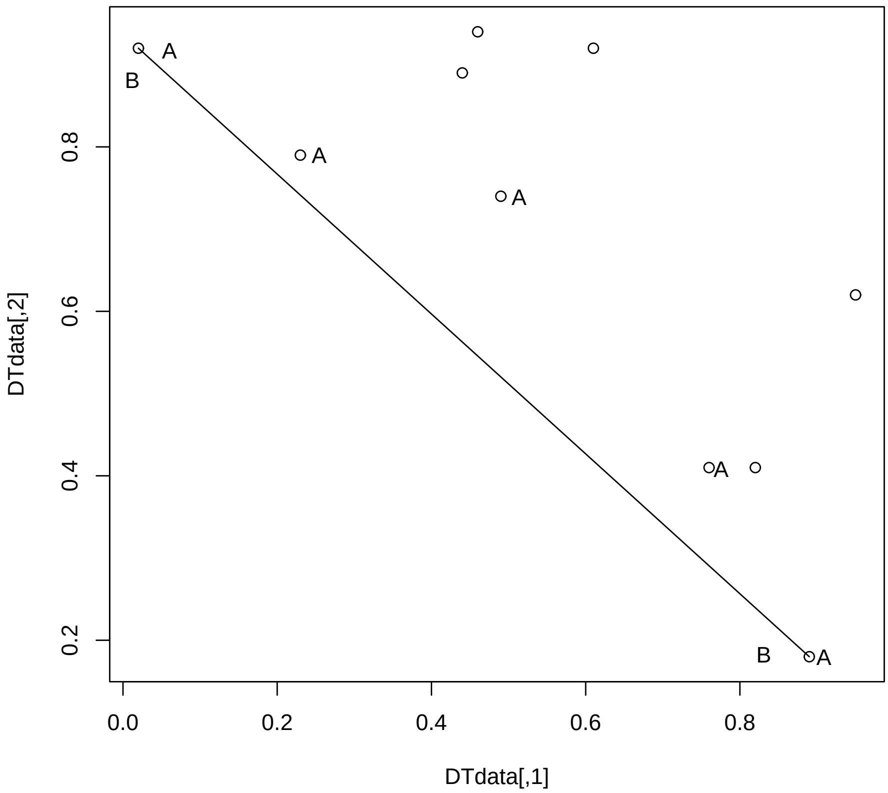
(1) Here is a plot to answer the question. Points labeled A are admissible. Points labeled B are Bayesian. If randomization were allowed, all points on the line shown connecting the two Bayesian points would be admissible, and all the points labeled A that lie above the line would become inadmissible.



(2) Suppose we have i.i.d. observations on $\{x_t, t = 1, \ldots, T\}$, with each x_t distributed as $N(\mu, \mu^2)$. That is, μ is both the mean and the standard deviation of the observations. It is known that $\mu > 0$. This exercise compares Bayesian with frequentist methods of constructing point estimates and interval estimates for μ.

(a) Show that $\sum x_t/T$ and $\sum x_t^2$ form a two-dimensional sufficient statistic here.

The pdf of the sample can be written as

$$(2\pi)^{-T/2} \exp\left(-\tfrac{1}{2}\sum_1^T \left(\frac{x_t^2}{\mu^2} - \frac{2x_t}{\mu} + 1\right)\right) ,$$

from which it is clear that the likelihood function depends on the data only through the two proposed sufficient statistics.

(b) Show that $\bar{x} = \sum x_t/T$ is an unbiased and consistent estimator for μ.

It is a standard result that for i.i.d. $N(\eta, \sigma^2)$ variables, $\bar{x} \sim N(\mu, \sigma^2/T)$. This implies $\bar{x}$ is unbiased and, since σ^2 is a constant, that the variance of $\bar{x}$ goes to zero as $T \to \infty$. This in turn implies that $\bar{x}$ converges in quadratic mean to μ and (therefore) also in probability.

