Exercises and Lecture Notes
STK 4090, Spring 2020

Version 0.57, 13-i-2020

Nils Lid Hjort

Department of Mathematics, University of Oslo

Abstract

These are Exercises and Lecture Notes for the new course on Statistical Large-Sample Theory, STK 4090 (Master level) or STK 9090 (PhD level), for spring semester 2020. They are partly taken from earlier collections, from other courses of mine, but are supplemented with new ones, also from empirical processes.

1. Illustrating the Central Limit Theorem (CLT)

Consider the variable

$$Z_k = (X_1 + \cdots + X_k - k\mu)/(\sqrt{k}\sigma) = \sqrt{k}(\bar X_k - \mu)/\sigma,$$

where the X_i are i.i.d. and uniform on the unit interval; here $\mu = 1/12$ and $\sigma = 1/\sqrt{12}$ are the mean and standard deviation, respectively. Your task is to simulate sim $= 10^4$ realisations of the variable Z_k, for say $k = 1, 2, 3, 5, 10, 25$, and display the corresponding histograms. Observe how the distribution of Z_k comes closer and closer to the standard normal, as k increases. To illustrate just how close, consider the case of $k = 6$, for example, and attempt to test the hypothesis that the 10^4 data points you have simulated come from the standard normal. Comment on your findings.

2. Illustrating the Law of Large Numbers (LLN)

Simulate say 10^4 variables $X_1, X_2, \ldots$ drawn from the unit exponential distribution. Compute and display the sequence

$$W_n = n^{-1}\sum_{i=1}^n (X_i - \bar X_n)^3 \quad \text{for } n = 1, 2, 3, \ldots,$$

where $\bar X_n = n^{-1}\sum_{i=1}^n X_i$. Comment on your picture, and show indeed that W_n converges in probability. Generalise your finding.

3. The continuity lemma for convergence in probability

There are actually two 'continuity lemmas' for convergence in probability.

(a) Suppose $X_n \to_{\rm pr} a$, with a being a constant. Show that if g is a function continuous at point $x = a$, then indeed $g(X_n) \to_{\rm pr} g(a)$.

(b) Suppose more generally that $X_n \to_{\rm pr} X$, with the limit being a random variable. Show that if g is a function that is continuous in the set in which X falls, then $g(X_n) \to_{\rm pr} g(X)$.

Comments: (i) To prove (b), use uniform continuity over closed and bounded intervals. (ii) In situations of relevance for this course, part (a) will be the more important. The typical application may be that consistency of $\widehat\theta_n$ for θ implies consistency of $g(\widehat\theta_n)$ for $g(\theta)$.

4. The maximum of uniforms

Let $X_1, \ldots, X_n$ be i.i.d. from the uniform $[0, \theta]$ distribution, and let $M_n = \max_{i \le n} X_i$.

(a) Show that $M_n \to_{\rm pr} \theta$ (i.e. the maximum observation is a consistent estimator of the unknown endpoint).

(b) Find the limit distribution of $V_n = n(\theta - M_n)$, and use this result to find an approximate 95% confidence interval for θ.

5. Distribution functions

For a real random variable X, consider its distribution function $F(t) = \Pr\{X \le t\}$. Show that F is right continuous, and that its set of discontinuities is at most countable (in particular, the set of continuity points is dense). Show also that $F(t) \to 1$ when $t \to \infty$ whereas $F(t) \to 0$ when $t \to -\infty$.

6. A 'master theorem' for convergence in distribution

[xx check ferguson's definition. xx] Let X_n and X be real random variables, with probability distributions P_n and P [so that $P_n(A) = \Pr\{X_n \in A\}$, etc.], and consider the following five statements:

(1) $X_n \to_d X$;

(2) for every open set A, $\liminf P_n(A) \ge P(A)$;

(3) for every closed set B, $\limsup P_n(B) \le P(B)$;

(4) for every set C that is P-continuous, in the sense that $P(\partial C) = 0$, where $\partial C = \bar C - C^0$ is the 'boundary' of C (the closure minus its interior), $\lim P_n(C) = P(C)$;

(5) for every bounded and continuous g, $\lim {\rm E}\, g(X_n) = {\rm E}\, g(X)$.

Show that these five statements are in fact all equivalent. Hints: For (1) implies (2), write $A = \cup_{j=1}^\infty A_j$ for open sets $A_j = (a_j, b_j)$, where a_j and b_j can be chosen to be among the continuity points for the distribution function F for X. Then show that (2) implies (3) [using that B is closed if and only if B^c is open], and that (3) implies (4). For (4) implying (5), take g to have its values inside $[0, 1]$, without loss of generality, and write

$$ {\rm E}\, g(X_n) = \int \int_0^1 I\{y \le g(x)\}\, {\rm d}y\, {\rm d}P_n(x) = \int_0^1 \Pr\{g(X_n) \ge y\}\, {\rm d}y, $$

along with a Lebesgue theorem for convergence of integrals. Finally, for (5) implies (1), construct for given F-continuity point x a continuous function g_ε that is close to $g_0(y) = I\{y \le x\}$.

7. The continuity lemma for convergence in convergence

Suppose $X_n \to_d X$ and that h is continuous (and not necessarily bounded). Show that $h(X_n) \to_d h(X)$. [Use e.g. statement (5) of the previous exercise.] Thus $\exp(tX_n) \to_d \exp(tX)$, etc.

8. Convergence in distribution for discrete variables

Let X_n and X take on values in the set of natural numbers, and let $p_n(j) = \Pr\{X_n = j\}$ and $p(j) = \Pr\{X = j\}$ for $j = 0, 1, 2, \ldots$. Show that $X_n \to_d X$ if and only if $p_n(j) \to p(j)$ for each j. To illustrate this, prove the classic 'law of small numbers' (first proven by Ladislaus Bortkiewicz in 1898), that a binomial is close to a Poisson, if the count number is high and the probability is small.

9. Convergence in probability in dimension two (and more)

We have defined $X_n \to_{\rm pr} X$ to mean that

$$\Pr\{|X_n - X| \ge \varepsilon\} \to 0 \quad \text{for each } \varepsilon > 0.$$

The natural generalisation for the two-dimensional (and higher) case is to say that

$$X_n = (X_{n,1}, X_{n,2}) \to_{\rm pr} X = (X_1, X_2)$$

provided

$$\Pr\{\|X_n - X\| \ge \varepsilon\} \to 0 \quad \text{for each } \varepsilon > 0,$$

where $\|X_n - X\|$ is the usual Euclidean distance. Prove that $X_n \to_{\rm pr} X$ (in such a two-dimensional situation) if and only if $X_{n,j} \to_{\rm pr} X_j$ for $j = 1, 2$ (i.e. ordinary one-dimensional convergence for each component). Generalise.

10. Moment generating functions and convergence in distribution

For a random variable X, its moment generating function (mgf) is

$$M(t) = {\rm E}\exp(tX),$$

defined for each t at which the expectation exists. Among its basic properties are the following; attempt to demonstrate these.

1. $M(0) = 1$, and when the mean is finite, then $M'(t)$ exists, with $M'(0) = {\rm E}\,X$.
2. More generally, if $|X|^r$ has finite mean, then $M^{(r)}(0) = {\rm E}\,X^r$ (the rth derivative of M, at the point zero).
3. When X and Y are independent, then

   $$M_{X+Y}(t) = M_X(t) M_Y(t)$$

   in the obvious notation. This generalises of course to the case of more than two independent variables.
4. If X and Y are two variables with identical mgfs, then their distributions are identical. [There are also 'inversion formulae' in the literature, giving the distribution as a function of M.]

5. If X_n and X have mgfs M_n and M, then $M_n(t) \to M(t)$ for all t in a neighbourhood around zero is sufficient for $X_n \to_d X$.

6. In particular, if $M_n(t) \to \exp(\frac{1}{2}t^2)$ for all t close to zero, then $X_n \to_d \mathrm{N}(0,1)$.

11. Finite moments

Show that if $\mathrm{E}\,X^2$ is finite, then necessarily $\mathrm{E}\,X$ is finite too. Show more generally that $\mathrm{E}\,|X|^q$ is finite, then also $\mathrm{E}\,|X|^p$ is finite for all $p < q$. Prove indeed that $(\mathrm{E}\,|X|^p)^{1/p}$ is a non-decreasing function of p.

12. Proving the CLT (under some restrictions)

Let $X_1, X_2, \ldots$ be i.i.d. with some distribution F having finite variance and mean, and assume for simplicity that the mean is zero.

(a) Show that if the mgf exists, in a neighbourhood around zero, then

$$M(t) = 1 + \tfrac{1}{2}\sigma^2 t^2 + o(t^2),$$

where σ is the standard deviation of X_i.

(b) Show that $\sqrt{n}\bar{X}_n$ has mgf of the form

$$M_n^*(t) = M(t/\sqrt{n})^n = \{1 + \tfrac{1}{2}\sigma^2 t^2/n + o(1/n)\}^n,$$

and conclude that the CLT holds.

13. Characteristic functions

The trouble with the approach to the CLT above is that is has somewhat limited scope, in that some distributions do not have a finite mgf (since $\exp(tX)$ may be too big with too high probability for its mean to be finite). The so-called characteristic functions (chf) provide a more elegant mathematical tool in this regard. For a random variable X, its chf is defined as

$$\phi(t) = \mathrm{E}\,\exp(itX) = \mathrm{E}\,\cos(tX) + i\,\mathrm{E}\,\sin(tX),$$

with $i = \sqrt{-1}$ the complex unit, and $t \in R$.

