# Exercises and Lecture Notes
# STK 4090, Spring 2020

Version 0.15, 24-Aug-2020

**Nils Lid Hjort**

**Department of Mathematics, University of Oslo**

**Abstract**

These are Exercises and Lecture Notes for the course Confidence, Likelihood, Probability, STK 4180 (Master level) or STK 9180 (PhD level), for the autumn semester 2020. I'll add on more exercises as the course progresses.

### 1. The probability transform

Some of the following facts are related to various operations for confidence distributions and confidence curves.

(a) Suppose $X$ has a continuous and increasing cumulative distribution function $F$, i.e. $F(x) = \Pr\{X \le x\}$. Show that $U = F(X)$ is uniform on the unit interval. Any continously distributed random variable can hence be transformed to uniformity, via this *probability transform.*

(b) Show that also $U_2 = 1 - F(X)$ and $U_3 = |1 - 2\,F(X)|$ have uniform distributions.

(c) Simulate a million copies of $x_i \sim \mathrm{N}(0,1)$, and check the histogram of $\Gamma_1(x_i^2)$, where $\Gamma_\nu$ is the cumulative distribution function of a $\chi^2_\nu$. Comment on what you find.

(d) Suppose $\widehat\theta$ is an estimator for the real parameter $\theta$, based on data $y$, with some continuous distribution function $K_\theta(x) = \Pr_\theta\{\widehat\theta \le x\}$; we are in particular assuming that the distribution of $\widehat\theta$ depends only on $\theta$, not on other aspects of the underlying model employed. Consider the construction

$$C(\theta, y_{\rm obs}) = \Pr_\theta\{\widehat\theta \ge \widehat\theta_{\rm obs}\} = 1 - K_\theta(\widehat\theta_{\rm obs}),$$

a curve that can be computed and plotted post-data, where $\widehat\theta_{\rm obs} = \widehat\theta(y_{\rm obs})$ is the observed estimate. Show that it has the property that the random $C(\theta, Y)$ is uniformly distributed, for each fixed $\theta$.

### 2. CD and cc for the normal standard deviation

Read Cunen and Hjort's *Confidence Curves for Dummies* (2020), a FocuStat Blog Post. Then do the details, regarding mathematics and implementation, for their introductory meant-to-be-simple example:

"Here's a simple example. You observe the data points 4.09, 6.37, 6.87, 7.86, 8.28, 13.13 from a normal distribution and wish to assess the underlying spread parameter, the famous standard deviation $\sigma$. We'll now introduce you to as many as two (2) curves: the confidence curve $\mathrm{cc}(\sigma)$ and the confidence distribution $C(\sigma)$. They're close cousins, actually, and it's not the case that both curves need to be displayed for each new statistical application."

(a) Here you might start with the classic fact concerning the empirical variance that $\widehat\sigma^2 \sim \sigma^2\chi^2_m/m$, where $m = n-1$, with $n$ the sample size. Then deduce that

$$C(\sigma, y_{\rm obs}) = \Pr_\sigma\{\widehat\sigma \ge \widehat\sigma_{\rm obs}\} = 1 - \Gamma_m(m\widehat\sigma^2_{\rm obs}/\sigma^2).$$

Here $y_{\rm obs}$ represents the observed data, and $\widehat\sigma_{\rm obs}$ the observed point estimate. Show that $C(\sigma, Y) \sim \mathrm{unif}$, where $Y$ represents a random data set $Y_1, \ldots, Y_n$, from the $\sigma$ in question. In particular, the distribution of $C(\sigma.Y)$ does not depend on $\sigma$.

(b) Reproduce versions of Cunen and Hjort's Figures A and B, with the confidence curve $\mathrm{cc}(\sigma)$, the CD $C(\sigma)$, the median confidence estimate, etc.

(c) Compute also the *confidence density* $c(\sigma, y_{\rm obs})$ associated with the CD. Compute also its mode, say $\sigma^*$, and briefly assess its properties as an estimator of $\sigma$.

