Chapter 17
Pure Likelihood Methods

17.1 Introduction

As we have seen in previous chapters use of the likelihood is important in frequentist methods and in Bayesian methods. In this chapter we explore the use of the likelihood function in another context, that of providing a self-contained method of statistical inference. Richard Royall in his book, Statistical Evidence: A Likelihood Paradigm, carefully developed the foundation for this method building on the work of Ian Hacking and Anthony Edwards. Royall lists three questions of interest to statisticians and scientists after having observed some data

1. What do I do?
2. What do I believe?
3. What evidence do I now have?

   In the context of the usual parametric statistical model where we have an observed value $\boldsymbol{x}_{obs}$ of random $\boldsymbol{X}$ having sample space $\mathcal{X}$, parameter space Θ, and probability density function $f(\boldsymbol{x}_{obs}; \theta)$ at the observed value, $\boldsymbol{x}_{obs}$ of $\boldsymbol{X}$ the first question is a decision theoretic problem re the actions to be taken on the basis of the model and the observed data and the second concerns what do I believe about θ given the observed data and presumably some prior knowledge about θ. The third question concerns characterizing what evidence the data has provided us about θ and requires no actions or beliefs. It is simply a question of "what do the data say" (about θ).

   We have already stated the Law of Likelihood:

Axiom 17.1.1. (*Law of Likelihood*). *For two parameter values, θ_1 and θ_0, in the model $\mathcal{X}, f(\boldsymbol{x}; \theta), \Theta)$, the magnitude of the likelihood ratio*

$$\boldsymbol{L}(\theta_1, \theta_0; \boldsymbol{x}_{obs}) = \frac{f(\boldsymbol{x}_{obs}; \theta_1)}{f(\boldsymbol{x}_{obs}; \theta_0)}$$

measures the statistical evidence for θ_1 vs θ_0. If the ratio is greater than 1 we have statistical evidence for θ_1 vs θ_0 while if less than 1 we have statistical evidence for θ_0 vs θ_1.

I have used the term statistical evidence so as to not conflict with the use of the word evidence in other contexts, e.g., in P-values. We say the statistical evidence for θ_1 vs θ_0 is of strength $k > 1$ if $\boldsymbol{L}(\theta_1, \theta_0; \boldsymbol{x}_{obs}) > k$.

17.2   Misleading Statistical Evidence

Since we are dealing with probability models it is possible to observe a value, $\boldsymbol{x}_{obs}$, for which $\boldsymbol{L}(\theta_1, \theta_0; \boldsymbol{x}_{obs}) > k$, and yet θ_0 is true. This is called **misleading evidence**. The following is called the **universal bound** and shows that the probability of misleading evidence can be kept small by choice of k.

Theorem 17.2.1. *The probability of misleading evidence is bounded by $1/k$, i.e.,*

$$\mathbb{P}_{\theta_0} \left\{ \frac{f(\boldsymbol{X}; \theta_1)}{f(\boldsymbol{X}; \theta_0)} \geq k \right\} \leq \frac{1}{k}$$

Proof. Let M be the set

$$M = \left\{ \boldsymbol{x} \ : \ \frac{f(\boldsymbol{X}; \theta_1)}{f(\boldsymbol{X}; \theta_0)} \geq k \right\}$$

Then

$$\int_M f(\boldsymbol{x}; \theta_0) d\mu(\boldsymbol{x}) \leq \int_M \frac{1}{k} f(\boldsymbol{x}; \theta_1) d\mu(\boldsymbol{x})$$

$$\leq \frac{1}{k} \int_{\mathcal{X}} f(\boldsymbol{x}; \theta_1) d\mu(\boldsymbol{x})$$

$$= \frac{1}{k}$$

In fact a much stronger result is true. Consider a sequence of observations

$$\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$$

such that if A is true then $\mathbf{X}_n \sim f_n$ and when B is true $\mathbf{X}_n \sim g_n$. The likelihood ratio

$$\frac{g_n(\mathbf{x}_n)}{f_n(\mathbf{x}_n)} = z_n$$

is the LR in favor of B after n observations. Then we have the following theorem.