(a) Show that the chf always exists, and that is is uniformly continuous. Show that the chf for the $\mathrm{N}(0, \sigma^2)$ is $\exp(-\frac{1}{2}\sigma^2 t^2)$.

(b) Assume $X_n \to_d X$. Show that

$$\phi_n(t) = \mathrm{E}\,\exp(itX_n) \to \phi(t) = \mathrm{E}\,\exp(itX) \quad \text{for all } t.$$

(c) The converse is also true (but harder to prove), and it is 'inside the curriculum' to know this: If

$$\phi_n(t) = \mathrm{E}\,\exp(itX_n) \quad \text{converges to some function} \quad \phi(t)$$

for all t in an interval around zero, and this limit function is continuous there, then (i) $\phi(t)$ is necessarily the chf of some random variable X, and (ii) there is convergence in distribution $X_n \to_d X$.

14. When is the sum of Bernoulli variables close to a normal?

Let $X_1, X_2, \ldots$ be independent Bernoulli variables (i.e. taking values 0 and 1 only), with $X_i \sim \mathrm{Bin}(1, p_i)$. We shall investigate when

$$Z_n = \frac{\sum_{i=1}^n (X_i - p_i)}{B_n} \to_d \mathrm{N}(0,1),$$

where $B_n = \{\sum_{i=1}^n p_i(1-p_i)\}^{1/2}$. Show, using mgfs or chfs, that this happens if and only $\sum_{i=1}^\infty p_i = \infty$ – and show, additionally, that this condition is equivalent to $B_n \to \infty$. Thus the cases $p_i = 1/i$ and $p_i = 1/i^2$, for example, are fundamentally different. For this second case, investigate the limit distribution of Z_n (which by the arguments given is not normal).

15. Proving the CLT (again)

Using chfs instead of mgfs gives a more elegant and unified proof of the CLT.

(a) Show that if X has a finite mean ξ, then its chf satisfies

$$\phi(t) = 1 + i\xi t + o(t) \quad \text{for } t \to 0.$$

Also, its derivative exists, and $\phi'(0) = \xi$.

(b) Show similarly that if X has a finite variance σ^2, then

$$\phi(t) = 1 + i\xi t - \tfrac{1}{2}(\xi^2 + \sigma^2 t^2) + o(t^2) \quad \text{for } t \to 0.$$

(c) If $X_1, X_2, \ldots$ are i.i.d. with mean zero and finite variance σ^2, then show that $Z_n = \sqrt{n}\bar{X}_n$ has chf of the form

$$\phi_n(t) = \{1 - \tfrac{1}{2}\sigma^2 t^2/n + o(1/n)\}^n.$$

Prove the CLT from this.

16. More on characteristic functions

Here are some more details and illustrations pertaining to characteristic functions.

(a) Find the characteristic function for a binomial distribution and for a Poisson distribution.

(b) Demonstrate the classical 'Gesetz der kleinen Zahlen' (cf. Exercise 8), that a binomial (n, p_n) tends to the Poisson λ, when $np_n \to \lambda$.

(c) Show that for the Cauchy distribution, with density $f(x) = (1/\pi)(1+x^2)^{-1}$, the chf is equal to $\exp(-|t|)$. Note that this function does not have a derivative at zero, corresponding to the fact that the Cauchy does not have a finite mean (cf. Exercise 15(a)).

(d) Let $X_1, \ldots, X_n$ be i.i.d. from the Cauchy. Show that the chf of $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ is identical to the chf of a single observation. Conclude, by the 'inversion theorem', the amazing fact that $\bar{X}_n =_d X_i$; the average has the same statistical distribution as each single component.

(e) There are several versions of 'inverse theorems', providing a mechanism for finding the distribution of a random variable from its chf; the perhaps primary aspect, defined as an 'inside curriculum fact', is that the chf indeed fully characterises the distribution (if X and Y have identical chfs, then their distributions are identical too). One such inversion formula is as follows: if X has a chf ϕ that is integrable (i.e. $\int |\phi(t)| \,\mathrm{d}t$ is finite), then X has a density f, for which a formula is

$$f(x) = \frac{1}{2\pi} \int \exp(-itx)\phi(t) \,\mathrm{d}t.$$

Write down what this means, in the cases of a normal and a Cauchy, and verify the implied formulae. Show that f in each such case of an integrable $\phi(t)$ necessarily becomes continuous.

(f) Show that the chf for the uniform $[-1, 1]$ distribution becomes $\phi(t) = (\sin t)/t$. Deduce that

$$\int \Big|\frac{\sin t}{t}\Big| \,\mathrm{d}t = \infty \quad \text{even though} \quad \int \frac{\sin t}{t} \,\mathrm{d}t = \pi.$$

(g) Point (e) above gives a formula for the density f of a variable, in the case of it having an integrable chf ϕ. One also needs a more general formula, for the case of variables that do not have densities, etc. Let X be any random variable, with cumulative distribution function F and chf ϕ (but with nothing assumed about it having a density), and add on to it a little bit of Gaußian noise:

$$Z_\sigma = X + Y_\sigma, \quad \text{with } Y \sim \mathrm{N}(0, \sigma^2).$$

Then Z has a density (even if X does not have one). Our intention is to let $\sigma \to 0$, to come back to X. Show that Z_σ has cdf of the form

$$F_\sigma(x) = \int F(x-y) \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp(-\tfrac{1}{2}y^2/\sigma^2) \,\mathrm{d}y$$

and chf equal to

$$\phi_\sigma(t) = \phi(t) \exp(-\tfrac{1}{2}\sigma^2 t^2).$$

Hence show that

$$f_\sigma(x) = \frac{1}{2\pi} \int \exp(-itx)\phi(t) \exp(-\tfrac{1}{2}\sigma^2 t^2) \,\mathrm{d}t.$$

and that, consequently,

$$\begin{aligned}
\Pr\{X + Y_\sigma \in [a, b]\} &= F_\sigma(b) - F_\sigma(a) \\
&= \frac{1}{2\pi} \int \frac{\exp(-itb) - \exp(-ita)}{-it} \phi(t) \exp(-\tfrac{1}{2}\sigma^2 t^2) \,\mathrm{d}t.
\end{aligned}$$

(h) Conclude with the following general inversion formula, valid for all continuity points a, b of F:

$$F(b) - F(a) = \lim_{\sigma \to 0} \frac{1}{2\pi} \int \frac{\exp(-itb) - \exp(-ita)}{-it} \phi(t) \exp(-\tfrac{1}{2}\sigma^2 t^2) \,\mathrm{d}t.$$

17. Scheffé's Lemma

There are situations where $g_n(y) \to g(y)$ for all y, for appropriate functions g_n and g, does not imply $\int g_n(y)\,\mathrm{d}y \to \int g(y)\,\mathrm{d}y$. However, it may be shown that this is not a problem when g_n and g are probability densities (due to certain 'dominated convergence' Lebesgue theorems from the theory of measure and integration): if g_n and g are the densities of Y_n and Y, and $g_n(y) \to g(y)$ for (almost) all y, then

$$\int |g_n - g|\,\mathrm{d}y \to 0,$$

and, in particular,

$$\Pr\{Y_n \in [a,b]\} = \int_a^b g_n(y)\,\mathrm{d}y \to \int_a^b g(y)\,\mathrm{d}y = \Pr\{Y \in [a,b]\}$$

for all intervals, and we have $Y_n \to_d Y$. This is Scheffé's Lemma, defined as an inside curriculum fact.

(a) Let $Y_n \sim t_n$, a t distribution with n degrees of freedom. Show that $Y_n \to_d \mathrm{N}(0,1)$, using this lemma. Can you prove this statement in a simpler fashion?

(b) If $X_1, \ldots, X_n$ are i.i.d. from a uniform on $[0,1]$, with $M_n = \max_{i \le n} X_i$, show using the Scheffé Lemma that $n(1 - M_n)$ tends to a unit exponential in distribution.

(c) Suppose $X_n \sim \chi^2_n$, and consider $Z_n = (X_n - n)/\sqrt{2n}$. Prove that $Z_n \to_d \mathrm{N}(0,1)$.

18. The median

'The median isn't the message', said Stephen Jay Gould (when he was diagnosed with a serious illness and looked at survival statistics). Let $X_1, \ldots, X_n$ be i.i.d. from a positive density f with true median $\theta = F^{-1}(\frac{1}{2})$.

(a) Suppose for simplicity that n is odd, say $n = 2m+1$. Show that M_n has density of the form

$$g_n(y) = \frac{(2m+1)!}{m!\,m!} F(y)^m \{1 - F(y)\}^m f(y).$$

(b) Show then that the density of $Z_n = \sqrt{n}(M_n - \theta)$ can be written in the form

$$h_n(z) = g_n(\theta + z/\sqrt{n})/\sqrt{n}.$$

Prove that

$$h_n(z) \to (2\pi)^{-1/2} 2 f(\theta) \exp\{-\tfrac{1}{2} 4 f(\theta)^2 z^2\},$$

which by the Scheff'e Lemma means that

$$\sqrt{n}(M_n - \theta) \to_d \mathrm{N}(0, \tau^2) \quad \text{with } \tau = \tfrac{1}{2}/f(\theta).$$

Why does this also prove that the sample median is consistent for the population median?

