Today

- HW 3 due April 1
- Project due April 15
- model choice — Cox & Donnelly, Ch. 7.3
- nonparametric and semi-parametric regression
- In the News:

In the News

- "Benefits gained, benefits lost: comparing baby boomers to other generations in a longitudinal cohort study of self-rated health". highlights; full article

  Badley et al., The Millbank Quarterly, Volume 93, Issue 1, 2015.
- "Cognitive control in media multitaskers". (HW 3)

  Ophir et al., PNAS, Volume 106, Sept 15 2009
- "A learning secret: don't take notes with a laptop".

  May, Scientific American, June 3, 2014
- "What is the question?" Leek & Peng, Science, Feb 26, 2015
- "A compendium of clean graphs in R" A Shiny App

  very helpful for classic R plots
- "Data science done well looks easy" Leek at simply statistics
- "Vitamin D supplements aren't all sunshine and lollipops".

  Picard, Globe & Mail, March 17
- " How the anti-vaccine movement lie with statistics"

  Johnson, Significance Magazine, March 17

"Vitamin D supplements aren't all sunshine and lollipops". Picard, Globe & Mail, March 17

- "In recent weeks, an advertisement has been running prominently in The Globe and Mail that makes some eye-popping claims, among them that vitamin D deficiency is causing widespread illness and premature deaths, costing the health system $20-billion a year"
- "The solution, according to the Pure North S'Energy Foundation, is to dramatically increase Canadians' intake of vitamin D from the current recommendation of 600 to 800 international units daily to 6,000 to 9,000 IUs a day"
- "the claims that it's a miracle drug that can prevent a wide range of illnesses ... have to be kept in context."
- "this observational research shows is that people with adequate vitamin D levels have lower rates of a wide range of chronic illnesses. Stated simply, healthy people tend to be healthy."

... vitamin D

- “True North’s campaign takes issue with the recommended daily allowance, claiming that it was established based on a mathematical error made by the IOM. That debate can be left to statisticians.”
- “A statistical error in the estimation of the recommended dietary allowance for Vitamin D

Veugelers & Ekwaru, Nutrients, Oct 20 2014

3/17/2015 PubMed Central, Figure 1: Nutrients. 2014 Oct; 6(10): 4472–4475. Published online 2014 Oct 20. doi: 10.3390/nu6104472

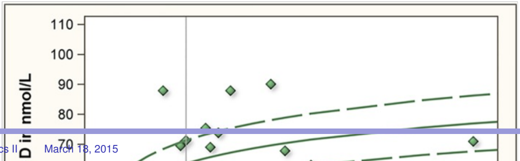
MDPI OPEN ACCESS nutrients

home | this article | this journal | instructions for authors | subscribe

PMC full text: Nutrients. 2014 Oct; 6(10): 4472–4475.
Published online 2014 Oct 20. doi: 10.3390/nu6104472
Copyright/License ► Request permission to reuse

<< Prev Figure 1 Next >>

Figure 1

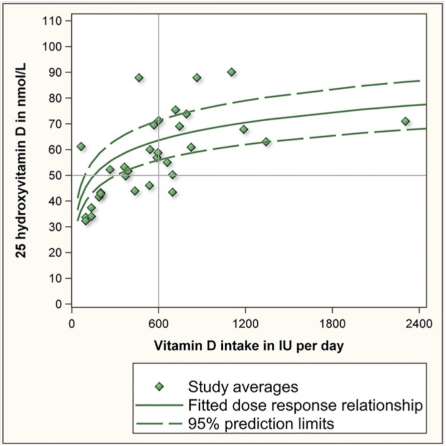


MDPI OPEN ACCESS nutrients
home | this article | this journal | instructions for authors | subscribe

PMC full text: Nutrients. 2014 Oct; 6(10): 4472–4475.
Published online 2014 Oct 20. doi: 10.3390/nu6104472

<< Prev Figure 1 Next >>

Figure 1



Dose response relationship of vitamin D intake and serum 25 hydroxyvitamin D.

Images in this article

... vitamin D

– The IOM regressed the 32 study averages against vitamin D intake to yield the dose response relationship of vitamin D intake and serum 25(OH)D (green solid line in Figure 1).

– On the basis of this, the IOM estimated that 600 IU of vitamin D would achieve an average 25(OH)D level of 63 nmol/L and a lower 95% confidence prediction limit (2.5 percentile) of 56 nmol/L.

– this data point (600 IU vitamin D, 50 nmol/L) is the basis for the current RDA and for the IOM's conclusion that an intake of 600 IU of vitamin D per day will achieve serum 25(OH)D levels of 50 nmol/L or more in 97.5% of individuals. This conclusion, however, is incorrect.