(c) Show that $s^2 = \sum(x_t - \bar{x})^2/T$ is an unbiased estimator for μ^2 and that its square root is a consistent estimate of μ.

The assertion in this question is not quite correct. $E[s^2] = \mu^2(T-1)/T$. To make it unbiased for μ^2 we have to multiply it by $T/(T-1)$ (or just define it with $T-1$ in the denominator). It is another standard result from normal sampling theory that $\sum(x_t - \bar{x})^2/\sigma^2 \sim \chi^2(T-1)$. The variance of a $\chi^2(T)$ distribution is T, so the variance of $\chi^2(T)/T$ is $1/T$ and therefore $s^2 \xrightarrow{P} \sigma^2$. Since for any continuous function f, $x_t \xrightarrow{P} z \Rightarrow f(x_t) \xrightarrow{P} f(z)$, $\sqrt{s^2} \xrightarrow{P} \mu$.

(d) Show that $\bar{x}/\mu$ and s^2/μ^2 are what is known as **pivotal quantities**, or just plain **pivots**, meaning that they each have a distribution that does not depend on the unknown parameter μ. Show also that the two are independent for each μ.

Since $\bar{x} \sim N(\mu, \mu^2/T)$, $\bar{x}/\mu \sim N(1, 1/T)$. And it is a standard result that Ts^2/μ^2 (with our definition of s^2) is $\chi^2(T-1)$. That these two quantities are independent across repeated samples with μ fixed is again a standard result. To prove it, though, note that

$$\vec{x} - \bar{x} = (I - (1/T)\underset{T\times T}{\mathbf{1}})\vec{x}\,.$$

Here $\mathbf{1}$ stands for a matrix full of ones. If we use M to denote the factor multiplying $\vec{x}$ on the right of this expression, we have

$$E[(\vec{x} - \bar{x})(\bar{x} - \mu)] = E[M\vec{x}\frac{1}{T}(\vec{x}'\underset{T\times 1}{\mathbf{1}} - \mu)]$$
$$= \frac{1}{T}M(\mu\underset{T\times T}{\mathbf{1}} + \mu^2\underset{T\times 1}{\mathbf{1}} - \mu^2\underset{T\times 1}{\mathbf{1}}) = 0\,,$$

since it easily checked that $M\underset{T\times 1}{\mathbf{1}} = 0$.

(e) Derive the form of three confidence intervals, based on $\bar{x}/\mu$, on s^2/μ^2, and on $\sum x_t^2/\mu^2$ (which is also pivotal).

$$\sqrt{T}\left(\frac{\bar{x}}{\mu} - 1\right) \sim N(0,1)$$
$$\therefore P\left[-1.96 < \sqrt{T}\left(\frac{\bar{x}}{\mu} - 1\right) < 1.96\right] = .5$$
$$\therefore P\left[1 - \frac{1.96}{\sqrt{T}} < \frac{\bar{x}}{\mu} < 1 - \frac{1.96}{\sqrt{T}}\right] = .95$$

What kind of an interval this gives depends on T and on the sign of $\bar{X}$. In our case below, $T = 3$, so $1 - 1.96/\sqrt{T} < 0$. This means that only one side of the interval on $\bar{x}/\mu$ binds. If $\bar{x} > 0$, the interval is $(\bar{x}/(1.96/\sqrt{T} + 1), \infty)$, while if $\bar{x} < 0$ the interval is $(\bar{x}/(1 - 1.96/\sqrt{T}), \infty)$.

$$T\frac{s^2}{\mu^2} \sim \chi^2(T-1)$$
$$\therefore T = 3 \Rightarrow P\left[.05064 < T\frac{s^2}{\mu^2} < 7.378\right] = .95$$
$$\therefore T = 3 \Rightarrow P\left[\frac{Ts^2}{7.378} < \mu^2 < \frac{Ts^2}{.05064}\right] = .95\,.$$

