Jaynesの切断指数分布モデルに関する不適切な議論

- 黒木玄

- 2022-09-28

要約: 文献 Jaynes (1976)

- E. T. Jaynes, Confidence Intervals vs Bayesian Intervals, 1976. [doi (https://doi.org/10.1007/978-94-010-1436-6_6)] [pdf (https://bayes.wustl.edu/etj/articles/confidence.pdf)]

の(b) Example 5にあるに切断指数分布モデルの下限に関する通常の信頼区間とBayesianの信用区間の不適切な比較の仕方を, 以下のリンク先で引用している:

- Frequentism and Bayesianism III: Confidence, Credibility, and why Frequentism and Science do not Mix, 2014. [html (http://jakevdp.github.io/blog/2014/06/12/frequentism-and-bayesianism-3-confidence-credibility/)]
- Alex Pizzuto and Austin Schneider, Statistical Methods for Analysis, IceCube Bootcamp Summer 2020. [pdf (https://events.icecube.wisc.edu/event/123/contributions/6441/attachments/5491/6311/Statistical_Methods_for_Analysis.pdf)]

信頼区間の側の構成の仕方がおかしいので, 非Bayesianの側に不公平な議論になってしまっている.

切断指数分布モデルにおける通常の信頼区間と平坦事前分布に関するBayesianの信用区間はぴったり一致する. ぴったり一致する区間に優劣を付けることは不可能である.

目次

In [1]:
```
using Distributions
using Optim
using QuadGK
using SpecialFunctions
using StatsPlots
default(fmt=:png, titlefontsize=10, plot_titlefontsize=10,
    guidefontsize=9, tickfontsize=6)
```

1 切断指数分布モデルのBayes統計

以下で扱う統計モデルのパラメータは実数 θ の1つである.

データ $x = (x_1, \dots, x_n)$ (n 個の実数の組)に関する切断指数分布モデルの尤度函数 $p(x|\theta)$ は次のように定義される:

$$p(x|\theta) = \prod_{i=1}^n p(x_i|\theta).$$

ここで,

$$p(x_i|\theta) = \begin{cases} 0 & (x_i < 0) \\ e^{-(x_i-\theta)} & (x_i \geq 0). \end{cases}$$

尤度函数を整理して書き直すために次のようにおく:

$$\min(x) = \min(x_1, \dots, x_n), \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i.$$

このとき,

$$p(x|\theta) = \begin{cases} 0 & (\min(x) < \theta) \\ e^{-n(\bar{x}-\theta)} & (\min(x) \geq \theta). \end{cases}$$

これは次のように書き直される:

$$p(x|\theta) = (\text{independent of } \theta) \times \begin{cases} 0 & (\min(x) < \theta) \\ ne^{-n(\min(x)-\theta)} & (\min(x) \geq \theta). \end{cases}$$

ここで

$$\int_{\min(x)}^{\infty} ne^{-n(\min(x)-\theta)}\,d\theta = 1$$

となっていることに注意せよ.

このことから, 平坦事前分布に関する以上のモデルの事後分布 $p(\theta|x)$ は次になることがわかる:

$$p(\theta|x) = \begin{cases} 0 & (\min(x) < \theta) \\ ne^{-n(\min(x)-\theta)} & (\min(x) \geq \theta). \end{cases}$$