Theorem 17.2.2. *If A is true then*

$$P_A(Z_n \geq k \text{ for some } n = 1, 2, \ldots) \leq \frac{1}{k}$$

Robbins [41]

In many circumstances the universal bound is far too conservative. Consider the situation where we have $X_1, X_2, \ldots, X_n$ where the X_i are iid as $N(\mu, \sigma^2)$ where, for simplicity, σ^2 is assumed known. The joint density is given by

$$f(\mathbf{y}; \mu) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{(y_i - \mu)^2}{2\sigma^2} \right\}$$

After some algebraic simplification the likelihood ratio for comparing μ_1 vs μ_0 is given by

$$\exp\left\{ \left(\bar{x} - \frac{\mu_0 + \mu_1}{2} \right) \frac{n(\mu_1 - \mu_0)}{\sigma^2} \right\}$$

It follows that the likelihood ratio exceeds k if and only if

$$\frac{n(\mu_1 - \mu_0)}{\sigma^2} \left(\bar{x} - \frac{\mu_1 + \mu_0}{2} \right) \geq \ln(k)$$

Thus, without loss of generality, if $\mu_1 - \mu_0 > 0$, the likelihood ratio exceeds k if and only if

$$\bar{x} \geq \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{n(\mu_1 - \mu_0)} \ln(k)$$

Thus the probability of misleading statistical evidence when $H_0 = \mu_0$ is assumed true is given by

$$\begin{aligned}
\text{PMLEV}_1 &= P_{\mu=\mu_0} \left\{ \frac{f(\mathbf{X}; \mu_1)}{f(\mathbf{X}; \mu_0)} \geq k \right\} \\
&= P_{\mu=\mu_0} \left\{ \overline{X} \geq \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{n\mu_1 - \mu_0} \ln(k) \right\} \\
&= P_{\mu=\mu_0} \left\{ \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \geq \frac{\sqrt{n}(\mu_1 - \mu_0)}{2\sigma} + \frac{\sigma \ln(k)}{\sqrt{n}(\mu_1 - \mu_0)} \right\} \\
&= P\left(Z \geq \frac{\sqrt{n}c}{2} + \frac{\ln(k_2)}{\sqrt{n}c} \right)
\end{aligned}$$

$$= \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)$$

where $\Phi(z)$ is the standard normal distribution function evaluated at z and

$$c = \frac{|\mu_1 - \mu_0|}{\sigma}$$

If $\mu_1 - \mu_0 < 0$ similar calculations show that the probability of misleading evidence when μ_0 is assumed true is given by the same expression. It follows that the probability of misleading evidence when $H_0 = \mu_0$ is true is

$$\text{PMLEV} = \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k_2)}{c\sqrt{n}}\right)$$

where

$$c = \frac{|\mu_2 - \mu_1|}{\sigma}$$

and Φ is the standard normal distribution function. The function

$$B(c, k, n) = \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)$$

has been called the **bump function** by Royall.

Also note that c is often called the **effect size** in the social science literature and represents the difference between μ_0 and μ_1 in standard deviation units. The following are rules of thumb for judging the magnitude of the effect size:

- $c \leq 0.1$ trivial
- $0.1 < c \leq 0.6$ small
- $0.6 < c \leq 1.2$ moderate
- $c \geq 1.2$ large

Note that the derivative with respect to c of the bump function is

$$\phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)\left(-\frac{\sqrt{n}}{2} + \frac{\ln(k)}{c^2\sqrt{n}}\right)$$

which vanishes when

$$\frac{\sqrt{n}}{2} = \frac{\ln(k)}{c^2\sqrt{n}}$$

i.e., when

$$c = \sqrt{\frac{2\ln(k)}{n}} = c^*$$

The second derivative with respect to c is

$$\phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)\left(-\frac{\sqrt{n}}{2} + \frac{\ln(k)}{c^2\sqrt{n}}\right)^2 \phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)$$
$$+ \left(-\frac{2\ln(k)}{c^3\sqrt{n}}\right)$$

which is negative when $c = c^*$ so that the bump function has a maximum at $c = c^*$ given by

$$B(c^*, k, n) = \Phi\left(-\frac{c^*\sqrt{n}}{2} - \frac{\ln(k)}{c^*\sqrt{n}}\right) = \Phi(-\sqrt{2\ln(k)})$$

It is well known that

$$\frac{t}{1+t^2}\phi(t) \le \Phi(-t) \le \frac{1}{t}\phi(t)$$

so that

$$\Phi(-\sqrt{2\ln(k)}) \le \frac{1}{\sqrt{2\ln(k)}}\phi(\sqrt{2\ln(k)})$$

$$= \frac{1}{2\sqrt{\pi\ln(k)}}\exp\left\{-(\sqrt{2\ln(k)})^2/2\right\}$$

$$= \frac{1}{2\sqrt{\pi\ln(k)}}\exp\left\{-\ln(k)\right\}$$

$$= \frac{1}{k2\sqrt{\pi\ln(k)}}$$

which is considerably less than the universal bound of $1/k$.