Most applications of the χ^2 distribution to frequentist testing use just the right tail. Here we use left and right tails because it seems reasonable to consider μ's that are too large as well as μ's that are too big. But there is no right answer to which tail or tails to use, unless one is explicit about what alternative hypotheses one has in mind.

$$\frac{\sum x_t^2}{\mu^2} \sim \chi^2(T;T)\,,$$

where the first parameter of the χ^2 distribution is the degrees of freedom and the second is the non-centrality parameter. The $\chi^2(n, \lambda)$ distribution is the distribution of the sum of n i.i.d. $N(\nu_i, 1)$ random variables, where $\lambda = \sum \nu_i^2$. This distribution is not as widely available in tables at the back of statistics books as is the ordinary χ^2 distribution, but it is available as a function call in R or S, and also (as the function `ncx2inv`) in matlab if the statistics toolbox has been installed.(which it probably is on the departmental machines, but probably isn't on student edition versions running on student laptops). In any case, with $T = 3$,

$$P\left[.5643 < \frac{\sum x_t^2}{\mu^2} < 16.521\right] = .95$$
$$\therefore P\left[\frac{\sum x_t^2}{16.521} < \mu^2 < \frac{\sum x_t^2}{.5643}\right] = .95\,.$$

(f) For each of the following samples, find $\bar{x}$, $\sqrt{s^2}$, the maximum likelihood estimate of μ, the flat-prior posterior mean of μ, and the posterior mean of μ when the prior is proportional to $\mu^{-2}\exp(-1/(10\mu))$. The posterior means probably require numerical integration. There are functions in Matlab and R that do numerical integration, or it is fairly easy to code this yourself in a couple of lines.

Also find for each sample the 95% confidence intervals you derived above and 95% HPD (highest posterior density) regions under the flat prior and the proper prior. If $p(\mu \mid \vec{x})$ is the posterior density function, the 95% HPD region for μ is a set of the form $\{\mu \mid p(\mu \mid \vec{x}) > \bar{p}\}$ for a $\bar{p}$ such that the set's posterior probability is .95. You will need the computer to find it numerically.

The samples:

(i) $\{-5, 0, 5\}$

(ii) $\{0, 1, 2\}$

(iii) $\{5, 5.1, 5.2\}$

(iv) $\{-2, -2, -2.1\}$

The pdf of the sample is

$$\mu^{-3} \exp\left(-\tfrac{1}{2} \sum \left(\frac{x}{\mu} - 1\right)^2\right)$$

from which it is not hard to derive a quadratic equation satisfied at the FOC for a maximum, whose unique positive solution is, for a sample of size 3,

$$\hat{\mu} = \frac{-\sum x_t + \sqrt{(\sum x_t)^2 + 12 \sum x_t^2}}{6}\,.$$

The answers for the estimators, then are

MLE's	4.082	.8844	3.153	3.290
Flat prior $\hat{\mu}$	8.868	1.471	4.841	8.685
Prop. prior $\hat{\mu}$	4.449	0.9669	3.397	3.525

Note that the prior, though it is proper, has an infinite mean. However it is very sharply peaked, much moreso than the likelihood itself, at $\mu = .05$. This does not matter as much as it might seem, because the likelihood is extremely small near $\mu = .05$. The prior's main effect is to strongly damp the long right tail in the likelihood and slightly shift the mode to the left. Even with this fairly strong prior belief in smaller μ's, however, the posterior is enough skewed to the right that the posterior mean lies considerably above the MLE in each case. This is entirely from the effect of taking the posterior mean rather than the mode. The posterior mode with the proper prior is in every case considerably *below* the MLE.