すなわち, θ が事後分布に従う確率変数ならば

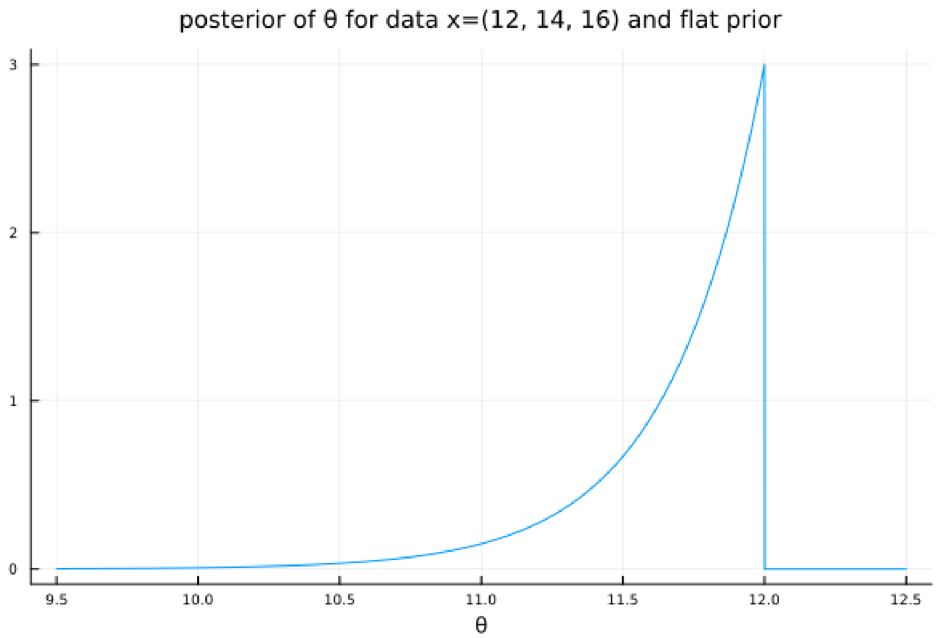
$$\theta \sim \min(x) - \mathrm{Exponential}(1/n).$$

$x = (12, 14, 16)$ の場合のこの事後分布のグラフを描くと次のセルのようになる:

In [2]:

```julia
function pdf_posterior(θ, x)
    n, t = length(x), minimum(x)
    t < θ ? float(zero(θ)) : n*exp(-n*(t - θ))
end

x = (12, 14, 16)
plot(θ -> pdf_posterior(θ, x), 9.5, 12.5; label="")
title!("posterior of θ for data x=$x and flat prior")
plot!(xguide="θ")
```

Out[2]:



$0 < \alpha \leq 1$ と仮定し, $c \geq 0$ を次の条件によって定める:

$$e^{-nc} = \alpha, \qquad c = -\frac{1}{n}\log\alpha$$

このとき, 上の事後分布の最高密度区間(HDI, highest density interval)としての90%Bayesianの信用区間(credible interval)は

$$[\min(x) - c, \min(x)]$$

になる.

$\alpha = 10\%$ のとき, $c \approx 0.7675\cdots$ なので, $x = (12, 14, 16)$ の場合には,

$$[\min(x) - c, \min(x)] \approx [11.23, 12].$$

これは次のセルに引用する文献Jaynes (1976)の結論に等しい.

it is proportional to $y^{N-1}\exp(-Ny)$ for $y>0$, where $y\equiv(\theta^*-\theta+1)$. Evidently, it will not be feasible to find the shortest confidence interval in closed analytical form, so in order to prevent this example from growing into another research project, we specialize to the case $N=3$, suppose that the observed sample values were $\{x_1, x_2, x_3\}=\{12, 14, 16\}$; and ask only for the shortest 90% confidence interval.

A further integration then yields the cumulative distribution function $F(y)=[1-(1+3y+9y^2/2)\exp(-3y)]$, $y>0$. Any numbers y_1, y_2 satisfying $F(y_2)-F(y_1)=0.9$ determine a 90% confidence interval. To find the shortest one, we impose in addition the constraint $F'(y_1)=F'(y_2)$. By computer, this yields the interval

$$\theta^* - 0.8529 < \theta < \theta^* + 0.8264 \tag{17}$$

or, with the above sample values, the shortest 90% confidence interval is

$$12.1471 < \theta < 13.8264. \tag{18}$$

The Bayesian solution is obtained from inspection of (15); with a constant prior density [which, as we have argued elsewhere (Jaynes, 1968) is the proper way to express complete ignorance of location parameter], the posterior density of θ will be

$$p(\theta \mid x_1 \ldots x_N) = \begin{Bmatrix} N\exp N(\theta - x_1), & \theta < x_1 \\ 0 \qquad , & \theta > x_1 \end{Bmatrix} \tag{19}$$

where we have ordered the sample values so that x_1 denotes the least one observed. The shortest posterior probability belt that contains 100 P percent of the posterior probability is thus $(x_1-q)<\theta<x_1$, where $q=-N^{-1}\log(1-P)$. For the above sample values we conclude (by slide-rule) that, with 90% probability, the true value of θ is contained in the interval

