Risk and Environmental Sustainability

Linda Mhalla

1 Introduction slide 2

1.1 Motivation slide 3

Sustainability?

- ☐ How robust are human activities to environmental hazards in a changing world?
  - – Sea level change?
  - – Earthquakes, tsunamis, major windstorms?
  - – Increases in air and water temperatures?
  - – Changes to permafrost?
  - – Changes in rainfall patterns — droughts and floods?
  - – ...
- ☐ Some examples, among many ...

slide 4

Irma, September 2017

slide 5

Bondo, August 2017

slide 6

Fukushima, March 2011

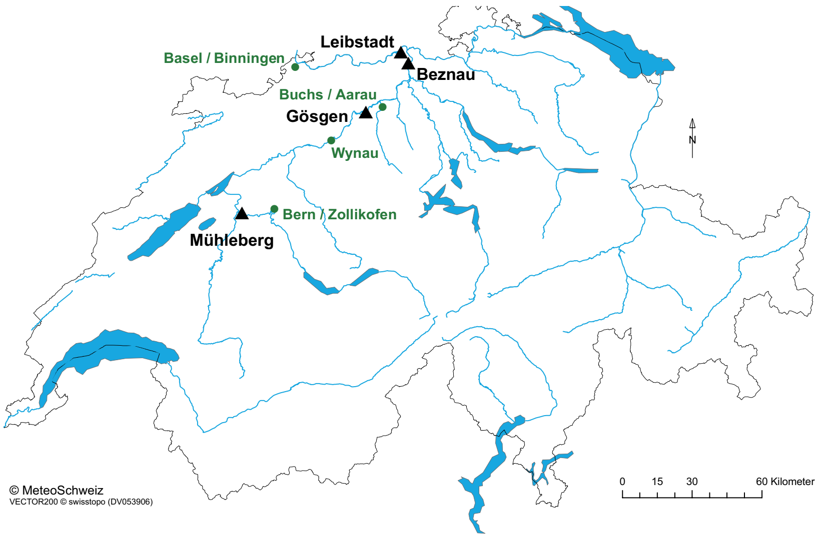
slide 7

Nuclear power safety

- ☐ Fukushima ⇒ nuclear power safety concerns worldwide
- ☐ Swiss nuclear regulator asked for (re-)assessment of vulnerability of the four nuclear plants to
  - high and low air temperatures
  - high and low river water temperatures
  - high winds (and tornados)
  - intense rainfall, snowload, lightning strikes,
  - earthquakes and any tsunamis are dealt with separately!
- ☐ Task: estimate quantiles for probabilities 10^{-4} per year (and 10^{-7} for high winds), and give their uncertainties
  - based on 25 years of data or so at the plants themselves, and (at very most, and only for comparison) 150 years of data nearby