(c) Generalise to the following quantilian result: if $Q_n(p) = F_n^{-1}(p)$ is the pth quantile of the data, then $Q_n(p)$ converges in probability to the corresponding population quantile $\xi_p = F^{-1}(p)$, and

$$\sqrt{n}\{Q_n(p) - \xi_p\} \to_d \mathrm{N}(0, \tau_p^2) \quad \text{with } \tau_p^2 = p(1-p)/f(\xi_p)^2.$$

(d) Constructing a nonparametric confidence interval for an unknown median is not that simple – the 'usual recipe' works, up to a point, and tells us that if we first find a consistent estimator $\widehat\kappa$ of the doubly unknown quantity $f(\theta)$ (f is unknown, and so is θ, its median), then we're in business. We would then have

$$Z_n = \frac{\sqrt{n}(M_n - \theta)}{\widehat\tau} \to_d \mathrm{N}(0,1), \quad \text{with } \widehat\tau = \tfrac{1}{2}/\widehat\kappa,$$

from which it then follows that

$$I_n = \widehat\theta \pm 1.96\,\widehat\tau/\sqrt{n} \quad \text{obeys} \quad \Pr\{\theta \in I_n\} \to 0.95.$$

The trouble lies in finding a satisfactory $\widehat\kappa$. Try to construct such a consistent estimator.

19. Limiting local power games

This exercise is meant to study a 'prototype situation' in some detail; the type of calculation and results will be seen to rather similar in a long range of different situations. – Let $X_1, \ldots, X_n$ be i.i.d. data from $\mathrm{N}(\theta, \sigma^2)$. One wishes to test $H_0\colon \theta = \theta_0$ vs. the alternative that $\theta > \theta_0$, where θ_0 is a known value (e.g. 3.14). Two tests will be considered, based on respectively

$$\bar X_n = n^{-1}\sum_{i=1}^n X_i \quad \text{and} \quad M_m = \mathrm{median}(X_1, \ldots, X_n).$$

(a) For given value of θ, prove that

$$\sqrt{n}(\bar X_n - \theta) \to_d \mathrm{N}(0, \sigma^2),$$
$$\sqrt{n}(M_n - \theta) \to_d \mathrm{N}(0, (\pi/2)\sigma^2).$$

Note that the first result is immediate and actually holds with exactness for each n; the second result requires more care, e.g. working with the required density, cf. Exercise xx.

(b) Working under the null hypothesis $\theta = \theta_0$, show that

$$Z_n = \sqrt{n}(\bar X_n - \sigma_0)/\widehat\sigma \to_d \mathrm{N}(0,1),$$
$$Z_n^* = \sqrt{n}(M_n - \theta_0)/\{(\pi/2)^{1/2}\widehat\sigma\} \to_d \mathrm{N}(0,1),$$

where $\widehat\sigma$ is any consistent estimator of σ.

– [xx Figure 1: Limiting local power functions for two tests for $\theta \le \theta_0$ against $\theta > \theta_0$, in the situation with $\mathrm{N}(\theta, \sigma^2)$ data. based on the mean (full line) and on the median (dotted line). xx]

(c) Conclude from this that the two tests that reject H_0 provided respectively

$$\bar X_n > \theta_0 + z_{0.95}\widehat\sigma/\sqrt{n} \quad \text{and} \quad M_n > \theta_0 + z_{0.95}(\pi/2)^{1/2}\widehat\sigma/\sqrt{n},$$

where $z_{0.95} = \Phi^{-1}(0.95) = 1.645$, have the required asymptotic significance level 0.05;

$$\alpha_n = \Pr\{\text{reject } H_0 \,|\, \theta = \theta_0\} \to 0.05.$$

(There is one such α_n for the first test, and one for the other; both converge however to 0.05.)

(d) Then our object is to study the *local power*, the chance of rejecting the null hypothesis under alternatives of the type $\theta_n = \theta_0 + \delta/\sqrt{n}$. In generalisation of (b), show that

$$Z_n = \sqrt{n}(\bar{X}_n - \sigma_0)/\widehat{\sigma} \to_d \mathrm{N}(\delta/\sigma, 1),$$
$$Z_n^* = \sqrt{n}(M_n - \theta_0)/\{(\pi/2)^{1/2}\widehat{\sigma}\} \to_d \mathrm{N}((\pi/2)^{1/2}\delta/\sigma, 1),$$

[xx check this xx] where the convergence in question takes place under the indicated $\theta_0 + \delta/\sqrt{n}$ parameter values. (You need to generalise the results of Exercise xx, to the $\delta \neq 0$ case.)

(e) Use these results to show that

$$\begin{aligned}\pi_n(\delta) &= \Pr\{\text{reject} \,|\, \theta_0 + \delta/\sqrt{n}\} \to \Phi(\delta/\sigma - z_{0.95}),\\ \pi_n^*(\delta) &= \Pr\{\text{reject} \,|\, \theta_0 + \delta/\sqrt{n}\} \to \Phi((2/\pi)^{1/2}\delta/\sigma - z_{0.95}),\end{aligned}$$

for the two power functions. Draw these in a diagram, and compare; cf. Figure xx.

(f) Assume one wishes n to be large enough to secure that the power function is at least at level β for a certain alternative point θ_1. Using the local power approximation, show that the required sample sizes are respectively

$$n_A \doteq \frac{\sigma^2}{(\theta_1 - \theta_0)^2}(z_{1-\alpha} + z_\beta)^2 \quad \text{and} \quad n_B \doteq \frac{\sigma^2/c^2}{(\theta_1 - \theta_0)^2}(z_{1-\alpha} + z_\beta)^2$$

for tests A (based on the mean) and B (based on the median), with $c = \sqrt{2/\pi}$. Compute these sample sizes for the case of $\beta = 0.05$ and $\theta_1 = \theta_0 + \frac{1}{2}\sigma$, when also $\alpha = 0.05$.

(g) Lehmann defines 'the ARE [asymptotic relative efficiency] of test B with respect to test A' as

$$\mathrm{ARE} = \lim \frac{n_A(\theta_1, \beta)}{n_B(\theta_1, \beta)},$$

the limit in question in the sense of alternatives θ_1 coming closer to the null hypothesi at speed $1/\sqrt{n}$. Show that indeed

$$\mathrm{ARE} = \frac{\sigma^2}{\sigma^2/c^2} = c^2 = 2/\pi = 0.6366$$

in this particular situation – test A needs only ca. 64% as many data points to reach the same detection power as B needs.

20. Testing the normal scale

We have essentially covered Exercise 19 in class [xx alter this xx], as a 'prototype illustration' of the themes developed in Chapter 3 [xx change this xx]. Here is another illustration, for you to check that you may develop similar results in a different situation. Data $X_1, \ldots, X_n$ are now taken to be i.i.d. $\mathrm{N}(0, \sigma^2)$, and the object is to construct and compare tests for $H_0 : \sigma = \sigma_0$ vs. $\sigma > \sigma_0$, where σ_0 is some known quantity.

(a) Show that $\mathrm{E}\, X_i^2 = \sigma^2$ and $\mathrm{E}\, |X_i| = b\sigma$, with $b = \sqrt{2/\pi}$. Show that the estimators

$$\widehat{\sigma}_A = \Big\{n^{-1}\sum_{i=1}^n X_i^2\Big\}^{1/2} \quad \text{and} \quad \widehat{\sigma}_B = n^{-1}\sum_{i=1}^n |X_i|/b$$

are both consistent for σ.

(b) Find the limit distributions for

$$Z_{n,A} = \sqrt{n}(\widehat{\sigma}_A - \sigma) \quad \text{and} \quad Z_{n,B} = \sqrt{n}(\widehat{\sigma}_B - \sigma),$$

and comment on your findings.

(c) Construct explicit tests A and B, based on respectively $\widehat{\sigma}_A$ and $\widehat{\sigma}_B$, that have asymptotic level $\alpha = 0.01$.

(d) Show that both tests are consistent.

(e) Then we need to compare the two tests in terms of local power. For alternatives of the type $\sigma = \sigma_0 + \delta/\sqrt{n}$, establish limit distributions of the type

$$\sqrt{n}(\widehat{\sigma}_A - \sigma_0) \to_d \mathrm{N}(\delta, \tau_A^2\sigma^2),$$
$$\sqrt{n}(\widehat{\sigma}_B - \sigma_0) \to_d \mathrm{N}(\delta, \tau_B^2\sigma^2),$$

with certain values (that you should find) for τ_A and τ_B.

(f) Establish the limiting local power functions $\pi_A(\delta)$ and $\pi_B(\delta)$, and plot them in a diagram (cf. Figure xx of the previous exercise).

(g) Compute the required sample sizes n_A and n_B for tests A and B to achieve detection power 0.99 when the true state of affairs is $\sigma = 1.333\,\sigma_0$.

(h) Compute the ARE for test A w.r.t. test B, and comment.

(i) Could there be other tests for H_0 here that would outperform test A?

21. Algebras of sets

Let $\mathcal{X}$ be a non-empty set, and let $\mathcal{A}$ be a class of subsets of $\mathcal{X}$. We say that $\mathcal{A}$ is an *algebra* if (i) both $\mathcal{X}$ and the empty-set is in $\mathcal{A}$; (ii) each time A is in $\mathcal{A}$, then also its complement A^c is in $\mathcal{A}$; (iii) whem $A_1, \ldots, A_n$ are sets in $\mathcal{A}$, then also their union $\cup_{i=1}^n A_i$ is in $\mathcal{A}$. In other words: an algebra is closed with respect to the formation of complements and finite unions.

(a) Are you yourself closed with respect to compliments?