$$11.23 < \theta < 12.0. \tag{20}$$

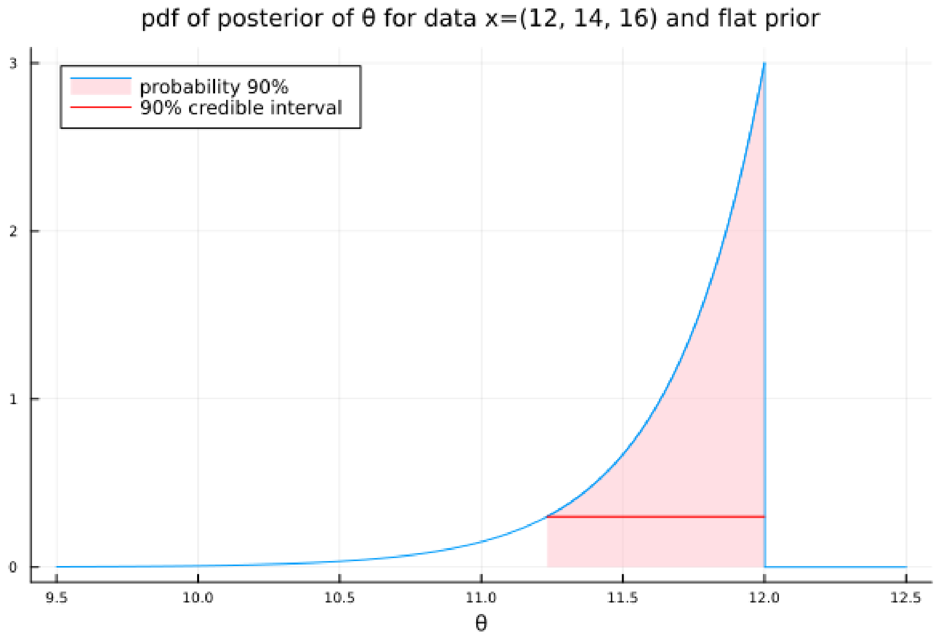
Now what is the verdict of our common sense? The Bayesian interval corresponds quite nicely to our common sense; the confidence interval (18) is over twice as wide, and *it lies entirely in the region $\theta > x_1$ where it is obviously impossible for θ to be*!.

I first presented this result to a recent convention of reliability and quality control statisticians working in the computer and aerospace

In [3]:
```julia
function credint_truncexp(x, α = 0.05)
    n, t = length(x), minimum(x)
    c = -(1/n)*log(α)
    [t - c, t]
end

@show n = length(x)
@show t = minimum(x)
@show α = 0.10
@show c = -1/n * log(α)
@show credint = credint_truncexp(x, α)

P_post = plot()
plot!(θ -> pdf_posterior(θ, x), 9.5, 12.5; label="")
title!("pdf of posterior of θ for data x=$x and flat prior")
plot!(xguide="θ")
plot!(θ -> pdf_posterior(θ, x), credint...;
    label="probability 90%", c=1, fc=:pink, fa=0.5, fillrange=0)
plot!(credint, fill(pdf_posterior(credint[begin], x), 2);
    label="90% credible interval", c=:red)
plot!(legend=:topleft)
```

```
n = length(x) = 3
t = minimum(x) = 12
α = 0.1 = 0.1
c = (-1 / n) * log(α) = 0.7675283643313484
credint = credint_truncexp(x, α) = [11.232471635668652, 12.0]
```

Out[3]:



2 Jaynesによる不適切な信頼区間の構成

以下, 確率変数 $X = (X_1, \dots, X_n)$ が従う分布の確率密度函数は $p(x|\theta)$ であると仮定する.

このとき, X の標本平均 $\bar{X}$ が従う確率分布は次のようになる:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim \theta + \mathrm{Gamma}(n, 1/n).$$

特に $E[\bar{X}] = \theta + 1$ となるので,

$$\theta^* = \bar{X} - 1$$

は θ の不偏推定量になる. さらに, $\bar{X}$ の分散は $\mathrm{var}(\bar{X}) = 1/n$ になので, $\theta^* = \bar{X} - 1$ は θ の一致推定量にもなっている.

