$$h(x) \approx \frac{\hat{x}_\theta}{\beta} + (\alpha - 1)\log(\hat{x}_\theta) + \frac{\alpha - 1}{2\hat{x}_\theta^2}(x - \hat{x}_\theta)^2$$

Now substituting into (3.22) yields the Laplace approximation

$$\int_a^b \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta} dx = \hat{x}_\theta^{\alpha-1} e^{\hat{x}_\theta/\beta} \sqrt{\frac{2\pi\hat{x}_\theta^2}{\alpha - 1}} \times \left\{ \Phi\left(\sqrt{\frac{\alpha - 1}{\hat{x}_\theta^2}}(b - \hat{x}_\theta) \right) - \Phi\left(\sqrt{\frac{\alpha - 1}{\hat{x}_\theta^2}}(a - \hat{x}_\theta) \right) \right\}$$

For $\alpha = 5$ and $\beta = 2$, $\hat{x}_\theta = 8$, and the approximation will be best in that area. In Table 3.6 we see that although the approximation is reasonable in the central region of the density, it becomes quite unacceptable in the tails. ‖

Interval	Approximation	Exact
$(7, 9)$	0.193351	0.193341
$(6, 10)$	0.375046	0.37477
$(2, 14)$	0.848559	0.823349
$(15.987, \infty)$	0.0224544	0.100005

Table 3.6. Laplace approximation of a Gamma integral for $\alpha = 5$ and $\beta = 2$.

Thus, we see both the usefulness and the limits of the Laplace approximation. In problems where Monte Carlo calculations are prohibitive because of computing time, the Laplace approximation can be useful as a guide to the solution of the problem. Also, the corresponding Taylor series can be used as a proposal density, which is particularly useful in problems where no obvious proposal exists. (See Example 7.12 for a similar situation.)

3.5 Problems

3.1 For the normal-Cauchy Bayes estimator

$$\delta(x) = \frac{\int_{-\infty}^{\infty} \frac{\theta}{1+\theta^2} e^{-(x-\theta)^2/2} d\theta}{\int_{-\infty}^{\infty} \frac{1}{1+\theta^2} e^{-(x-\theta)^2/2} d\theta}$$

(a) Plot the integrand and use Monte Carlo integration to calculate the integral.
(b) Monitor the convergence with the standard error of the estimate. Obtain three digits of accuracy with probability .95.

3.2 (Continuation of Problem 3.1)
(a) Use the Accept–Reject algorithm, with a Cauchy candidate, to generate a sample from the posterior distribution and calculate the estimator.

(b) Design a computer experiment to compare Monte Carlo error when using (i) the same random variables θ_i in numerator and denominator, or (ii) different random variables.

3.3 (a) For a standard normal random variable Z, calculate $P(Z > 2.5)$ using Monte Carlo sums based on indicator functions. How many simulated random variables are needed to obtain three digits of accuracy?

(b) Using Monte Carlo sums verify that if $X \sim \mathcal{G}(1,1)$, $P(X > 5.3) \approx .005$. Find the exact .995 cutoff to three digits of accuracy.

3.4 (a) If $X \sim \mathcal{N}(0, \sigma^2)$, show that

$$\mathbb{E}[e^{-X^2}] = \frac{1}{\sqrt{2\sigma^2 + 1}} .$$

(b) Generalize to the case $X \sim \mathcal{N}(\mu, \sigma^2)$.

3.5 Referring to Example 3.6:

(a) Verify the maximum of the likelihood ratio statistic.

(b) Generate 5000 random variables according to (3.7), recreating the left panel of Figure 3.2. Compare this distribution to a null distribution where we *fix* null values of p_1 and p_2, for example, $(p_1, p_2) = (.25, .75)$. For a range of values of (p_1, p_2), compare the histograms both with the one from (3.7) and the χ_1^2 density. What can you conclude?

3.6 An alternate analysis to that of Example 3.6 is to treat the contingency table as two binomial distributions, one for the patients receiving surgery and one for those receiving radiation. Then the test of hypothesis becomes a test of equality of the two binomial parameters. Repeat the analysis of the data in Table 3.2 under the assumption of two binomials. Compare the results to those of Example 3.6.