slide 8

Swiss nuclear plants

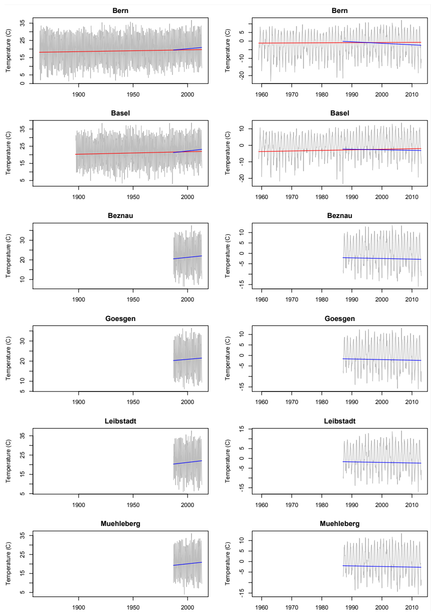


slide 9

Muhleberg

slide 10

Air temperature maxima and minima

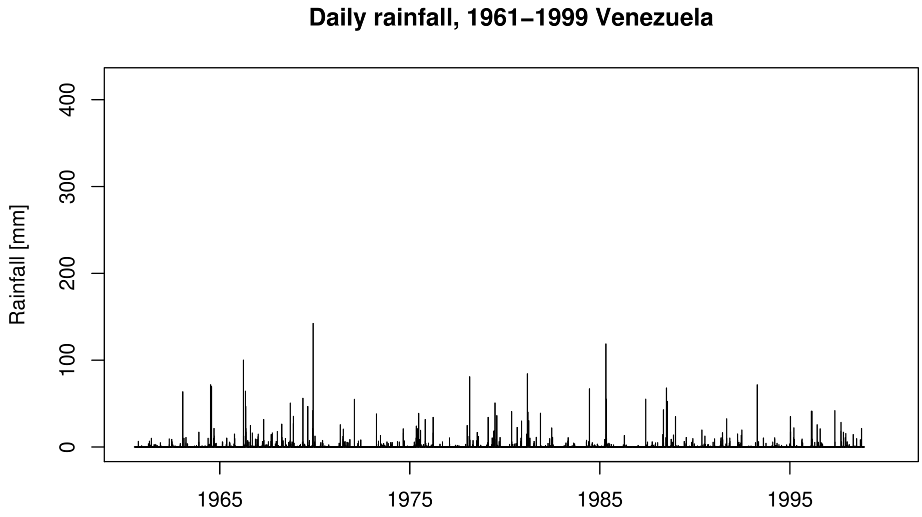


slide 11

Tanaguarena, 1999

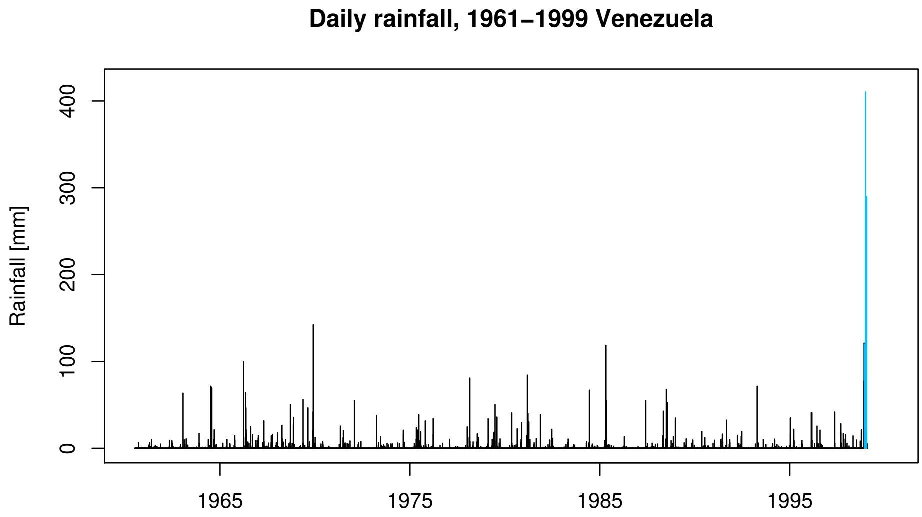
- ☐ Following two weeks of intermittent rainfall, torrential rainfall on 14–16 December 1999 spawned landsides throughout the upper watersheds of the Cerro Grande River near the coast of Venezuela.
- ☐ Mud floods, debris flows and flood surges then destroyed much of Tanaguarena and other coastal tourist towns. Perhaps 30,000 people died.
- ☐ The data are from the airport at Maiquetia: the estimated recurrence time for the three-day rainfall is between 250 years and 6 million years!
- ☐ Similar events, fortunately with less loss of life, have occurred nearby.

Rainfall at Maiquetia



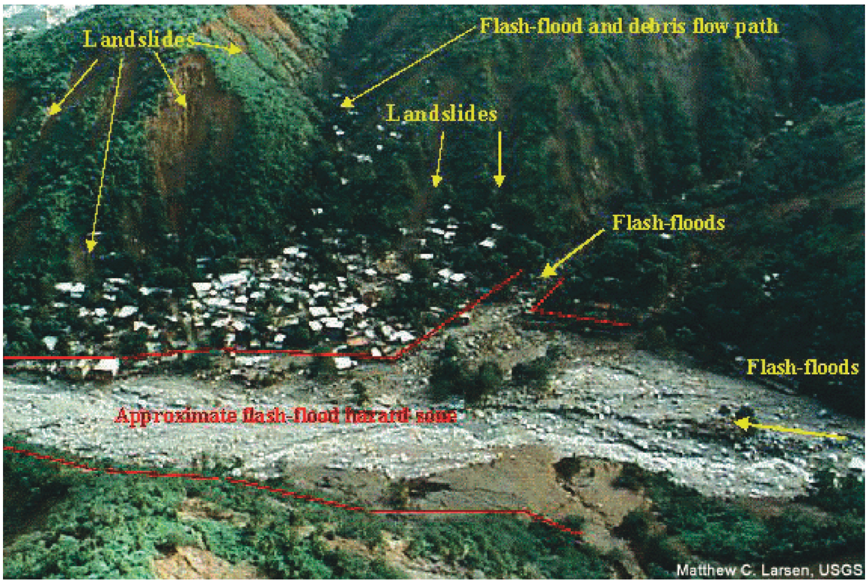
slide 13

Rainfall at Maiquetia



slide 14

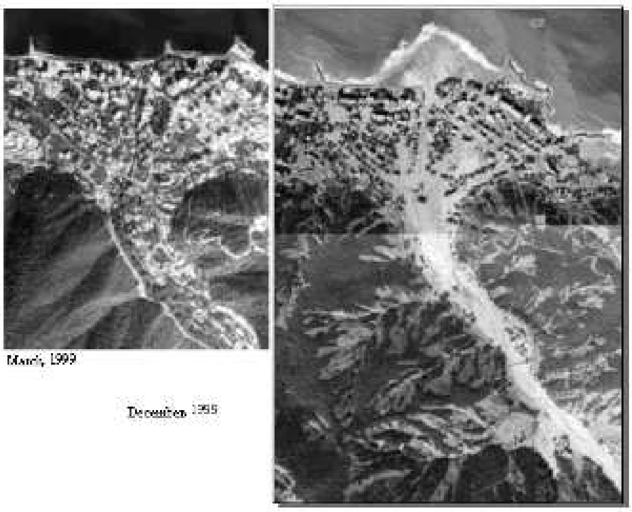
Tanaguarena



slide 15

Cerro Grande rivermouth

Comparison of Cerro Grande fan before and after the Dec. 1999 flood disaster.



slide 16

Risk

- From the Oxford English Dictionary:

  > (Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance; a chance or situation involving such a possibility.

- Risk R can be expressed as

$$R = (A, C, U, P, K),$$

  where

A	is an event that might occur,
C	is the consequences of the event,
U	is an assessment of uncertainties,
P	is a knowledge-based probability of the event
K	is the background knowledge that U and P are based on.

