Appendix 1: Standard Families for Internal Distributions

This appendix contains useful formulas for internal sampling distribution specification for both exponential tilting and translation families. In each case we present the distribution and density functions for the internal variable (denoted by Z), the transformation $V = \tau(U)$, and the resulting weights as a function of V. In selected cases we present plots of the transformations, with parameter values chosen so that $\tau(k/10) = 1-k/10$, for $k = 1, 2, \ldots 9$.

These families are used in importance sampling by replacing the uniform random numbers U_i that are used to generate X with the values V_i, as described in section 6.4.

A1.1 Tilting Families

Exponential families discussed include the uniform, tilted uniform, normal, exponential, reverse exponential (power family), gamma, reverse gamma, and general discrete distributions, and for translation for the normal, cauchy, and logistic distributions. The tilted uniform family is equivalent to a truncated exponential distribution (or truncated reverse exponential distribution). The gamma, reverse gamma, and tilted uniform families require specification of an additional parameter that determines the shape of the base distribution.

General form

$$v = \tau(u) = F_0(F_\alpha^{-1}(u))$$

$$z = F_\alpha^{-1}(u) = F_0^{-1}(v)$$

$$W(v) = \frac{f_0(z)}{f_\alpha(z)} = \frac{1}{\tau'(\tau^{-1}(v))}$$

$$f_\alpha(z) = \frac{1}{\Psi(\alpha)} e^{\alpha z} f_0(z)$$

Uniform Distribution

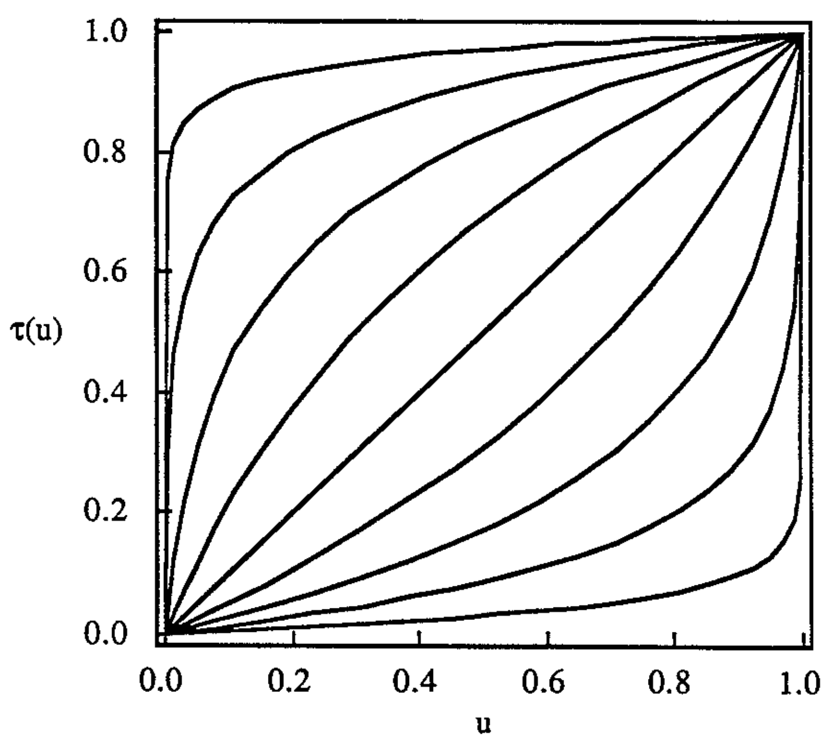
$0 < z < 1, -\infty < \alpha < \infty$

$$F_\alpha(z) = \begin{cases} z & \alpha = 0 \\ \dfrac{e^{\alpha z}-1}{e^{\alpha}-1} & \alpha \neq 0 \end{cases}$$

$$f_\alpha(z) = \begin{cases} 1 & \alpha = 0 \\ \dfrac{\alpha\, e^{\alpha z}}{e^{\alpha}-1} & \alpha \neq 0 \end{cases}$$

Figure A.1

Internal Exponential Tilting with Uniform Basis Distribution



$$v = \tau(u) = \frac{\log\left(1 + u(e^{\alpha} - 1)\right)}{\alpha}$$

$$W(v) = \frac{e^{\alpha} - 1}{\alpha} e^{-\alpha v}$$

Tilted Uniform Family (fixed shape parameter ζ)

For the case ζ=0 see the uniform distribution above.

$$0 < z < 1, -\infty < \alpha < \infty, -\infty < \zeta < \infty, \zeta \neq 0$$

$$F_{\alpha}(z) = \begin{cases} z & \alpha + \zeta = 0 \\ \dfrac{e^{(\alpha + \zeta)x}}{e^{\alpha + \zeta} - 1} & \alpha + \zeta \neq 0 \end{cases}$$

$$f_{\alpha}(z) = \begin{cases} 1 & \alpha + \zeta = 0 \\ \dfrac{(\alpha + \zeta)\, e^{(\alpha + \zeta)z}}{e^{\alpha + \zeta} - 1} & \alpha + \zeta \neq 0 \end{cases}$$

$$v = \tau(u) = \frac{\left(1+u(e^{\zeta+\alpha} - 1)\right)^{\frac{\zeta}{\zeta+\alpha}} - 1}{e^{\zeta} - 1}$$

$$W(v) = \frac{\zeta\left(e^{\zeta+\alpha} - 1\right)}{\left(\zeta + \alpha\right)\left(e^{\zeta} - 1\right)} \left(1 + v(e^{\zeta} - 1)\right)^{-\alpha/\zeta}$$

Gaussian Distribution

$$-\infty < \alpha < \infty, -\infty < z < \infty$$

$$F_\alpha(z) = \Phi(z - \alpha)$$

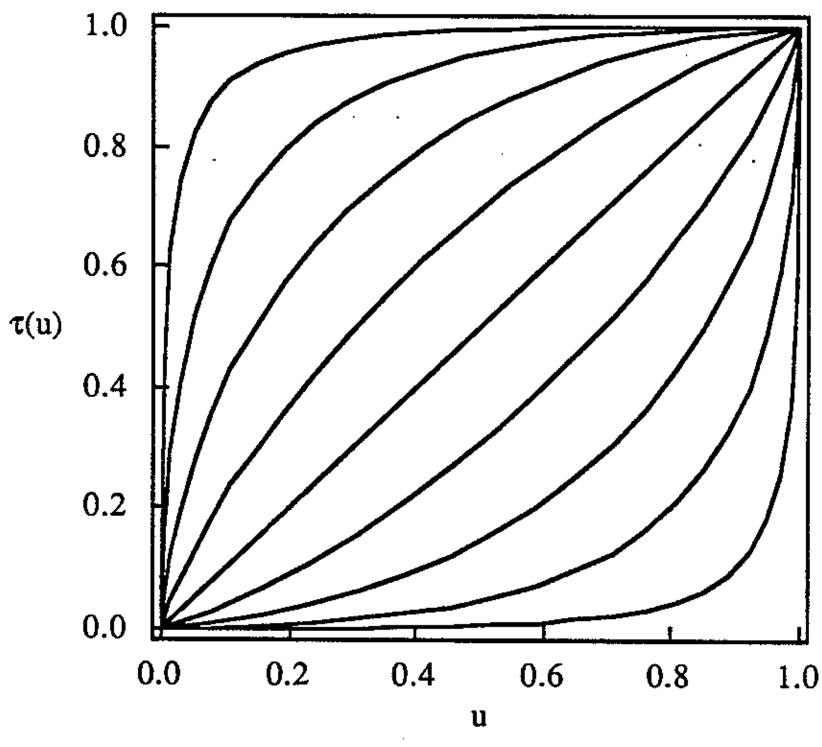
$$f_\alpha(z) = \phi(z - \alpha) = \frac{1}{\sqrt{2\pi}} e^{-(z-\alpha)^2/2}$$

$$v = \tau(u) = \Phi\left(\Phi^{-1}(u) + \alpha\right)$$

$$W(v) = e^{-\alpha\Phi^{-1}(v) + \alpha^2/2}$$

Figure A.2

Internal Exponential Tilting with Gaussian Basis Distribution



Exponential Distribution

$\alpha < 1, z > 0$

$F_\alpha(z) = 1 - e^{(\alpha-1)z}$

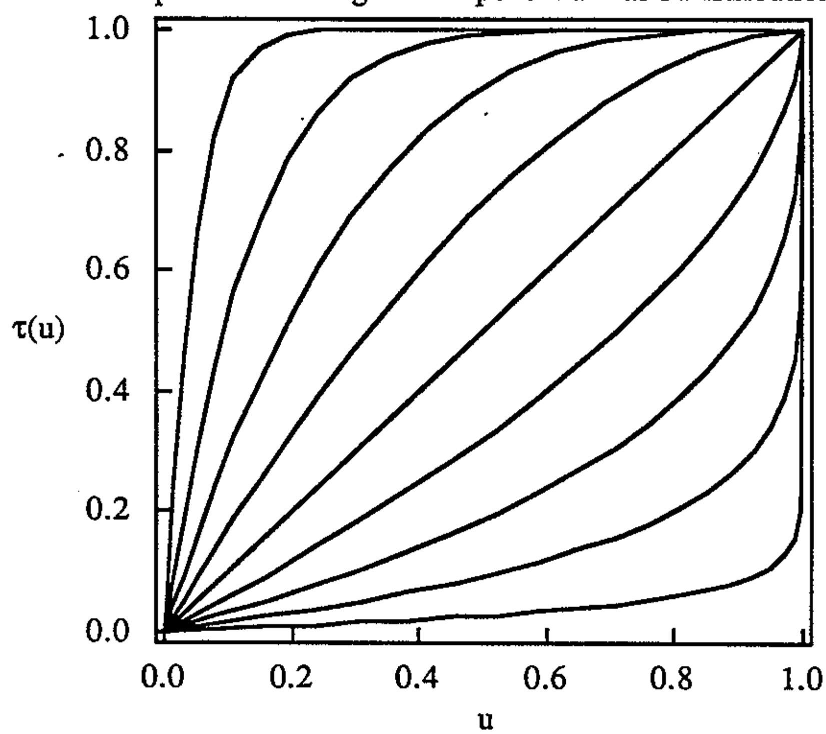
$f_\alpha(z) = (1 - \alpha)\, e^{(\alpha-1)z}$

$v = \tau(u) = 1 - (1 - u)^{1/(1-\alpha)}$

$W(v) = \dfrac{1}{1-\alpha}(1 - v)^{\alpha}$

Figure A.3

Internal Exponential Tilting with Exponential Basis Distribution



Reverse Exponential Distribution (Power Family)

$\alpha > -1,\ z < 0$

$F_\alpha(z) = e^{(\alpha+1)z}$

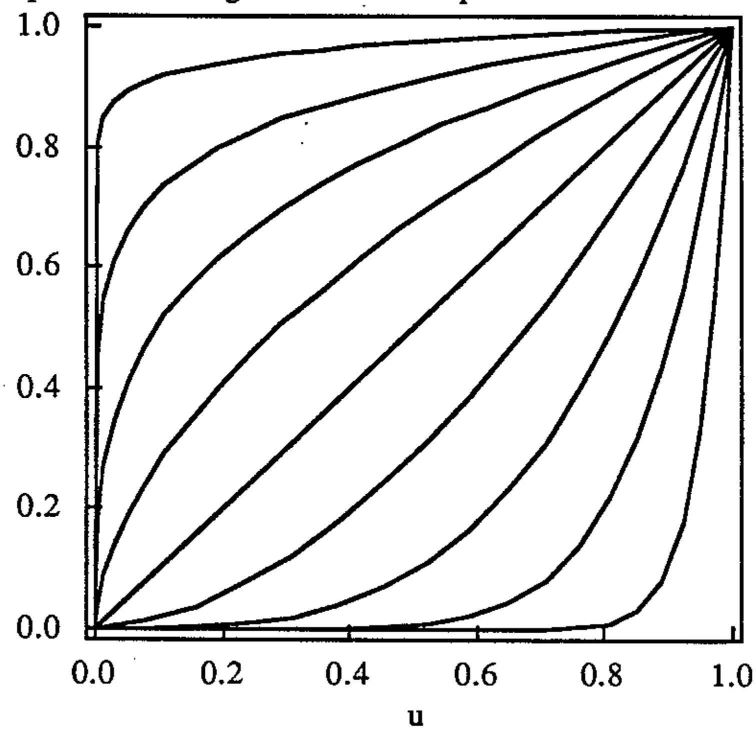
$f_\alpha(z) = (1 + \alpha)\, e^{(1+\alpha)z}$

$v = \tau(u) = u^{1/(1+\alpha)}$

$$W(v) = \frac{1}{1+\alpha} v^{\alpha}$$

Figure A.4

Internal Exponential Tilting with Reverse Exponential Basis Distribution



Gamma Family (fixed shape parameter ζ)