(d) A Bayesian approach to the same problem, i.d. finding a posterior distribution for $\sigma$, is to start with a prior $\pi(\sigma)$ and then compute $\pi(\sigma\,|\,y_{\rm obs}) \propto \pi(\sigma)g(\widehat\sigma, \sigma)$, where $g(\widehat\sigma, \sigma)$ is the likelihood, here the density function for $\widehat\sigma$ as a function of $\sigma$. When does such a Bayesian approach agree with the confidence density?

(e) Suppose there are two independent normal samples, with standard deviations $\sigma_1$ and $\sigma_2$. Construct a CD for $\rho = \sigma_1/\sigma_2$. Invent a second simple small dataset, to complement the first dataset given above, and then compute and display the confidence curve $\mathrm{cc}(\rho, \mathrm{data})$.

### 3. An often (but not always) useful CD construction

In Exercise 1 we saw that the simple construction $C(\theta, y) = \Pr_\theta\{\widehat\theta \ge \widehat\theta_{\rm obs}\}$ gives a CD, in the case of one-dimensional setups with a well-defined estimator $\widehat\theta$.

(a) More generally, assume $Y_1, \ldots, Y_n$ come from some distribution, depending on a single parameter $\theta$, and that $Z$ is a statistic with mean increasing in $\theta$. Then study $C(\theta, y) = \Pr_\theta\{Z \ge z_{\rm obs}\}$. Show that this is a bona fide CD.

(b) Show also that the construction works, if there are other parameters at play too, as long as the distribution of the chosen $Z$ only depends on $\theta$. Go through the details for the case of the $Y_i$ being $\mathrm{N}(\mu, \sigma^2)$, with $Z = \sum_{i=1}^n (Y_i - \bar Y)^2$, and also for $Z' = \sum_{i=1}^n |Y_i - M_n|$, where $M_n$ is the empirical median. Compute, display, compare both CDs, based on $Z$ and on $Z'$, for the simple dataset of Exercise 2 (with $n = 6$). For the $Z$ case, there is a formula, but for the $Z'$ case you would need simulation, for a grid of $\sigma$ values.

(c) For a normal sample from $\mathrm{N}(\mu, \sigma)$, we see that several $\Pr_{\mu,\sigma}\{Z \ge z_{\rm obs}\}$ schemes work, in that the $Z$ in question has a distribution depending on $\sigma$, but not $\mu$. Attempt to work with $C^*(\mu, y) = \Pr_{\mu,\sigma}\{\bar Y \ge \bar y_{\rm obs}\}$ ... and show that it will not really work (unless $\sigma$ is known).

(d) But of course there *are* natural CD constructions for $\mu$ here. What is needed is a *pivot*, say $A = \text{piv}(\mu, y)$, a function binding the focus parameter and data together in a way which makes its distribution not depend on the parameters. Study indeed

$$t_n = t_n(\mu, Y) = \frac{\bar{Y} - \mu}{\widehat{\sigma}/\sqrt{n}},$$

with $\widehat{\sigma}^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ the classical empirical variance. Pretend that you in all your cleverness has not seen this $t_n$ before, and are unaware of its relation to a t distribution – but show that the distribution of $t_n$, call it $K_n$, does not depend on $(\mu, \sigma)$.

(e) Then show that $C(\mu, y_{\text{obs}}) = K_n(t_n(\mu, y_{\text{obs}}))$ is a CD for $\mu$. Even if you do not see the connection to the classic t of Student (1908), you may still carry through this, by simulating $B = 10^5$ realisations of $t_n$, and use

$$C(\mu, y_{\text{obs}}) = K_n^*(t_n(\mu, y_{\text{obs}})) = \frac{1}{B}\sum_{j=1}^{B} I\{t_{n,j} \le t_n(\mu, y_{\text{obs}})\}.$$

But show that by all means $K_n$ is a $t_m$, with $m = n-1$, so the canonical CD for $\mu$ is and remains $C(\mu, y_{\text{obs}}) = G_m(\sqrt{n}(\mu - \bar{y}_{\text{obs}})/\widehat{\sigma}_{\text{obs}})$, with $G_m$ the cdf for the $t_m$.

