chains with large K, there could be some $\widehat{p}_{ij} = 0$

BAYESIAN INFERENCE

- Dirichlet density: $f(x_1, \ldots, x_K; \alpha_1, \ldots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$, with $B(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}$

- Conjugate prior for $\mathbf{P}$ defined by letting $\mathbf{p}_i = (p_{i1}, \ldots, p_{iK}) \sim \mathcal{D}ir(\boldsymbol{\alpha}_i)$, where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iK})$ for $i = 1, \ldots, K \Rightarrow$ matrix beta prior distribution

- $\Rightarrow$ posterior $\mathbf{p}_i | \mathbf{x} \sim \mathcal{D}ir\left(\boldsymbol{\alpha}'_i\right)$ where $\alpha'_{ij} = \alpha_{ij} + n_{ij}$ for $i, j = 1, \ldots, K$

- Jeffreys prior: matrix beta prior with $\boldsymbol{\alpha}_{ij} = 1/2$ for all $i, j = 1, \ldots, K$

- Other improper prior: $f(\mathbf{p}_i) \propto \prod_{j=1}^{K} \frac{1}{p_{ij}}$, i.e. with $\alpha_{ij} \to 0$ for all $i, j = 1, \ldots, K$
  - Posterior distribution $\mathbf{p}_i | \mathbf{x} \sim \mathcal{D}ir(n_{i1}, \ldots, n_{ik})$ would imply a posterior mean $E[p_{ij} | \mathbf{x}] = n_{ij} / n_{i\cdot}$ equal to MLE
  - Improper posterior distribution if there are any $n_{ij} = 0$

- A proper Dirichlet prior

SYDNEY BOTANIC GARDENS WEATHER CENTER

Rainfall levels (from `weatherzone.com.au`) illustrate occurrence (2) or non occurrence (1) of rain between February 1st and March 20th 2008. The data are to be read consecutively from left to right. Thus, it rained on February 1st and did not rain on March 20th.

2	2	2	2	2	2	2	2	2	2
1	1	2	1	1	1	1	1	1	1
2	2	1	1	1	1	2	2	2	1
2	1	1	1	1	1	2	1	1	1
1	1	1	1	1	1	1	1	1	

The daily occurrence of rainfall is modeled as a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}.$$

Given a Jeffreys prior, $p_{ii} \sim \mathcal{B}e\,(1/2, 1/2)$, for $i = 1, 2$, then conditioning on the occurrence of rainfall on February 1st, the posterior distribution is

$$p_{11}|\mathbf{x} \sim \mathcal{B}e\,(25.5, 5.5) \quad p_{22}|\mathbf{x} \sim \mathcal{B}e\,(12.5, 6.5).$$

The expectation of the transition matrix is $E[\mathbf{P}|\mathbf{x}] = \begin{pmatrix} 0.823 & 0.177 \\ 0.342 & 0.658 \end{pmatrix}$

FORECASTING SHORT TERM BEHAVIOR

- Suppose that we wish to predict future values of the chain
- For example we can predict the next value of the chain, at time $n+1$, using

$$\begin{aligned} P(X_{n+1}=j|\mathbf{x}) &= \int P(X_{n+1}=j|\mathbf{x},\mathbf{P}) f(\mathbf{P}|\mathbf{x})\, d\mathbf{P} \\ &= \int p_{x_n j} f(\mathbf{P}|\mathbf{x})\, d\mathbf{P} = \frac{\alpha_{x_n j}+n_{x_n j}}{\alpha_{x_n \cdot}+n_{x_n \cdot}} \end{aligned}$$

  where $\alpha_{i\cdot} = \sum_{j=1}^{K} \alpha_{ij}$.

- Prediction of state at $t>1$ steps is slightly more complex. For small t, use

$$P(X_{n+t}=j|\mathbf{x}) = \int \left(\mathbf{P}^t\right)_{x_n j} f(\mathbf{P}|\mathbf{x})\, d\mathbf{P}$$

  which gives a sum of Dirichlet expectation terms. However, as t increases, the evaluation of this expression becomes computationally infeasible.

- A simple alternative is to use a Monte Carlo algorithm based on simulating future values of the chain

FORECASTING SHORT TERM BEHAVIOR

- `For` $s = 1, \ldots, S$:

  `Generate` $\mathbf{P}^{(s)}$ `from` $f(\mathbf{P}|\mathbf{x})$.

  `Generate` $x_{n+1}^{(s)}, \ldots, x_{n+t}^{(s)}$ `from the Markov chain with` $\mathbf{P}^{(s)}$ `and initial state` x_n.

- Then, $P(X_{n+t} = j|\mathbf{x}) \approx \frac{1}{S}\sum_{s=1}^{S} I\left(x_{n+t}^{(s)} = j\right)$ where $I(\cdot)$ is an indicator function and $E[X_{n+t}|\mathbf{x}] \approx \frac{1}{S}\sum_{s=1}^{S} x_{n+t}^{(s)}$.

- Assume that it is now wished to predict the Sydney weather on the 21st and 22nd of March. Given that it did not rain on the 20th March, then immediately, we have

$$\begin{aligned} P(\text{no rain on 21st March}|\mathbf{x}) &= E[p_{11}|\mathbf{x}] = 0.823, \\ P(\text{no rain on 22nd March}|\mathbf{x}) &= E\left[p_{11}^2 + p_{12}p_{21}|\mathbf{x}\right] = 0.742, \\ P(\text{no rain on both}) &= E[p_{11}^2|\mathbf{x}] = 0.681. \end{aligned}$$