For the confidence intervals, the answers are

$\bar{x}$	(0, ∞)	(.469, ∞)	(2.393, ∞)	(15.45, ∞)
s^2	(20.33, 2962)	(.8132, 118.5)	(.00813, 1.1185)	(.00271, .3949)
$\sum x_t^2$	(3.026, 88.61)	(.3026, 8.861)	(4.724, 138.3)	(.7511, 21.99)

For the HPD regions the answers are

flat	(1.82, 22.2)	(0.467, 3.17)	(1.75, 9.84)	(1.22, 23.5)
proper	(1.74, 8.54)	(0.454, 1.697)	(1.689, 5.77)	(1.14, 7.29)

Note that the confidence intervals for a given sample differ sharply among themselves in some samples and are generally wider than the Bayesian intervals. Their erratic behavior reflects our having chosen samples in which the sample means and variances are in unlikely relations to each other, except for the second sample $\{0, 1, 2\}$. The Bayesian intervals are sensitive to the rather strong prior mainly in cutting back on the right tail when the prior is applied. The left ends of the interval are not very sensitive to the prior.

A set of graphs that display all the answers to this question is at the end of this answer sheet. The flat prior cases come first, for samples 1-4, then the proper prior cases. The points labeled "ml" in the graphs are maximum likelihood estimators for the flat prior graphs, but for the proper prior graphs they are just the posterior mode. The graphs are generated by Tamas Papp's R code, which is available on the course web site.

(g) In deciding which of the three confidence intervals to use, would it make sense in a given sample to pick whichever is smallest, on the grounds that that is the one that is giving the most precise information? Why or why not?

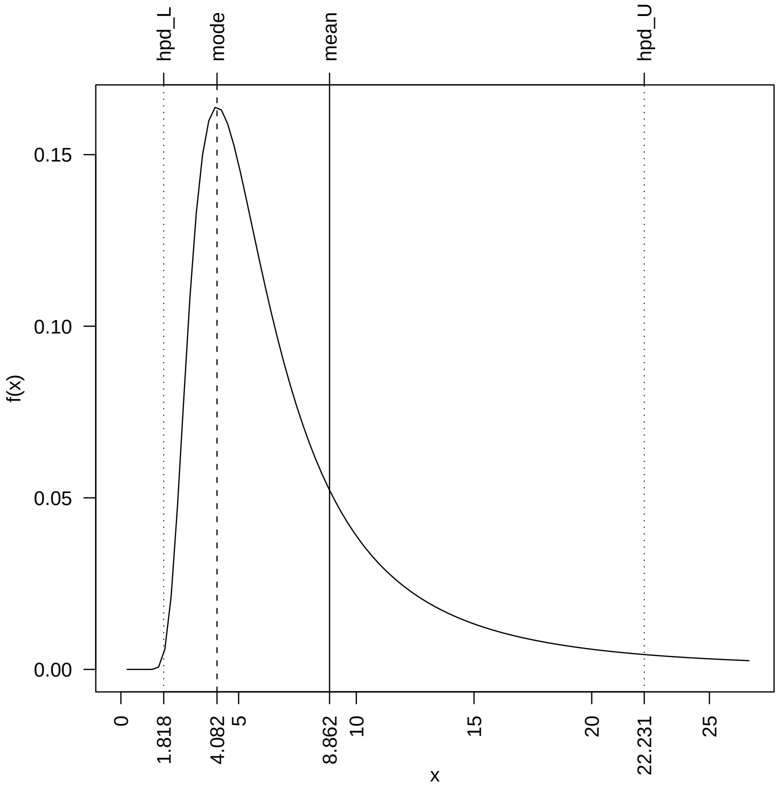
As discussed in class, short intervals can indicate that the model fits well for hardly any parameter values. Usually we want confidence intervals to behave like HPD regions for some prior. If a confidence interval is much shorter than an HPD region with the same probability, this then suggests it is not a good indicator of uncertainty about the parameter's location.

