A. Mira, P. Tenconi, D. Bressanini

# Variance reduction for MCMC

2003/29

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

# Variance reduction in MCMC

Antonietta Mira[*], Paolo Tenconi [†], Dario Bressanini[‡]
Università dell'Insubria, Varese, Italy.

**Abstract**

We propose a general purpose variance reduction technique for MCMC estimators. The idea is obtained by combining standard variance reduction principles known for regular Monte Carlo simulations (Ripley, 1987) and the Zero-Variance principle introduced in the physics literature (Assaraf and Caffarel, 1999). The potential of the new idea is illustrated with some toy examples and an application to Bayesian estimation.
**Keywords**: Markov chain Monte Carlo, Metropolis-Hastings algorithm, Variance reduction, Zero-Variance principle.

## 1 Main idea

We are interested in estimating the expected value of a function $f$ with respect to a, possibly unnormalized, probability distribution $\pi$:

$$\mu_f = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx}. \tag{1}$$

Markov chain Monte Carlo methods (MCMC, Metropolis et al. 1953, Hastings 1970, Tierney, 1994) estimate integrals using a large but finite set of sample points, $x^i, i = 1, \cdots, N$ collected along the path of an ergodic Markov chain, $P$, having $\pi$ (normalized) as its unique stationary and limiting distribution:

$$\hat{\mu}_f = \frac{1}{N}\sum_{i=1}^{N} f(x^i). \tag{2}$$

We have that

$$\mu_f = \hat{\mu}_f + \Delta\mu_f$$

where $\Delta\mu_f$ is the statistical error associated with the fact that $N$ is finite. For large enough $N$, standard statistical arguments lead to the following expression of the error:

$$\Delta\mu_f = K\frac{\sigma_f}{\sqrt{N}}$$

*Dip. di Economia Università dell'Insubria, Via Ravasi 2, 21100 Varese, Italy.
Email: `antonietta.mira@uninsubria.it`
†Istituto di Finanza, Università della Svizzera Italiana, Via Buffi 1, 6900 Lugano, Svizzera.
Email: `paolo.tenconi@lu.unisi.ch`
‡Dip. di Sc. Chimiche Fisiche e Matematiche, Università dell'Insubria, Via Lucini 3, 22100 Como, Italy.
Email: `dario.bressanini@uninsubria.it`

where the constant $K$ is proportional to the amount of correlation along the sampled chain and $\sigma_f$ is the standard deviation of $f$ under $\pi$ (assumed to be finite).

Recent literature (Peskun, 1973; Liu, 1996; Tierney, 1998; Tierney and Mira, 1999; Mira and Geyer, 2000; Green and Mira, 2001) aimed to reduce the statistical MCMC error, $\Delta\mu_f$, by reducing the correlation along the Markov chain, that is, by reducing $K$.

In this paper we suggest instead to reduce the error by replacing $f$ with a different function, $\tilde{f}$, obtained by properly renormalizing $f$. The function $\tilde{f}$ is constructed so that its expectation under $\pi$ equals $\mu$ (this is a standard variance reduction technique used in Monte Carlo simulation, see Ripley, 1987). To define $\tilde{f}$ an operator, $H$, and a trial function $\phi$ are introduced. We require that $H$ is Hermitian (symmetric for finite state spaces, and real in all practical applications) and

$$\int H(x,y)\sqrt{\pi(y)}dy = 0. \tag{3}$$

The trial function $\phi(x)$ is a rather arbitrary function which is only required to be integrable. We define the renormalized function to be

$$\tilde{f}(x) = f(x) + \frac{\int H(x,y)\phi(y)dy}{\sqrt{\pi(x)}} = f(x) + \Delta f(x). \tag{4}$$