TIME SERIES

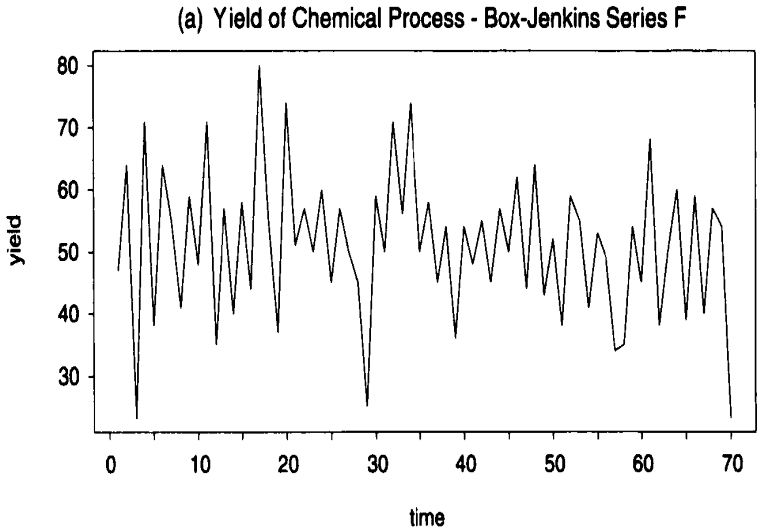
- Data in business, economics, engineering, environment, medicine, and other areas of scientific investigations are often collected in the form of time series, that is, a sequence of observations taken at regular intervals of time such as hourly temperature readings, daily stock prices, weekly traffic volume, monthly beer consumption, and annual growth rates

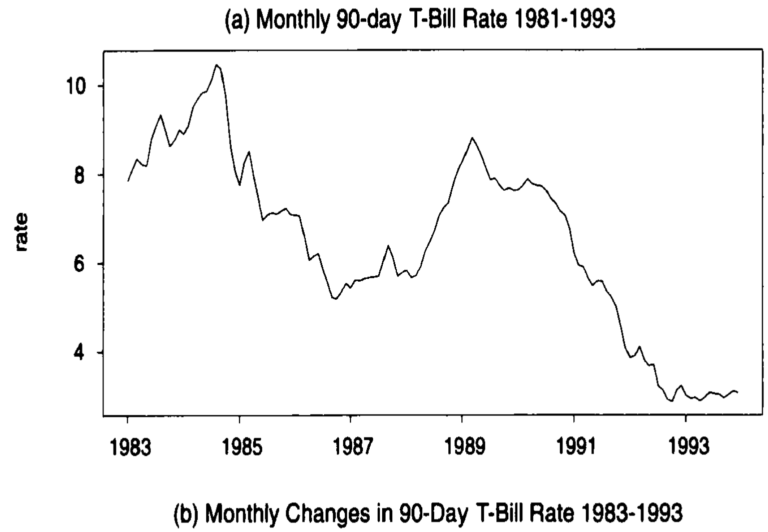
- The main objectives of time series modeling and analysis are
  - understanding the dynamic or time-dependent structure of the observations of a single series (univariate time series analysis)
  - ascertaining the leading, lagging, and feedback relationships among several series (multivariate time series analysis)

- Knowledge of the dynamic structure will help produce accurate forecasts of future observations and design optimal control schemes

- We now present some plots about time series, briefly discussing some relevant features

TIME SERIES: DIFFERENT FEATURES





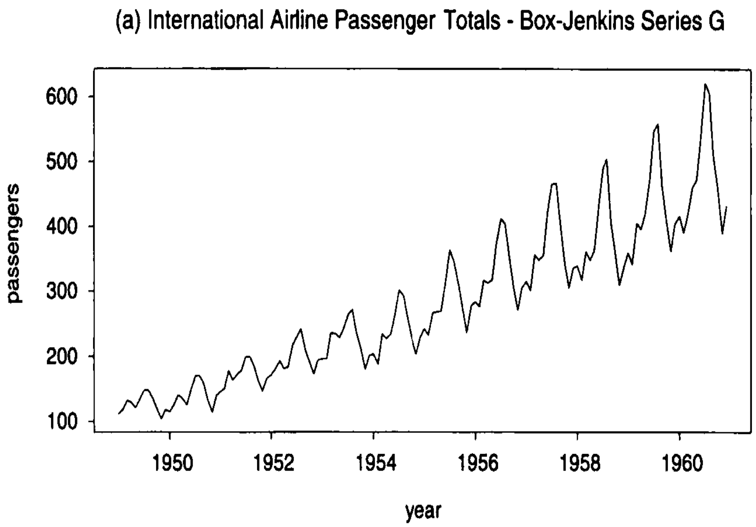
- Left: Yield of 70 consecutive batches of a chemical process

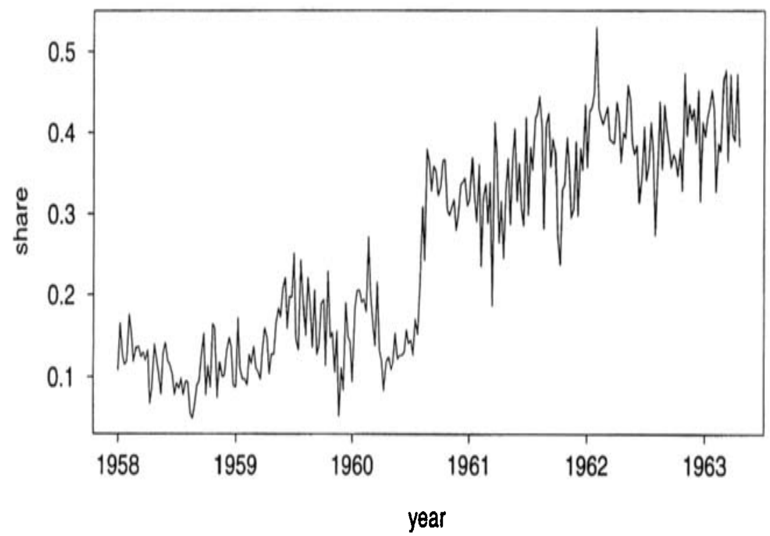
  $\Rightarrow$ **Stationary series** since the behaviour of the series remains the same over time, fluctuating about a fixed mean level with constant variance

- Right: Interest rates of 90-day U.S. Treasury bills (T-bills)

  $\Rightarrow$ **Nonstationary series** since the series does not seem to have a mean level and exhibits a drifting or wandering behaviour