– the correct interpretation of the lower prediction limit is that 97.5% of study averages are predicted to have values exceeding this limit. This is essentially different from the IOM's conclusion that 97.5% of individuals will have values exceeding the lower prediction limit

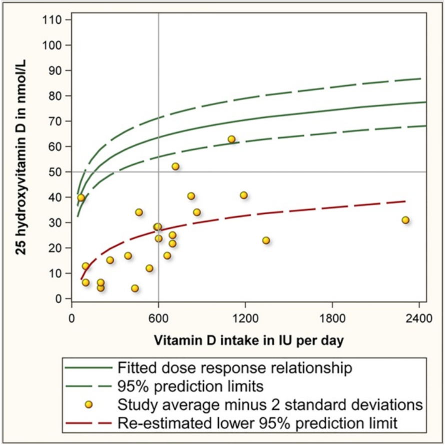
– we estimated how much vitamin D is needed to achieve that 97.5% of individuals achieve serum 25(OH)D values of 50 nmol/L or more.

MDPI OPEN ACCESS nutrients
home | this article | this journal | instructions for authors | subscribe

PMC full text: Nutrients. 2014 Oct; 6(10): 4472–4475.
Published online 2014 Oct 20. doi: 10.3390/nu6104472
Copyright/License ► Request permission to reuse

<< Prev Figure 2 Next >>

Figure 2



Dose response relationship of vitamin D intake and serum 25 hydroxyvitamin D.

Images in this article

For each of these 23 study averages we calculated the 2.5th percentile by subtracting 2 standard deviations from the average (depicted by yellow dots in Figure 2). Next, we regressed these 23 values against vitamin D intake to yield the lower prediction limit (red line in Figure 2). This regression line revealed that 600 IU of vitamin D per day achieves that 97.5% of individuals will have serum 25(OH)D values above 26.8 nmol/L rather than above 50 nmol/L which is currently assumed. It also estimated that 8895 IU of vitamin D per day may be needed to accomplish that 97.5% of individuals achieve serum 25(OH)D values of 50 nmol/L or more.

... vitamin D

NIH book, published by IOM
Dietary Reference Intakes for Adequacy:
Calcium and Vitamin D

– need to check this source, but it's very detailed!

– note also, from the Nutrients paper: "As this dose is far beyond the range of studied doses, caution is warranted when interpreting this estimate."

– and note, from a later Nutrients paper by the same authors

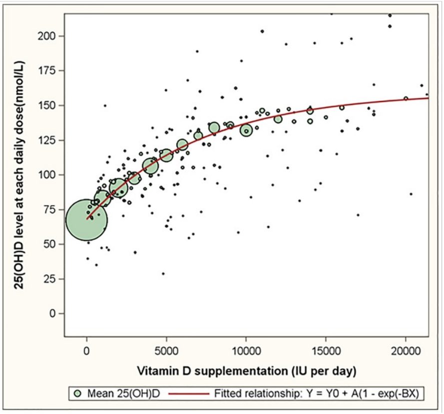
– This study is based on information from healthy volunteers participating in a preventive health program provided by the Pure North S'Energy Foundation (PN), a not-for-profit charitable organization providing free services since October 2007.

– which is the Foundation publishing the advertisements in the Globe & Mail

PMC full text: PLoS One. 2014; 9(11): e111265.
Published online 2014 Nov 5. doi: 10.1371/journal.pone.0111265

<< Prev Figure 1 Next >>

Figure 1

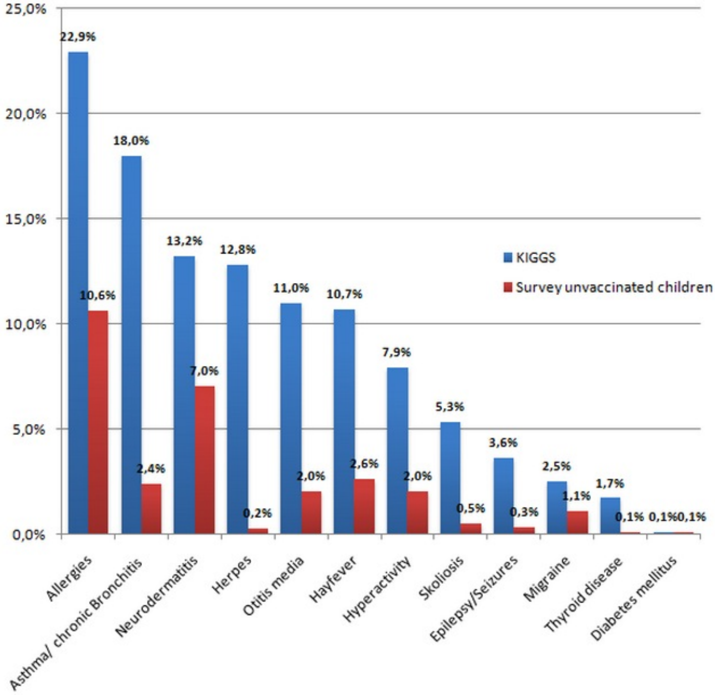


The dose response relationship between oral vitamin D supplementation and serum 25(OH)D levels based on 22,214 observations of healthy volunteers.

