Statistical Inference:
666 Exercises, 66 Stories, and Solutions

Nils Lid Hjort
University of Oslo

Emil Aas Stoltenberg
BI Norwegian Business School

– This version, with 0.90 versions of a subset of 21 Statistical Stories, last touched by Nils and by Emil, 23-Aug-2023 –

Some technical stuff

ISBN - Numbers numbers

The Kioskvelter Project

1234-5678

To my somebody
– N.L.H.

To my somebody
– E.A.S.

Preface

This book builds on Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

(xx then several crisp paragraphs here, on the carrying ideas behind and structure of the book: *exercises* and *stories.* a partly flipped classroom, with direct participation from the first pages of each chapter, also on prerequisties: linear algebra, with matrix theory, etc.; calcululs, with functions of one or more variables, partial derivatives, etc.; programming, in R or Python or other appropriate language, for simulation etc.)

The authors owe special thanks to Céline Cunen, Gudmund Hermansen, Tore Schweder, for having contributed significantly to several of our Statistical Stories, and also for always pleasant and inspiring long-term collaborations. Many thanks are also due to a long list of colleagues and friends, who have taken part in discussions and rounds of clarification of relevance to various exercises and stories in our book: Marthe Aastveit, Patrick Ball, Bear Braumoeller, Aaron Clauset, Dennis Cristensen, Ingrid Dæhlen, Åsa Engelstad, Arnoldo Frigessi, Ingrid Glad, Håvard Hegre, Aliaksandr Hubin, Ingrid Hobæk Haff, Kristoffer Hellton, Bjørn Jamtveit, Martin Jullum, Vinnie Ko, Alexander Koning, Per Mykland, Jonas Moss, Håvard Mokleiv Nygård, Lars Olsen, Catharina Stoltenberg, Gunnar Taraldsen, Ingunn Fride Tvete, Sam-Erik Walker, Lars Walløe, Jonathan Williams, Lan Zhang.

We have also benefited, directly and indirectly, through the collective efforts of grander wide-horizoned funded projects: the *FocuStat: Focus Driven Statistical Inference with Complex Data* 2014-2019 project (led by Hjort) at the Department of Mathematics, University of Oslo, funded by the Norwegian Research Council; the *Stability and Change* 2022-2023 project (led by Hegre and Hjort) at the Centre for Advanced Study, Academy of Science and Letters, Oslo; the grand *Integreat: The Norwegian Centre for Knowledge-Driven Machine Learning* 2023-2033 Centre of Excellence (led by Frigessi and Glad), Oslo, funded by the Norwegian Research Council. We finally acknowledge with gratitude a partial support stipend from the Norwegian Non-Fiction Writers and Translators Association (Norsk faglitterær forfatter- og oversetterforening).

(xx Then current time plan, as of 23-Aug-2023, possibly optimistic xx)

Nils Lid Hjort and Emil Aas Stoltenberg
Blindern, some day in 2023

Contents

Part I

Short & crisp

I.14

Bootstrapping

Part II

Stories

II.viii

Statistical Stories: a Subset (for stk4011)

(xx WELL: many yhings to clean, as of 23-Aug-2023. In the full Hjort-Stoltenberg book, there will be around 66 Stories, sorted into categories (i) Epidemiology and Medicine; (ii), Art, Literature, Music; (iii) Social Sciences; (iv) Biology, Climate, Ecology; (v) Sports; (vi) Simulation. xx)

Story viii.1 *Cooling of newborns.* Seminal work carried out by Marianne Thoresen and coworkers (xx put in reference here praps 2005 paper xx) has demonstrated that when a newborn has been deprived of oxygen during birth, an emergency intervention involving cooling (therapeutic hypothermia) can save its life, with no loss of motoric or mental abilities later on – provided this is implemented within six hours. Is it still helpful, or not at all, when the cooling scheme starts later than six hours? Laptook (2017) report on a wide and elaborate study, combining information from many registries across several U.S. state, pertaining to this and related question. In particular, one counts the number of events, in the cooled and non-cooled groups, the event in question being death or disability (with a precise definition of disability, assessed when the child is about 18 months old). The essential relevant summary, from all these life-and-death efforts, lies in the two times two table

```
non-cooled infants   m0: y0 and m0-y0:   79: 22 and 57
hypothermic infants: m1: y1 and m1-y1:   78: 19 and 59
```

Seeing these as two biomial experiments, $y_0 \sim \text{binom}(m_0, p_0)$ and $y_1 \sim \text{binom}(m_1, p_1)$, the statistical question is what inferences we may make, for comparing p_0 and p_1.

(a) First, give ordinary (and perhaps approximate) 95 percent confidence intervals for p_0 and p_1, and comment. Then compute and display in the same diagram the confidence distributions $\text{cc}_0(p_0)$ and $\text{cc}_1(p_1)$, associated with the optimal binomial confidence distributions $C(p) = P_p(Y > y_{\text{obs}}) + \frac{1}{2}P_p(Y = y_{\text{obs}})$, as for the left panel pf Figure viii.1; see Ex. 7.25.

(b) To analyse the degree to which p_0 and p_1 might be different, transform to the logistic scale, with $p_0 = \exp(\theta_0)/\{1 + \exp(\theta_0)\}$ and $p_1 = \exp(\theta_0 + \gamma)/\{1 + \exp(\theta_0 + \gamma)\}$. Note that γ can be seen as the log-odds difference $\log(p_1/(1 - p_1)) - \log(p_0/(1 - p_0))$; results

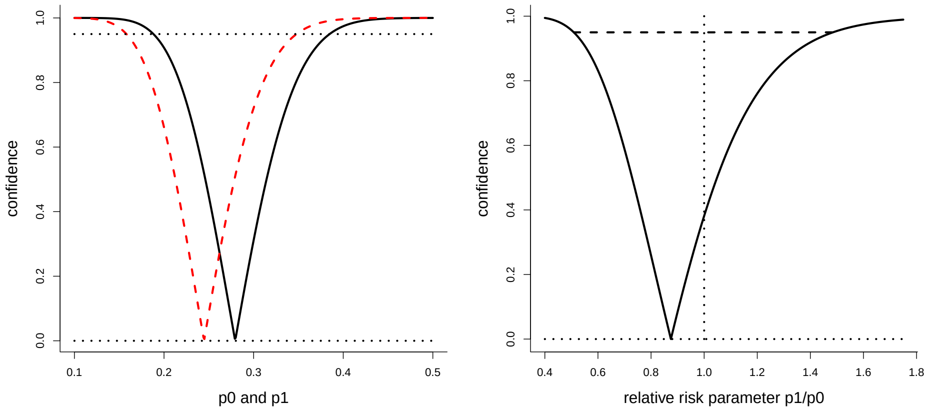

Figure viii.1: *Left panel: confidence curves for p_0 and p_1, with 95 intervals [0.189,0.384] and [0.159,0.347]. Right panel: confidence curve for $\mathrm{rr} = p_1/p_0$, with 95 interval [0.509,1.485]. There is no indication that p_0 and p_1 really differ.*

may also be given in terms of the odds ratio $\rho = \exp(\gamma)$. Use Ex. 7.27 to compute the optimal confidence curve $\mathrm{cc}(\rho)$, and give a 95 percent interval for the parameter. The Laptook (2017) article reported mainly in terms of the relative risk parameter $\mathrm{rr} = p_1/p_0$, however. Construct therefore a confidence distribution cc(rr) also for that parameter, using the Wilks theorem based recipe of Ex. **??**, as in the right panel of Figure viii.1. Give the median confidence estimate $\widehat{\mathrm{rr}}_{0.50}$ and a 95 percent interval.

(c) The Laptook (2017) report framed results in terms of Bayesian priors and posteriors. With priors $p_0 \sim \mathrm{Beta}(a_0, b_0)$ and $p_1 \sim \mathrm{Beta}(a_1, b_1)$, show that $\mathrm{rr} = p_1/p_0$ given data is a ratio of independent $\mathrm{Beta}(a_0 + y_0, b_0 + m_0 - y_0)$ and $\mathrm{Beta}(a_1 + y_1, b_1 + m_1 - y_1)$. Deduce that the posterior distribution has cumulative

$$F(v \,|\, \mathrm{data}) = \int_0^1 G(vp_0, a_1 + y_1, b_1 + m_1 - y_1) g(p_0, a_0 + y_0, b_0 + m_0 - y_0)\, \mathrm{d}p_0,$$

in terms of the density g and c.d.f. G for Beta distributions. Compute and display this $F(\mathrm{rr} \,|\, \mathrm{data})$, using the Jeffreys prior $\mathrm{Beta}(\frac{1}{2}, \frac{1}{2})$ (xx pointer to Ch5 with this detail xx), Show that it becomes very close to the prior-free CD $C(\mathrm{rr})$, constructed from the cc(rr) above as $\frac{1}{2} - \frac{1}{2}\mathrm{cc}(\mathrm{rr})$ for $\mathrm{rr} \le \widehat{\mathrm{rr}}_{0.50}$ and $\frac{1}{2} + \frac{1}{2}\mathrm{cc}(\mathrm{rr})$ for $\mathrm{rr} \ge \widehat{\mathrm{rr}}_{0.50}$ (xx pointer to that thing in Ch7 xx).

(d) Laptook (2017) used several informative priors for their analyses, including one called by them a neutral prior, with mean zero and standard deviataion 0.35 for $\log \mathrm{rr}$. Translate this to two equal Beta priors $(a, a), (a, a)$, finding the a matching their 0.35 standard deviation, perhaps via simulations; you should find $a \doteq 8.95$. For this neutral prior, display the posterior c.d.f. and density for rr. Compute also $P(\mathrm{rr} \le v \,|\, \mathrm{data})$ for $v = 0.90, 0.95, 1.00$.

(e) Above we have computed $P(\mathrm{rr} < 1 \,|\, y_0 = 22, y_1 = 19) = 0.664$ with the neutral prior. Compute $P(\mathrm{rr} < 1 \,|\, y_0 = 22, y_1)$ for imagined data sets with $y_1 = 19, 18, \ldots, 5$, say, keeping the other aspects of the data fixed, including $y_0 = 22$. How small ought y_1 to have been, in order for the $rr < 1$ scenario to have posterior probabilty above 0.95?

(f) (xx round off. point to Hjort blog story Hjort (2017a), big JAMA paper Laptook (2017), short critical follow-up papers Walløe et al. (2019a,b). xx)

Story viii.2 *Suicide attempt rates for Paroxetine vs. placebo.* There are several studies of the effects and side effects of the antidepressant Paroxetine (sold under brand names Seroxat, Paxil, and yet others, since 1992). While beneficial for hundreds of thousands of users, serious concerns are also part of the broader picture, with one particularly disturbing aspect being its potential association with suicidal thoughts and actions. Here we use data and information from Aursnes et al. (2005, 2006), who used Bayesian analyses with informative priors, based on data and other information available to those authors in respectively 2005 and 2006. Below we discuss these priors to posteriors calculations, but also include other non-Bayesian methods.

The data are as simple as two Poisson counts, $Y_0 \sim \mathrm{Pois}(m_0\theta_0)$ and $Y_1 \sim \mathrm{Pois}(m_1\theta_1)$, for the placebo and the drug groups, with m_0 and m_1 cumulative exposure time, here conveniently counted as patient years. The parameter of primary interest is $\gamma = \theta_1/\theta_0$. The articles pointed to concentrate on the probability that $\gamma > 1$, or, equivalently, that $\kappa > 0$, where $\kappa = \log(\theta_1/\theta_0)$ is a more convenient scale for computation and summary reporting, due the inherent strong right skewnesses involved on the γ scale.

For the studies in question, the 2005 article had $(y_0, y_1) = (1, 7)$, after $(m_0, m_1) = (73.3, 190.7)$ patient years, and used informative priors based on previous literature to conclude that it was rather likely that $\theta_1 > \theta_0$, i.e. an increased suicide attempt risk in the Paroxetine group. This was followed by media exposure and debate, along with critical comments from both individual researchers and from GlaxoSmithKline plc, the multinational pharmaceutical and biotechnology company manufacturing the drug. This again led to the 2006 article, by the same four authors, with more extensive data collection and also further care for accuracy. In summary, the data now had $(y_0, y_1) = (1, 11)$, after $(m_0, m_1) = (333.0, 601.0)$ patient years. The 2005 data are to be seen as part of the extended and more accurately curated 2006 dataset.

(a) With $Z = Y_0 + Y_1$, show that $Y_1 \,|\, z \sim \mathrm{binom}(z, m_1\gamma/(m_0 + m_1\gamma))$, with $\gamma = \exp(\kappa)$. Compute and display what is according to Ex. 7.24 the optimal confidence distribution,

$$C(\kappa) = P_\kappa(Y_1 > y_{1,\mathrm{obs}} \,|\, Y_0 + Y_1 = z_{\mathrm{obs}}) + \tfrac{1}{2} P_\kappa(Y_1 = y_{1,\mathrm{obs}} \,|\, Y_0 + Y_1 = z_{\mathrm{obs}}),$$

with the 2005-information and the 2006-information; construct versions of Figure viii.2, both on the κ and γ scales. Verify in particular that $C_{2005}(0) = 0.188$ and $C_{2006}(0) = 0.022$. Explain how these can be seen as p-values for testing $\theta_0 \le \theta_1$ against the drastic alternative that the antidepressant in question increases the suicide attempt risk. Discuss also how the complementary numbers 0.812 and 0.978 can be seen as epistemic probabiities for $\theta_1 > \theta_0$. Give also 95 percent confidence intervals, first for κ and then transformed back to the scale of θ_1/θ_0.

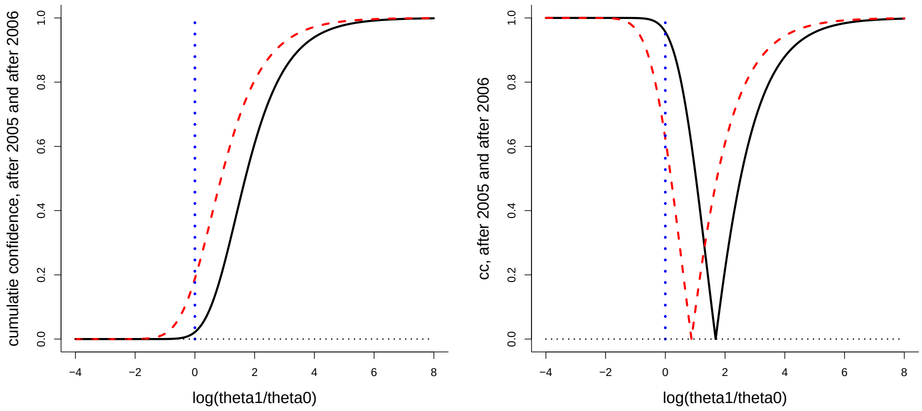

Figure viii.2: *Cumulative confidencec distributions (left panel) and confidence curves (right panel), for* $\kappa = \log(\theta_1/\theta_0)$*, based on information in the 2005 article (red, slanted) and in the 2006 article (black, full). The central question is whether* $\theta_1 > \theta_0$*, i.e. whether* $\kappa > 0$*. The evidence for this is much clearer with the 2006 information.*

(b) Suppose now that adequate prior distributions are set of the type $\theta_0 \sim \mathrm{Gam}(a_0, b_0)$ and $\theta_1 \sim \mathrm{Gam}(a_1, b_1)$. Show that this leads to clear posterior distributions

$$\theta_0 \,|\, \mathrm{data} \sim \mathrm{Gam}(a_0 + y_0, b_0 + m_0), \qquad \theta_1 \,|\, \mathrm{data} \sim \mathrm{Gam}(a_1 + y_1, b_1 + m_1).$$

Show that the posterior cumulative and density functions for $\kappa = \log(\theta_1/\theta_0)$ can be expressed as

$$\begin{aligned} F(\kappa \,|\, \mathrm{data}) &= \int_0^\infty G(\exp(\kappa)\theta_0, a_1 + y_1, b_1 + m_1)\, g(\theta_0, a_0 + y_0, b_0 + m_0)\,\mathrm{d}\theta_0, \\ f(\kappa \,|\, \mathrm{data}) &= \int_0^\infty g(\exp(\kappa)\theta_0, a_1 + y_1, b_1 + m_1)\exp(\kappa)\theta_0\, g(\theta_0, a_0 + y_0, b_0 + m_0)\,\mathrm{d}\theta_0, \end{aligned}$$

in terms of the cumulative and density $G(\cdot, a, b)$ and $g(\cdot, a, b)$ of the $\mathrm{Gam}(a, b)$. In particular, explain that $p_B = 1 - F(0 \,|\, \mathrm{data})$ is the posterior probability for the dramatic $\theta_1 > \theta_0$ scenario, building on both the priors and the Poisson counts (y_0, y_1) with (m_0, m_1) patient years. Give also corresponding expressions for the posterior cumulative and density on the direct scale of $\gamma = \theta_1/\theta_0$.

(c) For each of the 2005-infomation and 2006-information cases, compute and display

$$p_B = 1 - \int_0^\infty G(\theta_0, a + y_1, 50 + m_1)\, g(\theta_0, a + y_0, 50 + m_1)\,\mathrm{d}\theta_0$$

as a function of a, a common parameter in gamma prior parameters $(a, 50), (a, 50)$ for θ_0, θ_1, interpreted as the expected number of suicide attempts in the course of 50 patient years, for either the placebo or drug groups of patients. Comment on your findings.

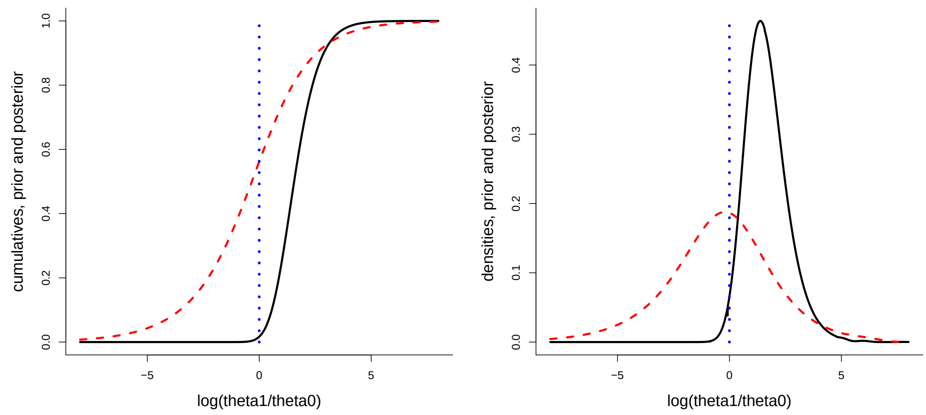

Figure viii.3: *From the informative slightly optimistic prior (red, slanted) to the posterior (black, full), using the 2006 data; cumulatives in left panel, densities in right panel. The evidence is very strong that* $\theta_1 > \theta_0$.

(d) Several informative priors are carefully argued for and worked with in Aursnes et al. (2005, 2006). "This does not mean that these parameters are to be interpreted as random variables, but our knowledge of the parameters is uncertain and we describe this uncertainty with the help of probability distributions," as they write, when setting their priors, in fact by attempting to match conclusions of earlier meta-analysis publications to gamma prior parameters. For illustration in the present story we are content with using one of these, called by them in their 2006 article the slightly optimistic prior, having $(a_0, b_0) = (0.71, 50)$ and $(a_1, b_1) = (0.58, 50)$. The idea was to quantify the expected number of suicide attempts for the placebo and the drug groups, in the course of 50 patient years, and with these expected numbers adding up to 1.29 attempts per 100 patient years, matching information recorded in previous literature. For this prior, work through the numerics, and display the prior and posterior densities, as well as the prior and posterior cumulatives, for the focus parameter $\log(\theta_1/\theta_0)$; construct versions of Figure viii.3. Find the 95 percent posterior interval for κ, and by transformation for θ_1/θ_0. Also, record the Bayesian answer p_B to the question of how likely we should think it is that the drug increases suicide attempt risk.

(e) Results reported on in Aursnes et al. (2005, 2006) were not reached via the precise integration tools above; the authors resorted rather to simulation. Carry out such work too, simulating say 10^5 realisations of $\log(\theta_2/\theta_1)$ from the posterior distribution, followed by simple density estimation, to reach a simulated version of Figure viii.3. As explained via the integration details above, however, there is no real need for simulation here.

(f) (xx something more. try noninformative priors, of the type $(0.1, 0.1)$ and $(0.1, 0.1)$ for θ_0 and θ_1. and something more neutral, like $(1, 50)$ and $(1, 50)$. xx)

(g) (xx something iiccff using Cunen and Hjort (2022). combining one of these informative priors with the new data. several paths possible. (i) using the informative priors for θ_0 and θ_1, then log-likelihoods for θ_0 and θ_1, i.e. four sources combined to find $\mathrm{cc}^*(\kappa)$. (ii) using the log-prior for κ coming out of two priors, and the converted log-likelihood for κ coming from $\mathrm{cc}(\kappa)$. both should work, but four sources in detail might be a bit more precise. xx)

(h) (xx the analyses above have presumed that θ_0 and θ_1 are somehow well-defined overall rate parameters, one for the Paroxetine users and one for the placebo group. more realistically, these suicide attempt rate parameters would vary in the population, e.g. with gender and age. argue that this could lead to negative binomial models for the final counts (Y_0, Y_1). perhaps are conclusions above too sharp. but we can't well answer this since we do not have data divided into any subcategories. xx)

Story viii.3 *Overdispersed children.* Some one and a half century ago, there were as many as $n = 38495$ plentiful 8-or-more children families living in Sachsen, with Geißler (1889) dutifully counting and reporting about them and the number of girls and boys. The little table to the left here gives the number $N(y)$ of these families having y girls and $8 - y$ boys, for $y = 0, 1, \ldots, 8$. In the course of this and the following Story i.4 we will work through models 1, 2, 3, say, producing expected numbers $E_1(y), E_2(y), E_3(y)$ to match the $N(y)$, along with what we term Pearson residuals $\{N(y) - E(y)\}/E(y)^{1/2}$.

y	N	E1	pear1	E2	pear2	E3	pear3
0	264	192.325	5.168	255.621	0.524	255.210	0.550
1	1655	1445.384	5.514	1657.032	-0.050	1655.181	-0.004
2	4948	4752.364	2.838	4909.686	0.547	4901.376	0.666
3	8498	8928.902	-4.560	8683.213	-1.988	8692.383	-2.085
4	10263	10484.952	-2.168	10024.863	2.378	10034.318	2.283
5	7603	7879.792	-3.118	7735.975	-1.512	7736.379	-1.516
6	3951	3701.205	4.106	3896.509	0.873	3890.978	0.962
7	1152	993.421	5.031	1171.238	-0.562	1167.280	-0.447
8	161	116.655	4.106	160.865	0.011	161.895	-0.070

(a) Compute the overall fraction of girls, among the $mn = 307860$ children, as $\widehat{p} = \sum_{y=0}^{m} N(y)y/(mn) = 0.4844$. Show that the null hypothesis $p = 0.50$ must be soundly rejected here.

(b) Of course a statistician can't always expect to see the difference between 0.500 and 0.485 as a clearly significant one – as this very much depends on the sample size. Suppose you go out on the street sampling, counting a binomial $B \sim \mathrm{binom}(k, p)$ after having studied k objects or persons. How large must k be, in order for your 0.05-level test of $p = 0.50$ against $p \neq 0.50$ to have detection power say 0.95, if the truth is $p = 0.485$? What if you use a 0.01-level test and need detection power 0.99?

(c) Assume that the binomial model $Y \sim \mathrm{binom}(8, p_0)$ holds, with the same p_0 across all families. Find point estimates and 99 percent confidence intervals for $nf_1(0, p_0)$, the expected number of all-boys families, and for $nf_1(8, p_0)$, the expected number of all-girls families, among the $n = 38495$ families with eight children. Then check with the real world.

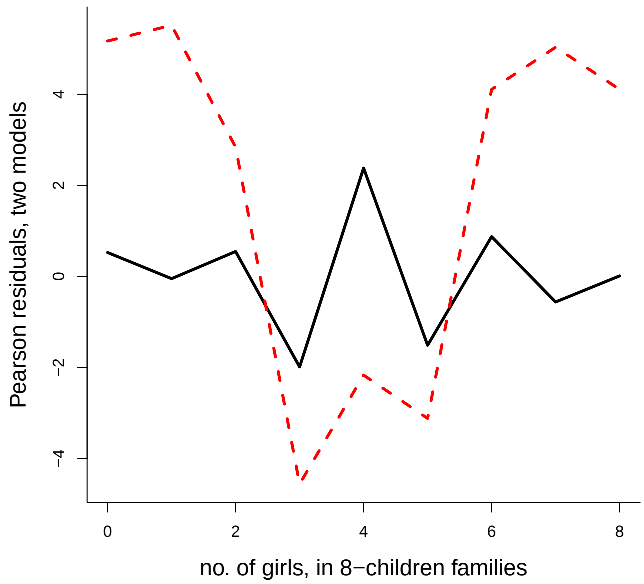

Figure viii.4: *Pearson residuals $\{N(y) - E(y)\}/E(y)^{1/2}$, for two models: the simple binomial (red, dashed curve) and the betabinomial (black, full curve). Here $N(y)$ is the observed number of girls y, and $E_1(y)$ and $E_2(y)$ the expected number under models 1 and 2.*

(d) Under the assumption that the girl-probability p is constant, across families, we would have $Y \sim \mathrm{binom}(8, p)$, for these $n = 38495$ Sachsen families. Compute $E_1(y) = nf_1(y, \widehat{p})$, the expected number of y-girls families, under this model, with $f_1(y, p)$ the usual binomial. Compute also the Pearson residual, say $P_1(y) = \{N(y) - E_1(y)\}/E_1(y)^{1/2}$, for $y = 0, 1, \ldots, 8$. These should roughly be standard normal, if the model used is good. Check with Figure viii.4. Discuss what you find: in particular, it appears that the real world exhibits significantly more 'extreme families', those with all boys or all girls, than what is predicted under the straight binomial model.