TIME SERIES: DIFFERENT FEATURES





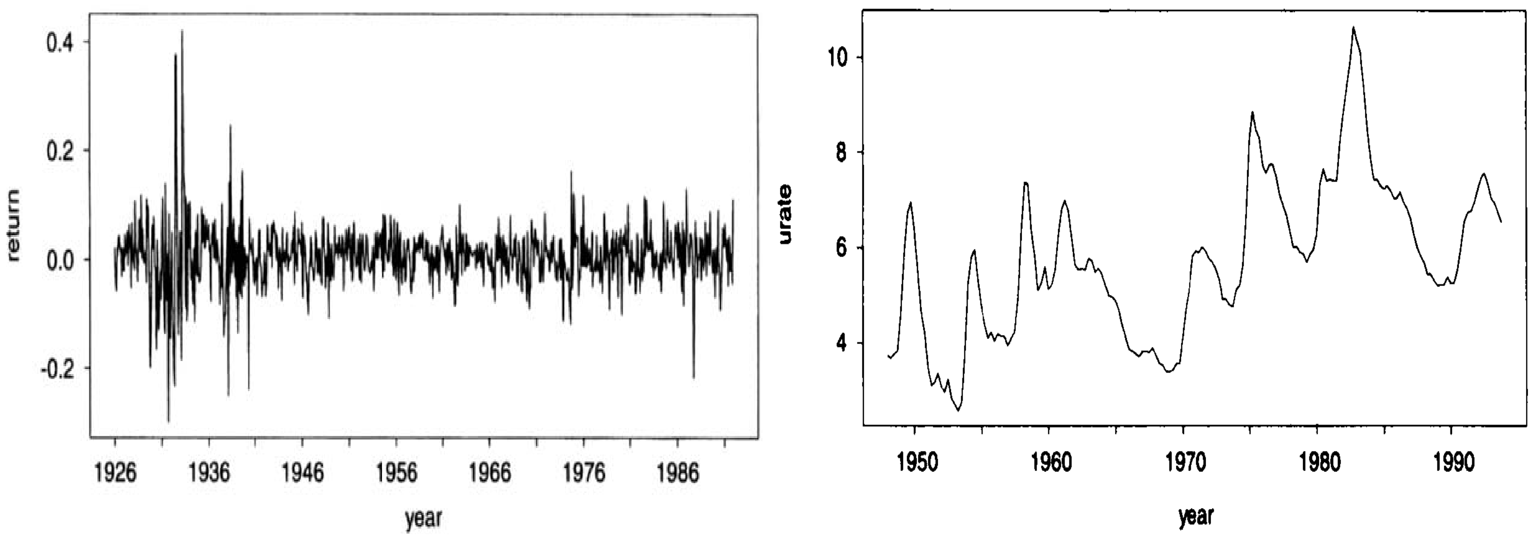
- Left: Monthly international airline passenger totals

  $\Rightarrow$ **Seasonality** as shown by the oscillations, repeated over time
  $\Rightarrow$ **Trend** as shown by the increasing pattern

- Right: Weekly market share data of Crest toothpaste

  $\Rightarrow$ **Level shift** in August 1960 due to the endorsement of Crest by the American Dental Association

TIME SERIES: DIFFERENT FEATURES



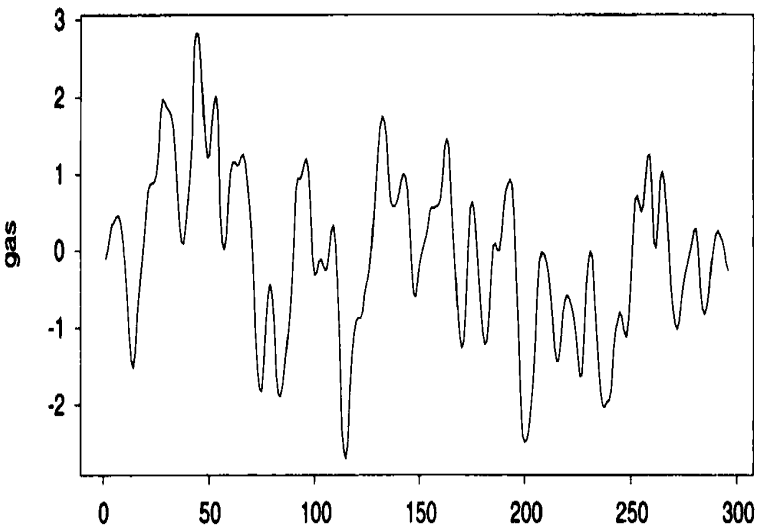
- Left: Monthly returns of value-weighted Standard and Poor 500 stocks

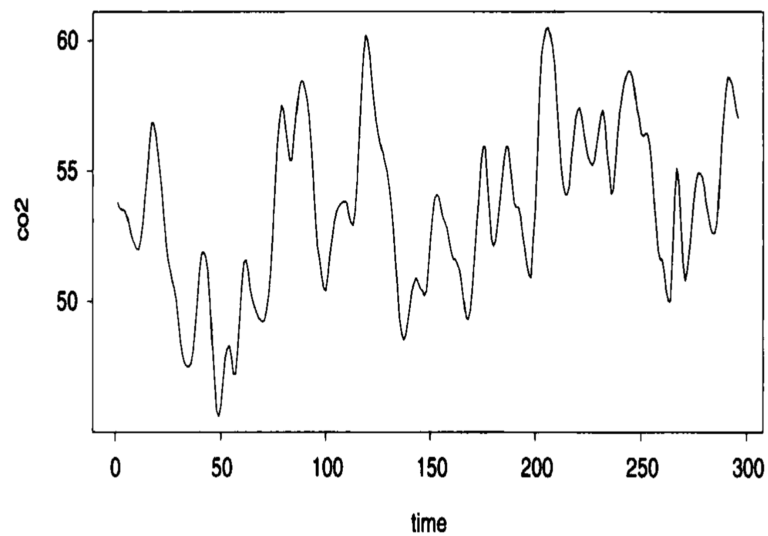
  $\Rightarrow$ **Variance changes** or volatility (in Finance) as shown by some jumps away from the zero mean every once in a while
  [Return: (current month value - past month value)/past month value]

- Right: Seasonally adjusted quarterly U.S. unemployment rates

  $\Rightarrow$ **Asymmetry** as shown by different behaviours in the rise and fall of the observations

