Chapter 10. Conditioning on the DDF statistic in cases that are not log-symmetric.

In this chapter we look at a number of different scenarios and compare the results produced by E. C. inference with those from conventional inference.

We are able to state the general form[1] of the DDF statistic for tests on the Exponential mean, two Gamma parameters, the variance of a Normal with known mean, and one of the Weibull parameters. Although we can always find the value of $\pi(y_0)$ for any particular y_0, in most of these cases we cannot find $\pi(\cdot)$ analytically. The exceptions to this are several scenarios involving the exponential model where we can find $\pi(\cdot)$ by solving quadratic and cubic equations; in these cases we are able to discuss the general nature of the cp-function. We identify a type of scenario where E. C. inference breaks down in the sense that it produces significantly small cp-values for non-significant likelihood ratios, and find that taking sufficiently large samples can solve this problem. Finally, we examine an artificial model (the Gradient model) designed to illuminate a number of the issues that arose in Welch's Uniform case. In that case, both of the rival methods breached the sufficiency principle whereas, in our example, they do not, this example provides a better basis for comparing the conditional and unconditional approaches.

10.1 Tests on the variance of a Normal population.

Let $X_1,\ldots,X_n$ be independent and identically distributed random variables with a $N(\mu,\sigma^2)$ distribution where μ is known and σ^2 is the unknown parameter of interest. Consider hypotheses of the form H: $\sigma^2=\sigma_1^2$ versus K: $\sigma^2=\sigma_2^2$, $(\sigma_1^2,\sigma_2^2\in\mathbb{R}^+)$. Let $q=\frac{\sigma_2}{\sigma_1}=\sqrt{\frac{\sigma_2^2}{\sigma_1^2}}>0$ and $V=\sum_i (X_i-\mu)^2\geq 0$.

[1] That is, the form that can be applied to any particular pair of hypotheses.

Then

$$y = LR(\underset{\sim}{x}) = \frac{(\frac{1}{\sigma_1})^n \exp\{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma_1^2}\}}{(\frac{1}{\sigma_2})^n \exp\{\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma_2^2}\}}$$

$$\Rightarrow \boxed{y = q^n \cdot \exp\{-\tfrac{1}{2}(\tfrac{1}{\sigma_1^2} - \tfrac{1}{\sigma_2^2})v\}.}$$

y is a one-to-one function of v, increasing when $\sigma_2 < \sigma_1$ ($q<1$) and decreasing when $\sigma_2 > \sigma_1$ ($q>1$). Also, $q<1 \Rightarrow y \in (q^n, \infty)$ and $q>1 \Rightarrow y \in (0, q^n)$.

We can use the fact that $\frac{v}{\sigma^2}$ has the χ_n^2 distribution to find the DDF statistic:

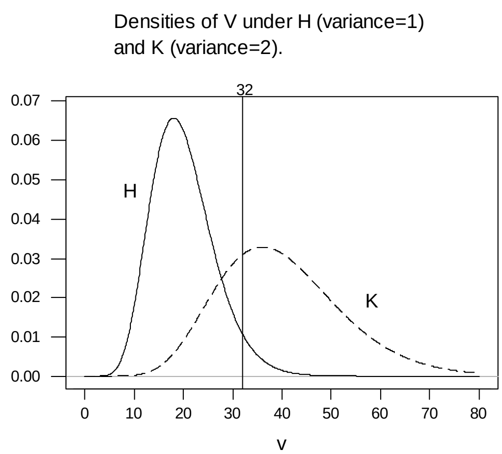
$$D(y) = \left| F\left(\frac{2(\ln y - n\ln q)}{(1-q^2)}\right) - F\left(\frac{2(\ln y - n\ln q)q^2}{(1-q^2)}\right)\right|,$$

where F is the distribution function of a χ_n^2 random variable.

Example 10.1

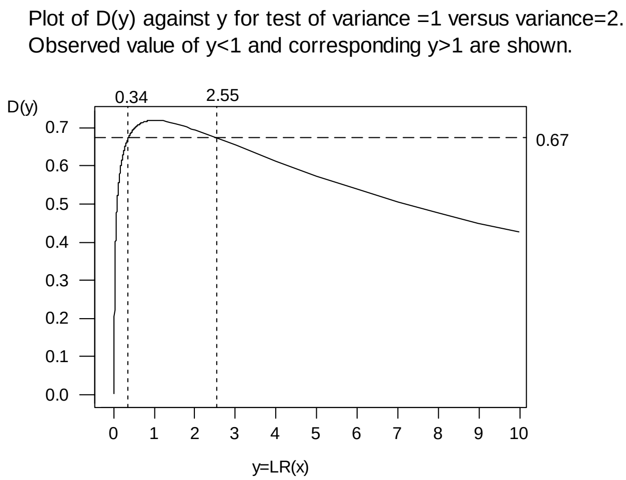
Let $X_1, \ldots, X_{20}$ be a random sample from a $N(0,\sigma^2)$ population and suppose we want to test the null hypothesis $\sigma^2 = 1$ against the alternative $\sigma^2 = 2$ (i.e. $q = \sqrt{2}$). The densities of $V = \sum_{i=1}^{20} X_i^2$ under H and K are shown below.

Figure 10.1



The LR, $y \in (0, 2^{10}) \equiv (0,1024)$. Suppose that we observe data such that $v = \sum_{i=1}^{20} x_i^2 = 32$ (shown on the above plot). From the plot, it is apparent that the likelihood ratio of this data is not particularly small: $y = 2^{10} e^{-8} = 0.3435 = \frac{1}{2.9}$ indicating weak evidence against H. However the conventional p-value is $P_H(V > 32) = 4.33\%$ usually interpreted as strong evidence against H relative to K. Inserting $q = \sqrt{2}$, $n = 20$ and $\ln y = \ln(0.3435)$ into the formula for $D(y)$, we find $D(0.3435) = 0.6733$. We can numerically derive the fact that $0.6733 = D(2.5468)$ (see below).

Figure 10.2



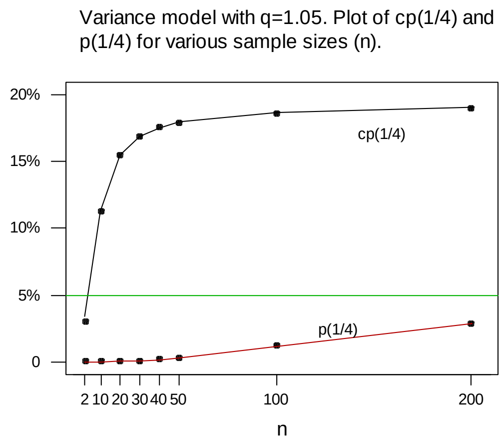
Thus $\pi(0.3435) = 2.5468$ and hence $cp(0.3435) = \frac{0.3435(2.5468-1)}{(2.5468-0.3435)} = 24.12\%$, leading us to a conclusion that is consistent with the observed value of y i.e. the data does not constitute strong evidence against H relative to K.

Large samples.

Right-sided tests on the Normal variance (i.e. $\sigma_2 > \sigma_1$) have likelihood ratios in the range $(0, q^n)$. Thus the range of y-values that are greater than *one* is very limited whenever n is small *and* the hypothesised values are close together (recall $q = \frac{\sigma_2}{\sigma_1}$). For example, suppose that $\sigma_2 = 1.05 \times \sigma_1$ and $n = 2$, then $y \in (0, 1.1025)$ and, for all $y < 1$, the value of $\pi(y)$ is bounded above by a value that is close to *one* (i.e. 1.1025). Since $cp(y) = \frac{y[\pi(y)-1]}{[\pi(y)-y]}$ is increasing in $\pi(y)$ for any fixed y, even the cp-values of moderately large y will tend to be low in such cases. For instance, the observation $y = \frac{1}{4}$ is not significant evidence against H relative to K (from a likelihood point of view), but $\pi(\frac{1}{4}) < 1.1025$ and hence $cp(\frac{1}{4}) < \frac{0.25(0.1025)}{0.8525} = 3\%$. This is a significant value, but not as significant us the unconditional p-value which is $(1.2 \times 10^{-5})\%$.

As n increases so does the bound, $q^n (>1)$. To see what affect this has on the test result, we find $\pi(\frac{1}{4})$ and hence $cp(\frac{1}{4})$ for various n between 2 and 10000 when $q = 1.05$; we also show the conventional p-value of $y = \frac{1}{4}$ in each case.

Figure 10.3



For n in the range 2 to 200, we note that both the cp-value of $\frac{1}{4}$ and the p-value increase as n increases although the p-value is a great deal smaller and is still less than 5% when $n = 200$ while the cp-value is greater than 10% for all $n \geq 10$. However the more interesting result appears when we let the sample size increase to 10000.

Figure 10.4

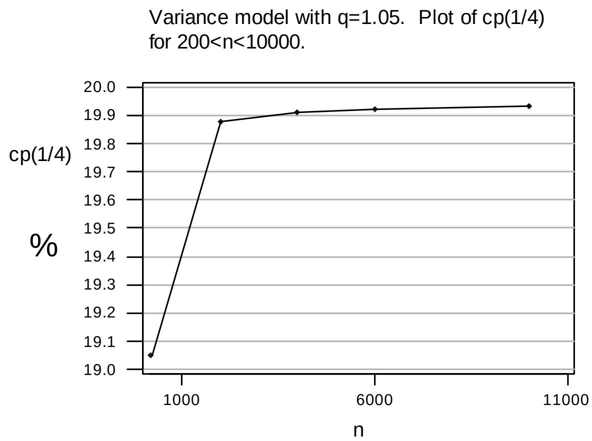
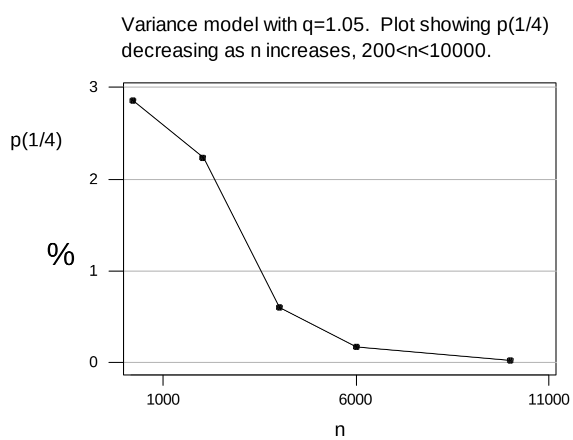


Figure 10.5



The cp-value of $y = \frac{1}{4}$ for $n > 200$ shows the same pattern as for smaller n, increasing (slightly) as n continues to increase. But the p-value of $y = \frac{1}{4}$ behaves quite differently: having achieved a maximum value (still less than 5%) at some n close to 200, it then declines steadily as n increases further, until it is (again) indistinguishable from *zero*. We can show why this happens.

Let F be the distribution function of a χ_n^2 variable, then the conventional p-value is defined as:

$$\begin{aligned} p(y) &= P_H(Y < y) \\ &= 1 - F\left(\frac{2q^2[\ln y - n\ln q]}{(1-q^2)}\right). \end{aligned}$$

When $q > 1$ and $y < q^n$, $\dfrac{2q^2[\ln y - n\ln q]}{(1-q^2)} \to \infty$ as $n \to \infty$, and hence $\lim_{n\to\infty} p(y) = 1 - F(\infty) = 0$.

Note that this is true for all finite values of y, no matter how large. This is an example of the phenomenon that we discussed in Chapter 3 for the Normal location model. As the sample size and, hence, the power of a test increase, the distributions under the two hypotheses become increasingly far apart with the result that we eventually obtain extremely small p-values even for data with a large likelihood ratio. By contrast, $cp(\frac{1}{4})$ appears to converge to some value $\leq 25\%$, as n increases[2]. (Both $cp(y)$ and $p(y)$ are bounded above by y – this is true generally.)

We have shown that, when $q = 1.05$, the poor quality of the conditional inference for small n can be overcome by increasing the sample size, which does not have to be very large before we get reasonable conditional inferences from $y = \frac{1}{4}$. In comparison, the conventional inference deteriorates as n increases and does not give an accurate interpretation of $y = \frac{1}{4}$ for any n.

[2] It is tempting to wonder if $cp(y) \to \frac{y}{(1+y)} = 20\%$. Note however that, although the distribution of V tends to a Normal as n increases, the limiting distribution is $N(n\sigma^2, 2n\sigma^4)$ and the distributions under the two hypotheses have different variances as well as different means. Thus we cannot, simply, use the log-symmetric case to show that the cp-value converges to this limit.

10.2 Tests on the mean of an Exponential population.

Let $X \sim Expo(\theta)$ where $\theta = E(X)$. The density of X is given by:

$$f_X(x;\theta) = \tfrac{1}{\theta} e^{-x/\theta}, \; x > 0, \; \theta > 0.$$

Exponential distributions are widely used to model time-related variables such as the length of telephone calls and they play a prominent role in the *Poisson Process*.

For any two hypotheses H: $\theta = \theta_1$ and K: $\theta = \theta_2$, the likelihood ratio statistic, based on a single observation, x, is:

$$\begin{aligned} y = LR(x) &= \tfrac{\theta_2}{\theta_1} \exp\{-x(\tfrac{1}{\theta_1} - \tfrac{1}{\theta_2})\} \\ &= q \cdot \exp\{(\tfrac{(1-q)}{\theta_1 q})x\}, \end{aligned}$$

where $q = \frac{\theta_2}{\theta_1} > 0$. y is a one-to-one function of x, increasing when $\theta_2 < \theta_1$ ($q < 1$) and decreasing when $\theta_2 > \theta_1$ ($q > 1$). When $q < 1$, $y \in (q, \infty)$ and when $q > 1$, $y \in (0, q)$.

The conventional p-value.

For this model, the conventional p-value for testing two simple hypotheses is:

$$\text{p-value}(x) = \begin{cases} 1 - e^{-x/\theta_1}, & \theta_2 < \theta_1 \\ e^{-x/\theta_1}, & \theta_2 > \theta_1. \end{cases}$$

This can be re-written in terms of y as:

$$p(y) = \begin{cases} 1 - (\frac{y}{q})^{\frac{q}{(q-1)}}, & q < 1 \\ (\frac{y}{q})^{\frac{q}{(q-1)}}, & q > 1. \end{cases}$$

The value y_c such that $p(y_c) = 5\%$ is the critical likelihood ratio for the conventional 5% test – we reject H at 5% whenever we observe $y \le y_c$. For our conclusions to be consistent with the standard interpretation of likelihood ratio, it must be the case that y_c is reasonably small (e.g. less than $\frac{1}{8}$ or $\frac{1}{16}$). The following plots show the value of y_c for a wide range of values of $q = \frac{\theta_2}{\theta_1}$.

Figure 10.6

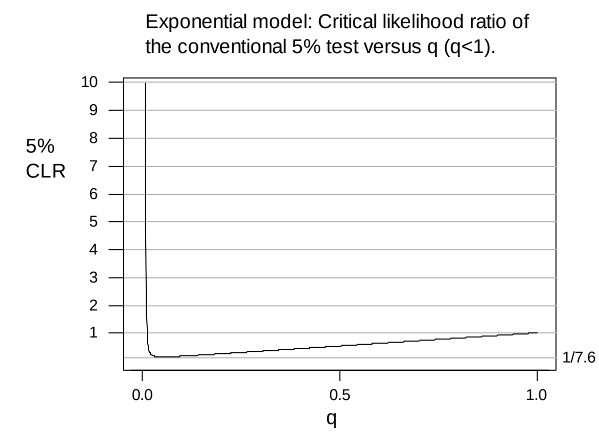
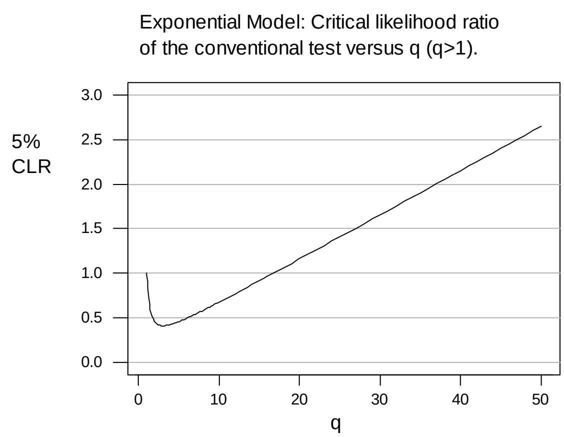


Figure 10.7



The smallest value of y ever required to get a p-value of 5% is $\frac{1}{7.6}$ when $q = 0.0513$. For other values of q, larger y-values – often greater than *one* – are sufficient to give a significant result. Note that as $q \to 1$ (from above or below) the value $y_c \to 1$ and this is true for all significance levels, not only 5%; as the hypothesised values become closer together, it is possible to get (any) significant p-value for likelihood ratios that are close to *one*, that is, close to neutral evidence. Also as $q \to 0$ or $q \to \infty$, the critical likelihood ratio increases without bound, from which it follows that any value of y, no matter how large, will be significant (at any value of α, no matter how small) for *some* hypothesis test.