- The consequences C are highly situation-specific, so we focus on methods for estimating the risks based on data.
- This course mostly concerns the estimation of the probabilities P of rare events A based on data K that leads to a robust assessment of their uncertainties U.

slide 17

Environmental sustainability

- Climate change, loss of biodiversity, population growth . . . all threaten our future.
- Change to average conditions are important — world GDP is estimated to drop by 12% for each 1°C of warming (WEF) — but many immediate impacts come from increases in the sizes and occurrence of (previously) rare events:
  - heat waves are dangerous for vulnerable human populations and can impact on food security;
  - hurricanes, typhoons and other major storms can have massive impacts on habitations and consequently on insurance premiums;
  - heavy rainfall leading to widespread flooding can make homes uninhabitable for months and lead to drastic reductions in their value;
  - wildfires can devastate large areas even in first world countries (e.g., Los Angeles earlier this year);
  - et cetera . . .
- Economic sustainability (major financial crashes, food prices, . . .) also involve (formerly) rare events.
- Many such events are **compound**, i.e., depend on a rare combination of several variables.

slide 18

Plan

- ☐ Many risky situations can be formulated in terms of the Poisson process, which is a basic stochastic model for point events — analogous to the Gaussian distribution in modelling continuous random variables.
- ☐ Draft plan . . .
  - Today: motivation, basics of statistical modelling, Poisson process
  - Weeks 2–6: Modelling rare events (extreme-value statistics)
  - Weeks 7–8: Point process and Poisson process
  - Weeks 9–10: Multivariate (compound) rare events
  - Weeks 11–14: Probabilistic forecasting
- ☐ Much of the course will use the contents of Coles (2001) *An Introduction to the Statistical Modeling of Extreme Values*, Springer.

slide 19

1.2 Revision

slide 20

Statistical models

- ☐ A **statistical model** is a set of probability distributions used to
  - describe the variation in (future or existing) data;
  - help understand underlying phenomena;
  - predict future data and answer 'what if' questions;
  - give a realistic assessment of the uncertainty of inferences.
- ☐ We suppose that observed data y are a realisation of a random variable Y from the model, so y might have been different.
- ☐ A model is **parametric** if the distributions can be indexed by a finite parameter vector θ; otherwise it is **nonparametric**.
  - $y_1, \ldots, y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, with $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, is a parametric model;
  - $y_1, \ldots, y_n \overset{\text{iid}}{\sim} F$, with F unknown, is a nonparametric model.
- ☐ In this course almost all the models will be parametric, and key steps are
  - formulation of appropriate models;
  - inference on the parameters, usually by likelihood methods.

slide 21

Likelihood

- Let y be a data set, assumed to be the realisation of a random variable Y from a parametric model $f_Y(y;\theta)$, where the unknown parameter θ lies in a **parameter space** $\Theta \subset \mathbb{R}^p$.
- The **likelihood** (for θ based on y) and the corresponding **log likelihood** are

$$L(\theta) = L(\theta; y) = f_Y(y;\theta), \quad \ell(\theta) = \log L(\theta), \quad \theta \in \Theta.$$

- The **maximum likelihood estimate** (MLE) $\widehat{\theta}$ satisfies $\ell(\widehat{\theta}) \geq \ell(\theta)$, for all $\theta \in \Theta$.
- Often $\widehat{\theta}$ is unique and in many cases it satisfies the **score (or likelihood) equation**

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0,$$

which is interpreted as a vector equation of dimension $p \times 1$ if θ is a $p \times 1$ vector.
- The **observed information** and **expected (Fisher) information** are defined as

$$\jmath(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}}, \quad \imath(\theta) = \mathrm{E}\left\{\jmath(\theta)\right\};$$

these are $p \times p$ matrices if θ has dimension p. The information matrices encode the curvature of the log likelihood and provide information about the variability of $\widehat{\theta}$.

slide 22

Log likelihood

- For both theoretical and numerical reasons we prefer to work with the log likelihood.
- If the data are a random sample, i.e., $y_1, \ldots, y_n \overset{\text{iid}}{\sim} f(y;\theta)$, then

$$L(\theta) = f(y;\theta) = f(y_1, \ldots, y_n; \theta) = \prod_{j=1}^{n} f(y_j;\theta), \quad \theta \in \Theta,$$

so

$$\ell(\theta) = \log L(\theta) = \sum_{j=1}^{n} \log f(y_j;\theta), \quad \theta \in \Theta.$$

- If the data are independent but not identically distributed, with $y_j \sim f_j(y_j;\theta)$, then

$$\ell(\theta) = \sum_{j=1}^{n} \log f_j(y_j;\theta), \quad \theta \in \Theta.$$

- If the data are dependent and ordered in time, then we can write

$$\ell(\theta) = \log f(y_1;\theta) + \sum_{j=2}^{n} \log f(y_j \mid y_1, \ldots, y_{j-1};\theta), \quad \theta \in \Theta.$$

- In each case the information matrices are sums and (under mild conditions) are of order n.

slide 23

Maximum likelihood estimator

- In large samples from a **regular model** in which the true parameter is $\theta^0_{p\times 1}$, the maximum likelihood estimator $\widehat{\theta}$ has an approximate normal distribution,

$$\widehat{\theta} \dot{\sim} \mathcal{N}_p\left\{\theta^0, \jmath(\widehat{\theta})^{-1}\right\},$$

  so we can compute an approximate $(1-2\alpha)$ confidence interval for the rth parameter θ^0_r as

$$\widehat{\theta}_r \pm z_\alpha v_{rr}^{1/2},$$

  where v_{rr} is the rth diagonal element of the matrix $\jmath(\widehat{\theta})^{-1}$.
- This approximation also holds under weaker conditions, for non-identically distributed and dependent data.
- This is easily implemented:
  - we (carefully!) code the negative log likelihood $-\ell(\theta)$;
  - we minimise $-\ell(\theta)$ numerically, ensuring that the routine returns $\widehat{\theta}$ and the Hessian matrix $\jmath(\widehat{\theta}) = -\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}|_{\theta=\widehat{\theta}}$
  - we compute $\jmath(\widehat{\theta})^{-1}$, and use the square roots of its diagonal elements, $v_{11}^{1/2}, \ldots, v_{pp}^{1/2}$, as standard errors for the corresponding elements of $\widehat{\theta}$.

