@p

Btry 694 Class Homework 3

10/2/98

1 Problem 3.4

Søren Hvidkjær

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

$$\vartheta = \int_1^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

with those of an estimator based on the optimal choice of instrumental distribution.

A natural approach would be to simulate from the $\mathcal{U}_{[0,1]}$, transform to the $\mathcal{N}(0,1)$ by way of the Box-Muller algorithm, and estimate ϑ by

$$\hat{\vartheta}_1 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{X_i \in [1,2]},$$

where n is the sample size and $\mathbf{I}_{X_i \in [1,2]}$ an indicator variable. The variance of $\hat{\vartheta}_1$ is $(1-\vartheta)\vartheta/n$, or $0.117/n$ since $\vartheta = 0.1359$. An alternative estimator, taking the symmetry of the normal distribution into account would be

$$\hat{\vartheta}_1 = \frac{1}{2n} \sum_{i=1}^{n} \mathbf{I}_{|X_i| \in [1,2]},$$

with the variance $(1 - 2\vartheta)\vartheta/2n$, or $0.049/n$. A third approach would be to view ϑ as the expectation of $g(Y) = 1/\sqrt{2\pi}\, exp(-Y^2/2)$ with $Y \sim \mathcal{U}_{[1,2]}$. The estimator then becomes

$$\hat{\vartheta}_3 = \frac{1}{\sqrt{2\pi} n} \sum_{i=1}^{n} e^{-U_i^2/2},$$

where $U_i \sim \mathcal{U}_{[1,2]}$.

Simulating 2 million times in Fortran on a Sun Sparc workstation yielded the estimates given in table 1, see also separate plot. The program took about 13 clock seconds to complete. From the table, and in particular from the figure, it is apparent that $\hat{\vartheta}_3$ is the preferred estimator, being precise to the fourth decimal after 1,000,000 iterations. Whether indeed this is *the* optimal instrumental distribution is a conceptual issue, but it certainly dominates the two other estimators tested here.

Table 1:

n	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\vartheta}_3$	Exact
50000	0.13586	0.136	0.135472	0.135905
100000	0.13602	0.136365	0.135663	0.135905
500000	0.136094	0.135931	0.13581	0.135905
1000000	0.136187	0.13599	0.135858	0.135905
1500000	0.136208	0.136083	0.135895	0.135905
2000000	0.136113	0.136047	0.135905	0.135905

Note: $\hat{\vartheta}_3$ appears to be precise to the sixth decimal after 2 mill. simulations, but this is a coincidence.

2 Problem 3.8

Krishanu Maulik

In this chapter, the imporance sampling method was developed for an iid sample $(Y_1, \dots, Y_n)$ 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 $(Y_1, Y_3, \dots, Y_{2n-1})$ and a secondary sample $(Y_2, Y_4, \dots, Y_{2n})$ such that $Y_{2i} \sim q(y_{2i}|Y_{2i-1})$ and $cov(Y_{2i}, Y_{2i-1}) <$

0. Show that the covariance

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

is null. Generalize.

(a) We continue to use the same estimator

$$\frac{1}{n}\sum_{i=1}^{n} h(Y_i)\frac{f(Y_i)}{g(Y_i)}.$$

It has the expectation

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\left[h(Y_i)\frac{f(Y_i)}{g(Y_i)}\right] \\
= \; & \mathbf{E}\left[h(Y_1)\frac{f(Y_1)}{g(Y_1)}\right] \text{ (as } Y_1,\ldots,Y_n \text{ have same marginal density } g) \\
= \; & \int h(y)\frac{f(y)}{g(y)}g(y)dy \\
= \; & \int h(y)f(y)dy
\end{aligned}$$

So the estimator remains unbiased.

(b) In this case, we are sampling Y_i conditionally on Y_{i-1} using the density $q(y|Y_{i-1})$. So we change the denominators accordingly and have the estimator

$$\frac{1}{n}\left\{h(Y_1)\frac{f(Y_1)}{g(Y_1)} + \sum_{i=2}^{n} h(Y_i)\frac{f(Y_i)}{q(Y_i|Y_{i-1})}\right\},$$

where g is the marginal density of Y_1.