(e) Suppose rather that each family has its own girl-probability p, but that this p varies across families, according to some distribution with overall mean p_0 and positive standard deviation τ_0. Show that $\mathrm{E}\, Y = mp_0$ and that the extra-binomial variability manifests itself by

$$\mathrm{Var}\, Y = mp_0(1 - p_0) + m(m-1)\tau_0^2.$$

Compute the empirical variance $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2/(n-1) = \sum_{y=0}^m N(y)(y - m\widehat{p})^2/(n-1)$, set the extra-binomial variance $S^2 - m\widehat{p}_0(1 - \widehat{p}_0)$ equal to $m(m-1)\tau_0^2$, and show that this leads to $\widehat{\tau}_0 = 0.0538$.

(f) Establish that this extra-variation, with $\widehat{\tau}_0 = 0.0538$, is indeed very significantly positive. Again, we would not always be able to identify a standard deviation of this size as being significantly present, but we are, of course, helped by the enormous sample size.

(g) A natural two-parameter model, to explain also the extra-binomial variability, is to take $Y \,|\, p \sim \mathrm{binom}(m, p)$ and $p \sim \mathrm{Beta}(a, b)$; see Ex. 1.25. Show that this leads to

$$f_2(y, a, b) = \binom{m}{y} \frac{\Gamma(a+y)\Gamma(b+m-y)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+b)}{\Gamma(a+b+m)} \quad \text{for } y = 0, 1, \ldots, m.$$

Representing (a, b) as $(kp_0, k(1-p_0))$, estimate k from the overdispersion number $\widehat{\tau}_0 = 0.0538$. Draw the resulting Beta density in a diagram. How many families in the world have their girl-probabilities outside the interval $[0.40, 0.60]$?

(h) Compute the expected numbers $E_2(y)$ and the Pearson residuals $P_2(y) = \{N(y) - E_2(y)\}/E_2(y)^{1/2}$ also for this two-parameter model, and reconstruct both the table above and Figure viii.4.

(i) (xx a thing or two more. point to later chapter where we also do Markov modelling, from child to child inside given families. comment on $\sum_y P_1(y)^2 = 159.41$, way too big for the binomial, but $\sum_y P_2(y)^2 = 13.55$, close to acceptable, for the beta-binomial. we have $\sum_y P_3(y)^2 = 13.41$ for the Markov model. xx)

Story viii.4 *Boys are born slightly bigger than girls.* (xx quantile things, to be told. perestroika required. first separate quantiles, boys and girls, then ratios of quantiles. xx) $n_b = 548$ boys and $n_g = 480$ girls born in oslo. ratio of quantiles. let $f_b(x)$ and $f_g(x)$ be the birthweight densities, for boys and for girls, with cumulative distribution functions $F_b(x)$ and $F_g(x)$. here we shall compare quantiles for boys and girls, $\mu_{b,q} = F_b^{-1}(q)$ and $\mu_{g,q} = F_g^{-1}(\mu_g)$, at different levels q. may give a figure of estimated densities, standard kernel methods from Ch. 13. see Figures.

(a) Show that boys are significantly bigger than girls, but that there is no clear indication that they have different variances in their birthweight distributions.

(b) For each of the five quantile levels $0.1, 0.3, 0.5, 0.7, 0.9$, construct a CD for the $F^{-1}(q)$, for boys and for girls, using the order statistic method of Ex. 7.20. Compute also the consequent confidence curves $\mathrm{cc}(\mu_q)$, and make a version of Figure viii.5. (xx should get a better plot from nils com16c. xx)

(c) In the following, keep one quantile level q fixed, to avoid a too heavily subscripted notation. From Ex. 2.20 we know that $n_b^{1/2}(Q_b - \mu_b) \to_d \mathrm{N}(0, \kappa_b^2)$ and $n_g^{1/2}(Q_g - \mu_g) \to_d \mathrm{N}(0, \kappa_g^2)$, with $\kappa_b = \{q(1-q)\}^{1/2}/f_b(\mu_{b,q})$ and with $\kappa_g = \{q(1-q)\}^{1/2}/f_g(\mu_{b,g})$. Show that this entails

$$Q_b = \mu_b + \kappa_b/n_b^{1/2} Z_b \quad \text{and} \quad Q_g = \mu_g + \kappa_g/n_g^{1/2} Z_g,$$

where $Z_b = Z_{b,n_b}$ and $Z_g = Z_{g,n_g}$ have distributions coming (very) close to the standard normal.

(d) Estimate the difference between the boys and girls distributions, as a function of the quantile q, along with a confidence band. Construct a version of Figure viii.6, using gram. The horizonal line represents the estimated overall difference $d = \xi_b - \xi_g$.

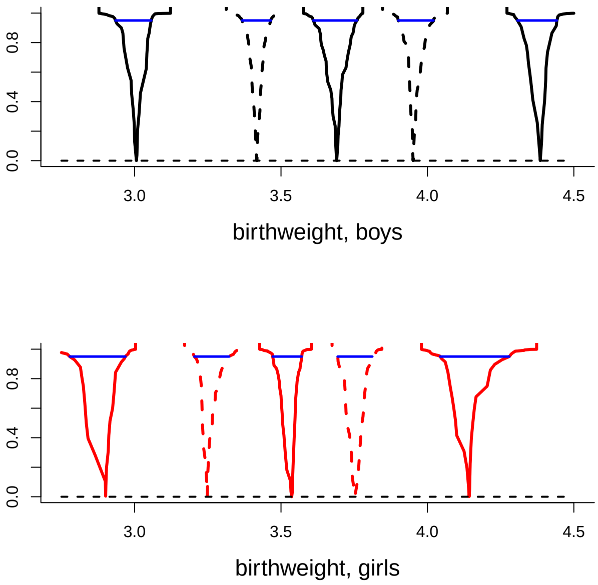

Figure viii.5: *Confidence curves for the five deciles* $F^{-1}(q)$*, for levels 0.1, 0.3, 0.5, 0.7, 0.9, for the birthweight distributions of boys (upper panel) and girls (lower panel), in kg. 95 percent confidence intervals for the 5 + 5 quantities are indicated with the blue horizontal lines.*

(e) We now wish to estimate the ratio of quantiles function $\rho = \mu_b/\mu_g$, nonparametrically, using $\widehat{\rho} = Q_b/Q_g$. Use delta method arguments to deduce that

$$\widehat{\rho} = \rho\{1 + (1/\mu_b)(\kappa_b/n_b^{1/2})Z_b - (1/\mu_g)\kappa_g/n_g^{1/2}Z_g\},$$

and from this that $\widehat{\rho} \approx_d \mathrm{N}(\rho, v^2)$, with variance

$$v^2 = \Big(\frac{\mu_b}{\mu_g}\Big)^2\Big(\frac{1}{\mu_b^2}\frac{\kappa_b^2}{n_b} + \frac{1}{\mu_g^2}\frac{\kappa_g^2}{n_g}\Big) = \frac{1}{\mu_g^2}\Big(\frac{\kappa_b^2}{n_b} + \rho^2\frac{\kappa_g^2}{n_g}\Big).$$

Construct a version of Figure viii.6. (xx conclude that across quantiles, boys tend to about 5 percent bigger than girls. Attempt to build a model for how $F_b^{-1}(q)$ for boys relates to $F_g^{-1}(q)$ for girls. xx)

(f) (xx just a bit more. xx) [xx can ask for estimates with bands of the ratio $f_1(y)/f_2(y)$, perhaps constructed by first estimating the log difference, finding band there, and exp-ing home. could also bake a Type B Story from the birthweights of Oslo boys and Oslo girls, 2001–2008, with other natural analyses. see (xx Data Story B.2.B xx). xx]

Story viii.5 *Lifelengths in Roman Era Egypt, 2100 years ago.* Intriguingly, archeologists have been able to learn the ages at death of 141 mummified individuals living in Roman

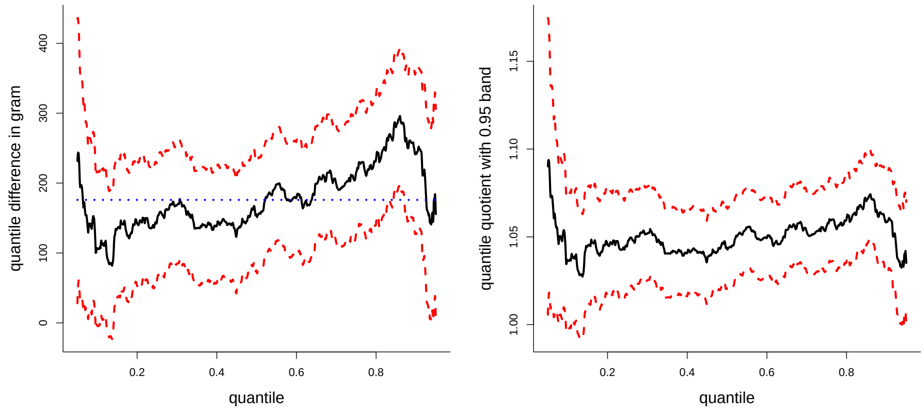

Figure viii.6: *For the 548 boys and 480 girls born at Rikshospitalet in Oslo, during 2001–2008, and for quantiles in [0.05,0.95], the plot displays the estimated difference of quantiles (left panel), and the estimated quotient of quantiles (right panel), along with a 95 percent confidence band.*

Era Egypt, some 2100 years ago, see Spiegelberg (1901). These lifelengths, varying from 1 to 96 years, for 82 men and 59 women, were discussed and analysed by Karl Pearson in the very first volume of *Biometrika*, see Pearson (1902). We treat them here as a random sample of lifelengths from the upper social class of Roman Era Egypt, during a period of relatove societal stability; more details are in Claeskens and Hjort (2008b, Ch. 2). (xx with data and details in Ch. B.2.B. when polishing, do the right pointing to Ch10 and to ML machinery of Ch5. xx)

Despite Pearson's not unreasonable comment that "in dealing with [these data] I have not ventured to separate the men from the women mortality, the numbers are far too insignificant" we shall work with parametric modelling of the men's and women's survival functions and hazard rates, and in that process illustrate the main practical uses of maximum likelihood machineries, both for model parameters and for natural parameter functions of these, and for model comparison and model selection.

(a) Go through as many as eight candidate models for these data, given below. For each model, estimate the parameters via maximum likelihood, along with estimated standard deviations for these. Here we use the general versatile machinery partly summed up in Ex. 5.12, involving programming the log-likelihood functions, finding their optima, and inverting the Fisher information matrices. Part of the learning experience here is that handling rather different parametric models does not take many extra forces, but may involve relatively small changes from script to script. (i) Use gamma distributions $\mathrm{Gam}(a_m, b_m)$ and $\mathrm{Gam}(a_w, b_w)$ for the men and the women, with parametrisation as in Ex. 1.9. Then use gamma distributions again, but take a common b for the shape parameter. (ii) Then use Weibull distributions (a_m, b_m) and (a_w, b_w), with parametrisation as in Ex. 1.40. Similarly, use a common shape parameter b for the two groups, but sep-

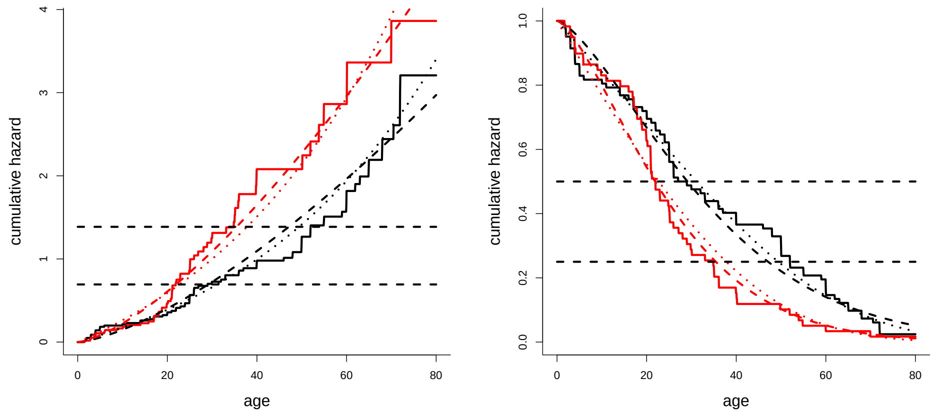

Figure viii.7: *Lifelengths in Roman Era Egypt, a century B.C., with men tending to have longer lives than women. Left panel: Nonparametric Nelson–Aalen plots for the cumulative hazards, for men (lower curves) and for women (upper curves), along with parametric fits from Models 2b (Weibull with equal 2nd parameter) and 3b (Gompertz with equal 2nd parameter) as per Ex. viii.5. Estimated 50 and 75 percent survival times can be read off from the two horizontal lines at* $\log 2$ *and* $\log 4$*. Right panel: Survival curves for the men (upper curves) and women (lower curves), nonparametric (rugged) along with parametric fits. Estimated 0.50 and 0.75 percent survival times are read off from where the curves cross the 0.50 and 0.25 lines.*

arate a_m and a_w. (iii) Then use Gompertz distributions with parameters (a_m, b_m) and (a_w, b_w), with parametrisation as in Ex. 1.41. Again allow the variation taking a common shape parameter b but different a_m, a_w for the two Gompertz distributions. (iv) Throw in also the lognormal distributions, first with four free parameters (ξ_m, σ_m), (ξ_w, σ_w), then with a common σ for the lognormals. For each of these 4 + 4 models, graph the estimated cumulative hazard functions $A(t, \widehat{\theta})$ for men and women, plotted alongside the nonparametric Nelson–Aalen curves, and also the estimated survival curves $S(t, \widehat{\theta})$ for men and women, alongside the nonparametric Kaplan–Meier curves. In other words, construct versions of Figure viii.7, left and right panels.

(b) After having fitted all the candidate models, and computed the log-likelihood maxima in question, it is a small extra step to count parameters and compute the AIC scores. Do this, organising your results into a table with the three first columns here, with ‘dim’ denoting the number of parameters in the model. Conclude that model 3B is the best (so far), the Gompertz model with parameters (a_m, b) and (a_w, b), as judged by the AIC. Incidentally, show that the log-normal models are decidely worse. We include them here for the sake of exercising the general maximum likelihood machinery, and since we could not have known a priori which models are good and which are not.

```
         dim logLmax   aic      men    women  delta sd     low    up
model 1A 4 -612.064 -1232.129 26.655 21.877 4.778 3.371 -0.766 10.323 gamma
```

```
model 1B 3 -614.922 -1235.844 26.055 22.725 3.330 3.439 -2.327  8.986
model 2A 4 -609.954 -1227.909 28.262 22.728 5.534 3.502 -0.226 11.294 weib
model 2B 3 -610.387 -1226.774 29.120 21.910 7.209 2.997  2.279 12.139
model 3A 4 -608.388 -1224.776 31.783 22.140 9.644 4.246  2.659 16.629 gomp
model 3B 3 -608.520 -1223.040 31.124 22.723 8.401 3.447  2.730 14.072
model 4A 4 -627.397 -1262.794 23.237 19.958 3.279 3.447 -2.391  8.949 logN
model 4B 3 -629.511 -1265.023 23.237 19.958 3.279 3.521 -2.514  9.071
```

(c) For the three best of the fitted models, compute and graph the estimated densities $\widehat{f}(t)$, survival curves $\widehat{S}(t)$, and cumulative hazard rates $\widehat{A}(t)$. Complement these with the nonparametric Nelson–Aalen estimators. Present also the estimated hazard rates $\widehat{\alpha}(t)$. Construct a version of Figure viii.7. Explain how estimated median survival time in Roman era Egypt can be read off from the horizontal $\log 2$ line, and similarly the estimated 75 percent quantile survival time via the $\log 4$ line.

(d) It appears clear that in Roman era Egypt, men tended to have longer lives than women. The direct nonparametric median lifetime estimates are 28 for men and 22 for women. For each of the eight candidate models, compute the implied median-life difference estimate, i.e. of $\delta = F_m^{-1}(0.50) - F_w^{-1}(0.50)$. Also use the maximum likelihood theory summarised in Ex. 5.12, specifically the use of the delta method for any smooth function of the model parameters, to compute the approximate standard deviation for these $\widehat{\delta}$ estimates, and give 90 percent confidence intervals. These estimates, with lower and upper confidence points, are given in the table above. Your code should be flexible enough to carry out similar analyses for e.g. the upper quartile difference $F_m^{-1}(0.75) - F_w^{-1}(0.75)$, a parameter of high interest for the five million Egyptians two thousand years ago. Attempt to pinpoint where the men and women of Roman Era Egypt started having different lifelength expectancies.

(e) (xx find an easy reference to the fact that a high proportion of women died in childbirth, in many socities. xx) The models worked through above are generic in character and do not take on board why or in which ways the lives of men and women might have been dfferent in old Egypt. The flexibility and versality of the maximum likelihood machinery should inspire building other models. Consider random lifetimes

$$T_m = \min\{t \ge 0\colon Z_m(t) \ge c\}, \quad T_w = \min\{t \ge 0\colon Z_w(t) \ge c\},$$

defined via cumulative risk processes Z_m and Z_w for men and women; when these cross threshold c, the individual dies. A natural class of such processes, amenable to further survival analysis for their threshold crossing times, is that of independent increment gamma processes, see Cunen and Hjort (2023). For the present purposes we take $Z_m(t)$ having mean function at whereas $Z_w(t)$ has mean function $at + d\,\mathrm{ex}(t)$, with an extra risk function $\mathrm{ex}(t)$ here taken to be the c.d.f. of a uniform distribution on $[15, 40]$. Show that this leads to survival functions $S_m(t) = G(c, at, 1)$ and $S_w(t) = G(c, at + d\,\mathrm{ex}(t), 1)$ for men and women, where $G(\cdot, u, 1)$ is the c.d.f. for $\mathrm{Gam}(u, 1)$. Show that the log-likelihood function becomes

$$\ell(a, c, d) = \sum_{i=1}^{n_m} \log f_m(t_{m,i}) + \sum_{i=1}^{n_w} \log f_w(t_{w,i}),$$

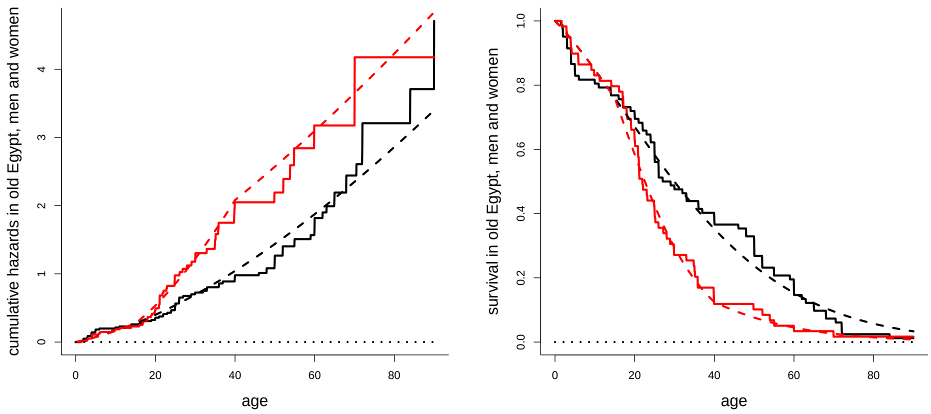

Figure viii.8: *(xx text to be coordinated and polished. xx) Left panel: Nelson–Aalen cumulative hazards for men and women of Roman Era Egypt, along with those fitted to the gamma process threshold crossing model; these are better than the 4 + 4 models worked with initially. Right panel: Associated survival curves, nonparametric and parametric.*

with the n_m and n_w lifelengths for men and women, and with densities $f_m(t) = -S'_m(t)$ and $f_w(t) = -S'_w(t)$ implied by the survival functions.

(f) Now programme and optimise the log-likelihood. You should find $(\widehat{a}, \widehat{c}, \widehat{d}) = (0.033, 0.687, 0.810)$, and with a much higher log-likelihood maximum -604.368 than for the eight models worked with above. Show also that this leads to an AIC score very clearly better than for the competitors. Display nonparametric and parametrically fitted cumulative hazard rates and survival curvess, as in Figure viii.8, left and right panels. The gamma process models provide much better fits than for models portrayed in Figure viii.7.

(g) To illustrate how the gamma process models work, simulate e.g. 25 Z_m processes, with mean function at, and Z_w processes, with mean function $at + d\,\mathrm{ex}(t)$, using the estimated $(\widehat{a}, \widehat{c}, \widehat{d})$. Death occurs when the process reaches c. More men than women survive the age of forty. Construct a version of Figure viii.9 (male processes in left panel, female in right panel). (xx put in somewhere: with our gamma process model, women and men have the same longer-time survival chances after the age of forty. xx)

Story viii.6 *Stride towards your bookshelves.* As part of the obligatory exercises work for a bachelor level course on statistical methodology at the Department of Mathematics, University of Oslo, we instructed each student to stride towards her or his bookshelves, to pick one book in Norwegian and one in English, then record the lengths of the first 100 words on page 51. The books could be novels, collections of short stories, poetry, or prose in general, but not technical material (as with mathematics or statistics); the students were also instructed to use page 52 if page 51 didn't have enough words. Do

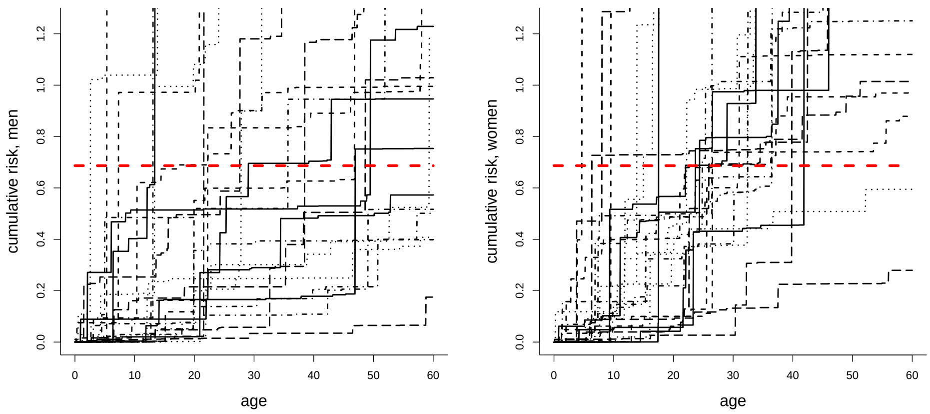

Figure viii.9: *(xx text to be coordinated and polished. xx) 25 simulated gamma processes, for mean functions at for the men (left panel) and* $at+d\,\mathrm{ex}(t)$ *for the women (right panel); an individual dies when his or her process crosses the threshold* $c = 0.687$*, the horizontal line.*

Fløgstad, Kjærstad, Solstad tend to use words with more or less the same lengths as do Miller, Lessing, Munro? And do some students have books tending to have longer words than those of other students?

The students were asked to summarise information and to compare their own two datasets in terms of means and standard deviations. This was expected to involve tests for equality of means and of variances, confidence intervals for differences, perhaps comments on skewnesses, etc. But the experiment also gave us an interesting combined data set, where we recorded the empirical mean and standard deviation for each dataset, for the two languages, for each student. In other words, we have summary statistics data $(x_{i,N}, \widehat\kappa_{i,N}, x_{i,E}, \widehat\kappa_{i,E})$ for $i = 1, \ldots, n$, for the $n = 64$ students, with

$$x_{i,N} = \text{average word-length for 100 Norwegian words for student } i,$$
$$x_{i,E} = \text{average word-length for 100 English words for student } i,$$

along with empirical standard deviations $\widehat\kappa_{i,1}$ and $\widehat\kappa_{i,2}$, say, for these 100 Norwegian and 100 English words, for student i.

(a) Construct a version of Figure viii.10, one panel with $(x_{i,N}, x_{i,E})$, a second panel with $(\widehat\kappa_{i,N}, \widehat\kappa_{i,E})$. Why and in which sense was it ok for Hjort and Stoltenberg to throw away the individual data samples, with $nm = 64 \cdot 100$ words in each of the two languages, and just keep the empirical means and standard deviations?

(b) Carry out a test to see if the mean word lengths are about the same, for the Norwegian and English books (in these students' bookshelves). For this point, suppose that $X_{i,N} \sim \mathrm{N}(\xi_{i,N}, \kappa^2_{i,N}/100)$ and $X_{i,E} \sim \mathrm{N}(\xi_{i,E}, \kappa^2_{i,E}/100)$. Then perform a second test, to see if the underlying spread in wordlength distributions are the same for the two languages. [xx polish a bit. answers are no for $\bar x$, but yes for the $\widehat\kappa$. xx]

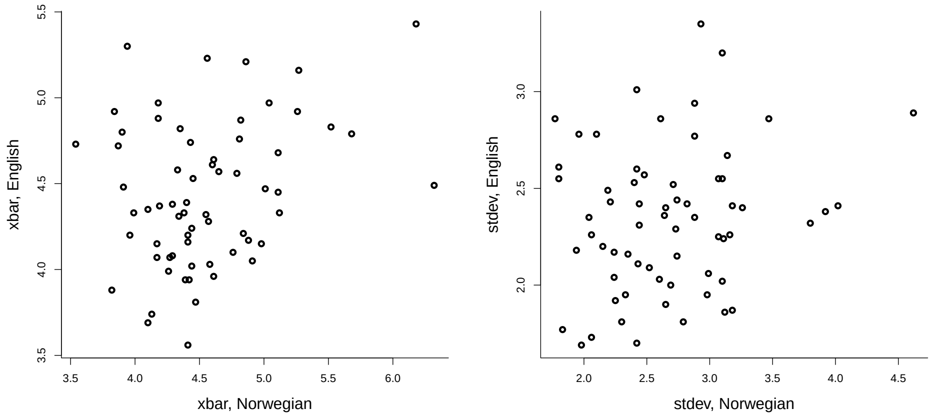

Figure viii.10: *Empirical means* $(x_{i,N}, x_{i,E})$ *(left panel), and empirical standard deviations* $(\widehat{\kappa}_{i,N}, \widehat{\kappa}_{i,E})$ *(right panel), for the Norwegian and English wordlengths found in 64 students' bookshelves, with each student sampling 100 words from their sampled books.*

(c) We then take an in interest in *the correlation* between the two wordlength distributions. But taking the ordinary correlation between the reported averages $x_{i,N}, x_{i,E}$ is less interesting than inference for *the real correlation*, between say $x_{i,\mathrm{real},N}, x_{i,\mathrm{real},E}$, these being the averages over the tens of thousands of Norwegian and English words on the bookshelves of student i. It turns out the ordinary correlation deflates this underlying real correlation, due to the measurement errors involved in sampling merely 100 words for the two corpora.