This is parameterized so $\alpha=0$ is the reference case. $\zeta = 1$ corresponds to the exponential distribution. Tilting is done by modifying the scale parameter.

$$0 < z < \infty, -\infty < \alpha < 1, 0 < \zeta < \infty$$

$$F_\alpha(z) = \Gamma\left(\zeta, (1-\alpha)z\right)$$

$$f_\alpha(z) = \frac{(1-\alpha)^\zeta}{\Gamma(\zeta)} z^{\zeta-1} e^{(\alpha-1)z}$$

$$v = \tau(u) = \Gamma(\zeta, z) = \Gamma\left(\zeta, \frac{1}{1-\alpha} \Gamma^{-1}(\zeta, u)\right)$$

$$z = \Gamma^{-1}(\zeta, v) = \frac{1}{1-\alpha} \Gamma^{-1}(\zeta, u)$$

$$W(v) = (1-\alpha)^{-\zeta} e^{-\alpha z}$$

Reverse Gamma Family (fixed shape parameter ζ)

This is parameterized so $\alpha=0$ is the reference case. $\zeta = 1$ corresponds to the exponential distribution. Tilting is done by modifying the scale parameter.

$$-\infty < z < 0, -1 < \alpha < \infty, 0 < \zeta < \infty$$

$$F_\alpha(z) = 1 - \Gamma\left(\zeta, -(1+\alpha)z\right)$$

$$f_\alpha(z) = \frac{(1+\alpha)^\zeta}{\Gamma(\zeta)} (-z)^{\zeta-1} e^{(\alpha+1)z}$$

$$v = \tau(u) = 1 - \Gamma(\zeta, -z)$$

$$z = -\Gamma^{-1}(\zeta, 1-v) = -\frac{1}{1+\alpha} \Gamma^{-1}(\zeta, 1-u)$$

$$W(v) = (1+\alpha)^{-\zeta} e^{-\alpha z}$$

Discrete Distribution

$$P_\alpha\{ Z = z_j \} = \frac{f_{0,j} e^{\alpha z_j}}{\sum f_{0,k} e^{\alpha z_k}}$$

$$F^{-1}(u) := \min_x \{x : F(x) \geq u\}$$

$$v = \tau(u) = F_0(F_\alpha^{-1}(u))$$

$$z = F_\alpha^{-1}(u) = F_0^{-1}(v)$$

$$W(v) = e^{-\alpha z_j} \sum f_{0,k} e^{\alpha z_l}$$

A1.2 Translation Families

General form:

$$v = \tau(u) = F(F^{-1}(u) + \alpha)$$

$$z = F^{-1}(u) + \alpha = F^{-1}(v)$$

$$W(v) = \frac{f(z)}{f(z-\alpha)} = \frac{1}{\tau'(\tau^{-1}(v))}$$

Normal Translation (same as normal tilting)

$$-\infty < \alpha < \infty, -\infty < x < \infty$$

$$F(x) = \Phi(x)$$

$$v = \tau(u) = \Phi\left(\Phi^{-1}(u) + \alpha\right)$$

$$W(v) = e^{-\alpha\Phi^{-1}(v) + \alpha^2/2}$$

Cauchy Translation

$$-\infty < \alpha < \infty, -\infty < x < \infty$$

$$F(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$$

$$v = \tau(u) = \frac{1}{2} + \frac{\arctan(z)}{\pi}$$

$$z = \tan(\pi\,(v - 1/2)) = \tan(\pi\,(u - 1/2)) + \alpha$$

$$W(v) = 1 - \frac{2\,z - \alpha}{1 + z^2}$$

Logistic Translation

$$-\infty < \alpha < \infty, -\infty < x < \infty$$

$$F(x) = \frac{1}{1 + e^{-x}}$$

$$v = \tau(u) = \frac{u\,e^{\alpha}}{1 + u\,(e^{\alpha} - 1)}$$

$$W(v) = \frac{e^{-\alpha}}{v + (1 - v)e^{\alpha}}$$

Appendix 2: One-Pass Formulas for Estimates and Variances

This appendix contains one-pass formulas for computing estimates and their standard errors. Let

(A2.1) $$S_Y = \sum_{i=1}^{n} Y_i$$

(A2.2) $$S_W = \sum_{i=1}^{n} W_i$$

(A2.3) $$\bar{Y} = \bar{Y}_n = \frac{S_Y}{n}$$

(A2.4) $$\bar{W} = \bar{W}_n = \frac{S_W}{n}$$

(A2.5) $$SS_{YY} = SS_{YY,n} = \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2$$

(A2.6) $$SS_{YW} = SS_{YW,n} = \sum_{i=1}^{n} (Y_i - \bar{Y}_n)(W_i - \bar{W}_n)$$

(A2.7) $$SS_{WW} = SS_{WW,n} = \sum_{i=1}^{n} (W_i - \bar{W}_n)^2$$

(A2.8) $$SS_c = \sum_{i=1}^{n} (Y_i - cW_i - \bar{Y} + c\bar{W})^2$$

Formulas (A2.5)-(A2.8) can be computed iteratively using the updating formulas

(A2.9) $$SS_{YY,n+1} = SS_{YY,n} + \frac{n}{n+1}(Y_{n+1} - \bar{Y}_n)^2$$

(A2.10) $$SS_{YW,n+1} = SS_{YW,n} + \frac{n}{n+1}(Y_{n+1} - \bar{Y}_n)(W_{n+1} - \bar{W}_n)$$

(A2.11) $$SS_{WW,n+1} = SS_{WW,n} + \frac{n}{n+1}(W_{n+1} - \bar{W}_n)^2$$

(A2.12) $$SS_{c,n+1} = SS_{c,n} + \frac{n}{n+1}(Y_{n+1} - cW_{n+1} - \overline{Y}_n + c\overline{W}_n)^2$$

where in every case $SS_{**,1} = 0$.

In addition, the change formula for SS_c is

(A2.13) $$SS_d = SS_c + 2(c-d)SS_{YW} + (d^2 - c^2)SS_{WW}$$

The integration estimate and its variance are:

$$\hat{\mu}_{int} = \overline{Y}$$

$$\text{var}(\hat{\mu}_{int}) = \text{var}(\overline{Y}) = \sigma^2_Y / n$$

Two estimates of the variance are:

$$\hat{\text{var}}_1(\hat{\mu}_{int}) = \frac{SS_{YY}}{n(n-1)}$$

$$\hat{\text{var}}_2(\hat{\mu}_{int}) = \frac{SS_{YY}}{n(n-1)} / \overline{W}^2$$

both of which can be computed using the updating formulas. The latter estimate corresponds to correcting the estimate based on the average value of W, as described in the discussion of confidence intervals in chapter two.

The ratio estimate and the first term in its asymptotic variance expansion are:

$$\hat{\mu}_{ratio} = \overline{Y} / \overline{W}$$

$$\text{var}(\hat{\mu}_{ratio}) \approx \text{var}(\overline{Y} - \mu\,\overline{W}) = \frac{1}{n}(\sigma^2_Y - 2\,\mu\,\sigma_{XY} + \mu^2\,\sigma^2_W)$$

Two estimates of the variance are:

$$\hat{\text{var}}_1(\hat{\mu}_{ratio}) = \frac{SS_{YY} - 2\,\hat{\mu}_{ratio}\,SS_{YW} + \hat{\mu}^2_{ratio}\,SS_{WW}}{n(n-1)}$$

$$= \frac{SS_c}{n(n-1)}$$

$$\hat{\text{var}}_2(\hat{\mu}_{ratio}) = \frac{SS_c}{n(n-1)\bar{W}^2}$$

where $c = \hat{\mu}_{ratio}$. If the updating formula (A2.12) is used it is necessary also to use the change formula (A2.13) with $c = \hat{\mu}_{ratio,n}$ and $d = \hat{\mu}_{ratio,n+1}$. The use of SS_c is more stable numerically, though this level of accuracy is rarely needed. The change formula need not be used every replication.

The regression estimate and the first term in its asymptotic variance expansion are:

$$\hat{\mu}_{reg} = \bar{Y} + \hat{\beta}\,(1 - \bar{W})$$

$$\text{var}(\hat{\mu}_{reg}) \approx \text{var}(\bar{Y} - \beta\,\bar{W}) = \frac{1}{n}(\sigma^2_Y - 2\,\beta\,\sigma_{XY} + \beta^2\,\sigma^2_W)$$

Two estimates of the variance are:

$$\hat{\text{var}}_1(\hat{\mu}_{reg}) = \frac{SS_{YY} - 2\,\hat{\beta}\,SS_{YW} + \hat{\beta}\,SS_{WW}}{n(n-1)}$$

$$= \frac{SS_c}{n(n-1)}$$

$$\hat{\text{var}}_2(\hat{\mu}_{reg}) = \frac{SS_c}{n(n-1)\bar{W}^2}$$

where $c = \hat{\beta}$. If the updating formula (A2.12) is used it is necessary also to use the change formula (A2.13) with $c = \hat{\beta}_n$ and $d = \hat{\beta}_{n+1}$. The use of SS_c is more stable numerically, though this level of accuracy is rarely needed. The change formula need not be used every replication.

Appendix 3. Proofs for Chapter Two

A3.1 Equivalence of Exponential and Iterated Regression Estimates

Theorem Equivalence of exponential weights and iterated linear regression weights. Suppose that $\min(W_i) < 1 < \max(W_i)$, let $\pi_i^{(0)} = \frac{1}{n}$, let $\bar{W}^{(k)} = \bar{W} + \frac{k}{K}(1-\bar{W})$, and

let

(A3.1) $$\pi_i^{(k+1)} = \pi_i^{(k)} \left(1 + \frac{(W_i - \bar{W}^{(k)})(\bar{W}^{(k+1)} - \bar{W}^{(k)})}{\hat{\sigma}^{(k)}}\right)$$

as in (2.35), i = 1, 2, ... n. Then as $K \rightarrow \infty$ the iterated weights

(A3.2) $$\pi_i^{(K)} = \frac{1}{n} \prod_{k=0}^{K-1} \left(1 + \frac{(W_i - \bar{W}^{(k)})(\bar{W}^{(k+1)} - \bar{W}^{(k)})}{\hat{\sigma}^{(k)}}\right)$$

converge to the exponential weights

(A3.3) $$\pi_{exp,i} = a\, e^{b\, W_i}$$

where a and b solve $\sum \pi_{exp,i} = \sum \pi_{exp,i} W_i = 0$. $\hat{\sigma}$ is the standard deviation of the distribution with weight $\pi_i^{(k)}$ on point W_i.

Proof Define the "iteration conditions"

$$\sum_{i=1}^{n} \pi_i^{(k)} = 1,$$

$$\sum_{i=1}^{n} \pi_i^{(k)} W_i = \bar{W}^{(k)}$$

and the "nonnegativity condition"

$$\pi_i^{(k)} > 0 \text{ for } 1 \le i \le n.$$

$\pi^{(0)}$ satisfies all conditions. Note that if $\pi^{(k)}$ satisfies the nonnegativity condition then $\hat{\sigma}^{(k)} > 0$, and $\pi^{(k+1)}$ satisfies the iteration conditions as well, though not necessarily the nonnegativity condition.

Without loss of generality assume that $\bar{W} > 1$ and let m denote the index of the minimum values of W, $W_m = \min(W_i)$, and assume that $\pi^{(l)}$ satisfies the nonnegativity condition for $0 \le l \le k$; then $\pi_m^{(k)} \ge 1/n$, and since $\bar{W}^{(k)} = 1$, $\hat{\sigma}^{(k)} > (1-W_m)/\sqrt{n}$. Then if $|\bar{W}^{(k+1)} - \bar{W}^{(k)}| \frac{\max(W_i) - 1}{(1-W_m)/\sqrt{n}} < 1$ $\pi^{(k+1)}$ also satisfies the nonnegativity condition. By iteration $\pi^{(k)}$ satisfies the nonnegativity condition for all k, as long as $K > \sqrt{n}\,|\bar{W} - 1|\frac{\max(W_i) - 1}{\min(W_i) - 1}$. Similarly, if $\bar{W} < 1$ $\pi^{(k)}$ is nonnegative if $K > \sqrt{n}\,|\bar{W} - 1|\frac{\min(W_i) - 1}{\max(W_i) - 1}$.