Footnote: Bubbles represent the mean plasma 25(OH)D level for all reported daily doses. The size of the bubbles is proportional to the number of assessments for each of the reported daily doses. The red line represents the fitted dose response curve.

... in the News

- How the anti-vaccine movement lie with statistics
- “Facebook recommended I read an article about health outcomes in unvaccinated children”
- “Health Impact News has all the markings of a crank site. For instance, its banner claims it is a site for ’News that impacts your health that other media sources may censor.’ ”
- “They have a pretty blue and red bar graph that’s just itching to be shredded, so let’s do it. This blue and red bar graph is designed to demonstrate that vaccinated children are more likely to develop certain medical conditions, such as asthma and seizures, than unvaccinated children. Pretty scary stuff, if their evidence were actually true.”
- “This study fails miserably at defining its population. The best I can tell, the comparison is between a population in an observation study called KIGGS and respondents to an open invitation survey conducted at vaccineinjury.info.”
- “They are comparing apples to rotten oranges.”


25,0%
20,0%
15,0%
10,0%
5,0%
0,0%
22,9%
10,6%
18,0%
2,4%
13,2%
7,0%
12,8%
0,2%
11,0%
2,0%
10,7%
2,6%
7,9%
2,0%
5,3%
0,5%
3,6%
0,3%
2,5%
1,1%
1,7%
0,1%
0,1%
0,1%
KIGGS
Survey unvaccinated children
Allergies
Asthma/ chronic Bronchitis
Neurodermatitis
Herpes
Otitis media
Hayfever
Hyperactivity
Skoliosis
Epilepsy/Seizures
Migraine
Thyroid disease
Diabetes mellitus

Aside: modelling epidemics

- Kendrick and McCormack Model: population of size N in three 'compartments'
- S, susceptible to infection; I, infectious; and R, removed/immune SIR model
- assumptions: susceptible individuals have no immunity; infectious individuals are currently infected, and can transmit this infection to susceptibles:
$$S \longrightarrow I \longrightarrow R. \tag{1}$$
- S, I and R change with time: dynamics
$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)I(t), \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \gamma I(t), \end{aligned}$$
- $\beta > 0$ and $\gamma > 0$ are two parameters special to the epidemic.
- β is the transmission rate per capita rate at which individuals come into contact with each other, and transmit the infection
- γ is the recovery rate, the rate at which infectious individuals move into the R category.

... aside

- $$\frac{dS(t)}{dt} = -\beta S(t)I(t), \quad \frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t),$$
- If $S(0) = N$, then the first infectious individual can be expected to infect
  $$R_0 = \beta N/\gamma$$
  individuals.
- R_0 is the basic reproduction number of the epidemic. different from $R(0)$.
- assumes the population is closed, i.e. $N = S + I + R$,
- doesn't allow for randomness in the infection rate, for example.
- Models for ebola add extra equations for various generalizations, allowing immigration and emigration from the population, changing rates of transmission, and so on.

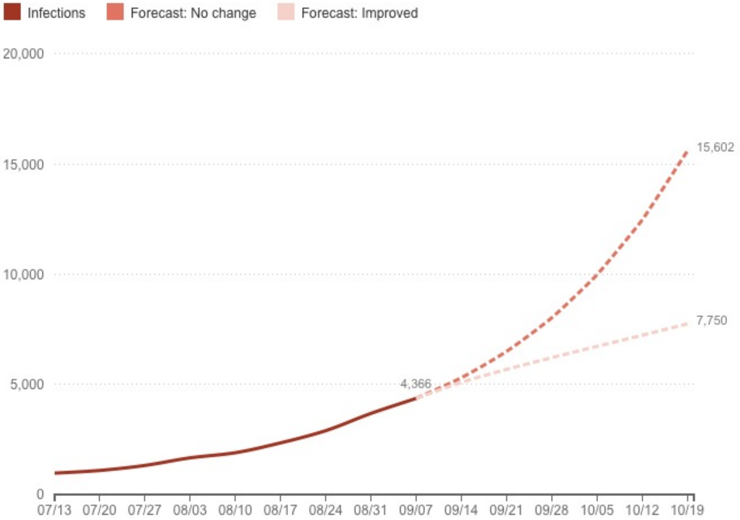
References: Senn, S. (2003). *Dicing with Death*, Ch. 9. Cambridge University Press, Cambridge.
Earn, D.J.D. (2008). A light introduction to modelling recurrent epidemics. In *Lecture Notes in Mathematical Epidemiology*, edited by F. Brauer, P. van den Driessche, J. Wu, (Springer), pp. 3-18.

Pi in the Sky: Volume 8.

Predicting How The Ebola Epidemic Will Grow

Link

Researchers at Columbia University developed a model to forecast how the current Ebola epidemic might continue through mid-October, based on the infection rates as of Sept. 7. The "no change" forecast assumes that current efforts at stopping the virus will continue at the same rate of effectiveness. The "improved" forecast assumes that interventions will become more effective.