Now,

$$\mathbf{E}\left[h(Y_1)\frac{f(Y_1)}{g(Y_1)}\right] = \int h(y)\frac{f(y)}{g(y)}g(y)dy = \int h(y)f(y)dy$$

and for $i > 1$,

$$\begin{aligned}
& \mathbf{E}\left\{h(Y_i)\frac{f(Y_i)}{q(Y_i|Y_{i-1})}|Y_{i-1}\right\} \\
= & \int h(y)\frac{f(y)}{q(y|Y_{i-1})}q(y|Y_{i-1})dy \\
= & \int h(y)f(y)dy \qquad (1)
\end{aligned}$$

So,

$$\mathbf{E}\left\{h(Y_i)\frac{f(Y_i)}{q(Y_i|Y_{i-1})}\right\} = \int h(y)f(y)dy$$

$$\therefore \mathbf{E}\left[\frac{1}{n}\left\{h(Y_1)\frac{f(Y_1)}{g(Y_1)} + \sum_{i=2}^{n} h(Y_i)\frac{f(Y_i)}{q(Y_i|Y_{i-1})}\right\}\right] = \int h(y)f(y)dy$$

(c)

$$\begin{aligned}
& cov\left(h(Y_{2i-1})\frac{f(Y_{2i-1})}{g(Y_{2i-1})}, h(Y_{2i})\frac{f(Y_{2i})}{g(Y_{2i})}\right) \\
= & \mathbf{E}\left[\mathrm{cov}\left\{h(Y_{2i-1})\frac{f(Y_{2i-1})}{g(Y_{2i-1})}, h(Y_{2i})\frac{f(Y_{2i})}{g(Y_{2i})}|Y_{2i-1}\right\}\right] \\
+ & cov\left[\mathbf{E}\left\{h(Y_{2i-1})\frac{f(Y_{2i-1})}{g(Y_{2i-1})}|Y_{2i}\right\}, \mathbf{E}\left\{h(Y_{2i})\frac{f(Y_{2i})}{g(Y_{2i})}|Y_{2i-1}\right\}\right] \\
= & 0 + \mathrm{c}ov\left[h(Y_{2i-1})\frac{f(Y_{2i-1})}{g(Y_{2i-1})}, \int h(y)f(y)dy\right] \qquad by \quad (1) \\
= & 0
\end{aligned}$$

Note that we do not need the information that $cov(Y_{2i-1}, Y_{2i}) < 0$.

In general, suppose we have an iid sample $(Y_1, Y_2, \ldots, Y_m)$ with density g and for each i, we have a secondary sample $(Z_1^i, \ldots, Z_{n_i}^i)$, which are iid conditioned on Y_i with conditional density $q(\cdot|Y_i)$.

Then we use the estimator

$$\frac{1}{N+m}\sum_{i=1}^{m}\left\{h(Y_i)\frac{f(Y_i)}{g(Y_i)} + \sum_{j=1}^{n_i} h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}\right\}, \text{ where } N = \sum_{i=1}^{m} n_i.$$

Here too, the estimator is unbiased.

$$\begin{aligned}
& cov\left(h(Y_i)\frac{f(Y_i)}{g(Y_i)}, h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}\right) \\
= \;& \mathbf{E}\left[cov\left\{h(Y_i)\frac{f(Y_i)}{g(Y_i)}, h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}|Y_i\right\}\right] \\
+ \;& cov\left[\mathbf{E}\left\{h(Y_i)\frac{f(Y_i)}{g(Y_i)}|Y_i\right\}, \mathbf{E}\left\{h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}|Y_i\right\}\right] \\
= \;& 0 + cov\left[h(Y_i)\frac{f(Y_i)}{g(Y_i)}, \int h(y)f(y)dy\right] && by \quad (1) \\
= \;& 0
\end{aligned}$$

and,

$$cov\left(h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}, h(Z_k^i)\frac{f(Z_k^i)}{q(Z_k^i|Y_i)}\right)$$

$$
\begin{aligned}
&= \mathbf{E}\left[\mathrm{cov}\left\{h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}, h(Z_k^i)\frac{f(Z_k^i)}{q(Z_k^i|Y_i)}|Y_i\right\}\right] \\
&+ \ cov\left[\mathbf{E}\left\{h(Z_j^i)\frac{f(Z_j^i)}{q(Z_j^i|Y_i)}|Y_i\right\}, \mathbf{E}\left\{h(Z_k^i)\frac{f(Z_k^i)}{q(Z_k^i|Y_i)}|Y_i\right\}\right] \\
&= 0 + \mathrm{cov}\left[\int h(y)f(y)dy, \int h(y)f(y)dy\right] \text{ as } Z_j^i, Z_k^i \text{ are independent given } Y_i \text{ and by (1)} \\
&= 0.
\end{aligned}
$$

So the terms in the estimator are uncorrelated even in this general case.

3 Problem 3.12

Baojun Song

(a).

By the **Law of Large Number**,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{f_{XY}(x^*, y_i)w(x_i)}{f_{XY}(x_i, y_i)} = E_f(\frac{f_{XY}(x^*, Y)w(X)}{f_{XY}(X, Y)})$$

$$= \int\int \frac{f_{XY}(x^*, y)w(x)}{f_{XY}(x, y)} * f_{XY}(x, y)dxdy = \int\int f_{XY}(x^*, y)w(x)dxdy$$

$$= \int w(x)dx \int f_{XY}(x^*, y)dy = f_X(x)$$

(b).