In general terms, suppose we have observations (x_i, y_i) for $i = 1, \ldots, n$, where these are really proxies for certain underlying $(x_{i,0}, y_{i,0})$, and where the measurement errors involved are normal with known variance levels. We have in mind situations where the correlation $\rho = \mathrm{corr}(x_0, y_0)$ between these underlying quantities is of higher concern than the deflated correlation $\mathrm{corr}(x, y)$ between the directly observed (x_i, y_i). We formalise a version of the setup described as

$$x_i = x_{i,0} + \delta_{i,1}, \quad y_i = y_{i,0} + \delta_{i,2},$$

where the not fully observed $(x_{i,0}, y_{i,0})$ have a binormal distribution, with correlation ρ, and where the measurement errors $\delta_{i,1}$ and $\delta_{i,2}$ are independent zero-mean normals with known or well estimated standard deviations τ_1 and τ_2. With (ξ_1, ξ_2) the means and (σ_1, σ_2) the standard deviations for $(x_{i,0}, y_{i,0})$, show that

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim \mathrm{N}_2(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, V), \quad \text{with} \quad V = \Sigma + D = \begin{pmatrix} \sigma_1^2 + \tau_1^2, & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2, & \sigma_2^2 + \tau_2^2 \end{pmatrix},$$

writing D for $\mathrm{diag}(\tau_1^2, \tau_2^2)$.

(d) First we sort out what happens with the traditional empirical correlation coefficient for the observed data, say $R_n = s_{1,2}/(s_1 s_2)$, where s_1 and s_2 are the empirical standard deviations for the x_i and the y_i, and $s_{1,2} = (n-1)^{-1}\sum_{i=1}^n (x_i-\bar x)(y_i-\bar y)/(s_1 s_2)$. Show that

$$R_n \to_{\rm pr} \frac{\rho\sigma_1\sigma_2}{(\sigma_1^2+\tau_1^2)^{1/2}(\sigma_2^2+\tau_2^2)^{1/2}},$$

i.e. the default operation R_n actually estimates a deflated version of the real ρ.

(e) Consider first repair operation 1, which is to estimate the σ_j by $\widehat\sigma_j^2 = \max(s_j^2-\tau_j^2, 0)$ for $j=1,2$. Show that $\widehat\rho = s_{1,2}/(\widehat\sigma_1\widehat\sigma_2)$ is consistent for ρ. Note however that its limit distribution is more complicated than for the classical case of no measurement error; see Ex. 5.13. (xx then spell out what this means for the bookshelves story. the point is to see τ_1 and τ_2 from the data, as precisely estimated. For the $x_{i,N}$, with individual variances $\kappa_{i,N}^2/m$, argue that $\tau_N = \{(1/n)\sum_{i=1}^n \widehat\kappa_{i,N}^2/m\}^{1/2} = 0.2728$ and $\tau_E = \{(1/n)\sum_{i=1}^n \widehat\kappa_{i,E}^2/m\}^{1/2} = 0.2370$ are precise estimates of the measurement errors here. From the directly observed standard deviations $s_N = 0.5285$ and $s_E = 0.4193$ show that these are reduced to $\widehat\sigma_N = (s_N^2-\tau_N^2)^{1/2} = 0.4526$ and $\widehat\sigma_E = (s_E^2-\tau_E^2)^{1/2} = 0.3459$. This adjusts the deflated $R_n = 0.2833$ to $\widehat\rho = 0.4010$. xx)

(f) Then consider repair operation 2, using likelihood methods. Show that the log-likelihood function for the observed data becomes

$$\ell_n = \sum_{i=1}^n \Big\{-\tfrac12\log|\Sigma+D| - \tfrac12\begin{pmatrix}x_i-\xi_1\\ y_i-\xi_2\end{pmatrix}^{\rm t}(\Sigma+D)^{-1}\begin{pmatrix}x_i-\xi_1\\ y_i-\xi_2\end{pmatrix}\Big\}.$$

Show that profiling over the means leads to $\ell_{n,\rm prof}(\sigma_1,\sigma_2,\rho) = -\tfrac12 nQ(\sigma_1,\sigma_2,\rho)$, where

$$Q(\sigma_1,\sigma_2,\rho) = \log|\Sigma+D| + {\rm Tr}\{(\Sigma+D)^{-1}S_n\},$$

in terms of the empirical variance matrix S_n for the (x_i, y_i) pairs. (xx do this for the bookshelves data. find and display a cc(ρ). point estimate 0.401, 95 percent interval $[0.064, 0.670]$. xx) (xx variations: could actually have different $\tau_{i,1}, \tau_{i,2}$ for the log-likelihood. xx) (xx careful with wording: We learn that a student having Norwegian books with long words tends to have English books with long words too, and vice versa. The reasons for this interesting finding are not clear, but it's interesting to do a bit of speculation – some readers prefer longer-worded books, others might like shorter-worded literature. We're also reminded that the students were not instructed to choose books from their bookshelves in a totally random fashion, so there's a limit to how far we should stretch our imagination here. xx)

(g) As is already apparent from the correlation analysis, the wordlengths exhibit not merely the obvious variation inside bookshelves, but also between students. Construct a version of Figure viii.11, left panel, displaying English wordlength averages $x_{i,E}$ along with their associated individual 90 percent intervals. To assess the degree of disparity between students, i.e. betweeen their bookshelves, model the $x_{i,E}$ as coming from a $N(\xi_E, \omega_E^2)$ distribution. Show that marginally, $x_{i,E} \sim N(\xi_E, \sigma_{i,E}^2 + \omega_E^2)$, with $\sigma_{i,E}^2 =$

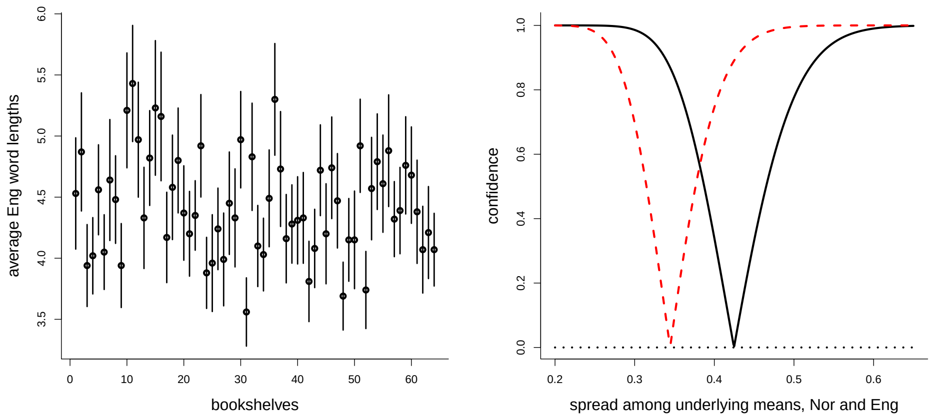

Figure viii.11: *Left panel: For the $n = 64$ students, average word lengths in their English books, with 90 percent confidence intervals. Right panel: Confidence curves for the spread parameters ω_N and ω_E, for the models where the averages $x_{i,N}$ and $x_{i,E}$ follow distributions $\mathrm{N}(\xi_N, \omega_N^2)$ and $\mathrm{N}(\xi_E, \omega_E^2)$.*

$\kappa_{i,E}^2/m$. Since these are well estimated, we take them as nearly known, set equal to $\widehat{\kappa}_{i,E}^2/m$. (xx then calibrate with what is in Ch7. xx) Using

$$Q_E(\omega_E) = \sum_{i=1}^n \frac{\{x_{i,E} - \widehat{\xi}_E(\omega_E)\}^2}{\sigma_{i,E}^2 + \omega_E^2} \sim \chi_{n-1}^2, \quad \text{with } \widehat{\xi}_E(\omega_E) = \frac{\sum_{i=1}^n x_{i,E}/(\sigma_{i,E}^2 + \omega_E^2)}{\sum_{i=1}^n 1/(\sigma_{i,E}^2 + \omega_E^2)},$$

construct the CD $C_E(\omega_E) = 1 - \Gamma_{n-1}(Q_E(\omega_E))$ and its associated confidence curve. Compute $\mathrm{cc}(\omega_E)$ and $\mathrm{cc}(\omega_N)$ and display these in a diagram, as with Figure viii.11. Find median confidence estimates and also 95 percent intervals for the spread parameters ω_E and ω_N, and comment on their sizes.

Story viii.7 *The children of Odin.* As we know, Odin had six male offspring – Thor, Balder, Vitharr, Váli, Heimdallr, Bragi – with the sources saying nothing about daughters. So how many children is it likely that he had, in total? With N the number of children, and y the number of boys, we assume $y \,|\, N \sim \mathrm{binom}(N, p)$, with $p = 0.514$ (a good overall point estimate human reproduction; see Story viii.3). So the data is that $y = 6$, and we can attempt confidence inference for N. The themes and details below expand on those given in Schweder and Hjort (2016, Example 3.11).

(a) A natural construction for a CD is

$$C(N, y) = P_N(Y > y) + \tfrac{1}{2} P_N(Y = y),$$

with the half-correction for discreteness, as in the partly parallel situation of Ex. 7.25. Compute and display this CD, and take differences to compute also the confidence point masses, $c(N, y)$.

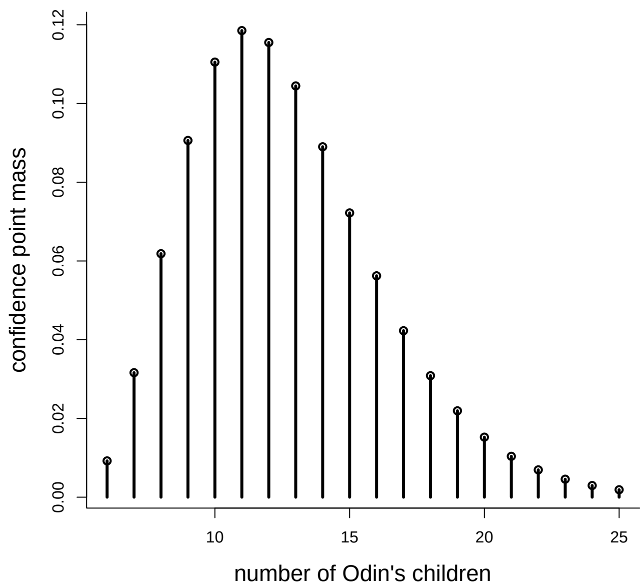

Figure viii.12: *Confidence point masses $c(N, y)$, for $N \geq 6$. (xx more here, in a separate point, checking $U = C(N, Y)$ and approximate uniformity, with $Y \sim \mathrm{binom}(N, p)$. xx)*

(b) A CD $C(\theta, y)$, for a parameter θ based on data y, should ideally have the uniformity property that $U = C(\theta_0, Y)$ has the uniform distribution, for any fixed θ_0, with Y a random dataset drawn from the model at that position in the parameter space. This is not quite possible here, since the situation is discrete, with not many values to attain for y. For a given $N_0 = 14$, simulate say 10^4 realisations of $U = C(N_0, Y)$, then compute and display the empirical distribution function $P(U \leq u)$. Comment on your findings.

(c) Carry out also a Bayesian analysis, using the prior proportional to $1/(N + 1)$ for $N \geq 0$. Compare the posterior distribution to the CD.

(d) Invent your own prior for N (formed before you learn in school that $y = 6$), and compare the posterior distribution withat that found above.

(e) The frequentist CD $C(N, y)$ above should be trusted as a good and neutral statistical summary function for the unknown N. Find and display the Bayesian prior that would give the same result.

(f) In some of the Snorri kennings there are also references to Týr and Höd as sons of Odin (and yet other names are mentioned in the somewhat apocryphical Skáldskaparmál). Adjust the calculations above to this revised case, with $y = 8$, and comment on your findings.

(g) Find or dream up another situation (not necessarily with full data) where the model above might be used, i.e. p is known, but the binomial N is unknown.

Story viii.8 *How many Abel envelopes from 1902?* Hundred years after the death of Niels Henik Abel (1802–1829), the Norwegian postal office issued a certain stamp and a 'first-day cover' envelope commemorating him; this was only the second time such an honour had been bestowed upon a person outside royalty (in 1928, a similar first-day cover had been issued for Henrik Ibsen, hundred years after his being born). As the facsimile below indicates, these carry 'R numbers' (as in 'rekommandert post'), and R numbers from five such Abel 1929 envelopes, from various philatelic sales lists and auctions in the 2003–2008 period, were 280, 304, 308, 310, 328. The operating assumption is that the Abelian first-day covers with stamps were produced in a running noninterrupted sequence, but one does not know when it started and neither when it ended. So how many were there? An answer to this curiosity question also enters the realm of philatelic market prices and speculations (one such specimen might fetch 5000 kroner, in 2022).

Figure viii.13: *A philatelic rarity: an Abel first-day envelope from 1929.*

(a) We allow ourselves a statistical detour, discussing natural setups and solutions for range estimation for the case of continuous data, before returning to Abel. So, for a concrete illustration, consider the numbers

```
4.712 6.412 7.043 7.141 7.245 7.379 7.602 8.417 8.671 8.702
```

We've simulated these $n = 10$ points, from a uniform distribution over $[a, b]$, and ordered them, for simplicity. But we won't tell you the values we used for a or b, or indeed the range $\gamma = b - a$. Your task will be to make inference about this γ. We come back to Bayesian solutions below, but now approach the problem using frequentist confidence distributions. With $Y_1, \ldots, Y_n$ from the uniform on $[a, b]$, explain that one may write $Y_i = a + (b - a)U_i$, with the U_i from the standard uniform over the unit interval. Deduce that $R_n = Y_{(n)} - Y_{(1)} = \gamma R_{n,0}$, with $R_{n,0} = U_{(n)} - U_{(1)}$, relating the range of data naturally to the range of a uniform sample. Explain that R_n/γ is a pivot, as defined in Ex. 7.7.

(b) With H_n the c.d.f. of the uniform range $R_{n,0}$ distribution, show that the canonical confidence distribution for γ becomes $C_n(\gamma, \text{data}) = P_{a,b}(R_n \geq R_{n,\text{obs}}) = 1 - H_n(R_{n,\text{obs}}/\gamma)$, for $\gamma \geq R_{n,\text{obs}}$ (here observed to be 3.990). Simulate say 10^4 realisations of $R_{n,0}$ in your computer, and use these to compute and display the CD $C_n(\gamma, \text{data})$, as well as the confidence curve $\text{cc}_n(\gamma, \text{data}) = |1 - 2\,C_n(\gamma, \text{data})|$. Use also the explicit knowledge from Ex. 2.19, that H_n actually is a $\text{Be}(n-1, 2)$, to show that the confidence distribution and its confidence density become

$$\begin{aligned} C_n(\gamma, \text{data}) &= 1 - n(R_{n,\text{obs}}/\gamma)^{n-1} + (n-1)(R_{n,\text{obs}}/\gamma)^n, \\ c_n(\gamma, \text{data}) &= n(n-1)R_{n,\text{obs}}^{n-1}(\gamma - R_{n,\text{obs}})/\gamma^{n+1}, \end{aligned}$$

for $\gamma \geq R_{n,\text{obs}}$. Compute the median confidence and maximum confidence estimates.

(c) We now approach the inference problem with Bayesian means. Starting with the likelihood function, show that it can be written as follows, expressed as a function of (a, γ) rather than of (a, b):

$$L_n(a, \gamma) = (1/\gamma)^n\, I(a \leq y_{(1)} \text{ and } y_{(n)} \leq a + \gamma),$$

in particular taking the value zero if $a > y_{(1)}$ or $y_{(n)} > a + \gamma$. Find the ML estimates for a and for γ. Then, with a flat prior on a, independently of a prior $p(\gamma)$ for γ, show that the posterior distribution of γ is

$$p(\gamma \,|\, \text{data}) \propto p(\gamma)(\gamma - R_{n,\text{obs}})(1/\gamma)^n \quad \text{for } \gamma \geq R_{n,\text{obs}}.$$

(d) Without clear prior knowledge concerning the range a natural prior is proportional to $1/\gamma$. Show that this leads to the Bayesian posterior distribution agreeing precisely with the frequentist CD above. In particular, explain that the Bayes machine, starting from the $1/\gamma$ prior, leads to credibility intervals with perfect frequentist coverage. The 95 percent interval becomes $[4.094, 7.189]$, for example. As an alternative, consider also using a flat prior for γ, and show that this leads to a posterior with density and cumulative equal to (xx check all details here xx)

$$\begin{aligned} g_n(\gamma \,|\, \text{data}) &= (n-1)(n-2)R_{n,\text{obs}}^{n-2}(\gamma - R_{n,\text{obs}})/\gamma^n, \\ G_n(\gamma \,|\, \text{data}) &= 1 - (n-1)(R_{n,\text{obs}}/\gamma)^{n-2} + (n-2)(R_{n,\text{obs}}/\gamma)^{n-1}, \end{aligned}$$

for $\gamma \geq R_{n,\text{obs}}$. Plot both posteriors (with one of these equal to the CD) for the dataset above.

(e) We now return to the Abel numbers 280, 304, 308, 310, 328, first with a natural CD approach. Take these to be a random sample $X_1, \ldots, X_n$ (without replacement) of size $n = 5$ from $\{a+1, \ldots, a+N\}$, with both a and N unknown. It is natural to base the inference on the range $R_n = V_n - U_n$, where $U_n = \min_{i \leq n} X_i$ and $V_n = \max_{i \leq n} X_i$. Show that its distribution is independent of a. Argue that this leads to the confidence distribution $C(N) = P_N(R_n > 48) + \frac{1}{2}\,P_N(R_n = 48)$; as usual, P_N signals probability calculations under the value N of this parameter.

(f) It remains to find expressions for the distribution of R_n. Consider first the joint distribution of (U_n^0, V_n^0), where U_n^0 and V_n^0 are as U_n and V_n, but in the situation where $a = 0$. Show that their joint probability distribution can be expressed as

$$f(u,v) = \binom{v-u-1}{n-2}\Big/\binom{N}{n} \quad \text{for } 1 \le u,\ u+n-1 \le v \le N.$$

Deduce that the distribution of $Z_n = V_n - U_n$ can be written

$$P_N(Z_n = z) = \sum_{v-u=z} f(u,v) = (N-z)\binom{z-1}{n-2}\Big/\binom{N}{n}$$

for $z = n-1, \ldots, N-1$. Compute and display the CD and the confidence curve $\mathrm{cc}(N)$ for N.

(g) We then work towards a Bayesian solution, based on data $X_1, \ldots, X_n$ as above, a random draw from $\{a+1, \ldots, a+N\}$. With independent priors $p_0(a)$ and $p(N)$ for the start-point a and sequence length N, show that

$$p(a, N \,|\, \text{data}) \propto p_0(a)p(N)\, I(a+1 \le U_n < V_n \le a+N)/\binom{N}{n}.$$

With a flat prior on the starting point a, show that this under some conditions leads to

$$p(N \,|\, \text{data}) \propto p(N)\frac{(N-R_n)}{N(N-1)\cdots(N-n+1)}\,;$$

note the partial similarity to the posterior $p(\gamma \,|\, \text{data})$ in the continuous case above. Work out the posterior distribution for N, over a suitable range of N values, starting with a flat prior. Find posterior median and a 95 percent interval. – One ought to be careful here, since the prior for a should be flat on $1, 2, 3, \ldots$, not including negative numbers. Show that the associated refinement of the direct result above becomes $p(N \,|\, \text{data}) \propto p(N)q(N)/\{N(N-1)\cdots(N-n+1)\}$, where $q(N)$ counts the number of $a \ge 1$ satisfying $V_n - N \le a \le U_n - 1$. Show that this means either $U_n - 1 - (V_n - N - 1) = N - R_n$, provided $V_n \ge N+1$, or $U_n - 1$, in the case of $V_n \le N$. In other words and symbols, $q(N) = (N - R_n)\, I(V_n > N) + (U_n - 1)\, I(V_n \le N)$. Show however that for the present occasion, the relevant values of N are smaller than $V_n = 328$, so this additional layer of care turns out not to be needed.

(h) In addition to the five R numbers 280, 304, 308, 310, 328 known as of 2008, five more such first-day Abel envelopes have been unearthed, the latest in 2022: 285, 314, 317, 327, 334. Update your inference, for the CD and the Bayesian posterior, and construct versions of Figure viii.14; CDs in the left panel and Bayesian cumulatives in the right. Compute also the median confidence and median Bayes estimates, along with 95 percent intervals. (Answers: for data up to 2008, point estimates are 69 and 75, with intervals $[51, 164]$ and $[51, 170]$, for the CD and the Bayes. With extended data up to 2022, point estimates are 64 and 64, with intervals $[55, 94]$ and $[55, 100]$.)

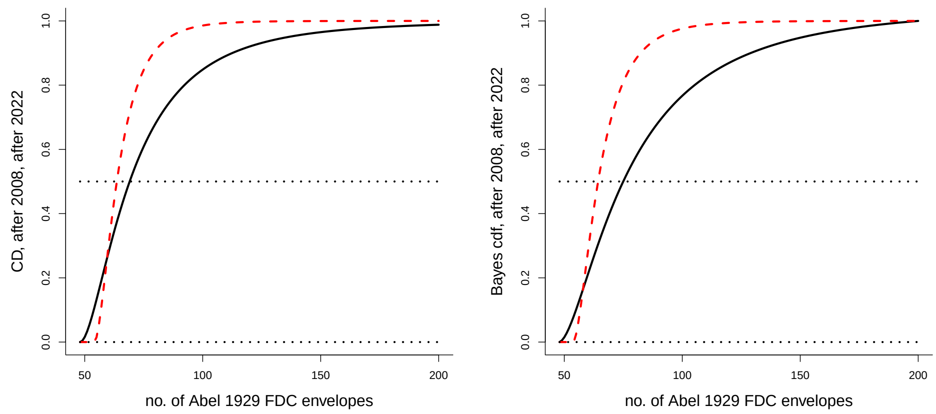

Figure viii.14: *Left panel: Confidence distribution for N, the number of Abel 1929 first-day cover envelopes, based on the 5 known numbers by 2008 (full curve), and on the now 10 known numbers by 2022 (dashed curve). Right panel: for the same information, the Bayesian posterior c.d.f.*

Story viii.9 *Bach, Reger, organ fugues, and Wohltemperierte I und II.* A fugue, whether for a piano, an organ, a choir, or an ensemble of instruments, starts with the principal fugue theme itself, before it is imitated and varied, perhaps in complex ways, in other voices; typical Bach fugues have from three to five voices. Rydén (2020) has studied such fugue themes from the organ works of Bach and other composers. He has accurately defined certain features, for quantitive analysis and comparisons. These can be identified and counted for each given fugue theme. In brief, these are

x_1, the length, number of notes, range 7 to 64;

x_2, the compass, range (in semitones), range 5 to 20;

x_3, the number of unique notes, range 4 to 12;

x_4, the initial interval (in semitones), range 0 to 12;

x_5, the number of unique intervals between successive notes, range 2 to 11;

x_6, the max interval (in semitones), range 2 to 12.

Further aspects of the data are briefly described in (xx data overview 2.B xx). (xx could mention Prout, 1891, Tovey, 1924. xx)

The musical here is to statistically describe and compare the fugues of J.S. Bach (1685–1750) and Max Reger (1873–1916). Figure viii.15 shows (x_1, x_6) for the $n_B = 47$ Bach fugues and $n_R = 45$ Reger fugues, indicating also that the distributions are not very different. (xx mention the Händel concerto gross, is it no. 7, with only a single note for the fugue theme, so $x_3 = 1$; for Bach and Reger the range is from 4 to 12, though. xx)

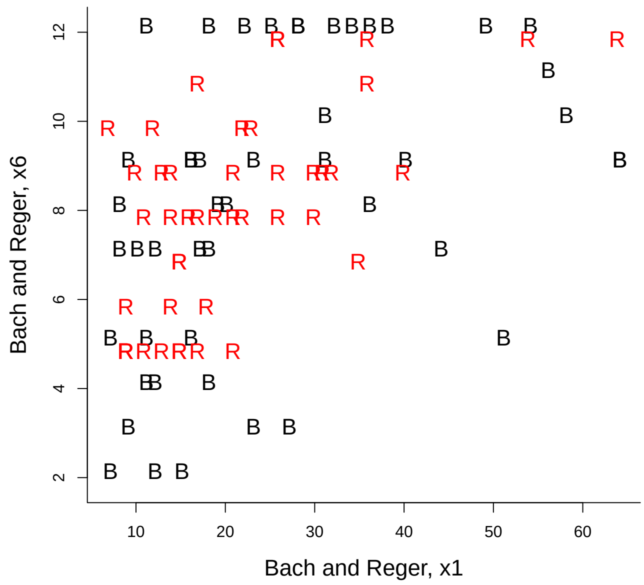

Figure viii.15: For the chief organ fugue themes of Bach (B, 47 fugues) and Reger (R, 45 fugues), the plot gives features x_1, the length, and x_6, the max interval.

(a) For an initial check of the data, take the $n_B + n_R = 92$ fugues together. Go through each of the fugue features $x_1, \ldots, x_6$, and give brief statistical descriptions. Identify also pairs of features with strong correlation, if any. Construct a version of Figure viii.15, which has (x_1, x_6) for Bach and Reger; construct a similar one for (x_1, x_3).