Source: Columbia Prediction of Infectious Diseases, World Health Organization

Credit: Alyson Hurt/NPR

Model Choice

Cox & Donnelly §7.3

- often this will involve at least two levels of choice, first between distinct separate families and then between specific models within a chosen family
- of course all choices are to some extent provisional
- example: survival data – gamma or weibull model both extend the exponential
- example: linear regression $E(Y) = \beta_0 + \beta_1 x$, or nonlinear regression $E(Y) = \gamma_0/(1 + \gamma_1 x)$
- neither, one, or both may be adequate

... choice of a specific model

- comparisons between models are sometimes made using Bayes factors, ... however, misleading if neither model is adequate
- for dependencies of y on x that are curved, a low-degree polynomial might be adequate
- but subject-matter may suggest an asymptote, in which case $E(Y) = \alpha + \gamma e^{-\delta x}$ may be preferred

... model choice with a natural hierarchy

- polynomials provide a flexible family of smooth relationships, although poor for extrapolation
- it will typically be wise to measure the x_i from a meaningful origin near the centre of the data
- example:
  $E(Y) = \beta_{00} + \beta_{10}x_1 + \beta_{01}x_2 + \beta_{20}x_1^2 + \beta_{11}x_1x_2 + \beta_{02}x_2^2$
- it would not normally be sensible to include β_{11}, and not β_{20}, β_{02}
- with qualitative (categorical) x's, this means models with interaction terms should include the corresponding main effects

... model choice

- example: $E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$
- example: time series $AR(p)$
  $y_t = \mu + \rho_1(y_{t-1} - \mu) + \cdots + \rho_p(y_{t-p} - \mu) + \epsilon_t$
- for a single set of data choose the smallest order compatible with the data, using standard tests
- for several sets of data, usually would choose the same order for each set

... choosing among explanatory variables

- response y, potential explanatory variables $x_1, \ldots, x_p$
- suppose interest focusses on the role of a particular variable or set of variables, x^*
- the value, standard error, and interpretation of the coefficient of x^* depends on which other variables are included
- variables prior to x^* in the generating process should be included in the model unless...
- unless these variables are conditionally independent of y, given x^* (and other variables in the model)
- OR unless they are conditionally independent of x^*, given other variables in the model
- variables intermediate between x^* and y are omitted in initial assessment of the effect of x^*
- but may be needed later to study the pathways of dependence

... choosing among explanatory variables

- relatively mechanical methods of choosing may be helpful in preliminary exploration, but are insecure as a basis for final interpretation
- explanatory variables not of direct interest, but known to have a substantial effect, should be included
- several different models may be equally effective
- if there are several potential explanatory variables on an equal footing, interpretation is particularly difficult

- A two-phase approach:
- First search among a large number of possibilities for a base for interpretation
- Second check the adequacy of that base

First phase: a broad strategy

- x^*, required explanatory variables; $\tilde{x}$ some potential further explanatory variables
- $\tilde{x}$ conceptually prior to x^*

- fit a reduced model with x^* only $\mathcal{M}_{\text{red}}$
- fit, if possible, a full model with x^* and $\tilde{x}$ $\quad \mathcal{M}_{\text{full}}$
- compare the estimated standard errors of the coefficients for x^* under the two models

- if these are of the same order, then $\mathcal{M}_{\text{full}}$ is safer
- if precision improvement under $\mathcal{M}_{\text{red}}$ seems substantial, then explore eliminating some of $\tilde{x}$
- for example with backwards elimination

- with emphasis on the effect of x^*

Second phase: adequacy of the model

- add back selected components of the omitted variables $\tilde{x}$
- to check that conclusions are not changed
- or to report on the differences if they are
- if the model to date has been linear, may be important now to check some curvature terms, for continuous xs, and interaction terms for categorical xs
- these provide a 'warning system', but not usually direct interpretation

- interpretation of coefficients, especially in observational studies, needs care
- example: x includes several measurements of smoking behaviour: yes/no; years since quitting; no. of cigarettes smoked; pipe/cigar; etc.
- role of these depends on the goal of the study – confounder? primary exposure?