Exhaustive conditional inference.

The distribution function of Y can be derived from that of X to give the ancillary DDF statistic:

$$\begin{aligned} D(y) &= \left| \left(\frac{y}{q}\right)^{\frac{1}{(q-1)}} - \left(\frac{y}{q}\right)^{\frac{q}{(q-1)}} \right| \\ &= \left| \left(\frac{y}{q}\right)^{\frac{1}{(q-1)}} \left\{1 - \frac{y}{q}\right\} \right|. \end{aligned}$$

In general we cannot solve $D(y) = D(\pi(y))$ for $\pi(y)$ analytically, except for certain values of q where $D(y)$ is a polynomial with known solutions. Note that the value of q defines an inference class for both exhaustive conditional and conventional inferences. The cp-value (or p-value) of any likelihood ratio, y, depends on θ_1 and θ_2 only through q.

Example 10.2

Test H: $\theta = \theta_1$ against K: $\theta = 2\theta_1$ (for any θ_1), i.e. $q = 2$.

Hence $D(y) = (\frac{y}{2})\{1 - \frac{y}{2}\} = -\frac{1}{4}y^2 + \frac{1}{2}y$ ($y \in (0,2)$) and the equation $D(y) = a$ is:

$$\begin{aligned} & y^2 - 2y + 4a = 0 \\ \Rightarrow\ & y_1 = 1 - \sqrt{1-4a} \text{ and } y_2 = 1 + \sqrt{1-4a} = 2 - y_1 \\ \Rightarrow\ & \pi(y) = 2 - y. \end{aligned}$$

Since we have a general formula for the pairing function, we can also obtain one for the cp-value as a function of y:

$$cp(y) = \begin{cases} \frac{y[(2-y)-1]}{[(2-y)-y]} = \frac{y}{2}, & y \in (0,1) \\ 100\%, & y \in (1,2). \end{cases}$$

This compares with the unconditional p-value, $p(y) = \frac{y^2}{4}, \ \ y \in (0,2)$.

The following table compares the p-value and cp-value for various outcomes.

Table 10.1

$y = LR(\text{data})$	Equivalent coin toss result[3].	$cp(y)\%$	$p(y)\%$
$\frac{1}{64}$	*hhhhhh*	0.78%	0.006%
$\frac{1}{16}$	*hhhh*	3.125%	0.1%
$\frac{1}{4}$	*hh*	12.5%	1.6%
$\frac{1}{2}$	*h*	25%	6.25%
$\frac{1}{1.6}$	*	31%	9.8%
$\frac{1}{1.1}$	*	45%	20.7%

The conventional p-values are small even where the likelihood ratio is moderately large and indicates that there is little evidence against H relative to K.

We take the opportunity afforded by this case to illustrate the point we made in the log-symmetric case, namely, that the conventional p-value is the mean of a number of conditional probabilities only one of which (the cp-value) corresponds to the observed value of the DDF statistic (the particular 'sub-experiment' performed). In this case, for any $y_0 \in (0,1)$, $p(y_0) = \frac{y_0^2}{4}$ and $cp(y_0) = \frac{y_0}{2}$. For convenience, we use an ancillary

[3] For testing H: *Coin is fair* against K: *Coin is double-headed*.

statistic equivalent[4] to the DDF statistic, i.e. $A = |Y-1| = \{1-4D(y)\}^{\frac{1}{2}}$, and note that $A \sim Uni(0,1)$, that is, $f_A(a)=1,\ a \in (0,1)$ (under both hypotheses).

We will show that

$$\begin{aligned} p(y_0) &= P_H(Y \le y_0) \\ &= \int_a \vec{P}_H(Y \le y_0 \mid A=a) \cdot f_A(a)da \\ &= E_A[\vec{P}_H(Y \le y_0 \mid A)]. \end{aligned}$$

First note that:

$$\vec{P}_H(Y \le y_0 \mid A=a) = \begin{cases} \vec{P}_H(Y = 1-a \mid A=a), & 1-a \le y_0 \\ 0, \text{ otherwise.} \end{cases}$$

For example, suppose that $y_0 = \frac{1}{3}$. If we condition on $A = |Y-1| = 0.8$, we find that $\vec{P}_H(Y \le \frac{1}{3} \mid A = |1-Y| = 0.8) = \vec{P}_H(Y = 0.2 \mid A = 0.8)$ since 0.2 is the only value less than $\frac{1}{3}$ consistent with $A = 0.8$. On the other hand, $\vec{P}_H(Y \le \frac{1}{3} \mid A = 0.6)$ must be *zero* since there is no value of y less than $\frac{1}{3}$ that is consistent with $A = |Y-1| = 0.6$.

By definition, $\vec{P}_H(Y = 1-a \mid A=a) = cp(1-a)$ and hence:

$$\vec{P}_H(Y \le y_0 \mid A=a) = \begin{cases} \frac{(1-a)}{2}, & 1-y_0 \le a < 1 \\ 0, & 0 < a < 1-y_0. \end{cases}$$

Thus,

$$\begin{aligned} &\int_a \vec{P}_H(Y \le y_0 \mid A=a) \cdot f_A(a)da \\ &= \int_{1-y_0}^{1} \{\tfrac{(1-a)}{2} \times 1\}da + \int_0^{1-y_0} \{0 \times 1\}da \\ &= \tfrac{y_0^2}{4} = p(y_0). \end{aligned}$$

---

[4] That is, A is a one-to-one function of $D(Y)$, which catorgorises the values of y the same way and produces the same $\pi(\cdot)$ and $cp(\cdot)$ functions.

The conventional p-value is the average (over a) of the probabilities $\vec{P}_H(Y \le y_0 \mid A = a)$ and this average is used to interpret the significance of the observation, y_0, despite the fact that whenever we observe $Y = y_0$, we must also have observed $A = 1 - y_0$ so that $\vec{P}_H(Y \le y_0 \mid A = 1 - y_0) = cp(y_0)$ is the only one of these conditional probabilities that is relevant. Note also how the *zeros* in the average pull the p-value *down*; these are associated with an increasingly large proportion of the distribution of A (or the DDF statistic) when y_0 is small. For example, when $y_0 = \frac{1}{4}$, $\vec{P}_H(Y \le y_0 \mid A = a) = 0$ for all $a \in (0, \frac{3}{4})$ and these values constitute $\frac{3}{4}$ of the probability mass of A. However, even when y_0 is not much less than *one*, the p-value, as an average, is dominated by those probabilities associated with unobserved values of a (and, in the case $q = 2$, all of these probabilities are less than $cp(y_0)$ since $cp(\cdot)$ is an increasing function).

Example 10.3

Test H: $\theta = \theta_1$ against K: $\theta = 1.5\theta_1$ (for any θ_1), i.e. $q = 1.5$.

Hence, $D(y) = (\frac{2y}{3})^2\{1 - \frac{2y}{3}\} = -\frac{8}{27}y^3 + \frac{4}{9}y^2$ ($y \in (0,1.5)$).

The equation $D(y) = a$ can be solved to find:

$$\pi(y) = \tfrac{1}{4}\{3 - 2y + \sqrt{3(3-2y)(2y+1)}\},$$

$$\text{and hence } cp(y) = \begin{cases} \dfrac{y\{1 + 2y - \sqrt{3(3-2y)(2y+1)}\}}{\{6y - 3 - \sqrt{3(3-2y)(2y+1)}\}}, & y \in (0,1) \\ 100\%, & y \in (1,1.5). \end{cases}$$

This compares with the unconditional p-value, $p(y) = \frac{8y^3}{27}$, $y \in (0,1.5)$.

The following table compares the p-value and cp-value for various outcomes.

Table 10.2

$y = LR(\text{data})$	Equivalent coin toss result.	$cp(y)\%$	$p(y)\%$
$\frac{1}{64}$	*hhhhhh*	0.53%	0.0001%
$\frac{1}{16}$	*hhhh*	2.17%	0.007%
$\frac{1}{4}$	*hh*	9.55%	0.46%
$\frac{1}{2}$	*h*	21.1%	3.7%
$\frac{1}{1.6}$	*	27.6%	7.2%
$\frac{1}{1.1}$	*	44%	22%

Again we see that the conventional p-values are unreasonably small when there is not much evidence against H: even $y = \frac{1}{2}$ returning a significant p-value. We might consider that the cp-value of less than 10% for $y = \frac{1}{4}$ is also too small and we will see that when q is close to, but larger than, *one*, the cp-value tends to overstate the significance of the result, though not as badly as the p-value. This will be further illustrated in the next example; in the section on Gamma distributions, we suggest a solution to this problem.

Example 10.4

Consider the case where $q = 1.01$, for example, a test of $\theta = 20.0$ against $\theta = 20.2$, based on a single observation. In this case, E. C. inference breaks down as it did when we tested some hypotheses about the Normal variance using $n = 1$.

Since $y \in (0,1.01)$, it follows that, for all $y < 1$, $\pi(y) \in (1,1.01)$ and hence $0 < cp(y) < \frac{0.01y}{(1.01-y)}$. Thus $cp(\frac{1}{2}) < 1\%$, seeming to indicate that a likelihood ratio of *one half* is significant evidence against H relative to K. This problem is insuperable as long as the hypothesised values are extremely close together ($\theta_2 > \theta_1$) and the data involves a sample of size $n = 1$. The cp-value is based not only on the observed likelihood ratio, y_0, but also on an unobserved value, $\pi(y_0)$ and, because of this, our method is not consistent with the likelihood principle. There is, thus, the potential for situations, such as the present example, where our interpretations of y and $cp(y)$ are not in agreement, though this can never happen when $y > 1$. (The conventional p-value is far worse; in this case, $p(\tfrac{1}{2})$ is of the order of $10^{-29}\%$.)

No small likelihood ratios.

When $q < 1$, no data has a likelihood ratio less than q since $y \in (q, \infty)$. If q is not very small, it follows that no data has a likelihood ratio that constitutes significant evidence against H relative to K. Any inference made in such circumstances should reflect this fact. In this section, we consider the cases $q = \tfrac{1}{2}$ and $q = \tfrac{2}{3}$ for which we have complete analytic results.

Example 10.5

(a) Let $q = \tfrac{1}{2}$.

We are testing hypotheses of the form H: $\theta = \theta_1$ versus K: $\theta = \tfrac{1}{2}\theta_1$ (for any θ_1) and $y \in (\tfrac{1}{2}, \infty)$ which indicates that there is no observable data that constitutes evidence against H as strong as the evidence from a single *head* in the coin toss case.

We already have the pairing function for the case where $q = 2$, i.e. $\pi_{q=2}(y) = 2 - y$. In the present case the hypotheses are the opposite way around, so we can use the result given in §9.5 for finding the pairing function for the swapped hypotheses. Thus $\pi(y) = \{\pi_{q=2}(\tfrac{1}{y})\}^{-1} = \{2 - \tfrac{1}{y}\}^{-1} = \tfrac{y}{2y-1}$ and it follows that

$$cp(y) = \begin{cases} \frac{y[\pi(y)-1]}{[\pi(y)-y]} = 50\%, & \tfrac{1}{2} < y < 1 \\ 100\%, & y > 1. \end{cases}$$

The conventional p-value is: $p(y) = 1 - \frac{1}{2y}$, $\forall y \in (\frac{1}{2}, \infty)$. When y is *relatively* small (i.e. $y \to \frac{1}{2}$ from above) the p-value approaches *zero*, but the cp-value is never less than 50% which is consistent with the fact that there can never be strong evidence against H.

(b) Let $q = \frac{2}{3}$.

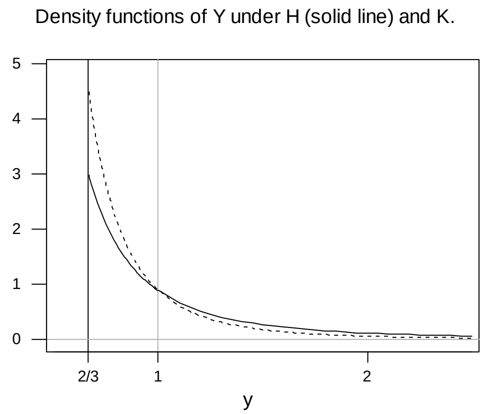
The hypotheses are of the form H: $\theta = \theta_1$ versus K: $\theta = \frac{2}{3}\theta_1$ (for any θ_1) and $y \in (\frac{2}{3}, \infty)$. A likelihood ratio of *two-thirds* is such weak evidence that it cannot be characterised by the paradigm coin-tossing case that we have generally used. Instead consider the following scenario.

There are two dice. Each of the dice is fair regarding each of its six faces but Die A has faces that are numbered $\{2,2,2,2,6,6\}$ and Die B has all six faces numbered '2'. One of the two dice is randomly selected (with a 50:50 probability[5]). The chosen die is rolled repeatedly. For testing H_A: 'The chosen die is Die A' against H_B: 'The chosen die is Die B', the outcome '$\underbrace{2,2,2,...2}_{n}$' has a likelihood ratio of $y = (\frac{2}{3})^n$.

Thus the value $y = \frac{2}{3}$ (which is the strongest evidence we can find against the null hypothesis in the Exponential case with $q = \frac{2}{3}$) is equivalent to the weight of evidence against H_A obtained from a *single* roll of the die resulting in the outcome '2'. This evidence is extremely weak and we can verify this by comparing the densities of Y under H and K when $q = \frac{2}{3}$.

[5] This probability does not affect the likelihood ratio and hence the evidence in the data. However different prior probabilities are capable of biasing our perception of the meaning of the data, hence we stipulate these values.

Figure 10.8



For $y<1$, the density under K is greater than that under H, but the two densities are so alike that no value of y is much more consistent with K than with H. However the conventional p-value of (say) $y=0.68$ is 3.9%.

We already have the pairing function for the case where $q=1.5$ from which we can derive the pairing function for this case:

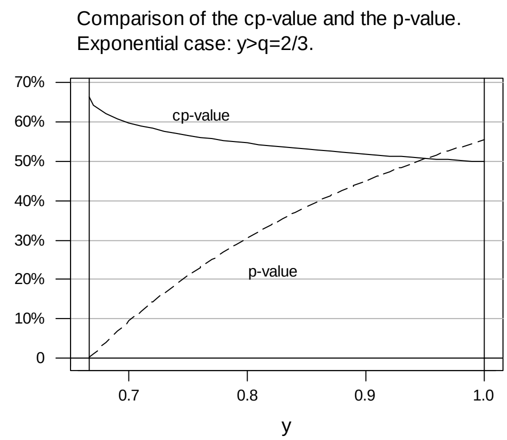
$$\pi(y)=\frac{4y}{\{3y-2+\sqrt{3(3y-2)(2+y)}\}}.$$

Hence, for $\frac{2}{3}<y<1$,

$$cp(y)=\frac{\{y+2-\sqrt{3(3y-2)(2+y)}\}}{\{6-3y-\sqrt{3(3y-2)(2+y)}\}}.$$

This compares with the conventional p-value: $p(y)=1-(\frac{4}{9y^2}),\ y\in(\frac{2}{3},\infty)$. The following plot shows the conventional and conditional p-values as functions of y.

Figure 10.9



The first point to note is that $cp(y)$ is a decreasing function[6] of y. This conflicts with the interpretation we have of the likelihood ratio as a measure of the evidence favouring H relative to K. However, the conflict is real at a practical level only if we insist that '$cp(y_1) < cp(y_2)$' must be interpreted as 'y_1 is stronger evidence against H relative to K than y_2' *regardless* of the actual cp-values, i.e. regardless of whether either of the values is at all significant. Despite the fact that the cp-value is decreasing, the statements '$50\% < cp(y) < 66.6\%$' and '$\frac{2}{3} < y < 1$' are consistent to the extent that both imply that any evidence against H relative to K is extremely weak.