(b) For each of the fugue features $x_1, \ldots, x_6$, compute means and standard deviations, for the 47 Bach fugues and 45 Reger fugues. Then for each feature, test equality of means, say $\xi_{B,j} = \xi_{R,j}$, using t testing; see Ex. 3.4. Comment both on the use of t testing for these data and on your findings. (You should find that for feature x_3, Reger has higher mean than Bach, whereas they are more or less equal, for the other five features.)

	mean B	mean R	sd B	sd R	kurt B	kurt R
x1	25.766	21.111	16.240	11.924	-0.336	2.571
x2	11.043	11.911	3.520	2.636	-0.642	-0.821
x3	6.872	8.556	1.541	1.791	0.729	-1.055
x4	2.617	2.222	2.524	1.894	2.323	13.201
x5	6.000	5.978	2.467	2.072	-0.579	-0.901
x6	8.021	8.089	3.267	2.275	-1.114	-1.053

(c) Then go on to testing equality of standard deviations, say $\sigma_{B,j} = \sigma_{R,j}$. Do this first by applying a traditional F test, as from Ex. 3.37, even though the data are not normal. This should give an indication that Bach intriguingly exhibits greater variability than Reger, for features x_1, x_2, x_4, x_6, with the means being about the same. Also carry out the somewhat more elaborate testing regime, for equality of standard deviations, from Ex. 3.37, which does not rely on normal data. Does this change the previous tentative findings?

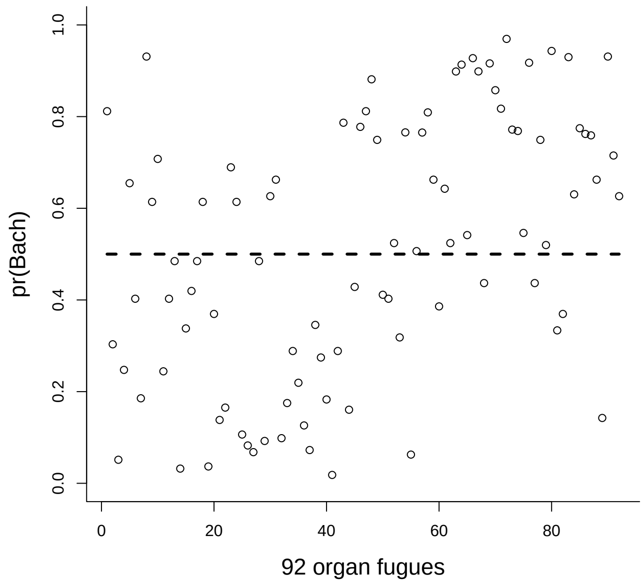

Figure viii.16: Using logistic regression on the basis of fugue features x_1, x_3, the figure shows the estimated $P(\text{Bach}\,|\,x_1, x_3)$, for the 92 fugues, listed with the 45 Reger ones first, the 47 Bach ones afterwards.

(d) The fugue features $x_1, \ldots, x_6$ devised by Rydén (2020) are meant as useful musicological descriptors, but as they concern merely the fugue theme itself, not the further compositional development, they cannot be expected to and do not pretend to discriminate between e.g. Bach and Reger to any high degree. Even amateur musicians are able to see or hear the difference between a Bach page and a Reger page, by looking through or playing the music, though it would be hard to translate such knowledge into algorithms. Leaving these musical considerations aside, we look here into the degree of discrimination afforded by the fugue theme features. This can clearly be done in several ways, but here we attempt to build a formula for the probability that the piece is by Bach, via logistic regression,

$$P(\text{Bach}\,|\,x_1, \ldots, x_6) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_6 x_6)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_6 x_6)}.$$

Carry out such an analysis, and check how well it works, when we tentatively sort fugues into Bach, if the probability is at least 0.50, and Reger, if the probability is less than

0.50. In this analysis, which features x_j are significantly present, according to the logistic regression? (xx nils will check whether this is an ok illustration or not: given the hope of expressing Pr(Bach) in this way, what's the uncertainty, how wide are confidence intervals? but we should have another illustration of this somewhere. xx)

(e) Search through submodels, where some but not all of the six features are being used, and check their AIC scores. You should find that including x_1, x_3, but excluding the other four, gives the best AIC score. Construct a version of Figure viii.16), which uses logistic regression for x_1, x_3. What is the appararent success rate, for the ensuing algorithm, which sorts fugues into Bach and Reger?

(f) The direct counting of how many of the 92 fugues are correctly sorted into Bach and Reger suffers from a certain bias (xx point to things in Ch. 11 xx), since the data are used both to construct a formula and to test that formula. To form a clearer picture, carry out leave-one-out cross validation (xx pointer xx), and estimate the success rate.

(g) Rydén (2020) concentrated on the organ fugues of Bach, Reger, and others, as discussed above. We recommend playing through also the 24 fugues of Wohltemperiertes Klavier I (from the Köthen period, c. 1722) and the 24 fugues of Wohltemperiertes Klavier II (from Leipzig, c. 1742). Has Bach stayed about the same, as a fugue theme composer, for the clavier? (xx might point to Hindemidt 1950. xx) One of us has actually played all $24 + 24$ fugues and carefully recorded a table of $x_1, \ldots, x_6$; see 2.B. Use this to construct a version of the table below, of means, standard deviations, skewness, kurtoses, for the six characteristics, for WTK I and WTK II:

	xi		sigma		skewness		kurtosis	
	I	II	I	II	I	II	I	II
x1	18.042	21.333	7.932	9.342	0.653	0.429	-0.098	-0.241
x2	11.083	11.000	2.749	2.874	-0.264	-0.158	-0.947	-0.767
x3	7.750	7.333	2.069	1.685	0.547	0.120	-0.466	-0.615
x4	2.667	2.750	1.903	1.800	1.149	0.745	-0.004	0.226
x5	5.083	5.500	1.767	1.445	0.921	-0.373	0.350	-0.833
x6	8.167	7.750	2.220	2.625	-0.201	0.316	0.107	-1.023

(h) To assess the grand hypothesis that Bach did not change much, as a fugue theme composer from 1722 to 1742, carry out tests for the hypotheses $\xi_{I,j} = \xi_{II,j}$ and $\sigma_{I,j} = \sigma_{II,j}$, for the means and standard deviations, for features $x_1, \ldots, x_6$.

(i) Then compute empirical correlations, say $r_{I,j,k}$ and $r_{II,j,k}$, for the two datasets, for $j < k$. To compare these, test equality, using the machinery of Ex. 5.13. Argue that since kurtoses values are relatively small, these simpler methods will suffice, without bringing in the somewhat heavier machinery of Ex. 5.14.

(j) Take the 48 WTK clavier fugues together, and compare these with the organ fugues. What might be notable differences?

(k) (xx briefly, other themes, other questions to briefly explore. distance between two distributions for (x, y), when these take on integer values. xx)

Story viii.10 *Power law scaling for academics and support staff.* Considering the world of science, and more particularly the people populating the world's many research institutions, there is a surprisingly clear relationship between x_0, the number of scientists, and y_0, the number of non-scientists or support staff (from administration and economists and lawyers to a range of technical positions). Here we use a dataset building on Jamtveit et al. (2009), with (x_0, y_0) for $n = 61$ institutions. These 2008 data range from smaller centres, like the Centre for Advanced Study of Theoretical Linguistics at the University of Tromsø, with 18 academics and 2 support staff; the bigger ones, like the Faculty of Mathematics and Natural Sciences at the University of Oslo, with 944 academics and 356 support; to the truly gargantuan ones, like the UK National Health System, with 230,000 in science but 1,130,000 in various support positions. Intriguingly, all these data dots of (x_0, y_0), from the tiny to medium to very big, follow a very clear regression line on the log-scale, as seen in Figure viii.17, left panel. We shall work through the relevant details and aspects to land the associated growth equation

$$\text{number of support people} = c\,(\text{number of research people})^b,$$

with b a positive growth parameter. (xx point to this phenomenon being at work in various other context and applications. growing cities. Story iv.4. mention Jamtveit et al. (2018) for an instance of growth parameter b shifting after political reform. nils emil: we use the jamtveit data, with $n = 61$, but amend it slightly, using FHI, BI, and perhaps a few updated numbers, for MN fakultetet, for CEES, for NR. we ask around for these. xx)

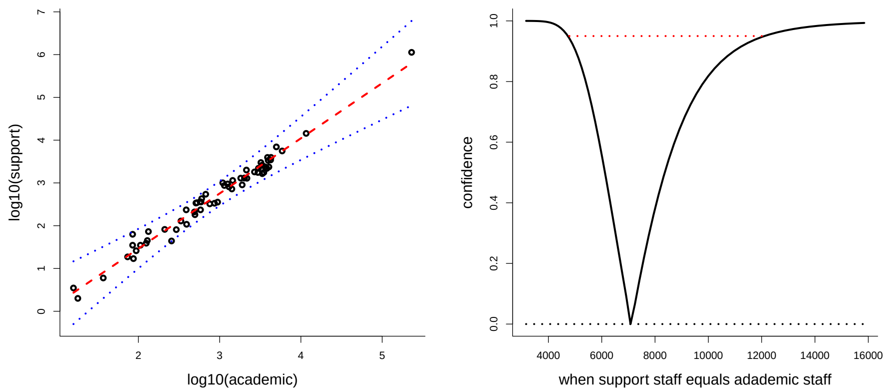

Figure viii.17: *Left panel: On* $\log_{10}$ *scale, the number of academics (x-axis) vs. the number of support employees (y-axis), for 61 research institutions, with regression line and 95 percent confidence band. Right panel: Confidence curve for the break-even size* $x_0 = 10^{-a/(b-1)}$ *at which the non-academic staff will equal the academic staff in size.*

(a) Transform to $x = \log_{10} x_0$ and $y = \log_{10} y_0$, and carry out linear regression analysis $y_i = a + bx_i + \varepsilon_i$ on those scales, with the ε_i seen as i.i.d. with mean zero and standard deviation σ. You should find $(\widehat{a}, \widehat{b}) = (-1.116, 1.289)$, with standard errors $(0.076, 0.025)$. Show that this leads to the power-law growth curve $\widehat{y}_0 = 0.077 \cdot x_0^{1.289}$ for relating the non-academic to the academic.

(b) In this context, searching for a universal statistical law valid along the full scale, from the smallest of apples to the colossal ones, argue that it makes sense to give each research institution equal weight. Back this up with inspection of the resiudals $\widehat{\varepsilon}_i = (y_i - \widehat{a} - \widehat{b} x_i)/\widehat{\sigma}$. For research institutions with 500 scientists, about how many non-scientists are there? Construct a version of Figure viii.17, left panel, with a 95 percent pointwise confidence band around the regression line. Estimate also the correlation, on the (x, y) scale, and give a confidence interval. Is the correlation on the (x_0, y_0) scale meaningful?

(c) So how big will a research environment need to be, in order for the number of non-scientists to equal the number of scientists? Argue that this concerns $\gamma = 10^{-a/(b-1)}$. Estimate this number, and construct a version of Figure viii.17, right panel, using ideas of Ex. 7.29 to work out a full confidence curve for this parameter.

(d) (xx to be finalised after having finalised the dataset. compare b for Norway, Denmark, Sweden, and Other. small differences, but significant. xx)

(e) (xx can bother to do this too, a fresh little change from logistic regression. xx) consider $R(x_0) = y_0/(x_0 + z_0)$, the fraction of non-academics in a research institution, by the above expected to be low for small but higher for bigger environments. explain that this leads to studying the parameter

$$\rho(x_0) = \frac{10^{a + b\log_{10} x}}{10^{\log_{10} x} + 10^{a + b\log_{10} x}} = \frac{10^a x^{b-1}}{1 + 10^a x^{b-1}},$$

and show that this is a logistic regression in $\log_{10} x_i$. plot $(\log_{10} x_0, R(x_0))$ along with the estimated $\widehat{\rho}(x_0)$, and give a 95 percent confidence band.

(f) We learn from the above that there is a bureaucratic growth parameter b at work for a long range of institutions, with $y_0 \doteq c x_0^b$. The growth parameter might however vary across societies, as we saw when comparing Norway, Denmark, Sweden, or over time, perhaps caused by political decisions. We now access the second dataset from Jamtveit et al. (2009), with information pertaining to the sizes of the Universities of Oslo, Bergen, Trondheim over the period 1960 to 2008. We may organise these data as triples $(t_i, x_{0,i}, y_{0,i})$, with t_i being calendar year minus 1960. Again transforming to $x_i = \log_{10} x_{0,i}$ and $y_i = \log_{10} y_{0,i}$, work through models 0, 1, 2, which have the y_i as respectively $\mathrm{N}(a_0 + b_0 x_i, \sigma_0^2)$, $\mathrm{N}(a_1 + (b_1 + c_1 t_i) x_i, \sigma_1^2)$, $\mathrm{N}(a_2 + (b_2 + c_2 t_i + d t_i^2) x_i, \sigma_2^2)$, the idea being to allow data to show us if b has not been constant over time. For the three candidate models, estimate the parameters, comparing in the end the $\widehat{\sigma}_j$ and the AIC scores aic_j, e.g. using Ex. 11.3. Show that model 2, with growth parameter seen as $b_2 + c_2 t + d_2 t^2$ over time $t = \text{year} - 1960$, is judged the best one. Plot the estimated growth parameter over this time window, and comment.

(g) Above the context and the natural interest in the growth parameter led naturally to a regression model with mean structure $a + (b + ct + dt^2)x$. Explain why and how this is different from the more traditional modelling with mean structure $a' + b'x + c't + d't^2$, say.

Story viii.11 *Statistical Sightings of Better Angels, I.* When is the next big interstate war coming? Why do the nations so furiously rage together, why do the people imagine a vayne thing? [xx nils edits this. need rus-ukr feb 2022 as datapoint, sadly. we go more quickly for ML in two-parameter model, but include also briefly moment-matching and quantile-matching. The Correlates of War story, here with emphasis on waiting times between wars, the $w_i = x_i - x_{i-1}$. They are approximately Expo, point to Lewis Fry Richardson volume editor Gleditsch, but the mixed expo works better; point to Pinker (2011), Hjort (2018b), Gleditsch (2020), Cunen et al. (2020a). point to data and description in 2.B. we ought to include Rus-Ukr too, where the CoW definition would say 2022, i suppose, not 2014. xx]

The dataset `allwars-data`, available at the book website, contains data pairs (x_i, z_i) for all $n_0 = 95$ gruesome interstate wars with at least 1000 battle deaths (as per well-maintained and publicly available databases for such matters, specifically the Correlates of War project), from the Franco-Spanish war in 1823 to the invasion of Iraque in 2003. Here x_i is the time where war i started, with dates transformed via months and days to decimals, so that the Korean war started at $x_{60} = 1950.483$, etc.; and z_i is the number of battle deaths. Check Figure iii.3 for seeing the $(x_i, \log z_i)$ data displayed, along with a horizontal line attempting to divide already big wars into the truly horrendously big ones and the relatively speaking less big ones. We return to several other aspects of these war data in Story iii.3, but presently focus attention on the x_i, and more specifically with *the between-times*

$$w_i = x_{i+1} - x_i \quad \text{for } i = 1, \ldots, n,$$

say, with $n = n_0 - 1 = 94$.

(a) There are both empirical studies and certain theoretical arguments, also for many other types of violence phenomena, pointing to the interesting and non-obvious supposition that the between-times ought to be approximately independent and identically exponentially distributed. In other words and terms, the w_i will behave as waiting times in a Poisson process with constant rate. Fit the model $f(w, \lambda) = \lambda \exp(-\lambda w)$ for $w > 0$ to the $w_1, \ldots, w_n$ data, via maximum likelihood. Assuming the model holds, give a 90 percent confidence interval for λ.

(b) For this one-parameter model, find a formula for the probability $p = p_1(\lambda)$ that the time between two consecutive wars is at least $w_0 = 3.00$ years. Estimate this probability, and find a 90 percent confidence interval.

(c) Perhaps the size of a war influences the eagerness with which cohorts of humankind again decide to embark on the next war? Fit the model where $w_i = x_{i+1} - x_i$ is an exponential with parameter $\lambda_i = \lambda_0 \exp(\beta v_i)$, where $v_i = \log z_i$, and comment on your findings.

(d) Broader models emerge by taking the w_i given λ to be exponential with this parameter λ, but to take the λ not as a single constant, but coming from a distribution of such rates. Assume that λ comes from a Gamma distribution with parameters (a, b), i.e. with density proportional to $\lambda^{a-1}\exp(-b\lambda)$. As in Ex. 1.10, show that this leads to c.d.f. and density

$$\begin{aligned} G(w, a, b) &= 1 - \{b/(b+w)\}^a = 1 - \exp\{-a\log(1 + w/b)\}, \\ g(w, a, b) &= ab^a/(b+w)^{a+1}, \end{aligned}$$

for $w > 0$. Starting from $\mathrm{E}\,(W \mid \lambda) = 1/\lambda$ and $\mathrm{Var}\,(W \mid \lambda) = 1/\lambda^2$, find explicit expressions for the mean and variance of W.

(e) You should now fit also this two-parameter model to the $w_1, \ldots, w_n$ data. We return to the more canonical maximum likelihood method below, but first we use the occasion to see two other estimation schemes in action. (i) Use moment matching, see Ex. 2.28, solving two equations with the two parameters, to find $(\widehat{a}_m, \widehat{b}_m)$. (ii) Use also quantile matching, see Ex. 2.30, finding $(\widehat{a}_q, \widehat{b}_q)$ by setting two sample quantiles $Q_n(q_1)$ and $Q_n(q_2)$ equal to the model based $G^{-1}(q_1)$ and $G^{-1}(q_2)$. Choose first $(q_1, q_2) = (1/4, 3/4)$, but use your code with other quantiles to check whether the resulting parameter estimates are reasonably stable. Then given such parameter estimates, construct figures plotting the empirical c.d.f. $F_n(w)$ along with the parametrically estimated versions, including also the simpler $1 - \exp(-\widehat{\lambda}w)$. Comment on your findings here.

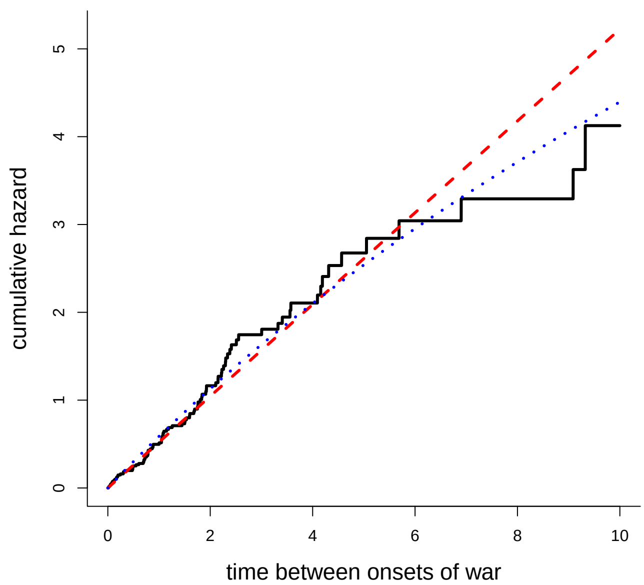

Figure viii.18: *The empirical cumulative hazard for the between-wars time data (black curve), along with the two fitted parametric hazard cumulatives* $A(w, \widehat{\lambda})$ *and* $A(w, \widehat{a}, \widehat{b})$.

(f) Find a formula $p = p_2(a, b)$ for the probability that the waiting time between two wars is at least $w_0 = 3.00$ years. Estimate this p, using the parameter estimates you've found above, and compare with $\widehat{p}_1 = p_1(\widehat{\lambda})$. [xx something more, about finding confidence interval, approximate standard deviation of $\widehat{p}_2$, etc. could ask for bootstrapping; will point to delta method. yes, we ask the readers to go through delta method for $p_2(\widehat{a},\widehat{b})$, using results from earlier exercises, about $(\bar{w},\widehat{\sigma})$. xx]

(g) We now turn to ML estimation of the two-parameter model. It is fruitful to parametrise the gamma mixing distribution via $(a,b)=(\lambda_0/c,1/c)$; show that the random λ then has mean λ_0 and variance $c\lambda_0$. Show that the density may be written $g(w,\lambda_0,c)=\lambda_0/(1+cw)^{1+\lambda_0/c}$; show that it is close to $\lambda_0\exp(-\lambda_0 w)$ for small c. Write down the log-likelihood function $\ell_n(\lambda_0,c)$ and find its maximisers $(\widehat{\lambda}_0,\widehat{c})$. Construct a version of Figure viii.18, with the nonparametric Nelson–Aalen estimate $\widehat{A}(w)$ alongside the parametric $\widehat{\lambda}w$ and $A(w,\widehat{\lambda}_0,\widehat{c})$. (xx polish this. xx)

(h) (xx then the things with the ML here. polish, calibrate with the above and with what we have in Ch7 with CDs for boundary parameters, and round off. xx) In this context we wish to have a clear test for $c=0$, corresponding to Poisson process behaviour for the waiting times, versus $c>0$. This requires more care than usual since $c=0$ sits at the boundary of the parameter space, as opposed to being an inner point. To study the ML estimator $\widehat{c}$ with the required care, show that the log-likelihood profile function becomes

$$\ell_{n,\rm prof}(c)=\max_{\text{all }\lambda_0}\ell_n(\lambda_0,c)=-n\{\log B_n(c)+cB_n(c)+1\},$$

with $B_n(c)=n^{-1}\sum_{i=1}^n(1/c)\log(1+cw_i)$. Plot it for the war onset waiting time data. For small c, show that $B_n(c)\doteq\bar{w}-\frac{1}{2}c(v_n^2+\bar{w}^2)$, where $\bar{w}=(1/n)\sum_{i=1}^n w_i$ and $v_n^2=(1/n)\sum_{i=1}^n(w_i-\bar{w})^2$ are the mean and variance of the w_i; with continuity, therefore, we have $B_n(0)=\bar{w}$ and $B_n'(0)=-\frac{1}{2}(v_n^2+\bar{w}^2)$. Show that

$$\ell_{n,\rm prof}'(0)=-n\{B_n'(0)/B_n(0)+B_n(0)\}=\tfrac{1}{2}n\bar{w}(v_n^2/\bar{w}^2-1).$$

Argue that the ML estimator $\widehat{c}$ is positive, provided $v_n/\bar{w}_n>1$, but zero, in the case of $v_n/\bar{w}\le 1$. Check that the derivative at zero is indeed positive for the war onset data. (xx then round off. note that $v_n^2/\bar{w}_n\to_{\rm pr}1$ if the data really come from an exponential. so prof half etc. xx) the approximation $\sqrt{n}(\widehat{c}-\delta/\sqrt{n})$ under $c=\delta/\sqrt{n}$ which makes it possible to have both a test, a p-value, different from the usual things, and a CD for c. under $c=0$, should land at $\sqrt{n}\widehat{c}/\widehat{\lambda}_0\to_d\max(0,N)$, half a normal, and $D_n=2(\ell_{n,\max}-\ell_{n,0})\to_d\max(0,N)^2$, half a chisquared. so pvalue is ... $1-\Phi(D_n^{1/2})$, which is 0.039; hence expo hypothesis is rejected. we need to crank out a good CD, and need exercise with $\sqrt{n}(\widehat{c}-\delta/\sqrt{n})$ limit, at the end of Ch5, to be used in Ch7.

(i) (xx rewrite and polish. xx) Above various analyses have been based on the observed between-war times, up to $w_{94}=x_{95}-x_{94}$. There is also information in the fact that since onset time $x_{95}=2003.219$, there have gudsigforbyde as of July 1, 2020, been no further interstate wars (well, according to the operative definitions of the Correlates of War project). Explain how this may be used to modify or update your previous analyses.

Story viii.12 *How many were killed in Srebrenica, 1995, and in Guatemala, 1978–1996?* In dramatic data analysed by Brunborg et al. (2003), numbers are reported for lists of killed Muslim men in Srebrenica 1995. They in particular go into the details of List A, by the International Committee of the Red Cross, and List B, by Physicians for Human Rights. We may draw up a simple Venn diagram, with 5,712 found on both lists, 1,586 on List A only, 192 on List B only. How can we estimate the number of people killed, outside both lists, i.e. outside the $A \cup B$ set in the Venn diagram? Similar challenges surface in connection with the complicated task of estimating the number of killed individuals in Guatemala, during the 1978–1995 period, where there are three lists A, B, C.

(xx need serious editing here. point to Ex. 4.45. basic reference Brunborg et al. (2003) for two-sources, then Lum et al. (2013) and patrick ball for Guatemala. in notes mention Bartolucci and Lupparelli (2008), Sanathanan (1972), Goudie and Goudie (2007), who point back to Laplace and Grant? we need a Venn diagram. see Figure viii.19. dealing with three sources means non-trivial generalisation from two to three. xx) (xx then this, appropriate placed, for Guatemala. emil constructs a Venn diagram for A, B, C. Sources are the Recovery of Historical Memory (REMHI), Commission for Historical Clarification (CEH), and the International Center for Human Rights Investigations (CIIDH), with acronyms reflecting project names in Spanish. The data are $n_{1,1,1} = 393$, $n_{1,1,0} = 3943$, $n_{1,0,0} = 15955$, $n_{1,0,1} = 634$, $n_{0,1,1} = 898$, $n_{0,1,0} = 19663$, $n_{0,0,1} = 6317$; the task is to estimate the full number $N = n_{0,0,0} + \cdots + n_{1,1,1}$ of individuals killed, hence also in the process the number $n_{0,0,0}$ of deads not captured on any of the three lists. (xx Ball xx) reports the overall estimate 132,174 for the total number of killed, with a standard error of 6,568; this agrees reasonably well with our likelihood analysis below. make clear here that our analysis is under independence. point to Ball and more. xx)

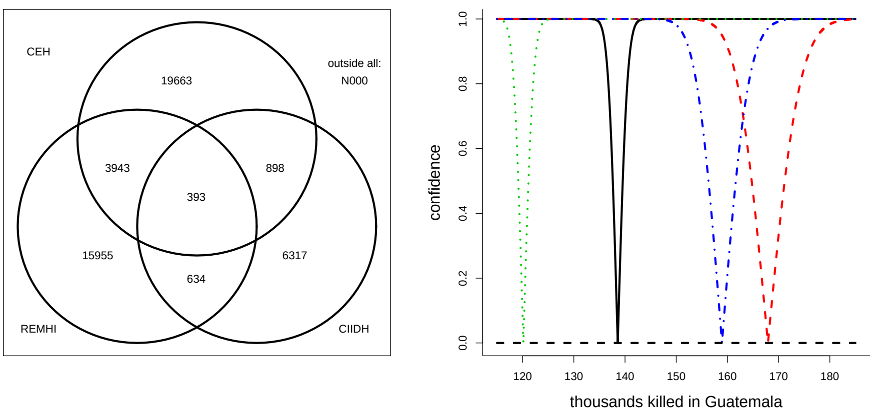

Figure viii.19: *Left panel: Venn diagram for the number of people killed and accounted for, for the three lists REMHI, CEH, CIIDH, and with N000 denoting those killed but not any of these lists. Right panel: Confidence curves for N, the total number of people killed in Guatemala 1978–1996, in thousands, based on all three sources (full black curve), and based on pairwise analyses.*

We start with the two-lists setup before we address the issues with three and more lists. Consider therefore the setup of Ex. 4.45, with a multinomial model for counts $n_{0,0}, n_{0,1}, n_{1,0}, n_{1,1}$ in a 2×2 table, but where $n_{0,0}$ and also the total population size $N = n_{0,0} + n_{0,1} + n_{1,0} + n_{1,1}$ are unknown. Assuming independence between the two underlying factors, so that the four probabilities $p_{0,0}, p_{0,1}, p_{1,0}, p_{1,1}$ can be expressed as $(1-p)(1-q), (1-p)q, (1-q)p, pq$, the simple estimator $N^* = n_{1,\cdot}n_{\cdot,1}/n_{1,1}$ was analysed; in particular, we found there that $(N^*-N)/N^{1/2} \to_d \mathrm{N}(0, \tau^2)$, with $\tau^2 = (1-p)(1-q)/(pq)$. Presently we use likelihood analysis for estimating N (and hence the hidden $n_{0,0}$), which also lends itself more easily to tables of higher order than two.