Nonparametric Regression

ELM, Ch. 11; SM, §10.7

- model $y_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, n \quad x_i$ scalar
- mean function $f(\cdot)$ assumed to be "smooth"
- introduce a kernel function $K(u)$ and define a set of weights
$$w_i = \frac{1}{\lambda} K\left(\frac{x_i - x_0}{\lambda}\right)$$
- estimate of $f(x)$, at $x = x_0$:
$$\hat{f}_\lambda(x_0) = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$
- Nadaraya-Watson estimator – local averaging

Kernel smoothers

ELM §11.1

- choice of bandwidth, λ controls smoothness of function
- larger bandwidth $=$ more smoothing
- kernel estimators are biased
- making the estimate smoother increases bias, decreases variance

- choice of kernel function, $K(\cdot)$, controls smoothness and "local-ness"
- Faraway recommends Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2), |x| \le 1$
- `ksmooth(base)` offers only uniform (box) or normal
- `bkde(KernSmooth)` offers `normal, box, epanech, biweight, triweight`
- `biweight`: $K(x) = (1 - |x|^2)^3, |x| \le 1$

Example 10.31; 6.38

SM

and has form $\lambda(y)$. This is the hazard function corresponding to the density of interval lengths, f. Statistical analysis for such a process is straightforward. Time series tools such as the correlogram and partial correlogram can be used to find serial dependence among successive intervals between events, though it may be clear from the context that these are independent. If independent and stationary, they can be treated as a random sample from f and inference performed in the usual way. ■

Example 6.37 (Birth process) In a birth process the intensity at time t depends on the number of previous events. Assuming that the number n of events up to t is finite, then $\lambda_{\mathcal{H}}(t) = \beta_0 + \beta_1 n$, where $\beta_0 > 0$, $\beta_1 \geq 0$. The complete intensity function is a step function which jumps β_1 at each event; if $\beta_1 = 0$ the process is a homogeneous Poisson process. ■

Before giving a numerical example, we briefly describe two functions useful for model checking and exploratory analysis of stationary processes.

The *variance-time curve* is defined as $V(t) = \text{var}\{N(t)\}$, for $t > 0$. A homogeneous Poisson process of intensity λ has $V(t) = \lambda t$, comparisons with which may be informative. Estimation of $V(t)$ is described in Problem 6.12.

The *conditional intensity function* is defined as

$$m_f(t) = \lim_{\delta s, \delta t \to 0} (\delta t)^{-1} \Pr\{N(t, t+\delta t) > 0 \mid N(-\delta s, 0) > 0\}, \quad t > 0,$$

which gives the intensity of events at t conditionally on there being an event at the origin. Evidently $m_f(t) = \lambda$ for a homogeneous Poisson process. An event at time t need not be the first event after that at the origin.

Example 6.38 (Japanese earthquake data) Figure 6.19 shows the times and magnitudes of earthquakes with epicentre less than 100km deep in an offshore region west of the main Japanese island of Honshū and south of the northern island of Hokkaidō. The figure shows all 483 earthquakes of magnitude 6 or more on the Richter scale in the period 1885–1980, about 5 tremors per year, in one of the most seismically active areas of Japan. A cumulative plot of the times rises fairly evenly and suggests that the data may be regarded as stationary; we shall assume this below. We take days as the units, giving $t_0 = 35{,}175$.

This is a *marked point process*, as in addition to the event times there is a *mark* — the magnitude — attached to each event. If we let the times be $0 < t_1 < \cdots < t_n < t_0$ and the associated magnitudes $m_1, \ldots, m_n$, their joint density may be written

$$\prod_{j=1}^{n} f(m_j \mid m_{(j-1)}, t_{(j)}) \prod_{j=1}^{n} f(t_j \mid m_{(j-1)}, t_{(j-1)}), \qquad (6.42)$$

where $t_{(j-1)}$ and $m_{(j-1)}$ represent $t_1, \ldots, t_{j-1}$ and $m_1, \ldots, m_{j-1}$. Here we concentrate on inference for the times using the second term, leaving the magnitudes to Examples 10.7 and 10.31. The lower panels of Figure 6.19 show the estimated variance-time curve and conditional intensity function for the times, which are are clearly far from Poisson. The variance-time curve grows more quickly than for a Poisson process, indicating clustering of events, and this is confirmed by the

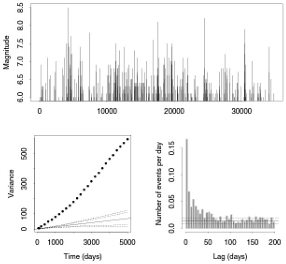


Figure 6.19 Japanese earthquake data (Ogata, 1988). The upper panel shows the times and magnitudes (Richter scale) of 483 shallow earthquakes. Lower left: estimated variance-time curve for earthquake times, with theoretical line for a Poisson process (solid) and two-sided 95% and 99% pointwise confidence limits (dots). Lower right: estimated conditional intensity, with baseline for Poisson process (solid) and two-sided 95% pointwise confidence limits (dots).

conditional intensity: for about 2–3 months after each shock the probability of another is increased.