Now for K sufficiently large we have $\hat{\sigma}^{(k)} > (1-W_m)/\sqrt{n}$ and both the nonnegativity and iteration conditions for $\pi^{(k)}$, $0 \le k \le K$. Let $\Delta := 1 - \bar{W}$, and the log of the product weights (A3.2) is:

$$\log(\pi_i^{(K)}) = -\log(n) + \sum_{k=0}^{K-1} \log\left(1 + \frac{(W_i - \bar{W}^{(k)})\,\Delta}{K\,\hat{\sigma}^{(k)}}\right)$$

$$= -\log(n) + \frac{\Delta}{K}\sum_{k=0}^{K-1} \frac{(W_i - \bar{W}^{(k)})}{\hat{\sigma}^{(k)}} + o(K^{-1})$$

$$= -\log(n) + W_i\ C_1(K) + C_2(K) + o(K^{-1}),$$

by a Taylor series expansion and inequalities on the log function. The $o(K^{-1})$ term disappears as K goes to infinity, and what remains is linear in W. Furthermore, the weights satisfy the iteration conditions at every step, and $C_1(K)$ determines the ratio $\Sigma\, \pi_i^{(K)} W_i / \Sigma \pi_i^{(K)}$, so $C_1(K)$ converges and hence $C_2(K)$ does as well. The result is that

$$\lim_{K \to \infty} \pi_i^{(K)} = \frac{1}{n} e^{W_i C_1 + C_2}$$

$$= a\, e^{b W_i}$$

for some a and b chosen to satisfy the iteration conditions, and so the iterated regression weights match the exponential weights.

A3.2 No Influence Function Paradox for the Regression Estimate

Proposition It is not possible to achieve a perfect negative association ($\theta < \mu \Leftrightarrow IF > 0$) between θ and the influence function for the regression estimate, given by:

$$IF_{reg}(X) = W\,(\,\theta - \beta\,) + \beta - \mu$$

when g dominates f and both W and Y have finite second moments.

Proof: The proof is by contradiction. Assume that $\theta < \mu \Leftrightarrow Y - \beta\,(W - 1) > \mu$, where β is the least-squares regression slope for Y on W. W.l.o.g. $\mu = 0$ (since the regression estimate is equivariant). β can not be zero, since then $\theta < 0 \Rightarrow Y - 0\,(W - 1) < 0$. Suppose β is positive, then $\theta < 0 \Rightarrow W < \beta/(\beta{-}\theta)$ and $0 < \theta < \beta \Rightarrow W > \beta/(\beta{-}\theta)$. $\theta \geq \beta \Rightarrow Y - \beta\,(W - 1) > 0$, which is a contradiction, so $\theta < \beta$.

Now we have $\theta < 0 \Rightarrow W < 1$ and $Y > \beta\,(W{-}1)$, and $\theta > 0 \Rightarrow W > 1$ and $Y < \beta\,(W{-}1)$. That is, Y is above the line $y = \beta\,(W{-}1)$ whenever $W < 1$ and Y is below the line whenever $W > 1$. But that (together with the fact that the expected value of W is 1) implies that the regression slope of Y on W is less than β, which is a contradiction. The case where β is negative is similar.

A3.3 Asymptotic Normality and Moments for Edgeworth Expansions

The asymptotic normality and variances and moments for the Edgeworth expansions for the distribution of the ratio and regression estimates are obtained by means of a delta-method estimate of the moments of these estimates. Let

$$C_i = Y_i - \mu,$$

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i$$

$$D_i = W_i - 1$$

$$\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i$$

A delta-method expansion for the ratio estimate is

$$\hat{\mu}_{ratio} = \frac{\bar{Y}}{\bar{W}}$$

$$= \mu \frac{1 + \bar{C} / \mu}{1 + \bar{D}}$$

$$= \mu \, (1 + \bar{C} / \mu) \, (1 - \bar{D} + \bar{D}^2 + \dots)$$

$$= \mu + \bar{C} - \mu\bar{D} - \bar{C}\bar{D} + \mu\bar{D}^2 + \dots$$

At each point terms of low order are dropped. The final expression is used in estimating moments of the distribution of the estimate, again dropping low-order terms

$$E[\hat{\mu}_{ratio}] \approx \mu - E[\,\bar{C}\bar{D} - \mu\bar{D}^2\,]$$

$$= \mu - \frac{1}{n}E[\,CD - \mu D^2\,]$$

$$= \mu - \frac{E[(W-1)(Y-\mu W)]}{n}$$

$$\mathrm{var}(\hat{\mu}_{ratio}) \approx \mathrm{var}(\bar{C} - \mu\bar{D})$$

$$= \frac{1}{n}\mathrm{var}(C - \mu D)$$

$$= \frac{\mathrm{var}(Y - \mu W)}{n}$$

$$E[\,(\hat{\mu}_{ratio}-\mu)^3\,] \approx E[\,(\bar{C} - \mu\bar{D} - \bar{C}\bar{D} + \mu\bar{D}^2\,)^3\,]$$

$$= E[\,(\bar{C} - \mu\bar{D})^3 - 3\bar{D}(\bar{C} - \mu\bar{D})^3 + \ldots\,]$$

$$= \frac{1}{n^2}E[\,(C - \mu D)^3\,] - \frac{9}{n^2}E[\,D(C - \mu D)\,]\,E[\,(C - \mu D)^2\,] + \ldots$$

$$= \frac{1}{n^2}E[\,(Y-\mu W)^3\,] - \frac{9}{n^2}E[\,(W-1)(Y-\mu W)\,]\,E[\,(Y-\mu W)^2\,]$$

$$E[\,(\hat{\mu}_{ratio} - E[\hat{\mu}_{ratio}])^3\,]$$

$$\approx E[\,(\hat{\mu}_{ratio}-\mu)^3\,] - 3\,\mathrm{var}(\hat{\mu}_{ratio})\,E[\hat{\mu}_{ratio} - \mu] + \ldots$$

$$= \frac{1}{n^2}E[\,(Y-\mu W)^3\,] - \frac{6}{n^2}E[\,(W-1)(Y-\mu W)\,]\,E[\,(Y-\mu W)^2\,]$$

A delta-method expansion for the regression estimate is similar, and uses the additional terms

$$E_i = C_i D_i - \sigma_{12}$$

$$\bar{E} = \frac{1}{n} \sum_{i=1}^{n} E_i$$

$$F_i = D_i^2 - \sigma_w^2$$

$$\bar{F} = \frac{1}{n} \sum_{i=1}^{n} F_i$$

$$\sigma_{12} = E[\, C\, D\,] = \mathrm{cov}(\, Y, W\,)$$

also note that $\beta = \sigma_{12} / \sigma_w^2$, that $E - \beta F = D\,(C - \beta D)$, and that $E[\, D\,(C - \beta D)\,] = 0$. The expansion is:

$$\hat{\mu}_{reg} = \bar{Y} - \hat{\beta}\,(\bar{W} - 1)$$

$$= \mu + \bar{C} - \frac{\sigma_{12} + \bar{E} - \bar{C}\,\bar{D}}{\sigma_w^2 + \bar{F} - \bar{D}^2} \bar{D}$$

$$= \mu + \bar{C} - \beta\,\bar{D}\,(1 + \frac{\bar{E}}{\sigma_{12}})\,(1 - \frac{\bar{F}}{\sigma_w^2}) + \ldots$$

$$= \mu + \bar{C} - \beta\,\bar{D} - \bar{D}\,(\bar{E} - \beta\,\bar{F}) / \sigma_w^2 + \ldots$$

At each point terms of low order are dropped. The final expression is used in estimating moments of the distribution of the estimate, again dropping low-order terms

$$E[\hat{\mu}_{reg}] \approx \mu - E[\bar{D}(\bar{E} - \beta\bar{F})] / \sigma_w^2$$

$$= \mu - \frac{1}{n\,\sigma_w^2} E[D(E - \beta F)]$$

$$= \mu - \frac{1}{n\,\sigma_w^2} E[D^2(C - \beta D)]$$

$$= \mu - \frac{1}{n\,\sigma_w^2} E[(W-1)^2 (Y - \beta W - \mu + \beta)]$$

$$\mathrm{var}(\hat{\mu}_{reg}) \approx \mathrm{var}(\bar{C} - \beta\bar{D})$$

$$= \frac{1}{n}\mathrm{var}(C - \beta D)$$

$$= \frac{\mathrm{var}(Y - \beta W)}{n}$$

$$E[(\hat{\mu}_{reg} - \mu)^3] \approx E[(\bar{C} - \beta\bar{D} - \bar{D}(\bar{E} - \beta\bar{F})/\sigma_w^2)^3]$$

$$= E[(\bar{C} - \beta\bar{D})^3 - 3(\bar{C} - \beta\bar{D})^2 \bar{D}(\bar{E} - \beta\bar{F})/\sigma_w^2] + \ldots$$

$$= \frac{1}{n^2}E[(C - \beta D)^3] - \frac{3}{n^2}E[(C - \beta D)^2]\,E[D(E - \beta F)]$$

$$- \frac{6}{n^2}E[D(C - \beta D)]\,E[(C - \beta D)(E - \beta F)] + \ldots$$

$$= \frac{1}{n^2}E[(C - \beta D)^3] - \frac{3}{n^2}E[(C - \beta D)^2]\,E[D(C - \beta D)]$$

$$E[\,(\hat{\mu}_{reg} - E[\hat{\mu}_{reg}])^3\,]$$

$$\approx E[\,(\hat{\mu}_{reg} - \mu)^3\,] - 3\,\mathrm{var}(\hat{\mu}_{reg})\,E[\hat{\mu}_{reg} - \mu] + \ldots$$

$$= \frac{1}{n^2} E[\,(C - \beta D)^3\,]$$

$$= \frac{1}{n^2} E[\,(Y - \beta W - \mu + \beta)^3\,]$$

To show the asymptotic normality of these estimates they are approximated by a second-order delta-method expansion,

$$\hat{\mu}_{ratio} \approx \mu + (\bar{Y} - \mu\bar{W})$$

$$\hat{\mu}_{reg} \approx \mu + (\bar{Y} - \beta\bar{W} - \mu + \beta)$$

The first term is fixed, and the second term is asymptotically normal by the central limit theorem, as long as $\bar{Y}$ and $\bar{W}$ have finite second moments. If the difference between these approximations and the estimates are $o_P(1/\sqrt{n})$ the estimate is asymptotically normal with mean μ and variance the same as the variance of the second term. Now for the ratio estimate the difference is

$$\hat{\mu}_{ratio} - (\mu + (\bar{Y} - \mu\bar{W}))$$

$$= -(\bar{Y} - \mu)(\bar{W} - 1) + \frac{\bar{Y}(\bar{W} - 1)^2}{\bar{W}}$$

$$= O_P\left(n^{-1}\right)$$

since $(\bar{W} - 1)$ and $(\bar{Y} - \mu)$ are both $O_p(n^{-1/2})$. For the regression estimate the difference is

$$\hat{\mu}_{reg} - (\bar{Y} - \beta(\bar{W}-1))$$

$$= (\beta - \hat{\beta})(\bar{W} - 1)$$

$$= o_p\left(n^{-1/2}\right)$$

since $(\bar{W} - 1)$ is $O_p(n^{-1/2})$ and $(\hat{\beta} - \beta) = o_p(1)$.

A3.4 Asymptotic Normality of Nonlinear Estimates

Theorem: The exponential and maximum likelihood estimates are asymptotically normal with the same variance as the regression estimate, if W is bounded and $E[Y^2] < \infty$.

Proof: Let $D_i := W_i - 1$, and write the weights in the form

(A3.4) $$\pi_{exp,i} = a\, e^{b\, D_i}$$

(A3.5) $$\pi_{mle,i} = \frac{a}{1 - b\ D_i}$$

where in each case a and b are chosen to satisfy the "moment constraints"

(A3.6) $$S(a,b) := \sum \pi_i = 1$$

(A3.7) $$T(a,b) := \sum \pi_i D_i = 0$$

Note that T(a,b) is increasing in b, for any a.