(b) What's the world's smallest algebra?

(c) Show that an algebra is also closed with respect to finite intersections.

(d) And show that $A - B = A \cap B^c$ is within the algebra if A and B are so.

(e) Construct an example of an algebra.

(f) What was Muhammad ibn Musa al-Khvarizmi [xx fix xx]?

22. Sigma-algebras of sets

A *sigma-algebra* is an algebra $\mathcal{A}$ which is also closed with respect to countably infinite formations of unions, that is, if $A_1, A_2, \ldots$ are in $\mathcal{A}$, then so is $\cup_{i=1}^\infty A_i$.

(a) Let $\mathcal{A}$ consist of all those subsets of $\mathcal{R}$, the real numbers, which are themselves either finite or have finite complements. Is $\mathcal{A}$ an algebra? A sigma-algebra?

(b) Show that a sigma-algebra is closed with respect to countably infinite intersection operations.

23. Inverse and direct images of functions

Let $f\colon \mathcal{X} \to \mathcal{Y}$ be an arbitrary function, from set $\mathcal{X}$ to set $\mathcal{Y}$. For subsets A of $\mathcal{X}$, define the *direct image* as $fA = f(A) = \{f(x)\colon x \in A\}$. And for subsets B of $\mathcal{Y}$, define the *inverse image* as $f^1 B = f^{-1}(B) = \{x\colon f(x) \in B\}$.

(a) Let $\{B_i\colon i \in I\}$ be a collection of subsets of $\mathcal{Y}$. Show that $f^{-1}(\cup_i B_i) = \cup_i f^{-1}(B_i)$.

(b) And that $f^{-1}(\cap_i B_i) = \cap_i f^{-1}(B_i)$.

(c) Then show $f^{-1}(\mathcal{Y} - B) = \mathcal{X} - f^{-1}(B)$.

(d) Show that $A \subset f^{-1}f(A)$ for all A.

(e) And that $B \supset ff^{-1}B$ for all B.

(f) For functions $f\colon \mathcal{X} \to \mathcal{Y}$ and $g\colon \mathcal{Y} \to \mathcal{Z}$, show that $(g \circ f)^{-1}(C) = f^{-1}g^{-1}C$.

24. Independence of complements

We say that $A_1, \ldots, A_n$ are independent if

$$P(A_{i_1} \cap \cdots \cap A_{i,m}) = P(A_{i_1}) \cdots P(A_{i_m})$$

for all subsets $\{i_1, \ldots, i_m\}$ of $\{1, \ldots, n\}$. Thus we demand quite a bit more than merely saying that $P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$.

Show that if $A_1, \ldots, A_n$ are independent, then so are $A_1^c, \ldots, A_n^c$.

25. The Borel–Cantelli emma

Let $A_1, A_2, \ldots$ denote events with probabilities $P(A_1), P(A_2), \ldots$. We are interested in the event that infinitely many of these A_j occur, i.e.

$$A_{\text{i.o.}} = \cap_{i \geq 1} \cup_{j \geq i} A_j.$$

(a) Show that if $\sum_{i=1}^{\infty} P(A_i) < \infty$, then $P(A_{\text{i.o.}}) = 0$. In other words, it is certain that only a finite number of the A_i will occur.

(b) Show under the additional assumption that the A_j are independent, that the previous result holds in the 'if and only if' sense, i.e. that if $\sum_{i=1}^{\infty} P(A_i) = \infty$, then $P(A_{\text{i.o.}}) = 1$. In particular, under independence, the probability of $A_{\text{i.o.}}$ is either 0 or 1, there is no 'middle ground' possibility.

26. Does this happen infinitely often?

Let $X_1, X_2, \ldots$ be independent with the same Expo(1) distribution, i.e. with density e^{-x} for $x \geq 0$.

(a) Will $X_n > 10 + 0.99 \log n$ infinitely often ?

(b) Will $X_n > 10 + 1.00 \log n$ infinitely often?

(c) Will $X_n > 10 + 1.01 \log n$ infinitely often?

(d) Will $X_n > 10^{12} + \log n$ infinitely often?

27. Normal deviations

Let X be standard normal, and write as usual $\Phi(x)$ for its cumulative distribution function and $\phi(x)$ for its density.

(a) Show that $\Pr\{X > x\} = 1 - \Phi(x) \doteq \phi(x)/x$ for large x.

(b) Let $X_1, X_2, \ldots$ be independent standard normals. Pray, will $X_n > 0.000001\sqrt{n}$ for infinitely many n?

(c) Let $\bar{X}_n$ be the average of the first n of these observations. Show that $|\bar{X}_n| > \varepsilon$ for at most a finite number of n.

(d) If $X_1, X_2, \ldots$ are independent and $\mathrm{N}(\xi, 1)$, what is the probability that $\bar{X}_n$ converges to ξ?

28. If you are sure about infinitely many things

Show that the event $\cap_{n=1}^{\infty} B_n$ is certain (i.e. it takes place with probability 1) if and only if each of the B_n is certain. Construct an example to show that this is *not* the case for uncountably many certain events.

29. At msot countably many discontinuities

Let F be a one-dimensional cumulative distribution function, and let D be the set of its discontinuities. Show that D is either empty, finite, or countably infinite.

30. Borel sets in dimensions one and two

Let $\mathcal{B}$ be the Borel sets in $\mathcal{R}$; it is the smallest sigma-algreba containing all intervals. Define then

$$\mathcal{B} \times \mathcal{B} = \sigma(\mathcal{C}),$$

the smallest sigma-algebra containing all $A \times B$, with A and B in $\mathcal{B}$. (This is the usual definition of a product-sigma-algebra.) Define also

$$\mathcal{B}^2 = \sigma(\mathcal{O}),$$

where $\mathcal{O}$ is the set of all open sets in $\mathcal{R}^2$ (This is the usual definition of a Borel-sigma-algebra.) Show that, luckily & conveniently, $\mathcal{B} \times \mathcal{B} = \mathcal{B}^2$.

31. Measurability of coordinate functions

Let $f, g \colon (\Omega, \mathcal{A}) \to (\mathcal{R}, \mathcal{B})$ be two functions, and let $h \colon \Omega \to \mathcal{R}^2$ be given by

$$h(\omega) = (f(\omega), g(\omega)).$$

Show that h is measurable if & only if both f and g are measureable. Generalise.

32. Normal mixtures

Let first X and Y be independent, with X a standard normal and Y very discrete, $\Pr\{Y = y\} = \frac{1}{2}$ for $y \in \{-1, 1\}$. Note that a sum of a continuous and a discrete variable will have a continuous distribution. Find the density for $X + Y$. Find also its mean and variance.

Generalise to finite normal mixtures, which may be done in several ways, with one path as follows. Start with the density

$$f(x) = \sum_{j=1}^{k} p_j \phi_{\sigma_j}(x - \mu_j),$$

defined via the triples (p_j, μ_j, σ_j) for $j = 1, \ldots, k$. Here the p_j make up a probability vector, i.e. nonnegative with sum 1, and $\phi_\sigma(x-u) = \sigma^{-1}\phi(\sigma^{-1}(x-\mu))$ is the density of the normal (μ, σ). One may now view X, drawn from f, as the result of the two-stage operation where the index $J = j$ is drawn from $\{1, \ldots, k\}$ first, with $\Pr\{J = j\} = p_j$, and $X \mid j \sim \mathrm{N}(\mu_j, \sigma_j^2)$. Use this to find $\mathrm{E}\,(X \mid j)$ and $\mathrm{Var}\,(X \mid j)$, and then the unconditional mean and variance for X.

The class of finite normal mixtures is a large one, and even with say $k \leq 5$ components a broad range of shapes may be attained – play a bit with this on your computer, drawing $f(x)$ curves on your screen, by mixing in different input vectors of p_j, μ_j, σ_j.

Find also a formula for the skewness of f, i.e. $\gamma = \mathrm{E}\,\{(X - \mu)/\sigma\}^3$, in terms of the overall mean and standard deviation μ and σ.

33. The Markov inequality, and bounding tails

Sometimes one wishes to bound tail probabilities, say $\Pr\{X \geq a\} \leq B(a)$, and there are several ways in which to do this.

(a) Let X be a nonnegative random variable, and let $h(x)$ be a nonnegative and nondecreasing function for $x \geq 0$. Demonstrate Неравенство Маркова (Markov's inequality), that

$$\Pr\{X \geq a\} \leq \mathrm{E}\,h(X)/h(a).$$

(b) If X is a random variable with mean ξ, show that

$$\Pr\{|X - \xi| \geq \varepsilon\} \leq \frac{\mathrm{E}\,|X - \xi|^p}{\varepsilon^p} \quad \text{for each } p > 0.$$

For $p = 2$ we have the famous special case of Неравенство Чебышёва (Chebyshov's inequality, from about 1853).

(c) Let $X_1, X_2, \ldots$ be independent normals $\mathrm{N}(\xi, 1)$, so that $\bar{X}_n \sim \mathrm{N}(\xi, 1/n)$. Writing N for a standard normal, show that

$$\Pr\{|\bar{X}_n - \xi| \geq \varepsilon\} \leq \frac{n^{-p/2}\mathrm{E}\,|N|^p}{\varepsilon^p} \quad \text{for each } p > 0.$$

For $n = 100$ and $\varepsilon = 0.05$, compute the exact probability in question and track the right hand bound as a function of p. Which p gives the sharpest bound, in this case?