TIME SERIES: DIFFERENT FEATURES

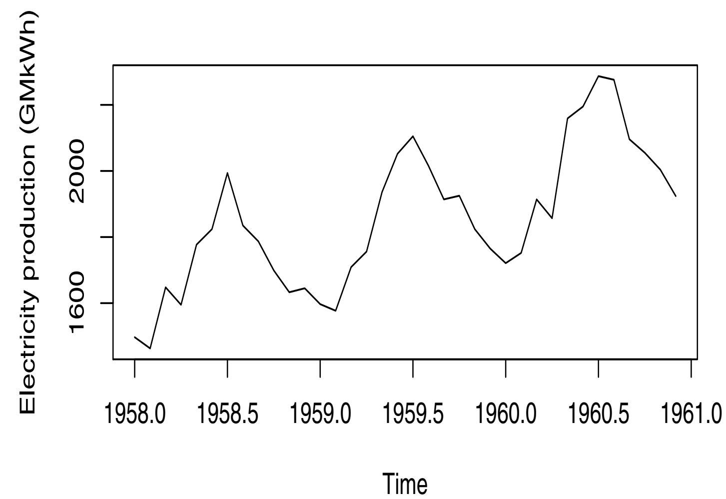




- Input gas rate (Left) and output CO_2 (Right) of a chemical reactor

  $\Rightarrow$ **Relation between series** since when one goes up the other comes down (actually, the input series anticipates the output series by some periods and, therefore, the former can be used to forecast the latter)

TIME SERIES: DIFFERENT FEATURES



- International air passengers (Left) and Australian electricity production (Right)

  $\Rightarrow$ **No causal relation between series** despite the similarity between the two plots, both showing an increasing trend and a seasonal cycle

TIME SERIES: AR(1)

- $x_t = \theta x_{t-1} + w_t$, with $w_t \sim \mathcal{N}(0,1)$
- What happens if I change θ?
- Try $\theta = 0.5, -0.5, 0.9, -0.9, 1, -1, 2, -2$
- θ plays a relevant role, as we will see later

```
x<-rnorm(100)
par(mfrow=c(2,1))
theta=0.5
z<-rep(0,101)
for (i in 1:100) {z[i+1]<-(theta*z[i]+x[i])}; plot(z,type='l')
theta=-theta
z<-rep(0,101)
for (i in 1:100) {z[i+1]<-(theta*z[i]+x[i])}; plot(z,type='l')
```

TIME SERIES: AR(1)

- A time series $\{x_t; t = 0, \pm 1, \pm 2, \ldots\}$ is an **autoregressive moving average model of order** (p, q), $ARMA(p, q)$ if it is stationary and

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q},$$

  with $\phi_p \neq 0, \theta_q \neq 0$, $w_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d. for all t and $\sigma^2 > 0$. The parameters p and q are called the autoregressive and the moving average orders, respectively

- We will consider one of the simplest models, $AR(1)$, **autoregressive of order 1**

- $\Rightarrow x_t = \theta x_{t-1} + w_t$, with $w_t \sim \mathcal{N}(0, \sigma^2)$

- This is also a Markov chain since x_t depends on the past only through x_{t-1}

- θ plays a relevant role, as seen earlier when using R
  - The time series is stationary if and only if $|\theta| < 1$
  - A random walk is obtained when $\theta = 1$
  - The correlation between x_t and x_{t-h} is θ^h, e.g. $Corr(x_t, x_{t-1}) = \theta$

TIME SERIES: AR(1)

- We suppose x_0 known and consider a sample $x_t, \dots, x_1$
- We model the data with an $AR(1)$ time series: $x_t = \theta x_{t-1} + w_t$, with $|\theta| < 1$ and $w_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d. for all t
- The parameters to be estimated are θ and σ^2
- Likelihood

$$
\begin{aligned}
f(x_t, \dots, x_1 | \theta, \sigma^2, x_0) &= \prod_{i=1}^{t} f(x_i | x_{i-1}, \dots, x_1, \theta, \sigma^2, x_0) \\
&= \prod_{i=1}^{t} f(x_i | x_{i-1}, \theta, \sigma^2, x_0) \\
&= \prod_{i=1}^{t} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(x_i - \theta x_{i-1})^2\} \\
&\propto \frac{1}{(\sigma^2)^{t/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t} x_i^2 - 2\theta \sum_{i=1}^{t} x_i x_{i-1} + \theta^2 \sum_{i=1}^{t} x_{i-1}^2\right]\right\}
\end{aligned}
$$

TIME SERIES: AR(1)

- We are considering a stationary time series, with $|\theta| < 1$, and Bayesians can imposing such condition very easily through a prior constrained to the interval $(-1, 1)$

- Priors: Uniform $\theta \sim \mathcal{U}(-1, 1)$ and Inverse Gamma $\sigma^2 \sim \mathcal{IG}(a, b)$

- Conditional posteriors:

  – $\theta | x_t, \ldots, x_0, \sigma^2 \sim \mathcal{N}(\frac{\sum_{i=1}^t x_i x_{i-1}}{\sum_{i=1}^t x_{i-1}^2}, \frac{\sigma^2}{\sum_{i=1}^t x_{i-1}^2})$ truncated on $(-1, 1)$

  – $\sigma^2 | x_t, \ldots, x_0, \theta \sim \mathcal{IG}(a + t/2, b + \sum_{i=1}^{t} (x_i - \theta x_{i-1})^2/2)$

- A Gaussian distribution with cdf $\Phi(x|\mu, \tau^2)$ and df $\phi(x|\mu, \tau^2)$ has a density, when truncated to the interval (a, b), given by $\frac{\phi(x|\mu, \tau^2)}{\Phi(b|\mu, \tau^2) - \Phi(a|\mu, \tau^2)}$

- The parameters can be estimated by using Gibbs sampling

POISSON PROCESS

- One of the simplest and most applied stochastic processes
- Used to model occurrences (and counts) of rare events in time and/or space, when they are not affected by past history
- Applied to describe and forecast incoming telephone calls at a switchboard, arrival of customers for service at a counter, occurrence of accidents at a given place, visits to a website, earthquake occurrences and machine failures, to name but a few applications
- Simple mathematical formulation and relatively straightforward statistical analysis
$\Rightarrow$ very practical, if approximate, model for describing and forecasting many random events

POISSON PROCESS

- Counting process $N(t)$, $t \geq 0$: stochastic process counting number of events occurred up to time t

- $N(s, t]$, $s < t$: number of events occurred in time interval $(s, t]$

- Poisson process with intensity function $\lambda(t)$: counting process $N(t), t \geq 0$, s.t.