As a consequence of (1) and (3) we have that

$$\mu_f = \mu_{\tilde{f}} \tag{5}$$

that is, both functions $f$ and $\tilde{f}$ can be used to estimate the desired quantity. However, the statistical error of the resulting MCMC estimator can be very different. The optimal choice for $(H, \phi)$ can be obtained by imposing that $\tilde{f}$ is constant and equal to its average that is by requiring

$$\sigma_{\tilde{f}} = 0$$

which is equivalent to require that

$$\tilde{f} = \mu_f.$$

The latter, together with (4), leads to the fundamental equation:

$$\int H(x,y)\phi(y)dy = -\sqrt{\pi(x)}[f(x) - \mu_f]. \tag{6}$$

In most practical applications equation (6) cannot be solved exactly still, we propose to find an approximate solution in the following way. First choose $H$ verifying (3) (in the sequel we will suggest two general recipe to construct $H$). Second, parametrize $\phi$ and optimally choose the parameters by minimizing $\sigma_{\tilde{f}}$ over a finite set of points generated according to $P$. Finally, a much longer MCMC simulation is performed using $\hat{\mu}_{\tilde{f}}$ instead of $\hat{\mu}_f$ as the estimator.

# 2 Choice of H

## 2.1 Discrete case

Denote with $P(x,y)$ a transition kernel reversible with respect to $\pi$:

$$\pi(x)P(x,y) = \pi(y)P(y,x) \qquad \forall x, y.$$

The following choice of $H$

$$H(x,y) = \sqrt{\frac{\pi(x)}{\pi(y)}}[P(x,y) - \delta(x-y)]$$

satisfies the requirements, where $\delta(x-y)$ is the Dirac delta function: $\delta(x-y) = 1$ if $x = y$ and zero otherwise.

With this choice of $H$, letting $\tilde{\phi} = \frac{\phi}{\sqrt{\pi}}$, equation (4) becomes:

$$\tilde{f}(x) = f(x) - \int P(x,y)[\tilde{\phi}(x) - \tilde{\phi}(y)]dy. \tag{7}$$

The main difficulty with (7) is the evaluation of the integral.

### 2.2 Continuous case

If $x \in \Re^d$ we can also consider the operator:

$$H = -\frac{1}{2}\sum_{i-1}^{d}\frac{\partial^2}{\partial x_i^2} + V(x) \tag{8}$$

where $V(x)$ is constructed to fulfill equation (3):

$$V(x) = \frac{1}{2\sqrt{\pi(x)}}\sum_{i=1}^{d}\frac{\partial^2\sqrt{\pi(x)}}{\partial x_i^2}. \tag{9}$$

In this setting we have that

$$\tilde{f}(x) = f(x) + \frac{H\phi(x)}{\sqrt{\pi(x)}}. \tag{10}$$

This is the function we will use in the examples considered in Section 5. To obtain the first and second order derivatives we used the R function “fdHess” from the library “nlme” which evaluates an approximate Hessian and gradient of a scalar function using finite differences.

## 3 Choice of $\phi$

The optimal choice of $\phi$ is the *exact solution* of the fundamental equation. In real applications typically only *approximate solutions*, obtained by minimizing $\sigma_{\tilde{f}}$, are available. The particular form of $\phi$ is very dependent on the problem at hand, on $f$ and on $H$. However an important point to notice is that if we parametrize $\phi$ in terms of $c = \int \phi(x)dx$ and then minimize $\sigma_{\tilde{f}}$ with respect to $c$, the optimal choice of $c$ is

$$c = -\frac{[E_\pi(f(x)\Delta f(x))]^2}{E_\pi(\Delta f(x))^2}$$

and, for this value of the parameter, from (4) we obtain

$$\sigma_{\tilde{f}}^2 = \sigma_f^2 - \frac{[E_\pi(f(x)\Delta f(x))]^2}{E_\pi(\Delta f(x))^2}. \tag{11}$$

Since the correction factor in (11) that leads from $\sigma_f^2$ to $\sigma_{\tilde{f}}^2$ is always negative, regardless of the choice of $\phi$, a variance reduction in the MCMC estimator is obtained by replacing $f$ with $\tilde{f}$ in (2).

The R function "optim" is used to estimate the parameters of $\phi$ minimizing $\sigma_{\tilde{f}}$.