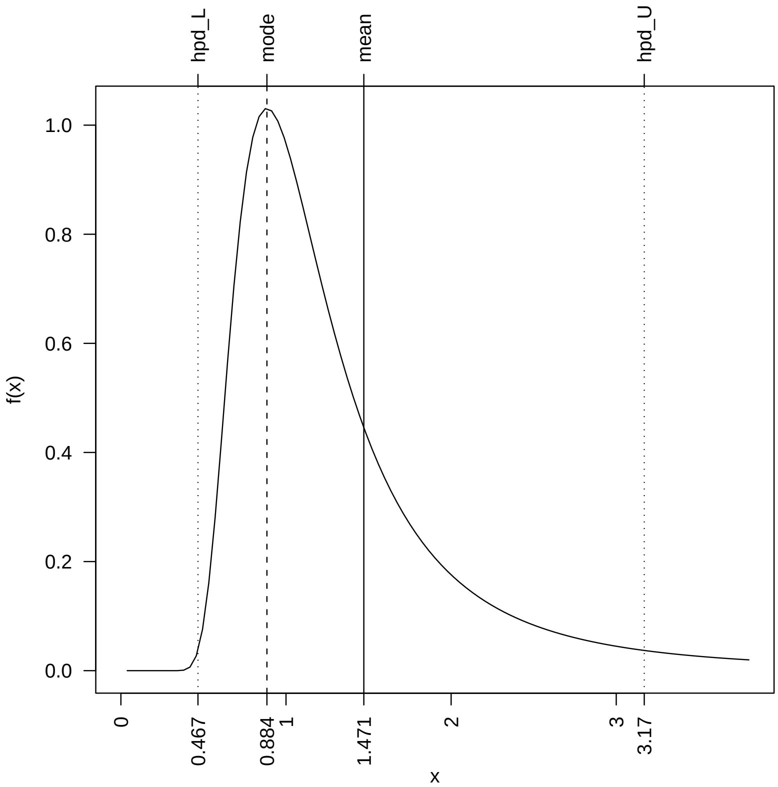
(h) How should empty confidence intervals be interpreted? Should there be a big difference between the interpretation of an empty confidence interval and of a very short (nearly empty?) confidence interval?

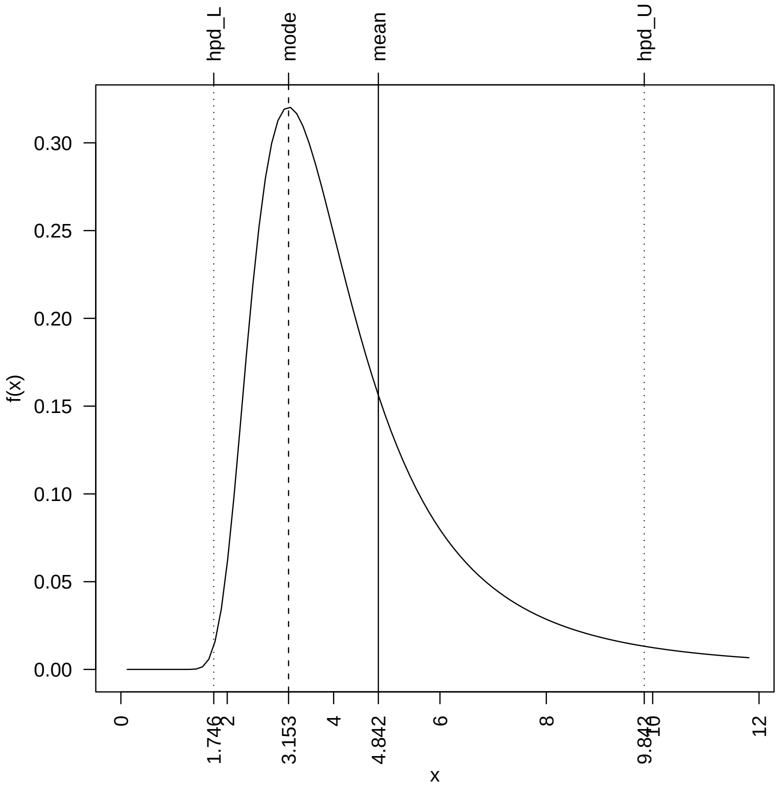
See previous answer.