$Y = \theta^* + 1 - \theta = \bar{X} - \theta$ が従う分布 $\mathrm{Gamma}(n, 1/n)$ の確率密度函数は次のように書ける:

$$p(y) = \begin{cases} 0 & (y < 0) \\ (n^n/\Gamma(n))\, y^{n-1} e^{-ny} & (y \geq 0). \end{cases}$$

$n = 3$ のとき, この $Y = \bar{X} - \theta$ が従う分布の90%最高密度区間は次になる:

$$0.14710896409415852 \le Y = \bar{X} - \theta \le 1.8263915697442574.$$

X の実現値が $x = (12, 14, 16)$ のとき, $\bar{X}$ の実現値は $\bar{x} = 14$ になる. そのとき, Y の90%最高密度区間に対応する θ の区間が

$$0.14710896409415852 \le 14 - \theta \le 1.8263915697442574$$

で定義され, それは

$$[12.173608430255742, 13.852891035905841]$$

になる. これはJaynesの結果とは異なるが, こちらが正しいように思われる.

しかし, 要点は上で求めた θ の区間内の値が $\min(x) = 12$ よりも真に大きいことである.

θ の推定値は常識的に考えて $\min(x) = 12$ 以下でなければいけない.

Jaynesはこのような θ の区間を信頼区間とみなし, 信頼区間の使用はナンセンスだと自信たっぷりに決めつけている. 原文 (https://bayes.wustl.edu/etj/articles/confidence.pdf) のp.198を見ると, この件について講演したときの様子をものすごく偉そうに説明している.

もしもJaynesが次の節で説明するより適切な信頼区間の構成法を発見できていれば, そのように偉そうな態度を取ることは不可能だったと思われる.

この場合も, 信頼区間を適切に構成すれば, 平坦事前分布のBayes統計の結果が非Bayesianの結果にぴったり一致する場合になっているのである. 次節でその一致について説明する.

industries; and at this point the meeting was thrown into an uproar, about a dozen people trying to shout me down at once. They told me, "This is complete nonsense. A method as firmly established and thoroughly worked over as confidence intervals couldn't possibly do such a thing. You are maligning a very great man; Neyman would never have advocated a method that breaks down on such a simple problem. If you can't do your arithmetic right, you have no business running around giving talks like this".

After partial calm was restored, I went a second time, very slowly and carefully, through the numerical work leading to (18), with all of them leering at me, eager to see who would be the first to catch my mistake [it is easy to show the correctness of (18), at least to two figures, merely by applying parallel rulers to a graph of $F(y)$]. In the end they had to concede that my result was correct after all.

To make a long story short, my talk was extended to four hours (all afternoon), and their reaction finally changed to: "My God – why didn't somebody tell me about these things before? My professors and textbooks never said anything about this. Now I have to go back home and recheck everything I've done for years".

This incident makes an interesting commentary on the kind of indoctrination that teachers of orthodox statistics have been giving their students for two generations now.

(c) WHAT WENT WRONG?

Let us try to understand what is happening here. It is perfectly true that, *if* the distribution (15) is indeed identical with the limiting frequencies of various sample values, and *if* we could repeat all this an indefinitely large number of times, then use of the confidence interval (17) *would* lead us, in the long run, to a correct statement 90% of the time. But it would lead us to a wrong answer 100% of the time in the subclass of cases where $\theta^* > x_1 + 0.85$; and *we know from the sample whether we are in that subclass.*

That there must be a very basic fallacy in the reasoning underlying the principle of confidence intervals, is obvious from this example. The difficulty just exhibited is generally present in a weaker form, where it escapes detection. The trouble can be traced to two different causes.