17.2.1   *Weak Statistical Evidence*

Again, since we are dealing with probability models, it is possible to observe a value, x_{obs}, for which

$$\frac{1}{k} < L(\theta_1, \theta_0; x_{obs}) < k$$

This is called **weak statistical evidence**. We have weak evidence in the example of normally distributed observations if and only if

$$-\ln(k) \le \left(\bar{x} - \frac{\mu_1 + \mu_0}{2}\right)\frac{n(\mu_1 - \mu_0)}{\sigma^2} \le \ln(k)$$

If $\mu_1 - \mu_0 > 0$ the condition for weak statistical evidence is that $\bar{x}$ must lie between

$$\frac{\mu_1 + \mu_0}{2} - \frac{\sigma^2}{n(\mu_1 - \mu_0)}\ \ln(k)$$

and

$$\frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{n(\mu_1 - \mu_0)}\ \ln(k)$$

If we define

$$g_n(\mu_0, \mu_1, k) = \frac{\mu_1 + \mu_0}{2} + \frac{\sigma^2}{n(\mu_1 - \mu_0)}\ \ln(k)$$

then the condition for weak evidence becomes

$$g_n(\mu_0, \mu_1, 1/k) \le \bar{x} \le g_n(\mu_0, \mu_1, k)$$

and the probability of weak evidence is given by

$$\Pr\left\{g_n(\mu_0, \mu_1, 1/k) \le \bar{X} \le g_n(\mu_0, \mu_1, k)\right\}$$

which is easily evaluated under H_0 and H_1 since $\bar{X}$ has an $N\left(\mu, \frac{\sigma^2}{n}\right)$ distribution.

Now we note that

$$\frac{\sqrt{n}}{\sigma}[g_n(\mu_0, \mu_1, k) - \mu_0] = \frac{c\sqrt{n}}{2} + \frac{\ln(k)}{c\sqrt{n}}$$

It follows that the probability of weak evidence, $P_{\mu=\mu_0}(\text{WEV})$, is given by

$$\Phi\left(\frac{c\sqrt{n}}{2} + \frac{\ln(k)}{c\sqrt{n}}\right) - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)$$

Similarly

$$\frac{\sqrt{n}}{\sigma}[g_n(\mu_0, \mu_1, k) - \mu_2] = \frac{-c\sqrt{n}}{2} + \frac{\ln(k)}{c\sqrt{n}}$$

It follows that the probability of weak evidence $P_{\mu=\mu_1}(\text{WEV})$ is given by

$$\Phi\left(\frac{c\sqrt{n}}{2} + \frac{\ln(k)}{c\sqrt{n}}\right) - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k)}{c\sqrt{n}}\right)$$

If we define

$$W(x,y) = \Phi\left(\frac{c\sqrt{n}}{2} + \frac{\ln(k_1)}{c\sqrt{n}}\right) - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k_2)}{c\sqrt{n}}\right)$$

then the two probabilities are given by

$$W_1 = P_1(\text{WEV}) = W(k_2, k_1)$$
$$\text{and}$$
$$W_2 = P_2(\text{WEV}) = W(k_1, k_2)$$

There is nothing that requires the same level of statistical evidence be the same for μ_1 vs μ_0 as for μ_0 vs μ_1. That is we say we have statistical evidence for μ_1 vs μ_0 of level k_1 if $\boldsymbol{L}(\theta_1, \theta_0; \boldsymbol{x}_{obs}) > k_1$ and statistical evidence for μ_0 vs μ_1 of level k_0 if $\boldsymbol{L}(\theta_0, \theta_1; \boldsymbol{x}_{obs}) > k_0$.

We then have the following summary of results for the normal distribution example.