We begin by showing that the parameters a and b agree to the first order with the corresponding parameters for the regression estimate, when the weights for that estimate are written in the form

(A3.8) $$\pi_{reg,i} = a\,(1 + b\, D_i)$$

and a and b satisfy the moment constraints (A3.6 and A3.7). Let

$$m_k = \frac{1}{n} \sum_{i=1}^{n} D_i^k$$

be the estimated k-th moment of D, and write $\bar{D} = m_1$. The solutions for the regression estimate are

$$a_{reg} = \frac{1}{n\ (1 + b_{reg}\ \bar{D})}$$

$$b_{reg} = \frac{-\bar{D}}{m_2}$$

Assume that D is bounded above by M; it is bounded below by –1, since $W \geq 0$. Consider the exponential estimate. Taylor series expansions for S and T are:

(A3.9) $$S(a,b) = n\ a\ (\ 1 + b\ \bar{D} + \sum_{k=2}^{\infty} \frac{b^k\ m_k}{k!})$$

(A3.10) $$T(a,b) = n\ a\ (\ \bar{D} + b\ m_2 + \sum_{k=2}^{\infty} \frac{b^k\ m_{k+1}}{k!})$$

All m_k are bounded by $\max(1,M)^k$, so both series converge absolutely for any b. If b solves (A3.7) then

$$b_{exp} = b = \frac{-1}{m_2}(\bar{D} + \sum_{k=2}^{\infty} \frac{b^k\ m_{k+1}}{k!})$$

$$= \frac{-\bar{D}}{m_2} + O_P(n^{-1})$$

The latter equality is a solution with $b = O_P(n^{-1/2})$, and is the unique solution since T(a,b) is increasing in b. Now from (A3.6) and (A3.7),

$$a_{exp} = \frac{1}{n}(1 - \frac{\bar{D}^2}{m_2} + O_P(n^{-1}))^{-1}$$

$$= \frac{1}{n}(1 + O_P(n^{-1}))$$

Similarly, the Taylor series expansions for S and T for the maximum likelihood estimate are:

$$S(a,b) = n\,a\,(\,1 + b\,\bar{D} + \sum_{k=2}^{\infty} b^k\, m_k\,)$$

$$T(a,b) = n\,a\,(\,\bar{D} + b\,m_2 + \sum_{k=2}^{\infty} b^k\, m_{k+1}\,)$$

All m_k are bounded by $\max(1,M)^k$, so both series converge absolutely if $b < \max(1,M)^k$. As before the solution has $b = O_P(n^{-1/2})$,

$$b_{mle} = b = \frac{-1}{m_2}(\bar{D} + \sum_{k=2}^{\infty} b^k\, m_{k+1}\,)$$

$$= \frac{-\bar{D}}{m_2} + O_P(n^{-1})$$

and

$$a_{mle} = \frac{1}{n}(1 - \frac{\bar{D}^2}{m_2} + O_P(n^{-1}))^{-1}$$

$$= \frac{1}{n}(1 + O_P(n^{-1}))$$

Now we have $b_{mle} = b_{reg} + O_P(n^{-1})$, $b_{exp} = b_{reg} + O_P(n^{-1})$, $a_{mle} = a_{reg} + O_P(n^{-1})$, and $a_{exp} = a_{reg} + O_P(n^{-1})$. With $-1 \le D \le M$ this implies that

$$\pi_{exp,i} = \pi_{reg,i}\,(1 + O_P(n^{-1}))$$

$$\pi_{mle,i} = \pi_{reg,i}\,(1 + O_P(n^{-1}))$$

uniformly in D (and in i), or

$$\max_i |\,\pi_{reg,i} - \pi_{est,i}\,| < A\,n^{-2}\,,$$

where $A = O_P(1)$, for either estimate "est", since $\pi_{reg,i} = O_P(n^{-1})$. Now

$$\sqrt{n}\,|\,\hat{\mu}_{est} - \hat{\mu}_{reg}\,| = \sqrt{n}\,|\sum_{i=1}^{n} (\pi_{est,i} - \pi_{reg,i})\, Y_i\,|$$

$$\leq A\, n^{-3/2} \sum_{i=1}^{n} |Y_i|$$

$$= O_p(n^{-1/2}) = op(1)$$

since Y has a finite absolute first moment. Both estimates are equivalent to the regression estimate to the order needed for them to have the same asymptotic variance and bias.

We conjecture that a finite second moment for W is a sufficient condition for asymptotic equivalence to the first order.

A3.5 Perfect Transformation Estimate

Theorem Suppose $\theta(X)$ has an unbounded distribution when $X \sim f$. Let $g_b(x)$ be the optimal sampling distribution for the integration estimate of $\theta - b$. Then

$$\lim_{b\to\infty} \text{var}_{g_b}(\,W\,(\,\theta - b\,)\,) = 0$$

iff

$$\lim_{b\to\infty} b \int_{\theta > b} (\,\theta(x) - b\,)\, f(x) \;=\; 0$$

Proof The optimal sampling distribution $g_b(x)$ given by (2.79) is

$$g_b(x) = \frac{|\theta(x) - b|\, f(x)}{\int |\theta(x) - b|\, f(x)},$$

which results in variance

(A3.11) $$n\,\mathrm{var}(\hat{\mu}_{int}(\theta-b)) = \int W(x)(\theta(x)-b)^2 f(x) - (\mu-b)^2$$

$$= \left(\int |\,\theta(x)-b\,|\, f(x)\right)^2 - (\mu-b)^2$$

$$= \left(b - \mu + 2 \int_{\theta>b} (\theta(x) - b)\, f(x)\right)^2 - (\mu-b)^2$$

$$= 4\,(b-\mu)\,H(b) + 4\,H(b)^2,$$

where

$$H(b) = \int_{\theta>b} (\theta(x) - b)\, f(x)$$

Now as $b \to \infty$ (A3.11) goes to zero iff $b\,H(b)$ does, since μ is fixed and if $b\,H(b) = o(1)$ then $H(b)^2 = o(b^{-2})$:

$$\mathrm{var}(\hat{\mu}_{int}(\theta-b)) \to 0$$

$$\Leftrightarrow b\,H(b) \to 0$$

The same method shows that

$$\lim_{b\to-\infty} \mathrm{var}_{g_b}(\,W\,(\,\theta - b\,)\,) = 0$$

if

$$\lim_{b\to-\infty} b \int_{\theta<b} (\,\theta(x) - b\,)\, f(x) \;=\; 0$$

Corollary If $E_f[\,\max(\theta, 0)^2\,] < \infty$ then

$$\lim_{b\to\infty} \mathrm{var}_{g_b}(\,W\,(\,\theta - b\,)\,) = 0$$

Proof of Corollary By the theorem the result holds iff

(A3.12) $$\lim_{b\to\infty} b \int_{\theta>b} (\theta(x) - b)\, f(x) = 0$$

Now

$$E_f[\max(\theta, 0)^2] = \int \max(0,\theta)^2\, dF_\theta(\theta)$$

$$= \int_{\tau>0} P_f\{\max(0,\theta)^2 > \tau\}\, d\tau$$

$$= \int_{v>0} 2\, v\, P_f\{\theta > v\}\, dv \qquad \text{(change of variable)}$$

and if $E_f[\max(\theta, 0)^2] < \infty$ the last interval is finite, and so

(A3.13) $$\lim_{b\to\infty} \int_b^\infty v\, P_f\{\theta > v\}\, dv = 0\,,$$

from which we see that

$$\lim_{b\to\infty} \int_b^\infty b\, P_f\{\theta > v\}\, dv = 0\,.$$

We can also rewrite the quantity under the limit in (A3.12) as

$$b \int_{\theta>b} (\theta(x) - b)\, f(x)$$

$$= b\, E[\max(0, \theta - b)]$$

$$= b \int_b P_f\{\theta > v\}\, dv$$

which by (A3.13) goes to zero as $b \to \infty$ if $E_f[\max(\theta, 0)^2] < \infty$.

Appendix 4. Proofs for Chapter Four

A4.1 Asymptotic Normality for Quantile Estimates

The proof of the asymptotic normality of quantile estimates will use the following lemma:

Lemma (Large Weight Lemma). If $E_g[W^2] < \infty$, then

$$\lim_{n \to \infty} P_g\{ \max(W_i\,;\, 1 \le i \le n) > c\sqrt{n} \} = 0$$

for any $c > 0$.

Proof: $E_g[W^2 < \infty] \Leftrightarrow E_g[W^2/c^2 < \infty] \Leftrightarrow \sum_{n=1}^{\infty} P\{W^2/c^2 > n\} < \infty \Rightarrow$

$P\{W > c\sqrt{n}\} = o(1/n)$. The second equivalence is by Chung (1974) page 44. Write $P_n = P\{W^2/c^2 > n\}$. Suppose the final implication does not hold, then there is an infinite subsequence n_k of $\{1, 2, \ldots\}$ such that $P_n \ge b/n$ for some $b > 0$, but then

$$\sum_{n=1}^{\infty} P_n \ge \sum_{k=1}^{\infty} \frac{n_k - n_{k-1}}{n_k} \tag{A4.1}$$

If $\lim_{k\to\infty} n_{k-1}/n_k = 1$, then there is some k_0 such that for $k > k_0$ $n_{k-1}/n_k > .9$ and for $n > n_{k_0}$, $P_n > .9/n$, and $\sum P_n = \infty$. But if there is not a limit for n_{k-1}/n_k then the summation (A4.1) is infinite, and there is a contradiction. Therefor the final implication holds. Finally, $P\{W > c/\sqrt{n}\} = o(1/n) \Rightarrow P\{\max(W_i) > c/\sqrt{n}\} = o(1)$. $\therefore$

Theorem (Asymptotic normality and variance of quantile estimates)

Suppose that θ has a continuous nonzero density $f_\theta(\xi)$ at $\xi := F_\theta^{-1}(\alpha)$, $0 < \alpha < 1$, that $\hat{F}_\theta(x)_{est}$ is the estimate obtained using estimate "est" (integration, ratio, or regression) of $F_\theta(z) = P_f\{\theta \le x\}$, and that

(A4.2) $$\hat{\xi} = \hat{F}_\theta^{-1}(\alpha) := \min_x \hat{F}_\theta(x) \geq \alpha ,$$

that g dominates f for the ratio and regression estimates and that g dominates f I{θ < ξ+c} for some c > 0, and that $E_g[W^2] < \infty$, and for the regression estimate that W is bounded. Then the limiting distribution of a quantile estimate is asymptotically normal,

(A4.3) $$\sqrt{n}\,(\hat{\xi}_{est} - \xi) \xrightarrow{dist} N(0,\tau^2)$$

with variance given by

(A4.4) $$\tau^2 = \frac{\text{var}(\hat{F}_{\theta,est}(\xi))}{f_\theta(\xi)^2}$$

where the numerator is the (leading term of the) variance of the asymptotically normal distribution of $\hat{F}_{\theta,est}(\xi)$ obtained using estimate "est."

Proof: Let

(A4.5) $$\xi^* := \xi - \frac{\hat{F}_{\theta,est}(\xi) - \alpha}{f_\theta(\xi)}$$

Then ξ^* has the variance given in the theorem. If we can show that

(A4.6) $$\hat{\xi} - \xi^* = o_p(1/\sqrt{n})$$

then the theorem holds. We proceed to show that for any fixed $c_1 > 0$ that $P\{\,|\hat{\xi} - \xi^*| \geq c_1/\sqrt{n}\,\} \to 0$ as $n \to \infty$, which implies (A4.6). Note that for any a

$$\hat{\xi} > a \Leftrightarrow \hat{F}(a) < a$$

from the definition.

We begin with the integration estimate. Consider the case where $\hat{F}(\xi) := \hat{F}_{\theta,est}(\xi) < \alpha$, so $\xi^* > \xi$. Then

(A4.7)
$$\hat{\xi} > \xi^* + c_1/\sqrt{n}$$

$$\Leftrightarrow \hat{F}(\xi^* + c_1/\sqrt{n}) < \alpha$$

$$\Leftrightarrow \hat{F}(\xi^* + c_1/\sqrt{n}) - \hat{F}(\xi) < \alpha - \hat{F}(\xi)$$

$$\Leftrightarrow \hat{F}(\xi^* + c_1/\sqrt{n}) - \hat{F}(\xi) < f_\theta(\xi)\,(\xi^* - \xi)$$

We proceed to bound the probability of this event using Chebychev's inequality. Let

$$M = \#\{i : \theta_i \le \xi\} .$$

M has a binomial distribution with mean $n\gamma$, where $\gamma = E_g[I\{\theta \le \xi\}]$. For fixed a, $a > \xi$, let

$$D(a) := \hat{F}(a) - \hat{F}(\xi)$$

$$= \frac{1}{n} \sum_{i=1}^{n-M} W(X_i)\, I\{\xi < \theta(X_i) \le a\}$$

where X has a distribution equal to the marginal distribution of g given that $\theta(X) \ge \xi$. Note that D is independent of $\hat{F}(\xi)$, given M. D has conditional expected value

$$E[\, D(a) \mid M \,] = \frac{n-M}{n(1-\gamma)} (\, F(a) - \alpha \,)$$

and conditional variance

$$\mathrm{var}(\, D(a) \mid M \,) = \frac{n-M}{n^2} \left(\frac{E_g W^2\, I\{\xi<\theta\le a\}}{1-\gamma} - \left(\frac{F(a) - \alpha}{1 - \gamma} \right)^2 \right)$$

$$\le \frac{1}{n} \frac{n-M}{n(1-\gamma)} E_g[\, W^2\, I\{\xi<\theta\le a\} \,] .$$

Define

$$\tau(\delta) := \max_{|y| \le \delta} |\, F(\xi+y) - \alpha - y\, f_\theta(\xi) \,|$$

$$\nu(\delta) := E_g[\, W^2 I\{\, |\theta-\xi| \le \delta \,\}\,]$$

$$\psi(M) := \left|\, 1 - \frac{n-M}{n(1-\gamma)} \,\right|$$

which will be used to bound the error for approximations for the expectation and variance.

$$|\, E[\, D(a) \mid M\,] - (a-\xi)\, f_\theta(\xi) \,| \le \tau(\delta) + \psi(M)\, (\, (a-\xi)\, f_\theta(\xi) + \tau(\delta)\,)$$

(A4.8) $$\mathrm{var}(\, D(a) \mid M\,) \le \frac{1}{n} \nu(\delta)\, (1+ \psi(M)\,)$$

for $0 < a-\xi \le \delta$

For fixed M and $\hat{F}(\xi)$, by (A4.7) and Chebychev's inequality

(A4.9) $$P\{\, \hat{\xi} > \xi^* + c_1/\sqrt{n} \mid M, \hat{F}(\xi)\,\} =$$

$$P\{\, D(\xi^* + c_1/\sqrt{n}\,) < f_\theta(\xi)\, (\xi^* - \xi) \mid M, \hat{F}(\xi)\,\}$$

$$< \frac{\mathrm{var}(D)}{\left(E[D] - f_\theta(\xi)\, (\xi^* - \xi)\right)^2}$$

$$\le \frac{\nu(\delta)\ (1+ \psi(M)\,)\ /\ n}{(\, f_\theta(\xi) c_1/\sqrt{n} - \tau(\delta) - \psi(M)\ \delta\ f_\theta(\xi) - \psi(M)\tau(\delta)\,)^2}$$

for $\xi^* + c_1/\sqrt{n} - \xi < \delta$. Note that this bound depends on M and $\hat{F}(\xi)$ only through δ and $\psi(M)$.