(a) When $n_{0,1}, n_{1,0}, n_{1,1}$ are observed, show that the likelihood function can be expressed as

$$L(N,p,q) = \frac{N!}{(N-s)!}\{(1-p)(1-q)\}^{N-s}\{(1-p)q\}^{n_{0,1}}\{p(1-q)\}^{n_{1,0}}(pq)^{n_{1,1}},$$

with $s = n_{0,1} + n_{1,0} + n_{1,1}$, so that $n_{0,0} + s = N$. Show that this leads to the profiled log-likelihood

$$\ell_{\mathrm{prof}}(N) = \log(N!) - \log((N-s)!) + NH(\widehat{p}_N) + NH(\widehat{q}_N),$$

in terms of the function $H(r) = r\log r + (1-r)\log(1-r)$, and where $\widehat{p}_N = n_{1,\cdot}/N$ and $\widehat{q}_N = n_{\cdot,1}/N$. We shall reach results for the ML estimator $\widehat{N}$.

(b) Use the Stirling approximation, see Ex. 4.31, to reach

$$\ell_{\mathrm{prof}}(N) \doteq (N + \tfrac{1}{2})\log N - (N - s + \tfrac{1}{2})\log(N-s) + NH(\widehat{p}_N) + NH(\widehat{q}_N).$$

With $r = c/N$, show that the derivative of $NH(r/N)$ becomes $H(c/N) - NH'(c/N)c/N^2 = H(r) - H'(r)r = \log(1-r)$. Use this to show that the first derivative becomes

$$U(N) = \log N - \log(N-s) + \log(1 - n_{1,\cdot}/N) + \log(1 - n_{\cdot,1}/N) + \varepsilon_N,$$

with $\varepsilon_N = O_{\mathrm{pr}}(1/N)$.

(c) Letting N_0 denote the true value, we shall use what we established in Ex. 4.45, that there is joint asymptotic normality $N_0^{1/2}(\widehat{p}_{i,j} - p_{i,j}) \to_d A_{i,j}$, with variances and covariances given in that exercise. Show now that $N_0^{1/2}U(N_0) \to_d U$, with

$$U = -\Big(\frac{A_{0,0}}{p_{0,0}} + \frac{A_{1,\cdot}}{1-p} + \frac{A_{\cdot,1}}{1-q}\Big).$$

Show that $A_{1,\cdot}$ and $A_{\cdot,1}$ have variances $p(1-p)$ and $q(1-q)$, that they are uncorrelated, and that (xx check with care xx) $\mathrm{cov}(A_{0,0}, A_{1,\cdot}) = -p_{0,0}p$, $\mathrm{cov}(A_{0,0}, A_{\cdot,1}) = -p_{0,0}q$. Use these facts to conclude that $U \sim \mathrm{N}(0, J)$, with

$$J = \frac{1 - p_{0,0}}{p_{0,0}} - \frac{p}{1-p} - \frac{q}{1-q} = \frac{pq}{(1-p)(1-q)}.$$

(d) Then show that the second derivative can be written

$$I(N) = 1/N - 1/(N-s) + \frac{n_{1,\cdot}/N^2}{1 - n_{1,\cdot}/N} + \frac{n_{\cdot,1}/N^2}{1 - n_{\cdot,1}/N},$$

and deduce that $-N_0 I(N_0) \to_{\rm pr} J$.

(e) Study the process $Z_{N_0}(d) = \ell_{\rm prof}(N_0 + dN_0^{1/2}) - \ell_{\rm prof}(N_0)$, and show that $Z_{N_0}(d) \to_d Z(d) = dU - \frac{1}{2}d^2 J$. Use this to show the two crucial results, (i) that $(\widehat{N} - N_0)/N_0^{1/2} \to_d U/J \sim {\rm N}(0, 1/J)$; (ii) that $D = 2\{\ell_{\rm prof,max} - \ell_{\rm prof}(N_0)\} \to_d U^2/J \sim \chi^2_1$. Explain how this may be used to form confidence intervals for N, hence also for the hidden $n_{0,0}$. Apply this method for the data given above for the two lists of killed Muslim men in Srebrenica 1995; display the $\ell_{\rm prof}(N)$ curve; give a 95 percent interval; and also a full confidence curve for N.

(f) (xx questions to nils emil, as of 23-Aug-2023. here we don't have any Bartlett identity, or Cramér–Rao, or Wilks, so we do things from scratch. but conclusions are as we're used to. is there a way of saying that the setup is sufficienty close to something more familiar, and then deduce normality and Wilks from this? xx)

(g) We then proceed to the setup with three lists, see Figure viii.19, left panel. Assuming independence, show that the log-likelihood function can be expressed as

$$\begin{aligned}\ell(N,p,q,r) &= \log(N!) - \log((N-s)!) + n_{1,\cdot,\cdot}\log p + n_{0,\cdot,\cdot}\log(1-p)\\ &\quad + n_{\cdot,1,\cdot}\log q + n_{\cdot,0,\cdot}\log(1-q) + n_{\cdot,\cdot,1}\log r + n_{\cdot,\cdot,0}\log(1-r),\end{aligned}$$

writing $s = n_{0,0,1} + \cdots + n_{1,1,1}$ for the sum over the seven observed cells, so that $N = n_{0,0,0} + s$. We use '$\cdot$' notation to indicate summing over the index or indexes in questions. Show that this leads to the profiled log-likelihood

$$\ell_{\rm prof}(N) = \log(N!) - \log((N-s)!) + NH(\widehat{p}_N) + NH(\widehat{q}_N) + NH(\widehat{r}_N),$$

in terms of the function $H(x) = x\log x + (1-x)\log(1-x)$, and with $\widehat{p}_N = n_{1,\cdot,\cdot}/N$, $\widehat{q}_N = n_{\cdot,1,\cdot}/N$, $\widehat{r}_N = n_{\cdot,\cdot,1}/N$.

(h) With Stirling approximation, reach first the approximation

$$\ell^*_{\rm prof}(N) = (N + \tfrac{1}{2})\log N - (N - s + \tfrac{1}{2})\log(N-s) + NH(\widehat{p}_N) + NH(\widehat{q}_N) + NH(\widehat{r}_N).$$

Then show that this leads to the score function

$$U(N) = \log N - \log(N-s) + \log(1-\widehat{p}_N) + \log(1-\widehat{q}_N) + \log(1-\widehat{r}_N) + \varepsilon_N$$

with $\varepsilon_N = O_{\rm pr}(1/N)$ (xx check this xx), in terms of $\widehat{p}_N = n_{1,\cdot,\cdot}/N$, $\widehat{q}_N = n_{\cdot,1,\cdot}/N$, $\widehat{r}_N = n_{\cdot,\cdot,1}/N$. Taking the second derivative, find also (xx check remainder size xx)

$$I(N) = 1/N - 1/(N-s) + \frac{n_{1,\cdot,\cdot}/N^2}{1 - n_{1,\cdot,\cdot}/N} + \frac{n_{\cdot,1,\cdot}/N^2}{1 - n_{\cdot,1,\cdot}/N} + \frac{n_{\cdot,\cdot,1}/N^2}{1 - n_{\cdot,\cdot,1}/N}.$$

(i) Letting N_0 denote the true number, show that as N_0 grows large there is a clear limit in probability

$$-N_0 I(N_0) \to_{\rm pr} J = \frac{1-p_{0,0,0}}{p_{0,0,0}} - \frac{p}{1-p} - \frac{q}{1-q} - \frac{r}{1-r} = \frac{pq+pr+qr-2pqr}{(1-p)(1-q)(1-r)}.$$

(j) Next on the agenda is working with the approximate distribution of $U(N_0)$. Writing $\widehat{p}_{i,j,k} = n_{i,j,k}/N$, use Ex. 4.44 to show that there is joint convergence in distribution $N_0^{1/2}(\widehat{p}_{i,j,k} - p_{i,j,k}) \to_d A_{i,j,k}$, the $2^3 = 8$ components of a zero-mean multinormal with variances $p_{i,j,k}(1-p_{i,j,k})$ and covariances $-p_{i,j,k}p_{i',j',k'}$ when the index sets are not identical. Show then that

$$\begin{aligned} N_0^{1/2}U(N_0) &= N_0^{1/2}\{-\log\widehat{p}_{0,0,0} + \log(1-\widehat{p}) + \log(1-\widehat{q}) + \log(1-\widehat{r})\} \\ &\to_d U = -\frac{1}{p_{0,0,0}}A_{0,0,0} - \frac{1}{1-p}A_{1,\cdot,\cdot} - \frac{1}{1-q}A_{\cdot,1,\cdot} - \frac{1}{1-r}A_{\cdot,\cdot,1}. \end{aligned}$$

This is a zero-mean normal, where it remains to find an expression for its variance. Use algebra to show that (xx check all this with care xx) $A_{1,\cdot,\cdot}, A_{\cdot,1,\cdot}, A_{\cdot,\cdot,1}$ are actually independent, with variances $p(1-p), q(1-q), r(1-r)$, and that ${\rm cov}(A_{0,0,0}, A_{1,\cdot,\cdot}) = -p_{0,0,0}p$, ${\rm cov}(A_{0,0,0}, A_{\cdot,1,\cdot}) = -p_{0,0,0}q$, ${\rm cov}(A_{0,0,0}, A_{\cdot,\cdot}, 1) = -p_{0,0,0}r$. Use these formulae to show that indeed ${\rm Var}\,U = J$. (xx note: there is no Bartlett identity here, so we need to this from scratch. xx)

(k) Work with the process $Z_{N_0}(d) = \ell_{\rm prof}(N_0 + dN_0^{1/2}) - \ell_{\rm prof}(N_0)$, and show that it converges in distribution to $Z(d) = Ud - \frac{1}{2}Jd^2$. Use this to derive (i) that $(N^* - N_0)/N_0^{1/2} \to_d U/J \sim {\rm N}(0, 1/J)$; (ii) that $D(N_0) = 2\{\ell_{\rm prof,max} - \ell_{\rm prof}(N_0)\} \to_d U^2/J \sim \chi^2_1$. Use this again to argue that ${\rm cc}(N) = \Gamma_1(D(N))$ is a valid confidence curve for N. Construct a version of Figure viii.19, right panel. Here the full black curve is the crucal one, using all three sources, centred at $N^* = 138,576$, with 95 percent interval 135,794 to 141,453 (length 5,659). This agrees reasonably well with (xx Ball xx). The three other confidence curves use two sources at a time (xx round off this xx)

	table 1	table 2	table 3	overall
n10	19,898	16,589	6,951	
n01	7,215	20,561	23,606	
n11	1,027	4,336	1,291	
ML	167,916	120,145	158,935	138,576
low	158,918	117,309	151,458	135,794
up	177,681	123,097	166,979	141,453
length	18,763	5,788	15,521	5,659

(l) (xx ask readers to simulate a bit, to estimate N with ${\rm cc}(N)$, using `union, setdiff, intersect` things. also to check that ${\rm cc}(N_0) \sim$ unif, in repeated sampling. xx)

(m) (xx we shall see, probably too much for us, depending on how smoothly can tall $\ell_{\rm prof}(N)$ things for general $p_{i,j,k}(\theta)$ models. heterogeneity for the lists' ability to find people: take $p \sim {\rm Beta}(k_1p_0, k_1(1-p_0))$, $q \sim {\rm Beta}(k_2q_0, k_2(1-q_0))$, $r \sim {\rm Beta}(k_3r_0, k_3(1-$

$r_0)$). this gives a six-parameter model for the mixed multinomial:

$$\begin{aligned}\bar f &= \frac{N!}{n_{0,0,0}!\cdots n_{1,1,1}!}\frac{\Gamma(k_1)}{\Gamma(k_1p_0)\Gamma(k_1(1-p_0))}\frac{\Gamma(k_1p_0+n-A)\Gamma(k_1(1-p_0)+A)}{\Gamma(k_1+N)}\\ &\times\frac{\Gamma(k_2)}{\Gamma(k_2q_0)\Gamma(k_2(1-q_0))}\frac{\Gamma(k_2q_0+n-B)\Gamma(k_2(1-q_0)+B)}{\Gamma(k_2+N)}\\ &\times\frac{\Gamma(k_3)}{\Gamma(k_3r_0)\Gamma(k_3(1-r_0))}\frac{\Gamma(k_3q_0+n-C)\Gamma(k_3(1-r_0)+C)}{\Gamma(k_3+N)},\end{aligned}$$

with $A = n_{1,\cdot,\cdot}, B = n_{\cdot,1,\cdot}, C = n_{\cdot,\cdot,1}$. For the Guatemala Venn diagram data, show that $(A, B, C) = (20925, 24897, 8242)$.

Story viii.13 *New Haven annual temperatures 1912-1971.* Figure viii.20 displays the annual average temperatures at New Haven, Connecticut, in Celcius, for the years 1912 to 1971. (xx point to 2.B, in Ch. B. xx) Our task here is to analyse these data using first a simple linear normal regression model, to assess whether the upward trend is significiant, and to construct 'prediction intervals' for years a bit before and a bit after the observation range 1912-1971. We also investigate whether the data support more sophisticated modelling, (i) by using t-distributed error terms, with heavier tails than those implied by the traditional normal assumption, and (ii) by allowing for autocorrelation in the yearly data.

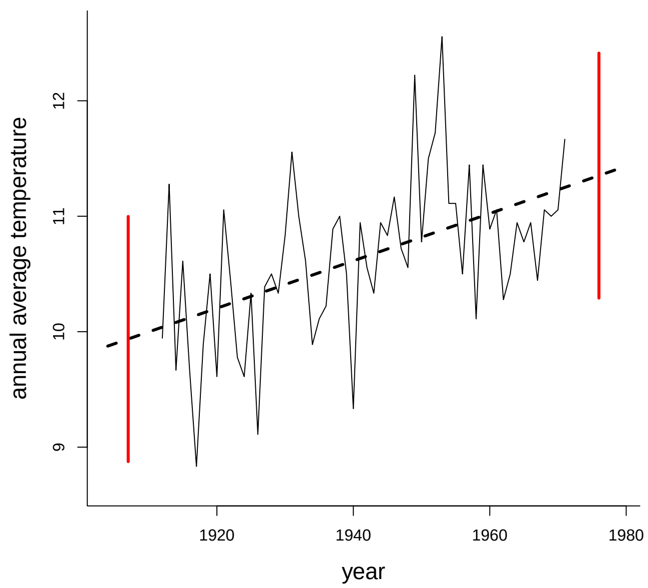

Figure viii.20: *Annual average temperatures (in Celsius) at New Haven, Connecticut, from 1912 to 1971. Also plotted is the linear trend, and 90 percent prediction intervals for average temperature for the years 1907 and 1976.*

(a) For simplicity of computation, write $t_i = x_i - 1912$, for year x_i. With y_i the average temperature in year x_i, fit the linear normal regression model $y_i \sim \mathrm{N}(a + bt_i, \sigma^2)$. Find confidence intervals for b and for σ, and show in particular that b is indeed significantly positive.

(b) For any year x_0 outside the 1912-1971 range, form a 90 percent prediction interval for the average temperature Y_0 in that year. In other words and symbols, construct $[L(x_0), U(x_0)]$ such that $P(Y_0 \in [L(x_0), U(x_0)]) = 0.95$. Construct a version of Figure viii.20, where the two extra years are 1907 and 1976. Comment on your findings. Try also with 1897 and 1986.

(c) Compute the estimated residuals $r_i = (y_i - \widehat{a} - \widehat{b}x_i)/\widehat{\sigma}$, and plot these as a function of year x_i. Use this to check aspects of the modelling assumptions, including the independence.

(d) Sometimes meteorological data like these exhibit heavier tails than those implied by the normality assumption. Look therefore into the extended four-parameter model which takes $y_i = a + bt_i + \sigma\varepsilon_i$, where the ε_i are i.i.d. t_ν, the t distribution with degrees of freedom ν. Compute and display the log-likelihood profile function $\ell_{\mathrm{prof}}(\nu)$, by maximising for each ν over (a, b, σ), and find the ML estimates. Also display a confidence curve for ν. (xx we give a figure here, askin readers to reproduce it; the point is that normality is inside the likely range. explain that if ν is moderate to small, it might not affect predictions so much, but the prediction intervals. xx)

(e) Then attempt another direction of sophistication, allowing autocorrelation. The model is now $y_i = a + bt_i + \sigma\varepsilon_i$, with the ε_i being jointly normal, with variance 1, and correlations $\rho^{|i-j|}$, for some ρ. Compute and display the profiled log-likeihood function $\ell_{\mathrm{prof}}(\rho)$, and give a confidence curve for ρ. How do you conclude?

Story viii.14 *Where are the snows of yesteryear?* Figure viii.23 is a potentitally dramatic one, for core segments of the Norwegian population, displaying the number of skiing days per year, from 1897 to 2015, at the location Bjørnholt in Nordmarka, a tram distance and a skiing hour north of central Oslo. A skiing day is defined as there being at least 25 cm snow on the ground. The data are in (xx in the Ch overview xx). How clear is the downward trend, will we still be able to ski, a dozen years from now?

(a) Since there is a gap in the time series, with no data from 1938 to 1954, we need a bit of care both with the notation and the analysis. With data index $t = 1, 2, \ldots, n$, write z_t for $\mathrm{year}_t - 1896$, these running from 1 to 119, though $n = 102$, due to the hole in the data. Fit the simple linear regression model to the skiing days data, with $y_t = \alpha_0 + \alpha_1 z_t + \sigma_0\varepsilon_{0,t}$, where the $\varepsilon_{0,t}$ are seen as i.i.d. $\mathrm{N}(0,1)$. Find confidence intervals for the slope α_1, for σ, and for the expected number of skiing days in 2023, given the available information up to 2015. Check the residuals $r_{0,t} = (y_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 z_t)/\widehat{\sigma}_0$, both for constancy of variance, and for autocorrelation, using the apprprincipate `acf` algorithm.

(b) To investigate whethere is is autocorrelation in the data, with possible consequences for both slope estimation and prediction, explore the four-parameter model

$$y_t = \beta_0 + \beta_1 z_t + \sigma\varepsilon_t, \quad \text{where } \mathrm{cov}(\varepsilon_s, \varepsilon_t) = \rho^{|s-t|}.$$

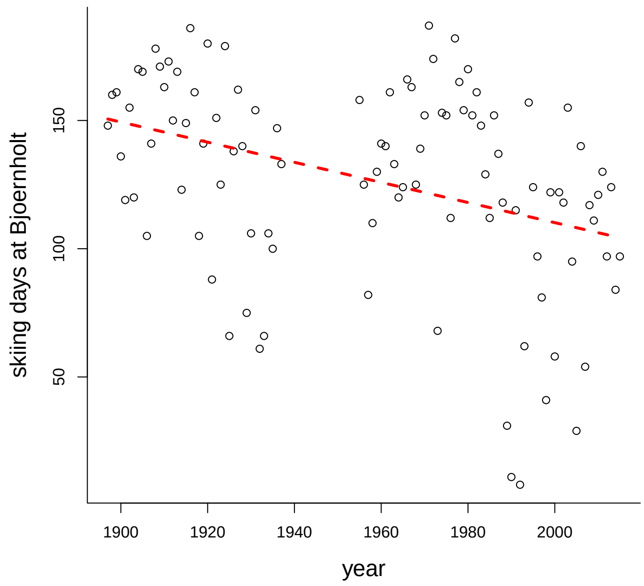

Figure viii.21: *The number of skiing days per year, at the location Bjørnholt in Nordmarka, from 1897 to 2015, though with a gap in the series, with no records from 1938 to 1954. The dashed line is the estimated regression from (xx the four-parameter model xx).*

Compute the profiled log-likelihood function $\ell_{\text{prof}}(\rho)$, and give the associated confidence curve $\text{cc}(\rho)$. (xx point to Ex. 12.21. and to cc recipe. xx)

(c) Compute AIC scores for the three-parameter and the four-parameter model, and comment. Also test $\rho = 0$ using (xx point to method of Ch3 xx).

(d) Use the four-parameter model to plot the data along with the estimated mean curve and a 90 percent pointwise confidence band.

(e) (xx can do even more. point to Cunen et al. (2018). xx)

Story viii.15 *Where are the snows of yesteryear?* Figure viii.23 is a potentitally dramatic one, for core segments of the Norwegian population, displaying the number of skiing days per year, from 1897 to 2015, at the location Bjørnholt in Nordmarka, a tram distance and a skiing hour north of central Oslo. A skiing day is defined as there being at least 25 cm snow on the ground. The data are in (xx in the Ch overview xx). How clear is the downward trend, will we still be able to ski, a dozen years from now?

(a) Since there is a gap in the time series, with no data from 1938 to 1954, we need a bit of care both with the notation and the analysis. With data index $t = 1, 2, \ldots, n$,

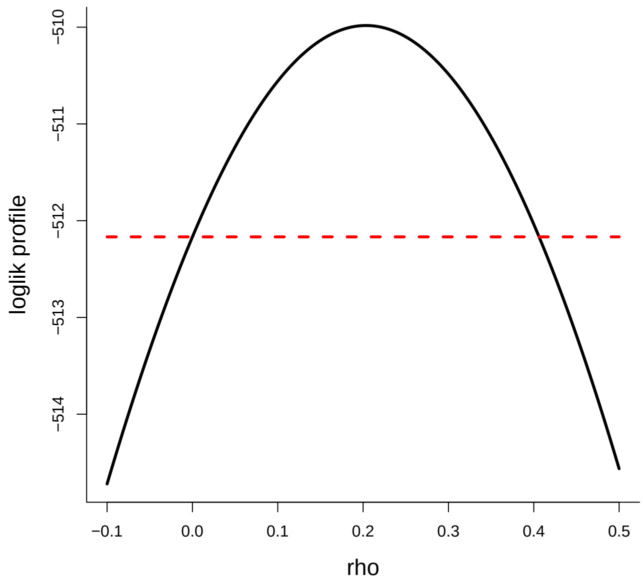

Figure viii.22: *The log-likelihood profile funcion $\ell_{\rm prof}(\rho)$, for the four-parameter model (xx). The horizontal dashed line indicates the level $\ell_{1,\max}$ obtained for the submodel of independence, where $\rho = 0$.*

write z_t for $\text{year}_t - 1896$, these running from 1 to 119, though $n = 102$, due to the hole in the data. Fit the simple linear regression model to the skiing days data, with $y_t = \alpha_0 + \alpha_1 z_t + \sigma_0 \varepsilon_{0,t}$, where the $\varepsilon_{0,t}$ are seen as i.i.d. $\mathrm{N}(0,1)$. Find confidence intervals for the slope α_1, for σ, and for the expected number of skiing days in 2023, given the available information up to 2015. Check the residuals $r_{0,t} = (y_t - \widehat\alpha_0 - \widehat\alpha_1 z_t)/\widehat\sigma_0$, both for constancy of variance, and for autocorrelation, using the approrpriate `acf` algorithm.

(b) To investigate whethere is is autocorrelation in the data, with possible consequences for both slope estimation and prediction, explore the four-parameter model

$$y_t = \beta_0 + \beta_1 z_t + \sigma \varepsilon_t, \quad \text{where } \mathrm{cov}(\varepsilon_s, \varepsilon_t) = \rho^{|s-t|}.$$

Compute the profiled log-likelihood function $\ell_{\rm prof}(\rho)$, and give the associated confidence curve $\mathrm{cc}(\rho)$. (xx point to Ex. 12.21. and to cc recipe. xx)

(c) Compute AIC scores for the three-parameter and the four-parameter model, and comment. Also test $\rho = 0$ using (xx point to method of Ch3 xx).

(d) Use the four-parameter model to plot the data along with the estimated mean curve and a 90 percent pointwise confidence band.

(e) (xx can do even more. point to Cunen et al. (2018). xx)

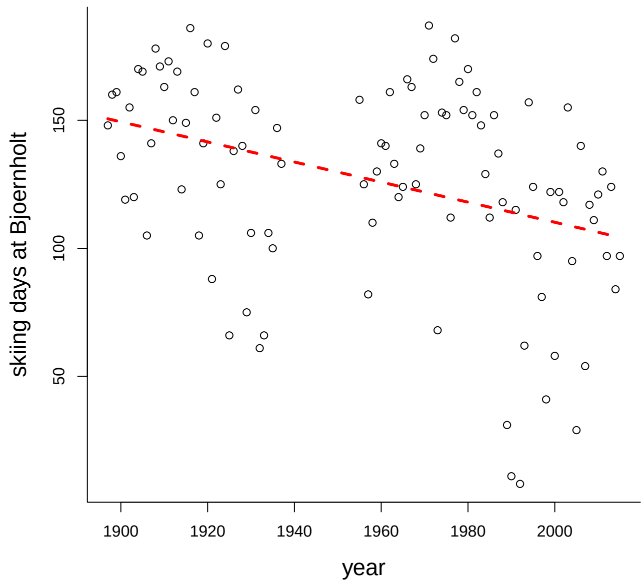

Figure viii.23: *The number of skiing days per year, at the location Bjørnholt in Nordmarka, from 1897 to 2015, though with a gap in the series, with no records from 1938 to 1954. The dashed line is the estimated regression from (xx the four-parameter model xx).*

Story viii.16 *How many Clethrionomys glareoli?* In work reported on in Blower et al. (1981, p. 83), a population of the bank vole *C. glareolus*, inside a certain area of biological interest, the voles were trapped, then marked and released (and potentially trapped again), over a six-month period. In total, 53 different voles were caught, in the course of 109 captures. So how many glareoli were there?