- When H_0 is true the probability of misleading evidence for H_1 at level k_1 defined by $(\text{L}_1/\text{L}_0 \geq k_1)$ is

$$M_0 = \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k_2)}{c\sqrt{n}}\right)$$

- When H_0 is true the probability of weak evidence is

$$W_0 = P_{\mu=\mu_0}\left(\frac{1}{k_0} \leq \frac{\text{L}_1}{\text{L}_0} \leq k_1\right)$$

$$= \Phi\left(\frac{c\sqrt{n}}{2} + \frac{\ln(k_1)}{c\sqrt{n}}\right) - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k_0)}{c\sqrt{n}}\right)$$

- When H_1 is true the probability of misleading evidence for H_0 at level k_0 defined by $(\text{L}_0/\text{L}_1 \geq k_0)$ is

$$M_1 = \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k_1)}{c\sqrt{n}}\right)$$

- When H_1 is true the probability of weak evidence is

$$W_1 = P_{\mu=\mu_1}\left(\frac{1}{k_0} \le \frac{\mathrm{L}_2}{\mathrm{L}_1} \le k_1\right)$$

$$= \Phi\left(\frac{c\sqrt{n}}{2} + \frac{\ln(k_0)}{c\sqrt{n}}\right) - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k_1)}{c\sqrt{n}}\right)$$

17.2.2   Sample Size

There is no doubt that one of the questions most asked of a statistician is

How many observations do I need?

Actually the usual question is how many subjects do I need to get a statistically significant result that is publishable? This question is easily answered, so let us consider refining the question.

Suppose that we will observe $X_1, X_2, \ldots, X_n$ assumed independent and identically distributed as normal with mean μ and variance σ^2 assumed known (usually based on past work with similar instruments, and so on). Of interest is a (null) hypothesis $H_0 : \mu = \mu_0$ and an alternative $H_1 : \mu = \mu_1$ where without loss of generality we assume that $\mu_1 > \mu_0$. It is assumed that μ_1 represents a value of μ which is of scientific importance,i.e., if μ_1 is true then a result of scientific or practical importance has been discovered.

The Neyman–Pearson theory has been used for decades to determine sample size is the default method. It is required in submitting grants to NIH, NSF, FDA, etc., as well as in reporting the results of published studies and dissertations. The Neyman–Pearson approach to sample size selection is as follows:

1. Choose a value α for the significance level (usually $\alpha = 0.05$).
2. Choose a value $1 - \beta$ for the power (usually $\beta = 0.20$ so that the power is 0.8).
3. Select the sample size n so that

$$P(\text{Type I error}) = P(\text{reject } H_0 | H_0 \text{ true}) = \alpha$$
$$1 - P(\text{Type II error}) = P(\text{reject } H_0 | H_1 \text{ true}) = 1 - \beta$$

In the case of a normal distribution with known variance we have that

$$P(\text{Type I error}) = P(\overline{X} \ge C | \mu = \mu_0)$$

$$= P\left(\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \ge \frac{\sqrt{n}(C - \mu_1)}{\sigma}\bigg| \mu = \mu_0\right)$$

$$= 1 - \Phi\left(\frac{C - \mu_1}{\sigma}\right)$$

and it follows that

$$C = \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = \mu_0 + 1.645\frac{\sigma}{\sqrt{n}} \text{ if } \alpha = 0.05$$

and

$$\text{Power} = P(\overline{X} \geq C | \mu = \mu_1)$$

$$= P\left(\overline{X} \geq \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\middle| \mu = \mu_1\right)$$

$$= P\left(\frac{\sqrt{n}(\overline{X} - \mu_1)}{\sigma} \geq z_{1-\alpha/2} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}\middle| \mu = \mu_1\right)$$

$$= 1 - \Phi\left(z_{1-\alpha/2} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}\right)$$

$$= 1 - \Phi(z_{1-\alpha/2} - c\sqrt{n})$$

In order to have power $1 - \beta$ we must have

$$\Phi\left(z_{1-\alpha/2} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}\right) = \beta$$

i.e.,

$$z_{1-\alpha/2} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma} = z_\beta$$

or

$$z_{1-\alpha/2} - z_\beta = (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}$$

and it follows that

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{c^2}$$

where

$$c = \frac{\mu_1 - \mu_0}{\sigma}$$

This is the prototype of sample size formulas.

