Using a probability distribution to characterize uncertainty is at the core of statistical inference. So, it seems natural to try to summarize the information about the parameters in statistical models with probability distributions. R. A. Fisher thought so. In fact, he expended a great deal of effort over more than thirty years, and put his professional reputation on the line trying to do so, with only limited success. Fisher's central difficulty was that, in the Frequentist tradition to which he was committed, parameters are not random variables. They are fixed and immutable constituents of the statistical models describing the behavior of populations, which we must estimate because we generally only have access to samples from populations, not to the full populations themselves. Now Bayesians, of course, characterize parameters with probability distributions from the get-go. Parameters are given prior distributions and combined with the likelihood function generated by the data to produce posterior distributions that characterize the parameters. Fisher wanted the posterior distributions without having to assume the priors. This was a key motivating idea for his work on Fiducial probability.

A few statisticians apparently quietly worked on this program throughout the twentieth century, even though Fisher's Fiducial ideas were mostly forgotten and not part of the mainstream statistical current. D. R. Cox (Cox (1958)), for example, pioneered the idea of constructing *confidence distributions* from confidence intervals, and Bradley Efron (Efron (1998)) expressed great optimism that Fisher's work in this area would become important in the twenty-first century. (Efron's paper is a masterpiece that summarizes a good bit of twentieth-century statistical research.)

Recently, however, there seems to have been a resurgence of Fisher's ideas among statisticians interested in chasing the idea of an orthodox Frequentist view of parameter distributions. The 2013 paper of Xie and Singh lays out the modern theory of confidence distributions as a fundamental idea that organizes a great deal of statistical practice. In the initial Summary, the authors write:

> . . .the concept of a confidence distribution subsumes and unifies a wide range of examples, from regular parametric (fiducial distribution) examples to bootstrap distributions, significance (p-value) functions, normalized likelihood functions, and, in some cases, Bayesian priors and posteriors.

Later in the paper, they go on to define a confidence distribution as:

> A function $H_n(\cdot) = H_n(x, \cdot)$ on $\mathbf{X} \times \Theta \to [0, 1]$ is called a confidence distribution (CD) for a parameter θ, if
> * R1) For each given $x \in \mathbf{X}$, $H_n(\cdot)$ is a cumulative distribution function on Θ;
> * R2) At the true parameter value $\theta = \theta_0$, $H_n(\theta_0) \equiv H_n(x, \theta_0)$, as a function of the sample

The simplest example of a confidence distribution I could find that is adequate to illustrate some of the key concepts comes from the book Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions by Schweder and Hjort. On page 62, the authors point out that the distribution of the p-value for the parameter θ describing the probability of success for a binomial trial can be considered as an approximate confidence distribution for θ. The distribution is approximate because the distribution is discrete and the "half-correction" is used to improve the approximation. Note that because the p-values follow uniform distributions under the null hypothesis, it should be clear that the assumptions of the definition above are satisfied.

Suppose $Y \sim Bin(n,\theta)$, then

$C(\theta) = P(Y > y_0) + .5 * P(Y = y_0)$
is a confidence distribution for θ.

To illustrate this, we consider the experiment of realizing 8 successes in 20 trials and write a short helper function.

```
CD <- function(theta,n=20,y0=8){
        1 - sum(dbinom(x = seq(from = 0, to = y0), size = n, prob = theta))
+
```

```
                 .5 * dbinom(x = y0, size = n, prob = theta)}
```

Here, we compute the CDF and plot it.

```
library(tidyverse)
library(highcharter)
conf_dist <-
    tibble(theta = seq(0, 1, by = .01)) %>%
    mutate(probability = map_dbl(theta, CD))

hchart(conf_dist, "line", hcaes(x = theta, y = probability)) %>%
hc_title(text = "Confidence Distribution for Binomial Model",
           margin = 20, align = "left",
           style = list(color = "black", useHTML = TRUE)) %>%
hc_tooltip(valueDecimals=4, valuePrefix="cum prob = ")
```

This is nice, but to really see how computing the confidence distribution might be useful, we compute and plot the confidence curve introduced by Birnbaum in his 1961 paper.

```
conf_curve <-
    conf_dist %>%
    mutate(confidence = 2 * abs(.5 - probability))

hchart(conf_curve, "line", hcaes(x = theta, y = confidence)) %>%
    hc_title(text = "Confidence Curve for Binomial Model",
             margin = 20, align = "left",
             style = list(color = "black", useHTML = TRUE)) %>%
    hc_tooltip(valueDecimals=4, valuePrefix="conf level = ")
```

Pick a point on the left branch of the curve. The y value gives you the level of confidence and the x value is the lower bound of the corresponding confidence interval. Move horizontally across to the right branch to read off the upper end of the confidence interval. So reading up and down the curve you can read off the confidence intervals for any value of confidence.

As a check, we compute the 95% confidence interval for θ using the normal approximation to the binomial.

```
ub <- 8 / 20 + (1.96 / 20) * sqrt(8 * 12 / 20)
lb <- 8 / 20 - (1.96 / 20) * sqrt(8 * 12 / 20)
cat("95% CI = [" , lb , "," , ub, "]")

## 95% CI = [ 0.1853 , 0.6147 ]
```

If you are a Bayesian, there is a really amusing side to confidence distributions. In order to be clear that what they are doing with confidence distributions is in fact different from what Bayesians do when they choose priors, the champions of confidence distributions appeal to epistomology, the study of knowledge and justified belief. On page (xiv) of their book, Schweder and Hjort write:

> The concept of confidence distribution is rather basic, but has proved difficult for statisticians to accept. The main reason is perhaps that confidence distributions represent epistemic probability obtained from the aleatory probability of the statistical model (i.e. the chance variation in nature and society), and to face both types of probability at the same time might be challenging. The traditional Bayesian deals only with subjective probability, which is epistemic when based on knowledge, and the frequentist of the Neyman-Wald school deals only with sampling variability, that is, aleatory probability.

Please indulge me while I unpack this. Aleatory probabilities are what nature and the world give us: the decay times of alpha particles, the valuable behaviors of large populations, etc. Hard-core frequentists will only allow themselves to compute aleatory probabilities. As soon as you compute a confidence distribution or even a confidence interval, you are working with epistemic probabilities: what you believe would be true under repeated sampling that may be impossible to actually carry out. But, you are justified in doing this because these epistemic probabilities are anchored in aleatory probabilities. When Bayesians base their choice of priors on rational beliefs based on plausible evidence, they are on the same epistemic footing as frequentists computing confidence intervals. When Bayesians are capricious in choosing their priors, they are not. Asking most statisticians to think about these things gives them headaches. Thus, quietly, in work that builds bridges, ends the Bayesian vs. Frequentist controversy.