One possible model for such data is a *self-exciting process* in which

$$\lambda_{\mathcal{H}}(t) = \mu + \sum_{j: t_j < t} w(t - t_j),$$

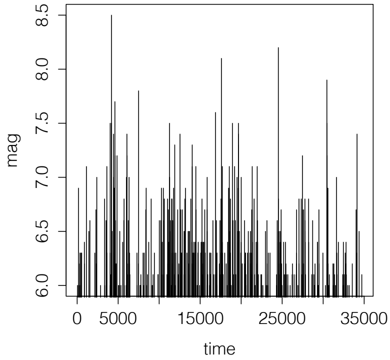
where μ is a positive constant and $w(u)$ is non-negative for $u > 0$ and otherwise zero. Here the intensity at any time is affected by the occurrence of previous events; often $w(u)$ is monotonic decreasing, so recent events affect the current intensity more than distant ones. This may be interpreted as asserting that events occur in clusters, whose centres occur as a Poisson process of rate μ. Subsidiary events are then spawned by the increase in intensity that occurs due to the superposition of the $w(t - t_j)$ for previous events. Seismological considerations suggest letting this function depend on m_j also, taking

$$w(t - t_j; m_j) = \frac{\kappa e^{\beta(m_j - 6)}}{(t - t_j + \gamma)^{\rho}}, \quad t > t_j,$$

where $\rho, \gamma, \kappa, \beta, \mu > 0$, with $\beta \doteq 2$. Under this formulation the increase in intensity depends not only on the time since an event but also on its magnitude.

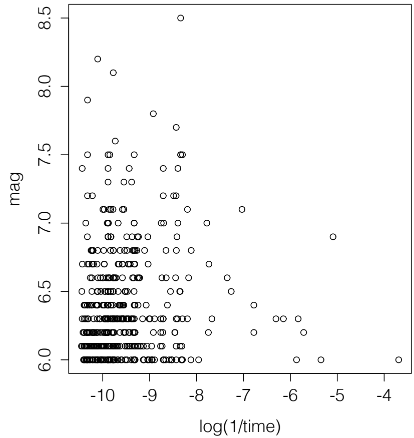
... example 10.31

SM




```
library(SMPracticals); data(quake)
with(quake, plot(time, mag, type="h"))
```

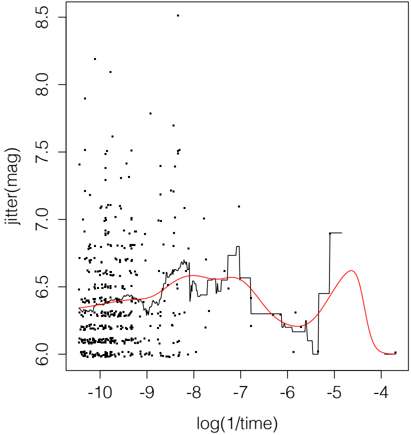
... example 10.31




```
library(SMPracticals); data(quake)
with(quake,plot(log(1/time),mag))
```

... example 10.31

SM



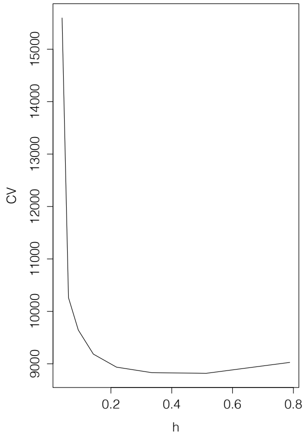

```
library(SMPracticals); data(quake)
with(quake,plot(log(1/time,jitter(mag)), pch = ".", cex = 2)
lines(ksmooth(log(1/quake$time),quake$mag))
lines(ksmooth(log(1/quake$time),quake$mag, kernel = "normal", bandwidth = 1), col = "red")
```

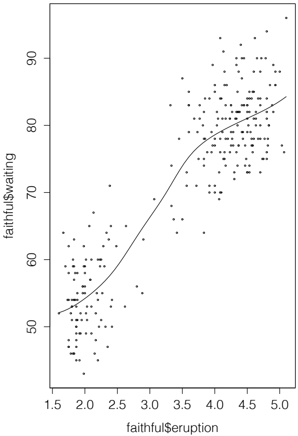
Bias and MSE

- $\hat{f}_\lambda(x)$ is biased: $\mathrm{E}\{\hat{f}_\lambda(x)\} \doteq \frac{1}{2}\lambda^2 f''(x)$
- $$\mathrm{var}\{\hat{f}_\lambda(x)\} \doteq \frac{\sigma^2}{n\lambda f_\lambda(x)} \int K^2(u)du$$
- could choose λ to minimize MSE $=$ bias^2 $+$ var, at x
- could choose λ to minimize integrated MSE
- more usual to use cross-validation
- $$CV(\lambda) = \sum_{i=1}^{n} \{y_i - \hat{f}_{-i}(x_i)\}^2$$

Cross-validation

```
library(faraway); data(faithful)
head(faithful)
 eruptions waiting
1     3.600      79
2     1.800      54
```





Local Polynomials

- better estimates can be obtained using local regression at point x
- $$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & (x_1 - x_0) & \cdots & (x_1 - x_0)^k \\ \vdots & \vdots & & \vdots \\ 1 & (x_n - x_0) & \cdots & (x_n - x_0)^k \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$
- $$\hat{\beta} = (X^T W X)^{-1} X^T W y$$
- $$\hat{f}_\lambda(x_0) = \hat{\beta}_0$$
- usually evaluate the function at sample points: $\hat{f}_\lambda(x_i), i = 1, \ldots, n$