The conditional density of X given Y is $f_{X|Y}(x|y) = G_a(y, 1) = \frac{x^{y-1}}{\Gamma(y)}e^{-x}$. The marginal density of Y is $f_Y(y) = e^{-y}$. The joint density, therefore, is

$$f_{XY}(x, y) = f_{X|Y}(x|y) * f_Y(y) = \frac{x^{y-1}}{\Gamma(y)}e^{-(x+y)}.$$

This will yield the marginal density of X is

$$f_X(x) = \int f_{XY}(x, y)dy = \int \frac{x^{y-1}}{\Gamma(y)}e^{-(x+y)}dy = e^{-x}\int \frac{x^{y-1}}{\Gamma(y)}e^{-y}dy$$

we cannot find out the explicity form for $f_X(x)$, we can, however, plot the marginal density by using *Mathematica*.

(c).

Based upon tghe spirit of Theorem 3.3.4, we may choose $w(x)$ which minimize $Var[\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)}]$. The approach is the same as the proof of that theorem.

$$Var[\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)}] = E_f(\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)})^2 - (E_f(\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)}))^2$$

The second term above is

$$(E_f(\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)}))^2 = (\int\int f_{XY}(x^*,y)w(x)dxdy)^2$$
$$= (\int w(x)dx \int f_{XY}(x^*,y)dy)^2 = (f_X(x^*))^2,$$

which is independent of $w(x)$. We only minimize the first term. Jensen's inequality is a powerful tool to work out this problem.

$$E_f(\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)})^2 \geq (E_f(\frac{f_{XY}(x^*,Y)w(X)}{f_{XY}(X,Y)}))^2 = (\int\int \frac{f_{XY}(x^*,y)w(x)}{f_{XY}(x,y)} * f_{XY}(x,y)dxdy)^2$$
$$= (\int\int f_{XY}(x^*,y)w(x)dxdy)^2 = (\int w(x)dx \int f_{XY}(x^*,y)dy)^2 = (f_X(x^*))^2.$$

The lower bound $(f_X(x^*))^2$ is independent of the choice of density function $w(x)$. If we could find a density function $w^*(x)$ that the lower bound is attained, $w^*(x)$ is the optimal choice. Generally, this is optimization problem as following,

$$min_w \int\int \frac{f^2_{XY}(x^*,y)w^2(x)}{f_{XY}(x,y)}dxdy,$$

$$subject\ to \int w(x)dx = 1, \quad and \quad w(x) \geq 0.$$

For the special case, we can work out the solution. For instance, if X and Y are indepedent, $w^*(x) = f_X(x)$ is the optimal choice.In fact,

$$\int\int \frac{f_{XY}^2(x^*, y)w^2(x)}{f_{XY}(x, y)} dxdy = \int\int \frac{f_X^2(x^*)f_Y^2(y)f_X^2(x)}{f_X(x)f_Y^2(y)} dxdy$$
$$= (f_X(x^*))^2 \int f_X(x)dx \int f_Y(y)dy = (f_X(x^*))^2.$$

This is a theoretical optimization, because we are looking for $(f_X(x)$.

4 Problem 3.13

Daniel Fink

The computation of ratios of normalizing constants arise in many statistical problems including computing Likelihood ratios for hypothesis testing, computing likelihood ratios in the presence(!) of missing (or latent) data, and the computation of Bayes factors.

In such cases the quantity of interest is $\rho = \frac{c_1}{c_2}$ where $p_i(\theta) = \frac{\tilde{p}_1(\theta)}{c_i}$ for $i = 1, 2$.

4.1 Part A

Show that ρ can be approximated by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{p}_1(\theta)}{\tilde{p}_2(\theta)}$$

where $\theta_i \sim p_2$. We know that $\int \tilde{p}_i(\theta)d\theta = c_i$ and by the LLN

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{p}_1(\theta)}{\tilde{p}_2(\theta)} &\rightarrow E_{p_2}(\frac{\tilde{p}_1(\theta)}{\tilde{p}_2(\theta)}) \\
&= \int(\frac{\tilde{p}_1(\theta)}{\tilde{p}_2(\theta)})p_2(\theta)d\theta \\
&= \frac{1}{c_2}\int \tilde{p}_1(\theta)d\theta \\
&= \frac{c_1}{c_2}.
\end{aligned}$$

4.2 Part B

Show that

$$\frac{\int \tilde{p}_1(\theta)p_2(\theta)\alpha(\theta)d\theta}{\int \tilde{p}_2(\theta)p_1(\theta)\alpha(\theta)d\theta} = \frac{c_1}{c_2}$$

holds for every function $\alpha(\theta)$ such that both integrals are finite.