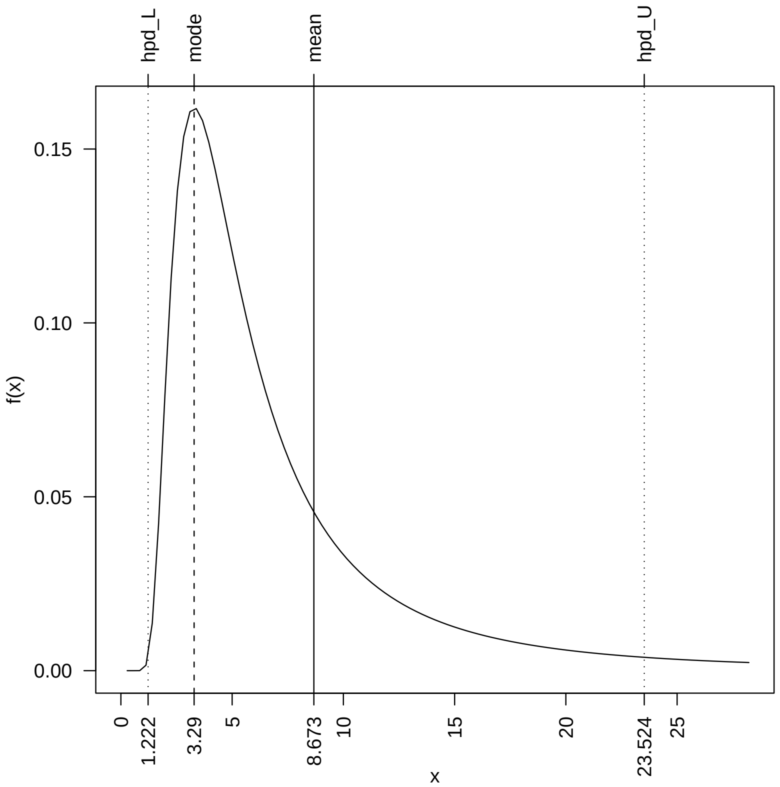
(i) Show that the Bayesian posterior mean and HPD region can't be computed for the flat prior, if the sample size is one.

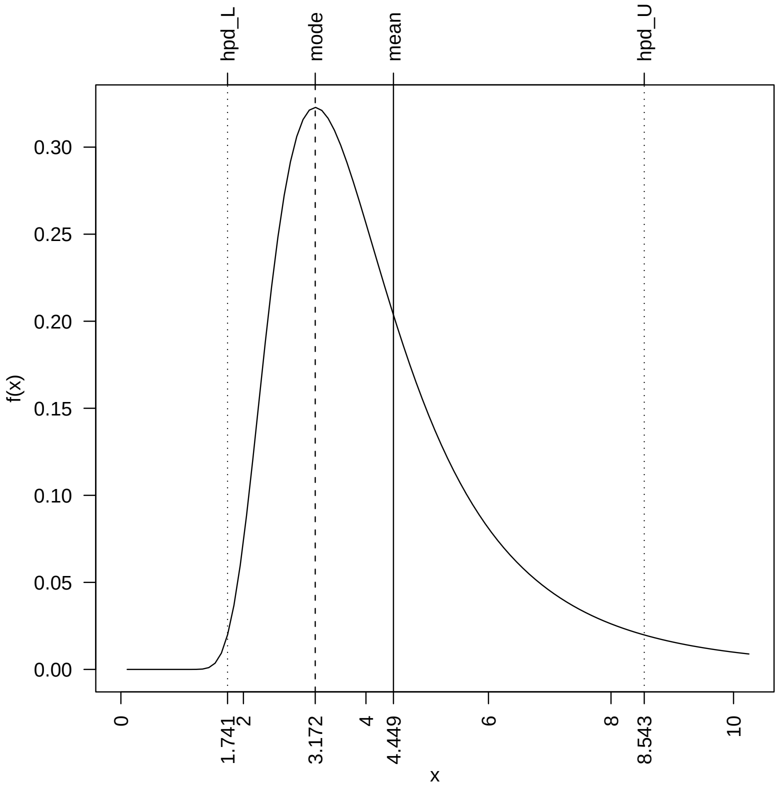
With a sample size of one, the likelihood is very close to e^{-1}/μ for large μ. But this is not integrable, so it can't be treated as a flat-prior posterior density.