(d) Let X have moment generating function $M(t) = \mathrm{E}\,\exp(tX)$, assumed to be finite for at least $0 \leq t \leq t_0$. Show that

$$\Pr\{X \geq a\} \leq \min_{0 \leq t \leq t_0} \exp(-ta)M(t).$$

(e) For the case of $\bar{X}_n \sim \mathrm{N}(\xi, 1/n)$ studied above, show that

$$\Pr\{\bar{X}_n - \xi \geq \varepsilon\} \leq \exp(-\tfrac{1}{2}n\varepsilon^2).$$

Compare this bound with the one reached via Chebyshov above.

(f) Let $X_1, X_2, \ldots$ be i.i.d. from the χ^2_b distribution, with $\mathrm{E}\,\bar{X}_n = b$ and $\mathrm{Var}\,\bar{X}_n = 2b/n$. Show that with $\varepsilon > 0$ given, there will with probability 1 be only finitely many n with $\bar{X}_n \geq b + \varepsilon$.

(g) [xx invent another application here. xx]

34. Amor's arrows sometimse miss

[From Nils Exam ST 200 December 1989, Exercise 1(e).] Amor shoots her arrows infinitely many times. Her shots are independent of each other, and shot no. n is (X_n, Y_n), measured from origo, where X_n and Y_n are independent and standard normal. The distance from origo is hence $R_n = (X_n^2 + Y_n^2)^{1/2}$, the square-root of a χ^2_2. Show that its density becomes $f(r) = r\exp(-\frac{1}{2}r^2)$. So how often does she miss, and by how much? Find the probabilities for these three events: that $R_n \geq 0.99\sqrt{2\log n}$ infinitely often; that $R_n \geq 1.00\sqrt{2\log n}$ infinitely often; that $R_n \geq 1.01\sqrt{2\log n}$ infinitely often.

35. Twins and paradigm shifts

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of independent standard normals. Say that X_{i-1} and X_i are *twins* if $|X_i - X_{i-1}| \leq c_i$, and that there is a *regime shift* if $|X_i - X_{i-1}| \geq d_i$. Such c_i and d_i will be specified below. Let A be the event that the sequence experiences infinitely many twins, and B the event that the history sees infinitely many regime shifts.

(a) Write up an exact formula for the expected number of twins in the course of the first $n = 10^{12}$ observations. Put up similarly a formula for the expected number of regime shifts over the same period.

(b) Find $P(A)$ for the cases $c_i = 1/i$ and $c_i = 1/i^2$.

(c) Find $P(B)$ for the cases $d_i = 2\sqrt{\log i}$ and $d_i = 2.001\sqrt{\log i}$.

(d) Construct a criterion, expressed in terms of the c_i and d_i, for the history to experience with probability 1 both infinitely many twins and infinitely many regime shifts. Here it many be convenient to first deal with the situations where $\inf_i c_i > 0$ and $\sup_i d_i < \infty$, and then focus on the cases where $c_i \to 0$ and $d_i \to \infty$.

36. Quickness of convergence of average to its mean

Assume that $X_1, X_2, \ldots$ is a sequence of i.i.d. variables with mean zero. Hence $\bar{X}_n$ will converge to 0 in probability, and even with probability 1, by the Law of Large Numbers. But *how fast* will $p_n(a) = \Pr\{\bar{X}_n \geq a\} \to 0$, for fixed $a > 0$?

(a) Assume $\mathrm{Var}\,X_i = \sigma^2$ is finite. Show that $p_n(a) \leq \sigma^2/(na^2)$, hence speed of order $1/n$.

(b) Assume that also the fourth order moment is finite, $\mathrm{E}\,X_i^4 < \infty$. Show that $p_n(a) \leq K\sigma^2/(n^2a^4)$, for a certain K, which gives speed of order $1/n^2$.

(c) Let us generalise: Assume that $\mathrm{E}\,|X_i|^p < \infty$, for a suitable $p \geq 2$. The central limit theorem says $\sqrt{n}\bar{X}_n/\sigma \to_d \mathrm{N}(0,1)$. One may show that

$$\mathrm{E}\,|\sqrt{n}\bar{X}_n/\sigma|^p \to \mathrm{E}\,|\mathrm{N}(0,1)|^p,$$

see e.g. von Bahr (1965). Show from this that

$$\mathrm{E}\,|\bar{X}_n|^p \leq c_p n^{-p/2}\,\mathrm{E}\,|\mathrm{N}(0,1)|^p\,\sigma^p \quad \text{for all } n,$$

for a suitable constant c_p – and one may use $c_p = 1.001$ if 'for all n' is replaced by 'for all large enough n'.

(d) Show that $p_n(a) \leq K_p/(n^{p/2}a^p)$ for a suitable constant K_p.

(e) Assume X_i has moments of all orders, such that (d) holds for each p. If you should succeed in proving that $p_n(a) \leq 0.999999^n$, is this a sharper result?

(f) Assume that the moment generating function $M(t) = \mathrm{E}\,\exp(tX)$ exists for (at least) $0 \leq t \leq t_0$. Show that

$$p_n(a) \leq \rho^n, \quad \text{where } \rho = \rho(a) = \min_{0\leq c\leq t_0} \frac{M(c)}{\exp(ac)},$$

and show that $\rho < 1$. (If $\rho = 1$ the result would still hold, but it would be a boring and rather unpublishable one.)

(g) Find $\rho = \rho(a)$ explicitly, when $X_i \sim \mathrm{N}(0,1)$, and when $X_i \sim \mathrm{N}(0,\sigma^2)$.

(h) It is practical to have explicit results also for $p_n(a) = \Pr\{\bar{X}_n \geq \xi + a\}$, of the type above, for the case of $\mathrm{E}\,X_i = \xi$. Establish such results.

(i) Find $\rho = \rho(a)$ explicitly for the cases (1) $X_i \sim \chi^2_m$; (2) $X_i \sim \mathrm{Bin}(1,p)$; and (3) $X_i \sim \mathrm{Pois}(\lambda)$.

37. The discrete and continuous parts of a cumulative distribution function

Let F be an arbitrary cumulative distribution function on $\mathcal{R}$. Show that one always may decompose F into $F = F_c + F_d$, where F_c is continuous and F_d is discrete.

38. Integrate and display your integrity

well

39. A probabilistic excursion into number theory

In this exercise we shall construct certain types of probability distributions on the natural numbers, via placing probabilities on the the exponents in their prime number factorisations. This becomes an excursion into the world of number theory, to show some their results and formulae, but with the probabilist's hat and spectacles. Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, $p_5 = 11$, etc., be the prime numbers.

(a) Find, like Gauß did when he was a little kid, all the prime numbers up tp 100. Gauß didn't stop there; as a 15 year old boy in 1792 he had essentially understood the fundamental prime number theorem $\pi(x) \doteq x/\log x$, where $\pi(x)$ is the number of primes below x, see point (xx) below. This was not formally proven until about 1896.

(b) Prove, as Euclid did about 2300 year ago, that there are infinitely many primes! (Later proofs of interest include those of Kummer, Pólya, Euler, Axel Thue, Perott, Auric, Métrod, Washington, and Fürstenberg. Even further proofs flow as corollaries of statements proved below, in points (g) and (k).)

(c) We do have $63 = 3^2 \cdot 7^1$, $104 = 2^3 \cdot 13^1$, $30\,141\,766 = 3^2 \cdot 5^1 \cdot 17^1 \cdot 31^2 \cdot 41$, $702\,958\,333 = 7^1 \cdot 11^4 \cdot 19^3$, right? Make it clear to you that each natural number n may be expressed in a unique prime factorisation fashion, in the form $n = p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}$. Here m is the number of the highest prime in n, and $x_1, x_2, \cdots, x_m$ are the exponents. We may also write n as the infinite product $\prod_{j=1}^{\infty} p_j^{x_j}$, where all x_j from a certain $j_0 + 1$ onwards are equal to zero.

(d) This opens a probabilistic door for us, creating a random natural number N by expressing it as

$$N = p_1^{X_1} p_2^{X_2} \cdots = \prod_{j=1}^{\infty} p_j^{X_j},$$

where $X_1, X_2, \ldots$ are random variables in $\{0, 1, 2, \ldots\}$, with the property that only a finite number of them are above 1. Let us try: assume the X_j are independent. Show that N is then a welldefined random variable if and only if

$$\sum_{j=1}^{\infty} \Pr\{X_j \geq 1\} = \sum_{j=1}^{\infty} [1 - \Pr\{X_j = 0\}] < \infty.$$

The division here is sharp: if the sum diverges, then not only is $N = \infty$ with positive probability, but with probability 1.

(e) As a preliminary example, let the X_j be independent with $X_j \sim \mathrm{Pois}(d_j)$. Show that N is welldefined if and only if $\sum_{j=1}^{\infty} d_j < \infty$. Find under this condition the expected values of N and $\log N$. Simulate say 10^4 such N, with $d_j = 1/i^{3/2}$.

(f) There's more beauty to be revealed for the case where the X_j are taken independent and geometrically distributed. Let $X_j \sim \mathrm{Geo}(c_j)$, which means

$$\Pr\{X_j = x\} = (1 - c_j)^x c_j \quad \text{for } x = 0, 1, 2, \ldots.$$

Find the mean, the variance, and the generating function for X_j:

$$\mathrm{E}\, X_j = \frac{1 - c_j}{c_j}, \quad \mathrm{Var}\, X_j = \frac{1 - c_j}{c_j^2}, \quad \mathrm{E}\, s^{X_j} = \frac{c_j}{1 - (1 - c_j)s}.$$

Show also that $\Pr\{X_j \geq x\} = (1 - c_j)^x$. Demonstrate that N is welldefined if and only if $\sum_{j=1}^{\infty} (1 - c_j) < \infty$.