slide 24

Delta method and transformations

The asymptotic normality result can be used to derive standard errors for other quantities of interest. If $\widehat{\phi} = g(\widehat{\theta})$, where $g : \mathbb{R}^p \to \mathbb{R}^k$ for $k \le p$ is a differentiable function of θ non-vanishing at θ^0 then

$$\widehat{\phi} \dot{\sim} \mathcal{N}_k(\phi^0, \nabla\phi^{\mathrm{T}}\jmath(\widehat{\theta})^{-1}\nabla\phi),$$

where

$$\nabla\phi = [\partial\phi/\partial\theta_1, \ldots, \partial\phi/\partial\theta_p]^{\mathrm{T}}.$$

The variance matrix and the jacobian are evaluated at the maximum likelihood estimate $\widehat{\theta}$.

slide 25

Maximum likelihood estimator

Example 1 (Weibull model)
Let $(y_1, \ldots, y_n)$ *be a data set, assumed to be a realisation of a Weibull random variable with scale* $\lambda > 0$ *and shape* $\alpha > 0$*. The corresponding density is*

$$f_Y(y; \lambda, \alpha) = \tfrac{\alpha}{\lambda^\alpha} y^{\alpha-1} \exp\left\{-(y/\lambda)^\alpha\right\}, \quad x \geq 0, \lambda > 0, \alpha > 0.$$

The Weibull distribution includes the exponential as special case when $\alpha = 1$*. The log likelihood for the Weibull(*λ, α*) model is*

$$\ell(\lambda, \alpha) = n \ln(\alpha) - n\alpha \ln(\lambda) + (\alpha - 1) \sum_{i=1}^{n} \ln y_i - \lambda^{-\alpha} \sum_{i=1}^{n} y_i^\alpha.$$

The gradient of this function is easily obtained by differentiation

$$\tfrac{\partial \ell(\lambda,\alpha)}{\partial \lambda} = -\tfrac{n\alpha}{\lambda} + \alpha \lambda^{-\alpha-1} \sum_{i=1}^{n} y_i^\alpha,$$

$$\tfrac{\partial \ell(\lambda,\alpha)}{\partial \alpha} = \tfrac{n}{\alpha} - n \ln(\lambda) + \sum_{i=1}^{n} \ln y_i - \sum_{i=1}^{n} \left(\tfrac{y_i}{\lambda}\right)^\alpha \times \ln\left(\tfrac{y_i}{\lambda}\right).$$

R demo to follow

slide 26

Likelihood ratio statistic

- ☐ Suppose that likelihood inference for model A is OK, so $\widehat{\theta}_A \dot\sim \mathcal{N}\{\theta_A, \jmath_A(\widehat{\theta}_A)^{-1}\}$.
- ☐ Model $f_B(y)$ is **nested** within model $f_A(y)$ if A reduces to B on restricting some parameters:
  - for example, $f_B \equiv \mathcal{N}(0, \sigma^2)$ is nested within $f_A \equiv \mathcal{N}(\mu, \sigma^2)$, because B is obtained by setting $\mu = 0$ in A;
  - the maximised log likelihoods satisfy $\widehat{\ell}_A \geq \widehat{\ell}_B$, because the maximisation for A is over a larger set than for B.
- ☐ The **deviance** for model A is defined to be $D_A = \text{const} - 2\widehat{\ell}_\text{A}$, and then $D_B > D_A$.
- ☐ The **likelihood ratio statistic** for comparing A and B is

$$W = 2(\widehat{\ell}_A - \widehat{\ell}_B) = D_B - D_A.$$

- ☐ If model B is true and the models have p_A and p_B parameters, then

$$W \dot\sim \chi^2_{p_A - p_B}.$$

- ☐ The deviance is often used to compare models, and so is the **Akaike information criterion**

$$\text{AIC} = 2p_A - 2\widehat{\ell}_A,$$

with smaller values of both D_A and AIC being preferred.

slide 27

Profile log likelihood

- Split $\theta = (\psi, \lambda)$ into a **parameter of interest** ψ and a **nuisance parameter** λ that are variation independent, i.e., $(\psi, \lambda) \in \Theta_\psi \times \Theta_\lambda$, and write the overall MLE as $\widehat{\theta} = (\widehat{\psi}, \widehat{\lambda})$.
- A $(1 - 2\alpha)$ confidence region for ψ can be based on the **profile log likelihood**

$$\ell_{\rm p}(\psi) = \max_{\lambda \in \Theta_\lambda} \ell(\psi, \lambda) = \ell(\psi, \widehat{\lambda}_\psi),$$

and is

$$\left\{\psi \in \Theta_\psi : 2\{\ell(\widehat{\psi}, \widehat{\lambda}) - \ell(\psi, \widehat{\lambda}_\psi)\} \le \chi^2_{\dim \psi}(1 - 2\alpha)\right\}.$$