The Neyman-Pearson approach is inadequate when we want to quantify statistical evidence for H_1 vs H_0 we now consider the selection of sample size necessary to quantify statistical evidence. Recall that there are four probabilities involved:

1. The probability of misleading statistical evidence for H_1 when H_0 is true
2. The probability of misleading statistical evidence for H_0 when H_1 is true
3. The probability of weak statistical evidence when H_0 is true
4. The probability of weak evidence when H_1 is true

The analogue to the Type I error probability is the probability of finding misleading evidence for H_1 when H_0 is true. For the normal distribution we have the correspondence for $\alpha = 0.05$ and M_0 given by

$$\alpha = 0.05 \quad M_0 = \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(8)}{c\sqrt{n}}\right)$$

and if we take $c = 0.5$, a moderate effect size, we have the correspondence

$$\alpha = 0.05 \quad M_0 = \Phi\left(-\frac{\sqrt{n}}{4} - \frac{4\ln(8)}{\sqrt{n}}\right)$$

For the analogue to the Type II error we must be more careful. The probability of failing to find evidence supporting H_1 when H_0 is true is composed of two parts:

1. The probability of misleading evidence in favor of H_0 when H_1 is true
2. The probability of weak evidence when H_1 is true

For the normal distribution we have the correspondence

$$\beta = P(\text{Type II error}) = \Phi\left(z_{1-\alpha/2} - (\mu_2 - \mu_1)\frac{\sqrt{n}}{\sigma}\right)$$

$$M_1 + W_1 = \Phi\left(-\frac{c\sqrt{n}}{2} - \frac{\ln(k_1)}{c\sqrt{n}}\right) + \Phi\left(\frac{c\sqrt{n}}{2} + \frac{\ln(k_1)}{c\sqrt{n}}\right) - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k_2)}{c\sqrt{n}}\right)$$

$$= 1 - \Phi\left(\frac{c\sqrt{n}}{2} - \frac{\ln(k_2)}{c\sqrt{n}}\right)$$

$$= \Phi\left(\frac{-c\sqrt{n}}{2} + \frac{\ln(k_2)}{c\sqrt{n}}\right)$$

and if $\beta = 0.2$, $c = 0.5$ and $k_2 = 8$ we have

$$\beta = 0.2 \ ; \ M_2 + W_2 = \Phi\left(-\frac{\sqrt{n}}{4} + \frac{2\ln(8)}{\sqrt{n}}\right)$$

For the Neyman Pearson sample size formula for $\alpha = 0.05$, $\beta = 0.20$ and $c = 0.5$ we get a sample size of

$$n = \frac{(1645 + 0.84)^2}{0.5^2} = 25$$

For this sample size we find that $M_1 + W_1$ is equal to

$$M_1 + W_1 = \Phi\left(-\frac{\sqrt{25}}{4} + \frac{2\ln(8)}{\sqrt{25}}\right) = \Phi(-0.418) = 0.34$$

Thus the conventional sample size formula does not lead to a small probability of finding weak evidence.

Exercises in Royall's book show that the this is true in general, i.e., conventional sample size formulas do not guarantee finding strong evidence.

17.3   Birnbaum's Confidence Concept

Recall Birnbaum's confidence concept which he advocated after becoming skeptical of the likelihood principle.

A concept of statistical evidence is not plausible unless it finds "strong evidence" for H_2 as against H_1 with small probability (α) when H_1 is true and with much larger probability ($1 - \beta$) when H_2 is true.

What the results in the sample size section show is that it is possible in certain cases to satisfy the confidence concept with sufficient observations.

17.4   Combining Evidence

Suppose that we have two independent estimators, t_1 and t_2 of a parameter θ where t_1 is normal with expected value θ and variance v_1 and t_2 is normal with expected value θ and variance v_2. Assume that v_1 and v_2 are known.

The joint density of t_1 and t_2 is

$$f(t_1, t_2; \theta) = \frac{1}{2\pi\sqrt{v_1 v_2}} \exp\left\{-\frac{(t_1 - \theta)^2}{2v_1} - \frac{(t_2 - \theta)^2}{2v_2}\right\}$$

which has logarithm

$$-\ln[2\pi\sqrt{v_1 v_2}] - \frac{(t_1 - \theta)^2}{2v_1} - \frac{(t_2 - \theta)^2}{2v_2}$$

The derivative with respect to θ is thus

$$\frac{(t_1 - \theta)}{v_1} - \frac{t_2 - \theta}{v_2}$$

and hence the maximum likelihood estimate of θ is

$$\widehat{\theta} = \frac{\frac{t_1}{v_1} + \frac{t_2}{v_2}}{\frac{1}{v_1} + \frac{1}{v_2}}$$