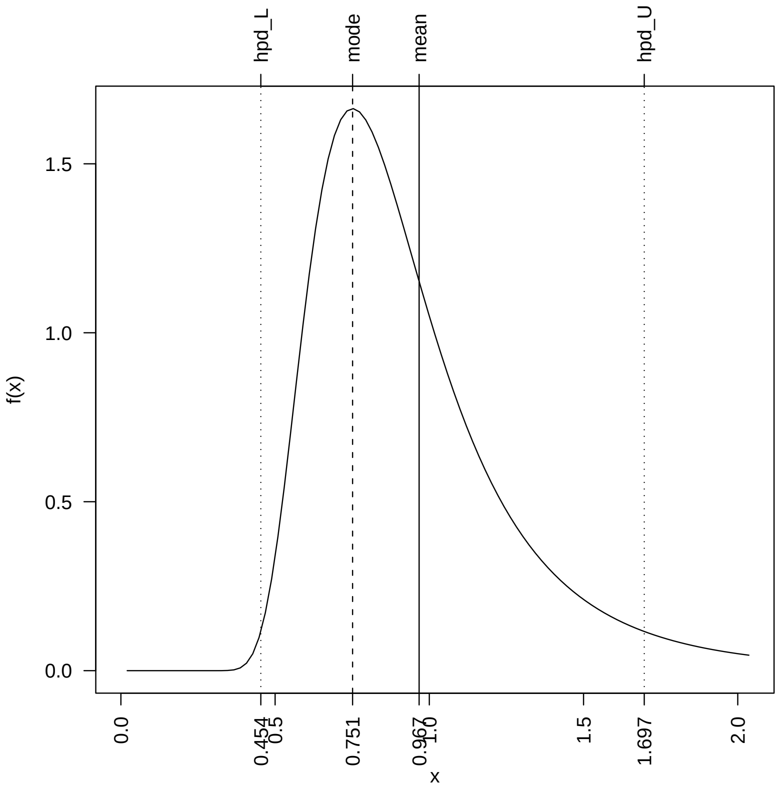

hpd_L
mode
mean
hpd_U
0.00
0.05
0.10
0.15
f(x)
0
1.818
4.082
5
8.862
10
15
20
22.231
25
x