Note that $\tau(\delta) = o(\delta)$ as $\delta \to 0$, since the distribution of θ has a continuous density at ξ, and that $\nu(\delta) \to 0$ as $\delta \to 0$, since $E_g[W^2 < \infty]$. Choose a sequence $\{b_n, n=1, 2, \ldots\}$, such that $\sqrt{n}\, b_n \to \infty$, $\tau(b_n) = o(1/\sqrt{n})$, and $\nu(b_n) \to 0$. This sequence may be chosen by letting $b_n = k_n / \sqrt{n}$, where $k_n \to \infty$ slowly enough that $\tau(b_n)$ and $\nu(b_n)$ still decrease fast enough. Also choose a sequence $\{a_n, n = 1, 2, \ldots\}$ such that $\sqrt{n}\, a_n \to \infty$ and $a_n b_n = o(1/\sqrt{n}\,)$.

Substitute b_n for δ and a_n for $\psi(M)$ in (A4.9), and remove the conditioning on M and $\hat{F}(\xi)$

$$P\{\ \hat{\xi} > \xi^* + c_1/\sqrt{n}\ \}$$

$$\le P\{\ \xi^* + c_1/\sqrt{n} - \xi \ge b_n\ \} + P\{\ \psi(M) > a_n\ \}$$

$$+ P\{\ \hat{\xi} > \xi^* + c_1/\sqrt{n}\ \mid\ \xi^* + c_1/\sqrt{n} - \xi < b_n\ ,\ \psi(M) \le a_n\ \}$$

$$\le P\{\ \xi^* + c_1/\sqrt{n} - \xi \ge b_n\ \} + P\{\ \psi(M) > a_n\ \}$$

$$+ \frac{v(b_n)\ (1+ a_n\ )\ /\ n}{(\ f_\theta(\xi)c_1/\sqrt{n} - \tau(b_n) - a_n b_n f_\theta(\xi) - a_n\tau(b_n)\ )^2}$$

$$\rightarrow 0 \text{ as } n \rightarrow \infty.$$

The first probability decreases to zero as $n \rightarrow \infty$, since $\xi^* - \xi = O_p(1/\sqrt{n})$ and $\sqrt{n}\, b_n \rightarrow \infty$. The second probability decreases to zero since M has expected value γ, M/n is $O_p(1/\sqrt{n})$ and $\sqrt{n}\, a_n \rightarrow \infty$. By the conditions on a_n and b_n, the third term has a numerator which is o(1/n) and a denominator which is O(1/n), so the fraction decreases to zero.

Similary,

$$\hat{\xi} \le \xi^* + c_1/\sqrt{n}$$

$$\Leftrightarrow \hat{F}(\ \xi^* - c_1/\sqrt{n}\ ) - \hat{F}(\xi) \ge f_\theta(\xi)\ (\xi^* - \xi)$$

$$P\{\ \hat{\xi} \le \xi^* - c_1/\sqrt{n}\ \mid M, \hat{F}(\xi)\ \} =$$

$$P\{\ D(\xi^* - c_1/\sqrt{n}\ ) \ge f_\theta(\xi)\ (\xi^* - \xi) \mid M, \hat{F}(\xi)\ \}$$

$$\le \frac{v(\delta)\ (1+ \psi(M)\ )\ /\ n}{(\ f_\theta(\xi)c_1/\sqrt{n} - \tau(\delta) - \psi(M)\delta f_\theta(\xi) - \psi(M)\tau(\delta)\ )^2}$$

for $0 \le \xi^* - \xi < \delta$.

$$P\{ \hat{\xi} \le \xi^* - c_1/\sqrt{n} \}$$

$$\le P\{ \xi^* - \xi \ge b_n \} + P\{ \psi(M) > a_n \}$$

$$+ P\{ \hat{\xi} \le \xi^* - c_1/\sqrt{n} \mid \xi^* - \xi < b_n , \psi(M) \le a_n \}$$

$$\le P\{ \xi^* + c_1/\sqrt{n} - \xi \ge b_n \} + P\{ \psi(M) > a_n \}$$

$$+ \frac{v(b_n)\ (1+ a_n\ )\ /\ n}{(\ f_\theta(\xi)c_1/\sqrt{n} - \tau(b_n) - a_n b_n f_\theta(\xi) - a_n \tau(b_n)\ )^2}$$

$$\rightarrow 0 \text{ as } n \rightarrow \infty.$$

Next consider the case where $\hat{F}(\xi) > \alpha$, so $\xi^* < \xi$. This case presents additional difficulties, since $\hat{F}(\xi^*) - \hat{F}(\xi)$ is not independent of $\hat{F}(\xi)$ for $\xi^* < \xi$. In contrast to the similar problem without importance sampling, the joint distribution of W and θ may be such that knowledge of $\hat{F}(\xi)$ may be sufficient to determine $\hat{F}(\xi^*)$ exactly. We take a different approach in this case, and note that $\hat{F}(\xi) - \hat{F}(\hat{\xi})$ depends on $\hat{F}(\hat{\xi})$ only through M, where M and γ are now defined as

$$M := \#\{ i : \theta_i \le \hat{\xi} \}$$

$$\gamma = \gamma(\hat{\xi}) = E[\ M \mid \hat{\xi}\ ]$$

We treat $\hat{F}(\hat{\xi})$ as fixed and put bounds on the random behavior of $\hat{F}(\xi) - \hat{F}(\hat{\xi})$ (and so on ξ^*). That is,

$$\hat{\xi} \le \xi^* - c_1/\sqrt{n}$$

$$\Leftrightarrow \hat{F}(\xi) \le \alpha + f_\theta(\xi)\ (\xi - \hat{\xi} - c_1/\sqrt{n}\ )$$

$$\Leftrightarrow \hat{F}(\xi) - \hat{F}(\hat{\xi}) \le f_\theta(\xi)\ (\xi - \hat{\xi} - c_1/\sqrt{n}\ ) + \alpha - \hat{F}(\hat{\xi})$$

D is defined as before, but now $D < 0$, and a slight modification of the variance formula must be made:

$$\mathrm{var}(\,D(a) \mid M\,) \le \frac{1}{n}\frac{n-M}{n(1-\gamma(a))}\, E_g[\, W^2\, I\{a<\theta\le\xi\}\,]\,.$$

For fixed M and $\hat{F}(\hat{\xi})$,

(A4.10) $$P\{\,\hat{\xi} \le \xi^* - c_1/\sqrt{n} \mid M, \hat{F}(\hat{\xi})\,\}$$

$$= P\{\,-D(\hat{\xi}) < f_\theta(\xi)\,(\xi - \hat{\xi} - c_1/\sqrt{n}\,) + \alpha - \hat{F}(\hat{\xi})\,\}$$

$$\le P\{\,-D(\hat{\xi}) < f_\theta(\xi)\,(\xi - \hat{\xi} - c_1/\sqrt{n}\,)\,\}$$

$$< \frac{\mathrm{var}(D)}{\left(E[-D] - f_\theta(\xi)\,(\xi - \hat{\xi} - c_1/\sqrt{n}\,)\right)^2}$$

$$\le \frac{v(\delta)\,(1+\psi(M,\hat{\xi})\,)\,/\,n}{(\,f(\xi)c_1/\sqrt{n} - \tau(\delta) - \psi(M,\hat{\xi})\,f_\theta(\xi) - \psi(M,\hat{\xi})\tau(\delta)\,)^2}$$

for $\xi - \hat{\xi} < \delta$. Note that this bound depends on M and $\hat{F}(\hat{\xi})$ only through δ and $\psi(M,\hat{\xi})$.

Substitute b_n for δ and a_n for $\psi(M,\hat{\xi})$ in (A4.10), and remove the conditioning on M and $\hat{F}(\hat{\xi})$

$$P\{\,\hat{\xi} \le \xi^* - c_1/\sqrt{n}\,\}$$

$$\le P\{\,\hat{\xi} < \xi - b_n\,\} + P\{\,\psi(M,\hat{\xi}) > a_n\,\}$$

$$+ P\{\,\hat{\xi} \le \xi^* - c_1/\sqrt{n} \mid \hat{\xi} \ge \xi - b_n\,,\ \psi(M,\hat{\xi}) \le a_n\,\}$$

$$\le P\{\,\hat{\xi} < \xi - b_n\,\} + P\{\,\psi(M,\hat{\xi}) > a_n\,\}$$

$$+ \frac{v(b_n)\,(1 + a_n\,)\,/\,n}{(\,f(\xi)c_1/\sqrt{n} - \tau(b_n) - a_n b_n f_\theta(\xi) - a_b\tau(b_n)\,)^2}$$

$$\to 0 \text{ as } n \to \infty.$$

Finally, we consider the case where $\hat{\xi} \le \xi$, and show that $P\{\hat{\xi} > \xi^* + c_1/\sqrt{n}\}$ decreases to zero. This case presents one additional difficulty, that in general $\hat{F}(\hat{\xi}) > \alpha$. If W is bounded this overshoot is not be a problem, but our conditions are only that $E_g[W^2] < \infty$. We will use the large-weight lemma here the put a probabilistic bound on the overshoot. If the distribution of θ has no atoms the overshoot is (a.e.) less than 1/n times the largest value of W, and by the large-weight lemma, for any $c_2 > 0$

$$P\{\hat{F}(\hat{\xi}) - \alpha > c_2/\sqrt{n}\} \to 0.$$

By the assumptions for the theorem θ has a density at ξ, and since $\hat{\xi}$ is consistent the probability that $\hat{\xi}$ occurs at an atom goes to zero. Choose c_2, $0 < c_2 < c_1 f_\theta(\xi)$, and proceed:

$$\hat{\xi} > \xi^* + c_1/\sqrt{n}$$

$$\Leftrightarrow \hat{F}(\xi) > \alpha + f_\theta(\xi)\,(\xi - \hat{\xi} + c_1/\sqrt{n})$$

$$\Leftrightarrow \hat{F}(\xi) - \hat{F}(\hat{\xi}) > f_\theta(\xi)\,(\xi - \hat{\xi} + c_1/\sqrt{n}) + \alpha - \hat{F}(\hat{\xi})$$

For fixed M and $\hat{F}(\hat{\xi})$,

$$P\{\hat{\xi} > \xi^* + c_1/\sqrt{n} \mid M, \hat{F}(\hat{\xi})\}$$

$$= P\{-D(\hat{\xi}) \ge f_\theta(\xi)\,(\xi - \hat{\xi} - c_1/\sqrt{n}) + \alpha - \hat{F}(\hat{\xi})\}$$

$$< \frac{\mathrm{var}(D)}{\left(E[-D] - f_\theta(\xi)\,(\xi - \hat{\xi} - c_1/\sqrt{n} - \alpha + \hat{F}(\hat{\xi}))\right)^2}$$

$$\le \frac{v(\delta)\,(1 + \psi(M,\hat{\xi}))\,/\,n}{(\,f(\xi)c_1/\sqrt{n} - \tau(\delta) - \psi(M,\hat{\xi})\delta f_\theta(\xi) - \psi(M,\hat{\xi})\tau(\delta) - \alpha + \hat{F}(\hat{\xi})\,)^2}$$

for $\xi - \hat{\xi} < \delta$. Substitute a_n and b_n and remove the conditioning as before

$$P\{\hat{\xi} \le \xi^* - c_1/\sqrt{n}\}$$

$$\le P\{\hat{\xi} < \xi - b_n\} + P\{\psi(M,\hat{\xi}) > a_n\} + P\{\alpha - \hat{F}(\hat{\xi}) > c_2/\sqrt{n}\}$$

$$+ P\{\hat{\xi} \le \xi^* - c_1/\sqrt{n} \mid \hat{\xi} \ge \xi - b_n, \psi(M,\hat{\xi}) \le a_n\}$$

$$\le P\{\hat{\xi} < \xi - b_n\} + P\{\psi(M,\hat{\xi}) > a_n\} + P\{\alpha - \hat{F}(\hat{\xi}) > c_2/\sqrt{n}\}$$

$$+ \frac{v(b_n)\,(1 + a_n)\,/\,n}{(f(\xi)c_1/\sqrt{n} - \tau(b_n) - a_n b_n f_\theta(\xi) - a_b\tau(b_n) - c_2/\sqrt{n})^2}$$

$$\to 0 \text{ as } n \to \infty.$$

This is the same as before except that now c_2 must be $< f(\xi)\, c_1$, as specified above.