# 4 Variance reduction in Bayesian inference

Bayesian inference is based on the posterior distribution, that is the distribution of the parameters of interest, $\theta \in R^d$, given the data, $y$. A model is postulated for the data, $l(y|\theta)$ and a prior distribution is assumed for the parameters, $h(\theta)$. Applying Bayes theorem we obtain that the posterior distribution, $\pi(\theta|y)$, is proportional to the product of the prior times the likelihood:

$$\pi(\theta|y) \propto h(\theta)l(y|\theta).$$

The normalizing constant of this distribution is a possibly complicated integral over a $d$-dimentional state space. Depending on the loss function adopted, different ways of synthesizing the posterior distribution are available. In particular we will focus on the square error loss function which leads to estimating a parameter of interest, say $\theta_j$, via its posterior mean:

$$\hat{\theta}_j = E[\theta_j|y], \qquad j = 1, \cdots, d.$$

In the notation of Section 1 we have that the role of $\pi(x)$ is payed, in the Bayesian setting, by $\pi(\theta|y)$ and, when adopting a square error loss function, we are mainly interested in evaluating the expected value with respect to $\pi$ of the family of functions defined as

$$f_j(\theta|y) = \theta_j, \qquad j = 1, \cdots, d.$$

The advantage is that these are typically one dimentional functions.

# 5 Examples

In this section we present a few toy examples to demonstrate the power of the proposed technique. In particular we first consider as target distributions, $\pi$, a univariate and bivariate Gaussian and a Student-T distribution. Finally a simple Bayesian model with conjugate prior is studied. The functions of interest, $f$, are either the mean, the variance or the covariance of the distributions of interest. Since all the target distributions considered in the examples are standard there is no need for MCMC simulation and we can resort to simple Monte Carlo. In the results presented we sample $T = 100$ values from $\pi$ and minimize the fluctuations of $\tilde{f}$ over this set of points finding the optimal values of the parameters entering $\phi$. We then estimate the mean of the target, via $\hat{\mu}_{\tilde{f}}$ over a new Monte Carlo simulation of length $N = 1\,000$. In Table 11 the estiamted parameters for each of the $\phi$ functions used are reported. An empty cell means that the corresponding parameter was not present in the analytic form of that specific function. A value of zero in a cell mean that the corresponding parameter has been estimated to be less than 0.0001.

## 5.1 Gaussian distribution

Consider, as a toy example, the case where $\pi(x) = exp(\frac{-x^2}{2})$, a standard normal distribution ($d = 1$). Let the functions of interest be $f_1(x) = x$, $f_2(x) = x^2$, that is we are interested in evaluating the first and second moment of the distribution. Let

$$\phi_1(x) = k_1 + k_2(x - k_3)\exp\{k_4(x - k_5)^2\}$$

and

$$\phi_2(x) = k_1 + k_2(x - k_3)^2 \exp\{k_4(x - k_5)^2\}$$

The functional form for $\phi$ is derived by obtaining the exact solution of (6) which, in this case is available, and considering the leading term of it. In Figure 1 and 2 we present the kernel density estimation of the empirical distribution functions obtained by Monte Carlo simulation of $f_1$, $\tilde{f}_1$ and $f_2$, $\tilde{f}_2$ respectively. Note that, in both cases, the distribution of $\tilde{f}$ is much more concentrated (almost point-mass) around the actual value of the parameter we are estimating. This is confirmed by Tables 1 and 2 where the mean and the variance of the empirical distribution of $f_1$, $\tilde{f}_1$ and $f_2$, $\tilde{f}_2$ are reported: both functions lead to unbiased estimators but $\tilde{f}$ presentes a much smaller variance and thus a smaller asymptotic mean square error. In Figure 3 and 4 we represent the functions $f, \phi, \Delta f$ and $\tilde{f}$ for the case $f(x) = x$ and $f(x) = x^2$ respectively. Notice that the goal of turning $f$ into an almost constant function, $\tilde{f}$, having the same mean is achieved.