(xx serious cleaning in a little while, after moving an earlier Ch5 exercise here. xx) This and related problems have a connection to the card collector problem studied in Ex. 4.65. Consider a version of that setup, with cards $X_1, X_2, \ldots$ being sampled from $\{1, \ldots, n\}$ with equal probabilities $1/n$. In the exercise pointed to we investigated the full time $T_1 + \cdots + T_n$ it takes to have the dull deck of cards, with T_r time needed to have seen new card no. r, having started clocking time again after having previously found $r - 1$ cards. Here we turn the table and ask how many n cards there are, based on having seen r different cards after V_r attempts. With many repetitions among the sampled cards one expects a low n, and if one needs many samples to reach a low r one expects the opposite.

(a) Via arguments discussed in Ex. 4.65, show that $V_r = T_1 + \cdots + T_r$, with independent

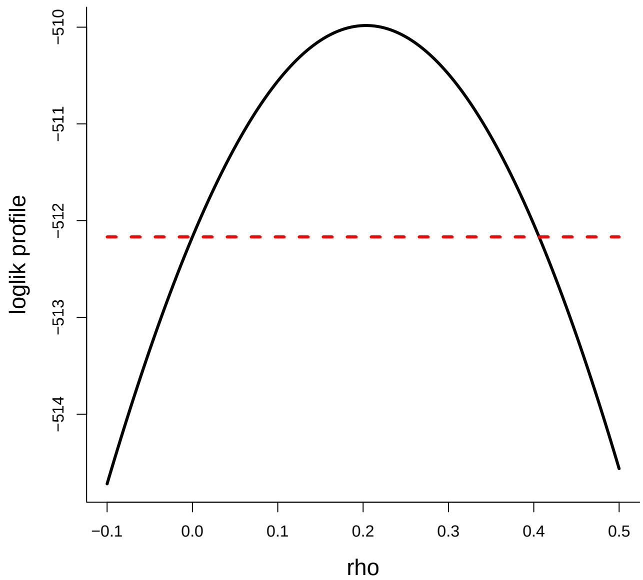

Figure viii.24: *The log-likelihood profile funcion* $\ell_{\rm prof}(\rho)$*, for the four-parameter model (xx). The horizontal dashed line indicates the level* $\ell_{1,\max}$ *obtained for the submodel of independence, where* $\rho = 0$.

waiting times $T_i \sim \text{geom}(p_i)$, and $p_i = (n-i+1)/n = 1-(i-1)/n$. Show that

$$\xi_r(n) = \mathrm{E}_n\,V_r = \sum_{i=1}^r 1/p_i = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{n-r+1} = n(H_n - H_{n-r}),$$

where $H_n = 1 + 1/2 + \cdots + 1/n$ is the harmonic series partial sum. If $V_r = 109$ captured-released-recaptured samples from a closed population of an unknown number n of animals yield $r = 53$ different animals, use this moment equation to estimate n.

(b) Show that the joint distribution of the observed $(T_1, \ldots, T_r)$ is

$$1 \cdot \Big(\frac{1}{n}\Big)^{t_2-1}\Big(1-\frac{1}{n}\Big)\cdots\Big(\frac{r-1}{n}\Big)^{t_r-1}\Big(1-\frac{r-1}{n}\Big)$$

with log-likelihood

$$\ell_r(n) = \sum_{i=2}^r \Big\{(t_i-1)\log\Big(\frac{i-1}{n}\Big) + \log\Big(\frac{1-i-1}{n}\Big)\Big\}.$$

Conclude also that $V_r = T_1 + \cdots + T_r$ is sufficient for n. The ML estimator is the maximiser of $\ell_r(n)$, rounded off to nearest integer, if required.

(c) Allowing ourselves taking the derivative with respecct to n, even though it is not a continuous parameter, show that $\partial\ell_r(n)/\partial n = -V_r/n + H_n - H_{n-r}$. Use this to show the ML estimator $\widehat{n}$ is the same as the moment estimator. The result is $\widehat{n} = 64$.

(d) (xx a bit more. something about normal approximation being ok for some (r, n) areas but not others. but $\mathrm{cc}(n)$ is perfect. xx)

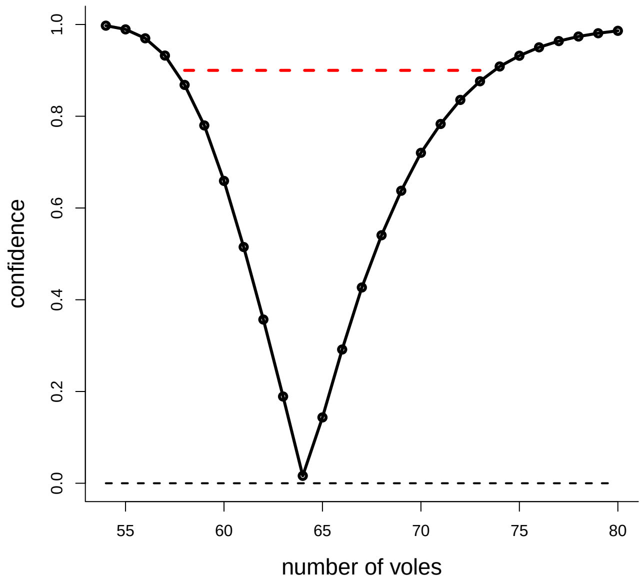

Figure viii.25: *Confidence curve* $\mathrm{cc}(n)$ *for the number of Clethrionymus glarealus, after having trapped* $r = 53$ *different animals in the course of* $V_r = 109$ *trappings.*

(e) Show that the likelihood function for $T_1, \ldots, T_r$ may be expressed as

$$n^{-V_r} a_r(n) = \exp\{-(V_r/r)\log(n/r) + b_r(n)\},$$

for suitable $a_r(n)$ and $b_r(n)$. Show that this is an exponential class situation, see Ex. 1.57 (xx check with care, also regarding canonical parameter, and uses in Ch7 xx), with $\log(n/r)$ the canonical parameter and V_r/r the sufficient statistic.

(f) Use theory from (xx point to exercise in Ch7) to argue that the confidence distribution

$$C_r(n) = P_n(V_r < V_{r,\mathrm{obs}}) + \tfrac{1}{2} P_n(V_r = V_{r,\mathrm{obs}}) \quad \text{for } n \geq V_{r,\mathrm{obs}}$$

is optimal (modulo half-correction for discreness). Implement this, computing $C_r(n)$ via a high number of simulations of V_r for each n in question, and display the confidence curve $\mathrm{cc}(n)$, as in Figure viii.25. Find also a 90 percent confidence interval, and compare the ML estimate with the median confidence estimate.

(g) (xx round off. xx)

Story viii.17 *Golf putting probabilities.* You're golfing, and when closer to the hole than some twenty feet need to focus on your putting. Drawn from databases of several hundreds of professional tournaments, the data below, from Gelman and Nolan (2002) with further discussion in Schweder and Hjort (2016, Ch. 14), give the number m_j of attempts and the number y_j of successful ones from these, at distances x_j, in feet, for say $j = 1, \ldots, k$. Our story concerns estimating the success probability $p(x_j)$, and also modelling the inherent variability at work. See Figure viii.26, which in particular also displays the raw estimates $\widetilde{p}_j = y_j/m_j$, with small vertical 90 binomial confidence intervals around them. It will be of relevance for a few of these models to factor in the radii R for the hole and r for the ball, which are respectively 4.252/2 inches and 1.680/2 inches, or 0.0177 and 0.070 on the foot scale.

```
feet away; number of tries; number of successes
                11  237   75
 2 1443 1346    12  202   52
 3  694  577    13  192   46
 4  455  337    14  174   54
 5  353  208    15  167   28
 6  272  149    16  201   27
 7  256  136    17  195   31
 8  240  111    18  191   33
 9  217   69    19  147   20
10  200   67    20  152   24
```

(a) We start out viewing the data as a sequence of independent binomial experiments, with $Y_j \sim \text{binom}(m_j, p_j)$ for $j = 1, \ldots, k$. The task is to model $p_1, \ldots, p_k$, as functions of the distances $x_1, \ldots, x_k$ to the hole. Show that with any such model, say $p_j = p(x_j, \theta)$, the log-likelihood function becomes $\sum_{j=1}^k [y_j \log p_j(\theta) + (m_j - y_j) \log\{1 - p_j(\theta)\}]$. Carry out logistic regressions, in x (order one); in $x, (x - \bar{x})^2$ (order two); in $x, (x - \bar{x})^2, (x - \bar{x})^3$ (order three); in $x, (x - \bar{x})^2, (x - \bar{x})^3, (x - \bar{x})^4$ (order four). As usual, $\bar{x}$ is the mean of the x_j. For each of these models, estimate and plot the curves

$$p_1(x) = H(a + bx) \quad \text{up to} \quad p_4(x) = H(a + bx + c(x - \bar{x})^2 + d(x - \bar{x})^3 + e(x - \bar{x})^4).$$

This can be achieved in R via `glm(cbind(y,m-y) ~ x + x2 + x3, family=binomial)`, and so on; for this standard type of model there is then no need to programme the log-likelihood function etc. For the four models, find the log-likelihood maxima and AIC scores, as per Chapter 11. In particular, you should find that the most traditional order one model does not work well here, and that AIC prefers the order four model among these.

(b) Considering the population of good golfers, and disregarding other geometric aspects of these thousands of putting situations, let Z be *the angle of the put*, from putting position to the hole. Not all attempts are perfect (Z close to zero), so we can translate

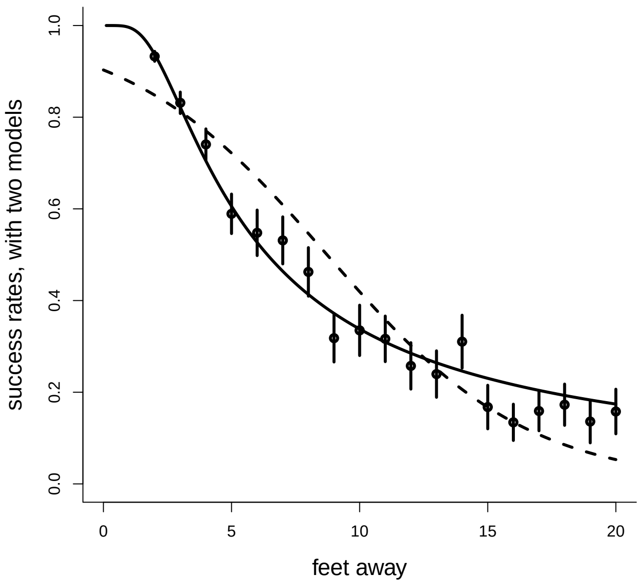

Figure viii.26: *The raw success estimates* $\widetilde{p}_j = y_j/m_j$ *at distances* $2, 3, \ldots, 20$ *feet from the hole, along with small vertical 90 percent binomial intervals. The full curve is the fitted* $p(x_j, a, b)$ *with the geometric model, and the dashed curve is the simple logistic regression curve.*

uncertainty and variability to a distribution of the random angle Z. In terms of such a distribution, show that to a good geometric approximation,

$$p(x) = P(\sin Z \in (-(R-r)/x, (R-r)/x)) \quad \text{for } x \geq R - r.$$

A natural model for the random angles is a normal $(0, \sigma^2)$. Fit the resulting model

$$p(x_j, \sigma) = P(|\sigma\, N| \in (-d_j, d_j)) \quad \text{for } j = 1, \ldots, k,$$

with N denoting a standard normal, and where we write $d_j = \arcsin((R-r)/x_j)$ for the bounds inside which successful putting angles must land at distance x_j. Compute the log-likelihood maximum, and compare with the logistic regressions above, using the AIC scores. In particular, demonstrate that this simple geometric one-parameter moel works better than logistic regressions of order one and two.

(c) (xx a little bit more with simple $p(x_i, \sigma)$ model before we go to variable σ. show that it works much better than standard first order two-parameter logistic regression; see also Figure viii.26. check $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_k$, fitted at the individual x_j. Can the $\widehat{\sigma}_j$ reasonably be taken as constant, across putting distances? can have a simple figure. xx)

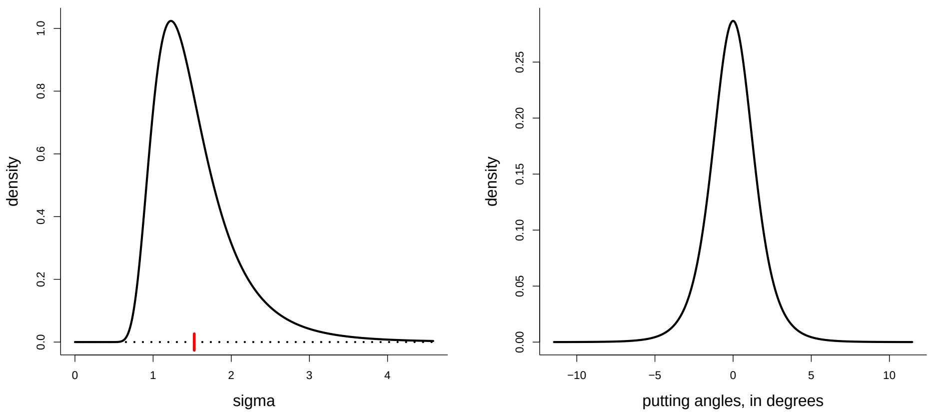

Figure viii.27: *Left panel: The estimated density for σ, on the scale of ordinary degrees (i.e. $90/(\pi/2)$ times radians). The point estimate 1.53 from the no-variability model is shown on the horizontal axis. Right panel: The estimated density of putting angles, again in ordinary degrees.*

(d) The simple model above somehow puts all angular uncertainty into one common σ. It might be better and more informative to view these σ as coming from a distribution, across golfers. There are several such possibilities, starting with $Z\,|\,\sigma \sim \mathrm{N}(0, \sigma^2)$, but here we take σ^2 to have an inverse gamma distribution, i.e. $\lambda = 1/\sigma^2 \sim \mathrm{Gam}(a, b)$; the model is flexible, and we can get through the mathematics to an explicit distribution for Z. Writing $g(\cdot, a, b)$ for that gamma density, show that this leads to a density for the random Z of the form

$$\bar{f}(z) = \int_0^\infty \phi(\lambda^{1/2} z)\lambda^{1/2} g(\lambda, a, b)\,\mathrm{d}\lambda = \frac{1}{(2\pi)^{1/2}} \frac{\Gamma(a + \frac{1}{2})}{\Gamma(a)} \frac{b^a}{(b + \frac{1}{2}z^2)^{a+1/2}}.$$

While this may be worked with directly, it is useful to transform the density to a member of the well-known distributions, to facilitate computations of probabilities etc. Show therefore that

$$V = (\tfrac{1}{2}Z^2/b)/(1 + \tfrac{1}{2}Z^2/b) \sim \mathrm{Beta}(\tfrac{1}{2}, a),$$

and express the c.d.f. of Z in terms of the c.d.f. Be of this Beta distribution. Demonstrate that all this leads to the model

$$p(x_j, a, b) = P(|Z| \le d_j) = \mathrm{Be}((\tfrac{1}{2}d_j^2/b)/(1 + \tfrac{1}{2}d_j^2/b), \tfrac{1}{2}, a) \quad \text{for } j = 1, \ldots, k.$$

Fit this model numerically, maximising the log-likelihood function; you should find $(\widehat{a}, \widehat{b}) = (2.8498, 0.00154)$. Compute also the log-likelihood maximum, and demonstrate that this two-parameter model has the best AIC score of the four plus two models considered (so far).

(e) Compute and display the estimated densities for σ, the variable normal scale for the putting angle, and for Z, the putting angle itself. Since most golfers prefer standard angular degrees to radians, present these densities on the degree scale $z' = 90/(\pi/2)\,z = (180/\pi)\,z$. Construct versions of the plots in Figure viii.27. With Z the random angle, use these fitted densities to demonstrate that 95 percent of all putting angles are inside ± 3.30 degrees, and 99 percent are inside ± 5.05 degrees. Note that this signifies rather heavier tails than for the normal; in a fair proportion of cases, the shot is off with more than say 4 degrees, which is enough to not hit the hole.

(f) It does perhaps not appear likely, but we may check statistically whether this population of players might have some systematic angular bias in their putting. The simplest check on this is to use the two-parameter normal $Z \sim (\xi, \sigma^2)$. Compute and display the profiled log-likelihood $\ell_{\rm prof}(\xi)$. It will indeed be seen to be very flat at the top, around zero, with no indication of such a bias. (xx point to previous stuff perhaps in Ch1 regarding noise in ξ being picked up in the σ. xx)

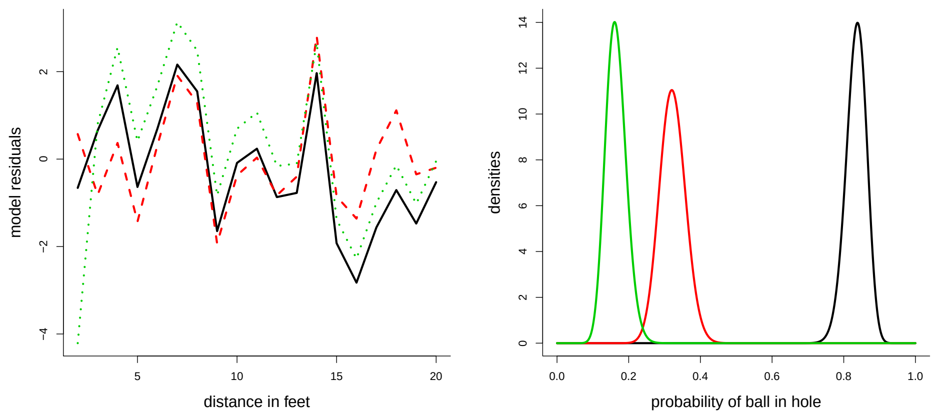

Figure viii.28: *Left panel: model residuals* $(y_j - m_j\widehat p_j)/\{m_j\widehat p_j(1-\widehat p_j)\}^{1/2}$, *for two-parameter geometric model (full black), simple geometric model (dotted), and for the third-order logistic model (dashed). These indicate good fit, but overdispersion. Right panel: At distances x equal to 3, 10, 20 feet, the Beta-binomial model leads to probability densities, centred around respectiely 0.701, 0.322, 0.165. The strict binomial models take these probabilities as fixed, for all golfers at a fixed distance from the holw.*

(g) (xx we do one more thing. binomial overdispersion. point to Story viii.3. xx) There is another tool for assessing and comparing model adequacy, with such data for a collection of tables, which is to monitor the model residuals $\widehat r_j = (y_j - \widehat p_j)/\{m_j\widehat p_j(1-\widehat p_j)\}^{1/2}$ for $j = 1,\ldots,k$, with $\widehat p_j$ that model's implied estimates of the p_j. If the model is adequate, explain why these should be distributed approximately as standard normals; also, the residual sum of squares $Q = \sum_{j=1}^k \widehat r_j^2$ should roughly have a $\chi^2_{\rm df}$ distribution, with ${\rm df} = n - p$, where p is the number of parameters estimated. Compute and display these residuals, for the models entertained so far; see the left panel of Figure viii.28. It will be seen that even when the fitted $\widehat{p}_j$ manage to be close to the raw estimates y_j/m_j, there is binomial overdispersion; the $(y_j - m_j\widehat{p}_j)^2$ tend to be bigger than $m_j\widehat{p}_j(1-\widehat{p}_j)$.

(h) The modelling and analyses above rest on the binomial assumption, for each y_j and position x_j, which means relying on all shots having the very same success probability p_j. This is not entirely realistic, as seen via the model residuals in the previous point. This invites modelling an extra little layer of uncertainty in p_j around some central value $p_{j,0}$. A natural way for this is the Beta-binomial setup, see Ex. 1.25, with $y_j \,|\, p_j \sim \mathrm{binom}(m_j, p_j)$, but $p_j \sim \mathrm{Beta}(cp_{j,0}, c(1-p_{j,0}))$. In other words, we use one of the parametric models for $p_{j,0}$, but then estimate the additional variability via c. Show that the log-likelihood becomes

$$\sum_{j=1}^{k} \log\Big[\frac{\Gamma(c)}{\Gamma(cp_{0,j}(\theta))\,\Gamma(c(1-p_{0,j}(\theta)))}\frac{\Gamma(cp_{0,j}(\theta)+y_j)\,\Gamma(c(1-p_{0,j}(\theta))+m_j-y_j)}{\Gamma(m_j+c)}\Big].$$

Analyse this binomial overdispersion model, for the case of the simple geometric $p_j(\sigma) = 2\,\Phi(d_j/\sigma)-1$ model, by maximising the log-likelihood over (σ, c). Show that the log-likelihood maximum increases very significantly, from the one-parameter binomial based to the two-oarameter overdispersion model. This does not necessarily influence the estimated overall curve $p(x)$, but aims at describing the probability mechanisms much better, e.g. for prediction. (xx can ask for a figure to complent the first. instead of binomial based 90 percent intervals around y_j/m_j, give 90 percent intervals induced by the estimated beta-binomial model. xx) (xx we might give a little table, for models 1, 2, 3, 4, then 5, 6, 7, with logLmax, AIC, Q. xx)

	dim	logLmax	aic	Q	
1	2	-3020.155	-6044.309	258.968	logistic order 1
2	3	-2929.041	-5864.082	74.356	logistic order 2
3	4	-2912.873	-5833.745	40.743	logistic order 3
4	5	-2904.889	-5819.778	25.288	logistic order 4
5	6	-2904.365	-5820.731	24.337	logistic order 5
6	1	-2922.639	-5847.279	62.436	one-para geometric
7	2	-2911.589	-5827.178	36.741	geometric with extra
8	2	-2910.926	-5825.852	71.528	two-para Beta-binomial

(i) (xx can at least point to a Tore-Nils thing with estimation and giving confidene curve for x_0, the distance at which the probability of success is e.g. 0.75. the point is to showcase than even with a complicated $x_0 = x_0(a,b)$ formula, we may crank through, find deviance and then $\mathrm{cc}(x_0)$. xx)

Story viii.18 *Who wins? Computing probabilities as a function match time.* (xx this ought to be good stuff, composed the day after Nor-Den 27-25 November 2022. data1: time points for goals; data2: 117 match results, for correlated Poissons. need to calibrate with what we write elsewhere on Poisson processes. xx) Watching a handball match, the two teams have at time t scored $A(t)$ and $B(t)$ goals. In our continuous excitement we speculate perhaps perplexidly about the final outcome, i.e. $A(60) = A(t) + A'$ and

$B(60) = B(t) + B'$. Below we find the dynamically evolving probabilities for team A winning and for team B winning, as a function of time t; see Figure viii.29 for how these dramatically panned out for the women's European Championship 2022, with Denmark taking an early lead but Norway prevailing in the end.

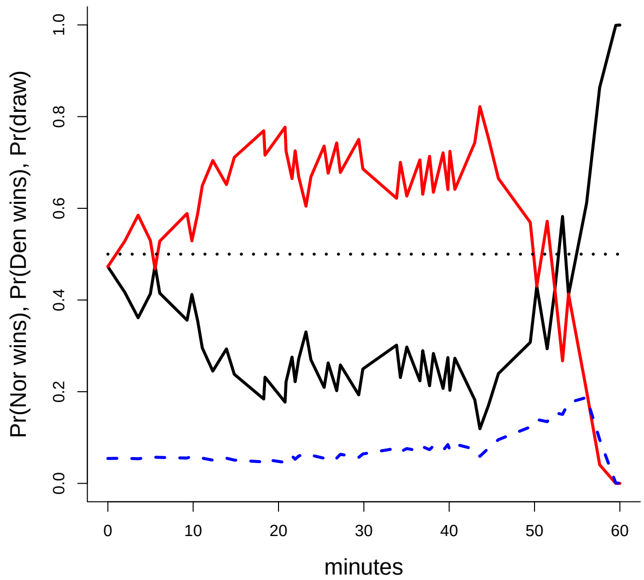

Figure viii.29: *Probabilities that Norway will win, that Denmark will win, or that it will be a draw, as a function of match time, in the women's European Finals November 2022.*

(a) Assume the teams are about equally strong, and that goals are scored according to independent Poisson processes with rate $\lambda = 27.00$; this is close to the average number of goals scored by teams in women's Olympic, World, European tournaments. What is the pre-match probability of a draw? What is the most likely result at halftime?

(b) Show that the relevant probabilities, at time t during the match, where $A(t)$ and $B(t)$ have just been observed, are

$$\begin{aligned} p_A(t) &= P(A(t) + A' > B(t) + B') = P(A' - B' > B(t) - A(t)), \\ p_B(t) &= P(A(t) + A' < B(t) + B') = P(A' - B' < B(t) - A(t)), \\ p_D(t) &= P(A(t) + A' = B(t) + B') = P(A' - B' = B(t) - A(t)), \end{aligned}$$

in which A' and B' are independent Poissons with means $\lambda(60 - t)$. Find formulae for these probabilities, in terms of sums. Then compute and plot these, from the beginning to the end of the match, for the case of Norway–Denmark in the European 2022 finals; produce indeed a version of Figure viii.29.