Note also that for values of y close to *one*, the cp-value is less than the p-value. In all the other examples we have considered, the p-value has been smaller than the cp-value and, therefore, sometimes significant when the cp-value is not. This might have lead us to conjecture that it is always so. However, this case and the gradient model (considered below) provide counter-examples to this conjecture. It is true that in all these cases, neither the p-value nor the cp-value is remotely significant so we have no

[6] It is impossible for the (unconditional) p-value to be a decreasing function of the likelihood ratio statistic, since the p-value is simply the (cumulative) distribution function, under the null hypothesis, of the likelihood ratio statistic at the point observed. However, when we condition on an ancillary statistic, we are no longer dealing with a single probability distribution. The distribution of $Y \mid A = a$ is different (under any given hypothesis) for each value of a, and, when we use an EAS, any two distinct values of y (<1) are associated with distinct values of a and, hence, with distinct distributions. If Y is continuous and $y_1 < y_2 < 1$, it follows that $P(Y < y_1)$ is less than $P(Y < y_2)$, but it does not follow that $P_{a_1}(Y < y_1)$ is necessarily less than $P_{a_2}(Y < y_2)$ when P_{a_1} and P_{a_2} refer to different (conditional) distributions.

counter-example to a claim that *it is not possible for the conditional test to yield a significant result when the unconditional test does not*. In fact, it can be shown that, in the gradient model, the cp-value is less than the p-value only when it is greater than 50%, so that neither measure is significant. The question of whether it is possible for the conditional test to deliver a significant result when the conventional test does not remains open: we have no instances of this happening but no proof that it cannot happen.

Despite these oddities, the cp-values are consistent with likelihood ratio values in showing that no data constitutes significant evidence against H relative to K when $q = \frac{2}{3}$. The conventional p-values, on the other hand, are significantly small (tending to *zero*) for y close to $\frac{2}{3}$.

Bounds on the cp-value for a range of Exponential cases.

For any observed likelihood ratio, y_0, we can find $cp(y_0)$ numerically, but since we cannot find $\pi(\cdot)$ and hence $cp(\cdot)$ analytically (for more than a few values of q), we cannot discover the general nature of the cp-functions. The following bounds are useful in providing some extra information.

The following general result compares the cp-value for the Exponential model ($0 < q < 1$) with that in the log-symmetric case.

***Claim*:** When $q < 1$, $cp(y) > y/(1+y)$.

Proof.

When $q < 1$, $D_q(y) = (\frac{y}{q})^{\frac{1}{q-1}}\{\frac{y}{q} - 1\}$.

For $y < 1$, let

$$\begin{aligned}\Delta_q(y) &= D_q(y) - D_q(y^{-1}) \\ &= (\tfrac{y}{q})^{\frac{1}{q-1}}\{\tfrac{y}{q} - 1\} - (\tfrac{y^{-1}}{q})^{\frac{1}{q-1}}\{\tfrac{y^{-1}}{q} - 1\}.\end{aligned}$$

Then $\frac{d}{dy}\Delta_q(y) = \Delta_q{}'(y)$ can be written as:

$$\Delta_q{}'(y) = \frac{q^{\frac{1}{1-q}}}{(1-q)}(y^{\frac{-1}{1-q}})(y^{-1}-1)(1-y^{\frac{1+q}{1-q}}).$$

Since $0 < y < 1$ and $q < 1$, all the components of this product are positive, and hence $\Delta_q{}'(y) > 0$ for all such y and q.

Thus $\forall q < 1$,

$$\max_{y<1} \Delta_q(y) < \Delta_q(1) = D_q(1) - D_q(1) = 0$$
$$\Rightarrow \Delta_q(y) < 0,\ \forall y < 1.$$

Hence $\forall y < 1$,

$$\mathrm{D}_q(y) < D_q(y^{-1})$$
$$\Leftrightarrow D_q(\pi_q(y)) < D_q(y^{-1})$$
$$\Leftrightarrow \pi_q(y) > y^{-1},$$

since $\pi(y)$ and y^{-1} are both greater than *one* and the function $D(\cdot)$ is decreasing in this range.

Hence also,

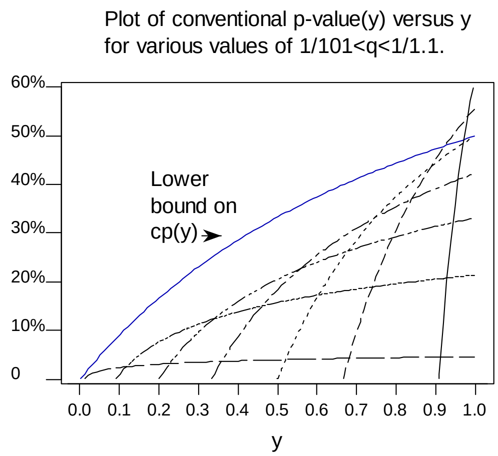
$$cp_q(y) = \frac{y(\pi_q(y)-1)}{(\pi_q(y)-y)} > \frac{y(y^{-1}-1)}{(y^{-1}-y)} = \frac{y}{1+y}.$$

From this it follows that, for all $q < 1$, $y/(1+y) < cp_q(y) < y$, since y is an upper bound on all conditional (and unconditional) p-values of y. This is a convenient result for approximating the cp-value, especially when y is small and the two bounds are close together. For example, for any test where $q < 1$, if we observe data with a likelihood ratio of $\frac{1}{10}$, it follows that the cp-value must lie in the interval $(9\%, 10\%)$, while the cp-value of data with a likelihood ratio of $\frac{1}{28}$ must be between 3.4% and

3.6%. When y is not so small, the lower bound can still establish the non-significance of the result, for example $\frac{y}{(1+y)} > 5\%$ whenever $y > \frac{1}{19}$ in which case we know that $cp(y)$ is also greater than 5%.

We can compare the lower bound on the cp-value (i.e. $\frac{y}{1+y}$) with the conventional p-values associated with various $q < 1$. In the following plot, q is in the range: $\frac{1}{101} < q < \frac{1}{1.1}$.

Figure 10.10



From the values displayed in this graph, it appears that it is only when $q > \frac{1}{2}$ that the p-value is greater than the cp-value for *any* values of y, and since y is always constrained to be greater than q, such values are close to *one* and both types of p-value are large.

As $q \to 0$, the range of y tends to $(0,\infty)$ and

$$\frac{D(y^{-1})}{D(y)} = (y^{-1})^{\frac{-2q}{(1-q)}} y \left[\frac{(1-yq)}{(y-q)}\right] \to 1,$$

hence $\pi_q(y) \to y^{-1}$ and $cp_q(y) \to \frac{y}{1+y}$. Thus the cp-value converges to the lower bound as q decreases.

It can also be shown (by derivations similar to the above) that when $q > 1$,

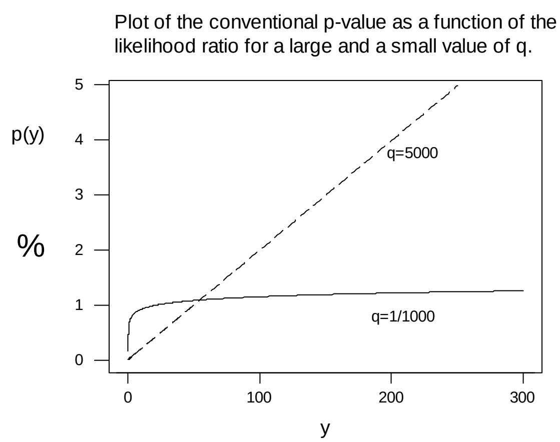
$cp(y) < \frac{y}{(1+y)}$ this being a more constraining *upper* bound on the cp-value than y.

This tells us that any data with a likelihood ratio $\leq \frac{1}{19}$ is significant at the 5% level.

Also, $\lim_{q \to \infty} \frac{D(y^{-1})}{D(y)} = 1$ and hence $cp_q(y) \to \frac{y}{(1+y)}$ (from below) as $q \to \infty$.

Exhaustive conditional inference, in the Exponential case, tends to that of the log-symmetric case when either $q = \frac{\theta_2}{\theta_1} \to 0$ or $q \to \infty$; in both cases the $cp(y)$ approaches the bound shown in the above plot, which is a lower bound for $q < 1$ and an upper bound for $q > 1$. By contrast, $p(y) \to 0$ for all finite y when either $q \to 0$ or $q \to \infty$ (see below).

Figure 10.11



Note that $p(y) < 5\%$ even for very large values of y that constitute strong evidence *in favour of* H relative to K.

10.3 Scenarios involving the Gamma distribution.

Larger samples from an Exponential population.

Let $X_1,\dots,X_n$ be independent and identically distributed random variables, each with an $Expo(\theta)$ distribution. Let $T=\sum_{i=1}^{n} X_i$, then $T \sim Gamma(n,\theta)$ (also called an *Erlang* distribution) with density:

$$f_n(t;\theta)=\tfrac{1}{\Gamma(n)}\theta^{-n}t^{n-1}e^{-t/n},\ t>0,\ \theta>0\ (n\geq 1,\ \text{known}).$$

Hence the likelihood ratio, for testing θ, is given by:

$$y=\tfrac{f_n(t;\theta_1)}{f_n(t;\theta_2)}=q^n\exp\{\tfrac{t(1-q)}{q\theta_1}\},$$

where $q=\frac{\theta_2}{\theta_1}>0$. When $q<1$, $y\in(q^n,\infty)$ and when $q>1$, $y\in(0,q^n)$, thus the support of Y tends to $\mathbb{R}^+$ as $n\to\infty$.

Using a well-known relationship between the Erlang distribution and the Poisson, it is easy to show that:

$$D(y)=|F(n-1;\lambda_1)-F(n-1;\lambda_2)|,$$

where $F(r;\lambda)$ is the value, at the point r, of the distribution function of a $Pois(\lambda)$ random variable and

$$\lambda_1=\frac{\ln(y\cdot q^{-n})}{(1-q)}$$
$$\lambda_2=q\cdot\lambda_1.$$

The special case $n=1$ reproduces the detail of the Exponential models examined in the previous section. In that section we gave the results of exhaustive conditional inferences for various q; we now examine the effect on those inferences of increasing the size of the sample. We examine the cases $q=2$ (which gave good results for

$n=1$), $q=\frac{1}{2}$ and $q=\frac{2}{3}$ (which gave reasonably good inferences but where the cp-value was found to be constant $(y<1)$ in the first case, and decreasing in the second) and $q=1.01$, which gave very poor results, producing significantly small cp-values for moderate y. Because the derivations are somewhat laborious, we consider only a small, representative number of likelihood ratio values, less than *one*. We also compare some of the cp-values with the corresponding conventional p-values which do not work well when $n=1$ (see previous section). Note, when comparing the plots of $cp(y)$ vs y, for various n, that the range of y varies depending on the value of n.

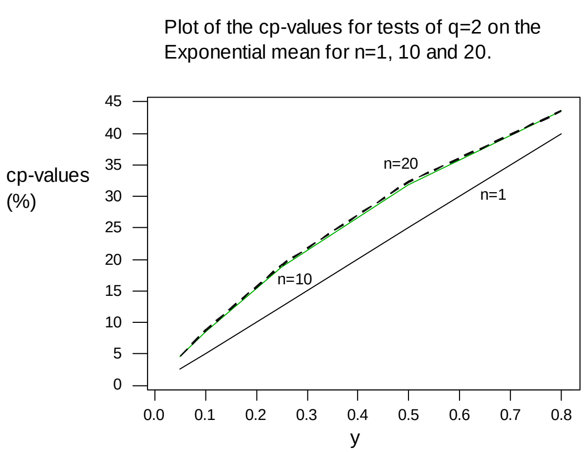
Results for the case q=2.

The following table shows the **cp-values** of various likelihood ratios for three sample sizes, also displayed in the following plot.

Table 10.3

$y=LR(t)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1.25}$ (0.8)
$n=1$	2.5%	5.0%	12.5%	25%	40%
$n=10$	4.4%	8.4%	18.7%	31.9%	43.6%
$n=20$	4.6%	8.7%	19.1%	32.4%	43.6%

Figure 10.12



The cp-values converge fairly soon with virtually no difference between the values when $n=10$ and $n=20$. The values are higher than when $n=1$ but do not, in any way, contradict the natural interpretation of the likelihood ratio values.

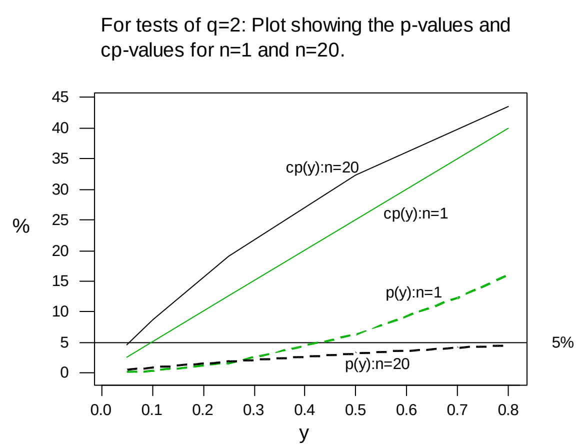
The conventional **p-values**, for $n=1$ and $n=20$ are as follows.

Table 10.4

$y=LR(t)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1.25}$
$n=1$	0.06%	0.25%	1.56%	6.25%	16%
$n=20$	0.43%	0.81%	1.78%	3.12%	4.5%

In contrast to the cp-values, the p-values associated with non-significant likelihood ratios (say $\geq \frac{1}{4}$) become *smaller* as n increases so that although the p-values of the moderately large likelihood ratios, $y=\frac{1}{2}$ and $y=\frac{1}{1.25}$, are not significant when $n=1$, they are significant when $n=20$. Thus increasing the sample size has made the conventional inference worse, rather than better (see **Figure 10.13**, below). We commented on the same phenomenon in the section on the Normal variance; in general, when the sample is large enough to produce a test with very high power, the test is biased in favour of the alternative hypothesis to the extent that it gives significant p-values to data with moderate, or even large, likelihood ratios.

Figure 10.13



Results for the case q = ½.

The **cp-values** for this case are as follows. Recall that when $n=1$, $cp(y)=50\%$, for all $\frac{1}{2}<y<1$.

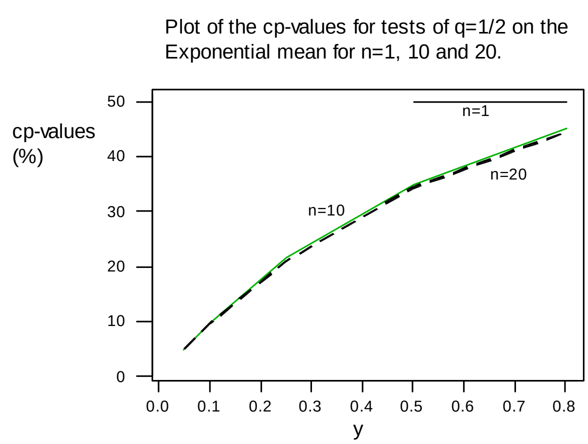
Table 10.5

$y=LR(t)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1.25}$ (0.8)
$n=1$	*	*	*	50%	50%
$n=10$	5.0%	9.7%	21.4%	34.8%	45.2%
$n=20$	4.9%	9.5%	20.9%	33.4%	44.4%

(* $y>q^n$ which equals $\frac{1}{2}$ when $q=\frac{1}{2}$, $n=1$.)

The main advantage of larger n, in this case, is that it increases the range of y so that we can get data that provides strong evidence against H. The cp-value is constant over $\frac{1}{2}<y<1$ when $n=1$ and this seems counter to the notion that the likelihood ratio can be regarded as a measure of the evidence in favour of H, relative to K, but when n is larger, this effect disappears – the cp-value is again increasing in y, so that it distinguishes between different likelihood ratios.

Figure 10.14



Again we see fairly fast convergence of the cp-values as n increases.

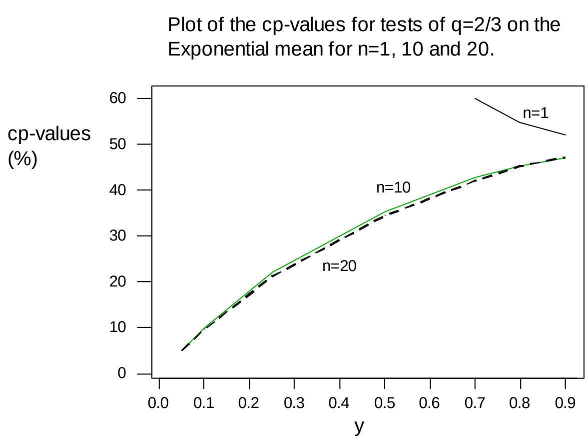
Results for the case q = 2/3.