Table 1: Normal target, $f_1(x) = x$, almost optimal $\phi$.

| | $f_1$ | $\tilde{f}_1$ |
|---|---|---|
| mean | 0.0065 | 0.0005 |
| variance | 0.94 | 0.008 |

Table 2: Normal target, $f_2(x) = x^2$, almost optimal $\phi$.

| | $f_2$ | $\tilde{f}_2$ |
|---|---|---|
| mean | 0.9018 | 1.0000 |
| variance | 1.66 | 2.80e-11 |

## 5.2 Student-T distribution

In this section we proceed as in the previous one but taking the Student-T distribution with $df$ degrees of freedom, $T(df)$, as the target distribution. Again we take $f_1(x) = x$ and $f_2(x) = x^2$ and consider the same auxiliary functions $\phi_1$ and $\phi_2$ as before (here we did not try to solve equation (6)). Figures 5, 6 and the Tables 3, 4, treat to the case of 5 degrees of freedom and refer to $f_1$ and $f_2$ respectively. Again the variance reduction obtained by substituting $f$ with $\tilde{f}$ is apparent, even if here the reduction is not as dramatic as in the Gaussian case where the optimal $\phi$ function was derived analytically.

Table 3: Student-T(5), $f_1(x) = x$.

| | $f_1$ | $\tilde{f}_1$ |
|---|---|---|
| mean | 0.0084 | 0.0174 |
| variance | 1.6910 | 0.7620 |

Table 4: Student-T(5), $f_2(x) = x^2$.

| | $f_2$ | $\tilde{f}_2$ |
|---|---|---|
| mean | 1.7570 | 1.6182 |
| variance | 11.7556 | 0.1146 |

In this setting we also performed a random walk Metropolis-Hastings algorithm (let $P$ be the corresponding transition kernel) to check that the substitution of $f$ with $\tilde{f}$ would not increase the integrated autocorrelation time $\tau = \sum_{k=-\infty}^{\infty} \rho_k$ where $\rho_k = \mathrm{Cov}_P[f(X^0), f(X^k)]/\sigma_f^2$. To estimate $\tau$ we used Sokal's adaptive truncated periodogram estimator (Sokal, 1989), $\hat{\tau} = \sum_{|k| \leq M} \widehat{\rho}_k$ with the window width $M$ chosen adaptively as the minimum integer with $M \geq 3\hat{\tau}$.

We used different the random walk proposals, namely normal distributions centered at the current position and with standard deviations, $\sigma_{RW}$, equal to 0.1; 0.2; 0.5; and 1 respectively. The results, obtained averaging 10 Monte Carlo simulations (we report mean and standard deviations in parenthesis) are presented in Table 5 and 6 for $f_1$ and $f_2$ respectively.

Table 5: Estimated $\tau$ for Student-T(5), $f_1(x) = x$.

| $\hat{\tau}$ | $\sigma_{RW} = 0.1$ | $\sigma_{RW} = 0.2$ | $\sigma_{RW} = 0.5$ | $\sigma_{RW} = 1$ |
|---|---|---|---|---|
| $f_1$ | 100.16 (33.2) | 80.39 (34.1) | 45.23 (23.1) | 13.32 (7.2) |
| $\tilde{f}_1$ | 79.73 (19.8) | 35.45 (21.6) | 18.48 (18.1) | 9.23 (7.3) |

It is clear, from this Monte Carlo study, that, at least in this setting (the same is true for all the examples presented in the sequel, simulation results not reported here), substituting $f$ with $\tilde{f}$ leads to a substantial gain in terms of asymptotic variance reduction of the resulting MCMC estimator.

## 5.3 Bivariate Gaussian distribution

We consider here a two dimentional vector, $x = (x_1, x_2)$, having a bivariate normal distribution where both marginals are standard normal and the correlation is 0.5. As functions of interest we take $f_1(x) = x_1$, $f_2(x) = x_1^2$ and $f_3(x) = x_1 x_2$. As for the $\phi$ function we take

$$\phi_1(x) = k_1(x_1 - k_2)\exp\{k_3(x_1 - k_4)^2\} + k_5 \qquad \phi_2(x) = k_1(x_1 - k_2)^2 \exp\{k_3(x_1 - k_4)^2\} + k_5$$

and

$$\phi_3(x) = k_1 + [k_2 x_1 + k_3 x_2 + k_4(x_1 x_2)]\exp[k_5(x_1 - k_8)^2 + k_6(x_2 - k_9)^2 + k_7(x_1 x_2)]$$

Table 6: Estimated $\tau$ for Student-T(5), $f_2(x) = x^2$.

| $\hat{\tau}$ | $\sigma_{RW} = 0.1$ | $\sigma_{RW} = 0.2$ | $\sigma_{RW} = 0.5$ | $\sigma_{RW} = 1$ |
|---|---|---|---|---|
| $f_2$ | 79.14 (20.8) | 63.66 (32.5) | 23.84 (11.5) | 14.18 (14.5) |
| $\tilde{f}_2$ | 1.86 (2.3) | 8.17 (20.7) | 1.30 (0.36) | 2.58 (2.3) |

respectively.