3.7 A famous medical experiment was conducted by Joseph Lister in the late 1800s to examine the relationship between the use of a disinfectant, carbolic acid, and surgical success rates. The data are

	Disinfectant	
	Yes	No
Success	34	19
Failure	6	16

Using the techniques of Example 3.6, analyze these data to examine the association between disinfectant and surgical success rates. Use both the multinomial model and the two-binomial model.

3.8 Referring to Example 3.3, we calculate the expected value of $\delta^\pi(x)$ from the posterior distribution $\pi(\theta|x) \propto \|\theta\|^{-2} \exp\left\{-\|x-\theta\|^2/2\right\}$, arising from a normal likelihood and noninformative prior $\|\theta\|^{-2}$ (see Example 1.12).

(a) Show that if the quadratic loss of Example 3.3 is normalized by $1/(2\|\theta\|^2 + p)$, the resulting Bayes estimator is

$$\delta^\pi(x) = \mathbb{E}^\pi\left[\frac{\|\theta\|^2}{2\|\theta\|^2 + p}\,\middle|\, x, \lambda\right] \Big/ \mathbb{E}^\pi\left[\frac{1}{2\|\theta\|^2 + p}\,\middle|\, x, \lambda\right] .$$

(b) Simulation of the posterior can be done by representing θ in polar coordinates $(\rho, \varphi_1, \varphi_2)$ $(\rho > 0, \varphi_1 \in [-\pi/2, \pi/2], \varphi_2 \in [-\pi/2, \pi/2])$, with $\theta = (\rho\cos\varphi_1,\ \rho\sin\varphi_1\cos\varphi_2,\ \rho\sin\varphi_1\sin\varphi_2)$. If we denote $\xi = \theta/\rho$, which depends only on (φ_1, φ_2), show that $\rho|\varphi_1, \varphi_2 \sim \mathcal{N}(x \cdot \xi, 1)$ and then integration of ρ then leads to

$$\pi(\varphi_1, \varphi_2|x) \propto \exp\{(x \cdot \xi)^2/2\} \sin(\varphi_1),$$

where $x \cdot \xi = x_1 \cos(\varphi_1) + x_2 \sin(\varphi_1) \cos(\varphi_2) + x_3 \sin(\varphi_1) \sin(\varphi_2)$.

(c) Show how to simulate from $\pi(\varphi_1, \varphi_2|x)$ using an Accept–Reject algorithm with instrumental function $\sin(\varphi_1) \exp\{||x||^2/2\}$.

(d) For $p = 3$ and $x = (0.1, 1.2, -0.7)$, demonstrate the convergence of the algorithm. Make plots of the iterations of the integral and its standard error.

3.9 For the situation of Example 3.10, recreate Figure 3.4 using the following simulation strategies with a sample size of 10,000 points:

(a) For each value of λ, simulate a sample from the $\mathcal{E}xp(1/\lambda)$ distribution and a separate sample from the log-normal $\mathcal{LN}(0, 2\log\lambda)$ distribution. Plot the resulting risk functions.

(b) For each value of λ, simulate a sample from the $\mathcal{E}xp(1/\lambda)$ distribution and then transform it into a sample from the $\mathcal{LN}(0, 2\log\lambda)$ distribution. Plot the resulting risk functions.

(c) Simulate a sample from the $\mathcal{E}xp(1)$ distribution. For each value of λ, transform it into a sample from $\mathcal{E}xp(1/\lambda)$, and then transform it into a sample from the $\mathcal{LN}(0, 2\log\lambda)$ distribution. Plot the resulting risk functions.

(d) Compare and comment on the accuracy of the plots.

3.10 Compare (in a simulation experiment) the performances of the regular Monte Carlo estimator of

$$\int_1^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = \Phi(2) - \Phi(1)$$

with those of an estimator based on an optimal choice of instrumental distribution (see (3.11)).

3.11 In the setup of Example 3.10, give the two first moments of the log-normal distribution $\mathcal{LN}(\mu, \sigma^2)$.