- When ψ is scalar, this yields

$$\left\{\psi \in \Theta_\psi : \ell(\psi, \widehat{\lambda}_\psi) \ge \ell(\widehat{\psi}, \widehat{\lambda}) - \tfrac{1}{2}\chi^2_1(1 - 2\alpha)\right\},$$

and $\chi^2_1(0.95) = 3.84$, $\chi^2_1(0.99) = 6.63$ and $\chi^2_1(0.999) = 10.83$.
- Such intervals are preferable to the standard interval $\widehat{\psi} \pm z_\alpha v_{\psi\psi}^{1/2}$ when the distribution of $\widehat{\psi}$ is asymmetric, but require more computation, since they involve many maximisations of ℓ.

slide 28

Profile log likelihood: back to Example 1

Consider the shape parameter $\psi \equiv \alpha$ as parameter of interest, and the scale λ as nuisance parameter. Using the gradient,

$$\frac{\partial \ell(\lambda, \alpha)}{\partial \lambda} = -\frac{n\alpha}{\lambda} + \alpha\lambda^{-\alpha-1}\sum_{i=1}^n y_i^\alpha$$

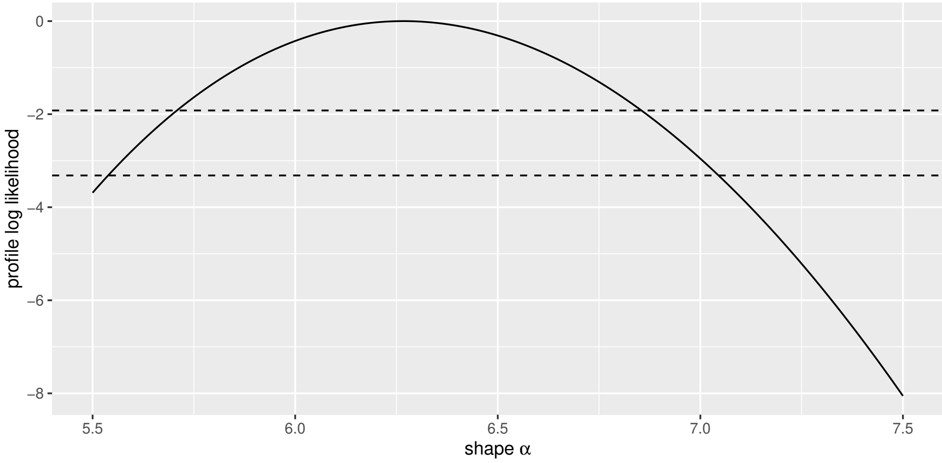
we find that the value of the scale that maximises the log likelihood for given α is

$$\widehat{\lambda}_\alpha = \left(\frac{1}{n}\sum_{i=1}^n y_i^\alpha\right)^{1/\alpha}.$$

and plugging in this value gives a function of α alone, thereby also reducing the optimisation problem for the Weibull to a line search along $\ell_{\rm p}(\alpha)$.

slide 29

Profile log likelihood: back to Example 1



slide 30

Regular models

The above approximate distributions hold under **regularity conditions**:

(C1) the true value θ^0 of θ is interior to the parameter space $\Theta \subset \mathbb{R}^p$ for some fixed p;

(C2) the densities defined by any two distinct values of θ are different;

(C3) there is a neighbourhood $\mathcal{N}$ of θ^0 within which the first three derivatives of ℓ with respect to θ exist almost surely, and for $r, s, t = 1, \dots, d$ satisfy

$$|\partial^3 \log f(Y;\theta)/\partial\theta_r\partial\theta_s\partial\theta_t| < m(Y),$$

with $\mathrm{E}_g\{m(Y)\} < \infty$; and

(C4) the first two **Bartlett identities** hold within $\mathcal{N}$, i.e., for $\theta \in \mathcal{N}$,

$$\begin{aligned} 0 &= \nabla_\theta \int f(y;\theta)\,\mathrm{d}y = \int \nabla_\theta \log f(y;\theta) \times f(y;\theta)\,\mathrm{d}y, \\ 0 &= \nabla_\theta^2 \int f(y;\theta)\,\mathrm{d}y \\ &= \int \nabla_\theta^2 \log f(y;\theta) \times f(y;\theta)\,\mathrm{d}y + \int \nabla_\theta \log f(y;\theta)\, \nabla_\theta^{\mathrm{T}} \log f(y;\theta) \times f(y;\theta)\,\mathrm{d}y, \end{aligned}$$

where $\nabla_\theta \cdot = \partial \cdot /\partial\theta$ and $\nabla_\theta^2 \cdot = \partial^2 \cdot /\partial\theta\partial\theta^{\mathrm{T}}$.

slide 31

Regularity conditions

- ☐ These conditions are sufficient (but not necessary) to prove theorems giving the limiting distributions for $\widehat{\theta}$ and W as the sample size (or more generally some measure of the information in the data) goes to infinity.
- ☐ Why they are needed:
  - (C1) ensures that $\widehat{\theta}$ can be 'on all sides' of θ^0 in the limit — if it fails, then any limiting distribution cannot be normal;
  - (C2) is essential for consistency, otherwise $\widehat{\theta}$ might not converge to a unique limit;
  - (C3) is needed to bound terms of a Taylor series — can be replaced by other conditions; and
  - (C4) ensures that $\widehat{\theta}$ is consistent for θ^0 and that the asymptotic variance of $\widehat{\theta}$ is the inverse Fisher information $\imath(\theta^0)^{-1}$.
- ☐ In some of the models arising later, (C4) may fail (or be close to failing), because the support of the data depends on a parameter.