In this case, we found that, when $n=1$, the cp-value decreases, as a function of y, on the interval $(\frac{2}{3},1)$. This result is very counter-intuitive although the cp-values themselves are reasonably consistent with the likelihood ratios, none of which can be regarded as constituting even moderate evidence against H relative to K. Here, we compare these results (**cp-values**) with those from samples of size 10 and 20.

Table 10.6

$y = LR(t)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1.43}$ (0.7)	$\frac{1}{1.25}$ (0.8)	$\frac{1}{1.11}$ (0.9)
$n=1$	*	*	*	*	60.0%	54.7%	52.0%
$n=10$	5.0%	9.8%	22.0%	35.3%	42.7%	45.2%	47.1%
$n=20$	4.9%	9.6%	21.1%	34.4%	42.0%	45.2%	47.1%

Figure 10.15



The same pattern emerges as when $q=\frac{1}{2}$. For larger n, the cp-values converge quickly and become an increasing function of the likelihood ratio; the values themselves are plausible with a likelihood ratio of $\frac{1}{20}$ being significant at the 5% level while a likelihood ratio of $\frac{1}{10}$ is not.

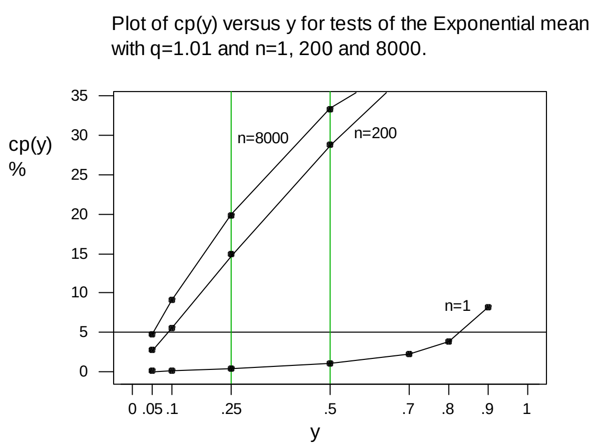
Results for the case q = 1.01.

In this case, when $n=1$, exhaustive conditional inference performs poorly assigning a cp-value of less than 1% to the non-significant likelihood ratio, $\frac{1}{2}$ and (see below) a cp-value of less than 4% even for $y=0.8$. When $q=1.01$, $y \in (0,1.01^n)$ which equals $(0,1.01)$ when $n=1$; this restrictive upper bound on $\pi(y)$ causes the cp-values to be unrealistically low. By increasing n, we can make the upper bound larger and this may improve matters. Since 1.01^n increases only slowly, we use larger values of n than in the previous sections. The **cp-values** for $n=1$, 200 and 8000 are given and plotted below.

Table 10.7

$y=LR(t)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1.43}$ (0.7)	$\frac{1}{1.25}$ (0.8)	$\frac{1}{1.11}$ (0.9)
$n=1$	0.05%	0.11%	0.33%	0.97%	2.24%	3.77%	8.11%
$n=200$	2.68%	5.50%	14.90%	28.89%	38.42%	43.05%	48.33%
$n=8000$	4.74%	9.04%	19.92%	33.40%	41.62%	45.29%	49.28%

Figure 10.16

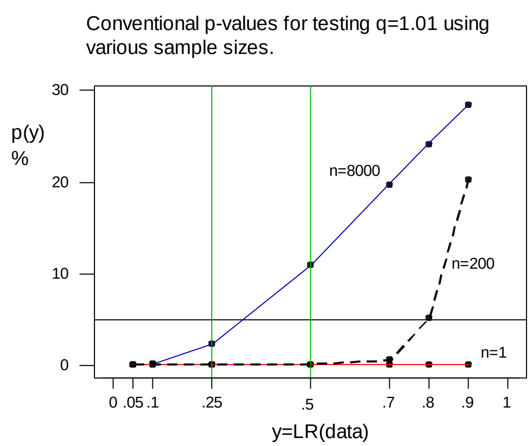


In this case the cp-values take longer to converge, as one would expect. The values for $n=4000$ (not shown) are very close to those for $n=8000$ suggesting that convergence has occurred by this point. The cp-values for large n are consistent with the likelihood ratios: $y = \frac{1}{4}$ and $y = \frac{1}{2}$ have cp-values of approximately 20% and 33% respectively; this contrasts with the highly significant cp-values produced when $n = 1$. Thus it seems that we can overcome the problem of poor results for close hypotheses by increasing the sample size. Below, we give the corresponding results for the conventional **p-value** ('0%' indicates a value less than $10^{-6}\%$).

Table 10.8

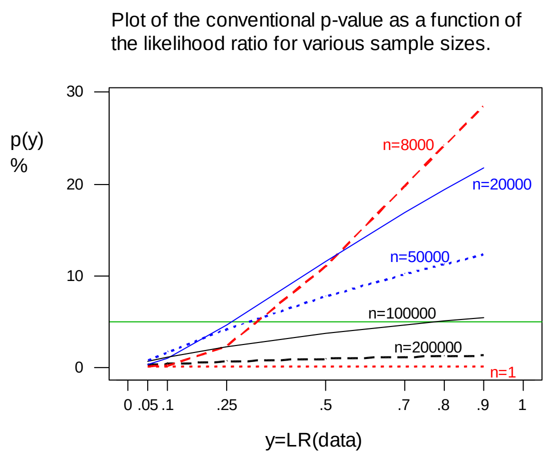
$y = LR(t)$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1.43}$ (0.7)	$\frac{1}{1.25}$ (0.8)	$\frac{1}{1.11}$ (0.9)
$n = 1$	0%	0%	0%	0%	0%	0%	$8.7 \times 10^{-4}\%$
$n = 200$	0%	0%	0%	$3.0 \times 10^{-4}\%$	0.63%	5.19%	20.30%
$n = 8000$	$7.9 \times 10^{-3}\%$	0.13%	2.28%	11.00%	19.78%	24.21%	28.53%

Figure 10.17



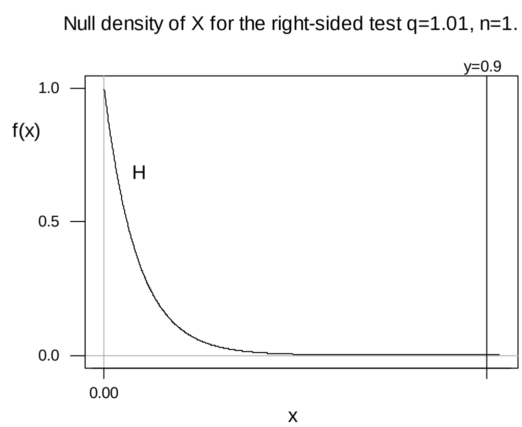
Although we see some improvement in the p-value when n is increased, the performance is still not good. A likelihood ratio of 0.7 is still rated highly significant when $n = 200$ and $y = \frac{1}{4}$ has a p-value well under 5% even when $n = 8000$. To investigate further, we increased the sample size to as high as 200,000 with the following results. (Note that $n = 1$ and 8000 are also displayed in this graph.)

Figure 10.18



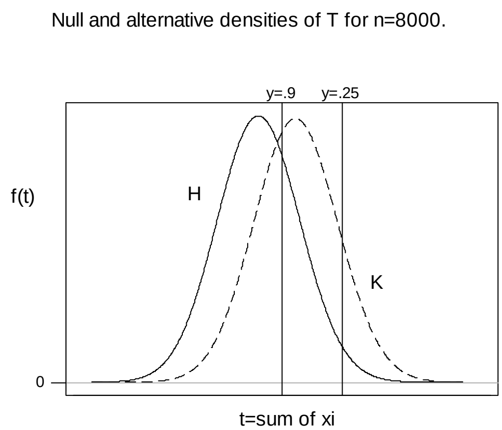
This shows the phenomenon that occurred in tests on the Normal variance. Initially the p-value improves when we increase the sample size, but as we increase it still further, the increasing bias in favour of K asserts itself so that the p-value decreases again. In this case there is little difference between the p-values of y based on $n=1$ and those based on $n=200000$ – both assign small values to likelihood ratios close to *one*. We can understand why this happens by looking at the distributions (under H and K) of the test statistic $T=\sum_{i=1}^{n} X_i$. (The following plots indicate those values of t where y equals 0.25 and 0.9. The p-value is right-sided.)

Figure 10.19



When $n = 1$, we are dealing with the single exponential distribution, i.e. $t = x$, (the value of x corresponding to $y = 0.25$ is off the graph to the right). The p-values of $y = 0.9$ and $y = 0.25$ are obviously close to *zero*. Because the distributions under H and K are so similar (if both were graphed on these axes we could not distinguish between them), the likelihood ratio is close to *one* over the bulk of the distribution(s) and smaller values are all in the extreme right-hand tail, hence the miniscule p-values.

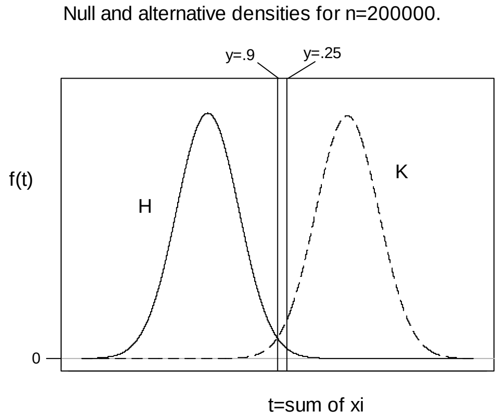
Figure 10.20



The test statistic, T, has mean of $n\theta$ and standard deviation of $|\theta|\sqrt{n}$ (and is approximately Normal for large n). Thus the distributions become further apart as n increases; we can confirm this by looking at the difference between the means measured in terms of either of the standard deviations, i.e. $\frac{|n\theta_1 - n\theta_2|}{|\theta_j|\sqrt{n}} = \sqrt{n} \cdot |\frac{\theta_1 - \theta_2}{\theta_j}|$. This value increases as n increases.

When $n = 8000$, the distributions of T, under the two hypotheses, are clearly distinguishable but overlap (in terms of probability) considerably. The densities cross over each other (i.e. $y = 1$) not far from the centres of the distributions and, as a result, the likelihood ratio is 0.9 not far from the centre of the null distribution giving it a reasonably large p-value.

Figure 10.21



By the time n equals 200000, the distributions overlap only a little and the point at which the densities cross is far from the centres of both and in the right tail of the null distribution, as is the point with a likelihood ratio of 0.9, thus the p-value is again very small. Evidently as the distributions get further apart this effect will become more pronounced: we could make the p-value of any likelihood ratio – no matter how large – very small by using a big enough sample.

Tests on the second Gamma parameter.

Let $X \sim Gamma(\alpha, \beta)$ where α and β are positive parameters, then:

$$f_X(x;\alpha,\beta) = \tfrac{1}{\Gamma(\alpha)}\beta^{-\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}},\ x > 0.$$

The parameter β gives the scale in the sense that, if $V \sim Gamma(\alpha,1)$, then $\beta V \sim Gamma(\alpha,\beta)$. X has mean and variance $\alpha\beta$ and $\alpha\beta^2$, respectively. The Exponential and Chi-squired distributions are subsets of the Gammas; the first corresponds to the $Gamma(1,\theta)$ while the χ_ν is equivalent to the $Gamma(\tfrac{1}{2}\nu, 2)$. In the previous section, we considered various tests on the parameter β (called θ) when α was a known integer, n, a situation that arises when we sample repeatedly from an Exponential population. This is sufficient to demonstrate exhaustive inference on β. We now consider using E. C. inference to perform tests on the unknown α when β

is known. (We could also use E. C. inference to perform tests on $\underset{\sim}{\theta} = (\alpha, \beta)^{\mathrm{T}}$ when neither parameter is known.)

For fixed, known β we test H: $\alpha = \alpha_1$ versus K: $\alpha = \alpha_2$ for which the likelihood ratio is:

$$\begin{aligned} y &= \tfrac{\Gamma(\alpha_2)}{\Gamma(\alpha_1)} \beta^{(\alpha_2 - \alpha_1)} x^{-(\alpha_2 - \alpha_1)} \\ &= \kappa x^{-\delta}, \; x > 0 \end{aligned}$$

where $\delta = \alpha_2 - \alpha_1$ and $\kappa = \beta^{\delta} \frac{\Gamma(\alpha_2)}{\Gamma(\alpha_1)} > 0$, hence $y \in (0, \infty)\ \forall \alpha_1,\ \alpha_2$ and β.

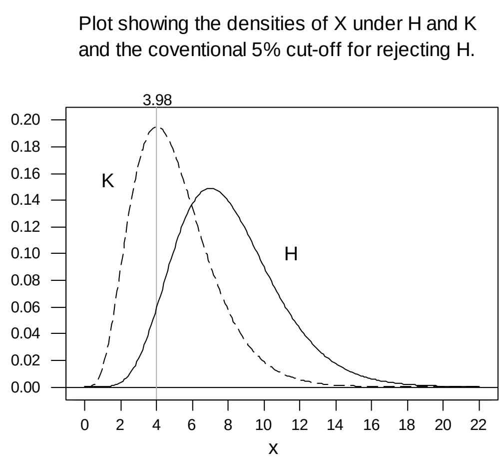
The DDF function is:

$$D(y) = \left| F_{X,K}[(\tfrac{y}{\kappa})^{-1/\delta}] - F_{X,H}[(\tfrac{y}{\kappa})^{-1/\delta}] \right|.$$

Example 10.6

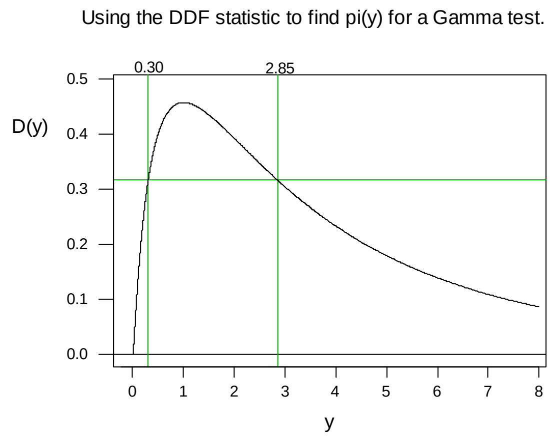
Let $\beta = 1$, in which case α is the mean and also the variance of the Gamma population. Suppose we want to test H: $\alpha = 8$ versus K: $\alpha = 5$ which is a left-sided test in x. The distributions are shown below.

Figure 10.22



A conventional test rejects H in favour of K at the 5% level whenever $x \le 3.98$, however the likelihood ratio of this value is $y = 0.30$ which is not at all significant[7]. Suppose $x = 3.98$ and thus has a conventional p-value of 5%, how does an exhaustive conditional test interpret this datum? In order to calculate $cp(0.30)$, we need to find $\pi(0.30)$ from the DDF function.

Figure 10.23



$D(0.30)$ equals 0.3174 as does $D(2.85)$, hence $\pi(0.30) = 2.85$ and $cp(0.30) = 21.8\%$. Again, the E. C. test result is far more consistent with the likelihood ratio than is the conventional test.

10.4 Tests on the Weibull model.

The various Weibull distributions are widely used in industrial contexts to model the failure rates of equipment and breaking strengths of materials.

If $X \sim Weibull(\alpha, \beta)$, then X has the following density:

$$f_X(x;\alpha,\beta) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp\{-(\tfrac{x}{\beta})^{\alpha}\},\ x>0,\ \alpha>0,\ \beta>0\,.$$

[7] Less evidence against H that *hh* in the coin toss case.

The distribution function is:

$$F_X(x;\alpha,\beta) = 1 - \exp\{-(\tfrac{x}{\beta})^{\alpha}\},\ x \in \mathbb{R}^{+}.$$

As in the Gamma case, β is a scale parameter. When $\alpha = 1$, X has an Exponential distribution with $\theta = \beta$. The hazard function (used in survival analysis) of a Weibull variable is a multiple of $x^{1-\alpha}$ and thus depends on x only through the value of α.