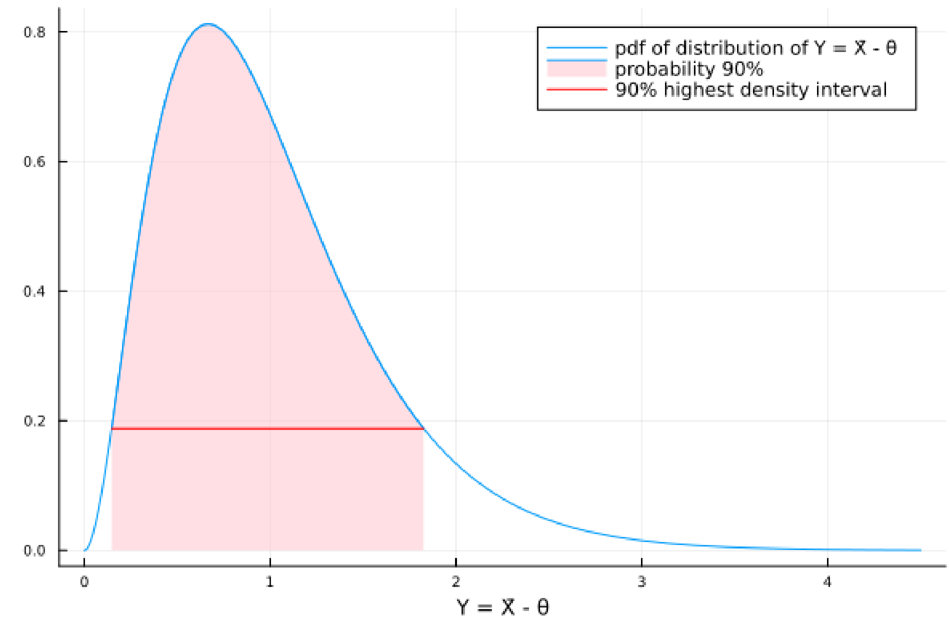
Firstly, it has never been a part of 'official' doctrine that confidence intervals must be based on sufficient statistics; indeed, it is usually held

In [4]:

```
function hdi(dist::ContinuousUnivariateDistribution, α = 0.05; alg = Brent())
    f(p) = quantile(dist, p + (1 - α)) - quantile(dist, p)
    o = optimize(f, 0, α, alg)
    p = o.minimizer
    quantile.(dist, [p, p + (1 - α)])
end

function confint_truncexp_jaynes(x, α = 0.05; alg=Brent())
    n = length(x)
    dist_y = Gamma(n, 1/n)
    y_int = hdi(dist_y, α; alg)
    mean(x) .- reverse(y_int)
end

@show x = (12, 14, 16)
@show n = length(x)
@show α = 0.10
@show dist_y = Gamma(n, 1/n);
@show y_int = hdi(dist_y, α)
@show cdf(dist_y, y_int[end]) - cdf(dist_y, y_int[begin])
@show pdf(dist_y, y_int[end]) ≈ pdf(dist_y, y_int[begin])
@show confint_jaynes = confint_truncexp_jaynes(x, α)

plot(y -> pdf(dist_y, y), 0, 4.5; label="pdf of distribution of Y = X̄ - θ")
plot!(y -> pdf(dist_y, y), y_int...;
    label="probability 90%", c=1, fc=:pink, fa=0.5, fillrange=0)
plot!(y_int, pdf.(dist_y, y_int);
    label="90% highest density interval", c=:red)
plot!(xguide="Y = X̄ - θ")
```

```
x = (12, 14, 16) = (12, 14, 16)
n = length(x) = 3
α = 0.1 = 0.1
dist_y = Gamma(n, 1 / n) = Gamma{Float64}(α=3.0, θ=0.3333333333333333)
y_int = hdi(dist_y, α) = [0.14710896409415852, 1.8263915697442574]
cdf(dist_y, y_int[end]) - cdf(dist_y, y_int[begin]) = 0.9
pdf(dist_y, y_int[end]) ≈ pdf(dist_y, y_int[begin]) = true
confint_jaynes = confint_truncexp_jaynes(x, α) = [12.173608430255742, 13.852891035905841]
```

Out[4]:



[Frequentism and Bayesianism III (http://jakevdp.github.io/blog/2014/06/12/frequentism-and-bayesianism-3-confidence-credibility/)](http://jakevdp.github.io/blog/2014/06/12/frequentism-and-bayesianism-3-confidence-credibility/) では $x = (10, 12, 15)$ の場合を扱っており, exact版の95%信頼区間の計算結果は $[10.2, 12.2]$ である.

In [5]:

```
@show hdi(Gamma(3, 1/3), 0.05)
@show mean((10, 12, 15))
@show confint_truncexp_jaynes((10, 12, 15), 0.05);
```

```
hdi(Gamma(3, 1 / 3), 0.05) = [0.10116685354387298, 2.133740683353059]
mean((10, 12, 15)) = 12.333333333333334
confint_truncexp_jaynes((10, 12, 15), 0.05) = [10.199592649980275, 12.232166479789461]
```

3 適切な信頼区間の構成法

前節と同様に確率変数 $X = (X_1, \dots, X_n)$ が従う分布の確率密度函数は $p(x|\theta)$ であると仮定する.