Use the relationship

$$\int \tilde{p}_i(\theta)p_j(\theta)\alpha(\theta)d\theta = \frac{1}{c_j}\int \tilde{p}_i(\theta)\tilde{p}_j(\theta)\alpha(\theta)d\theta$$

to simplify both numerator and denominator:

$$\frac{\int \tilde{p}_1(\theta)p_2(\theta)\alpha(\theta)d\theta}{\int \tilde{p}_2(\theta)p_1(\theta)\alpha(\theta)d\theta} = \frac{c_1}{c_2}\frac{\int \tilde{p}_1(\theta)\tilde{p}_2(\theta)\alpha(\theta)d\theta}{\int \tilde{p}_2(\theta)\tilde{p}_1(\theta)\alpha(\theta)d\theta} = \frac{c_1}{c_2}.$$

Thus, as long as both integrals are finite, this equality holds.

4.3 Part C

Deduce that

$$\frac{\frac{1}{n_2}\sum_{i=1}^{n_2}\tilde{p}_1(\theta_{2i})\alpha(\theta_{2i})}{\frac{1}{n_1}\sum_{i=1}^{n_1}\tilde{p}_2(\theta_{1i})\alpha(\theta_{1i})}$$

is a convergent estimator of ρ where $\theta_{ji} \sim p_j$, $i = 1, ..., n_j$ and $j = 1, 2$. Again we invoke the LLN:

$$\begin{aligned}
\frac{\frac{1}{n_2}\sum_{i=1}^{n_2}\tilde{p}_1(\theta_{2i})\alpha(\theta_{2i})}{\frac{1}{n_1}\sum_{i=1}^{n_1}\tilde{p}_2(\theta_{1i})\alpha(\theta_{1i})} \quad &\to \quad \frac{E_{p_2}\tilde{p}_1(\theta)\alpha(\theta)}{E_{p_1}\tilde{p}_2(\theta)\alpha(\theta)} \\
&= \quad \frac{c_1\int \tilde{p}_2(\theta)\tilde{p}_1(\theta)\alpha(\theta)}{c_2\int \tilde{p}_1(\theta)\tilde{p}_2(\theta)\alpha(\theta)}
\end{aligned}$$

$$= \frac{c_1}{c_2}.$$

4.4 Part D

Show that part (B) covers the Newton and Raftery (1994) representation

$$\rho = \frac{E_{p_2}[\tilde{p}_2(\theta)^{-1}]}{E_{p_1}[\tilde{p}_1(\theta)^{-1}]}.$$

$$\begin{aligned}
\frac{E_{p_2}[\tilde{p}_2(\theta)^{-1}]}{E_{p_1}[\tilde{p}_1(\theta)^{-1}]} &= \frac{\int p_2(\theta)/\tilde{p}_2(\theta)d\theta}{\int p_1(\theta)/\tilde{p}_1(\theta)d\theta} \\
&= \frac{c_1}{c_2}
\end{aligned}$$

where $\alpha(\theta) = (\tilde{p}_1(\theta)\tilde{p}_2(\theta))^{-1}$.

4.5 Part E

Show that the optimal choice (in terms of MSE) of α in part B) is

$$\alpha(\theta) = c\frac{n_1 + n_2}{n_1\pi_1 + n_2\pi_2}.$$

Following Meng and Wong (1996), we look at the Relative MSE,

$$RMSE(\hat{\rho_\alpha}) = \frac{E(\hat{\rho}_\alpha - \rho)^2}{\rho^2},$$

where the expectation is taken over all random draws. The exact calculation of the RMSE depends on α, and is difficult, if not intractable because $\hat{\rho_\alpha}$

is a ratio estimator. However, with a large number of draws from each density (n_1 and n_2 from p_1 and p_2 respectively) we can approximate the RMSE with its first order term. This essentially ignores the bias term (which should be small because of CLT).

Notation:

$$A_1 = \int \tilde{p}_1(\theta)p_2(\theta)\alpha(\theta)$$
$$\hat{A}_1 = \frac{1}{n_2}\sum_{i=1}^{n_2}\tilde{p}_1(\theta_{2i})\alpha(\theta_{2i})$$
$$A_2 = \int \tilde{p}_2(\theta)p_1(\theta)\alpha(\theta)$$
$$\hat{A}_2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\tilde{p}_2(\theta_{1i})\alpha(\theta_{1i})$$

Following the derivation in Casella and Berger (1990 section 7.4.2 pg 328), the first order approximation yields,

$$RMSE(\hat{\rho_\alpha}) \approx \frac{Var(\hat{A_1})}{A_1^2} + \frac{Var(\hat{A_2})}{A_2^2} + O(\frac{1}{n^2}).$$

Substituting the "A's" back into the equation above we find that

$$RMSE(\hat{\rho_\alpha}) \approx n\frac{\int q_1q_2(q_1+q_2)\alpha^2 d\theta}{(\int q_1q_2\alpha d\theta)^2} - \frac{1}{n_1} - \frac{1}{n_2} + O(\frac{1}{n^2}),$$

where $q_i = \frac{n_i p_i}{n}$.