(g) You recall $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$, Euler's sensational finding from about 1734? Consider the choice $c_j = 1 - 1/p_j^2$. Find the probability that N is equal to $1, 11, 63, 103\,141\,766$. Show that

$$\Pr\{N = n\} = \frac{6}{\pi^2} \frac{1}{n^2} \quad \text{for } n = 1, 2, 3, \ldots. \tag{0.1}$$

Then you have also essentially deduced the following intriguing formula:

$$\frac{\pi^2}{6} = \prod_{j=1}^{\infty} \frac{p_j^2}{p_j^2 - 1} = \frac{4}{3} \frac{9}{8} \frac{25}{24} \frac{49}{48} \frac{121}{120} \cdots.$$

As a low-hanging fruit in this garden: *If* there had been merely a finite number of primes, then π^2 would have been rational. Hence (fill in!).

(h) Show also, conversely, that if N is given the (0.1) distribution, then by necessity this leads to independent X_j which are geometrically distributed with parameters $c_j = 1 - 1/p_j^2$.

(i) With this distribution for N, find the following probabilities:

(i) that N is odd [answer: $\frac{3}{4}$];

(ii) that N is a prime numbers;

(iii) that N is a a 'prime potens', of the form p^y, for some $y \geq 1$;

(iv) that N is a factor in 100;

(v) that 100 is a factor in N [answer: $1/100^2$];

(vi) that N turns out to be a square [answer: $\pi^2/15\,!$];

(vii) invent something yourself.

(j) Find the mean for N and for $\log N$. And their variances, unless your willpower is strong enough to resist.

(k) *Riemann's zeta function* is defined as $\zeta(\alpha) = \sum_{n=1}^{\infty} 1/n^\alpha$, for $\alpha > 1$. Thus $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$, etc. Agree to say that N is zeta distributed with parameter α provided

$$\Pr\{N = n\} = \frac{1}{\zeta(\alpha)} \frac{1}{n^\alpha} \quad \text{for } n = 1, 2, 3, \ldots.$$

Assume from this point (k) onwards, up to point (y) below, that N has this distribution. Show that this is equivalent to having the X_j independent and geometric, with $X_j \sim \text{Geo}(1 - 1/p_j^\alpha)$. Derive in particular the following intriguing representation for the zeta function:

$$\zeta(\alpha) = \prod_{\text{prime}} \frac{p^\alpha}{p^\alpha - 1} = \prod_{j=1}^{\infty} \frac{p_j^\alpha}{p_j^\alpha - 1}.$$

This formula was first derived by Euler. So now we know that

$$\frac{\pi^4}{90} = \frac{16}{15} \frac{81}{80} \frac{625}{624} \frac{2401}{2400} \cdots.$$

Show also that $\zeta(\alpha) \to \infty$ as $\alpha \to 1$, which would not have been true if God had given us only a finite number of prime numbers.

(l) Generalise the questions and solutions from point (i) to the more general situation with parameter α rather than 2. Replace also '100' with an arbitrary $n = p_1^{x_1} \cdots p_m^{x_m}$ for sub-points 4 and 5. [A few answers: (l1) $1 - 1/2^\alpha$; (l2) $\zeta(\alpha)^{-1} \sum_1^\infty 1/p_j^\alpha$; (l3) $\zeta(\alpha)^{-1} \sum_1^\infty 1/(p_j^\alpha - 1)$; (l4) $\Pr\{N \text{ is a factor in } n\} = \zeta(\alpha)^{-1} n^{-\alpha} \prod_{j=1}^m (1 + p_j^\alpha + \cdots + p_j^{\alpha x_j})$; (l5) $\Pr\{n \text{ is a factor in } N\} = 1/n^\alpha$; (l6) $\zeta(2\alpha)/\zeta(\alpha)$; (l7) go confidently in the direction of your dreams.]

(m) Say that the number n is *modest* if all prime exponents x_j for n are 0 or 1. Show us three modest and three immodest numbers. Show that the probability that N is modest is $\zeta(2\alpha)^{-1}$. Demonstrate also that

$$B(\alpha) = \sum_{n \text{ modest}} \frac{1}{n^\alpha} = \frac{\zeta(\alpha)}{\zeta(2\alpha)} = \prod_{p \text{ primtall}} \frac{p^\alpha + 1}{p^\alpha}.$$

(n) Say that n is *second-order modest* if all prime exponents are less than or equal to 2. Show that the probability that N is such a second-order modest number is $\zeta(3\alpha)^{-1}$.

(o) Show that the events $\{63$ is a factor in $N\}$ and $\{100$ is a factor in $N\}$ are independent, whereas $\{18$ is a factor in $N\}$ and $\{52$ is a factor in $N\}$ are dependent. Generalise – ask the right questions, and find the right answers.

(p) Show, by studying $\mathrm{E}\,N$ for $\alpha = 2$, that $\prod_{p\ \mathrm{prime}}(1 + 1/p) = \infty$, and deduce from this that $\sum_{p\ \mathrm{prime}} 1/p = \infty$. This was first proven by Euler.

(q) Let $M = \max\{j\colon X_j \ge 1\}$ be the last prime factor present in the random N. FInd the probability distribution of M, and show that it has expected value

$$\sum_{m=1}^{\infty}\Big[1 - \prod_{j=m}^{\infty}\Big(1 - \frac{1}{p_j^{\alpha}}\Big)\Big].$$

(r) Let f and g be functions definde on the natural numbers. Define the *Dirichlet convolution* or *Dirichlet product* $f * g$ by

$$(f * g)(n) = \sum_{d|n} f(d)g(n/d), \quad n \ge 1,$$

with the sum taken over those d in $\{1, \ldots, n\}$ which are factors in n. Show that

$$\sum_{n=1}^{\infty}\frac{f(n)}{n^{\alpha}}\sum_{n=1}^{\infty}\frac{g(n)}{n^{\alpha}} = \sum_{n=1}^{\infty}\frac{(f * g)(n)}{n^{\alpha}}, \quad \text{or} \quad \mathrm{E}\,(f * g)(N) = \zeta(\alpha)\,\mathrm{E}\,f(N)\,\mathrm{E}\,g(N),$$

if the two series converge.

(s) Let $\sigma(n)$ be the number of d in $\{1, \ldots, n\}$ which are factors in n. Show that $\sum_{n=1}^{\infty}\frac{\sigma(n)}{n^{\alpha}} = \zeta(\alpha)^2$; (i) by working with $\mathrm{E}\,\sigma(N)$, (ii) by Dirichlet convolution.

(t) Let $\phi(n)$ be the so-called *Euler totient function*, defined as the number of numbers in $\{1, \ldots, n\}$ which are reciprocally prime with n. It is an important tool in mathematical number theory. Show that $\phi(p) = p - 1$ if p is a prime; that more generally $\phi(p^x) = p^x - p^{x-1}$ if p is a prime; that the function is so-called multiplicative, which means that $\phi(mn) = \phi(m)\phi(n)$ for reciprocally primeish numbers; that $n = \sum_{d|n}\phi(d)$; that $(1 * \phi)(n) = n$; and that $\phi(n) = n\prod_{p|n}(1 - 1/p)$. Prove the formulae

$$\sum_{n=1}^{\infty}\frac{\phi(n)}{n^2} = \infty, \quad \sum_{n=1}^{\infty}\frac{\phi(n)}{n^{\alpha}} = \frac{\zeta(\alpha - 1)}{\zeta(\alpha)};$$

(1) by working with $\mathrm{E}\,\phi(N)$, (2) by working with $\mathrm{E}\,\phi(N)/N$; (3) by using Dirichlet convolutions.

(u) Another number theoretic function of importance is the *Möbius function*, defined by $\mu(1) = 1$; $\mu(p_{j_1}\cdots p_{j_r}) = (-1)^r$ if the number is over distinct prime numbers; and $\mu(n) = 0$ for all other n. Show that $\mu(n) \neq 0$ only for the modest numbers studied in point (m). Prove the glamorous formula

$$\sum_{n=1}^{\infty}\frac{\mu(n)}{n^{\alpha}} = \frac{1}{\zeta(\alpha)}, \quad \text{or} \quad \sum_{n=1}^{\infty}\frac{1}{n^{\alpha}}\sum_{n=1}^{\infty}\frac{\mu(n)}{n^{\alpha}} \equiv 1,$$

by working with the mean of the random $\mu(N)$ in a couple of different ways. This point may also be solved by conditioning a zeta distribution on the event that the outcome is modest; check point ($\sqrt{\pi}$).

(v) It follows without too much efforts that $\lim_{\alpha\to 1}\sum_{n=1}^{\infty}\frac{\mu(n)}{n^\alpha} = 0$; mathematical finesse is however called for to really prove that $\sum_{n=1}^{\infty}\frac{\mu(n)}{n} = 0$. Attempt to come up with such finesse. Then attempt to attach The Fundamental Prime Number Theorem, which says that if $\pi(x)$ is the number of primes in $\{1, 2, \ldots, x\}$, then $\pi(x) \doteq x/\log x$. [One may prove that this implies and is implied by $\sum_{n=1}^{\infty}\frac{\mu(n)}{n} = 0$; see Amitsur's 'On arithmetic functions' in *Journal of Analytic Mathematics*, 1956.]