(c) When $A(t)$ is a Poisson process with constant rate λ, show that $A(t)$ given $A(60) = m$ is a binomial $(m, t/60)$. Show more generally that with match time $[0, 60]$ split into disjoint intervals $C_1, \ldots, C_k$, with lengths $\ell_1, \ldots, \ell_k$, then the goal counts $(A(C_1), \ldots, A(C_k))$ for these intervals have a multinomial distribution $m, (\ell_1/60, \ldots, \ell_k/60)$. For a finished handball match, having observed the $A(t)$ process, explain how a Pearson chi-squared test (see Story vii.1) can be put up to test the constant rate Poisson modelling hypothesis. Carry out such a test, for Norway and for Denmark, in the European 2022 finale, counting the number of goals scored in the six time windows $[0, 10], \ldots, [50, 60]$.

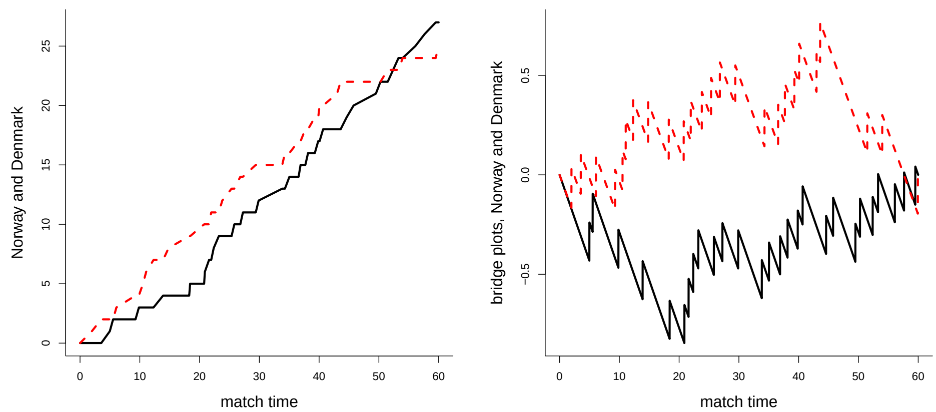

Figure viii.30: *Two summary views of the Norway–Denmark European finals 2022. Left panel: the number of goals scored, as a function of match time. Right panel: bridge plots, to assess the Poisson constant rate hypohtesis. These will under that modelling assumption be inside* ± 1.358 *in 95 percent of all cases.*

(d) Another view of the scoring-of-goals processes is as follows. If team A has scored $A(60) = m$ goals, show that the time points $T_1 < \cdots < T_m$ at which goals have been scored follows the joint density $m!/60^m$ on the set $t_1 < \cdots < t_m$. Explain that this also means that $(T_1, \ldots, T_m)$ behaves as an ordered sample from the uniform distribution on $[0, 60]$. Use this again to argue that with $F_m(t)$ the empirical c.d.f. for the data, the process $Z_m(t) = m^{1/2}\{F_m(t) - t/60\}$ is close in distribution to that of $Z(t) = W^0(t/60)$, with W^0 a Brownian bridge; see Ex. 9.9. To check the constant rate Poisson process assumption, therefore, compute and display these Z_m processes, for Norway and Denmark in their 27–25 European finals match. Construct a version of Figure viii.30.

(e) There is perhaps a feeling among spectators and handball followers that the top teams to a high degree follow each other during matches; the final scores $A(60), B(60)$ are often close. This motivates Poisson models with positive dependence. For a parameter $a \in [0, 1]$, consider $A = C + A_0$ and $B = C + B_0$, where A_0, B_0, C are independent Poissons, with parameters $a\lambda, a(1 - \lambda), a(1 - \lambda)$, and where only (A, B) are observed.

Show that A and B are Pois(λ), but dependent, with correlation a. Show that

$$P(A=a,B=b)=\sum_{c\le\min(a,b)} g(c,a\lambda)\,g(a-c,(1-a)\lambda)\,g(b-c,(1-a)\lambda),$$

in terms of the point mass function $g(x,\theta)$ for the Poisson with parameter θ. Now access the table of results (y_1,y_2) from 117 top-level women's handball matches (xx described in dataoverview xx). Plot the differences y_1-y_2 divided by standard deviation to argue that matches 33 (Norway vs. Slovenia, 41-18) and 54 (Greece vs. China, 13-33) are clear outliers, and work then with the resulting cleaned table of 115 match results. Programme and graph the profiled log-likelihood function for the a parameter, say $\ell_{\rm prof}(a)=\max_{{\rm all}\ \lambda}\{\ell(a,\lambda)\}$. Estimate the correlation parameter a, and construct a confidence curve ${\rm cc}(a)$. This should give the ML estimate $\widehat a=0.218$, with 95 percent interval $[0.012,0.390]$. Thus handball matches at the top international level are positively correlated Poisson processes.

(f) (xx round off. more sophisticated plots for $p_A(t),p_B(t),p_D(t)$, using this dependence model. probably not very different. xx)

Story viii.19 *The turn-around operation: from 0-2 to 3-2.* (xx nils ranting a bit before deciding whether this is an ok story or whether things should be given to exercises, Poisson and waiting times with integrals etc. Briefly point to Belgium vs. Japan, with Japan leading 2-0, but Belgium able to come back and win 3-2, in the round-of-16 match World Cup 2018, 'overcoming a curse', according to media. xx) Football is a simple game: 22 men chase a ball for 90 minutes and, in the end, the Germans win. Consider a match between two essentially equally strong teams A and B. The score at time t is $(X(t),Y(t))$, with independent Poisson processes with the same rate λ. In a detailed story in Claeskens and Hjort (2008b, Ch. 6), analysing 254 matches to see how FIFA ranking scores may influence these Poisson rates, 627 goals were scored, which means an average of 2.468 goals per match, which we here translate to a common rate of $\lambda=1.234/90$ per minute, up to match length $T=90$ minutes (sometimes extended with a few extra minutes for so-called injury time). This little story aims at assessing how small the probability is, for experiencing a match with first has 0-2 and then is turned around to a 3-2 or even better.

(a) What is the probability that the score is still 0-0 after five minutes, and after time t? Plot this probability, for $t\in[0,T]$.

(b) What is the probability that B some time during the match will be leading 2-0 over A? Show that this is

$$p_0=\int_0^T g_2(s,\lambda)\exp(-\lambda s)\,{\rm d}s,$$

with $g_2(s,\lambda)=\lambda^2 s\exp(-\lambda s)$ the Gam$(2,\lambda)$ density, a sum of two Expo(λ). Carry out the integration to find $p_0=(1/4)\{1-(1+2\lambda T)\exp(-2\lambda T)\}$. This is 0.176. Argue that the frequency of matches where a 2-0 lead will occur, some time during the event, is $2p_0=0.353$. – Note that if the teams had been allowed to play on, with T increasing beyond the 90 minutes, the p_0 tends to 1/4, the chance that when observing two independent undisturbed Poisson processes $X(t)$ and $Y(t)$ over time, with the same intensity, the two first events will occur in the X process.

(c) Show that in games where team A experiences a 0-2 situation against team B, the random timepoint S where this occurs has probability density,

$$h_2(s,\lambda) = g_2(s,\lambda)\exp(-\lambda s)/p_0 = \frac{\lambda^2 s\exp(-2\lambda s)}{(1/4)\{1-(1+2\lambda T)\exp(-2\lambda T)\}}$$

for $s \in [0,T]$. Show that it peaks at $s_0 = 1/(2\lambda)$, here in about 36 and a half minute. Construct a figure showing this.

(d) Team B now leads 2-0, at time point S, and team A better hurry up. Show that the probability that team A will actually accomplish the 3-2 feat, given that there is a 0-2 time point in the first place, may be expressed as

$$\begin{aligned} p^* &= \int_0^T P(\text{hurry up from } s \text{ to } T \,|\, S = s)h_2(s,\lambda)\,\mathrm{d}s \\ &= \int_0^T G_3(T-s,\lambda)\exp\{-\lambda(T-s)\}h_2(s,\lambda)\,\mathrm{d}s, \end{aligned}$$

with $G_3(T-s,\lambda)$ the cumulative gamma $(3,\lambda)$ distribution function for the sum of three exponential waiting times, evaluated at match time minus s, and $\exp\{-\lambda(T-s)\}$ the probability that team B doesn't score during this remaining time.

(e) Show first that the probability that a Poisson with mean $\lambda(T-s)$ is less than or equal to 2 is $Q(2,\lambda(T-s)) = \exp\{-\lambda(T-s)\}\{1+\lambda(T-s)+\frac{1}{2}\lambda^2(T-s)^2\}$. Use this to show that the probabiity of experiencing a 0-2 followed by a 3-2 operation is

$$p^* = \int_0^T \{1-Q(2,\lambda(T-s))\}\exp\{-\lambda(T-s)\}h_2(s,\lambda)\,\mathrm{d}s.$$

(xx check. we find $p^* = 0.014$. xx)

(f) When Belgium see Genki Haraguchi and Takashi Inui score, in the 48th and 52nd minute, they ought to be forgiven for being merely moderately interested in the overall $p^* = 0.014$, but more concerned with the imminent chance that they can still manage, given that they face 0-2 after precisely $s = 52$ minutes. Show that this probability is

$$p^*(s) = G_3(T-s,\lambda)\exp\{-\lambda(T-s)\}.$$

Plot that probability curve, as a function of 0-2 occurrence time s. At time $s = 0$, this is the chance of winning 3-0 or more, namely 3.7 percent, and after 52 minutes, it is about 1.0 percent.

(g) (xx could round off with one or two supplementing questions, using the fifa scores database nils built up for gerda-nils Ch6, to make it statistical too. there $\lambda_{i,j} = h(x_i/x_j,\theta)$, in terms of pre-tournament fifa scores. point finally to magne aldrin and anders løland, their prediction machines at NR, during tournaments. xx)

Story viii.20 *Bolt from heaven.* (xx to come. 195 sub-Hary races 2000 to 2007. how surprised ought we to have been, when Usain Bolt ran 9.72, in May 2008? data in 2.B. mention Hary (1960), Bolt (2013). See Figure viii.31. xx) On 31 May 2008, Usain Bolt burst upon us, with his first world record, 9.72. How surprised were we? To approach that question, along with those which followed as the Bolt From Heaven did 9.69 (August 2008) and then 9.58 (August 2009), we compare the 9.72 performance with the $n = 195$ sub-10.00 races of 2000–2007; these are given in data description 2.B (xx check that data description says these are bona fide races; dopers pushed out of dataset xx).

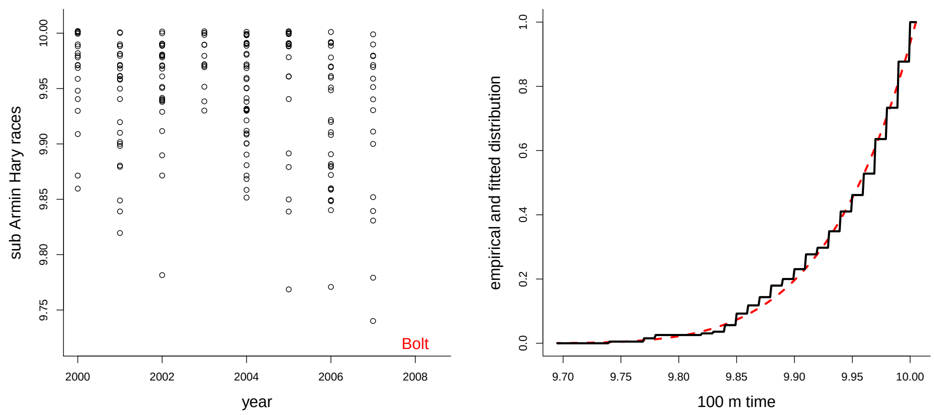

Figure viii.31: Left panel: displayed are all the 195 sub-Hary races achieved during the eight seasons 2000 to 2007, along with the new word record 9.72 ran by Bolt in May 2018. Right panel: the empirical distribution function (black, rugged) for these 195 races, along with the fitted two-parameter distribution from extreme values theory.

(a) To readily access a body of literature on extreme values theory (xx references here, embrechts xx), we transform these race times r_i to $y_i = 10.005 - r_i$. Such theory predicts that the y_i should follow the distribution

$$G(y, a, \sigma) = 1 - (1 - ay/\sigma)^{1/a} \quad \text{for } y > 0,$$

for parameters (a, σ). Show that the log-likelihood function takes the form

$$\ell(a, \sigma) = \sum_{i=1}^{n} \{-\log \sigma + (1/a - 1)\log(1 - ay_i/\sigma)\}.$$

Fit the model, which should give ML estimates $(\widehat{a}, \widehat{\sigma}) = (0.1821, 0.0701)$, and produce a version of Figure viii.31. As we see, the model works very well.

(b) For a season with N top races, below the Hary threshold 10.00, consider $p = p(a, \sigma, N) = P(\max(Y_1', \ldots, Y_N') \geq w)$. With N being a Poisson λ, show that

$$p = p(a, \sigma) = 1 - \exp\{-\lambda(1 - aw/\sigma)^{1/a}\}.$$

(c) Use $\lambda = 198/8 = 24.375$, the rate of top races per year. For each threshold w we may estimate $p(a, \sigma)$. With $w = 10.005 - 9.72 = 0.285$, for 31 May 2008, compute $\widehat{p} = 0.035$; the estimated probabiity of seeing a 9.72 better in the course of 2008, as judged from 1 January 2008, was 3.5 percent.

(d) (xx then about standard error. and CD for $p = p(a, \sigma)$. xx)

(e) xx

Story viii.21 *Monetary pre-WW2 US policy and its effects.* C.A. Sims won the Sveriges Riksbank Prize in Economic Sciences im Memory of Alfred Nobel for 2011. In his prize acceptance lecture he related his own contributions to the fundamental statistical economics theory work of Trygve Haavelmo, winner of the same Sveriges Riksbank Prize for 1989, see e.g. Haavelmo (1943), and used the occasion to analyse a certain dataset, given below, concerning US macroeconomics for the pre-WW2 years 1929–1940. Specifically, the variables examined amount to the multivariate time series of *consumption* (C), *investment* (I), *government spending* (G). From basic economics theory he constructed a certain vector time series model, with six regression coefficients and three variance parameters. For this Stockholm occasion, Sims (2012a,b) advocated and showcased the use of Bayesian methodology, setting up priors for the nine parameters followed by MCMC computation, assessment, interpretation of posterior summaries. Below we re-analyse the same data, using the very same model and with the same constraints on its nine parameters; we use however the frequentist methodology of Ch. 7, and derive confidence distributions for the crucial parameters. These clash significantly with Sims's findings, and we shall see how and why below. (xx we point to Ex. 7.14 and 7.15, and also to Ex. 7.11. xx)

often inaccurately called the Nobel Prize of Economics

year	C	I	G
1929	736.3	101.4	146.5
1930	696.8	67.6	161.4
1931	674.9	42.5	168.2
1932	614.4	12.8	162.6
1933	600.8	18.9	157.2
1934	643.7	34.1	177.3
1935	683.0	63.1	182.2
1936	752.5	80.9	212.6
1937	780.4	101.1	203.6
1938	767.8	66.8	219.3
1939	810.7	85.9	238.6
1940	752.7	119.7	245.3

(a) Consider in general terms a model for vectors $y_1, \ldots, y_n$, of dimension say p, progressing in time in a one-step memory fashion, via

$$H_0 y_t = c + H_1 y_{t-1} + \varepsilon_t \quad \text{for } t = 1, \ldots, n,$$

with the ε_t being i.i.d. from some error distribution density f_0. Here H_0 and H_1 are $p \times p$ matrices, perhaps constructed via regression parameters, with H_0 being invertible; also, there is a given start observation y_0 from which the process then develops. Using

$y_t = H_0^{-1} z_t$, with $z_t = c + H_1 y_{t-1} + \varepsilon_t$ given y_{t-1} having density $f_t(z_t \,|\, y_{t-1})$, say, show that the joint probability distribution for $(Y_1, \ldots, Y_n)$, given the start y_0, can be written

$$L = \prod_{t=1}^n g(y_t \,|\, y_{t-1}) = \prod_{t=1}^n f_t(H_0 y_t \,|\, y_{t-1}) |H_0| = \prod_{t=1}^n f_0(H_0 y_t - c - H_1 y_{t-1}) |H_0|.$$

For the case where the $\varepsilon_t \sim \mathrm{N}_p(0, D)$, with a diagonal $\sigma_1^2, \ldots, \sigma_p^2$ variance structure, show that this leads to log-likelihood

$$\begin{aligned} \ell &= \sum_{t=1}^n \Big[\log |H_0| + \sum_{j=1}^p \{-\log \sigma_j - \tfrac12 \widetilde\varepsilon_{t,j}^2/\sigma_j^2\}\Big] \\ &= n \log |H_0| + \sum_{j=1}^p \{-n \log \sigma_j - \tfrac12 \sum_{t=1}^n \widetilde\varepsilon_{t,j}^2/\sigma_j^2\}, \end{aligned}$$

where $\widetilde\varepsilon_t = H_0 y_t - c - H_1 y_{t-1}$. Supposing regression coefficients α go into the c and the H_0 and H_1 matrices, show that the log-likelihood profile, maximising over $\sigma_1, \ldots, \sigma_p$, becomes

$$\ell_{\rm prof}(\theta) = n \log |H_0(\alpha)| + \sum_{j=1}^p \{-n \log \widehat\sigma_j(\alpha) - \tfrac12 n\}, \quad \text{where } \widehat\sigma_j(\alpha)^2 = Q_j(\alpha)/n,$$

writing $Q_j(\alpha) = \sum_{t=1}^n \widetilde\varepsilon_{t,j}(\alpha)^2$. This reduces the log-likelihood optimisation problem from dimension $p_0 + p$ to dimension $p_0 = \dim(\alpha)$.

(b) (xx let's see. xx) The vector autoregressive model used in Sims (2012b) takes

$$\begin{aligned} C_t &= \beta_0 + \beta_1 (C_t + I_t + G_t) + \sigma_C Z_{1,t}, \\ I_t &= \theta_0 + \theta_1 (C_t - C_{t-1}) + \sigma_I Z_{2,t}, \\ G_t &= \gamma_0 + \gamma_1 G_{t-1} + \sigma_G Z_{3,t}, \end{aligned}$$

with the error terms $Z_{j,t}$ being i.i.d. standard normal. With $Y_t = (C_t, I_t, G_t)^{\rm t}$, show that this can be translated to the general form above, with

$$H_0 = \begin{pmatrix} 1-\beta_1, & -\beta_1, & -\beta_1 \\ -\theta_1, & 1, & 0 \\ 0, & 0, & 1 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0, & 0, & 0 \\ -\theta_1, & 0, & 0 \\ 0, & 0, & \gamma_1 \end{pmatrix}, \quad c = \begin{pmatrix} \beta_0 \\ \theta_0 \\ \gamma_0 \end{pmatrix}.$$

This leads to a clearly defined log-likelihood function of six regression coefficients and three standard deviation parameters. Show that $|H_0| = 1 - \beta_1(1 + \theta_1)$ here, and it is part of the prior constraints of the parameters that this determinant must be positive. Programme this log-likelihood function and find its optimisers, i.e. the unrestricted ML estimates. (xx For the unconstrained ML, nils finds the following. with approximate normality for $\widehat\theta_1$, there is a pointmass 0.904 at zero. Sims says $\theta_1 \ge 0$, $\gamma_1 \le 1.03$, $1 - \beta_1(1+\theta_1) > 0$. mention that $(\widehat\beta_0, \widehat\beta_1)$ as well as $(\widehat\gamma_0, \widehat\gamma_1)$ have strong negative correlations, about -0.99, so the model is not well parametrised. this is seen also for the mcmc. – Attention is now on θ_1, which Sims explains is a priori nonnegative. xx)

```
   ML       se           sims reports
201.5721 33.0779  beta0   166.0
  0.5246  0.0341  beta1     0.566
 63.8808 13.1022  theta0   63.0
 -0.5664  0.4347  theta1    0.0
 10.7936 23.5020  gamma0   10.7
  0.9902  0.1259  gamma1    0.991
```

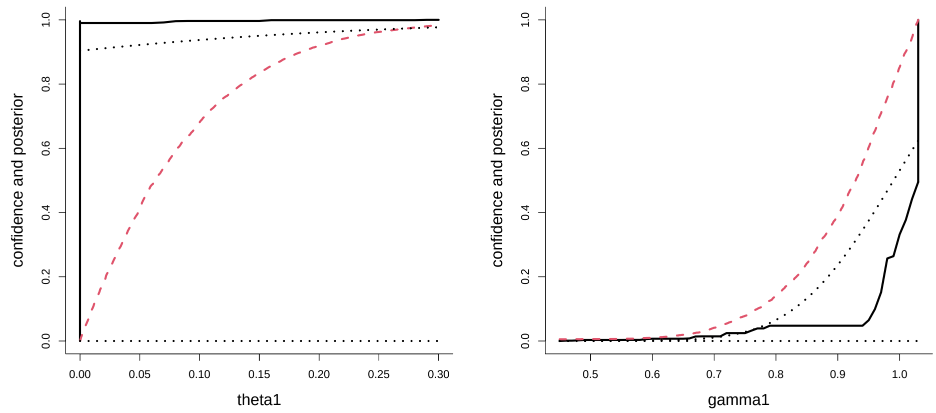

Figure viii.32: *(xx polish, with the last details, com34* of nilswork23. xx) Left panel, for the crucial θ_1 parameter: posterior cumulative, with the Simps prior (slanted), and 95 percent interval xxxx; the normal approximation CD (dotted); and the more carefully computed CD using t-bootstrapping (full curve). The confidence pointmass at $\theta_1 = 0$ is 0.989, whereas the Bayesian posterior does not detect that θ_1 very likely is zero. Right panel: similarly, for the γ_1 parameter, where Sims uses the upper bound 1.03. The posterior distribution (slanted) does not detect that there is a considerable probability that $\gamma_1 = 1.03$; the CD pointmass there is 0.501. (xx more, round off; see com31* and com34* of nilswork23. xx)*

(c) With $\alpha = (\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1)^{\rm t}$ the regression coefficients and $\sigma = (\sigma_1, \sigma_2, \sigma_3)^{\rm t}$, having independent priors π_a and π_s, say, show that the posterior distribution becomes

$$\pi(\alpha, \sigma \,|\, \text{data}) \propto \pi_a(\alpha)\pi_s(\sigma) \exp\{n \log |H_0(\alpha)|\} \prod_{j=1}^{3} (1/\sigma_j)^n \exp\{-\tfrac{1}{2} Q_j(\alpha)/\sigma_j^2\}$$

With independent noninformative priors $1/\sigma_j$ for the σ_j, show that

$$\pi(\alpha \,|\, \text{data}) \propto \pi_a(\alpha) \exp\{n \log |H_0(\alpha)|\} \prod_{j=1}^{3} 1/Q_j(\alpha)^{n/2}.$$

With flat priors on α, show that maximising this posterior density, over the regression coefficients α, is equivalent to finding the ML estimates. Then set up an MCMC to simulate posterior realisations of $\alpha = (\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1)$, using the prior Sims advocates here; it is flat, but with built-in constraints $\theta_1 \ge 0$, $\gamma_1 \in [0, 1.03]$, and $1 - \beta_1(1+\theta_1) > 0$. Of particular interest is the posterior $\pi(\theta_1 \,|\, \mathrm{data})$, which can then be read off from the MCMC. Construct a version of viii.32, left panel, with the c.d.f. for θ_1, alongside the confidence distribution $C(\theta_1) = \Phi((\theta_1 - \widehat\theta_1)/\widehat\kappa_1)$. (xx nils needs a bit more care with CD for θ_1. simulate lots of Sims datasets at positions $\widehat\alpha$, but with θ_1 on a little grid. need to verify that $t = (\widehat\theta_1 - \theta_1)/\widehat\kappa_1$ is approximately a standard normal. xx)

(d) (xx yet other points can be worked with. round off. we use t-bootstrapping methods of Ex. 7.11 for more accurate CDs for θ_1 and for γ_1. simulations are a bit expensive, so we use the isotonic repair trick of Ex. 7.4. perhaps one more parameter. computationally this is moderately costly. push the view that a quite likely submodel actually holds, a significant simplification of the original nine-parameter model:

$$\begin{aligned} C_t &= \beta_0 + \beta_1(C_t + I_t + G_t) + \sigma_C Z_{1,t}, \\ I_t &= \theta_0 + \sigma_I Z_{2,t}, \\ G_t &= \gamma_0 + G_{t-1} + \sigma_G Z_{3,t}. \end{aligned}$$

interpret this simpler model. In the pre-war US economy, investment I_t was independent of consumption and of its changes over time, and government spending acted like a random walk. round off. xx)

(e) (xx to be moved from here to solutions section. we not in passing that Sims is a bit sloppy with the log-likelihood things. anyway, this is to verify the basic vector autoregressive structure, with the H_0 and H_1 matrices. xx)

$$\begin{pmatrix} 1-\beta_1, & -\beta_1, & -\beta_1 \\ -\theta_1, & 1, & 0 \\ 0, & 0, & 1 \end{pmatrix} \begin{pmatrix} C_t \\ I_t \\ G_t \end{pmatrix} - \begin{pmatrix} 0, & 0, & 0 \\ -\theta_1, & 0, & 0 \\ 0, & 0, & \gamma_1 \end{pmatrix} \begin{pmatrix} C_{t-1} \\ I_{t-1} \\ G_{t-1} \end{pmatrix} = \begin{pmatrix} C_t - \beta_1(C_t + I_t + G_t) \\ I_t - \theta_1(C_t - C_{t-1}) \\ G_t - \gamma_1 G_{t-1} \end{pmatrix}.$$

Part III

Solutions

Part IV

Appendix

References

Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, 2:951–972.

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and Event History Analysis. A process point of view*. Springer, New York.

Aalen, O. O. and Gjessing, H. K. (2004). Survival models based on the Ornstein–Uhlenbeck process. *Lifetime Data Analysis*, 10:407–423.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press, Princeton.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, Berlin.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24:3–30.

Ashworth, S., Berry, C. R., and de Mesquita, E. B. (2021). *Theory and Credibility: Integrating Theoretical and Empirical Social Science*. Princeton University Press, Princeton.

Aursnes, I., Tvete, I. F., Gåsemyr, J., and Natvig, B. (2005). Suicide attempts in clinical trials with paroxetine randomised against placebo. *BMC Medicine*, xx:1–5.

