3.1.2 Confidence regions

It is often desirable to identify regions of the parameter space that are likely to contain the true value of the parameter. To do this, after observing the data $Y = y$ we can construct an interval $[l(y), u(y)]$ such that the probability that $l(y) < \theta < u(y)$ is large.

Definition 5 (Bayesian coverage) *An interval* $[l(y), u(y)]$*, based on the observed data* $Y = y$*, has 95% Bayesian coverage for* θ *if*

$$\Pr(l(y) < \theta < u(y) | Y = y) = .95.$$

The interpretation of this interval is that it describes your information about the location of the true value of θ after you have observed $Y = y$. This is different from the frequentist interpretation of coverage probability, which describes the probability that the interval will cover the true value *before* the data are observed:

Definition 6 (frequentist coverage) *A random interval* $[l(Y), u(Y)]$ *has 95% frequentist coverage for* θ *if, before the data are gathered,*

$$\Pr(l(Y) < \theta < u(Y) | \theta) = .95.$$

In a sense, the frequentist and Bayesian notions of coverage describe pre- and post-experimental coverage, respectively.

You may recall your introductory statistics instructor belaboring the following point: Once you observe $Y = y$ and you plug this data into your confidence interval formula $[l(y), u(y)]$, then

$$\Pr(l(y) < \theta < u(y) | \theta) = \begin{cases} 0 & \text{if } \theta \notin [l(y), u(y)]; \\ 1 & \text{if } \theta \in [l(y), u(y)]. \end{cases}$$

This highlights the lack of a post-experimental interpretation of frequentist coverage. Although this may make the frequentist interpretation seem somewhat lacking, it is still useful in many situations. Suppose you are running a large number of unrelated experiments and are creating a confidence interval for each one of them. If your intervals each have 95% frequentist coverage probability, you can expect that 95% of your intervals contain the correct parameter value.

Can a confidence interval have the same Bayesian and frequentist coverage probability? Hartigan (1966) showed that, for the types of intervals we will construct in this book, an interval that has 95% Bayesian coverage additionally has the property that

$$\Pr(l(Y) < \theta < u(Y) | \theta) = .95 + \epsilon_n$$

where $|\epsilon_n| < \frac{a}{n}$ for some constant a. This means that a confidence interval procedure that gives 95% Bayesian coverage will have approximately 95% frequentist coverage as well, at least asymptotically. It is important to keep in mind that most non-Bayesian methods of constructing 95% confidence intervals also only achieve this coverage rate asymptotically. For more discussion of the similarities between intervals constructed by Bayesian and non-Bayesian methods, see Severini (1991) and Sweeting (2001).

Quantile-based interval

Perhaps the easiest way to obtain a confidence interval is to use posterior quantiles. To make a $100 \times (1-\alpha)\%$ quantile-based confidence interval, find numbers $\theta_{\alpha/2} < \theta_{1-\alpha/2}$ such that

1. $\Pr(\theta < \theta_{\alpha}|Y = y) = \alpha/2$;
2. $\Pr(\theta > \theta_{1-\alpha/2}|Y = y) = \alpha/2$.

The numbers $\theta_{\alpha/2}, \theta_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ posterior quantiles of θ, and so

$$\begin{aligned}
\Pr(\theta \in [\theta_{\alpha/2}, \theta_{1-\alpha/2}]|Y = y) &= 1 - \Pr(\theta \notin [\theta_{\alpha/2}, \theta_{1-\alpha/2}]|Y = y) \\
&= 1 - [\Pr(\theta < \theta_{\alpha/2}|Y = y) + \Pr(\theta > \theta_{1-\alpha/2}|Y = y)] \\
&= 1 - \alpha\,.
\end{aligned}$$

Example: Binomial sampling and uniform prior

Suppose out of $n = 10$ conditionally independent draws of a binary random variable we observe $Y = 2$ ones. Using a uniform prior distribution for θ, the posterior distribution is $\theta|\{Y = 2\} \sim \text{beta}(1+2, 1+8)$. A 95% posterior confidence interval can be obtained from the .025 and .975 quantiles of this beta distribution. These quantiles are 0.06 and 0.52 respectively, and so the posterior probability that $\theta \in [0.06, 0.52]$ is 95%.

```
> a<-1 ; b<-1   #prior
> n<-10 ; y<-2   #data

> qbeta( c(.025,.975), a+y,b+n-y)

[1] 0.06021773 0.51775585
```

Highest posterior density (HPD) region

Figure 3.5 shows the posterior distribution and a 95% confidence interval for θ from the previous example. Notice that there are θ-values *outside* the quantile-based interval that have higher probability (density) than some points *inside* the interval. This suggests a more restrictive type of interval:

Definition 7 (HPD region) *A $100 \times (1-\alpha)\%$ HPD region consists of a subset of the parameter space, $s(y) \subset \Theta$ such that*

1. $\Pr(\theta \in s(y)|Y = y) = 1 - \alpha$;

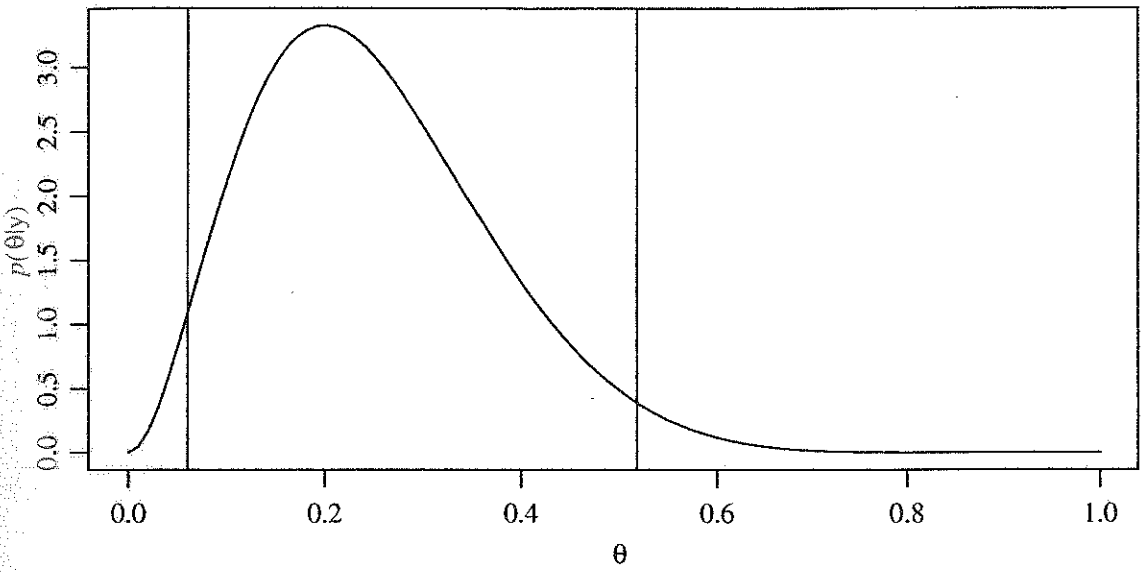

Fig. 3.5. A beta posterior distribution, with vertical bars indicating a 95% quantile-based confidence interval.

2. *If* $\theta_a \in s(y)$*, and* $\theta_b \notin s(y)$*, then* $p(\theta_a|Y = y) > p(\theta_b|Y = y)$.

All points in an HPD region have a higher posterior density than points outside the region. However, an HPD region might not be an interval if the posterior density is multimodal (having multiple peaks). Figure 3.6 gives the basic idea behind the construction of an HPD region: Gradually move a horizontal line down across the density, including in the HPD region all θ-values having a density above the horizontal line. Stop moving the line down when the posterior probability of the θ-values in the region reaches $(1 - \alpha)$. For the binomial example above, the 95% HPD region is $[0.04, 0.048]$, which is narrower (more precise) than the quantile-based interval, yet both contain 95% of the posterior probability.

3.2 The Poisson model

Some measurements, such as a person's number of children or number of friends, have values that are whole numbers. In these cases our sample space is $\mathcal{Y} = \{0, 1, 2, \ldots\}$. Perhaps the simplest probability model on $\mathcal{Y}$ is the Poisson model.

Poisson distribution

Recall from Chapter 2 that a random variable Y has a Poisson distribution with mean θ if

$$\Pr(Y = y|\theta) = \text{dpois}(y, \theta) = \theta^y e^{-\theta}/y! \quad \text{for } y \in \{0, 1, 2, \ldots\} .$$