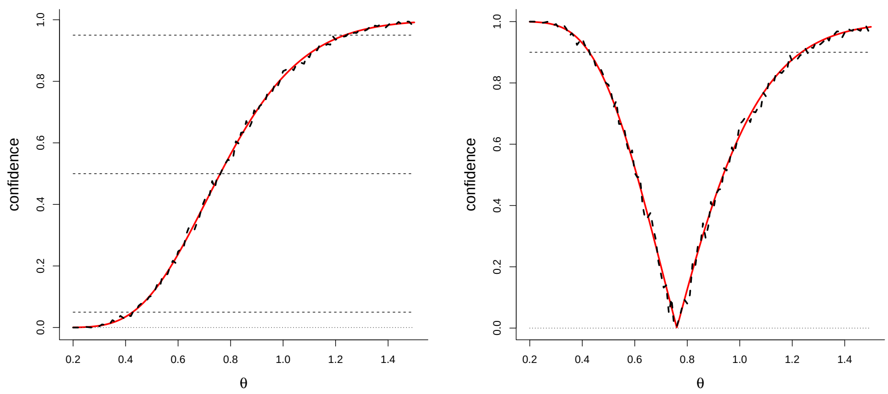


Figure 0.1: Left panel: confidence distribution $C(\theta)$, via simulations (black and wiggly curve) and via exact calculations (red and smooth curve); right panel: the two versions of the associated confidence curve $\text{cc}(\theta)$.

### 4. A skewed distribution on the unit interval

Consider the model $F(y, \theta) = y^\theta$ for observations on $[0, 1]$, where $\theta$ is an unknown positive parameter.

(a) Write down the log-likelihood function and find a formula for the maximum likelihood estimator $\widehat{\theta}$.

(b) Use theory of CLP, Chapter 2, to write down a normal approximation to the distribution of $\widehat{\theta}$.

(c) Consider the data set

`0.013 0.054 0.234 0.286 0.332 0.507 0.703 0.763 0.772 0.920`

Estimate $\theta$ and compute the confidence distribution $C(\theta) = \Pr_\theta\{\widehat\theta \geq \widehat\theta_{\rm obs}\}$, along with the confidence curve ${\rm cc}(\theta) = |1 - 2\,C(\theta)|$, (i) using simulations, (ii) using exact probability calculus. Reproduce a version of Figure 0.1.

(d) Supplement these two curves with approximations based (i) on the normal approximation for $\widehat\theta$ and (ii) on the chi-squared approximation for the deviance.

### 5. 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 $y\,|\,N \sim {\rm Bin}(N, p)$, with $p = 0.514$ (a good point estimate for today's overall figure for human reproduction). So the data is that $y = 6$, and we can attempt confidence inference for $N$. The questions below expand on those given in CLP, Example 3.11.

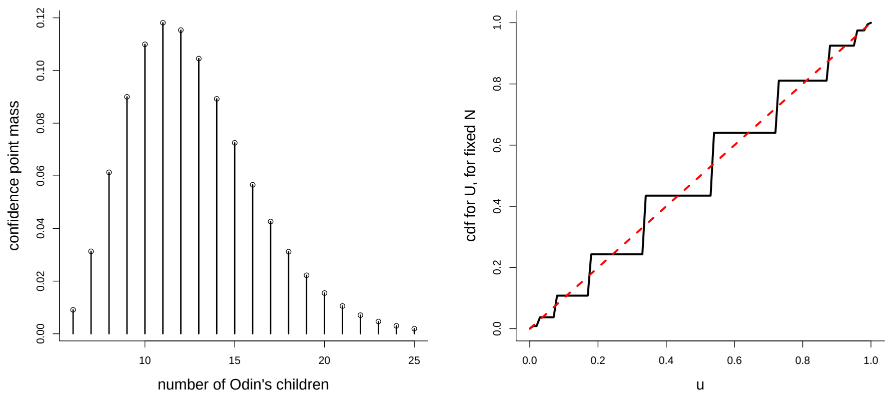


Figure 0.2: Left panel: confidence point masses $c(N, y)$, for $N \geq 6$; right panel: for a fixed $N = 14$ (in this example), the empirical cumulative distribution for $U = C(N, Y)$, with $Y \sim {\rm Bin}(N, p)$.