In Figures 7, 8, 9 and Tables 7, 8, 9 we report the kernel density estimates of the distributions of $f$ and $\tilde{f}$ and the relative means and variances. Again, in all cases, the distribution function estimated from the Monte Carlo simulation is more concentrated around the true value when $f$ is substituted with $\tilde{f}$.

Figures 10 and 11, where the 99 % confidence interval of the target distribution and the contour plots of $f$ and $\tilde{f}$ are presented, show how $\tilde{f}$ is more flat than $f$ in the relevant region of the state space.

Table 7: Bivariate normal, $f_1(x) = x_1$.

| | $f_1$ | $\tilde{f}_1$ |
|---|---|---|
| mean | -0.0376 | 0.0002 |
| variance | 0.9483 | 0.0072 |

Table 8: Bivariate normal, $f_2(x) = x_1^2$.

| | $f_2$ | $\tilde{f}_2$ |
|---|---|---|
| mean | 1.0161 | 0.9824 |
| variance | 2.1313 | 1.6956 |

## 5.4 Bayesian model

Consider the following model for $s$ iid observations $y_i$:

$$l(y_i|\theta) \sim N(\theta, \sigma_y^2) \qquad i = 1, \cdots, s$$

where $\sigma_y^2$ is the known variance and $\theta$ is the parameter of interest. We assume a conjugate Normal prior:

$$h(\theta) \sim N(\mu_\theta, \tau_\theta^2)$$

where $\mu_\theta$ and $\tau_\theta^2$ are known hyperparameters. It is well known that posterior distribution of the parameter of interest is

$$\pi(\theta|y_1, \cdots, y_s) = N(\mu_\pi, \sigma_\pi^2)$$

where

$$\mu_\pi = \frac{\mu_\theta \sigma_y^2 + s\tau_\theta^2 \overline{y}}{\sigma_y^2 + s\tau_\theta^2}$$

Table 9: Bivariate normal, $f_3(x) = x_1 x_2$.

| | $f_3$ | $\tilde{f}_3$ |
|---|---|---|
| mean | 0.504 | 0.502 |
| variance | 1.324 | 0.249 |

and

$$\sigma_\pi^2 = \frac{\sigma_y^2 \tau_\theta^2}{\sigma_y^2 + s\tau_\theta^2}$$

here $\overline{y}$ is the sample mean. In this setting we considered $f(\theta) = \theta$ and:

$$\phi(\theta) = \phi_1(\theta).$$

As a concrete example we used $\sigma_y = 3, \mu_\theta = 0, \tau_\theta = 3$ and generated the actual sample of size $s = 10$ from a Gaussian distribution with mean equal to one and standard deviation equal to 3. The posterior distribution has $\mu_\pi = 1.7487$ and $\sigma_\pi = 0.904$. In Figure 12 we report the empirical distribution function of $f$ and $\tilde{f}$. The mean and standard deviations of these distributions are presented in Table 10. Again, the advantage in terms of variance reduction of the method proposed is clear. Notice that, from Table 11, the estimated values of the parameters $k_1$ and $k_5$ are different from zero, while, in a similar setting (same $f$ and $\phi$), for the standard gaussian distribution these location parameters where estimated to be approximately equal to zero.