To test competing values of β where α is a fixed, known value, we use the likelihood ratio:

$$y = (\tfrac{\beta_2}{\beta_1})^{\alpha} \exp\{x^{\alpha}(\beta_2^{-\alpha} - \beta_1^{-\alpha})\}.$$

When $\frac{\beta_2}{\beta_1} < 1$, y is increasing in x and $y \in ((\frac{\beta_2}{\beta_1})^{\alpha}, \infty)$, otherwise y is decreasing in x and $y \in (0, (\frac{\beta_2}{\beta_1})^{\alpha})$. Note that, for a given value of $(\frac{\beta_2}{\beta_1})$, the support of Y is larger when α is larger. Using the fact that $X^{\alpha} \sim Expo(\beta^{\alpha})$, we can find the DDF statistic:

$$D(y) = \left| \left(\frac{y}{r}\right)^{\frac{1}{(r-1)}} \left\{ \left(\frac{y}{r}\right) - 1 \right\} \right|,$$

where $r = (\frac{\beta_2}{\beta_1})^{\alpha}$.

Since this has the same form as the Exponential DDF (substituting $r = (\frac{\beta_2}{\beta_1})^{\alpha}$ for $q = \frac{\theta_2}{\theta_1}$) it follows that $cp(y) \to \frac{y}{1+y}$ as $r \to 0$ or $r \to \infty$ and hence this occurs, for any value of $\frac{\beta_2}{\beta_1}$, as $\alpha \to \infty$.

When $(\frac{\beta_2}{\beta_1})^{\alpha}$ is greater than *one* but very close to it, we will have the same problem – small cp-values for moderate likelihood ratios – that we found in the Exponential case with $q = 1.01$. This can be overcome by taking a sufficiently large sample of Weibull observations. The Exponential model is a special case of this model, as shown below.

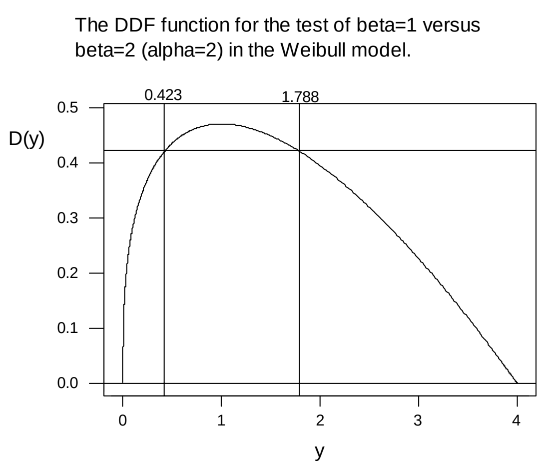
If $X_1,\ldots,X_n$ are independent and identically distributed $Weibull(\alpha,\beta)$ random variables where α is known, then $V=\sum_{i=1}^{n} X_i^{\alpha}$ is a sufficient statistic for β and has a $Gamma(n,\beta^{\alpha})$ distribution, of which β is the only unknown component. In the section on large samples from Exponential populations we carried out tests on β^{α} for the special case $\alpha=1$. Therefore the conditional tests that we performed for $q=1.01$ (i.e. $\beta_2=1.01\beta_1$) are also valid for testing $\beta_2=1.005\beta_1$ given $\alpha=2$, or for testing $\beta_2=1.001\beta_1$ given $\alpha=10$, in the Weibull context[8]. This shows that problems caused by hypotheses that are close together can be solved by increasing the sample size.

The following example involves less extreme values of $(\frac{\beta_2}{\beta_1})^{\alpha}$.

Example 10.7

Suppose $\alpha=2$; we want to test H: $\beta=1$ versus K: $\beta=2$, thus $r=4$ and $y=4\exp(-\frac{3x^2}{4})$ with $D(y)=4^{-\frac{4}{3}}y^{\frac{1}{3}}(4-y)$, $y\in(0,4)$.

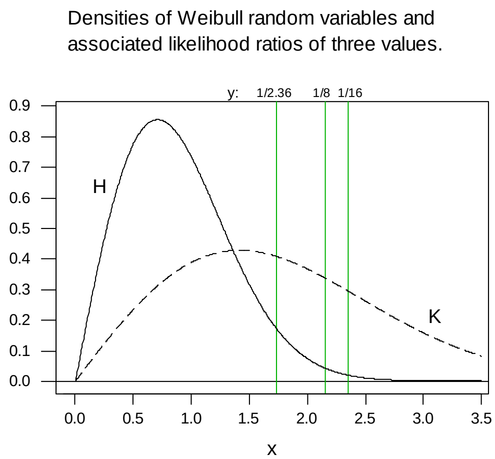
Figure 10.24



[8] Note: $1.005=1.01^{\frac{1}{2}}$ and $1.001=1.01^{\frac{1}{10}}$.

This is a right-sided test in x and the conventional 5% rejection region is $(1.731, \infty)$, however $x = 1.731$ has a likelihood ratio that is not much less than $\frac{1}{2}$ $(0.423 = \frac{1}{2.36})$, as shown in the plot below.

Figure 10.25



For this value and the values associated with likelihood ratios $\frac{1}{8}$ and $\frac{1}{16}$, we have derived the cp-values and p-values.

Table 10.9

x	$y = LR(x)$	$\pi(y)$	$p(y)$	$cp(y)$
1.731	0.423 $(\approx \frac{1}{2})$	1.788	5.0%	21.4%
2.150	0.125 $(\frac{1}{8})$	2.589	1.0%	8.06%
2.355	0.0625 $(\frac{1}{16})$	2.905	0.4%	4.19%

The cp-value is significant at 5% only for the most extreme of the three values, that with a likelihood ratio of $\frac{1}{16}$, in contrast to the p-value. Probably the results would be better still were $n > 1$.

For testing hypotheses about the value of α when β is known, we find exhaustive inferences by using the likelihood ratio:

$$y = \frac{f_X(x;\alpha_1,\beta)}{f_X(x;\alpha_2,\beta)} = (\tfrac{\alpha_1}{\alpha_2})\beta^{-(\alpha_1-\alpha_2)}x^{(\alpha_1-\alpha_2)}\exp\{-(\tfrac{x}{\beta})^{\alpha_1} + (\tfrac{x}{\beta})^{\alpha_2}\}.$$

The bounds on the likelihood ratio are: $y \in (0, \frac{\alpha_1}{\alpha_2})$ when $\frac{\alpha_1}{\alpha_2} > 1$ and $y \in (\frac{\alpha_1}{\alpha_2}, \infty)$ when $\frac{\alpha_1}{\alpha_2} < 1$. As in the Cauchy case, y is not a one-to-one function of x, but instead has a turning point at $x = \beta$. This makes the task of finding the distribution of Y, and hence the DDF function, more laborious and we do not give the details here.

10.5 A better model for studying the Welch phenomena.

In the Welch case, the unconditional approach has been shown to have very undesirable, conditional properties when examined in the light of the universal ancillary[9] statistic R. However, we have shown that the conditional inference, based on R, also has undesirable features because it breaches the sufficiency principle. This only becomes obvious when we look, in detail, at the inferences as they apply to tests of two simple hypotheses, i.e. cases involving binary parameter spaces. The statistic R is not ancillary, in the restricted sense, with respect to any binary parameter space because it is not a function of the likelihood ratio statistic, which is the MSS in such cases. The conditional approach produces different results from data with the same likelihood ratio and is thus in breach of the sufficiency principle. However the situation is complicated by the fact that the unconditional inference shares the same flaw.

If we are dealing with a binary parameter space and an ancillary statistic is not a function of Y (the MSS), then any frequentist conditional method based upon it will

[9] We use the term "universal ancillary" here in order to emphasise the fact that this statistic is ancillary over a large parameter space whereas most of the ancillary statistics we discuss are ancillary only for a given binary parameter space.

breach the SP even if the unconditional approach does not. It is for this reason that frequentists restrict the application of the conditional principle to (ancillary) statistics that are functions of the MSS.

The following artificial example shares the technical simplicity and interesting features of Welch's case, without the problems created by a discrete likelihood ratio statistic.

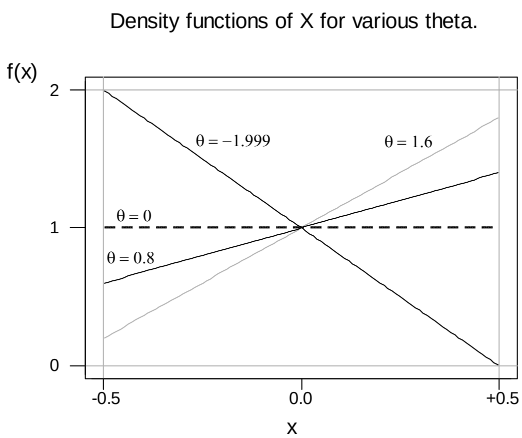
The Gradient Model.

Let X be a variable with density

$$f_X(x;\theta) = \theta x + 1,\ x \in (-\tfrac{1}{2}, \tfrac{1}{2}),\ \theta \in (-2, 2).$$

For all θ, the density is a straight line passing through the point $(0,1)$ and having a gradient of θ. The parameter, θ, is neither a simple location nor a scale parameter, although $E(X)$ is a linear, increasing function of θ. When $\theta = 0$, $X \sim Uni(-\tfrac{1}{2}, \tfrac{1}{2})$.

Figure 10.26



Large values of x are more likely when θ is larger and the likelihood ratio $y = LR(x) = \frac{f(x;\theta_1)}{f(x;\theta_2)}$ is increasing (decreasing) in x when $\theta_2 < \theta_1$ $(\theta_2 > \theta_1)$; the bounds on the value of y are $\frac{(2-\theta_1)}{(2-\theta_2)}$ and $\frac{(2+\theta_1)}{(2+\theta_2)}$. For any $\{\theta_1, \theta_2\}$, Y is a *continuous* variable, unlike in the Welch case, hence it is possible to find a unique, non-arbitrary most powerful rejection region for a test at any significance level.

When $x = 0$, the likelihood (as a function of θ) equals *one* for all θ. Since the likelihood is flat, the only non-arbitrary likelihood interval (LI) for θ is the $\frac{1}{1}$LI containing the entire natural parameter space $(-2, 2)$. By contrast conventional confidence intervals, based on $x = 0$, contain varying amounts of the parameter space depending on the specified coverage. These conventional confidence intervals have some worrying features even when $x \neq 0$.

Unconditional confidence intervals.

The unique, optimal (Neyman-Pearson), γ-level rejection regions for left-sided $(\theta_2 < \theta_1)$ and right-sided $(\theta_2 > \theta_1)$ tests of H: $\theta = \theta_1$ are, respectively:

$$(-\tfrac{1}{2}, l(\theta_1, \gamma)] \text{ and } [u(\theta_1, \gamma), \tfrac{1}{2}),$$

where

$$l(\theta, \gamma) = \tfrac{1}{\theta}\{\sqrt{1 - \theta(1-2\gamma) + \theta^2/4} - 1\}$$
$$u(\theta, \gamma) = \tfrac{1}{\theta}\{\sqrt{1 + \theta(1-2\gamma) + \theta^2/4} - 1\}.$$

It follows that H can not be rejected, in favour of *any* simple alternative hypothesis, at the $\frac{\alpha}{2}$-level as long as:

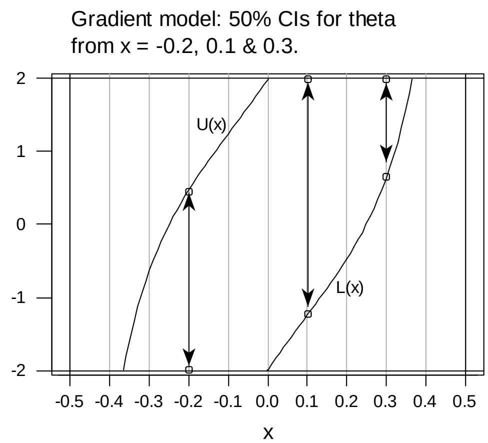
$$l(\theta_1, \tfrac{\alpha}{2}) < x < u(\theta_1, \tfrac{\alpha}{2}).$$

This fact can be used to produce exact, unique, optimal (uniformly most accurate unbiased) $100(1-\alpha)\%$ confidence intervals for θ. The CI based on x is given by $\mathbb{C}(x) \equiv \{\theta : l(\theta, \frac{\alpha}{2}) < x < u(\theta, \frac{\alpha}{2})\ \&\ \theta \in (-2, 2)\}$ which is equivalent to:

$$\mathbb{C}(x) = \left(\frac{2x - (1-\alpha)}{(\frac{1}{4} - x^2)}, \frac{2x + (1-\alpha)}{(\frac{1}{4} - x^2)} \right) \cap (-2, 2).$$

The following plot shows the bounds that produce 50% confidence intervals ($\alpha = 0.5$).

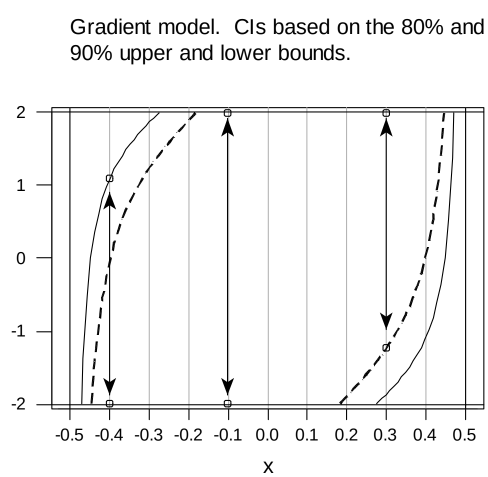
Figure 10.27



The endpoints of each confidence interval can be read from the vertical axis. Note that the intervals are wider when x is close to *zero*.

The following plot shows the bounds for constructing 80% (dashed lines) and 90% (solid lines) confidence intervals.

Figure 10.28



The three intervals shown are (from left to right):

i. The 90% CI for θ based on $x = -0.4$.
ii. The 90% CI for θ based on $x = -0.1$ which is equivalent to the 80% CI for θ based on $x = -0.1$.
iii. The 80% CI based on $x = 0.3$.

We can see that these intervals have a number of counter-intuitive features similar to those observed in the unconditional confidence intervals in Welch's case. Thus, for many values of x, the confidence interval contains all possible values of θ ($\text{CI} \equiv (-2,2)$) even though the stated coverage is not necessarily 100%. Specifically, this is true whenever x is in the interval: $(-\tfrac{1}{2}+\sqrt{\tfrac{\alpha}{2}}, \tfrac{1}{2}-\sqrt{\tfrac{\alpha}{2}})$; for instance, the 90% CI contains all possible values of θ whenever x is in $(-0.276, 0.276)$. The interval $(-\tfrac{1}{2}+\sqrt{\tfrac{\alpha}{2}}, \tfrac{1}{2}-\sqrt{\tfrac{\alpha}{2}})$ is wide, when the coverage is high (since α is small), so there will be many values of x that produce CIs of this kind. Note, from the above plot, that, when $x = -0.1$, both the 80% and 90% confidence intervals contain the entire parameter space and are thus equal to each other as well as to the 100% CI. In fact, the data $x = -0.1$ produces the same interval when the coverage is specified anywhere in the range 68% to 100%. The results are even more dramatic when x is further away from *zero*; for example, when $x = 0.4$, the 50% CI is the empty set (see **Figure 10.27**), as is the 90% CI whenever x is sufficiently close to $\tfrac{1}{2}$ or $-\tfrac{1}{2}$ (**Figure 10.28**). In such cases, there is no chance that the confidence interval contains θ.

Despite these strange properties, the intervals are optimal, in the sense intended by Neyman and Pearson; for example, the 50% intervals do indeed contain θ 50% of the time (in the long run). The reasons for this are as follows. When x is close to *zero*, the 50% confidence intervals are very long and contain all, or nearly all, of the parameter space; they, therefore, contain θ all, or most, of the time and this is enough to offset those occasions when the observation is far from *zero* and the interval is very short or empty and contains θ only rarely, or never.

The Welch case came to be interpreted as showing that an average long-run success or failure rate does not necessarily reveal the aspects of the experimental result that we would regard as most relevant: if a ‘50%’ CI based on (say) $x = 0.2$ falls in a category where the success rate is much higher than 50%, it may be this value, not the ‘50%’ figure, that is more relevant to our interests. Clearly the same issue arises in the present example.

An ancillary statistic.

In Welch’s case, these phenomena are connected to the existence of a universal ancillary statistic (R); this is true in the gradient example as well. If we let $A = |X|$, then $A \sim Uni[0, \tfrac{1}{2})\ \ \forall \theta \in (-2, 2)$, and, since X is the MSS for this model and parameter space, A is a function of the MSS and thus ancillary (in the restricted sense) on the natural parameter space, $(-2, 2)$. We can use the distribution of $X \mid A = a$ to find the conditional coverage of the conventional confidence intervals; this may accord more closely with our intuition about the true meaning of such intervals. Note that A is a continuous variable on $[0, \tfrac{1}{2})$ and the distribution of $X \mid A = a$ is dichotomous (in the limit), since when $A = a$, X is either $\pm a$.