For such a random variable,

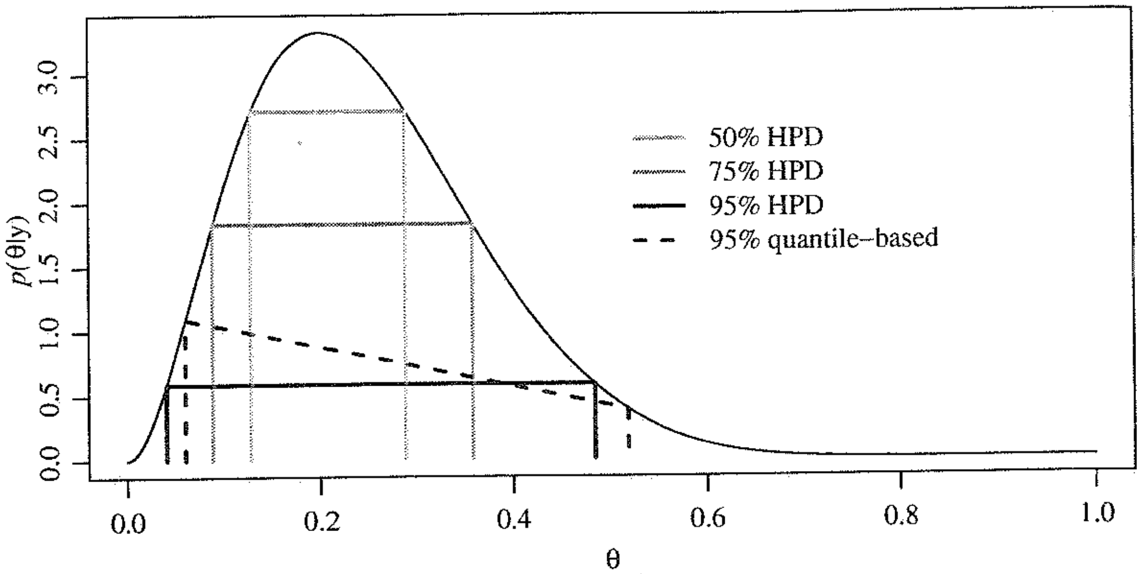

Fig. 3.6. Highest posterior density regions of varying probability content. The dashed line is the 95% quantile-based interval.

- $\mathrm{E}[Y|\theta] = \theta$;
- $\mathrm{Var}[Y|\theta] = \theta$.

People sometimes say that the Poisson family of distributions has a "mean-variance relationship" because if one Poisson distribution has a larger mean than another, it will have a larger variance as well.

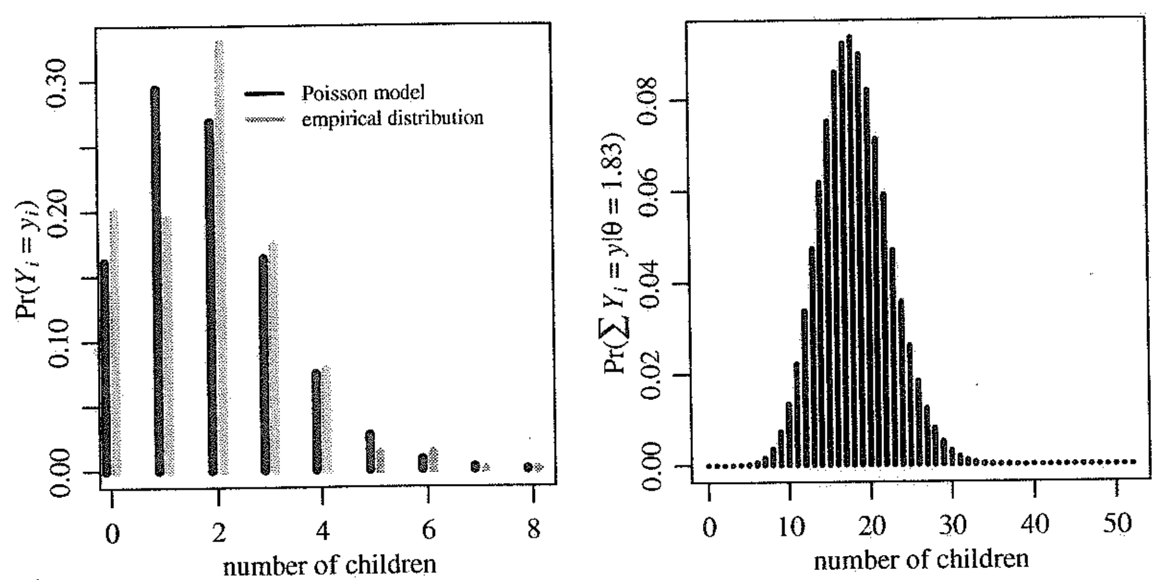

Fig. 3.7. Poisson distributions. The first panel shows a Poisson distribution with mean of 1.83, along with the empirical distribution of the number of children of women of age 40 from the GSS during the 1990s. The second panel shows the distribution of the sum of 10 i.i.d. Poisson random variables with mean 1.83. This is the same as a Poisson distribution with mean 18.3