slide 32

Model checking

- ☐ To check whether an assumed model for data is suitable we often use graphs, because
  - they show the data directly;
  - unexpected features may be visible.
- ☐ If the data are assumed to be a random sample $y_1, \ldots, y_n \stackrel{\text{iid}}{\sim} F$, with order statistics

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)},$$

  then a **quantile-quantile plot (Q-Q plot)** shows

$$(F^{-1}\{1/(n+1)\}, y_{(1)}), \ldots, (F^{-1}\{n/(n+1)\}, y_{(n)})$$

  where $F^{-1}\{1/(n+1)\}, \ldots, F^{-1}\{n/(n+1)\}$ are called the **plotting positions** for F.
- ☐ Ideally this plot
  - should be a straight line if the assumption is correct;
  - shows model failure as systematic curvature;
  - shows outliers as isolated points,

  but variation can be expected even if the assumption is correct!
- ☐ In practice F is often unknown and must be replaced by an estimate $\widehat{F}$.

slide 33

1.3 Example

Poisson process in the line

- A simple model for times of events (earthquakes, typhoons, heatwaves, . . .).
- Write $N(\mathcal{A})$ for the number of events in a set $\mathcal{A} \subset [0, t_0]$, where t_0 is fixed and known.
  - let $N(w, w+t)$ denote the number of events in $(w, w+t]$, and set
  - $N(t) = N(0, t)$, $t > 0$.
- Let $\dot{\mu}(t)$ be a non-negative **intensity function** giving the rate of events around t (picture!), and whose integral $\mu(0, t_0) = \int_0^{t_0} \dot{\mu}(t)\,\mathrm{d}t < \infty$, and suppose that
  - events in disjoint subsets of $[0, t_0]$ are independent, i.e., $N(\mathcal{A}_1)$ is independent of $N(\mathcal{A}_2)$ whenever $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$;
  - $\mathrm{P}\{N(t, t+\delta t) = 0\} = 1 - \dot{\mu}(t)\delta t + o(\delta t)$ for small δt; and
  - $\mathrm{P}\{N(t, t+\delta t) = 1\} = \dot{\mu}(t)\delta t + o(\delta t)$ for small δt.
- The last two properties imply that

$$\mathrm{P}\{N(t, t+\delta t) > 1\} = o(\delta t) \to 0, \quad \delta t \to 0,$$

so the process is **orderly**: multiple occurrences at the same time cannot occur.

Poisson process in the line, II

- Under these assumptions,
  - the **void probability** of the set $(w, w+t]$ is

$$\mathrm{P}\left\{N(w, w+t) = 0\right\} = \exp\left\{-\mu(w, w+t)\right\},$$

  - the random **waiting time** T from w to the next event has PDF

$$f_T(t) = \dot{\mu}(w+t)\exp\left\{-\mu(w, w+t)\right\}, \quad t > 0,$$

    i.e., $\mu(w, w+T) \sim \exp(1)$, independent of waiting times on other intervals;
  - the joint density of events at $0 < t_1 < \cdots < t_n < t_0$ is

$$\exp\left\{-\mu(0, t_0)\right\} \prod_{j=1}^{n} \dot{\mu}(t_j), \quad 0 < t_1 < \cdots < t_n < t_0,$$

  - and $N(0, t_0) \sim \mathrm{Poiss}\{\mu(0, t_0)\}$.
- Hence if the sets $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are disjoint, the corresponding numbers of events satisfy

$$N(\mathcal{A}_j) \stackrel{\text{ind}}{\sim} \mathrm{Poiss}\{\mu(\mathcal{A}_j)\}.$$

Note: Poisson process in the line

- ☐ To find the probability of no events in $(w, w+t]$ we divide it into k subintervals of length $\delta t = t/k$, and then let $\delta t \to 0$. Then as events in disjoint intervals are independent,

$$\begin{aligned} \mathrm{P}\{N(w, w+t)=0\} &= \prod_{i=0}^{k-1} \mathrm{P}\left[N\{w+i\delta t, w+(i+1)\delta t\}=0\right] \\ &= \prod_{i=0}^{k-1}\{1-\dot\mu(w+i\delta t)\delta t + o(\delta t)\} \end{aligned}$$

has negative logarithm

$$-\sum_{i=0}^{k-1}\log\{1-\dot\mu(w+i\delta t)\delta t+o(\delta t)\} = \sum_{i=0}^{k-1}\dot\mu(w+i\delta t)\delta t + o(k\delta t) \to \int_w^{w+t}\dot\mu(u)\,\mathrm{d}u = \mu(w, w+t),$$

where the limit follows because as $\delta t \to 0$ with t fixed, $o(k\delta t) = k\,o(\delta t)/\delta t \to 0$. Hence

$$\mathrm{P}\{N(w, w+t)=0\} = \exp\{-\mu(w, w+t)\}, \quad t>0.$$

- ☐ The time T after w to the next event exceeds t if and only if $N(w, w+t)=0$, so

$$\mathrm{P}(T>t) = \mathrm{P}\{N(w, w+t)=0\} = \exp\{-\mu(w, w+t)\},$$

and thus T has PDF

$$f_T(t) = -\frac{\mathrm{dP}\{N(w, w+t)=0\}}{\mathrm{d}t} = \dot\mu(w+t)\exp\{-\mu(w, w+t)\}. \quad t>0.$$

Put another way, $\mu(w, w+T) \sim \exp(1)$.