3.12 In the setup of Example 3.13, examine whether or not the different estimators of the expectations $\mathbb{E}_f[h_i(X)]$ have finite variances.

3.13 Establish the equality (3.18) using the representation $b = \beta a/\alpha$.

3.14 (Ó Ruanaidh and Fitzgerald 1996) For simulating random variables from the density $f(x) = \exp\{-\sqrt{x}\}[\sin(x)]^2$, $0 < x < \infty$, compare the following choices of instrumental densities:

$$\begin{aligned} g_1(x) &= \tfrac{1}{2} e^{-|x|}, & g_2(x) &= \tfrac{1}{2\sqrt{2}} \operatorname{sech}^2(x/\sqrt{2}), \\ g_3(x) &= \tfrac{1}{2\pi} \tfrac{1}{1+x^2/4}, & g_4(x) &= \tfrac{1}{\sqrt{2\pi}} e^{-x/2} . \end{aligned}$$

(a) For $M = 100, 1000$, and $10,000$, compare the standard deviations of the estimates based on simulating M random variables.

(b) For each of the instrumental densities, estimate the size of M needed to obtain three digits of accuracy in estimating $\mathbb{E}_f X$.

3.15 Use the techniques of Example 3.11 to redo Problem 3.3. Compare the number of variables needed to obtain three digits of accuracy with importance sampling to the answers obtained from Problem 3.3.

3.16 Referring to Example 3.11:

(a) Show that to simulate $Y \sim \mathcal{TE}(a, 1)$, an exponential distribution left truncated at a, we can simulate $X \sim \mathcal{E}(1)$ and take $Y = a + X$.
(b) Use this method to calculate the probability that a χ^2_3 random variable is greater that 25, and that a t_5 random variable is greater than 50.
(c) Explore the gain in efficiency from this method. Take $a = 4.5$ in part (a) and run an experiment to determine how many random variables would be needed to calculate $P(Z > 4.5)$ to the same accuracy obtained from using 100 random variables in an importance sampler.

3.17 In this chapter, the importance sampling method is developed for an iid sample $(Y_1, \ldots, Y_n)$ from g.
(a) Show that the importance sampling estimator is still unbiased if the Y_i's are correlated while being marginally distributed from g.
(b) Show that the importance sampling estimator can be extended to the case when Y_i is generated from a conditional distribution $q(y_i|Y_{i-1})$.
(c) Implement a scheme based on an iid sample $(Y_1, Y_3, \ldots, Y_{2n-1})$ and a secondary sample $(Y_2, Y_4, \ldots, Y_{2n})$ such that $Y_{2i} \sim q(y_{2i}|Y_{2i-1})$. Show that the covariance

$$\operatorname{cov}\left(h(Y_{2i-1})\frac{f(Y_{2i-1})}{g(Y_{2i-1})}, h(Y_{2i})\frac{f(Y_{2i})}{q(Y_{2i}|Y_{2i-1})}\right)$$

is null. Generalize.

3.18 For a sample $(Y_1, \ldots, Y_h)$ from g, the weights ω_i are defined as

$$\omega_i = \frac{f(Y_i)/g(Y_i)}{\sum_{j=1}^h f(Y_j)/g(Y_j)} .$$

Show that the following algorithm (Rubin 1987) produces a sample from f such that the empirical average

$$\frac{1}{M}\sum_{m=1}^{M} h(X_m)$$

is asymptotically equivalent to the importance sampling estimator based on $(Y_1, \ldots, Y_N)$:

`For` $m = 1, \ldots, M$`,`
`take` $X_m = Y_i$ `with probability` ω_i

(*Note:* This is the SIR algorithm.)

3.19 (Smith and Gelfand 1992) Show that, when evaluating an integral based on a posterior distribution

$$\pi(\theta|x) \propto \pi(\theta)\ell(\theta|x),$$

where π is the prior distribution and ℓ the likelihood function, the prior distribution can always be used as instrumental distribution (see Problem 2.29).
(a) Show that the variance is finite when the likelihood is bounded.
(b) Compare with choosing $\ell(\theta|x)$ as instrumental distribution when the likelihood is proportional to a density. (*Hint*: Consider the case of exponential families.)
(c) Discuss the drawbacks of this (these) choice(s) in specific settings.
(d) Show that a mixture between both instrumental distributions can ease some of the drawbacks.