  1. $N(0) = 0$
  2. Independent number of events in non-overlapping intervals
  3. $P\left(N(t, t + \Delta t] = 1\right) = \lambda(t)\Delta t + o(\Delta t)$, as $\Delta t \to 0$
  4. $P\left(N(t, t + \Delta t] \geq 2\right) = o(\Delta t)$, as $\Delta t \to 0$

- Definition $\Rightarrow P(N(s, t] = n) = \dfrac{(\int_s^t \lambda(x)dx)^n}{n!} e^{-\int_s^t \lambda(x)dx}$, for $n \in \mathbb{Z}^+$

  $\Rightarrow N(s, t] \sim \mathcal{P}\left(\int_s^t \lambda(x)dx\right)$

POISSON PROCESS

- Intensity function: $\lambda(t) = \lim_{\Delta t \to 0} \dfrac{P(N(t, t + \Delta t] \geq 1)}{\Delta t}$
  - HPP (homogeneous Poisson process): constant $\lambda(t) = \lambda$, $\forall t$
  - NHPP (nonhomogeneous Poisson process): o.w.

- HPP with rate λ
  - $N(s, t] \sim \mathcal{P}\left(\lambda(t - s)\right)$
  - Stationary increments (distribution dependent only on interval length)
  - Interarrival times, and first arrival time, have an exponential distribution $\mathcal{E}\left(\lambda\right)$ ($\Rightarrow$ HPP renewal process)
  - n-th arrival time, T_n, has a gamma distribution $\mathcal{G}\left(n, \lambda\right)$, for each $n \geq 1$

POISSON PROCESS

- Mean value function $m(t) = E[N(t)], t \geq 0$
- $m(s,t] = m(t) - m(s)$ expected number of events in $(s,t]$
- If $m(t)$ differentiable, $\mu(t) = m'(t)$, $t \geq 0$, Rate of Occurrence of Failures (ROCOF)
- $P\left(N(t, t + \Delta t] \geq 2\right) = o(\Delta t)$, as $\Delta t \to 0$
  $\Rightarrow$ orderly process
  $\Rightarrow \lambda(t) = \mu(t)$ a.e.
- $\Rightarrow m(t) = \int_0^t \lambda(x)dx$ and $m(s,t] = \int_s^t \lambda(x)dx$
- $\Rightarrow m(t) = \lambda t$ and $m(s,t] = \lambda(t-s)$ for HPP with rate λ

POISSON PROCESS

Poisson process $N(t)$ with intensity function $\lambda(t)$ and mean value function $m(t)$

- $T_1 < \ldots < T_n$: n arrival times in $(0, T] \Rightarrow P(T_1, \ldots, T_n) = \prod_{i=1}^{n} \lambda(T_i) \cdot e^{-m(T)}$

  $\Rightarrow$ likelihood

- $\Rightarrow P(T_1, \ldots, T_n) = \lambda^n e^{-\lambda T}$ for HPP with rate λ

- n events occur up to time $t_0 \Rightarrow$ distributed as order statistics from cdf $m(t)/m(t_0)$, for $0 \leq t \leq t_0$ (uniform distribution for HPP)

POISSON PROCESS: INFERENCE

- $N(t)$ HPP with parameter λ
- n events observed in the interval $(0, T]$
- Likelihood for two possible experiments
  - Times $T_1, \ldots, T_n$ available

    Theorem on Poisson processes $\Rightarrow l(\lambda|data) = \lambda^n e^{-\lambda T}$
  - Only number n available

    Properties of $\mathcal{P}(\lambda T) \Rightarrow l(\lambda|data) = \dfrac{(\lambda T)^n}{n!} e^{-\lambda T}$
- Proportional likelihoods $\Rightarrow$ same inferences (Likelihood Principle, Berger and Wolpert, 1988)
- In both cases, likelihood not dependent on the actual occurrence times but only on their number

POISSON PROCESS: INFERENCE

- Gamma priors conjugate w.r.t. λ in the HPP
- Prior $\mathcal{G}(\alpha, \beta)$
- $\Rightarrow f(\lambda|n,T) \propto \lambda^n e^{-\lambda T} \cdot \lambda^{\alpha-1} e^{-\beta\lambda}$
- $\Rightarrow$ posterior $\mathcal{G}(\alpha+n, \beta+T)$
- Posterior mean $\widehat{\lambda} = \dfrac{\alpha+n}{\beta+T}$
- Posterior mean combination of
  - Prior mean $\widehat{\lambda}_P = \dfrac{\alpha}{\beta}$
  - MLE $\widehat{\lambda}_M = \dfrac{n}{T}$

ACCIDENTS IN THE CONSTRUCTION SECTOR

Ríos Insua et al (1999)

- Interest in number of accidents in some companies in the Spanish construction sector
- 75 accidents and an average number of workers of 364 in 1987 for one company
- Number of workers constant during the year
- Times of all accidents of each worker are recorded
- Accidents occur *randomly* $\Rightarrow$ HPP model justified
- Each worker has the same propensity to have accidents $\Rightarrow$
  - HPP with same λ for all of them
  - If one year corresponds to $T = 1$ $\Rightarrow$ number of accidents for each worker follows the same Poisson distribution $\mathcal{P}(\lambda)$

ACCIDENTS IN THE CONSTRUCTION SECTOR

- Accidents of different workers are independent
  - Apply Superposition Theorem ($\Rightarrow$ total intensity as sum of individual intensities)
  - $\Rightarrow$ Number of accidents for all workers given by an HPP with rate 364λ
- Gamma prior $\mathcal{G}(1, 1)$ on λ
  - Likelihood $l(\lambda|data) = (364\lambda)^{75}e^{-364\lambda}$
  - Posterior gamma $\mathcal{G}(76, 365)$
  - Posterior mean $76/365 = 0.208$
  - Prior mean 1
  - MLE $75/364 = 0.206$
  - Posterior mean closer to MLE

ACCIDENTS IN THE CONSTRUCTION SECTOR

- Prior $\mathcal{G}(1,1) \Rightarrow$ mean 1 and variance 1
  - *large* variance in this experiment
  - $\Rightarrow$ scarce confidence on the prior assessment of mean equal to 1
- Prior $\mathcal{G}(1000,1000) \Rightarrow$ mean 1 and variance 0.001
  - *Small* variance in this experiment
  - $\Rightarrow$ strong confidence on the prior assessment of mean equal to 1
- $\Rightarrow$ Posterior $\mathcal{G}(1075,1364)$
- Posterior mean $1075/1364 = 0.79$
- Prior mean 1
- MLE $75/364 = 0.21$
- Posterior mean $10075/10364 = 0.97$ for a $\mathcal{G}(10000,10000)$ prior