We want to find the function $\alpha(\theta)$ that minimizes the first order approximation to the RMSE. There are several approaches to this problem, like the Calculus of Variations or finding a minimum bound. Pursuing the latter, note that

$$(\int q_1 q_2 \alpha d\theta)^2 \leq (\int \sqrt{\frac{q_1 q_2}{q_1 + q_2}} \sqrt{q_1 q_2 (q_1 + q_2)} |\alpha|^2 d\theta)^2.$$

The Cauchy-Schwartz inequality yields

$$(\int q_1 q_2 \alpha d\theta)^2 \leq (\int \frac{q_1 q_2}{q_1 + q_2} d\theta)(\int q_1 q_2 (q_1 + q_2) \alpha^2 d\theta).$$

Thus,

$$\frac{\int q_1 q_2 (q_1 + q_2) \alpha^2 d\theta}{(\int q_1 q_2 \alpha d\theta)^2} \geq \frac{1}{\int \frac{q_1 q_2}{q_1 + q_2} d\theta},$$

where equality holds iff

$$\begin{aligned} \alpha &\propto \frac{1}{q_1 + q_2} && (2) \\ &\propto \frac{n_1 + n_2}{n_1 p_1 + n_2 p_2}. && (3) \end{aligned}$$

5 Problem 3.27

Bárbara Patricia González

Let $X \sim f$, and suppose that we want to evaluate $P(X > a) = \int_a^\infty f(x)dx$. The natural place to start is with $\delta_1 = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(X_i > a)$, where the X_i's are iid from f.

Suppose now that f is symmetric or, more generally, that for some parameter μ we know the value of $P(X > \mu)$ (where we assume that $a > \mu$). We can then take $\delta_3 = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(X_i > \mu)$ and form the control variate estimator

$$\delta_2 = \delta_1 + \beta(\delta_3 - P(X > \mu)).$$

Since $Var\delta_2 = Var\delta_1 + \beta^2 Var\delta_3 + 2\beta Cov(\delta_1, \delta_3)$, and

$$Cov(\delta_1, \delta_3) = \frac{1}{n}P(X > a)(1 - P(X > \mu))$$
$$Var\delta_3 = \frac{1}{n}P(X > \mu)(1 - P(X > \mu)),$$

it follows that δ_2 will be an improvement over δ_1 if

$$\beta < 0 \text{ and } |\beta| < \frac{2Cov(\delta_1, \delta_3)}{Var\delta_3}.$$

a) Verify that $Var\delta_3 = \frac{1}{n}P(X > \mu)(1 - P(X > \mu))$.

b) Verify the conditions on β in order for δ_2 to improve on δ_1.

c) For $f = N(0,1)$, find $P(X > a)$ for $a = 3, 5, 7$.

d) For $f = T_5$, find $P(X > a)$ for $a = 3, 5, 7$.

e) For $f = T_5$, find a such that $P(X > a) = .01, .001, .0001$.

a) We want to calculate $Var\delta_3$, where $\delta_3 = \frac{1}{n}\sum_{i=1}^{n} I\!I(X_i > \mu)$. But $Var\delta_3 = E\delta_3^2 - (E\delta_3)^2$, and

$$\delta_3^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I\!I(X_i > \mu) I\!I(X_j > \mu) \tag{4}$$

So we have that

$$E\delta_3 = \frac{1}{n}\sum_{i=1}^{n} E(I\!I(X_i > \mu)) = \frac{1}{n}\sum_{i=1}^{n} P(X > \mu) = P(X > \mu) \tag{5}$$

and

$$E\delta_3^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E(I\!I(X_i > \mu) I\!I(X_j > \mu)) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E(I\!I(X_i > \mu)) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j \neq i}}^{n} E(I\!I(X_i > \mu) I\!I(X_j > \mu)) =$$

$$= \frac{1}{n^2}(nP(X > \mu) + (n^2 - n)P(X > \mu)^2) \tag{6}$$

So we finally have

$$Var\delta_3 = \frac{1}{n^2}(nP(X>\mu)+(n^2-n)P(X>\mu)^2)-P(X>\mu)^2 = \frac{1}{n}P(X>\mu)(1-P(X>\mu)) \tag{7}$$

b) Since $Var\delta_3 \geq 0$ and $Cov(\delta_1, \delta_3) \geq 0$ then $\beta < 0$ in order for δ_2 to improve on δ_1. So, given that $\beta < 0$ we want to prove that $\beta^2 Var\delta_3 + 2\beta Cov(\delta_1, \delta_3) < 0$ if and only if $|\beta| < \frac{2Cov(\delta_1,\delta_3)}{Var\delta_3}$.