... local polynomials

- odd-order polynomials work better than even; usually local linear fits are used
- kernel function is often a Gaussian density, or the tricube function
- as with Nadaraya-Watson estimators, choice of bandwidth, λ controls smoothness of function
- `loess` computes local linear regression (robustified)

Inference after fitting local polynomials

SM §10.7

- $\hat{\beta} = (X^T W X)^{-1} X^T W y$
- $W = \text{diag}(w_1, \ldots, w_n)$
- $\hat{f}_\lambda(x_0) = \hat{\beta}_0 = \sum_{i=1}^n S(x_0; x_i, \lambda) y_i$
- $S(x_0; x_1, \lambda), \ldots, S(x_0; x_n, \lambda)$ first row of "hat" matrix

  $(X^T W X)^{-1} X^T W$
- $\text{E}\{\hat{f}_\lambda(x_0)\} = \sum_{i=1}^n S(x_0; x_i, \lambda) f_\lambda(x_i)$
- $\text{var}\{\hat{f}_\lambda(x_0)\} = \sigma^2 \sum_{i=1}^n S(x_0; x_i, \lambda)^2$
- similarly $\hat{f} = (\hat{f}_\lambda(x_1), \ldots, \hat{f}_\lambda(x_n)) = \mathcal{S}_\lambda y$
- $\nu_1 = \text{tr}(S_\lambda)$, $\nu_2 = \text{tr}(S_\lambda^T S_\lambda)$

  potential estimates of 'degrees of freedom'

... cross-validation

- $$CV(\lambda) = \sum_{i=1}^{n} \{y_i - \hat{f}_{-i}(x_i)\}^2$$
- for local polynomials
$$CV(\lambda) = \sum_{i=1}^{n} \left\{ \frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_{ii}(\lambda)} \right\}^2$$
- simpler
$$GCV(\lambda) = \sum_{i=1}^{n} \left\{ \frac{y_i - \hat{f}_\lambda(x_i)}{1 - \text{tr}(S_\lambda)/n} \right\}^2$$
- $$\hat{f}_\lambda(x_0) = \hat{\beta}_0 = \sum_{i=1}^{n} S(x_0; x_i, \lambda) y_i$$
- $S(x_0; x_1, \lambda), \ldots, S(x_0; x_n, \lambda)$ is first row of $(X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W$

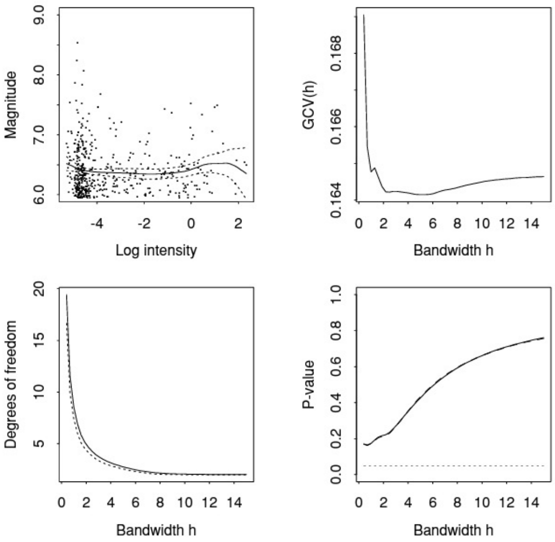


Figure 10.16 Smooth analysis of earthquake data. Upper left: local linear regression of magnitude on log intensity just before quake (solid), with 0.95 pointwise confidence bands (dots). Upper right: generalized cross-validation criterion GCV(h) as a function of bandwidth h. Lower left: relation between degrees of freedom ν_1 (solid), ν_2 (dots), and h. Lower right: significance traces for test of no relation between magnitude and log intensity, based on chi-squared approximation (dots) and saddlepoint approximation (solid). The horizontal line shows the conventional 0.05 significance level.

and has form $\lambda(v)$. This is the hazard function corresponding to the density of interval lengths, f. Statistical analysis for such a process is straightforward. Time series tools such as the correlogram and partial correlogram can be used to find serial dependence among successive intervals between events, though it may be clear from the context that these are independent. If independent and stationary, they can be treated as a random sample from f and inference performed in the usual way. ■

Example 6.37 (Birth process) In a birth process the intensity at time t depends on the number of previous events. Assuming that the number n of events up to t is finite, then $\lambda_{\mathcal{H}}(t) = \beta_0 + \beta_1 n$, where $\beta_0 > 0$, $\beta_1 \geq 0$. The complete intensity function is a step function which jumps β_1 at each event; if $\beta_1 = 0$ the process is a homogeneous Poisson process. ■

Before giving a numerical example, we briefly describe two functions useful for model checking and exploratory analysis of stationary processes.