POISSON PROCESS: INFERENCE

- Computation of quantities of interest
  - analytically (e.g. posterior mean and mode)
  - using basic statistical software (e.g. posterior median and credible intervals)

- Accidents in the construction sector
  - Gamma prior $\mathcal{G}\,(100, 100)$ for the rate λ
  - Posterior mean: $175/464 = 0.377$
  - Posterior mode: $174/464 = 0.375$
  - Posterior median: 0.376
  - $[0.323, 0.435]$: 95% credible interval $\Rightarrow$ quite concentrated distribution
  - Posterior probability of interval $[0.3, 0.4]$: 0.789

IMPROPER PRIORS

- Improper priors
  - Controversial, although rather common, choice, which might reflect lack of knowledge
  - Possible choices
    * $f(\lambda) \propto 1$: Uniform prior
      $\Rightarrow$ posterior $\mathcal{G}\left(n+1, T\right)$
    * $f(\lambda) \propto 1/\lambda$: Jeffreys prior given the experiment of observing times between events
      $\Rightarrow$ posterior $\mathcal{G}\left(n/2, T\right)$
    * $f(\lambda) \propto 1/\sqrt{\lambda}$: Jeffreys prior given the experiment of observing the number of events in a fixed period
      $\Rightarrow$ posterior $\mathcal{G}\left(n+1/2, T\right)$

NON CONJUGATE ANALYSIS

- Given the meaning of λ (expected number of events in unit time interval or inverse of mean interarrival time), it may often be considered that λ is bounded

- $\Rightarrow$ Prior on a bounded set

- Uniform prior on the interval $(0, L]$

- $\Rightarrow$ Posterior $f(\lambda|n, T) \propto \lambda^n e^{-\lambda T} I_{(0,L]}(\lambda)$

- Normalizing constant $\gamma(n+1, LT)/T^{n+1}$, with $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$ lower incomplete gamma function

- Posterior mean $\widehat{\lambda} = \dfrac{1}{T} \dfrac{\gamma(n+2, LT)}{\gamma(n+1, LT)}$

FORECASTING

- n events observed in the interval $(0, T]$
- Interest in forecasting number of events in next intervals: $P(N(T, T+s] = m)$,
- For $s > 0$ and integer m

$$
\begin{aligned}
P(N(T, T+s] = m) &= \int_0^\infty P(N(T, T+s] = m|\lambda)\, f(\lambda|n, T)\, d\lambda \\
&= \int_0^\infty \frac{(\lambda s)^m}{m!} e^{-\lambda s}\, f(\lambda|n, T)\, d\lambda
\end{aligned}
$$

Posterior $\mathcal{G}\left(\alpha + n, \beta + T\right)$

$$\Rightarrow P(N(T, T+s] = m) = \frac{s^m}{m!} \frac{(\beta + T)^{\alpha+n}}{(\beta + T + s)^{\alpha+n+m}} \frac{\Gamma(\alpha + n + m)}{\Gamma(\alpha + n)}$$

FORECASTING

- Expected number of events in the subsequent interval

$$\begin{aligned} E[N(T,T+s]] &= \int_0^\infty E[N(T,T+s]|\lambda] f(\lambda|n,T)d\lambda \\ &= \int_0^\infty \lambda s\, f(\lambda|n,T)d\lambda \end{aligned}$$

Posterior $\mathcal{G}\,(\alpha+n,\beta+T)$

$\Rightarrow E[N(T,T+s]] = s\dfrac{\alpha+n}{\beta+T}$

ACCIDENTS IN THE CONSTRUCTION SECTOR

- Gamma prior $\mathcal{G}\,(100, 100)$ for the rate λ
- Posterior gamma $\mathcal{G}\,(175, 464)$, having observed 75 accidents with 346 workers in 1987
- Interest in number of accidents during the first six months of 1988 (i.e. $s = 0.5$), when the number of workers has increased to 400 (i.e. $m = 400$)
- T_{1987} denotes December, 31st, 1987
- $\Rightarrow N(T_{1987}, T_{1987} + 0.5] \sim \mathcal{P}\,(400\lambda \cdot 0.5)$
- $E[N(T_{1987}, T_{1987} + 0.5]] = 400 \cdot 0.5 \dfrac{175}{464} = 75.431$
- Interested in probability of 100 accidents in the six months:
  $$P(N(T_{1987}, T_{1987} + 0.5] = 100) = \frac{200^{100}}{100!} \frac{464^{175}}{664^{275}} \frac{\Gamma(275)}{\Gamma(175)} = 0.003$$
- Probability of no accidents in the six months: $(464/664)^{175} \approx 0$

NONHOMOGENEOUS POISSON PROCESS

- NHPPs characterized by intensity function $\lambda(t)$ varying over time

- $\Rightarrow$ NHPPs useful to describe (*rare*) events whose rate of occurrence evolves over time (e.g. gas escapes in steel pipelines)

  – Life cycle of a new product

    $*$ initial elevated number of failures (*infant mortality*)

    $*$ almost steady rate of failures (*useful life*)

    $*$ increasing number of failures (*obsolescence*)

    $\Rightarrow$ NHPP with a *bathtub* intensity function

- NHPP has no stationary increments unlike the HPP

- Elicitation of priors raises similar issues as before

INTENSITY FUNCTIONS

Many intensity functions $\lambda(t)$ proposed in literature (see McCollin (ESQR, 2007))