(a) A natural construction for a CD is

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

a version of the general method of Exercise 1(e), but with so-called half-correction for the discreteness. Compute and display this CD, and take differences to compute also the confidence point masses, $c(N, y)$. Construct a version of Figure 0.2, left panel.

(b) For the sport of it, carry out a Bayesian analysis too. Start with a reasonable prior $\pi(N)$ (formed before you read in a book that $y = 6$), compute the posterior distribution $\pi(N\,|\,y = 6)$, and compare with the CD analysis.

(c) A CD $C(\theta, y)$, for a parameter $\theta$ based on data $y$, should ideally have the uniformity property that $U = C(\theta_0, Y)$ has the uniform distribution, for any fixed $\theta_0$, 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 $y$. Construct a verion of Figure 0.2, right panel; here I took $N_0 = 14$, simulated say $10^4$ realisations of $U = C(N_0, Y)$, and computed the empirical distribution function $\Pr\{U \leq u\}$. Comment on your findings.

(d) 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.

## 6. Guess my range

I've simulated these points in my computer, from a uniform distribution over $[a, b]$, and I've ordered them, for simplicity. But I won't tell you the values I used for $a$ or $b$, or indeed the range $\delta = b - a$. Your task will be to make inference about the $\delta$ – with a CD, a cc, and a median confidence estimate.

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

(a) 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

$$R_n = Y_{(n)} - Y_{(1)} = \delta R_{n,0}, \quad \text{with } R_{n,0} = U_{(n)} - U_{(1)},$$

relating the range of data naturally to the rannge of a uniform sample.

(b) Explain that $R_n/\delta$ is a pivot (as defined in Exercise 3). Simulate say $10^5$ realisations in your computer from this distribution, say $G_n$.

(c) Show that

$$C(\delta, y) = \Pr_{a,b}\{R_n \geq R_{n,\text{obs}}\} = 1 - G_n(R_{n,\text{obs}}/\delta) \quad \text{for } \delta > R_{n,\text{obs}}$$

is a CD for $\delta$. Compute it, using your simulations from $G_n$, and display as many as three curves: the CD $C(\delta, y_{\text{obs}})$; the $\text{cc}(\delta, y) = |1 - 2\,C(\delta, y_{\text{obs}})|$; and the confidence density $c(\delta, y_{\text{obs}})$. Also find the median confidence and maximum confidene point estimates Comment on your findings.

## References


Cunen, C. and Hjort, N.L. (2020). Confidence Curves for Dummies. FocuStat Blog Post, April 2020.

Hermansen, G., Stoltenberg, E.Aa., and Cunen, C. (2017). Bokmelding: Confidence, Likelihood, Probability: Statistical Inference With Confidence Distributions (Schweder og Hjort, CUP, 2016). FocuStat Blog Post, November 2017.

Hjort, N.L. and Schweder, T. (2018). Confidence distributions and related themes. [General introduction article to a Special Issue of the Journal of Statistical Planning and Inference dedicated to this topic, with eleven articles, and with Hjort and Schweder as guest editors; vol. 195, 1-13.

Schweder, T. (2017). Bayesian Analysis: Always and Everywhere? Confidence curves for dummies. FocuStat Blog Post, November 2017.

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

Schweder, T. and Hjort, N.L. (2016). *Confidence, Likelihood, Probability.* Cambridge University Press, Cambridge.