The conditional distribution of $X \mid A = a$ is shown below.

Table 10.10

x	$-a$	$+a$
$\vec{P}_\theta(X = x \mid A = a)$	$\dfrac{(1-a\theta)}{2}$	$\dfrac{(1+a\theta)}{2}$

(Note that $(1-a\theta) > (1+a\theta)$ when $\theta < 0$.)

This can also be written as: $\vec{P}_\theta(X = x \mid A = |x|) = \frac{(1+x\theta)}{2}$.

Conditioning on A.

Since θ is bounded by –2 and 2, it follows that there are bounds on the above probabilities for any given a (and also overall). For all values of θ, the two probabilities must both lie in the interval $(\frac{1}{2}-a, \frac{1}{2}+a)$ and when a is small (x is close to *zero*), both probabilities are close to $\frac{1}{2}$ and to each other. In such a case, the conditional p-value[10] (for any null hypothesis) is close to 50% and there is no basis for rejecting any of the possible values of θ; this is consistent with our observation that the likelihood function is close to flat in such a case. The greatest difference between the two probabilities occurs as $a \to \frac{1}{2}$ and the two probabilities approach $\frac{1}{2}-\frac{\theta}{4}$ and $\frac{1}{2}+\frac{\theta}{4}$ respectively; however, even in this case, we will only achieve a small conditional p-value for some observation if θ_1 is close to 2 or -2; for example, if $a \approx \frac{1}{2}$, one of the possible observations has a probability less than 5% only if $|\theta_1| > 1.8$. Thus it is only when θ_1 is an extreme value (in the context of the natural parameter space for the model) that it is possible to observe any data that has a conditional p-value sufficiently low to justify rejecting H in favour of some alternative.

Is this consistent with the evidence based on the likelihood ratio value? For any given hypothesised values θ_1 and θ_2, the likelihood ratio corresponding to data, x, is $y = \frac{(1+\theta_1 x)}{(1+\theta_2 x)}$ where $x \in (-\frac{1}{2}, \frac{1}{2})$ and sufficiently small values of y will count as evidence against θ_1 relative to θ_2. Consider (WLOG) the case $\theta_2 < \theta_1$, so that the minimum possible value of y is $\frac{(2-\theta_1)}{(2-\theta_2)}$. If θ_1 (the null value) is fixed but we consider different possible values of θ_2 satisfying $\theta_2 < \theta_1$, we can show that $\forall \theta_2 < \theta_1$, $0 < \frac{(2-\theta_1)}{4} < \frac{(2-\theta_1)}{(2-\theta_2)} < 1$; thus, for any x and any $\theta_2 < \theta_1$, y can never be less than $\frac{(2-\theta_1)}{4}$. It is possible to observe y that is significantly small, (say) $y \le \frac{1}{16}$, for *some* θ_2, only when $\frac{(2-\theta_1)}{4} < \frac{1}{16}$, and thus $\theta_1 > 1.75$; again we find that, under this model, we can find strong evidence against H (relative to some alternative), only when the value of θ_1 is

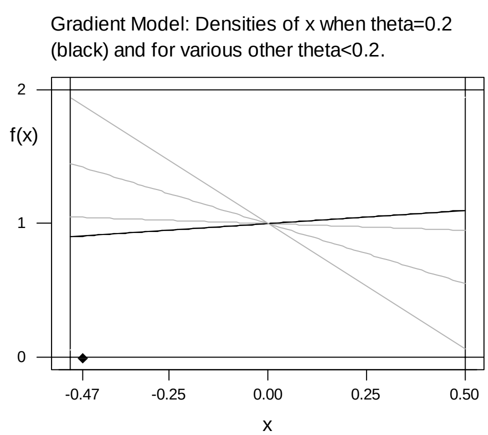
[10] I.e. the p-value conditional on the observed value of *a*.

itself extreme. This makes sense because the densities based on any two non-extreme θ-values, are very similar. Despite this, we can find data with an arbitrarily low *unconditional* (i.e. conventional) p-value for *any* value of θ_1, because Y is a continuous variable.

Example 10.8

Consider left-sided tests (i.e. $\theta_2 < \theta_1$) of the null hypothesis $\theta = 0.2$ with data $x = -0.47$. The following plot shows the density of X under H and some left-sided alternatives. The data is marked on the x-axis. Note that, although the data is clearly more consistent with the alternatives than with H, the difference is not great.

Figure 10.29



The conventional p-value of this data is $F_{X,\theta=0.2}(-0.47) = 2.7\%$, which is significant. On the other hand, for *all* left-sided alternative hypotheses, the likelihood ratio of the data $x = -0.47$ is in the range (0.467, 1). Thus the evidence against H (relative to any $\theta_2 < \theta_1$) is never much more than that which we get from observing a single coin toss resulting in a *head*, in the paradigm case. Does conditioning on the ancillary statistic, $A = |X|$, improve the inference?

The observed value of A is $a = |-0.47| = 0.47$. Conditional upon this value, X has the following distribution if $\theta = 0.2$.

Table 10.11

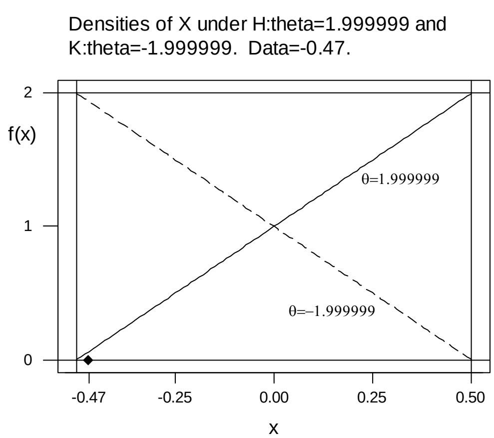
x	-0.47	$+0.47$
$\vec{P}_{0.2}(X = x \mid A = 0.47)$	0.453	0.547

Thus the conditional p-value of the data, under this hypothesis, is 45.3% in contrast to the p-value of 2.7%. The conditional p-value is much more consistent with the range of likelihood ratios.

Example 10.9

Although most of the densities consistent with this model are similar to each other (as a result of the model constraints), it is possible to observe strong evidence with respect to the extreme case hypotheses, $\theta \to -2$ and $\theta \to 2$. For instance, using the data $x = -0.47$, from the previous example, we find that the likelihood ratio for comparing H: $\theta = 1.999999$ with K: $\theta = -1.999999$ is small and has the same significance as the outcome ‘5 *heads*’ in the coin toss example, that is, $LR(-0.47) = \frac{(1-0.47\theta_1)}{(1-0.47\theta_2)} = \frac{1}{32.34}$.

Figure 10.30



We use the conditional distribution of $X \mid A = 0.47$, under H, to find the conditional p-value, as follows.

Table 10.12

x	–0.47	+0.47
$\vec{P}_{2-10^{-6}}(X = x \mid A = 0.47)$	0.030000	0.9700000

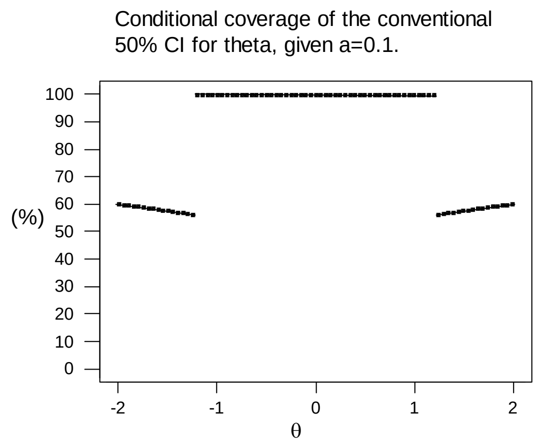
Thus the conditional p-value of $x = -0.47$, for testing H: $\theta = 1.999999$ against K: $\theta = -1.999999$, is 3% and significant. (The p-value is 0.09%.)

The conditional coverage of the conventional confidence intervals.

We have noted that the optimal 50% Neyman-Pearson confidence intervals for θ are very varied in width, even including intervals that cover all or none of the values in the parameter space (-2,2). We may examine the features of these confidence intervals by conditioning on the observed value of $A = |X|$. This can be done with no loss of information since A has the same distribution for all θ.

Consider the case where we observe $A = 0.1$. This value is associated with x being either –0.1 or 0.1, both of which produce a 50% CI that is relatively wide: $(-2.00, 1.25)$ and $(-1.25, 2.00)$ respectively (in a parameter space of $(-2, 2)$). Intuitively we tend to think that the real coverage of such intervals is more than 50%. Suppose we were to simulate a large number of samples (from the model and θ), producing a CI from each sample, and then observing the long-run proportion of instances where the CI contains θ. This proportion is the *coverage* of the CI procedure; in many well-known cases, this value is common to all θ. When we look for the conditional coverage, we consider data arising from the conditional distribution of $X \mid A$ instead of from the unconditional distribution of X. If $a = 0.1$, all of the confidence intervals that are produced will be one or other of the two intervals: $(-2.00, 1.25)$ and $(-1.25, 2.00)$. Since values in the range $(-1.25, +1.25)$ occur in both of these intervals, any θ between –1.25 and +1.25 is contained in all the intervals that can be produced from data consistent with $a = 0.1$; thus, the conditional coverage for these values of θ is 100%. Values of θ between –2.00 and –1.25 are only contained in confidence intervals produced from $x = -0.1$, and the proportion of the samples where $x = -0.1$ depends on θ, thus the coverage is itself a function of θ; the same applies to values of θ between +1.25 and +2.00.

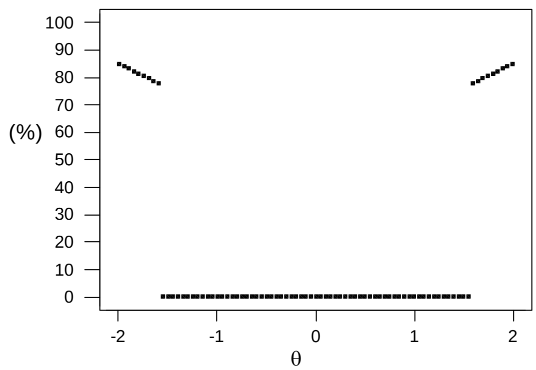
Figure 10.31



It is clear that the figure ‘50%’ does not well describe the reliability of this method since, $\forall\theta$, the conditional coverage > 50%. (It is the average coverage *over* a that is 50%, not the average over θ, so this is no contradiction.) We can compare this with one of the short 50% confidence intervals.

When $a = 0.35$, the intervals associated with the two possible values of x (-0.35 and 0.35) are very short: $(-2.00, -1.57)$ and $(1.57, 2.00)$ respectively. If $-1.57 < \theta < 1.57$, then the coverage of θ, by the confidence intervals, is 0%, since neither of the intervals contains any of these values. On the other hand, the more extreme values of θ are covered more than 50% of the time. That is, if $|\theta|$ is close to *two*, a high proportion of the intervals will contain θ, if not, none of them will. Since we do not know θ, we cannot really use the conditional coverage results (and, if we did know θ, we would not need intervals), but, they do serve to show how unsatisfactory the conventional intervals are.

Figure 10.32



When a is sufficiently large, both of the intervals that can be produced are empty and thus the conventional 50% method has a conditional coverage of *zero* for all possible θ. While there is a sense in which it is obvious that the empty intervals and whole parameter space intervals have a ‘confidence’ of *zero* and 100%, rather than 50%, we have shown that these values can be derived as the conditional coverage rates of the conventional intervals. The fact that these values accord with our intuition provides support for the view that conditioning brings us closer to a realistic interpretation of the data.

Conditioning on the DDF statistic.

In the previous sections, we conditioned on the ancillary statistic $A = |X|$, which has the same distribution for all $\theta \in (-2, 2)$. In our discussion of the Welch example, we showed that the ancillary statistic, R, which has the same distribution $\forall \theta \in \mathbb{R}$, should not be used for tests of two simple hypotheses because it is not a function of the MSS for any binary parameter space (BPS); conditioning upon it breaches the SP and produces nonsensical results. We also showed that, for each pair of hypothesised values $\{\theta_i, \theta_j\}$, there exists a statistic (A_{ij}) that is ancillary on the BPS and can be used to get effective conditional results and we used these statistics to derive a superior conditional confidence interval. In the present example, we have found that conditioning upon $A = |X|$ seems to produce results that are much better than the unconditional Neyman-Pearson results. Nevertheless, we must remember that conditioning upon the range, R, is still regarded as the right approach to take, in the Welch case. Does A, in the gradient example, have the same shortcomings as R in the Welch case? How well does it work for tests of two simple hypotheses; is it a function of the MSS for any BPS; and how does it compare with the exhaustive DDF statistics?

Since, $\forall \theta_1, \theta_2$, $Y = LR(X;\theta_1,\theta_2) = \frac{(1+\theta_1 X)}{(1+\theta_2 X)}$ is a continuous variable, it follows that we can apply the theory of Chapter 9 to this model. For any BPS, $\{\theta_1, \theta_2\}$, we can identify the ancillary DDF statistic, $D(y)$, which is ancillary in the restricted sense, (i.e. it is a function of Y), and partitions the sample space of Y exhaustively. An ancillary statistic defined over a larger parameter space, often lacks one or both of these features, however, in this case, the statistic $A = |X|$ partitions the sample space of X into subsets of two elements (x & $-x$), except that $x = 0$ is in a subset by itself. Since (in this model) y is always a one-to-one function of x, A must partition the sample space of Y similarly. Also $LR(x=0) = 1, \ \forall \theta_1, \theta_2$, thus A is an exhaustive ancillary statistic on every BPS. In order to compare A with the various $D(Y)$ statistics, we derive the details of those variables.

For all θ_1, θ_2 and y,

$$D(y) = F_{\theta_2}(y) - F_{\theta_1}(y),$$

where $F(\cdot)$ is the distribution function of $Y = LR(X;\theta_1,\theta_2)$. Using the relationship between Y and X, and the known distribution function of X, we can derive the following:

$$D(y) = \tfrac{1}{2}|\theta_1 - \theta_2| \cdot \left\{ \tfrac{1}{4} - \frac{(y-1)^2}{(\theta_1 - \theta_2 y)^2} \right\}, \ \forall y, \theta_1, \theta_2.$$

For fixed θ_1 and θ_2, this depends on y only through the expression $\tilde{D}(y) = \frac{(y-1)^2}{(\theta_1 - \theta_2 y)^2}$, which is a one-to-one function of $D(Y)$ and thus an equivalent ancillary statistic.

D and $\tilde{D}$ both partition the sample space of Y into subsets of the form $\{y, \pi(y)\}$ where $\tilde{D}(y) = \tilde{D}(\pi(y))$. It is straightforward to show that the pairing function is:

$$\pi(y) = \frac{\{(\theta_1 + \theta_2)y - 2\theta_1\}}{\{2\theta_2 y - (\theta_1 + \theta_2)\}}.$$

(This is the same for any hypothesis pair such that $\frac{\theta_2}{\theta_1} = q$, since $\pi(y) = \frac{[(1+q)y-2]}{[2qy-(1+q)]}$, thus all scenarios with the same value of $\frac{\theta_2}{\theta_1}$ are in the same E. C. inference class.)

We want to compare the partition produced by $\tilde{D}$ with the equally fine partition produced by the exhaustive ancillary statistic, $A = |X|$. A pairs x with $-x$; what does $\tilde{D}$ pair x with?

Since y is a one-to-one function of x, we can invert this function to find x as a function of y, hence:

$$y = LR(x) = \frac{(1+\theta_1 x)}{(1+\theta_2 x)} \Leftrightarrow x = \frac{(y-1)}{(\theta_1 - \theta_2 y)}.$$

Thus $\tilde{D}(y) = \frac{(y-1)^2}{(\theta_1 - \theta_2 y)^2} = x^2$, but this is also equal to $(-x)^2$, and hence[11] it follows that $\pi(y) = LR(-x)$, and pairing y with $\pi(y)$ is equivalent to pairing x with $-x$.