(w) Time has come to introduce the *von Mangholdt function*, definde by $\Lambda(n) = \log p$ for prime potens numbers $n = p^x$ for $x \geq 1$, and $\Lambda(n) = 0$ for all numbers not being prime potenses. Work with $\mathrm{E}\,\Lambda(N)$ and show that

$$\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^\alpha} = \sum_{p \text{ primtall}}\frac{\log p}{p^\alpha - 1};$$

(x) and show that

$$\sum_{p \text{ primtall}}\frac{\log p}{p^\alpha - 1} = \sum_{n=1}^{\infty}\frac{\log n}{n^\alpha}\Big/\sum_{n=1}^{\infty}\frac{1}{n^\alpha} = \frac{-\zeta'(\alpha)}{\zeta(\alpha)},$$

by working with $\log N$. Prove also that $(1 * \Lambda)(n) = \log n$.

(y) Find a numerical value for B, the Viggo Brun constant. [Answer: 1.90216054 ...]

(z) Let N_1 and N_2 be independent and zeta distributed with the same parameter α. Find the distribution for the product $N_1 N_2$.

(æ) If $n_1, \ldots, n_k$ are given numnbers, let $\gamma\{n_1, \ldots, n_k\}$ be their greatest common divisor; for instance, $\gamma\{20, 30\} = 10$ and $\gamma\{18, 24, 36\} = 6$. If N_1 and N_2 are independent and zeta distributed with parameters α_1 and α_2, show that $\gamma\{N_1, N_2\}$ becomes zeta distributed with parameter $\alpha_1 + \alpha_2$. Generalise.

(ø) Find also the probability distribution for $\lambda\{N_1, N_2\}$, the smallest common multiplum for N_1 and N_2, when $\alpha_1 = \alpha_2$. [The answer is more complicated than for $\gamma\{N_1, N_2\}$.]

(å) Back to semi-reality, or perhaps pseudo-reality, for a little while: The zeta distribution has been applied in certain loinguistic studies; it has e.g. been tenatively shown that the frequency of words, in long text corpora, to a certain degree of accuracy follows a zeta distribution. Assume you read V words by Shakespeare, that V_1 words are seen only once, that V_2 words are seen precisely twice, etc. Then the relative frequencies V_n/V should be fitted to the zeta model's $\zeta(\alpha)^{-1}/n^\alpha$. Estimate α for a few of your favourite authors. Who has the lowest α, Anne-Catharine Vestly or Knud Pedersen Hamsun? – The zeta distribution is also partly like a discretised Pareto distribution, and will perhaps fit sufficiently well to distributions of income in different socio-economic groups. Try it out, for a group you know.

(ß) Assume $N_1, \ldots, N_k$ are independent numbers drawn from the zeta distribution with parameter α. Show that the geometric mean $(N_1 \cdots N_k)^{1/k}$ is sufficient and complete. Explain how you can find the maximum likelihood estimator.

(ő̃o) I have simulated 25 realisations from a zeta distribution, using a simple R programme, and found

```
2, 3, 3, 1, 8, 1, 1, 1, 3, 1, 12, 29,
1, 37, 10, 2, 5, 1, 1, 6, 10, 1, 4, 1, 6.
```

Only I know the value of α being used. Estimate this value, and give a confidence interval.

(ă̈) Show that the maximum likelihood estimator is strongly consistent, and find its limit distribution.

(ç̈) Show that every even number (except 2) can be expressed as a sum of two primes, e.g. by studying the behaviour of an analytic continuation of the zeta function near zero.

($\sqrt{\pi}$) Let us attempt another type of distributions for the X_j than the geomtric ones. Let X_j be 0 or 1, with probabilities $1 - a_j$ and a_j. Then N is accordingly a random modest number (see point (m)). Show that N is welldefined if and only of $\sum_{j=1}^{\infty} a_j < \infty$. Show that if a_j is taken to be $1/(p_j^{\alpha} + 1)$, then $\Pr\{N = n\} = B(\alpha)^{-1}/n^{\alpha}$, for modest n. Show again that $B(\alpha) = \prod_{p\ \mathrm{prime}}(p^{\alpha} + 1)/p^{\alpha} = \zeta(\alpha)/\zeta(2\alpha)$. Show that this model may be characterised as the conditional zeta distribution given that N is modest, and, alternatively, as the conditional zeta distribution given that all the geometric X_j are in $\{0, 1\}$. Do a little formula excursion by finding expressions for natural quantities in two ways; in one way, working with the N distribution directly, in another way, using the X_j distributions. You may e.g. impress yourself by showing

$$\sum_{n\ \mathrm{modest}} \frac{\log n}{n^{\alpha}} = \frac{\zeta(\alpha)}{\zeta(2\alpha)} \sum_{p\ \mathrm{primtall}} \frac{\log p}{p^{\alpha} + 1},$$

and your surroundings by proving

$$\Pr\Big\{\sum_{j=1}^{\infty} \mathrm{Bin}\{1, 1/(1 + p_j^2)\} \text{ becomes even}\Big\} = 0.70.$$

[Consider $\mathrm{E}\,\mu(N)$.]

(ő̃ı) Then try out Poisson distributed prime number exponents. Say that N is Poisson prime number exponentially distributed with parameters $\{d_1, d_2, d_3, \ldots\}$ provided $X_j \sim \mathrm{Pois}(d_j)$, where these are still independent. Let in particular $d_j = d/p_j^{\alpha}$, and show that

$$\Pr\{N = n\} = e^{-dA(\alpha)} \frac{d^{s(n)}}{n^{\alpha} g(n)}, \quad n = 1, 2, 3, \ldots,$$

where $s(n) = \sum_{j=1}^{m} x_j$ and $g(n) = \prod_{j=1}^{m} x_j!$, for given n with factorisation as in (c), and where $A(\alpha) = \sum_{p\ \mathrm{primtall}} 1/p^{\alpha}$. Show, for example, that

$$\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}} \frac{1}{g(n)} = \exp\{A(\alpha)\}, \quad \sum_{n=1}^{\infty} \frac{\log n}{n^{\alpha} g(n)} = \exp\{A(\alpha)\} \sum_{p\ \mathrm{primtall}} \frac{\log p}{p^{\alpha}}.$$

Show that the probability of having a prime number for N is $A(\alpha)\exp\{-A(\alpha)\}$ when $d_j = 1/p_j^{\alpha}$. Find some further formulae in the flow created. Show that products of independent Poisson prime number exponentially distributed variables stay being Poisson prime number exponentially distributed. Find a sufficient and complete statistic based on $N_1, \ldots, N_k$ when d and α are unknown parameters. Study the large-sample properties of the maxium likelihood estimators.

(γ) We know that $\prod_p p^2/(p^2-1) = \pi^2/6$, but what is $\prod_p p^2/(p^2-0.99)$? – Allow me to show you my *gemneralised zeta function*:

$$\zeta_d(\alpha) = \sum_{n=1}^{\infty} \frac{d^{s(n)}}{n^\alpha}, \quad 0 < d \le 2,\ \alpha > 1,$$

where $s(n) = x_1 + x_2 + \cdots$ is the *extravaganza* for the number n. Show taht this de facto exists for $0 < d \le 2$ and $\alpha > 1$. Give probabilistic proofs for the following formulae, which all reduce to previous results when d is set equal to 1:

$$\zeta_d(\alpha) = \prod_{p \text{ primtall}} \frac{p^\alpha}{p^\alpha - d},$$

$$\sum_{n=1}^{\infty} \frac{d^{s(n)}\mu(n)}{n^\alpha} \sum_{n=1}^{\infty} \frac{d^{s(n)}}{n^\alpha} \equiv 1,$$

$$\sum_{n=1}^{\infty} \frac{d^{s(n)}\sigma(n)}{n^\alpha} = \zeta_d(\alpha)^2,$$

$$\sum_{n \text{ beskjeden}} \frac{d^{s(n)}}{n^\alpha} = \prod_{p \text{ primtall}} \frac{p^\alpha + d}{p^\alpha} = \frac{\zeta_d(\alpha)}{\zeta_{d^2}(2\alpha)},$$

$$\sum_{n=1}^{\infty} \frac{d^{s(n)}\phi(n)}{n^\alpha} = \frac{\zeta_d(\alpha-1)}{\zeta_d(\alpha)},$$

$$\sum_{n=1}^{\infty} \frac{d^{s(n)}f(n)}{n^\alpha} \sum_{n=1}^{\infty} \frac{d^{s(n)}h(n)}{n^\alpha} = \sum_{n=1}^{\infty} \frac{d^{s(n)}(f*h)(n)}{n^\alpha},$$

$$\sum_{n=1}^{\infty} \frac{d^{s(n)}\log n}{n^\alpha} = \zeta_d(\alpha) \sum_{n=1}^{\infty} \frac{d^{s(n)}\Lambda(n)}{n^\alpha},$$

$$\Pr\Big\{\sum_{j=1}^{\infty} \mathrm{Bin}\{1, d/(p_j^\alpha + d)\} \text{ becomes even}\Big\} = \frac{1}{2} + \frac{1}{2}\frac{\zeta_{d^2}(2\alpha)}{\zeta_d(\alpha)^2}.$$

Employ as probabilitistic tools (1) $X_j \sim \mathrm{Poisson}(d/p_j^\alpha)$; (2) $X_j \sim \mathrm{Bin}\{1, d/(p_j^\alpha + d)\}$; (3) $X_j \sim \mathrm{Geo}(1 - d/p_j^\alpha)$. Discuss relations between these models.