Jaynesの誤りは標本平均 $\bar{X}$ の分布を信頼区間の構成に安易に用いたことである.

その代わりに標本中の最小値 $\min(X)$ の分布を使えば, より適切な信頼区間が得られ, そのようにして得られた信頼区間が平坦事前分布から得られるBayesianな信用区間にぴったり一致することも確かめられる.

$X_1 \sim \theta + \mathrm{Exponential}(1)$ より, $X_1 \geq t$ となる確率は

$$P(X_1 \geq t|\theta) = \begin{cases} 1 & (t \leq \theta) \\ e^{-(t-\theta)} & (t \geq \theta) \end{cases}$$

になる. ゆえに, $\min(X) \geq t$ となる確率は次のようになる:

$$P(\min(X) \geq t|\theta) = P(X_1 \geq t|\theta) \cdots P(X_n \geq t|\theta) = P(X_1 \geq t|\theta)^n = \begin{cases} 1 & (t \leq \theta) \\ e^{-n(t-\theta)} & (t \geq \theta). \end{cases}$$

したがって, 確率変数 $T = \min(X)$ が従う分布の確率密度函数は次のようになる:

$$p(t|\theta) = -\frac{\partial}{\partial t} P(\min(X) \geq t|\theta) = \begin{cases} 0 & (t < \theta) \\ ne^{-n(t-\theta)} & (t > \theta). \end{cases}$$

以上の結果は次が成立していることを意味している:

$$\min(X) \sim \theta + \mathrm{Exponential}(1/n).$$

この結果とデータ x と平坦事前分布から得られる θ の事後分布が

$$\theta \sim \min(x) - \mathrm{Exponential}(1/n).$$

になることとの類似性に注意せよ.

P値の一般的(かつ大雑把)な定義は「統計モデル内でデータの数値以上に極端な値が生じる確率(もしくはその近似値)」であった.

Jaynesが採用した信頼区間はこれを「切断指数分布モデル内でデータの数値 $\bar{x} - 1$ 以上に $\bar{X} - 1$ の値が θ から離れる確率」と解釈した場合に対応している. その解釈は適切ではない.

ここでは, P値を「切断指数分布モデル内でデータの数値 $\min(x)$ 以上に $\min(X)$ の値が θ から離れる確率」によって定義しよう:

$$\mathrm{pvalue}(x|\theta) = \begin{Bmatrix} P(\min(X) \leq \min(x)|\theta) & (\min(x) < \theta) \\ P(\min(X) \geq \min(x)|\theta) & (\min(x) \geq \theta) \end{Bmatrix} = \begin{cases} 0 & (\min(x) < \theta) \\ e^{-n(\min(x)-\theta)} & (\min(x) \geq \theta). \end{cases}$$

$c \geq 0$ が $e^{-nc} = \alpha$ によって定義されていたことより, このP値の定義に対応する $100(1-\alpha)\%$ 信頼区間は次のように書ける:

$$\mathrm{confint}(x|\alpha) = \{\, \theta \in \mathbb{R} \mid \mathrm{pvalue}(x|\theta) \geq \alpha \,\} = [\min(x) - c, \min(x)].$$

この信頼区間は平坦事前分布から得られるBayesianな信用区間にぴったり一致している.