- ☐ If events in $(0, t_0]$ have been observed at times $t_1, \ldots, t_n$, where $0<t_1<\cdots<t_n<t_0$, then, as events in disjoint sets are independent, the joint probability density of the data is

$$\dot\mu(t_1)e^{-\mu(0,t_1)} \times \dot\mu(t_2)e^{-\mu(t_1,t_2)} \times \cdots \times \dot\mu(t_n)e^{-\mu(t_{n-1},t_n)} \times e^{-\mu(t_n,t_0)},$$

where the final term is the probability of no events in $(t_n, t_0]$. This joint density reduces to

$$\exp\{-\mu(0,t_0)\}\prod_{j=1}^{n}\dot\mu(t_j), \quad 0<t_1<\cdots<t_n<t_0. \tag{1}$$

note 1 of slide 36

Poisson process in the line, III

- Without further assumptions on μ, the Poisson process is a nonparametric model.
- The simplest parametric version is the **homogeneous Poisson process**, with $\dot{\mu}(t) \equiv \dot{\mu}$ a positive constant, under which the times between events are independent with PDF

$$f_T(t) = \dot{\mu}(t) \exp\{-\mu(w, t+w)\} = \dot{\mu} \exp(-\dot{\mu}t), \quad t > 0,$$

  i.e., the intervals $T_1, \ldots, T_n \overset{\text{iid}}{\sim} \exp(\dot{\mu})$.
- A simple parametric model for trend might set

$$\dot{\mu}(t) = \exp(\beta_0 + \beta_1 t), \quad \beta_0, \beta_1 \in \mathbb{R},$$

  which reduces to the homogeneous model when $\beta_1 = 0$.
- In principle we could model more complex trends by replacing $\beta_0 + \beta_1 t$ by linear combinations of basis functions,

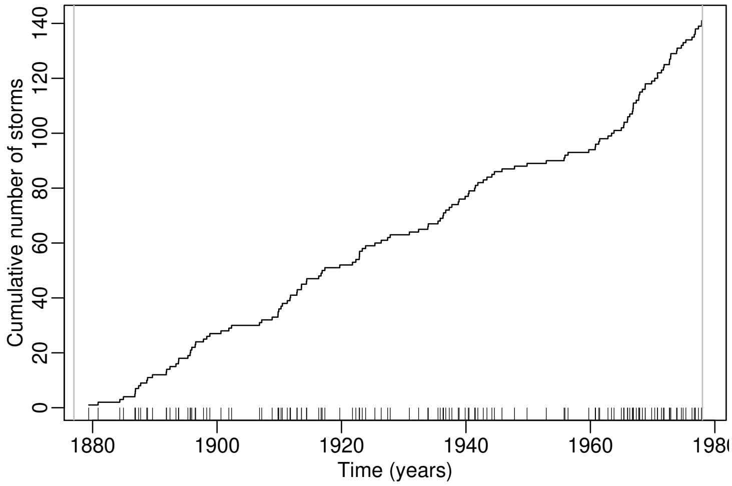
$$\beta_0 + \beta_1 b_1(t) + \cdots + \beta_p b_p(t),$$

  e.g., using trigonometric functions for seasonality, but we must compute $\int_0^{t_0} \dot{\mu}(t)\,\mathrm{d}t$.
- Such models are linear exponential families, so theory from the second year applies ...

slide 37

Cyclones

Times of major cyclonic storms striking the Bay of Bengal from 1877–1977; jittered vertically for visualisation (Mooley, 1981, *Sankhyā*). In November 1970, Cyclone Bhola, the deadliest storm in world history, occurred in the Bay of Bengal and killed around half a million people. It brought a storm surge estimated at 10.4m to the coast.



slide 38

Cyclones II

- The storm times don't look very even, but perhaps that's just randomness . . .
- Take $[0, t_0] \equiv$ [1 January 1877, 31 December 1977], so the t_j are measured in years after the start of 1877 and run up to $t_0 = 101$.
- Under the simplest possible model, the data are a homogeneous Poisson process with $n = 141$ events in $[0, 101]$. Then $\mu(t) = \dot\mu t$, so (writing $\lambda = \dot\mu$ for simpler notation) the likelihood is

$$L(\lambda) = f(t_1, \ldots, t_n; \lambda) = \exp\{-\mu(0, t_0)\} \prod_{j=1}^{n} \dot\mu(t_j) = \exp(-t_0\lambda)\lambda^n,$$

giving maximised log likelihood, MLE and corresponding observed information

$$\ell(\widehat\lambda) = -93.96, , \quad \widehat\lambda = n/t_0 = 141/101 \doteq 1.4 \text{ events/year}, \quad \jmath(\widehat\lambda) = n/\widehat\lambda^2 = t_0^2/n \doteq 72.3,$$

and the approximate 95% confidence interval based on $\widehat\lambda$ has limits

$$\widehat\lambda \pm 1.96\jmath(\widehat\lambda)^{-1/2} \doteq (1.17, 1.63) \text{ events/year}.$$