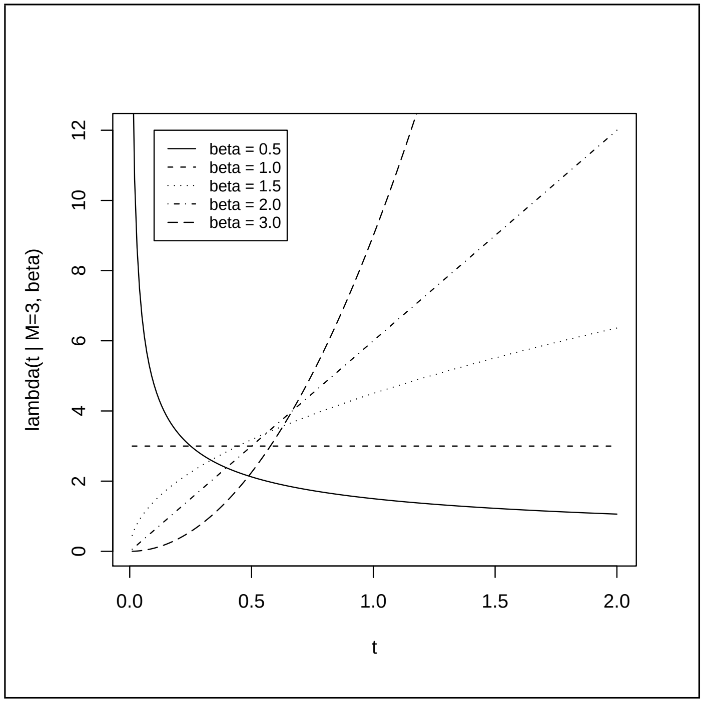
- Different origins
  - Polynomial transformations of HPP constant rate
    - $\lambda(t) = \alpha t + \beta$ (linear ROCOF model)
    - $\lambda(t) = \alpha t^2 + \beta t + \gamma$ (quadratic ROCOF model)
  - Actuarial studies (from hazard rates)
    - $\lambda(t) = \alpha \beta^t$ (Gompertz)
    - $\lambda(t) = \alpha \beta^t + \gamma t + \delta$
    - $\lambda(t) = e^{\alpha + \beta t} + e^{\gamma + \delta t}$
  - Reliability studies
    - $\lambda(t) = \alpha + \beta t + \dfrac{\gamma}{t + \delta}$ (quite close to *bathtub* for adequate values)
    - $\lambda(t) = \alpha \beta (\alpha t)^{\beta - 1} \exp\{\alpha t^\beta\}$ (Weibull software model)

INTENSITY FUNCTIONS

- Different origins
  - Logarithmic transformations
    - $\lambda(t) = \frac{\alpha}{t}$ ($\Rightarrow$ logarithmic $m(t)$)
    - $\lambda(t) = \alpha \log t + \alpha + \beta$
    - $\lambda(t) = \alpha \log (1 + \beta t) + \gamma$
    - $\lambda(t) = \frac{\alpha \log (1 + \beta t)}{1 + \beta t}$ (Pievatolo et al, underground train failures)
  - Associated to distribution functions
    - $\lambda(t) = \alpha f(t; \beta)$, with $f(\cdot)$ density function

NONHOMOGENEOUS POISSON PROCESS

- Different mathematical properties
  - Increasing, decreasing, convex or concave
    - $\lambda(t) = M\beta t^{\beta-1}$, $M, \beta > 0$ (Power Law Process)
    - Different behavior for different βs



NONHOMOGENEOUS POISSON PROCESS

- Different mathematical properties
  - Periodicity (Lewis)
    - $\lambda(t) = \alpha \exp\{\rho \cos(\omega t + \varphi)\}$
    - Earthquake occurrences (Vere-Jones and Ozaki, 1982)
    - Train doors' failures (Pievatolo et al., 2003)
  - Unimodal, starting at 0 and decreasing to 0 when t goes to infinity
    - Ratio-logarithmic intensity
    - $\lambda(t) = \dfrac{\alpha \log(1 + \beta t)}{1 + \beta t}$
    - Train doors' failures (Pievatolo et al., 2003)

NONHOMOGENEOUS POISSON PROCESS

- Properties of the system under consideration
  - Processes subject to faster and faster (slower and slower) occurrence of events
    $\Rightarrow$ increasing (decreasing) $\lambda(t)$
  - Failures of doors in subway trains, with no initial problems, then subject to an increasing sequence of failures, which later became more rare, possibly because of an intervention by the manufacturer
    $\Rightarrow$ ratio-logarithmic $\lambda(t)$ (Pievatolo et al., 2003)
  - New product $\Rightarrow$ life cycle described by *bathtub* intensity
  - Finite number of bugs to be detected during software testing
    $\Rightarrow$ $m(t)$ finite over an infinite horizon
  - Unlimited number of death in a population
    $\Rightarrow$ $m(t)$ infinite over an infinite horizon (as a good approximation)

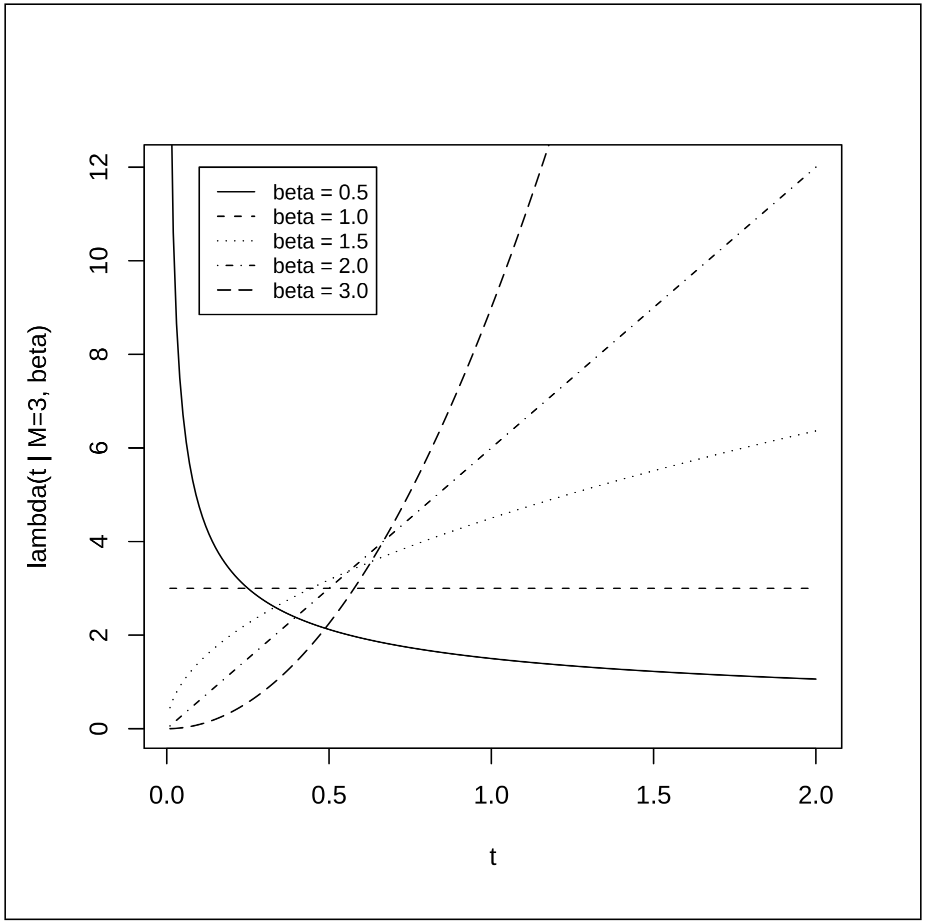
NONHOMOGENEOUS POISSON PROCESS

$N(t)$ Power Law process (PLP) (or Weibull process)