Aursnes, I., Tvete, I. F., Gåsemyr, J., and Natvig, B. (2006). Even more suicide attempts in clinical trials with paroxetine randomised against placebo. *BMC Psychiatry*, xx:1–3.

Barber, R. F., Candes, E. J., Ramdas, A., and Tibshirani, R. J. (2022). Conformal prediction beyond exchangeability. *arXiv preprint arXiv:2202.13415*.

Barfort, S., Klemmensen, R., and Larsen, E. G. (2020). Longevity returns to political office. *Political Science Research and Methods*, 9:658–664.

Bartolucci, F. and Lupparelli, M. (2008). Focused Information Criterion for capture-recapture models for closed populations. *Scandinavian Journal of Statistis*, 9:658–664.

Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998). Robust and efficient estimation by minimising a densithy power divergence. *Biometrika*, 85:549–559.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Chicago University Press, Chicago.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Blower, J. G., Cook, L. M., and Bishop, J. A. (1981). *Estimating the Size of Animal Populations*. Allen & Unwin, Kondon.

Boitsov, V. D., Karsakov, A. L., and Trofimov, A. G. (2012). Atlantic water temperature and climate in the barents sea, 2000–2009. *ICES Journal of Marine Science*, 69:833–840.

Bolt, U. (2013). *Faster Than Lightning: My Autobiography.* HarperSport, London.

Borgan, Ø., Fiaccone, R. L., Henderson, R., and Barreto, M. L. (2007). Dynamic analysis of recurrent event data with missing observations, with application to infant diarrhoea in brazil. *Scandinavian Journal of Statistics*, 34:53–69.

Borgan, Ø. and Keilman, N. (2019). Do Japanese and Italian women live longer than women in Scandinavia? *European Journal of Population*, 35:87–99.

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71:353–360.

Breiman, L. (2001). Statistical modeling: The two cultures [with comments and a rejoinder by the author]. *Statistical Science*, 16:199–231.

Brown, L. D. (1986). Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. *Lecture Notes-Monograph Series*, 9:i–279.

Brunborg, H., Lyngstad, T. H., and Urdal, H. (2003). Accounting for genocide: How many were killed in Srebrenica? *European Journal of Population*, 19:229–248.

Candès, E. J., Lei, L., and Ren, Z. (2021). Conformalized survival analysis. *arXiv preprint arXiv:2103.09763.*

Card, D. and Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review*, 84:772–793.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *American Statistician*, 46:167–174.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). Distributional conformal prediction. *Proceedings of the National Academy of Sciences*, 118:e2107794118.

Claeskens, G. and Hjort, N. L. (2008a). Minimizing average risk in regression. *Econometric Theory*, 24:493–527.

Claeskens, G. and Hjort, N. L. (2008b). *Model Selection and Model Averaging.* Cambridge University Press, Cambridge.

Clauset, A. (2018). Trends and fluctuations in the severity of interstate wars. *Science Advances*, 4:1–9.

Clauset, A. (2020). On the frequency and severity of interstate wars. In Gleditsch, N. P., editor, *Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences*, pages 113–128. Springer, Berlin.

Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws. *Journal of the American Statistical Association*, 70:698–705.

Cox, D. R. (1958). Some problems with statistical inference. *The Annals of Mathematical Statistics*, 29:357–372.

Cox, D. R. (1972). Regression models and life-tables [with discussion]. *Journal of the Royal Statistical Society: Series B*, 34:187–202.

Cox, D. R. and Brandwood, L. (1959). On a discriminatory problem connected with the worlks of Plato. *Journal of the Royal Statistical Society Series B*, 21:195–200.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes.* Chapman & Hall, London.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

Cramér, H. (1976). Half a century with probability theory: some personal reflections. *Annals of Probability Theory*, 4:509–546.

Cunen, C. (2015). Mortality and Nobility in the Wars of the Roses and Game of Thrones. *FocuStat Blog, University of Oslo*, iv.

Cunen, C., Hermansen, G. H., and Hjort, N. L. (2018). Confidence distributions for change points and regime shifts. *Journal of Statistical Planning and Inference*, 195:14–34.

Cunen, C. and Hjort, N. L. (2015). Optimal inference via confidence distributions for two-by-two tables modelled as Poisson pairs: fixed and random effects. In Nair, V., editor, *Proceedings of the 60th World Statistics Congress, ISI Rio*, pages xx–xx. Springer, Rio.

Cunen, C. and Hjort, N. L. (2022). Combining information from diverse sources: the II-CC-FF paradigm. *Scandinavian Journal of Statistics*, 49:625–656.

Cunen, C. and Hjort, N. L. (2023). Survival and event history models and methods via Gamma processes. Technical report, University of Oslo. Technical report.

Cunen, C., Hjort, N. L., and Nygård, H. M. (2020a). Statistical sightings of better angels. *Journal of Peace Research*, 57:221–234.

Cunen, C., Hjort, N. L., and Schweder, T. (2020b). Confidence in confidence distributions! *Proceedings of the Royal Society, A*, 476:1–5.

Cunen, C., Walløe, L., and Hjort, N. L. (2020c). Focused model selection for linear mixed models, with an application to whale ecology. *Annals of Applied Statistics*, 14:872–904.

De Blasi, P. and Hjort, N. L. (2007). Bayesian survival analysis in proportional hazard models with logistic relative risk. *Scandinavian Journal of Statistics*, 34:229–257.

DeGroot, M. H. (1970). *Optimal Statistical Decisions.* John Wiley & Sons, Hoboken, N.J.

Efron, B. (2023). *Exponential Families in Theory and Practice.* Cambridge University Press, Cambridge.

Efron, B. and Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, 236:119–127.

Fagerland, M., Lydersen, S., and Laake, P. (2017). *Statistical Analaysis of Contingency Tables.* Chapman and Hall/CRC, New YOrk.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman & Hall, London.

Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26:528–535.

Franklin, B. (1793). *The Autobiography of Benjamin Franklin.* Dover, New York. Reprinted from Dover, New York, 1996.

Friesinger, A. (2004). *Mein Leben, mein Sport, meine besten Fitness-Tipps.* Goldmann, Berlin.

Frigessi, A. and Hjort, N. L. (2002). Statistical methods for discontiniuous phenomena. *Journal of Nonparametric Statistics*, 14:1–5.

Galton, F. (1889). *Natural Inheritance.* Macmillan, London.

Geißler, A. (1889). Beiträge zur Frage des Geschlechts verhältnisses der Geborenen. *Zeitschrift des königlichen sächsischen statistischen Bureaus*, 35:1–24.

Gelman, A., Hill, J., and Vehtari, A. (2022). *Regression and Other Stories.* Cambridge University Press, Cambridge.

Gelman, A. and Nolan, D. (2002). A probability model for golf putting. *Teaching Statistics*, 24:93–95.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314.

Gjessing, H. K., Aalen, O. O., and Hjort, N. L. (2003). Frailty models based on Lévy processes. *Advances in Applied Probability*, 35:532–550.

Glad, I. K., Hjort, N. L., and Ushakov, N. U. (2003). Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, 30:415–427.

Gleditsch, N. P. (2020). *Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences (edited book).* Springer, Berlin.

Goudie, I. B. J. and Goudie, M. (2007). Who captures the marks for the Petersen estimator? *Journal of the Royal Statistical Society, Series A*, 170:825–839.

Gran, J. M. and Stensrud, M. J. (2022). Hva er forventet levealder? *Tidsskrift for Den norske legeforening*, page 245.

Grønneberg, S. and Hjort, N. L. (2012). On the errors committed by sequences of estimator functionals. *Mathematical Methods of Statistics*, 20:327–346.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrics*, 11:1–12.

Hall, P. (1927). The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika*, 19:240–245.

Hall, P. G. (1983). Large-sample optimality of least squares cross-validation in density estimation. *Annals of Statistics*, 11:1156–1174.

Halmos, P. R. and Savage, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *The Annals of Mathematical Statistics*, 20:225–241.

Hanche-Olsen, H. and Holden, H. (2010). The Kolmogorov-Riesz compactness theorem. *Expositiones Mathematicase*, 28:385–394.

Hary, A. (1960). *10,0.* Copress, München.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition.* Springer, New York.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 9:97–109.

Haug, K. K. (2019). Focused model selection for Markov chain models, with an application to armed conflict data. Technical report, University of Oslo. Master Thesis.

Heger, A. (2011). Jeg og jordkloden. *Dagsavisen*, Dec. 16.

Hermansen, G. H., Hjort, N. L., and Kjesbu, O. S. (2016). Modern statistical methods applied on extensive historic data: Hjort liver quality time series 1859-2012 and associated influential factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 73:273–295.

Hjort, J. (1914). *Fluctuations in the Great Fisheries of Northern Europe, Viewed in the Light of Biological Research.* Conseil Permanent International Pour l'Exploration de la Mer, Copenhagen.

Hjort, N. L. (1976). The Dirichlet Process Applied to Nonparametric Estimation Problems. Technical report, University of Tromsø, University of Tromsø. Cand. real. thesis, Universities of Tromsø and Oslo.

Hjort, N. L. (1986a). Bayes estimators and asymptotic efficiency in parametric counting process models. *Scandinaavian Journal of Statistics*, 13:63–85.

Hjort, N. L. (1986b). *Notes on the Theory of Statistical Symbol Recognition.* Norwegian Computing Centre, Oslo.

Hjort, N. L. (1988). The eccentric part of the non-central chi square. *The American Statistician*, 42:130–132.

Hjort, N. L. (1990a). Goodness of fit tests for life history data based on cumulative hazard rates. *Annals of Statistics*, 18:1221–1258.

Hjort, N. L. (1990b). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Annals of Statistics*, 18:1259–1294.

Hjort, N. L. (1992). On inference in parametric survival data models. *International Statisical Review*, xx:355–387.

Hjort, N. L. (1994). The exact amount of t-ness that the normal model can tolerate. *Journal of the American Statistical Association*, 89:665–675.

Hjort, N. L. (2007). And quiet does not flow the Don: Statistical analysis of a quarrel between Nobel laureates. In Østreng, W., editor, *Concilience*, pages 134–140. Centre for Advanced Research, Oslo.

Hjort, N. L. (2008). Discussion of P.L. Davies' article 'Approximating data'. *Journal of the Korean Statistical Society*, 37:221–225.

Hjort, N. L. (2017a). Cooling of Newborns and the Difference Between 0.244 and 0.278. *FocuStat Blog, University of Oslo*, xv.

Hjort, N. L. (2017b). The Semifinals Factor for Skiing Fast in the Finals. *FocuStat Blog, University of Oslo*, xv.

Hjort, N. L. (2018a). Overdispersed Children. *FocuStat Blog, University of Oslo*, xxi.

Hjort, N. L. (2018b). Towards a More Peaceful World [insert '!' or '?' here]. *FocuStat Blog, University of Oslo*, xvii.

Hjort, N. L. (2019a). The Magic Square of 33. *FocuStat Blog, University of Oslo*, xxi.

Hjort, N. L. (2019b). Sudoku Solving by Probability Models and Markov Chains. *FocuStat Blog, University of Oslo*, xxi.

Hjort, N. L. and Fenstad, G. (1992). On the last time and the number of times an estimator is more than ε from its target value. *The Annals of Statistics*, 20:469–489.

Hjort, N. L. and Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *The Annals of Statistics*, 23:882–904.

Hjort, N. L. and Jones, M. C. (1996). Locally parametric nonparametric density estimation. *The Annals of Statistics*, 24:1619–1647.

Hjort, N. L. and Koning, A. J. (2002). Tests for constancy of model parameters over time. *Journal of Nonparametric Statistics*, 14:113–132.

Hjort, N. L. and Lumley, T. (1993). Normalised local hazard plots. Technical report, Department of Statistics, University of Oxford, Oxford.

Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Annals of Statistics*, 37:1079–1111.

Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2018). Hybrid combinations of parametric and empirical likelihoods. *Statistica Sinica*, 27:2389–2407.

Hjort, N. L. and Petrone, S. (2007). Nonparametric quantile inference using Dirichlet processes. In Nair, V., editor, *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell Doksum*, pages 463–492. World Scientific, New Jersey.

Hjort, N. L. and Pollard, D. B. (1993). Asymptotics for minimisers of convex processes. Technical report, Department of Mathematics, University of Oslo.

Hjort, N. L. and Schweder, T. (2018). Confidence distributions and related themes: introduction to the special issue. *Journal of Statistical Planning and Inference*, 195:1–13.

Hjort, N. L. and Stoltenberg, E. A. (2021). The partly parametric and partly nonparametric additive risk model. *Lifetime Data Analysis*, 27:1–31.

Hjort, N. L. and Varin, C. (2008). ML, PL, QL in Markov chain models. *Scandinavian Journal of Statistics*, 35:64–82.

Hjort, N. L. and Walker, S. G. (2009). Quantile pyramids for Bayesian nonparametrics. *Annals of Statistics*, 37:105–131.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81:945–960.

Holum, D. (1984). *The Complete Handbook of Speed Skating.* High Peaks Cyclery, Lake Pacid.

Hosmer, D. W. and Lemeshow, S. (1999). *Applied Logistic Regression.* Wiley, New York.

Hveberg, K. (2019). *Lene din ensomhet langsomt mot min.* Aschehoug, Oslo.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences.* Cambridge University Press, Cambridge.

Inlow, M. (2010). A moment generating function proof of the Lindeberg–Lévy central limit theorem. *American Statistician*, 64:228–230.

Irwin, J. O. (1927). On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's type II. *Biometrika*, 19:225–239.

Jacod, J. and Shiryaev, A. (2013). *Limit Theorems for Stochastic Processes. Second Edition.* Springer, Berlin.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R. Second Edition.* Springer, New York.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 361–379.

Jamtveit, B., Jacobsen, A. U., and Wyller, T. B. (2018). Utvikling i andel administrativt personale i norske helseforetak. *Samfunnsøkonomen*, 6:17–21.

Jamtveit, B., Jettestuen, E., and Mathiesen, J. (2009). Scaling properties of European research units. *Proceedings of the National Academy of Sciences*, 106:13160–13163.

Jansen, D. (1994). *Full Circle.* Villard Books, New York.

Jones, M. C. (1991). The roles of ISE and MISE in density estimation. *Statistics and Probability Letters*, 12:51–56.

Jones, M. C., Hjort, N. L., Harris, I. R., and Basu, A. (2001). A comparison of related density-based minimum divergence estimators. *Biometrika*, 88:865–873.

Jullum, M. and Hjort, N. L. (2017). Parametric or nonparametric: The FIC approach. *Statistica Sinica*, 27:951–981.

Jullum, M. and Hjort, N. L. (2019). What price semiparametric Cox regression? *Lifetime Data Analysis*, 25:406–438.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux, New York.

Kahneman, D., Sibony, O., and Sunstein, C. R. (2020). *Noise: A Flaw in Human Judgment.* William Collins, London.

Kjesbu, O. S., Opdal, A. F., Korsbrekke, K., Devine, J. A., and Skjæraasen, J. E. (2014). Making use of Johan Hjort's 'unknown' legacy: reconstruction of a 150-year coastal time-series on northeast Arctic cod (Gadus morhua) liver data reveals long-term trends in energy allocation patterns. *ICES Journal of Marine Science*, 71:2053–2063.

Kjetsaa, G., Gustavson, S., Beckman, B., and Gil, S. (1984). *The Authorship of The Quiet Don [also published in Russian].* Solum/Humanities Press, Oslo.

Klotz, J. (1972). Markov chain clustering of births by year. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability Theory*, 4:173–185.

Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence. *Annals of Statistics*, 1:373–379.

Koehler, J. J. and Conley, C. A. (2003). The "hot hand" myth in professional basketball. *Journal of Sport and Exercise Psychology*, 25:253–259.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn Ist Ital Attuar*, 4:83–91.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer, New York.

Kusolitsch, N. (2010). Why the theorem of Scheffé should be rather called a theorem of Riesz. *Periodica Mathematica Hungarica*, 61:225–229.

Laptook, A. e. a. (2017). Effect of therapeutic hypothermia initiated after 6 hours of age on death and disability among newborns with hypoxic-ischemic encephalopathy: A randomized clinical trial. *Journal of the American Medical Association*, 318:1550–1560.

Larkey, P. D., Smith, R. A., and Kadane, J. B. (1989). It's okay to believe in the "hot hand". *Chance*, 2:22–30.

Le May Doan, C. (2002). *Going For Gold.* McClelland & Stewart Publisher, Toronto.

Lehmann, E. L. (1950). *Notes on the Theory of Estimation.* Berkeley University Press, Berkeley. Notes recorded by Colin Blyth.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111.

Leike, A. (2001). Demonstration of the exponential decay law using beer froth. *European Journal of Physics*, 23:1–21.

Lessing, D. (1997). *Walking in the Shade: Volume Two of My Autobiography, 1949 to 1962.* xx, xx.

Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225.

Lindqvist, B. H. (1978). A note on Bernoulli trials with dependence. *Scandinavian Journal of Statistics*, 5:205–208.

Loader, C. (1996). Local likelihood density estimation. *Annals of Statistics*, 67:1602–1618.

Lum, K., Price, M. E., and Banks, D. (2013). Applications of multiple systems estimation in human rights research. *American Statistician*, 24:191–200.

Markov, A. A. (1906). Распространение закона больших чисел на величины, зависящие друг от друга [Extending the law of large numbers for variables that are dependent of each other]. Известия Физико-математического общества при Казанском университете (2-я серия), 15:124–156.

Markov, A. A. (1913). Пример статистического исследования над текстом "Евгения Онегина", иллюстрирующий связь испытаний в цепь [Example of a statistical investigation illustrating the transitions in the chain for the 'Evgenii Onegin' text]. Известия Академии Наук, Санкт-Петербург (6-я серия), 7:153–162.

Marron, S. and Wand, M. P. (1992). Exact mean integrated squared error. *Annals of Statistics*, 20:712–736.

McCloskey, R. (1943). *Homer Price.* Scholastic Inc., New York.

McCullagh, P. (2002). What is a statistical model? [with discussion]. *Annals of Statistics*, 30:1225–1310.

Miller, J. B. and Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, 86:2019–2047.

Miller, J. B. and Sanjurjo, A. (2021). Is it a fallacy to believe in the hot hand in the NBA three-point contest? *European Economic Review*, 138:103771.

Mykland, P. A. and Zhang, L. (2012). The econometrics of high frequency data. In Kessler, M., Lindner, A., and Sørensen, M., editors, *Statistical Methods for Stochastic Differential Equations*, pages 109–190. CRC Press.

Mykland, P. A., Zhang, L., and Chen, D. (2019). The algebra of two scales estimation, and the S-TSRV: High frequency estimation that is robust to sampling times. *Journal of Econometrics*, 208:101–119.

Neyman, J. and Pearson, E. (1933). On the problem of the most efficient statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, 68:289–337.

Normand, S.-L. T. (1999). Tutorial in biostatistics: Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18:321–359.

O'Neill, B. (2014). Some useful moment results in sampling problems. *American Statistician*, A 231:282–296.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series*, 5(302):157–175.

Pearson, K. (1902). On the change in expectation of life in man during a period of circa 2000 years. *Biometrika*, 1:261–264.

Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Limfjord from the German Sea. *Report of the Danish Biological Station*, 6:5–84.

Peterson, A. V. (1975). Nonparametric estimation in the competing risks problem. Technical report, Department of Statistics, Stanford University.

Phadia, E. G. (1973). Minimax estimation of a cumulative distribution function. *Annals of Statistis*, 6:1149–1157.

Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined*. Viking Books, Toronto.

Price, R. M. and Bonett, D. G. (2001). Estimating the variance of the sample median. *Journal of Statistical Computation and Simulation*, 68:xx–xx.

Price, R. M. and Bonett, D. G. (2002). Distribution-free confidence intervals for difference and ratio of medians. *Journal of Statistical Computation and Simulation*, 72:xx–xx.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletins of the Calcutta Mathematical Society*, pages 81–91.

Reeves, R. V. (2022a). *Of Boys and Men: Why the Modern Male is Struggling, Why it Matters, and What to Do About It*. Brookings Institution Press, Washington, D.C.

Reeves, R. V. (2022b). Redshirt the boys. *The Atlantic*, October.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Royden, H. L. and Fitzpatrick, P. M. (2010). *Real Analysis [4th ed.]*. Pearson Education Asia, Beijin.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimation. *Scandinavian Journal of Statistics*, 9:65–78.

Rydén, J. (2020). On features of fugue subjects: A comparison of J.S. Bach and later composers. *Journal of Mathematics and Music*, pages 1–20.

Saleh, J. H. (2019). Statistical reliability analysis for a most dangerous occupation: Roman emperor. *Palgrave Communication*, 5:1–7.

Sanathanan, L. (1972). Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, 43:142–1542.

Scheffé, H. (1947). A useful convergence theorem for probability distributions. *Annals of Mathematical Statistics*, 18:434–438.

Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.

Schervish, M. J. (1995). *Theory of Statistics*. Springer, New York.

Schömig, A., Mehili, J., de Waha, A., Seyfarth, M., Pahce, J., and Kastrati, A. (2008). A meta-analysis of 17 randomized trials of a percutaneous coronary intervention-based strategy in patients with stable coronary artery disease. *Journal of the American College of Cardiology*, 52:894–904.

Schweder, T. (1980). Scandinavian statistics, some early lines of development. *Scandinavian Journal of Statistics*, 7:113–129.

Schweder, T. (1999). Early statistics in the Nordic countries – when did the Scandinavians slip behind the British? *Bulletin of the International Statistical Institute*, 58:1–4.

Schweder, T. (2017). Bayesian Analysis: Always and Everywhere? *FocuStat Blog, University of Oslo*, iii.

Schweder, T. and Hjort, N. L. (2016). *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*. Cambridge University Press, Cambridge.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, London.

Shafer, G. and Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9.

Shao, J. (1991). Second-order differentiability and jackknife. *Statistica Sinica*, 1:185–202.

Shumway, R. H. and Stoffer, D. S. (2016). *Time Seires Analysis and Its Applications [4th ed.]*. Springer, Heidelberg.

Silver, N. (2012). *The Signal and the Noise: Why so Many Predictions Fail, but Some Don't*. Penguin.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Simpson, R. J. S. and Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3:1243–1246.

Sims, C. A. (2012a). Appendix: inference for the Haavelmo model. Technical report, Puplic Policy & Finance, Princeton University, Princeton, NJ.

Sims, C. A. (2012b). Statistical modeling of monetary policy and its effects [Sveriges Riksbank Prize in Memory of Alfred Nobel lecture]. *American Economic Review*, xx:1–22.

Singh, K., Xie, M., and Strawderman, W. E. (2005). Combining information from independent sources through confidence distributions. *Annals of Statistics*, 33:159–183.

Slud, E. (1989). Clipped Gaussian processes are never M-step Markov. *Journal of Multivariate Analysis*, 29:1–14.

Smith, T. D. (1994). *Scaling Fisheries: The Science of Measuring the Effects of Fishing 18551955*. Cambridge University Press, Cambridge.

Spiegelberg, W. (1901). *Aegyptische und Griechische Eigennamen aus Mumientiketten der Römischen Kaiserzeit*. Greek Inscriptions, Cairo.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 197–206.

Stigler, S. M. (1990). The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators. *Statistical Science*, 5:147–155.

Stoltenberg, E. A. (2019). An MGF proof of the Lindeberg theorem. Technical report, Department of Mathematics, University of Oslo.

Stoltenberg, E. A. and Hjort, N. L. (2021). Models and inference for on-off data via clipped Ornsten–Uhlenbeck processes. *Scandinavian Journal of Statistics*, 48:908–929.

Stout, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

Student (1908). The probable error of a mean. *Biometrika*, 6:1–25.

Swensen, A. R. (1983). A note on convergence of distributions of conditional moments. *Scandinavian Journal of Statistics*, 10:41–44.

Tibshirani, R. J., Foygel Barber, R., Candes, E., and Ramdas, A. (2019). Conformal prediction under covariate shift. *Advances in Neural Information Processing Systems*, 32.

Tversky, A. and Gilovich, T. (1989). The cold facts about the "hot hand" in basketball. *Chance*, 2:16–21.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Varian, H. R. (1975). Distributive justice, welfare economics, and the theory of fairness. *Philosophy and Public Affairs*, 4:223–247.

Voldner, B., Frøslie, K. F., Haakstad, L., Hoff, C., and Godang, K. (2008). Modifiable determinants of fetal macrosomia: role of lifestyle-related factors. *Acta Obstetricia et Gynecologica Scandinavica*, 87:423–429.

von Bahr, B. (1965). On the convergence of moments in the central limit theorem. *Annals of Mathematical Statistics*, xx:808–818.

von Bortkiewicz, L. (1898). *Das Gesetz der kleinen Zahlen.* B.G. Teubner, Berlin.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic learning in a random world.* Springer Science & Business Media, Berlin/Heidelberg.

Walløe, L., Hjort, N. L., and Thoresen, M. (2019a). Major concerns about late hypothermia study. *Acta Paediatrica*, 108:588–589.

Walløe, L., Hjort, N. L., and Thoresen, M. (2019b). Why results from Bayesian statistical analyses of clinical trials with a strong prior and small sample sizes may be misleading: The case of the NICHD Neonatal Research Network Late Hypothermia Trial. *Acta Paediatrica*, 108:1190–1191.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, 49:24–28.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D., Riffe, T., Boe, C., Bubenheim, M., Philipov, D., Shkolnikov, V., Vachon, P., C, W., and M, B. (2021). Methods protocol for the Human Mortality Database. University of California, Berkeley, US, and Max Planck Institute for Demographic Research, Rostock, Germany. `https://www.mortality.org/` [Version 6. Last revised January 26, 2021].

Wissner-Gross, Z. (2020). Can you feed the hot hand? https://fivethirtyeight.com/features/can-you-feed-the-hot-hand/. Accessed: December 12, 2020.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: a review [with discussion and a rejoinder]. *International Statistical Review*, 81:3–39.

Zabriskie, B. N., Corcoran, C., and Senchaudhuri, P. (2021). A comparison of confidence distribution approaches for rare event meta-analysis. *Statistics in Medicine*, 40:5276–5297.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81:446–451.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100:1394–1411.