slide 39

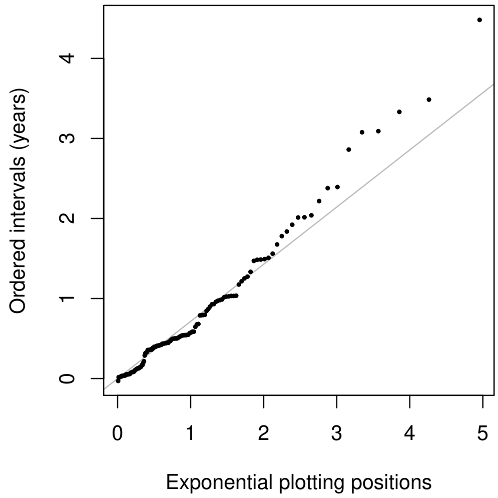
Cyclones V

```
# numerical optimisation of the likelihood of the simplest model
# bengal has data in units of years
load("bengal.dat")  # data available on Moodle page
bengal # look at event times

nlogL <- function(lambda, t, t0=101)
{ # negative log likelihood
  t0*lambda - log(lambda)*length(t)
}
(fit <- optim(par=c(1.2), fn=nlogL, hessian=T, t=bengal-1877, t0=101))
$par
[1] 1.395937
$value
[1] 93.95685
...
$hessian
      [,1]
[1,] 72.35818
```

slide 40

Cyclones III

- ☐ Under this model, and setting $t_0 = 101$, the intervals $t_1 - t_0, t_2 - t_1, \ldots t_n - t_{n-1} \stackrel{\text{iid}}{\sim} \exp(\lambda)$, so a QQ-plot of these intervals against exponential plotting positions should be a straight line.



The grey line corresponds to $y = x/\widehat{\lambda}$.

slide 41

Cyclones IV

- ☐ The QQplot shows departures from the exponential distribution (the larger values are systematically too big), so the basic model may be too simple.
- ☐ Let $\dot{\mu}(t) = \lambda \exp(t\beta)$, so $\mu(0, t_0) = \lambda(e^{t_0\beta} - 1)/\beta$, where $\beta > 0$ would lead to increases in the rate, and conversely.
- ☐ The code on the next slide fits this model and computes the standard errors, giving

$$\ell(\widehat{\lambda}, \widehat{\beta}) = -89.65, \quad \widehat{\lambda} = 0.88_{0.17}, \quad \widehat{\beta} = 0.0086_{0.0030}.$$

- ☐ The likelihood ratio statistic for comparing the models is

$$2\{-89.65 - (-93.96)\} = 8.62 \dot{\sim} \chi^2_{2-1},$$

which gives (approximate) significance level 0.0034, fairly strong evidence of an increase in numbers of cyclones.

- ☐ Looking at the original data, we might query this model of smooth increase. As $\mu(w, w+T) \sim \exp(1)$, we could try a QQplot of

$$\widehat{\mu}(t_{j-1}, t_j) = \widehat{\lambda}\left(e^{\widehat{\beta}t_j} - e^{\widehat{\beta}t_{j-1}}\right)/\widehat{\beta}, \quad j = 1, \ldots, n.$$

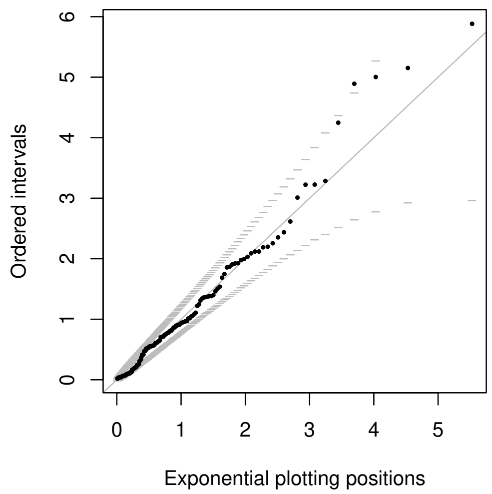
slide 42

Cyclones V

```
# comparison of homogeneous model with log-linear trend
# bengal has data in units of years

nlogL <- function(th, t, t0=101)
{ # negative log likelihood
  int <- th[1]*(exp(t0*th[2])-1)/th[2]
  int - sum( log(th[1]) + t*th[2] )
}

(fit <- optim(par=c(1.4,0.1), fn=nlogL, hessian=T, t=bengal-1877, t0=101 ))
$par
[1] 0.881341438 0.008567385

$value
[1] 89.64509

...

$hessian
          [,1]       [,2]
[1,]  181.5231   9273.082
[2,] 9273.0820 588285.232

(se <- sqrt(diag(solve(fit$hessian))))
[1] 0.168186263 0.002954354
```

slide 43

Cyclones VI



The grey line corresponds to $y = x$, and the grey minus signs show the 95% ranges for individual order statistics.

slide 44

Cyclones VII

- ☐ Visual guidance about 'acceptable' variation is useful . . .
- ☐ The new QQplot is better but even if they are individually (mostly) inside the limits,
  - – the largest intervals still seem too long, and
  - – the smallest intervals seem too short?
- ☐ The original data variation looks more like a change in slope around 1960 than a smooth increase in rate . . . .
- ☐ Maybe we could explain this variation by allowing
  - – seasonality?
  - – (random?) changes in the rate?
  - – external climatic factors such as the El Niño-Southern Oscillation (ENSO)?
- ☐ The latter would be preferable — if we could predict how climate change would influence the ENSO, we could then make an educated guess about the likely future frequencies of cyclones . . . If ENSO does influence the occurrence of cyclones, then it would be a so-called **causal variable**, unlike time, or pure randomness.

slide 45

Summary

- ☐ The course will mostly concern statistical modelling for rare events that could have big impacts.
- ☐ We've now:
  - – seen some basic modelling ideas that will be used repeatedly;
  - – met the simplest Poisson process for the occurrence of random point events;
  - – applied that model to a small dataset.
- ☐ The Poisson process is a key ingredient in rare event modelling, so next week we shall look at it in more generality.

slide 46