3.20 In the setting of Example 3.13, show that the variance of the importance sampling estimator associated with an importance function g and the integrand $h(x) = \sqrt{x/(1-x)}$ is infinite for all g's such that $g(1) < \infty$.

3.21 Monte Carlo marginalization is a technique for calculating a marginal density when simulating from a joint density. Let $(X_i, Y_i) \sim f_{XY}(x, y)$, independent, and the corresponding marginal distribution $f_X(x) = \int f_{XY}(x, y)dy$.

(a) Let $w(x)$ be an arbitrary density. Show that

$$\lim_n \frac{1}{n} \sum_{i=1}^n \frac{f_{XY}(x^*, y_i)w(x_i)}{f_{XY}(x_i, y_i)} = \int\int \frac{f_{XY}(x^*, y)w(x)}{f_{XY}(x, y)} f_{XY}(x, y)dxdy = f_X(x^*)$$

and so we have a Monte Carlo estimate of f_X, the marginal distribution of X, from only knowing the form of the joint distribution.

(b) Let $X|Y = y \sim \mathcal{G}a(y, 1)$ and $Y \sim \mathcal{E}xp(1)$. Use the technique of part (a) to plot the marginal density of X. Compare it to the exact marginal.

(c) Choosing $w(x) = f_{X|Y}(x|y)$ works to produce the marginal distribution, and it is optimal. In the spirit of Theorem 3.12, can you prove this?

3.22 Given a real importance sample $X_1, \ldots, X_n$ with importance function g and target density f,

(a) show that the sum of the weights $\omega_i = f(X_i)/g(X_i)$ is only equal to n in expectation and deduce that the weights need to be renormalized even when both densities have know normalizing constants.

(b) Assuming that the weights ω_i have been renormalized to sum to one, we sample, with replacement, n points $\tilde{X}_j$ from the X_i's using those weights. Show that the $\tilde{X}_j$'s satisfy

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n h(\tilde{X}_j)\right] = \mathbb{E}\left[\sum_{i=1}^n \omega_i h(X_i)\right] .$$

(c) Deduce that, if the above formula is satisfied for $\omega_i = f(X_i)/g(X_i)$ instead, the empirical distribution associated with the $\tilde{X}_j$'s is unbiased.

3.23 (Evans and Swartz 1995) Devise and implement a simulation experiment to approximate the probability $P(Z \in (0, \infty)^6)$ when $Z \sim \mathcal{N}_6(0, \Sigma)$ and

$$\Sigma^{-1/2} = \text{diag}(0, 1, 2, 3, 4, 5) + e \cdot e^t,$$

with $e^t = (1, 1, 1, 1, 1, 1)$:

(a) when using the $\Sigma^{-1/2}$ transform of a $\mathcal{N}_6(0, I_6)$ random variables;

(b) when using the Choleski decomposition of Σ;

(c) when using a distribution restricted to $(0, \infty)^6$ and importance sampling.

3.24 Using the facts

$$\int y^3 e^{-cy^2/2} dy = \frac{-1}{2c}\left[y^2 + \frac{1}{c}\right] e^{-cy^2/2} ,$$

$$\int y^6 e^{-cy^2/2} dy = \frac{-1}{2c}\left[y^5 + \frac{5y^3}{2c} + \frac{15y}{4c}\right] e^{-cy^2/2} + 30\sqrt{\frac{\pi}{c^7}}\Phi(\sqrt{2c}y) ,$$

derive expressions similar to (3.22) for the second- and third-order approximations (see also Problem 5.6).