$$\beta^2 Var\delta_3 + 2\beta Cov(\delta_1, \delta_3) < 0 \Longleftrightarrow \beta^2 Var\delta_3 < -2\beta Cov(\delta_1, \delta_3) \Longleftrightarrow$$
$$\beta Var\delta_3 > -2Cov(\delta_1, \delta_3) \Longleftrightarrow \beta > \frac{-2Cov(\delta_1, \delta_3)}{Var\delta_3} \Longleftrightarrow |\beta| < \frac{2Cov(\delta_1, \delta_3)}{Var\delta_3} \tag{8}$$

c) In order to find $P(X > a)$ for $a = 3, 5, 7$ we first simulated 10^6 random variables, distributed $N(0, 1)$. And we had the following results for the acceleration method described before:

a	δ_1	μ	δ_3	β	$P(X > a)$
3	$1.3346x10^{-3}$	.5	.4999876	$-.004$	$1.33x10^{-3}$
5	$1.2x10^{-6}$	$1.33x10^{-3}$	$1.3346x10^{-3}$	$-.004$	$1.18x10^{-6}$
7	$.2x10^{-6}$	$1.18x10^{-6}$	$1.2x10^{-6}$	$-.004$	$2x10^{-7}$

d) In order to find $P(X > a)$ for $a = 3, 5, 7$ we first simulated $50,000$ random variables, distributed T_5. And we had the following results for the acceleration method described before:

a	δ_1	μ	δ_3	β	$P(X > a)$
3	$2.73x10^{-2}$	.5	$4.99x10^{-1}$	$-.03$	$2.733x10^{-2}$
5	$8.56x10^{-3}$	$2.733x10^{-2}$	$2.73x10^{-2}$	$-.03$	$8.56x10^{-3}$
7	$3.82x10^{-3}$	$8.56x10^{-3}$	$8.56x10^{-3}$	$-.03$	$3.82x10^{-3}$

e) In order to obtain a the same simulation from part (d) will be used, so we have the following results:

$P(X > a)$	μ	β	δ_3	δ_1	a
.01	.5	$-.03$	$4.99x10^{-1}$	$9.97x10^{-3}$	4.677
.001	.5	$-.003$	$4.99x10^{-1}$	$9.97x10^{-4}$	13.22
.0001	.5	$-.0003$	$4.99x10^{-1}$	$9.97x10^{-5}$	43.25

6 Problem 3.31

Jianghong Wang

For the estimator δ_2 of §3.7.3

(a) : Verify the expression (3.7.5) for ρ_i.

Proof: In accept-reject algorithm,

$$(X_1, \ldots, X_m) \sim f$$

$$(U_1, \ldots, U_N) \sim U_{[0,1]}, iid$$

$$(Y_1, \ldots, Y_N) \sim g, iid$$

accept $X_j = Y_j$ if $U_j \leq w_j$, where $w_j = \frac{f(Y_j)}{Mg(Y_j)}$. m Y_j's are accepted with different w_j since Y_j and U_j are random variables.

N is the stopping time associated with the acceptance of m variables Y_j, that is the Y_N must be accepted as X_m, there are totally (m-1) Y_j's accepted as X_j before Y_N.

$$\begin{aligned}
\rho_i &= P(U_i \leq w_i | N = n, Y_1, \ldots, Y_n) \\
&= \frac{P(U_i \leq w_i, N = n, Y_1, \ldots, Y_n)}{P(N = n, Y_1, \ldots, Y_n)}.
\end{aligned}$$

where

$$\text{Numerator } = P(U_i \leq w_i, N = n, Y_1, \ldots, Y_n)$$

which is the Probability(the Y_N is accepted as X, the Y_i is accepted as X, there are (m-2) X's are choosen from the (n-2) Y_j's)

$$\text{P(each } Y_j \text{ is accepted)} = P(U_j \leq w_j) = w_j$$

thus,

$$\text{Numerator} = w_i \sum_{(i_1,\ldots,i_{m-2})} \Pi_{j=1}^{m-2} w_{i_j} \Pi_{j=m-1}^{n-2} (1 - w_{i_j})$$

where

(i). w_i denotes the probability that the ith Y is accepted as X.

(ii). $\Pi_{j=1}^{m-2} w_{i_j}$ denotes the probability that in N Y_j's , in addition to the Y_N and Y_i are accepted as X, there are other $(m-2)$ Y_j's are accepted as X, their joint probability is $\Pi_{j=1}^{m-2} w_{i_j}$.

(iii). $\Pi_{j=m-1}^{n-2}(1 - w_{i_j})$ denotes the probability that there are totally (n-m) Y_j's are rejected, since the Y_i and Y_N are accepted, the sum of j is from m-1 to n-2.

(iv). Of cause P(Y_N is accepted)=1.

(v) The numerator sum is over all subsets of (1, ..., i-1, i+1, ..., n) since except the Y_i and Y_N are fixed, other (m-2) Y_j's are choosen freely from (n-2)

Y_j's.