The proof for the ratio and regression estimates is verbal, based on the proof for the integration estimate. For $-b_n < \hat{\xi} < b_n$ the expected value of the estimates of $D(\xi^*)$ and $D(\hat{\xi})$ is the same, modulo a low-order term, and the variances have the same upper bound given by (A4.8). Thus the same limiting bounds hold. The final case, where overshoot must be considered, presents no problems for the ratio estimate, since the ratio of the overshoot in the ratio case to the overshoot in the integration case is given by $\bar{W}$, which goes to 1 since $E[W] = 1$ and $\mathrm{var}(W) < \infty$. For the regression estimate it is a problem, since the ratio between the integration weights and the regression weights is linear in W, and for $\bar{W} < 1$ the ratio is not bounded unless W is, hence the additional condition.

A4.2 Chi-Square Majorization

Proposition: A chi-square distribution with m+3d+1 degrees of freedom is a majorizing distribution (suitable for acceptance-rejection random variable generation) for

$$f(x) = c(d,m) \prod_{j\neq k} |x_j - x_k| \prod_{j=1}^{d} x_j^{(m-d-1)/2}\, e^{-.5\sum x_j} \tag{A4.11}$$

Proof: The middle term in (A4.11) is bounded by

(A4.12)
$$\prod_{j\neq k}|x_j - x_k|$$

$$\leq \prod_{j\neq k}\max(x_j, x_k)$$

$$= \left(\prod_{j=1}^{d} \prod_{k:x_k<x_j} x_j\right)^2$$

$$\leq \prod_{j=1}^{d} x_j^{2d}$$

and the whole product is bounded by

$$f(x) \leq c(d,m) \prod_{j=1}^{d} x_j^{(m+3d-1)/2} \; e^{-.5 \, \Sigma xj}$$

which is

$$c(d,m) \left(\Gamma(\frac{m+3d+1}{2}) \, 2^{(m+3d+1)/2}\right)^d$$

times a product of independent chi-square distributions with degrees of freedom m+3d+1.

Appendix 5. Proofs for Chapter Six

A5.1 Inequalities for the Mixing Parameter in Mixture Sampling

The proofs of the inequalities use the following "Mixing Lemma":

Lemma ("Mixing Lemma") If $C(x) \geq 0$, $C(x) > 0$ when $A(x) \neq 0$, $D(x) \geq 0$, and $D(x) > 0$ when $B(x) \neq 0$, then

$$\int \frac{(A(x) + B(x))^2}{C(x) + D(x)} dx \leq \int \frac{A(x)^2}{C(x)} dx + \int \frac{B(x)^2}{D(x)} dx$$

with equality iff

$$\frac{A(x)}{C(x)} = \frac{B(x)}{D(x)},$$

except possibly on a set of measure 0 (with respect to the measure $A(x)^2 + B(x)^2$).

Proof of Lemma

$$\frac{(A + B)^2}{C + D} \leq \frac{A^2}{C} + \frac{B^2}{D}$$

$$\Leftrightarrow (A + B)^2 C D \leq A^2 D (C + D) + B^2 C (C + D)$$

$$\Leftrightarrow 2 A B C D \leq A^2 D^2 + B^2 C^2$$

$$\Leftrightarrow (AD - BC)^2 \geq 0$$

so the inequality holds. The inequality is strict if $(AD - BC) \neq 0$, and the integral inequality is strict if $\int (A(x)D(x) - B(x)C(x))^2 > 0$, which is true under the additional condition. $\therefore$

Theorem (Inequalities for the Mixing Parameter in Mixture Sampling)

Let

(A5.1) $$V(\lambda,c) = \int \frac{f(x)^2 \, (\theta(x) - c)^2}{g_\lambda(x)}.$$

where

(A5.2) $$g_\lambda(x) = \lambda\, f(x) + (1-\lambda)\, g_0(x)$$

and $0 \le \lambda \le 1$. Then the following inequalities hold:

If $V(0) < \infty$ then

(A5.3) $$V(\lambda) \le \frac{V(0)}{1-\lambda}$$

If $V(1) < \infty$ then

(A5.4) $$V(\lambda) \le \frac{V(1)}{\lambda}$$

If $V(0) < \infty$ and $V(1) < \infty$ then

(A5.5) $$V(\lambda) \le \frac{V(0)\, V(1)}{\lambda\, V(0) + (1-\lambda)\, V(1)}$$

If $0 < \lambda^* < 1$ then

(A5.6) $$V(\lambda) \le V(\lambda^*) \left(\frac{\lambda^{*2}}{\lambda} + \frac{(1-\lambda^*)^2}{1-\lambda} \right)$$

If $0<\lambda^*<1$ and $V(0) < \infty$ then

(A5.7) $$V(\lambda) \le V(\lambda^*) \left(\frac{a\, \lambda^{*2} - \lambda\, (\, \lambda^{*2} + 2(\lambda^*-1)(a-1)\,)}{(1-\lambda)\, (\lambda(a-1) + \lambda^{*2})} \right)$$

where $a := V(0) / V(\lambda^*)$.

If $0<\lambda^*<1$ and $V(1) < \infty$ then

(A5.8) $$V(\lambda) \le V(\lambda^*) \left(\frac{a\, (1-\lambda^*)^2 - (1-\lambda)\, (\, \lambda^{*2} - 2a\lambda^* + a)}{\lambda\, (\, (1-\lambda)(a-1) + (1-\lambda^*)^2\,)} \right)$$

where $a := V(1) / V(\lambda^*)$.

Proof Take m fixed, and let

$$m(x) = f(x)^2\, (\theta(x) - c\,)^2 ,$$

then

$$V(\lambda) = \int \frac{m(x)}{\lambda f(x) + (1\text{-}\lambda)g(x)}$$

where f and g are the two components in the mixture distributions. The first two inequalities are trivial; the first inequality holds since $f(x) \geq 0$, and the second holds since $g(x) \geq 0$. The third uses the mixing lemma. Let $\nu = \lambda\, V(0) / (\lambda\, V(0) + (1-\lambda)\, V(1))$ and apply the mixing lemma with $A(x) = \nu \sqrt{m(x)}$, $B(x) = (1-\nu)\sqrt{m(x)}$, $C(x) = \lambda f(x)$, and $D(x) = (1-\lambda)g_0(x)$. Algebraic simplification yields the desired result.

Now consider the cases where $0 < \lambda^* < 1$. Let

(A5.9) $$h(x) := \lambda^* f(x) + (1\text{-}\lambda^*)\, g_0(x)$$

be the sampling distribution corresponding to the optimal mixture, and define

$$q(x) = \frac{m(x)\, h(x)}{\int m(x)\, h(x)\, dx}$$

$$r(x) = g_0(x) / h(x)$$

$$s(x) = f(x) / h(x)$$

Then

$$V(\lambda) = \int \frac{m(x) / h(x)}{\lambda s(x) + (1-\lambda) r(x)} dx$$

$$= \int \frac{q(x)}{\lambda s(x) + (1-\lambda) r(x)} dx \int \frac{m(x)}{h(x)} dx$$

$$= V(\lambda^*)\, E_q\left[\frac{1}{\lambda s(x) + (1-\lambda) r(x)} \right]$$

Note that $s = (1 - (1-\lambda^*) r) / \lambda^*$, substitution and simplification give

$$V(\lambda) = V(\lambda^*)\, E_q\left[\frac{1}{\lambda/\lambda^* + r\,(1 - \lambda/\lambda^*)} \right]$$

Differentiating and setting equal to zero at the optimum $\lambda = \lambda^*$

(A5.10) $$\frac{\partial V(\lambda)}{\partial \lambda} |_{\lambda=\lambda^*} = 0 = E_q[1-r(x)]$$

The derivative and integral can be interchanged because the integral is finite for $0 < \lambda < 1$ and the integrand is positive since $0 \le r \le 1/(1-\lambda^*)$.

Now think of r as a random variable with $E_q[r] = 1$ (from (A5.10)) and range $[0, (1-\lambda^*)^{-1}]$, and note that:

(A5.11) $$\frac{V(\lambda)}{V(\lambda^*)} = E_q\left[\frac{1}{\lambda/\lambda^* + r\,(1 - \lambda/\lambda^*)} \right]$$

The remaining inequalities are are bounds on the ratio $V(\lambda)/V(\lambda^*)$, and are found by functional analysis with respect to r.

If there are no other constraints then $V(\lambda)/V(\lambda^*)$ is maximized when r has support on the endpoints of its distribution, $P\{r = 0\} = \lambda^* = 1 - P\{r = 1/(1-\lambda^*)\}$, which yields (A5.6).

If $V(0) < \infty$ then $V(\lambda)$ is maximized (for any λ) by a two-point distribution for r with support on its maximum value and one other value which is at its maximum possible value subject to the constraint on $V(0)/V(\lambda^*)$. Let $a = V(0) / V(\lambda^*)$, then the maximum $V(\lambda)$ is when $P\{r = 1/(1-\lambda^*)\} = p = (1 - \lambda^*)(a - 1) / (a - 2\lambda^* + \lambda^{*2})$ and $P\{r = (1 - p/(1-\lambda^*))/(1-p)\} = 1 - p$. Substituting these probabilities and values into (A5.11) gives (A5.7). The case (A5.8) is obtained by a similar argument, or by symmetry.

A5.2 Variance Under Stratified Mixtures

Proposition: If distribution allocations in mixture sampling are fixed at $n\lambda$ and $n(1-\lambda)$ (rounded to the nearest integer) then as $n \to \infty$ the leading term of the asymptotic variance is

(A5.12) $$\sigma^2_{fix} = \lambda\, var_f(\, Y - c\, W\,) + (1-\lambda)\, var_{g_0}(\, Y - c\, W\,) - (\mu - c)^2$$

where c = 0, μ, and β for the integration, ratio, and regression estimates.

Proof: The proof uses the same kind of delta-method argument as in the unstratified case, but with moments computed separately for observations from f and g_0. We ignore the rounding of nλ, which is of small order. Let

$$\mu^{(1)} = E_f[Y]$$

$$\mu^{(2)} = E_{g_0}[Y]$$

$$m^{(1)} = E_f[W]$$

$$m^{(2)} = E_{g_0}[W]$$

$$C^{(k)}{}_i = Y_i - \mu^{(k)},$$

$$D^{(k)}{}_i = W_i - m^{(k)}$$

and let $\bar{C}^{(k)}$ and $\bar{D}^{(k)}$ be the average of the C and D values over observations taken from the corresponding sampling distribution. Note that $\lambda\mu^{(1)} + (1-\lambda)\mu^{(2)} = \mu$, $\lambda m^{(1)} + (1-\lambda)m^{(2)} = 1$, $\lambda\bar{C}^{(1)} + (1-\lambda)\bar{C}^{(2)} = \bar{C} = \bar{Y} - \mu$, and $\lambda\bar{D}^{(1)} + (1-\lambda)\bar{D} = \bar{D} = \bar{W} - 1$. First-order delta-method expansions for the three estimates yield:

(A5.13) $$\hat{\mu}_{int} = \bar{Y}$$

$$\approx \mu + \lambda\ \bar{C}^{(1)} + (1-\lambda)\ \bar{C}^{(2)}$$

(A5.14) $$\hat{\mu}_{ratio} = \frac{\bar{Y}}{\bar{W}}$$

$$\approx \mu + \lambda\ (\bar{C}^{(1)} - \mu\bar{D}^{(1)}) + (1-\lambda)\ (\bar{C}^{(2)} - \mu\bar{D}^{(2)})$$

(A5.15) $$\hat{\mu}_{reg} = \bar{Y} - \hat{\beta}\,(\bar{W} - 1)$$

$$\approx \mu + \lambda\,(\bar{C}^{(1)} - \beta\,\bar{D}^{(1)}) + (1-\lambda)\,(\bar{C}^{(2)} - \beta\,\bar{D}^{(2)})$$

where $\hat{\beta}$ is computed from the joint sample. These approximations have variances

$$\mathrm{var}(\,\lambda\,(\bar{C}^{(1)} - c\bar{D}^{(1)}) + (1-\lambda)\,(\bar{C}^{(2)} - c\bar{D}^{(2)})\,)$$

$$= \frac{\lambda^2\,\mathrm{var}_f(C - cD)}{n\lambda} + \frac{(1-\lambda)^2\,\mathrm{var}_{g_0}(C - cD)}{n(1-\lambda)}$$

$$= \frac{1}{n}[\,\lambda\,\mathrm{var}_f(C - cD) + (1-\lambda)\mathrm{var}_{g_0}(C - cD)\,]$$

$$= \frac{1}{n}[\,\lambda\,\mathrm{var}_f(Y - cW) + (1-\lambda)\mathrm{var}_{g_0}(Y - cW)\,]$$

where $c = 0$, μ, and β for the integration, ratio, and regression estimates. Multiply by n to normalize, and the results are the variances given in the statement of the proposition. The proof that the error in the approximations (A5.13-A5.15) follows the same methods as used in the unstratified case, and is not repeated here.