We have shown that, regardless of the values of θ_1 and θ_2, A is always a one-to-one function[12] of $\tilde{D}(y;\theta_1,\theta_2)$ (and of $D(y;\theta_1,\theta_2)$) and partitions the sample space the same way. Thus A is equivalent to each and every one of the DDF statistics ($\{\theta_1,\theta_2\} \in \mathbb{R}^2$). This is possible because y is a function of θ_1 and θ_2 as is $D(y;\theta_1,\theta_2)$; it so happens that, in this case, the appearance of θ_1 and θ_2 (along with x) in the formula for y, and the appearance of θ_1 and θ_2 in the function $D(y;\theta_1,\theta_2)$, act to exactly cancel each other out so that $D(y;\theta_1,\theta_2)$ is a function of x alone, $\forall \theta_1,\theta_2$. While we have found that it is usually more helpful to structure a test in

[11] This reasoning is dependent on the fact that y is a one-to-one function of x which we know to be the case.
[12] Not the *same* one-to-one function.

terms of the likelihood ratio, y, this is a case where the natural variable describes the same results more simply. (However there has to be serendipity in identifying A as an ancillary statistic, whereas the DDF algorithm produces that ancillary statistic automatically.)

This equivalence of ancillary statistics makes another feature of $A = |X|$ evident; we noted earlier that A is a function of the MSS for $\theta \in (-2,2)$, which is X; however, it is now clear that, for any binary parameter space $\{\theta_1, \theta_2\} \subset (-2,2)^2$, A is also a function of the MSS on that binary parameter space (i.e. Y). Thus $A = |X|$ is an ancillary statistic, in the restricted sense, on all possible binary parameter spaces, as well as on the natural parameter space, and is equivalent to the statistic $D(y;\theta_1,\theta_2)$, *for all* θ_1 and θ_2.

Since we can use A in place of the DDF statistic and A is function of X alone and the conditional distribution of $X \mid A = a$ (under H) depends only on x and θ_1, we can formulate the conditional p-value (which we can now call the *cp-value* since it is based on the DDF statistic) for the left $(\theta_2 < \theta_1)$ and right-sided $(\theta_2 > \theta_1)$ cases without specifying the exact value of θ_2. Thus:

$$\text{Right-sided cp-value}(x) = \begin{cases} \frac{(1+\theta_1 x)}{2}, & x > 0 \\ 100\%, & x \le 0. \end{cases}$$

$$\text{Left-sided cp-value}(x) = \begin{cases} 100\%, & x \ge 0 \\ \frac{(1+\theta_1 x)}{2}, & x < 0. \end{cases}$$

Although these look similar to typical p-values, they are substantially different because (like all cp-values based on DDF statistics) they assign to all data with a likelihood ratio of more than *one*, a cp-value of 100%. The cp-value's insensitivity to the exact value of θ_2 reflects the fact that, when we change θ_2 to (say) $\theta_2 + \varepsilon$, then (as long as y is still on the same side of *one*) the change in y (from a function of (x,θ_1,θ_2) to the same function of $(x,\theta_1,\theta_2+\varepsilon)$) is cancelled out by the change in the cp-value (from a function of (y,θ_1,θ_2) to the same function of $(y,\theta_1,\theta_2+\varepsilon)$). Any method that breaches the likelihood principle, as ECI does, may be insensitive to the exact value of K because insensitive to the exact value of the LR. For given data and null hypothesis, changing K will generally change the likelihood ratio of the data but, in many cases (such as the present one), it also moves us from one inference class to another. Because this changes $cp(y)$ as a function of y, it is possible for the cp-value to remain the same even though y has changed. The very nice results that we found for the Normal location model arise because all the possible tests are in the same inference class and, hence, changing K and y inevitably changes $cp(y)$.

Conditional and unconditional p-values.

In log-symmetric models, the conventional p-value is always less than, and therefore more significant than, the cp-value; this is a corollary of the fact that $cp(y)$ is an increasing function of $y < 1$. By contrast, we find that the gradient model, applied to certain binary parameter spaces, presents us with cases where $cp(y)$ is a decreasing function of y $(y < 1)$ and hence $cp(y) < p(y)$ for some y; this does not, however, result in instances where the conditional inference produces a significant result and the unconditional inference does not.

The general formula for the cp-value obtained by conditioning on the observed value of $D(Y)$ is:

$$cp(y) = \frac{y(\pi(y)-1)}{(\pi(y)-y)},\ y < 1$$

where y is the observed likelihood ratio and $\pi(\cdot)$ is the pairing function defined by $D(\pi(y)) = D(y)$. Thus, for the gradient model:

$$cp(y) = \frac{(\theta_1-\theta_2)y}{2(\theta_1-\theta_2 y)},\ y < 1.$$

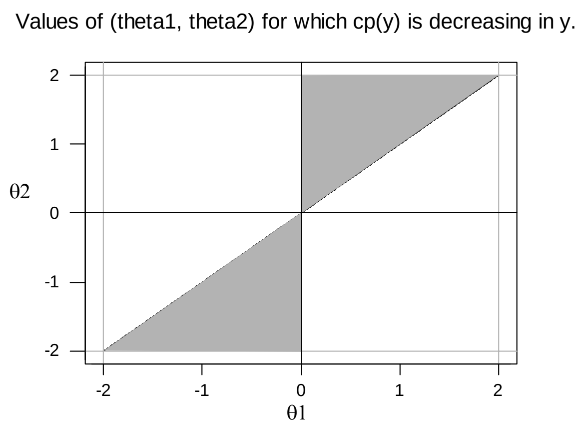
It follows that $\frac{d}{dy}cp(y) < 0$ if and only if $\theta_1(\theta_1-\theta_2) < 0$, and is otherwise positive. Note also that $\frac{d}{dy}cp(y) = 0$, and the cp-value is constant, if and only if $\theta_1 = 0$; in that case $\pi(y) = \frac{y}{2y-1}$ and $cp(y) = 50\%$ for $y < 1$ exactly as in the Exponential case with

$q = \frac{2}{3}$. It is not hard to show that this $\pi(y)$ is the only pairing function that can produce a constant cp-value.

$p(y)$ can only be greater than $cp(y)$ if the latter is a decreasing function of y on some interval in $(0,1)$. For the Gradient model, the function $cp(y)$ has no turning points in the interval $(0,1)$, thus, if $cp(y)$ is decreasing anywhere in $(0,1)$, it is decreasing everywhere in $(0,1)$. Since $cp(y) \to 50\%$ as $y \to 1$ (see §9.5), it follows that $cp(y) > 50\%$ wherever $p(y) > cp(y)$, thus neither type of p-value is significant.

The following plot highlights those points in the (θ_1, θ_2)-plane where $cp(y)$ is a decreasing function of $y < 1$.

Figure 10.33



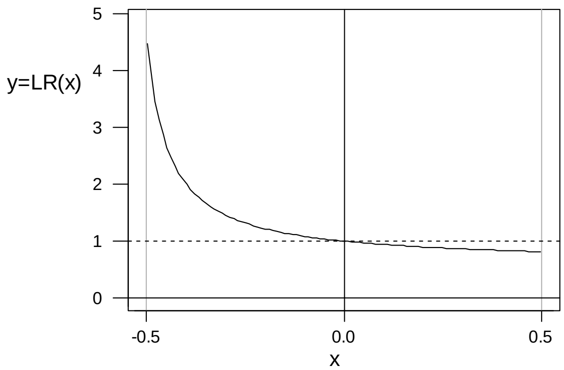
We will look at a particular instance where the cp-value decreases as the likelihood ratio increases.

Example 10.10

Consider testing H: $\theta = 1.1$ versus K: $\theta = 1.8$ so that (θ_1, θ_2) lies in the highlighted region of the above plot.

The following plot shows the relationship between x and $y = LR(x)$, for this test.

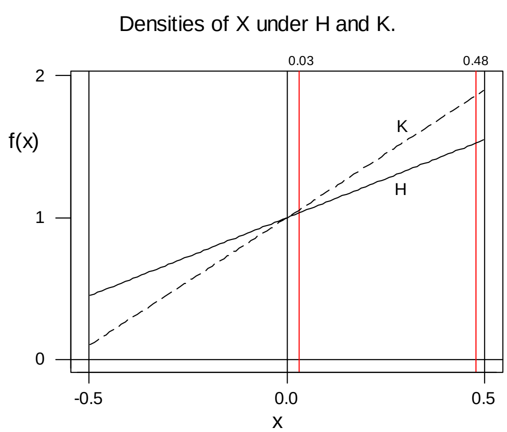
Figure 10.34



We note that y is a decreasing function of x (so this is a right-sided test, in terms of x) and that y is less than *one*, providing some evidence against H, only when $x > 0$. Also the likelihood ratio is always greater than $0.816 = \frac{1}{1.225}$ – it is not *much* less than *one* for any value of x. This shows (as does the plot below) that the two distributions are very similar to each other where $x > 0$.

Consider the two observations $x_1 = 0.03$ and $x_2 = 0.48$; these values are at opposite extremes among the positive values of x.

Figure 10.35



The two values are marked on the above plot showing the densities of X under the two hypotheses. Since this is a right-sided test, the conventional p-value of each point is the area to the right of the point under the null density. Although $x = 0.48$ is very large in terms of the possible values of x, it does not appear to provide strong evidence against H relative to K. The likelihood ratios, p-values and cp-values of these observations are shown below.

Table 10.13

x	$y = LR(x)$	$p(y)$	$cp(y)$
0.03	$0.98 = \frac{1}{1.02}$	60.70%	51.65%
0.48	$0.82 = \frac{1}{1.22}$	3.08%	76.40%

Both of the likelihood ratios are very close to *one*, indicating that the evidence against H, from both observations, is extremely weak. Counter-intuitively, the cp-value is higher when the likelihood ratio is lower, but, despite this anomaly, it is clear that the cp-values give a much more accurate idea of the significance of the data than the p-values. In particular, the observation $x = 0.48$, which has a LR only slightly less than *one*, has a significant p-value of 3.08% whereas the cp-value is 76.40% indicating insignificant evidence against H.

To find out what happens when $n > 1$, we simulated $N = 2500$ sets of data, with $n = 20$ observations each, under both H ($\theta = 1.1$) and K ($\theta = 1.8$). From this we were able to obtain likelihood ratio values and, hence, the empirical distribution functions of Y under H and K, and the empirical DDF statistic. The results indicated that, when $n = 20$, the $cp(y)$ function is increasing in y, thus, it seems that (as in the Exponential case) the anomalous relationship between y and $cp(y)$ disappears when the sample is larger.

Conditional confidence intervals.

Since we know how to use the data, x, to test all possible pairs of hypotheses, we can derive a *conditional confidence interval* for θ from x. The $\bar{\alpha}$-level conditional confidence interval (CCI) contains all and only those values of θ that would not be rejected (as the null value) in favour of any alternative in $(-2,2)$ at a conditional significance level less than or equal to $\bar{\alpha}$. Since $A=|X|$ is the exhaustive ancillary statistic, equivalent to the DDF statistic, for each and every binary parameter space, we can derive the conditional confidence intervals in terms of x and a rather than in terms of y and $\pi(y)$.

The result of any test of H: $\theta=\theta_1$ can be found from the cp-value, below.

$$\text{Right-sided cp-value}(x)=\begin{cases}\frac{(1+\theta_1 x)}{2}, & x>0\\ 1, & x\le 0.\end{cases}$$
$$\text{Left-sided cp-value}(x)=\begin{cases}1, & x\ge 0\\ \frac{(1+\theta_1 x)}{2}, & x<0.\end{cases}$$

The value of $a=|x|$ does not depend on the hypotheses in question. Note that $(\forall\theta_i,\theta_j)$ cp-value$(x)>\frac{1}{2}-a$ for all x, since $-2<\theta_1<2$. It follows that $\bar{\alpha}$ [13] must either be *zero* or be greater than $\frac{1}{2}-a$, since for no test does any value of x yield a cp-value in the range $(0,\frac{1}{2}-a]$. When $\bar{\alpha}=0$, *zero* is the common conditional significance level of *all* the tests (i.e. for all $(\theta_i,\theta_j)\in(-2,2)^2$) and thus the CCI is based on tests that all have the same (conditional) significance level, just as conventional confidence intervals are usually based on tests that all have the same unconditional level. Note that most of the values of $a\in[0,\frac{1}{2})$ require us to use $\bar{\alpha}=0$ since $\frac{1}{2}-a$ is unreasonably large for a significance level: $\frac{1}{2}-a<5\%$ only when $a>0.45$.

[13] $\bar{\alpha}=\max_{(\theta_i,\theta_j)\in(-2,2)^2}\alpha_a(\theta_i,\theta_j)$.

From the above formula for cp-value, we can easily derive the $\bar{\alpha}$ -level conditional confidence interval for θ as follows.

$$\text{For any } \bar{\alpha} \in (\tfrac{1}{2}-a,1), \quad CCI(x) = \begin{cases} (\frac{-(1-2\bar{\alpha})}{a},\ 2), & x>0 \\ (-2,\ \frac{(1-2\bar{\alpha})}{a}), & x<0. \end{cases}$$

$$\text{For } \bar{\alpha}=0, \quad CCI(x) = (-2,2), \text{ for all } x.$$

When $x=0$, cp-value$(x)=100\%$ $(\forall\theta_i \,\&\, \theta_j)$ and hence $\bar{\alpha}<1 \Rightarrow \bar{\alpha}=0$ and the CCI is the whole parameter space. When we use $\bar{\alpha}=0$, we are entitled to describe $(-2,2)$ as the 100% CCI for θ, since *zero* is the common (conditional) significance level of all the tests on which the CCI is based and 100% is clearly the coverage of the interval. No other achievable value of $\bar{\alpha}$ produces this interval, whereas the conventional approach produces $(-2,2)$ as the $100(1-\alpha)\%$ CI (based on a given x) for a wide range of α -values (not just $\alpha=0$).

Only values of a that are close to $\frac{1}{2}$ allow us to exclude any part of the parameter space from the CCI with small $\bar{\alpha}$. This is consistent with the fact that non-extreme x are reasonably consistent with all the θ values. Thus if $a=0.3$ ($x=-0.3$ or 0.3), $\frac{1}{2}-a=20\%$ and we cannot use any $\bar{\alpha}<20\%$ except $\bar{\alpha}=0$ which gives the 100% conditional confidence interval $(-2,2)$.

If $a=0.46$ ($x=-0.46$ or 0.46) then $\frac{1}{2}-a=4\%$ and we may use (for instance) $\bar{\alpha}=5\%$ to obtain the following non-trivial CCI:

$$\begin{cases} (-1.9565, 2.00), & \text{if } x=0.46 \\ (-2.00, 1.9565), & \text{if } x=-0.46. \end{cases}$$

More generally, for any $a>0.45$, there is a valid CCI at the 5% level given by:

$$(\tfrac{-0.9}{a}, 2), \quad x=a$$
$$(-2, \tfrac{0.9}{a}), \quad x=-a.$$

The width of this interval is $2+\frac{0.9}{a}$ which tends to a minimum of 3.8 as $a \to \frac{1}{2}$. Thus even the most extreme data can only exclude a small range of θ-values with any confidence. This is consistent with the nature of the model.

Summary.

For the Gradient model, the DDF statistic for any binary parameter space, in the natural space $(-2, 2)$, is equivalent to the universal ancillary statistic $A = |X|$, which can thus be used to produce conditional tests and confidence intervals. The conventional 'optimal' approach produces confidence intervals and test results that are intuitively unsatisfactory. We can get a formal account of the flaws in the conventional approach by considering the conditional distributions of $X \mid A = a$, in preference to the unconditional distribution of X, (for example, the conditional coverage of a 'long' 50% confidence interval is greater than 50%, $\forall \theta$). The inferences obtained by conditioning on A are more consistent with our understanding of the model, which operates under considerable restrictions not reflected in the conventional results. This model produces scenarios that show the superiority of the conditional approach much more clearly than those produced by Welch's Uniform example.

10.6 Tests on general Normal hypotheses.

Let $X \sim N(\mu, \sigma^2)$ and consider two hypothesised values of the two-dimensional parameter of interest $\theta = (\mu, \sigma^2)^{\mathrm{T}}$, i.e. H: $\theta = \theta_1 = (\mu_1, \sigma_1^2)^{\mathrm{T}}$ and K: $\theta = \theta_2 = (\mu_2, \sigma_2^2)^{\mathrm{T}}$. Each hypothesis specifies both the mean and variance of X. If $\sigma_1^2 = \sigma_2^2$, we can view σ^2 as known theoretically and use the log-symmetric structure, discussed in Chapter 8, to test hypotheses about the mean. When the variances differ, $\ln y$ is a quadratic function[14] of x and this can be used to derive the distribution functions of Y.