Similarly, we can explain the denominator.

$$\begin{aligned} \text{Denominator} &= P(N=n, Y_1, \ldots, Y_n) \\ &= \sum_{(i_1,\ldots,i_{m-1})} \Pi_{j=1}^{m-1} w_{i_j} \Pi_{j=m}^{n-1}(1-w_{ij}) \end{aligned}$$

which is the Probability (the Y_N is accepted as X_m, the other (m-1) X_j's are freely choosen from (n-1) Y_j's).

Of cause, it should also times P(the Y_N is accepted)=1.

thus,

$$\begin{aligned} \rho_i &= P(U_i \leq w_i | N=n, Y_1, \ldots, Y_n) \\ &= w_i \frac{\sum_{(i_1,\ldots,i_{m-2})} \Pi_{j=1}^{m-2} w_{i_j} \Pi_{j=m-1}^{n-2}(1-w_{i_j})}{\sum_{(i_1,\ldots,i_{m-1})} \Pi_{j=1}^{m-1} w_{i_j} \Pi_{j=m-1}^{n-1}(1-w_{i_j})} \end{aligned}$$

(b) Verify the recursion (3.7.6).

$$S_k(m) = w_m S_{k-1}(m-1) + (1-w_m) S_k(m-1)$$

$$S_k^i(m) = w_m S_{k-1}^i(m-1) + (1-w_m) S_k^i(m-1)$$

Proof:

$$S_k(m) = \sum_{(i_1,\ldots,i_k)} \Pi_{j=1}^{k} w_{i_j} \Pi_{j=k+1}^{m} (1 - w_{i_j})$$

(i). $S_k(m)$=P(any k Y_j's are accepted as X in totally m Y_j's, (m-k) Y_j are rejected, each Y_j is accepted with probability w_{i_j}, is rejected with probability $(1 - w_{i_j})$).

(ii). $S_{k-1}(m-1)$=P(any (k-1) Y_j are accepted as X in totally (m-1) Y_j)

(iii). $S_k(m-1)$=P(any k Y_j are accepted as X in totally (m-1) Y_j).

(iv). w_m=P(the Y_m is accepted as X).

(v). $(1 - w_m)$=P(Y_m is rejected).

Let $S_k(m) = P(A)$, then event A can be decomposed into two independent disjoint events B, C.

B=(the Y_m is accepted, select (k-1) Y_j accepted as X from (m-1) Y_j before Y_m).

C=(the Y_m is rejected, select k Y_j from (m-1) Y_j before Y_m).

Thus,

$$P(B) = w_m S_{k-1}(m-1)$$

$$P(C) = (1 - w_m) S_k(m-1)$$

so,

$$S_k(m) = w_m S_{k-1}(m1) + (1 - w_m) S_k(m-1)$$

$S_k^i(m) = S_k(m-1)$=P(any k Y_j's are accepted in (m-1) Y_j)=p(A)

$S_{k-1}^i(m-1) = S_{k-1}(m-2)$=P(any (k-1) Y_j's are acceted in (m-2) Y_j's).

$S_k^i(m-1) = S_k(m-2)$=P(any k Y_j's are accepted in (m-2) Y_j's).

We can decompose event A into two independent disjoint events. We can imagine m Y_j's $Y_1, \ldots, Y_{m-1}, Y_m$, we select k Y_j's accepted from m Y_j's but discard the Y_{m-1} and don't consider it, thus it is equivalent to two independent disjoint events B, C.

B=the Y_m is accepted, discard the Y_{m-1}, select (k-1) Y_j's accepted from the (m-2) Y_j's (Because 2 Y_j are deleted, one is Y_m, the other is Y_{m-1}).

C=the Y_m is rejected, discard the Y_{m-1}, select k Y_j accepted from the (m-2) Y_j.

thus,

$$P(A) = S_k^i(m)$$

$$P(B) = w_m S_{k-1}(m-2) = w_m S_{k-1}^i(m-1)$$

$$P(C) = (1-w_m) S_k(m-2) = (1-w_m) S_k^i(m-1)$$

so,

$$P(A) = P(B) + P(C)$$

$$S_k^i(m) = w_m S_{k-1}^i(m-1) + (1-w_m) S_k^i(m-1)$$

(c) Prove Proposition 3.7.6

Proof:

$$\delta_1 = \frac{1}{m}\sum_{i=1}^{m} h(X_i) = \frac{1}{m}\sum_{j=1}^{N} h(Y_j)1_{U_j<=w_j}$$

$$\delta_2 = \frac{1}{m}\sum_{j=1}^{N} E[1_{U_j\le w_j}|N, Y_1, \ldots, Y_N]h(Y_j) = \frac{1}{m}\sum_{i=1}^{N} \rho_i h(Y_i)$$