A5.3 Superiority of Single Ratio Estimate

Theorem 6.3 When using the ratio estimate, $\sigma_c^2 \le \sigma_a^2$, for $0 \le a \le 1$, with equality iff $g_2 / g_1 = \zeta$ for all x for which $f(x)\,(\theta(x)-\mu)$ is nonzero and for some constant ζ, and $a = n_1 / (n_1 + \zeta\, n_2)$.

Proof The asymptotic variance of the linear combination $\hat{\mu}_a = a\,\hat{\mu}_1 + (1-a)\,\hat{\mu}_2$ is

(A5.16) $$\sigma_a^2 = a^2 \int \frac{(\theta-\mu)^2 f^2(x)}{n_1 g_1(x)} + (1-a)^2 \int \frac{(\theta-\mu)^2 f^2(x)}{n_2 g_2(x)}$$

No covariance term is needed because $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent. By (6.18) the asymptotic variance of the combined estimate is

(A5.17) $$\sigma_c^2 \le \int \frac{(\theta-\mu)^2 f^2(x)}{n_1 g_1(x) + n_2 g_2(x)}$$

with equality iff $E_{g_1}[Y-\mu W] = 0$. Apply the mixing lemma from section 5.1 of the appendix with $C(x) = n_1\, g_1(x)$, $D(x) = n_2\, g_2(x)$, $A(x) = a\, (\theta-\mu)\, f(x)$, and $B(x) = (1-a)\, (\theta-\mu)\, f(x)$, to obtain

(A5.18) $$\int \frac{(\theta-\mu)^2 f^2(x)}{n_1 g_1(x) + n_2 g_2(x)} \le \sigma_a^2 ,$$

with equality iff $a\, n_2\, g_2 = (1-a)\, n_1\, g_1$ everywhere that $(\theta-\mu)\, f(x) \neq 0$, which holds iff the additional condition on the theorem holds. This also implies the condition for equality in (A5.17). The combination of (A5.18) and (A5.17) gives the stated result. ∴

A5.4 Mixture vs Combination for Integration and Regression Estimates

Proposition 6.1 (variance of the modified integration estimate)

The variance of the modified integration estimate (6.36) is:

(A5.19) $$\mathrm{var}(\hat{\mu}_{int}^{(k:m)}) = \frac{1}{n} \int \pi^{(k)2}\, (Y^{(k)}-\mu)^2\, g(x)\, dx$$

where $Y^{(k)} = Y^{(k)}(x)$ and $\pi^{(k)} = \pi^{(k)}(x)$.

Proof: A delta-method expansion for the estimate gives

$$\hat{\mu}_{int}^{(k:m)} = \Sigma p_i^{(k)}\, Y_i^{(k)}$$

$$= \frac{\Sigma \pi_i^{(k)}\, Y_i^{(k)}}{\Sigma \pi_i^{(k)}}$$

$$\approx \mu + \Sigma \pi_i^{(k)}\, Y_i^{(k)} - \mu\, \Sigma \pi_i^{(k)}$$

$$= \mu + \Sigma \pi_i^{(k)}\, (Y_i^{(k)} - \mu)$$

which has variance (A5.19). The summations are over all n observations.

Proposition 6.2 (variance of the modified regression estimate)

The variance of the modified regression estimate (6.38) is:

(A5.20) $$\text{var}(\hat{\mu}_{reg}^{(k:m)}) = \frac{1}{n}\int \pi^{(k)2}\,(Y^{(k)} - \mu - \beta^{(k)}\,(W^{(k)} - 1))^2\, g\, dx\ .$$

Proof: A delta-method expansion gives

$$\hat{\mu}_{reg}^{(k:m)} = \Sigma\, p_i^{(k)}\, Y_i^{(k)} - \hat{\beta}^{(k:m)}\, (\Sigma\, p_i^{(k)}\, W_i^{(k)} - 1)$$

$$= \frac{\Sigma \pi_i^{(k)}\, Y_i^{(k)}}{\Sigma \pi_i^{(k)}} - \hat{\beta}^{(k:m)} \left(\frac{\Sigma\, \pi_i^{(k)}\, W_i^{(k)}}{\Sigma\, \pi_i^{(k)}} - 1\right)$$

$$\approx \mu + \Sigma \pi_i^{(k)}\, Y_i^{(k)} - \mu \Sigma \pi_i^{(k)} - \beta^{(k)}(\Sigma \pi_i^{(k)}\, W_i^{(k)} - \Sigma \pi_i^{(k)})$$

$$= \mu + \Sigma\, \pi_i^{(k)}\, (Y_i^{(k)} - \beta^{(k)} W_i^{(k)} - \mu + \beta^{(k:m)}\,)$$

which has variance (A5.20).

Proposition 6.3 (variance of the combined integration estimate)

Define the combined integration estimate, with parameter a, as:

(A5.21) $$\hat{\mu}_{int}{}^{(a:m)} = a\, \hat{\mu}_{int}{}^{(1:m)} + (1-a)\, \hat{\mu}_{int}{}^{(2:m)}$$

The asymptotic variance of $\hat{\mu}_{int}{}^{(a:m)}$ is

(A5.22) $$\text{var}(\hat{\mu}_{int}{}^{(a:m)}) = \frac{1}{n}\int g\, (Y - \mu(a\, \pi^{(1)} + (1-a)\pi^{(2)}))^2\, dx$$

Proof: A delta-method expansion for the estimate gives

$$\hat{\mu}_{int}^{(a:m)} = a\, \frac{\overline{Y}}{\overline{\pi}^{(1)}} + (1-a)\, \frac{\overline{Y}}{\overline{\pi}^{(2)}}$$

$$\approx \overline{Y} - \mu\, (a\, \overline{\pi}^{(1)} + (1-a)\, \overline{\pi}^{(2)} - 1)$$

which has variance (A5.22).

Proposition 6.4 (variance of the combined regression estimate)

Define the combined regression estimate, with parameter a, as:

(A5.23) $$\hat{\mu}_{reg}^{(a:m)} = a\,\hat{\mu}_{reg}^{(1:m)} + (1-a)\,\hat{\mu}_{reg}^{(2:m)}$$

The asymptotic variance of $\hat{\mu}_{reg}^{(a:m)}$ is

(A5.24) $$\mathrm{var}(\,\hat{\mu}_{reg}^{(a:m)}\,) = \frac{1}{n}\int g\,(\,a\,h^{(1)} + (1-a)\,h^{(2)}\,)^2\,dx$$

where

$$h^{(k)}(x) = Y(x) - W(x)\,\beta^{(k)} - \pi^{(k)}(x)\,(\mu - \beta^{(k)})$$

Proof: A delta-method expansion for the estimate gives

$$\hat{\mu}_{reg}^{(a:m)} = a\left(\frac{\bar{Y}}{\bar{\pi}^{(1)}} - \hat{\beta}^{(1:m)}\left(\frac{\bar{W}}{\bar{\pi}^{(1)}} - 1\right)\right)$$

$$+ (1-a)\left(\frac{\bar{Y}}{\bar{\pi}^{(2)}} - \hat{\beta}^{(2:m)}\left(\frac{\bar{W}}{\bar{\pi}^{(2)}} - 1\right)\right)$$

$$\approx \bar{Y} + a\,(\,-\beta^{(1)}\,\bar{W} - \bar{\pi}^{(1)}\,(\mu - \beta^{(1)}))$$

$$+ (1-a)\,(\,-\beta^{(2)}\,\bar{W} - \bar{\pi}^{(2)}\,(\mu - \beta^{(2)}))$$

which has variance (A5.24).

The preceeding proofs use a delta-method approximation without showing that the approximation is sufficiently accurate. This is done using the same methods as in the proofs of the asymptotic variances of the three usual estimates, and is not repeated here.

Theorem 6.4 (superiority of the combined integration estimate)

The combined integration estimate defined in proposition 6.3 has smaller asymptotic variance than the corresponding weighted average of results from each replication, i.e.

(A5.25) $$\text{var}(\hat{\mu}_{int}^{(a:m)}) \leq \text{var}(a\,\hat{\mu}_{int}^{(1)} + (1-a)\,\hat{\mu}_{int}^{(2)})\,.$$

Equality holds if $a = n_2/(n_1 + n_2)$ and $g_1(x) = g_2(x)$ (except possibly on a set of measure 0).

Proof: The proof uses the mixing lemma.

$$\text{var}(\hat{\mu}_{int}^{(a:m)})$$

$$= \frac{1}{n}\int g\,(Y - \mu(a\,\pi^{(1)} + (1-a)\pi^{(2)}))^2\,dx$$

$$= \int \frac{(\theta f - \mu(a\,g_1 + (1-a)g_2\,))^2}{n\,g}\,dx$$

$$= \int \frac{(\,a(\theta f - \mu\,g_1) + (1-a)(\theta f - \mu\,g_2))^2}{n_1 g_1 + n_2 g_2}\,dx$$

$$\leq \int \frac{(a(\theta f - \mu\,g_1))^2}{n_1 g_1}\,dx + \int \frac{((1-a)(\theta f - \mu\,g_2))^2}{n_2 g_2}\,dx$$

$$= \text{var}(a\,\hat{\mu}_{int}^{(1)} + (1-a)\,\hat{\mu}_{int}^{(2)}).$$

The inequality is by the mixing lemma from section 5.1 of the appendix. If the additional condition is satisfied then the conditions required for equality in the lemma are satisfied.

Theorem 6.5 (superiority of the combined regression estimate)

The combined regression estimate defined in proposition 6.4 has smaller asymptotic variance than the corresponding weighted average of results from each replication, i.e.

(A5.26) $$\text{var}(\hat{\mu}_{reg}^{(a:m)}) \leq \text{var}(a\,\hat{\mu}_{reg}^{(1)} + (1-a)\,\hat{\mu}_{reg}^{(2)})\,.$$

Equality holds if $a = n_2/(n_1 + n_2)$ and $g_1(x) = g_2(x)$ (except possibly on a set of measure 0).

Proof: The proof uses the mixing lemma.

$$\mathrm{var}(\hat{\mu}_{reg}(a{:}m))$$

$$= \frac{1}{n} \int g \, (a\, h^{(1)} + (1-a)\, h^{(2)})^2 \, dx$$

$$= \int \frac{(a\, h^{(1)}\, g + (1-a)\, h^{(2)}\, g)^2}{n\, g} dx$$

$$\leq \int \frac{(a\, h^{(1)}\, g)^2}{n_1 g_1} dx + \int \frac{((1-a)\, h^{(2)}\, g)^2}{n_2 g_2} dx$$

$$= \frac{a^2}{n_1} \int (h^{(1)} \frac{g}{g_1})^2 \, g_1 \, dx + \frac{a^2}{n_2} \int (h^{(2)} \frac{g}{g_2})^2 \, g_2 \, dx$$

$$= a^2\, \mathrm{var}(\hat{\mu}_{reg}(1)) + (1-a)^2\, \mathrm{var}(\hat{\mu}_{reg}(2))$$

$$= \mathrm{var}(a\, \hat{\mu}_{reg}(1) + (1-a) \hat{\mu}_{reg}(2))$$

The fourth equality is since

$$\frac{g}{g_k} h^{(k)} = (Y - W\, \beta^{(k)} - \pi^{(k)}\, (\mu - \beta^{(k)})) \,/\, \pi^{(k)}$$

$$= Y^{(k)} - W^{(k)}\, \beta^{(k)} - (\mu - \beta^{(k)})$$

and the leading variance term of a regression estimate based only on observations from distribution k is:

$$\mathrm{var}(\hat{\mu}_{reg}(k)) = \frac{1}{n_k} \mathrm{var}_{g_k}(Y^{(k)} - W^{(k)}\, \beta^{(k)})$$

A5.6 Uniqueness of Exponential Tilting

Theorem 6.6 (Uniqueness of Exponential Tilting for Memoryless Weights)
Let f(x) and g(x) be such that $f(x) = \prod_{j=1}^{d} f_j(x_j)$, $g(x) = \prod_{j=1}^{d} g_j(x_j)$, $d>1$, $g(x)>0$ when $f(x)>0$,

(A5.27) $$g(x) = f(x)\, r(S(x)) ,$$

where $S(x) = \sum_{j=1}^{d} s_j(x_j)$, s_j is a real-valued function of x_j and r is a real-valued function. If the support of each s_j is a set of consecutive lattice points with the same lattice size or if each s_j has a nonzero density on a single (possibly infinite) interval, then

(A5.28) $$g(x) = \alpha\, e^{\beta S} f(x)$$

for some α and β.