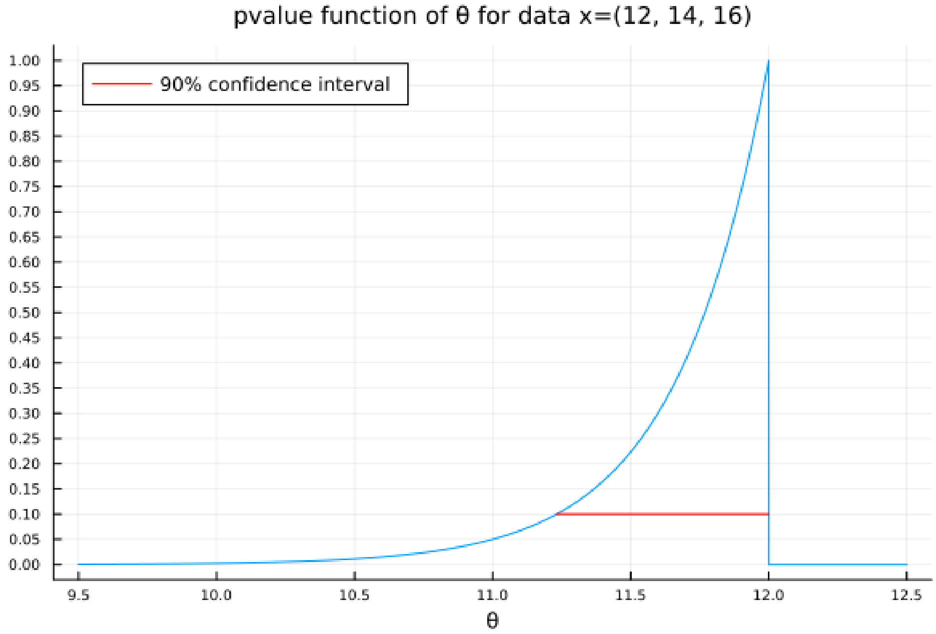
平坦事前分布のBayes統計の結果が非Bayesianの場合の対応する結果にぴったり一致する場合が少なくないが, この場合もそのような場合の例になっている.

In [6]:

```julia
function pvalue_truncexp(x, θ)
    n, t = length(x), minimum(x)
    t < θ ? float(zero(θ)) : exp(-n*(t - θ))
end

function confint_truncexp(x, α = 0.05)
    n, t = length(x), minimum(x)
    c = -(1/n)*log(α)
    [t - c, t]
end

x = (12, 14, 16)
@show α = 0.10
@show confint = confint_truncexp(x, α)
@show credint = credint_truncexp(x, α)

P_pval = plot()
plot!(θ -> pvalue_truncexp(x, θ), 9.5, 12.5; label="")
title!("pvalue function of θ for data x=$x")
plot!(xguide="θ", ytick=0:0.05:1)
plot!(confint, fill(α, 2); label="90% confidence interval", c=:red)
plot!(legend=:topleft)
```

```
α = 0.1 = 0.1
confint = confint_truncexp(x, α) = [11.232471635668652, 12.0]
credint = credint_truncexp(x, α) = [11.232471635668652, 12.0]
```

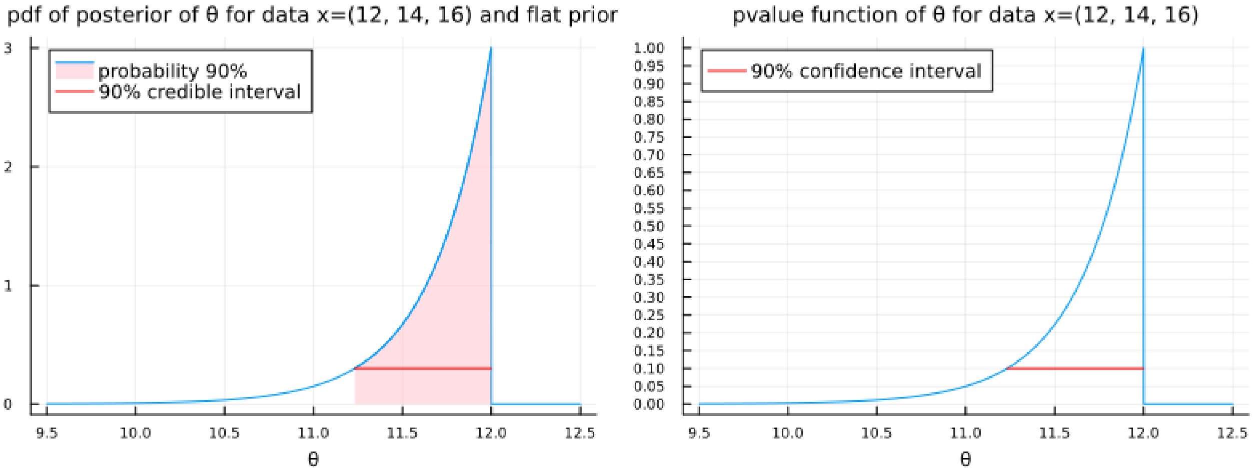
Out[6]:



In [7]:

```
plot(P_post, P_pval; size=(800, 300))
plot!(titlefontsize=9, guidefontsize=8, bottommargin=4Plots.mm)
```

Out[7]:



In [ ]:

```
```