- Two parameterizations:
  - $\lambda(t|\alpha,\beta) = \dfrac{\beta}{\alpha}\left(\dfrac{t}{\alpha}\right)^{\beta-1}$ and $m(t;\alpha,\beta) = \left(\dfrac{t}{\alpha}\right)^{\beta}$, $\alpha,\beta,t > 0$
  - $\lambda(t;M,\beta) = M\beta t^{\beta-1}$ and $m(t;M,\beta) = Mt^{\beta}$, $M,\beta > 0$
  - Link: $\alpha^{-\beta} = M$

- Parameters interpretation
  - $\beta > 1 \Rightarrow$ reliability decay
  - $\beta < 1 \Rightarrow$ reliability growth
  - $\beta = 1 \Rightarrow$ constant reliability
  - $M = m(1)$ expected number of events up to time 1

POWER LAW PROCESS



FREQUENTIST ANALYSIS

Failures $\underline{T} = (T_1, \ldots, T_n) \Rightarrow$ likelihood

$$l(\alpha, \beta \mid \underline{T}) = (\beta/\alpha)^n \prod_{i=1}^{n} (T_i/\alpha)^{\beta-1} \, e^{-(y/\alpha)^\beta}$$

- *Failure truncation* $\Rightarrow y = T_n$

  MLE: $\widehat{\beta} = n / \sum_{i=1}^{n-1} \log(T_n/T_i)$ and $\widehat{\alpha} = T_n / n^{1/\widehat{\beta}}$

  C.I. for β : $\left(\widehat{\beta}\chi^2_{\gamma/2}(2n-2)/(2n), \widehat{\beta}\chi^2_{1-\gamma/2}(2n-2))/(2n)\right)$

- *Time truncation* $\Rightarrow y = T$

  MLE: $\widehat{\beta} = n / \sum_{i=1}^{n} \log(T/T_i)$ and $\widehat{\alpha} = T / n^{1/\widehat{\beta}}$

  C.I. for β : $\left(\widehat{\beta}\chi^2_{\gamma/2}(2n)/(2n), \widehat{\beta}\chi^2_{1-\gamma/2}(2n)/(2n)\right)$

Unbiased estimators, $\widehat{\lambda}(t)$, approx. C.I., hypothesis testing, goodness-of-fit, etc.

BAYESIAN ANALYSIS

Failure truncation $\equiv$ Time truncation

$$l(\alpha, \beta \mid \underline{T}) = (\beta/\alpha)^n \prod_{i=1}^{n} (T_i/\alpha)^{\beta-1} \, e^{-(y/\alpha)^{\beta}}$$

- $\pi(\alpha, \beta) \propto (\alpha\beta^{\gamma})^{-1} \quad \alpha > 0, \beta > 0, \gamma = 0, 1 \Rightarrow \beta|\underline{T} \sim \widehat{\beta}\chi^2_{2(n-\gamma)}/(2n)$
  - Posterior exists, except for $\gamma = 0$ and $n = 1$
  - $\widehat{\beta} = n/\sum_{i=1}^{n} \log(T/T_i)$
  - Posterior mean $\tilde{\beta} = (n-\gamma)/\sum_{i=1}^{n} \log(T/T_i)$
  - Credible intervals easily obtained with standard statistical software
- $\pi(\alpha) \propto \alpha^{-1}$ and $\beta \sim \mathcal{U}(\beta_1, \beta_2) \Rightarrow \pi(\beta|\underline{T}) \propto \beta^{n-1} \prod_{i=1}^{n} T_i^{\beta} I_{[\beta_1,\beta_2]}(\beta)$
- $\pi(\alpha|\beta) \propto \beta s^{a\beta} \alpha^{-a\beta-1} e^{-b(s/\alpha)^{\beta}} \quad a, b, s > 0$ and $\beta \sim \mathcal{U}(\beta_1, \beta_2)$

$$\Rightarrow \pi(\beta|\underline{T}) \propto \beta^n \prod_{i=1}^{n} \left(\frac{T_i}{s}\right)^{\beta} \left[(\frac{T_n}{s})^{\beta} + b\right]^{-n-a} I_{[\beta_1,\beta_2]}(\beta)$$

- In all case $\alpha|\underline{T}$ by simulation (but $\alpha|\beta, \underline{T}$ inverse of a Weibull)

BAYESIAN ANALYSIS

Other parametrization

- $l(M, \beta \mid T_1, \ldots, T_n) = M^n \beta^n \prod_{i=1}^{n} T_i^{\beta-1} e^{-MT^\beta}$
- Independent priors $M \sim \mathcal{G}(\alpha, \delta)$ and $\beta \sim \mathcal{G}(\mu, \nu)$
- Possible dependent prior: $M|\beta \sim \mathcal{G}\left(\alpha, \delta^\beta\right)$
- $\Rightarrow$ posterior conditionals (in red changes for dependent prior)

$$
\begin{aligned}
M|T_1, \ldots, T_n \beta &\sim \mathcal{G}\left(\alpha + n, \delta^\beta + T^\beta\right) \\
\beta|T_1, \ldots, T_n M &\propto \beta^{\mu+n-1} \exp\{\beta(\sum_{i=1}^{n} \log T_i - \nu) - MT^\beta - M\delta^\beta\}
\end{aligned}
$$

- Sample from posterior applying Metropolis step within Gibbs sampler

Interest in posterior $\mathcal{E}\beta, \mathcal{P}\{\beta < 1\}$, *modes, C.I.'s,* $\mathcal{E}M$ *(for* $\lambda(t) = M\beta t^{\beta-1}$*)*