Proof: First, $r_j(x_j) := g_j(x_j) / f_j(x_j)$ must be a function only of $s_j(x_j)$, otherwise r(x) is not a function of S(x). That is, if there are two values $x_{1,j}$ and $x_{2,j}$ such that $s_j(x_{1,j}) = s_j(x_{2,j})$ and $r_j(x_{1,j}) \neq r_j(x_{2,j})$, then for other dimensions of x fixed r(x) takes on two values for the same value of S(x). Let

(A5.29) $$\rho_j(s_j) := \log\left(\frac{g_j(x_j \mid s_j(x_j) = s_j)}{f_j(x_j \mid s_j(x_j) = s_j)} \right)$$

be the log likelihood ratio associated with s_j, and note that

(A5.30) $$r(S(x)) = \exp\left(\sum_{j=1}^{d} \rho_j(s_j) \right).$$

We proceed to show that for r(S) to be a single-valued function of $\sum_{j=1}^{d} r_j(s_j)$ requires that r_j be linear and that each r_j have the same slope. To do this requires the conditions on the distributions.

Consider the lattice case first. Suppose there are two dimensions of x for which the log-likelihood ratios are not linear with the same slope. Then there are two lattice points in each marginal distribution such that

(A5.31) $$\rho_1(a + \delta) - \rho_1(a) \neq \rho_2(b + \delta) - \rho_2(b)$$

(we require that distributions have support on consecutive lattice points). Rearranging terms yields

(A5.32) $$\rho_1(a + \delta) + \rho_2(b) \neq \rho_2(b + \delta) + \rho_1(a) .$$

But then $\rho_1(s_1) + \rho_2(s_2)$ is not a single-valued function of $s_1 + s_2$, and by (A5.30) r is not a single-valued function of S. Thus all dimensions have log-likelihood ratios with the same slope, and by substitution in (A5.30) and (A5.27) we see that each marginal distribution is obtained by exponential tilting with the same slope, which gives (A5.28).

In the density case suppose there two dimensions of x for which the log-likelihood ratios are not linear with the same slope. There must be two equally-spaced points in each for which (A5.31) holds. To see this consider two cases. In case one both dimensions have linear log-likelihood ratio, but with different slopes. Choose a, b, and δ for which a + δ and a are in the support of s_1 and b and b + δ are in the support of s_2, then (A5.32) holds. In case two one of the dimensions, say x_1, has a nonlinear log-likelihood ratio. Choose $\alpha > 0$ for which $[b,b+\alpha]$ lies in the range of s_2, and choose $0 < \beta \leq \alpha$ and γ such that ρ_1 is not linear on $[\gamma-\beta, \gamma+\beta]$. If ρ_1 is not linear on the three points $\gamma-\beta$, γ, and $\gamma+\beta$ then (A5.32) holds for $\delta = \beta$ and $a = \gamma$ or $a = \gamma - \beta$. Otherwise check the triplet $(\gamma-\beta, \gamma-\beta/2, \gamma)$ for nonlinearity, and the triplet $(\gamma, \gamma+\beta/2, \gamma+\beta)$. Repeat this bisection process until three equi-spaced points are found for which ρ_1 is nonlinear. This bisection process checks points which are a dense subset of $[\gamma-\beta, \gamma+\beta]$, so eventually finds a nonlinear triple, since ρ_1 is piecewise continuous. Finally, once a nonlinear triple is found at least one of the pairs has a pairwise slope not equal to the pairwise slope for two points in the range of s_2. In both density cases the difference in pairwise slopes for two dimensions leads to a contradiction by the same argument as in the lattice case. Hence the log-likelihood ratios are all linear with the same slope, and the distribution is exponentially tilted.

References

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J, Rogers, W. H., Tukey, J. W. (1972) *Robust Estimates of Location: Survey and Advances*, Princeton University Press.

Bauwens, L. (1984), *Bayesian Full Information Analysis of Simultaneous Equation Models Using Integration by Monte Carlo*, New York: Springer Verlag.

Becker, R. A., and Chambers, J. M. (1984), *S: An Interactive Environment for Data Analysis and Graphics*, Belmont, CA: Wadsworth.

Beckman, R. J. and McKay, M. D. (1987), "Monte Carlo Estimation Under Different Distributions Using the Same Simulation," *Technometrics* 29 153-160.

Bhattacharya, R. N., and Ghosh, J. K. (1978), "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics* 6, 434-451.

Booth, T. E. (1986), "A Monte Carlo Learning/Biasing Experiment with Intelligent Random Numbers", *Nuclear Science and Engineering* 92, 465-81.

Bratley, P., Fox, B. L., and Schrage, L. E. (1983), *A Guide to Simulation*, New York: Springer Verlag.

Butler, J. W. (1956), "Machine Sampling From Given Probability Distributions", *Symposium on Monte Carlo Methods*, ed H. A. Meyer, New York: Wiley, pp. 249-264.

Chambers, J. M., Mallows, C. L., and Stuck B. W. (1976), "A Method for Simulating Stable Random Variables", *Journal of the American Statistical Association*, 71, 340-44.

Chung, K.L. (1974), *A Course in Probability*, New York: Academic Press.

Clark, F. H. (1966), "The Exponential Transform as an Importance-Sampling Device: A Review", Technical Report ORNL-RSIC-14, Oak Ridge National Laboratory.

Cochran, W. G. (1977), *Sampling Techniques*, New York: John Wiley.

Conway, A. E. & Goyal, A. (1987), "Monte Carlo Simulation of Computer System Availability/Reliability Models" *Proceedings of the Seventeenth Symposium on Fault-Tolerant Computing*, Pittsburgh, Pennsylvania, 230-235.

Cramér, H. (1945), *Mathematical Methods of Statistics*, Princeton University Press.

Daniels, H. E. (1954), "Saddlepoint Approximations in Statistics," *Annals of Mathematical Statistics*, 25, 631-650.

Davis, B. R. (1987), "An Improved Importance Sampling Method for Digital Communication System Simulations", *IEEE Transations on Communications*, COM-34, 715-719.

DiCiccio, T. and Tibshirani, R. (1987), "Bootstrap Confidence Intervals and Bootstrap Approximations," *Journal of the American Statistical Association*, 82, 163-170.

Efron, B. (1981), "Non-parametric Standard Errors and Confidence Intervals," *The Canadian Journal of Statistics*, 9, 139-172.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1987), "Better Bootstrap Confidence Intervals", *Journal of the American Statistical Association*, 82, 171-185.

Glynn, P. W. (1986), "Stochastic Approximation for Monte Carlo Optimization", *Proceedings of the 1986 Winter Simulation Conference*, 356-364

Glynn, P. W. (1987), "Likelihood Ratio Gradient Estimation: An Overview", *Proceedings of the 1987 Winter Simulation Conference*, 366-374.

Glynn, P. W., and Iglehart, D. L. (1987), "Importance Sampling for Stochastic Simulations", Technical Report No. 49, Department of Operations Research, Stanford University.

Goyal A., Heidelberger P., and Shahabuddin P. (1987) "Measure Specific Dynamic Importance Sampling for Availability Simulations" *Proceedings of the 1987 Winter Simulation Conference*, 351-357.

Hahn, P. H., Jeruchim, M. C. (1987), "Developments in the Theory and Application of Importance Sampling", *IEEE Transations on Communications*, COM-35, 706-714.

Hammersley, J. M., and Hanscomb, D. C. (1964), *Monte Carlo Methods,* London: Methuen.

Hampel, F. (1968) "Contributions to the Theory of Robust Estimation", Ph.D. Thesis, Berkeley.

Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383-393.

Hesterberg, T. C. (1987), "Importance Sampling in Multivariate Problems," *Proceedings of the Statistical Computing Section, American Statistical Association 1987 Meeting*, 412-417.

Hopmans, A. and Kleijnen, J.P.C., (1979) "Importance Sampling in Systems Simulation: a Practical Failure?," *Mathematics and Computers in Simulation* 21, 209-220.

International Mathematical and Statistical Libraries, Inc. (1984), *IMSL Library User's Manual.*

Johns, M. V. (1987), "Importance Sampling for Bootstrap Confidence Intervals," Department of Statistics, Stanford University, to appear in *Journal of the American Statistical Association.*

Johnson, N. J. (1978), "Modified t Tests and Confidence Intervals for Asymmetrical Populations," *Journal of the American Statistical Association*, 73, 536-544.

Kahn, H. (1950), Nucleonics, 6(5), 27-37 and 6(6) 60-65.

Kahn, H. and Marshall, A. W. (1953), "Methods of Reducing Sample Size in Monte Carlo Computations," *Journal of the Operations Research Society of America*, 1, 263-278.

Kioussis, L. C., and Miller, D. R. (1983), "An Importance Sampling Scheme for Simulating the Degradation and Failure of Complex Systems During Finite Missions", *Proceedings of the 1983 Winter Simulation Conference*, 631-639.

Kloek, T. and Van Dijk, H. K. (1978) "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica* 46(1) 1-19.

Kennedy, W. J. Jr. and Gentle, J. E. (1980), *Statistical Computing*, New York: Marcel Dekker.

Kleijnen, J.P.C. (1974), *Statistical Techniques in Simulation, Part 1*, New York: Marcel Dekker.

Knuth, D. E. (1981), *The Art of Computer Programming, Volume 2, Seminumerical Algorithms*, 2nd Edition, Reading Massachusetts: Addison Wesley.

Luzer, V. and Olkin, I. (1988), "Estimation of the Ordered Characteristic Roots of a Random Covariance Matrix," Technical Report, Department of Statistics, Stanford University.

Marsaglia, G. (1961), "Expressing a Random Variable in Terms of Uniform Random Variables," *Annals of Mathematical Statistics*, 32, 894-898.

Moy, W. A. (1965), "Sampling Techniques for Increasing the Efficiency of Simulations of Queueing Systems", Ph.D. dissertation, Industrial Engineering and Management Science, Northwestern University.

Murthy, K. P. N. and Indira, R. (1986), "Analytical Results of Variance Reduction Characteristics of Biased Monte Carlo for Deep-Penetration Problems," *Nuclear Science and Engineering* 92, 482-487.

Reiman, M. I., and Weiss, A. (1986), "Sensitvity Analysis Via Likelihood Ratios", *Proceedings of the 1986 Winter Simulation Conference*, 285-289.

SAS Institute, Inc. (1985), *SAS User's Guide: Basics, Version 5 Edition*, Cary, NC: SAS Institute Inc.

Siegmund, D. (1976) "Importance Sampling in the Monte Carlo Study of Sequential Tests," *The Annals of Statistics* 4, 673-684.

Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C. and Dransfield, M., (1985) "The Implementation of the Bayesian Paradigm," *Communications in Statististics-Theory and Methods*, 14, 1079-1102.

Stewart L. (1979), "Multiparameter Univariate Bayesian Analysis" *Journal of the American Statistical Association*, 74, 684-693.

Stewart L. (1983), "Multiparameter Bayesian Inference Using Monte Carlo Integraiton — Some Techniques for Bivariate Analysis," *Bayesian Statistics 2*, eds J.BM. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith, Elsevier Science Publishers B.V. (North-Holland), 495-510.

Therneau, T. M. (1983), "Variance Reduction Techniques for the Bootstrap," *Technical Report No. 200 (PhD. Thesis)* Department of Statistics, Stanford University.

Tibshirani, R. J., (1984) "Bootstrap Confidence Intervals", Technical Report LCS-3, Department of Statistics, Stanford University.

Tukey, J. W. (1987), "Configural Polysampling," *SIAM REVIEW* 29, 1-20.

Van Dijk, H.K. and Kloek, T. (1983), "Experiments with Some Alternatives for Simple Importance Sampling in Monte Carlo Integration," Bayesian Statistics 2, eds J.BM. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith, Elsevier Science Publishers B.V. (North-Holland), 511-530.

Von Neumann, J. (1951), "Various Techniques Used in Connection with Random Digits", National Bureau of Standards symposium, NBS Applied Mathematics Series 12, National Bureau of Standards, Washington, D.C.

Wilson, J. R. (1984), "Variance Reduction Techniques for Digital Simulation," *American Journal of Mathematical and Management Sciences* 4, 277-312.