[14] $\ln y = ax^2 + bx + c$, where: $a = \frac{1}{2}(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2})$, $b = (\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2})$ and $c = \ln(\frac{\sigma_2}{\sigma_1}) - \frac{1}{2}(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2})$.

Let

$$
\begin{aligned}
\gamma_1 &= \tfrac{\mu_1\sigma_2^2-\mu_2\sigma_1^2}{\sigma_1^2-\sigma_2^2}\\
\gamma_2 &= \tfrac{\mu_1-\mu_2}{\sigma_1^2-\sigma_2^2}\\
\gamma_3 &= \tfrac{2}{\sigma_1^2-\sigma_2^2}\\
\gamma_4 &= \tfrac{\sigma_1}{\sigma_2}\\
m &= \ln(\tfrac{\sigma_2}{\sigma_1}) - \tfrac{(\mu_1-\mu_2)^2}{2(\sigma_1^2-\sigma_2^2)} = -\ln\gamma_4 - \gamma_2^2/\gamma_3\\
\gamma_{5,i} &= -\tfrac{1}{\sigma_i}\{\mu_i+\gamma_1\},\quad i\in\{1,2\}.
\end{aligned}
$$

Note that $\sigma_1 < \sigma_2 \Rightarrow m > 0$ and $y \in (0, e^m)$, while $\sigma_1 > \sigma_2 \Rightarrow m < 0$ and $y \in (e^m, \infty)$.

Let $h_i(y) = \sigma_j\sqrt{\gamma_2^2 + \gamma_3\ln(\gamma_4 \cdot y)}$ where $i \in \{1,2\}$ and $j = \begin{cases} 2, & i=1 \\ 1, & i=2 \end{cases}$.

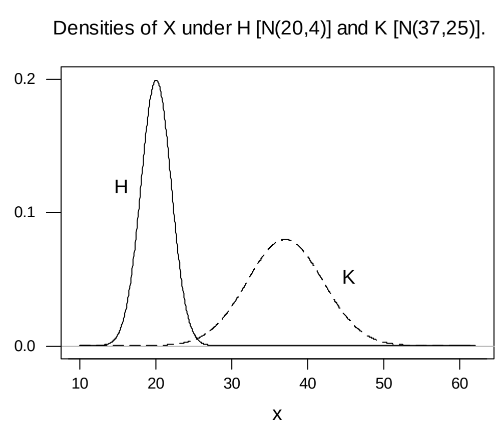
Then

$$
\boxed{
\begin{aligned}
D(y) &= F_K(y) - F_H(y)\\
&= |\,\Phi[\gamma_{5,2}+h_2(y)] - \Phi[\gamma_{5,2}-h_2(y)] - \Phi[\gamma_{5,1}+h_1(y)] + \Phi[\gamma_{5,1}-h_1(y)]\,|.
\end{aligned}}
$$

Example 10.11.

Given that $X \sim N(\mu,\sigma^2)$, the null hypothesis (H) states that $(\mu,\sigma^2)^T = (20,4)$ and the alternative hypothesis (K) that $(\mu,\sigma^2)^T = (37,25)$.

Figure 10.36

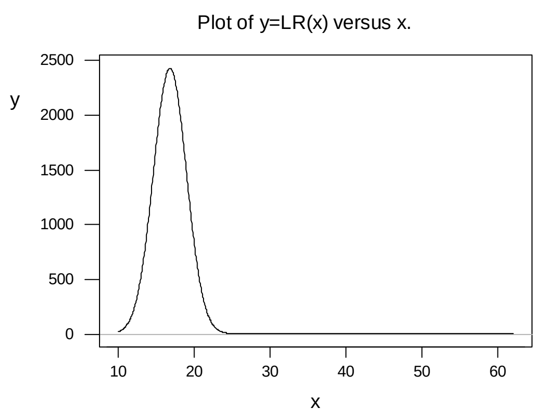


When the variances are different, $y = LR(x)$ is no longer a one-to-one function of x, and the conventional p-value is (in terms of y):

$$p(y) = P_H(Y < y) = P_H(X < x_1(y)) + P_H(X > x_2(y)),$$

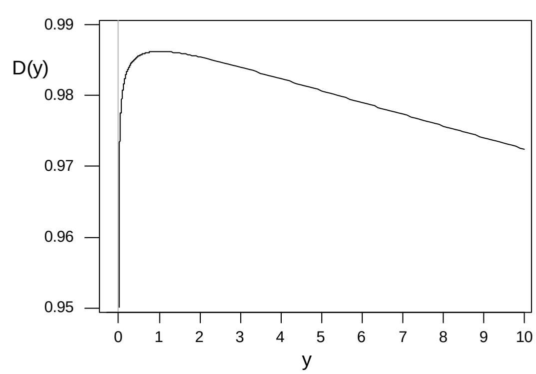
where $x_1(y)$ and $x_2(y)$ are the roots of the equation $y = LR(x)$ (shown below).

Figure 10.37



When $\mu_1 = 20,\ \sigma_1 = 2,\ \mu_2 = 37,\ \sigma_2 = 5$, $y \in (0, 2433.88)$. The DDF statistic is shown below for values of y in $(10^{-9}, 10)$.

Figure 10.38



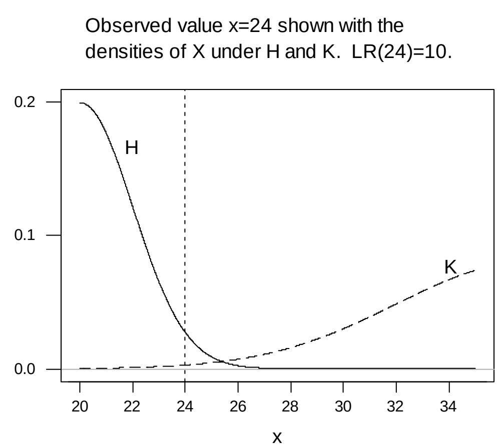
For given values of y we can find the corresponding $\pi(y)$ to any given level of accuracy, and hence the cp-values. For six values of y we have derived $p(y)$ and $cp(y)$.

Table 10.14

$y = LR(x)$	$0.1 = \frac{1}{10}$	$0.3 = \frac{1}{3.33}$	$0.5 = \frac{1}{2}$	$0.8 = \frac{1}{1.25}$	$1.2 = \frac{1}{0.83}$	$9.94 = \frac{1}{0.10}$
$x_1(y)$	6.95	7.52	7.77	8.02	8.25	9.52
$x_2(y)$	26.55	26.00	25.75	25.50	25.28	24.00
$\pi(y)$	4.9	2.6	1.8	1.2	*	*
$\boxed{p(y)}$	0.05%	0.15%	0.20%	0.30%	0.41%	2.28%
$\boxed{cp(y)}$	8.13%	20.87%	30.77%	40.00%	100%	100%

As on some former occasions, we note that the conventional approach produces highly significant results even when the likelihood ratio is greater than *one* and the data has a higher likelihood under H than under K. Looking in more detail at the data represented by the last column of the table, we see that, when we observe $x = 24$, the p-value is 2.28% but the data is surely more consistent with H than with K (see below).

Figure 10.39



10.7 Conjectures on the asymptotic properties of exhaustive conditional inference.

Consider a random sample[15] $\underset{\sim}{X} = X_1, \ldots, X_n$ providing evidence about the sole unknown parameter, θ. In some cases, such as the Exponential, we have seen that, when n is large, the cp-functions, $cp(\cdot)$, produced for different scenarios are very similar to each other, although they differ dramatically when $n = 1$. This raises the possibility that there may exist a reasonably general ECI in the limit as $n \to \infty$. The fact that the range of $Y = LR(\underset{\sim}{X})$ approaches $\mathbb{R}^+$, in all cases, as $n \to \infty$, lends some support to this conjecture; it is only in cases where the range is severely restricted that we have seen the ECI in conflict with the LR, and with the ECI for other scenarios.

If convergence of the cp-function to the same limit occurs across a large number of models, then there exists a comprehensive *asymptotic* EC inference-class, and it follows that exhaustive conditional inference is (to a large degree) asymptotically consistent with likelihood theory. Examining the asymptotic behaviour of ECI in any depth is beyond the scope of this work; in this section we consider briefly the possible implications of the asymptotic Normality of many maximum likelihood estimators.

Under reasonably common regularity conditions[16], θ has a unique maximum likelihood estimator, $\hat{\theta}$, that is asymptotically Normal with a mean of θ and variance equal to the minimum variance (Cramér-Rao) bound. That is, as $n \to \infty$, the distribution of $\hat{\theta} \to N(\theta, \frac{d(\theta)}{n})$, where $d(\theta) = -[E_\theta\{\frac{\partial^2}{\partial\theta^2} L(X_i;\theta)\}]^{-1}$ and $L(x;\theta)$ is the common likelihood function of the X variables. Thus, in the limit, any hypothesis of the form $\theta = \theta'$ entails $\hat{\theta} \sim N(\theta', \frac{d(\theta')}{n})$, and any two simple hypotheses specifying θ give rise to the type of scenario discussed in the previous section, where each hypothesis defines a different mean and variance for a Normal variable, i.e. $\hat{\theta}$ plays the role of X in §10.6. $\hat{\theta}$ is asymptotically sufficient and hence $LR(\hat{\theta}) \to Y (= LR(\underset{\sim}{X}))$ as $n \to \infty$. If n is large and the regularity conditions are met,

[15] Independent and identically distributed variables,
[16] See Kendall & Stuart, pp. 39-43.

it is reasonable to use the asymptotic distribution of $\hat{\theta}$ to perform an ECI, in the manner illustrated in the previous section.

Consider the model $X \sim N(\mu,\sigma^2)$ with two simple hypotheses, as in §10.6, but now suppose that $\sigma_i^2 = \frac{d_i}{n}$. For each value of n, the scenario is completely defined giving rise to a pairing function and hence a cp-function. We have not derived the form of the pairing function for this type of scenario – the particular values of $\pi(y)$ used to produce the cp-values in ***Example 10.11*** were derived numerically – nevertheless a pairing function exists. If we let $n \to \infty$ but hold $\mu_1,\ \mu_2,\ d_1$ and d_2 constant, what happens to the pairing function? Does it converge and, if so, does the limit depend on any of $\mu_1,\ \mu_2,\ d_1$ and d_2, or is there a general limit that applies regardless of the values taken by these variables? If it converges, then the limit of the pairing function also applies to any case where the maximum likelihood estimator of θ has the necessary asymptotic characteristics. For example, if the limit is independent of $\mu_1,\ \mu_2,\ d_1$ and d_2, it will apply to any case where the MLE of θ is asymptotically Normal; if the limit depends on (say) d_1 and d_2, then it will apply to any case where the MLE is asymptotically Normal and $d(\theta_1) = d_1$ and $d(\theta_2) = d_2$. The former possibility is the more interesting since it would imply that, whenever n is large enough, the ECI on any θ where $\hat{\theta}$ is asymptotically Normal is the same and, thus, there exists a huge asymptotic E. C. inference class.

The Normal location case, $X_i \sim N(\mu,\sigma^2)$ (σ^2 known), is a particular case of the above type, where the limiting properties are completely known since $\hat{\theta} = \overline{X} \sim N(\mu,\frac{\sigma^2}{n})$. The sufficiency is exact for all n, and the model is log-symmetric for all n. Thus $\pi_n(y)$ is the same for all n and trivially constitutes the limit, i.e. $\pi_L(y) = y^{-1}$. Thus, we know that whenever $d_1 = d_2$, the pairing functions are independent of the values $\mu_1,\ \mu_2$ and d (for all n as well as in the limit). If $\pi_n(y)$ converges for the more general Normal case, and the limit to which it converges is independent of $\mu_1,\ \mu_2,\ d_1$ and d_2, then this unique limit must be y^{-1}, since it is applicable to the Normal location case. We consider two examples of the general Normal model to see whether these conjectures have any plausibility.

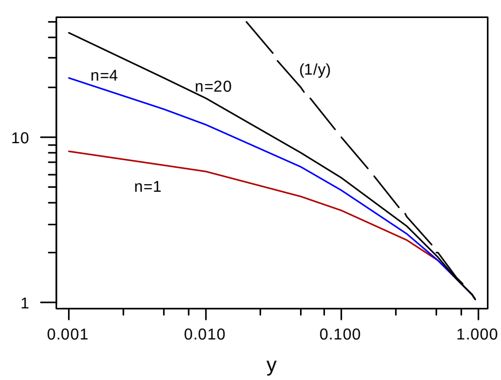
Example 10.12

Consider the particular cases where $X \sim N(\mu, \sigma^2)$ and $\mu_1 = 20$, $\mu_2 = 37$, $d_1 = 16$, and $d_2 = 100$ (thus $\sigma_1^2 = \frac{16}{n}$ and $\sigma_2^2 = \frac{100}{n}$). In ***Example 10.11***, we performed an ECI for this case when $n = 4$. We want to see what happens to the pairing function as n increases.

As $n \to \infty$, $D(y) \to 1$, for all finite y; this makes it difficult to find $\pi_n(y)$ numerically, for large n, since we need to evaluate $D(y)$ to a high level of accuracy in order to distinguish between $D(y_1)$ and $D(y_2)$ when both are very close to *one*. For this reason, we have derived the pairing functions only for the cases $n = 1$, $n = 4$ and $n = 20$. The following plot shows the pairing functions for these three cases compared with the log-symmetric pairing function $\pi(y) = y^{-1}$.

Figure 10.40

Plot of the pairing functions for n=1, 4, and 20 and the log-symmetric pairing function, 1/y.

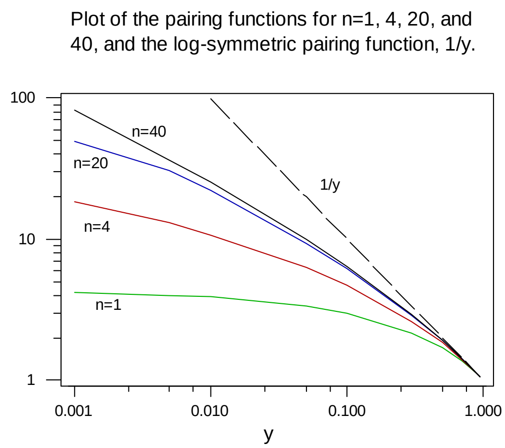


It seems likely that the pairing function converges and y^{-1} is a possible limit – more than this we cannot say.

Example 10.13

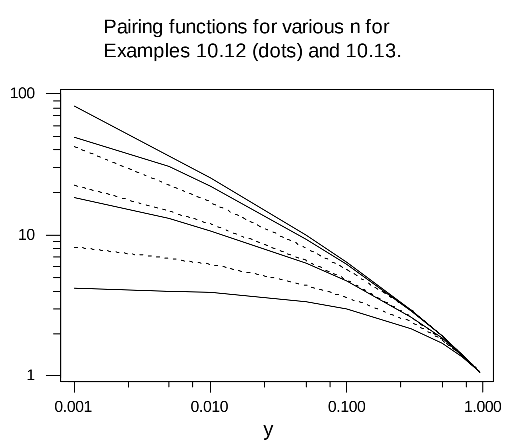
Now, consider the cases where $X \sim N(\mu, \sigma^2)$ and $\mu_1 = 50$, $\mu_2 = 40$, $d_1 = 25$, and $d_2 = 81$. (No specific examples can establish a general result, but we have attempted to pick two examples with little in common. Thus this example has $\mu_1 > \mu_2$ and the smaller variance associated with the larger mean, in contrast to the previous case.)

Figure 10.41



As in the previous case, convergence seems likely and the log-symmetric pairing function is a possible limit. The following plot shows the two sets of pairing functions together. There is nothing in this plot to indicate that the two examples produce *different* limits.

Figure 10.42



On the basis of this very limited investigation, we may conjecture that the pairing functions for the general Normal model converge to the same limit in all cases. If this were shown to be true, it would follow that the limit is y^{-1} and $cp(y) \to \frac{y}{(1+y)}$ as $n \to \infty$ $(\forall y < 1)$. It would also follow that this asymptotic ECI would be applicable to a wide range of non-Normal data whenever n is sufficiently large. Intuitively we would expect our inferences to improve with increased sample sizes and the above relationship between the cp-function and the likelihood ratio certainly ensures that the cp-values make sense from a likelihood point of view. By comparison, we have frequently noted that $p(y) \to 0$ as $n \to \infty$, for all finite y. This result is consistent across models but is patently ridiculous since it means that any likelihood ratio, no matter how large, is interpreted as significant evidence against H relative to K once the sample size is large enough.