(œ) Investigate consequences for the distribution of primes among the natural numbers, from $\sum_{n=1}^{\infty} d^{s(n)}\mu(n)/n = 0$; as mentioned this statement, for the special case of $d = 1$, implies the glorious prime number distribution theorem.

(α) Put a probability distribution on the modest numers by taking the X_j to form a time inhomogeneous Markov chain on $\{0, 1\}$. Grei ut.

(ω) Find out a wholde deal on how the prime numbers and their cousins are distributed among the natural numbers, by studying distributions of the type $\mathcal{D}\{N | N \le n_0\}$, where n_0 is big, and by moving this threshold for the α parameter to the left of 1. Meld fra hvor du går.

40. Quartile and quantile differences

One way of assessing the spread of a distribution F, based on data $X_1, \ldots, X_n$, is via the *quartile difference* $Q_3 - Q_1$, the difference between the upper and lower quartiles. Often this difference is multiplied with a well chosen constant, such that the resulting spread estimate becomes approximately unbiased for the the standard deviation parameter in the case of F being normal.

What is this constant? How clever is this estimator, compared with the usual one under normal conditions? Which cons and pres does the estimator have, compared to others? How do yet other naturally generalised competitors behave, where one uses upper and lower ε quantile, instead of upper and lower 25 percent quantiles? Which of these is best, on Gauß's home turf?

(a) Attempt to make your own exam type exercise, containing progressively more detailed questions, based on the above sentences.

(b) Define $Q_3 = X_{[0.75\,n]}$ and $Q_3 = X_{[0.25\,n]}$, where $X_{(1)} < \cdots < X_{(n)}$ are the order statistics. Speculate a little regarding suitable interpolation tricks to make them better.

(c) For a few of the points below we shall take F to be the normal $\mathrm{N}(\xi, \sigma^2)$. Assume for this point only that F is strictly increasing with a continuous density f. Show that $Q_3 - Q_1$ converges almost surely to $q_3 - q_1 = F^{-1}(0.75) - F^{-1}(0.25)$. With which constant do we need to multiply $Q_3 - Q_1$ in order to get a consistent estimator of σ, in the case where F is a normal?

(d) Show that

$$\begin{pmatrix} \sqrt{n}(Q_1 - q_1) \\ \sqrt{n}(Q_3 - q_3) \end{pmatrix} \to_d \begin{pmatrix} (F^{-1})'(0.25)\, U \\ (F^{-1})'(0.75)\, V \end{pmatrix},$$

where

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathrm{N}_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3/16 & 1/16 \\ 1/16 & 3/16 \end{pmatrix}.$$

(e) Let $z(\varepsilon) = \Phi^{-1}(1 - \varepsilon)$ be the upper ε quantile for the standard normal, and let

$$\widetilde{\sigma} = \frac{Q_3 - Q_1}{2z(0.25)} = \frac{Q_3 - Q_1}{1.349}.$$

Show that $\sqrt{n}(\widetilde{\sigma} - \sigma)$ tends to $\mathrm{N}(0, \kappa^2)$, with $\kappa = 1.1664\,\sigma$.

(f) Here it is natural to compare with the traditional estimator $\widehat{\sigma}$, the empirical standard deviation. Show (which is more standard, right?) that $\sqrt{n}(\widehat{\sigma} - \sigma) \to_d \mathrm{N}(0, (0.7071\,\sigma)^2)$.

(g) Then generalise! That is, consider

$$\widetilde{\sigma}(\varepsilon) = \frac{X_{[(1-\varepsilon)n]} - X_{[\varepsilon n]}}{2z(\varepsilon)} = \frac{F_n^{-1}(1 - \varepsilon) - F_n(\varepsilon)}{2z(\varepsilon)},$$

where F_n is the empirical cumulative distribution function, and find the limit distribution for $\sqrt{n}(\widetilde{\sigma} - \sigma)$ under normal conditions. The answer should becomes $\mathrm{N}(0, \kappa(\varepsilon)^2)$, where

$$\kappa(\varepsilon) = \frac{\sqrt{2\pi}}{2\varepsilon}\sqrt{2\varepsilon(1-\varepsilon)}\exp\{\tfrac{1}{2}z(\varepsilon)^2\}\,\sigma.$$

(h) Investigate how the precision of $\widetilde{\sigma}(\varepsilon)$ changes when ε varies between 0 and $\frac{1}{2}$. Show in particular that the asymptotically speaking very best estimator of this type, under normality, is

$$\sigma^* = \frac{F_n^{-1}(0.931) - F_n^{-1}(0.069)}{2.9666},$$

with limit distribution $\mathrm{N}(0, (0.8755\,\sigma)^2)$, a loss of 1.2382 compared with the optimal value $\sigma/\sqrt{2}$.

(i) Investigate the behavious of such estimators outside normality.

41. something

well

42. something

well

43. something

well

References

Amitsur (1956). On arithmetic functions. *Journal of Analytic Mathematics.*

von Bahr, B. (1965). [xx ... xx] *Annals of Mathematical Statistics.*

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

Clauset, A. (2017). The enduring threat of a large interstate war. Technical Report, One Earth Foundation.

Clauset, A. (2018). Trends and fluctuations in the severity of interstate wars. *Science Advances* **4**, xx-xx.

Cunen, C., Hjort, N.L., and Nygård, H. (2019). Statistical sightings of better angels. *Journal of Peace Research* [to appear].

Doxiadis, A.K. (1992). *Uncle Petros and Goldbach's Conjecture: A Novel of Mathematical Obsession.*

Ferguson. [xx the book xx]

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Annals of Statistics* **2**, 615–629.

Ferguson, T.S. and Klass, M.J. (1972). A representation of independent increment processes without Gaussian components. *Annals of Mathematical Statistics* **43**, 1634–1643.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference.* Cambridge University Press, Cambridge.

Ghosh, A. (2017). Robust inference under the Beta regression model with application to health care studies. `arXiv`

Gould, S.J. [xx Adam's Navel. xx]

Heger, A. (2007). Jeg og jordkloden. Dagsavisen.

Hjort, N.L. and Fenstad, G.U. (1992). *Annals.*

Hjort, N.L. (1976). *The Dirichlet Process Applied to Some Nonparametric Problems.* Cand. real. thesis [in Norwegian], Department of Mathematics, Nordlysobservatoriet, University of Tromsø.

Hjort, N.L. (1985). Discussion contribution to P.K. Andersen and Ø. Borgan's 'Counting process models for life history data: A review'. *Scandinavian Journal of Statistics* **12**, xx–xx.

Hjort, N.L. (1985). An informative Bayesian bootstrap. Technical Report, Department of Statistics, Stanford University.

Hjort, N.L. (1986). Discussion contribution to P. Diaconis and D. Freedman's paper 'On the consistency of Bayes estimators'. *Annals of Statistics* **14**, 49–55.

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

Hjort, N.L. (1991). Bayesian and empirical Bayesian bootstrapping. Statistical Research Report, Department of Mathematics, University of Oslo.

Hjort, N.L. (2003). Topics in nonparametric Bayesian statistics [with discussion]. In *Highly Structured Stochastic Systems* (eds. P.J. Green, N.L. Hjort, S. Richardson). Oxford University Press, Oxford.

Hjort, N.L. (2018). Towards a More Peaceful World [Insert '!' or '?' Here]. FocuStat Blog Post.

Hjort, N.L. (2010). An invitation to Bayesian nonparametrics. In *Bayesian Nonparametrics* (by Hjort, N.L., Holmes, C.C., Müeller, P., and Walker, S.G.), 1–21.

Hjort, N.L., Holmes, C.C., Müller, P., and Walker, S.G. (2010). *Bayesian Nonparametrics.* Cambridge University Press, Cambridge.

Hjort, N.L. and Kim, Y. (2013). Beta processes and their applications and extensions. Statistical Research Report, Department of Mathematics, University of Oslo.

Hjort, N.L. and Ongaro. A. (2005). Exact inference for random Dirichlet means. *Statistical Inference for Stochastic Processes* **8**, 227–254.

Hjort, N.L. and Ongaro. A. (2006). On the distribution of random Dirichlet jumps. *Metron* **LXIV**, 61–92.

Hjort, N.L. and Petrone, S. Nonparametric quantile inference using Dirichlet processes. In *Festschrift for Kjell Doksum* (ed. V. Nair).

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

Lehmann, E.L. (1951). Notes on the Theory of Point Estimation. (Mimeographed by C. Blyth.) Department of Statistics, University of Berkeley, California.

Müller, O., Quintana, F.A., Jara, A., and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis.* Springer-Verlag, Berlin.

Ottosen, K. *Theta Theta.*

Ottosen, R. *ML.*

Rubin, D. (1981). The Bayesian bootstrap. *Annals of Statistics* **9**, 130–134.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Sethuraman, J. and Tiwari, R. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In: *Proceedings of the Third Purdue Symposium on Statistical Decision Theory and Related Topics* (eds. S.S. Gupta and J. Berger), 305–315. Academic Press, New York.

Stoltenberg, E.Aa. and Hjort, N.L. (2019a). Simultaneous estimation of Poisson parameters. *Journal of Multivariate Analysis*, in its way.

Stoltenberg, E.Aa. and Hjort, N.L. (2019b). Modelling and analysing the Beta- and Gamma Police Tweetery data. [Manuscript, in progress.]

Wolpert, R.L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.