The *variance-time curve* is defined as $V(t) = \text{var}\{N(t)\}$, for $t > 0$. A homogeneous Poisson process of intensity λ has $V(t) = \lambda t$, comparisons with which may be informative. Estimation of $V(t)$ is described in Problem 6.12.

The *conditional intensity function* is defined as

$$m_f(t) = \lim_{\delta s, \delta t \to 0} (\delta t)^{-1} \Pr\{N(t, t + \delta t) > 0 \mid N(-\delta s, 0) > 0\}, \quad t > 0,$$

which gives the intensity of events at t conditionally on there being an event at the origin. Evidently $m_f(t) = \lambda$ for a homogeneous Poisson process. An event at time t need not be the first event after that at the origin.

Example 6.38 (Japanese earthquake data) Figure 6.19 shows the times and magnitudes of earthquakes with epicentre less than 100km deep in an offshore region west of the main Japanese island of Honshū and south of the northern island of Hokkaidō. The figure shows all 483 earthquakes of magnitude 6 or more on the Richter scale in the period 1885–1980, about 5 tremors per year, in one of the most seismically active areas of Japan. A cumulative plot of the times rises fairly evenly and suggests that the data may be regarded as stationary; we shall assume this below. We take days as the units, giving $t_0 = 35{,}175$.

This is a *marked point process*, as in addition to the event times there is a *mark* — the magnitude — attached to each event. If we let the times be $0 < t_1 < \cdots < t_n < t_0$ and the associated magnitudes $m_1, \ldots, m_n$, their joint density may be written

$$\prod_{j=1}^{n} f(m_j \mid m_{(j-1)}, t_{(j)}) \prod_{j=1}^{n} f(t_j \mid m_{(j-1)}, t_{(j-1)}), \tag{6.42}$$

where $t_{(j-1)}$ and $m_{(j-1)}$ represent $t_1, \ldots, t_{j-1}$ and $m_1, \ldots, m_{j-1}$. Here we concentrate on inference for the times using the second term, leaving the magnitudes to Examples 10.7 and 10.31. The lower panels of Figure 6.19 show the estimated variance-time curve and conditional intensity function for the times, which are are clearly far from Poisson. The variance-time curve grows more quickly than for a Poisson process, indicating clustering of events, and this is confirmed by the

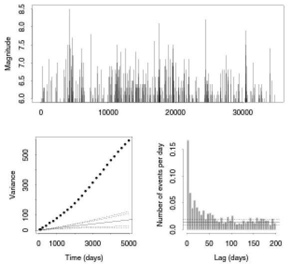


Figure 6.19 Japanese earthquake data (Ogata, 1988). The upper panel shows the times and magnitudes (Richter scale) of 483 shallow earthquakes. Lower left: estimated variance-time curve for earthquake times, with theoretical line for a Poisson process (solid) and two-sided 95% and 99% pointwise confidence limits (dots). Lower right: estimated conditional intensity, with baseline for Poisson process (solid) and two-sided 95% pointwise confidence limits (dots).

conditional intensity: for about 2–3 months after each shock the probability of another is increased.

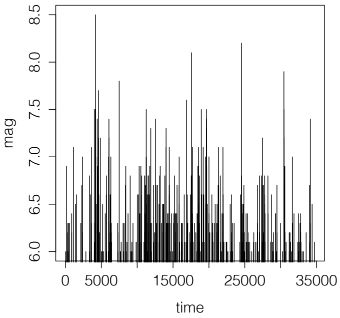
One possible model for such data is a *self-exciting process* in which

$$\lambda_{\mathcal{H}}(t) = \mu + \sum_{j: t_j < t} w(t - t_j),$$

where μ is a positive constant and $w(u)$ is non-negative for $u > 0$ and otherwise zero. Here the intensity at any time is affected by the occurrence of previous events; often $w(u)$ is monotonic decreasing, so recent events affect the current intensity more than distant ones. This may be interpreted as asserting that events occur in clusters, whose centres occur as a Poisson process of rate μ. Subsidiary events are then spawned by the increase in intensity that occurs due to the superposition of the $w(t - t_j)$ for previous events. Seismological considerations suggest letting this function depend on m_j also, taking

$$w(t - t_j; m_j) = \frac{\kappa e^{\beta(m_j - 6)}}{(t - t_j + \gamma)^{\rho}}, \quad t > t_j,$$

where $\rho, \gamma, \kappa, \beta, \mu > 0$, with $\beta \doteq 2$. Under this formulation the increase in intensity depends not only on the time since an event but also on its magnitude.




```
quake$intens = log10(1/quake$time)
with(quake,plot(intens,mag))
quake.lo <- loess(mag ~ intens, data = quake)
try <- predict(quake.lo, data.frame(intens = seq(-5,-1.5,.1)),se = TRUE)
lines(seq(-5,-1.5,.1),try$fit+ 2*try$se.fit)
```