now wish to examine various properties of $p(\boldsymbol{\theta} \mid \boldsymbol{x})$ as the number of observations increases; i.e., as $n \to \infty$. Intuitively, we would hope that beliefs about $\boldsymbol{\theta}$ would become more and more concentrated around the "true" parameter value; i.e., the corresponding strong law limit. Under appropriate conditions, we shall see that this is, indeed, the case.

5.3.1 Discrete Asymptotics

We begin by considering the situation where $\Theta = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \}$ consists of a countable (possibly finite) set of values, such that the parametric model corresponding to the true parameter, $\boldsymbol{\theta}_t$, is "distinguishable" from the others, in the sense that the logarithmic divergences, $\int p(x \mid \boldsymbol{\theta}_t) \log[p(x \mid \boldsymbol{\theta}_t)/p(x \mid \boldsymbol{\theta}_i)]\, dx$ are strictly larger than zero, for all $i \neq t$.

Proposition 5.12. (***Discrete asymptotics***). *Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be observations for which a belief model is defined by the parametric model $p(x \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots\}$, and the prior $p(\boldsymbol{\theta}) = \{p_1, p_2, \ldots\}$, $p_i > 0$, $\sum_i p_i = 1$. Suppose that $\boldsymbol{\theta}_t \in \Theta$ is the true value of $\boldsymbol{\theta}$ and that, for all $i \neq t$,*

$$\int p(x \mid \boldsymbol{\theta}_t) \log \left[\frac{p(x \mid \boldsymbol{\theta}_t)}{p(x \mid \boldsymbol{\theta}_i)}\right] dx > 0;$$

then

$$\lim_{n\to\infty} p(\boldsymbol{\theta}_t \mid \boldsymbol{x}) = 1, \quad \lim_{n\to\infty} p(\boldsymbol{\theta}_i \mid \boldsymbol{x}) = 0,\ i \neq t.$$

Proof. By Bayes' theorem, and assuming that $p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^n p(x_i|\boldsymbol{\theta})$,

$$\begin{aligned} p(\boldsymbol{\theta}_i \mid \boldsymbol{x}) &= p_i \, \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta}_i)}{p(\boldsymbol{x})} \\ &= \frac{p_i \{p(\boldsymbol{x} \mid \boldsymbol{\theta}_i)/p(\boldsymbol{x} \mid \boldsymbol{\theta}_t)\}}{\sum_i p_i \{p(\boldsymbol{x} \mid \boldsymbol{\theta}_i)/p(\boldsymbol{x} \mid \boldsymbol{\theta}_t)\}} \\ &= \frac{\exp\{\log p_i + S_i\}}{\sum_i \exp\{\log p_i + S_i\}}, \end{aligned}$$

where

$$S_i = \sum_{j=1}^n \log \frac{p(x_j \mid \boldsymbol{\theta}_i)}{p(x_j \mid \boldsymbol{\theta}_t)}\,.$$

Conditional on $\boldsymbol{\theta}_t$, the latter is the sum of n independent identically distributed random quantities and hence, by the strong law of large numbers (see Section 3.2.3),

$$\lim_{n\to\infty} \frac{1}{n} S_i = \int p(x \mid \boldsymbol{\theta}_t) \log \left[\frac{p(x \mid \boldsymbol{\theta}_i)}{p(x \mid \boldsymbol{\theta}_t)}\right] dx.$$

The right-hand side is negative for all $i \neq t$, and equals zero for $i = t$, so that, as $n \to \infty$, $S_t \to 0$ and $S_i \to -\infty$ for $i \neq t$, which establishes the result. ◁

An alternative way of expressing the result of Proposition 5.12, established for countable Θ, is to say that the posterior distribution function for $\boldsymbol{\theta}$ ultimately degenerates to a step function with a single (unit) step at $\boldsymbol{\theta} = \boldsymbol{\theta}_t$. In fact, this result can be shown to hold, under suitable regularity conditions, for much more general forms of Θ. However, the proofs require considerable measure-theoretic machinery and the reader is referred to Berk (1966, 1970) for details.

A particularly interesting result is that if the true $\boldsymbol{\theta}$ is *not* in Θ, the posterior degenerates onto the value in Θ which gives the parametric model closest in logarithmic divergence to the true model.

5.3.2 Continuous Asymptotics

Let us now consider what can be said in the case of general Θ about the forms of probability statements implied by $p(\boldsymbol{\theta} \mid \boldsymbol{x})$ for large n. Proceeding heuristically for the moment, without concern for precise regularity conditions, we note that, in the case of a parametric representation for an exchangeable sequence of observables,

$$\begin{aligned} p(\boldsymbol{\theta} \mid \boldsymbol{x}) &\propto p(\boldsymbol{\theta}) \prod_{i=1}^{n} p(x_i \mid \boldsymbol{\theta}) \\ &\propto \exp\left\{\log p(\boldsymbol{\theta}) + \log p(\boldsymbol{x} \mid \boldsymbol{\theta})\right\}. \end{aligned}$$

If we now expand the two logarithmic terms about their respective maxima, $\boldsymbol{m}_0$ and $\hat{\boldsymbol{\theta}}_n$, assumed to be determined by setting $\nabla \log p(\boldsymbol{\theta}) = 0$, $\nabla \log p(\boldsymbol{x} \mid \boldsymbol{\theta}) = 0$, respectively, we obtain

$$\begin{aligned} \log p(\boldsymbol{\theta}) &= \log p(\boldsymbol{m}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m}_0)^t \boldsymbol{H}_0 (\boldsymbol{\theta} - \boldsymbol{m}_0) + R_0 \\ \log p(\boldsymbol{x} \mid \boldsymbol{\theta}) &= \log p(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}_n) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)^t \boldsymbol{H}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) + R_n, \end{aligned}$$

where R_0, R_n denote remainder terms and

$$\boldsymbol{H}_0 = \left(-\frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{m}_0} \qquad \boldsymbol{H}(\hat{\boldsymbol{\theta}}_n) = \left(-\frac{\partial^2 \log p(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} \right)\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}.$$

Assuming regularity conditions which ensure that R_0, R_n are small for large n, and ignoring constants of proportionality, we see that

$$\begin{aligned} p(\boldsymbol{\theta} \mid \boldsymbol{x}) &\propto \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m}_0)^t \boldsymbol{H}_0 (\boldsymbol{\theta} - \boldsymbol{m}_0) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)^t H(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) \right\} \\ &\propto \exp\left\{ -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m}_n)^t \boldsymbol{H}_n (\boldsymbol{\theta} - \boldsymbol{m}_n) \right\}, \end{aligned}$$