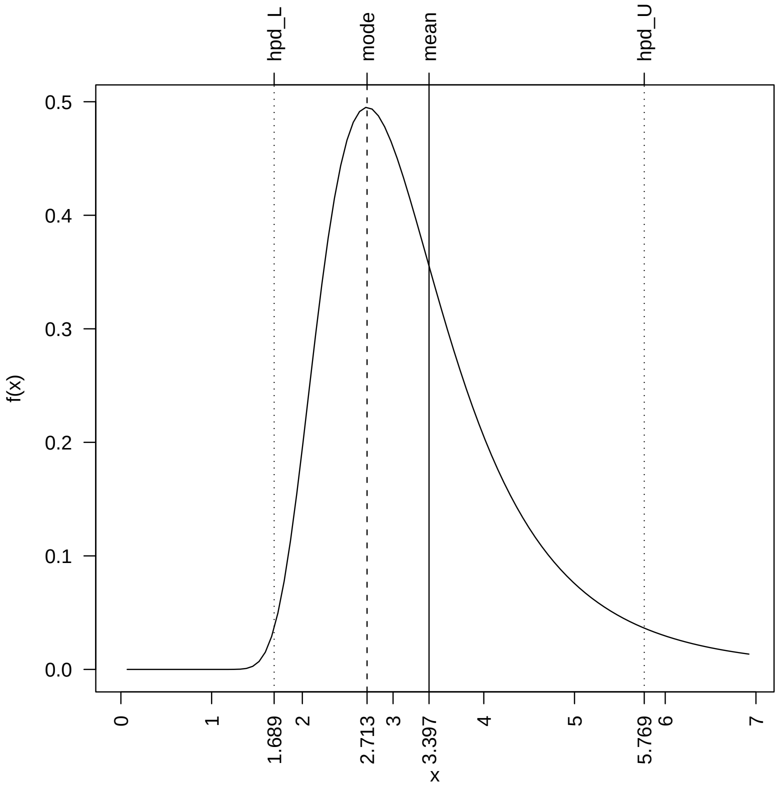

hpd_L
mode
mean
hpd_U
0.0
0.2
0.4
0.6
0.8
1.0
f(x)
0
0.467
0.884
1
1.471
2
3
3.17
x

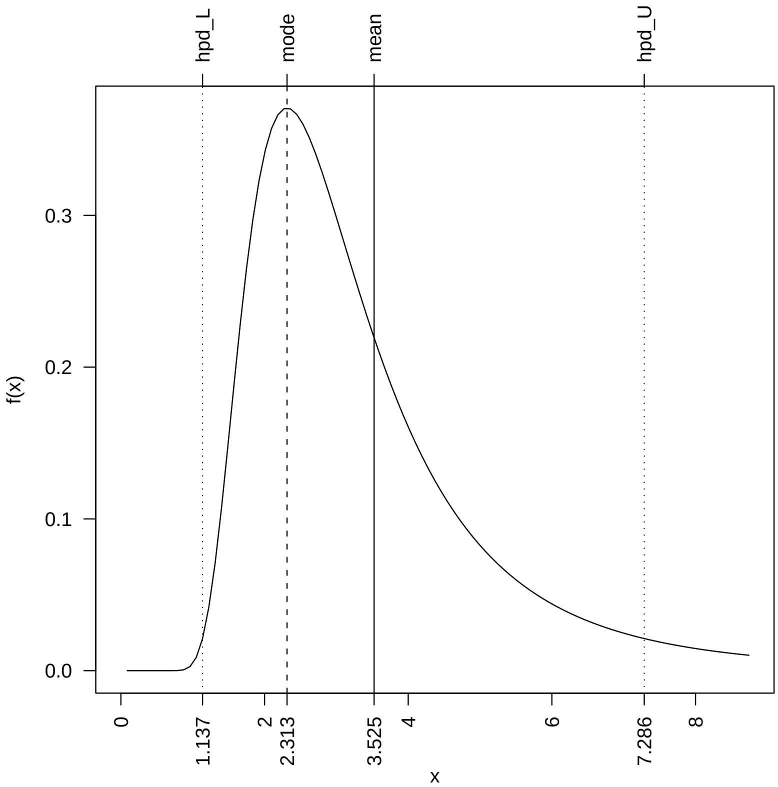

hpd_L
mode
mean
hpd_U
f(x)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1.746
2
3.153
4
4.842
6
8
9.842
10
12
x


hpd_L
mode
mean
hpd_U
f(x)
0.00
0.05
0.10
0.15
0
1.222
3.29
5
8.673
10
15
20
23.524
25
x


hpd_L
mode
mean
hpd_U
f(x)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1.741
2
3.172
4
4.449
6
8
8.543
10
x


hpd_L
mode
mean
hpd_U
f(x)
0.0
0.5
1.0
1.5
0.0
0.454
0.5
0.751
0.967
1.0
1.5
1.697
2.0
x


hpd_L
mode
mean
hpd_U
f(x)
0.0
0.1
0.2
0.3
0.4
0.5
0
1
1.689
2
2.713
3
3.397
4
5
5.769
6
7
x


hpd_L
mode
mean
hpd_U
f(x)
0.0
0.1
0.2
0.3
0
1.137
2
2.313
3.525
4
6
7.286
8
x