3.25 By evaluating the normal integral for the first order approximation from (3.21), establish (3.22).

3.26 Referring to Example 3.16, derive the Laplace approximation for the Gamma density and reproduce Table 3.6.

3.27 (Gelfand and Dey 1994) Consider a density function $f(x|\theta)$ and a prior distribution $\pi(\theta)$ such that the marginal $m(x) = \int_\Theta f(x|\theta)\pi(\theta)d\theta$ is finite a.e. The marginal density is of use in the comparison of models since it appears in the Bayes factor (see Section 1.3).

(a) Give a Laplace approximation of m and derive the corresponding approximation of the Bayes factor. (See Tierney et al. 1989 for details.)

(b) Give the general shape of an importance sampling approximation of m.

(c) Detail this approximation when the importance function is the posterior distribution and when the normalizing constant is unknown.

(d) Show that for a proper density τ,

$$m(x)^{-1} = \int_\Theta \frac{\tau(\theta)}{f(x|\theta)\pi(\theta)}\pi(\theta|x)d\theta \ ,$$

and deduce that when the θ_i^*'s are generated from the posterior,

$$\hat{m}(x) = \left\{ \frac{1}{T} \sum_{t=1}^{T} \frac{\tau(\theta_i^*)}{f(x|\theta_i^*)\pi(\theta_i^*)} \right\}^{-1}$$

is another importance sampling estimator of m.

3.28 (Berger et al. 1998) For Σ a $p \times p$ positive-definite symmetric matrix, consider the distribution

$$\pi(\theta) \propto \frac{\exp\left(-(\theta-\mu)^t \Sigma^{-1}(\theta-\mu)/2\right)}{||\theta||^{p-1}} \ .$$

(a) Show that the distribution is well defined; that is, that

$$\int_{\mathbb{R}^p} \frac{\exp\left(-(\theta-\mu)^t \Sigma^{-1}(\theta-\mu)/2\right)}{||\theta||^{p-1}} d\theta < \infty.$$

(b) Show that an importance sampling implementation based on the normal instrumental distribution $\mathcal{N}_p(\mu, \Sigma)$ is not satisfactory from both theoretical and practical points of view.

(c) Examine the alternative based on a Gamma distribution $\mathcal{G}a(\alpha, \beta)$ on $\eta = ||\theta||^2$ and a uniform distribution on the angles.

Note: Priors such as these have been used to derive *Bayes minimax estimators* of a multivariate normal mean. See Lehmann and Casella (1998).

3.29 From the Accept–Reject Algorithm we get a sequence $Y_1, Y_2, \ldots$ of independent random variables generated from g along with a corresponding sequence $U_1, U_2, \ldots$ of uniform random variables. For a fixed sample size t (i.e. for a fixed number of accepted random variables), the number of generated Y_i's is a random integer N.

(a) Show that the joint distribution of $(N, Y_1, \ldots, Y_N, U_1, \ldots, U_N)$ is given by

$$\begin{aligned}
&P(N = n, Y_1 \le y_1, \ldots, Y_n \le y_n, U_1 \le u_1, \ldots, U_n \le u_n) \\
&\quad = \int_{-\infty}^{y_n} g(t_n)(u_n \wedge w_n)dt_n \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_{n-1}} g(t_1) \ldots g(t_{n-1}) \\
&\quad \times \sum_{(i_1, \cdots, i_{t-1})} \prod_{j=1}^{t-1} (w_{i_j} \wedge u_{i_j}) \prod_{j=t}^{n-1} (u_{i_j} - w_{i_j})^+ dt_1 \cdots dt_{n-1},
\end{aligned}$$

where $w_i = f(y_i)/Mg(y_i)$ and the sum is over all subsets of $\{1, \ldots, n-1\}$ of size $t-1$.