At this value of θ the joint density is

$$f(t_1, t_2; \widehat{\theta}) = \frac{1}{2\pi\sqrt{v_1 v_2}} \exp\left\{ -\frac{(t_1 - \widehat{\theta})^2}{2v_1} - \frac{(t_2 - \widehat{\theta})^2}{2v_2} \right\}$$

and hence the likelihood for θ is

$$\mathscr{L}(\theta; t_1, t_2) = \frac{f(t_1, t_2; \theta)}{f(t_1, t_2; \widehat{\theta})} = \frac{\frac{1}{2\pi\sqrt{v_1 v_2}} \exp\left\{ -\frac{(t_1 - \theta)^2}{2v_1} - \frac{(t_2 - \theta)^2}{2v_2} \right\}}{\frac{1}{2\pi\sqrt{v_1 v_2}} \exp\left\{ -\frac{(t_1 - \widehat{\theta})^2}{2v_1} - \frac{(t_2 - \widehat{\theta})^2}{2v_2} \right\}}$$

or

$$\mathscr{L}(\theta; t_1, t_2) = \exp\left\{ \frac{t_1\theta}{v_1} + \frac{t_2\theta}{v_2} - \frac{\theta^2}{2v_1} - \frac{\theta^2}{2v_2} - \frac{t_1\widehat{\theta}}{v_1} - \frac{t_2\widehat{\theta}}{v_2} + \frac{\widehat{\theta}^2}{2v_1} + \frac{\widehat{\theta}^2}{2v_2} \right\}$$

$$= \exp\left\{ \theta\left(\frac{t_1}{v_1} + \frac{t_2}{v_2} \right) - \frac{\theta^2}{2}\left(\frac{1}{v_1} + \frac{1}{v_2} \right) - \widehat{\theta}\left(\frac{t_1}{v_1} + \frac{t_2}{v_2} \right) \right.$$
$$\left. + \frac{\widehat{\theta}^2}{2}\left(\frac{1}{v_1} + \frac{1}{v_2} \right) \right\}$$

$$= \exp\left\{ \theta\widehat{\theta}\left(\frac{1}{v_1} + \frac{1}{v_2} \right) - \frac{\theta^2}{2}\left(\frac{1}{v_1} + \frac{1}{v_2} \right) - \frac{\widehat{\theta}^2}{2}\left(\frac{1}{v_1} + \frac{1}{v_2} \right) \right\}$$

$$= \exp\left\{ -\frac{1}{2}\left(\frac{1}{v_1} + \frac{1}{v_2} \right)\left(\theta^2 - 2\theta\widehat{\theta} + \widehat{\theta}^2 \right) \right\}$$

$$= \exp\left\{ -\left(\frac{1}{v_1} + \frac{1}{v_2} \right)\frac{(\theta - \widehat{\theta})^2}{2} \right\}$$

$$= \exp\left\{ \frac{(\theta - \widehat{\theta})^2}{2v} \right\}$$

where

$$v = \frac{1}{\frac{1}{v_1} + \frac{1}{v_2}}$$

which is a normal likelihood centered at $\widehat{\theta}$ and curvature v.

This is a likelihood version of the standard result that to combine unbiased uncorrelated estimators weight inversely as their variance and divide the result by the sum of the weights.

In general note that if we have $\boldsymbol{x}_1$ observations from $f(x; \theta)$ and independent observations $\boldsymbol{x}_2$ from $f(x; \theta)$ then the evidence for θ_1 vs θ_0 based on $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is

$$\frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2; \theta_1)}{f(\boldsymbol{x}_1, \boldsymbol{x}_2; \theta_0)} = \left[\frac{f(\boldsymbol{x}_1; \theta_1)}{f(\boldsymbol{x}_1; \theta_0)}\right] \left[\frac{f(\boldsymbol{x}_2; \theta_1)}{f(\boldsymbol{x}_2; \theta_0)}\right]$$

i.e., evidence is multiplicative.

17.5   Exercises

1. Suppose that $X_1, X_2, \ldots, X_n$ are iid, each Poisson with parameter λ. Let $k = 8$, $n = 1, 10, 25$. Draw graphs of the probability of misleading evidence for $\lambda_1 = 2$ vs $\lambda_0 = 1$.
2. Repeat Exercise 1 for the binomial with $n = 10, 25, 100, 1000$ and $p = 0.6$ vs $p = 0.5$.