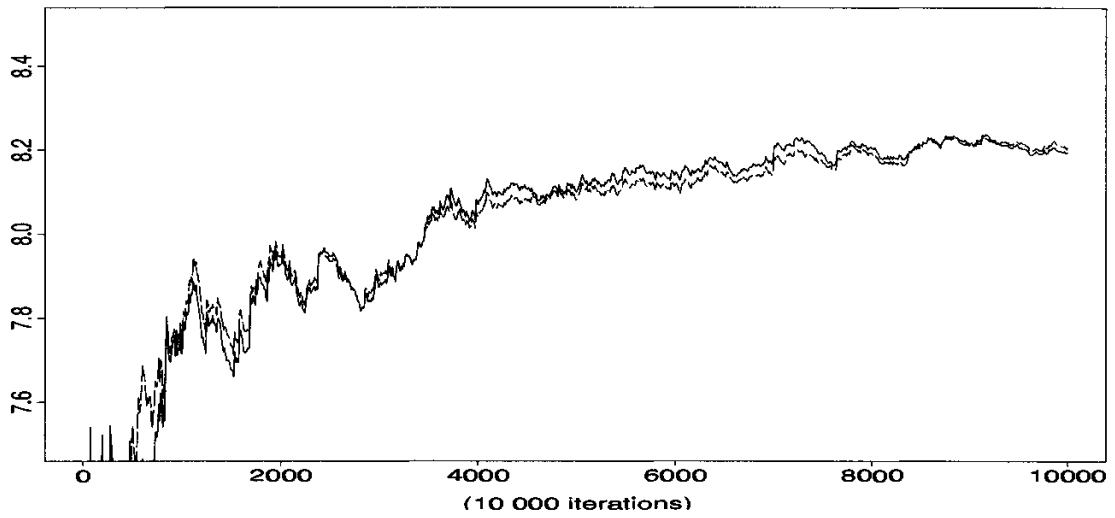

Fig. 7.2. Convergence to $\mathbb{E}_f[X^2] = 8.33$ of Accept–Reject (full line) and Metropolis–Hastings (dots) estimators for 10,000 acceptances in [*A*.27], the same sequence of y_i's simulated from $\mathcal{G}a(2, 2/2.43)$ being used in [*A*.27] and [*A*.26]. The final values of the estimators are 8.20 for [*A*.27] and 8.21 for [*A*.26].

t of values accepted by [*A*.27] is then random. Figure 7.1 describes the convergence of the estimators of $\mathbb{E}_f[X^2]$ associated with both algorithms for the same sequence of y_i's and exhibits strong agreement between the approaches, with the estimator based on [*A*.26] being closer to the exact value 8.33 in this case.

On the other hand, the number t of values accepted by [*A*.27] can be fixed and [*A*.26] can then use the resulting sample of random size n, $y_1, \ldots, y_n$. Figure 7.2 reproduces the comparison in this second case and exhibits a behavior rather similar to Figure 7.1, with another close agreement between estimators and, the scale being different, a smaller variance (which is due to the larger size of the effective sample).

Note, however, that both comparisons are biased. In the first case, the sample of $X^{(i)}$ produced by [*A*.27] does not have the distribution f and, in the second case, the sample of Y_i's in [*A*.26] is not iid. In both cases, this is due to the use of a stopping rule which modifies the distribution of the samples.||

Example 7.11. Logistic Regression. We return to the data of Example 1.13, which described a logistic regression relating the failure of O-rings in shuttle flights to air temperature. We observe (x_i, y_i), $i = 1, \ldots, n$ according to the model

$$Y_i \sim \text{Bernoulli}(p(x_i)), \quad p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)},$$

where $p(x)$ is the probability of an O-ring failure at temperature x. The likelihood is

$$L(\alpha, \beta|\mathbf{y}) \propto \prod_{i=1}^{n} \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)} \right)^{1-y_i}$$

and we take the prior to be

$$\pi_\alpha(\alpha|b)\pi_\beta(\beta) = \frac{1}{b} e^{\alpha} e^{-e^{\alpha}/b} d\alpha d\beta,$$

which puts an exponential prior on $\log \alpha$ and a flat prior on β, and insures propriety of the posterior distribution (Problem 7.25). To complete the prior specification we must give a value for b, and we choose the data-dependent value that makes $\mathbb{E}\alpha = \hat{\alpha}$, where $\hat{\alpha}$ is the MLE of α. (This also insures that the prior will not have undue influence, as it is now centered near the likelihood.) It can be shown that

$$\mathbb{E}[\alpha] = \int_0^\infty \frac{1}{b} e^{\alpha} e^{-e^{\alpha}/b} d\alpha = \int_0^\infty \log(w) \frac{1}{b} e^{-w/b} dw = \log(b) - \gamma,$$

where γ is *Euler's Constant*, equal to .577216. Thus we take $\hat{b} = e^{\hat{\alpha}+\gamma}$.

The posterior distribution is proportional to $L(\alpha, \beta|\mathbf{y})\pi(\alpha, \beta)$, and to simulate from this distribution we take an independent candidate

$$g(\alpha, \beta) = \pi_\alpha(\alpha|\hat{b})\phi(\beta),$$

where $\phi(\beta)$ is a normal distribution with mean $\hat{\beta}$ and variance $\hat{\sigma}_\beta^2$, the MLEs. Note that although basing the prior distribution on the data is somewhat in violation of the formal Bayesian paradigm, nothing is violated if the candidate depends on the data. In fact, this will usually result in a more effective simulation, as the candidate is placed close to the target.

Generating a random variable from $g(\alpha, \beta)$ is straightforward, as it only involves the generation of a normal and an exponential random variable. If we are at the point (α_0, β_0) in the Markov chain, and we generate (α', β') from $g(\alpha, \beta)$, we accept the candidate with probability

$$\min \left\{ \frac{L(\alpha', \beta'|\mathbf{y})}{L(\alpha_0, \beta_0|\mathbf{y})} \frac{\phi(\beta_0)}{\phi(\beta')}, 1 \right\}.$$

Figure 7.3 shows the distribution of the generated parameters and their convergence. ||

Example 7.12. Saddlepoint tail area approximation. In Example 3.18, we saw an approximation to noncentral chi squared tail areas based on the regular and renormalized saddlepoint approximations. Such an approximation requires numerical integration, both to calculate the constant and to evaluate the tail area.

An alternative is to produce a sample $Z_1, \ldots, Z_m$, from the saddlepoint distribution, and then approximate the tail area using