(b) There is also interest in the joint distribution of $(Y_i, U_i)|N = n$, for any $i = 1, \ldots, n-1$, as we will see in Problem 4.17. Since this distribution is the same for each i, we can just derive it for (Y_1, U_1). (Recall that $Y_n \sim f$.) Show that

$$\begin{aligned} &P(N = n, Y_1 \le y, U_1 \le u_1) \\ &= \binom{n-1}{t-1}\left(\frac{1}{M}\right)^{t-1}\left(1-\frac{1}{M}\right)^{n-t-1} \\ &\times\left[\frac{t-1}{n-1}(w_1 \wedge u_1)\left(1-\frac{1}{M}\right)+\frac{n-t}{n-1}(u_1-w_1)^+\left(\frac{1}{M}\right)\right]\int_{-\infty}^{y} g(t_1)dt_1. \end{aligned}$$

(c) Show that part (b) yields the negative binomial marginal distribution of N,

$$P(N = n) = \binom{n-1}{t-1}\left(\frac{1}{M}\right)^{t}\left(1-\frac{1}{M}\right)^{n-t},$$

the marginal distribution of Y_1, $m(y)$,

$$m(y) = \frac{t-1}{n-1}f(y) + \frac{n-t}{n-1}\frac{g(y)-\rho f(y)}{1-\rho},$$

and

$$P(U_1 \le w(y)|Y_1 = y, N = n) = \frac{g(y)w(y)M\frac{t-1}{n-1}}{m(y)}.$$

3.30 If $(Y_1, \ldots, Y_N)$ is the sample produced by an Accept–Reject method based on (f, g), where $M = \sup(f/g)$, $(X_1, \ldots, X_t)$ denotes the accepted subsample and $(Z_1, \ldots, Z_{N-t})$ the rejected subsample.

(a) Show that both

$$\delta_2 = \frac{1}{N-t}\sum_{i=1}^{N-t} h(Z_i)\frac{(M-1)f(Z_i)}{Mg(Z_i)-f(Z_i)}$$

and

$$\delta_1 = \frac{1}{t}\sum_{i=1}^{t} h(X_i)$$

are unbiased estimators of $I = \mathbb{E}_f[h(X)]$ (when $N > t$).

(b) Show that δ_1 and δ_2 are independent.

(c) Determine the optimal weight β^* in $\delta_3 = \beta\delta_1 + (1-\beta)\delta_2$ in terms of variance. (*Note*: β may depend on N but not on $(Y_1, \ldots, Y_N)$.)

3.31 Given a sample $Z_1, \ldots, Z_{n+t}$ produced by an Accept–Reject algorithm to accept n values, based on (f, g, M), show that the distribution of a rejected variable is

$$\left(1-\frac{f(z)}{Mg(z)}\right)g(z) = \frac{g(z)-\rho f(z)}{1-\rho},$$

where $\rho = 1/M$, that the marginal distribution of Z_i $(i < n+t)$ is

$$f_m(z) = \frac{n-1}{n+t-1}f(z) + \frac{t}{n+t-1}\frac{g(z)-\rho f(z)}{1-\rho},$$

and that the joint distribution of (Z_i, Z_j) $(1 \le i \ne j < n+t)$ is

$$\begin{aligned}
&\frac{(n-1)(n-2)}{(n+t-1)(n+t-2)} f(z_i) f(z_j) \\
&\quad + \frac{(n-1)t}{(n+t-1)(n+t-2)} \left\{ f(z_i)\, \frac{g(z_j)-\rho f(z_j)}{1-\rho}\, \frac{g(z_i)-\rho f(z_i)}{1-\rho}\, f(z_j) \right\} \\
&\quad + \frac{n(n-1)}{(n+t-1)(n+t-2)}\, \frac{g(z_i)-\rho f(z_i)}{1-\rho}\, \frac{g(z_j)-\rho f(z_j)}{1-\rho}\,.
\end{aligned}$$

3.32 (Continuation of Problem 3.31) If $Z_1, \ldots, Z_{n+t}$ is the sample produced by an Accept–Reject algorithm to generate n values, based on (f, g, M), show that the Z_i's are negatively correlated in the sense that for every square integrable function h,

$$\begin{aligned}
\mathrm{cov}(h(Y_i), h(Y_j)) &= -\mathbb{E}_g[h]^2 \mathbb{E}_N \left[\frac{(t-1)(n-t)}{(n-1)^2(n-2)} \right] \\
&= -\mathbb{E}_g[h]^2 \{ \rho^t {}_2F_1(t-1, t-1; t-1; 1-\rho) - \rho^2 \},
\end{aligned}$$

where ${}_2F_1(a, b; c; z)$ is the confluent hypergeometric function (see Abramowitz and Stegun 1964 or Problem 1.38).