Since, $E[E(X|Y)] = EX$, thus,

$$\begin{aligned}
E\delta_2 &= E(\frac{1}{m}\sum_{j=1}^{N} E[1_{U_j\le w_j}|N, Y_1, \ldots, Y_N]h(Y_j)) \\
&= \frac{1}{m}\sum_{j=1}^{N} E(E[1_{U_j\le w_j}|N, Y_1, \ldots, Y_N]h(Y_j)]) \\
&= \frac{1}{m}\sum_{j=1}^{N} E(1_{U_j\le w_j})h(Y_j) \\
&= E[\frac{1}{m}\sum_{j=1}^{N} h(Y_j)1_{U_j\le w_j}] \\
&= E\delta_1.
\end{aligned}$$

Under quadratic loss, the risk of δ_1 and δ_2 are:

$$\begin{aligned}
R(\delta_1) &= E[\delta_1 - Eh(X)]^2 \\
&= E[\delta_1^2 + E(Eh(X))^2 - 2\delta_1 Eh(X)] \\
&= E\delta_1^2 + E(Eh(X))^2 - 2E\delta_1 Eh(X) \\
&= Var\delta_1 - (E\delta_1)^2 + E(Eh(X))^2 - 2E\delta_1 Eh(X)
\end{aligned}$$

$$
\begin{aligned}
R(\delta_2) &= E[\delta_2 - Eh(X)]^2 \\
&= E[\delta_2^2 + E(Eh(X))^2 - 2\delta_2 Eh(X)] \\
&= E\delta_2^2 + E(Eh(X))^2 - 2E\delta_2 Eh(X) \\
&= Var\delta_2 - (E\delta_2)^2 + E(Eh(X))^2 - 2E\delta_2 Eh(X)
\end{aligned}
$$

Since $E\delta_1 = E\delta_2$, we only need to compare $Var\delta_1$ and $Var\delta_2$. From the definition of δ_1 and δ_2, we have:

$$\delta_2(X) = E[\delta_1(X)|Y]$$

we know that,

$$Var(E[X|Y]) <= Var(X)$$

so,

$$Var(E[\delta_1|Y]) = Var\delta_2 <= Var\delta_1$$

thus,

$$R(\delta_1) <= R(\delta_2)$$

Then, the estimator $\delta_2 = \frac{1}{m}\sum_{i=1}^{N}\rho_i h(y_i)$ dominates the estimator δ_1 under quadratic loss.

7 Problem 3.48

Mitch Warachka

The advantage of the stratified estimator is that allows observations to be assigned to segments of the integral which have the most variablilty. Imagine the integral partitioned into p segments with unequal variablity. Rather than allocate the same number of draws to each partition, sample more frequently from partitions with greater variance to improve the variance of the overall estimation. Thus, the relationship below for relating the number of observations to variablility of the partition

$$n_j \propto p_j^2 \int_{\chi_j} (h(x) - I_j)^2 f_j(x) dx \tag{9}$$

with I_j defined as

$$\int_{\chi_j} h(x) f_j(x) dx \tag{10}$$

here the $f_j(x)$ just means the f distribution restricted to the interval χ_j. In the stratified estimator given in the question, n_j is set equal to n. Thus, no improvement is possible, it is as if the varaince of each partiton were the same. The $\propto$ can be replaced with some constant and an equality. Using the fact that the χ_j's form a partition, the following can be done

$$
\begin{aligned}
&\sum_{j=1}^{p} \frac{1}{n_j} p_j^2 \int_{\chi_j} (h(x) - I_j)^2 f_j(x) dx \\
&\qquad = \frac{1}{n} \sum_{j=1}^{p} \left(\int_{\chi_j} f(x) dx \right)^2 \int_{\chi_j} (h(x) - I_j)^2 f_j(x) dx \\
&\qquad = \frac{1}{np} \left(\int_{\chi} f(x) dx \right)^2 \int_{\chi} (h(x) - I)^2 f(x) dx \\
&\qquad = \frac{1}{np} \int_{\chi} (h(x) - I)^2 f(x) dx
\end{aligned}
$$

which is exactly the variance for the normal Monte Carlo Estimator. This would not be possible if $n_j \neq n$. If the p_j's are unknown and must be estimated the formula for conditional variance written below can be employed to show there is no improvement

$$
Var(x) = Var(E[x|y]) + E[Var(x|y)] \tag{11}
$$

where the y's are the p_j's and denumerate the various partitions. These can be thought of as indicating what partition is under consideration with a multinomial distribution. The x represents the integral of χ_j. The $Var(x|y)$ is the variance due to the partition with $E[Var(x|y)]$ being its mean. The $Var(E[x|y])$ is the variance of the former accross the partitions. Since $E[Var(x|y)]$ is nonnegative, having the p_j's a unknowns can't improve variance estimation.