3.33 Given an Accept–Reject algorithm based on (f, g, ρ), we denote by

$$b(y_j) = \frac{(1-\rho) f(y_j)}{g(y_j) - \rho f(y_j)}$$

the importance sampling weight of the rejected variables $(Y_1, \ldots, Y_t)$, and by $(X_1, \ldots, X_n)$ the accepted variables.

(a) Show that the estimator

$$\delta_1 = \frac{n}{n+t}\, \delta^{AR} + \frac{t}{n+t}\, \delta_0,$$

with

$$\delta_0 = \frac{1}{t} \sum_{j=1}^{t} b(Y_j) h(Y_j)$$

and

$$\delta^{AR} = \frac{1}{n} \sum_{i=1}^{n} h(X_i),$$

does not uniformly dominate δ^{AR}. (*Hint:* Consider the constant functions.)

(b) Show that

$$\delta_{2w} = \frac{n}{n+t}\, \delta^{AR} + \frac{t}{n+t} \sum_{j=1}^{t} b(Y_j) h(Y_j) \Big/ \sum_{j=1}^{t} b(Y_j)$$

is asymptotically equivalent to δ_1 in terms of bias and variance.

(c) Deduce that δ_{2w} asymptotically dominates δ^{AR} if (4.20) holds.

3.34 For the Accept–Reject algorithm of Section 3.3.3:

(a) Show that conditionally on t, the joint density of $(Y_1, \ldots, Y_t)$ is indeed

$$\prod_{j=1}^{t-1} \left(\frac{Mg(y_j) - f(y_j)}{M-1} \right) f(y_t)$$

and the expectation of δ_2 of (3.15) is given by

$$\mathbb{E}\left[\frac{t-1}{t} \left\{ \frac{M}{M-1} \mathbb{E}_f[h(X)] - \frac{1}{M-1} \mathbb{E}_f \left[h(X) \frac{f(X)}{g(X)} \right] \right\} \right.$$
$$\left. + \frac{1}{t} \mathbb{E}_f \left[h(X) \frac{f(X)}{g(X)} \right] \right] .$$

(b) If we denote the acceptance probability of the Accept–Reject algorithm by $\rho = 1/M$ and assume $\mathbb{E}_f[h(X)] = 0$, show that the *bias* of δ_2 is

$$\left(\frac{1}{1-\rho} \mathbb{E}[t^{-1}] - \frac{\rho}{1-\rho} \right) \mathbb{E}_f \left[h(X) \frac{f(X)}{g(X)} \right] .$$

(c) Establish that for $t \sim \mathcal{G}eo(\rho)$, $\mathbb{E}[t^{-1}] = -\rho \log(\rho)/(1-\rho)$, and that the bias of δ_2 can be written as

$$-\frac{\rho}{1-\rho} (1 + \log(\rho)) \mathbb{E}_f \left[h(X) \frac{f(X)}{g(X)} \right] .$$

(d) Assuming that $\mathbb{E}_f[h(X)] = 0$, show that the variance of δ_2 is

$$\mathbb{E}\left[\frac{t-1}{t^2} \right] \frac{1}{1-\rho} \mathbb{E}_f \left[h^2(X) \frac{f(X)}{g(X)} \right]$$
$$+ \mathbb{E}\left[\frac{1}{t^2} \left\{ 1 - \frac{\rho(t-1)}{1-\rho} \right\} \right] \mathrm{var}_f(h(X)f(X)/g(X)).$$

3.35 Using the information from Note 3.6.1, for a binomial experiment $X_n \sim \mathcal{B}(n,p)$ with $p = 10^{-6}$, determine the minimum sample size n so that

$$P\left(\left| \frac{X_n}{n} - p \right| \le \epsilon p \right) > .95$$

when $\epsilon = 10^{-1}, 10^{-2}$, and 10^{-3}.

3.36 When random variables Y_i are generated from (3.25), show that $J^{(m)}$ is distributed as $\lambda(\theta_0)^{-n} \exp(-n\theta J)$. Deduce that (3.26) is unbiased.

3.37 Starting with a density f of interest, we create the exponential family

$$\mathcal{F} = \{ f(\cdot|\tau); f(x|\tau) = \exp[\tau x - K(\tau)] f(x) \} ,$$

where $K(\tau)$ is the cumulant generating function of f given in Section 3.6.2. It immediately follows that if $X_1, X_2, \ldots, X_n$ are iid from $f(x|\tau)$, the density of $\bar{X}$ is

$$f_{\bar{X}}(x|\tau) = \exp\{ n[\tau x - K(\tau)] \} f_{\bar{X}}(x), \tag{3.24}$$

where $f_{\bar{X}}(x)$ is the density of the average of an iid sample from f

(a) Show that $f(x|\tau)$ is a density.

(b) If $X_1, X_2, \ldots, X_n$ is an iid sample from $f(x)$, and $f_{\bar{X}}(x)$ is the density of the sample mean, show that $\int e^{n[\tau x - K(\tau)]} f_{\bar{X}}(x) dx = 1$ and hence $f_{\bar{X}}(x|\tau)$ of (3.24) is a density.

(c) Show that the mgf of $f_{\bar{X}}(x|\tau)$ is $e^{n[K(\tau+t/n)x - K(\tau)]}$

(d) In (3.29), for each x we choose τ so that μ_τ, the mean of $f_{\bar{X}}(x|\tau)$, satisfies $\mu_\tau = x$. Show that this value of τ is the solution to the equation $K'(\tau) = x$.

3.38 For the situation of Example 3.18:

(a) Verify the mgf in (3.32).

(b) Show that the solution to the saddlepoint equation is given by (3.33).

(c) Plot the saddlepoint density for $p = 7$ and $n = 1, 5, 20$. Compare your results to the exact density.

3.6 Notes

3.6.1 Large Deviations Techniques

When we introduced importance sampling methods in Section 3.3, we showed in Example 3.8 that alternatives to direct sampling were preferable when sampling from the tails of a distribution f. When the event A is particularly rare, say $P(A) \le 10^{-6}$, methods like importance sampling are needed to get an acceptable approximation (see Problem 3.35). Since the optimal choice given in Theorem 3.12 is formal, in the sense that it involves the unknown constant I, more practical choices have been proposed in the literature. In particular, Bucklew (1990) indicates how the *theory of large deviations* may help in devising proposal distributions in this purpose.

Briefly, the theory of large deviations is concerned with the approximation of tail probabilities $P(|\bar{X}_n - \mu| > \varepsilon)$ when $\bar{X}_n = (X_1 + \cdots + X_n)/n$ is a mean of iid random variables, n goes to infinity, and ε is large. (When ε is small, the normal approximation based on the Central Limit Theorem works well enough.)

If $M(\theta) = \mathbb{E}[\exp(\theta X_1)]$ is the moment generating function of X_1 and we define $I(x) = \sup_\theta \{\theta x - \log M(\theta)\}$, the large deviation approximation is

$$\frac{1}{n} \log P(S_n \in F) \approx -\inf_F I(x).$$

This result is sometimes called *Cramér's Theorem* and a simulation device based on this result and called *twisted simulation* is as follows.

To evaluate

$$I = P\left(\frac{1}{n}\sum_{i=1}^n h(x_i) \ge 0\right),$$

when $\mathbb{E}[h(X_1)] < 0$, we use the proposal density

$$t(x) \propto f(x) \exp\{\theta_0 h(x)\}, \tag{3.25}$$

where the parameter θ_0 is chosen such that $\int h(x) f(x) e^{\theta_0 h(x)} dx = 0$. (Note the similarity with exponential tilting in saddlepoint approximations in Section 3.6.2.) The corresponding estimate of I is then based on blocks ($m = 1, \ldots, M$)

$$J^{(m)} = \frac{1}{n}\sum_{i=1}^n h(Y_i^{(m)}),$$