Table 10: Bayesian model, $f(\theta) = \theta$, $N = 500$.

| | $f$ | $\tilde{f}$ |
|---|---|---|
| mean | 1.7736 | 1.7399 |
| variance | 0.8838 | 0.0362 |

Table 11: Estimated parameters for $\phi$ functions

| model | functions | $k_1$ | $k_2$ | $k_3$ | $k_4$ | $k_5$ | $k_6$ | $k_7$ | $k_8$ | $k_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| N(0,1) | $f(x) = x$ | 0 | -1.99 | -0.01 | -0.50 | 0 | | | | |
| N(0,1) | $f(x) = x^2$ | 0 | -0.99 | 0 | -0.25 | 0 | | | | |
| T(5) | $f(x) = x$ | -0.036 | -1.162 | -0.105 | -0.176 | 0.012 | | | | |
| T(5) | $f(x) = x^2$ | -2.621 | -0.197 | 0.017 | -0.033 | -0.019 | | | | |
| biv. normal | $f(x) = x_1$ | 0 | -1.498 | -1.693 | -0.333 | 0 | | | | |
| biv. normal | $f(x) = x_1^2$ | 0 | -0.750 | 0 | -0.333 | 0 | | | | |
| biv. normal | $f(x) = x_1 x_2$ | -0.080 | 0.073 | 0.011 | -0.920 | -0.326 | -0.325 | 0.375 | 0.049 | -0.013 |
| Bayesian | $f(x) = x$ | 0.262 | 1.831 | -4.037 | -0.309 | 1.197 | | | | |

# 6 Rao-Blackwellization

Rao-Blackwellization (Casella and Robert, 1996) can be seen as a special case of the variance reduction technique proposed in this paper. The idea is to replace $f(x^i)$ in $\hat{\mu}$ by a conditional expectation, $E_\pi[f(x^i)|h(x^i)]$, for some function $h$ or to condition on the previous value of the chain thus using $E[f(x^i)|x^{i-1} = x]$ instead. Changing an expectation with a conditional expectation naturally reduces the variance of the resulting MCMC estimator. The functions $E_\pi[f(x^i)|h(x^i)]$ and $E[f(x^i)|x^{i-1} = x]$ can be considered as special instances of $\tilde{f}$ which do not minimize $\sigma_{\tilde{f}}$ but certainly reduce it. This suggests general guidelines that can be adopted to construct $\phi$ based on which we obtain $\tilde{f}$. In real applications, typically $E_\pi[f(x^i)|h(x^i)]$ or $E[f(x^i)|x^{i-1} = x]$ are not available in closed form, still, the researcher may have some intuition on the parametric form of such functions (or estimate them via pilot runs of the Markov chain). This intuition might aid the design of $\phi$.

# 7 Conclusions

We have presented a general purpose variance reduction technique which has been originally suggested in the physics literature (Assaraf and Caffarel, 1999). The extent by which the variance of MCMC estimators can be reduced is illustrated by some toy examples. Connections with the Rao-Blackwellization principle known in the MCMC literature are explored and exploited to better apply the zero-variance technique in a Bayesian setting.

# References


Assaraf, R. and Caffarel, M. (1999), Zero-Variance principle for Monte Carlo Algorithms. *Physical Review letters*, **83**, (23), 4682–4685.

Casella, G. and Robert, C. P. (1996), Rao-Blackwellization of Sampling Schemes. *Biometrika*, **83** (1), 81—94.

Green, P. J. and Mira, A. (2001), Delayed Rejection in Reversible Jump Metropolis-Hastings. *Biometrika*, **88**, 1035–1053.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Liu, J. S. (1996), Peskun theorem and a modified discrete-state Gibbs sampler. *Biometrika*, **83** 681–682.

Metropolis, N. and Rosenblutt, N. and Rosenblutt, A. W. and Teller, M. N. and Teller, A. H. (1953), Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, **21**, 1087-1092.

Mira, A. and Geyer, C. J. (2000), On non-reversible Markov chains. *Fields Inst. Communic.: Monte Carlo Methods*, **26** 93–108.

Peskun, P. H. (1973), Optimum Monte Carlo sampling using Markov chains. *Biometrika*, **60**, 607–612.

Ripley, B. D. (1987), *Stochastic Simulation.* John Wiley and Sons, New York,

Sokal, A. D. (1989) *Monte Carlo methods in statistical mechanics: foundations and new algorithms.* Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne.

Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *Annals of Statistics*, **22**, 1701-1762

Tierney, L. (1998), A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, **8**, 1–9.

Tierney, L. and Mira, A. (1999), Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, **18**, 2507–2515.

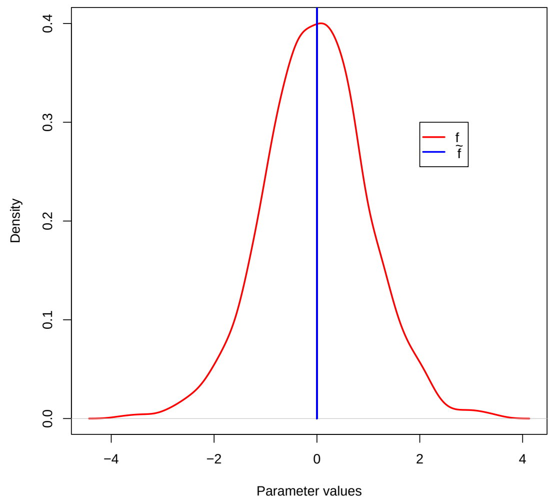


Figure 1: Normal target: kernel density estimation of the distribution of $f_1$ and $\tilde{f}_1$.

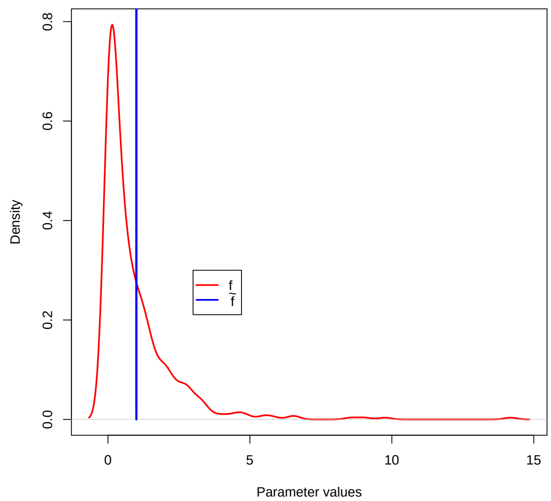


Figure 2: Normal target: kernel density estimation of the distribution of $f_2$ and $\tilde{f}_2$.

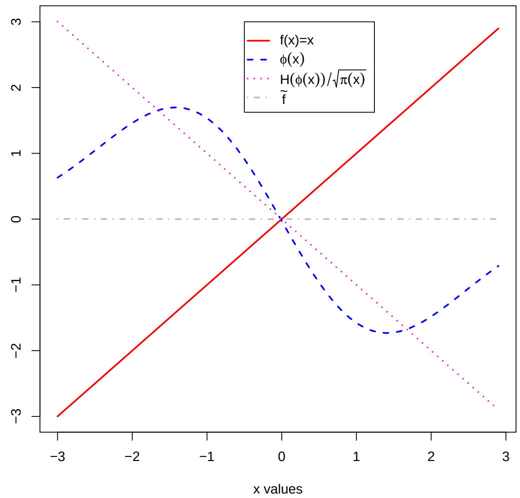


Figure 3: Normal target: $f(x) = x$

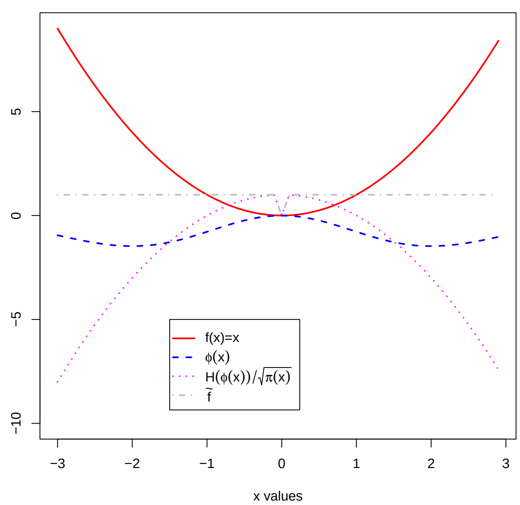


Figure 4: Normal target: $f(x) = x^2$

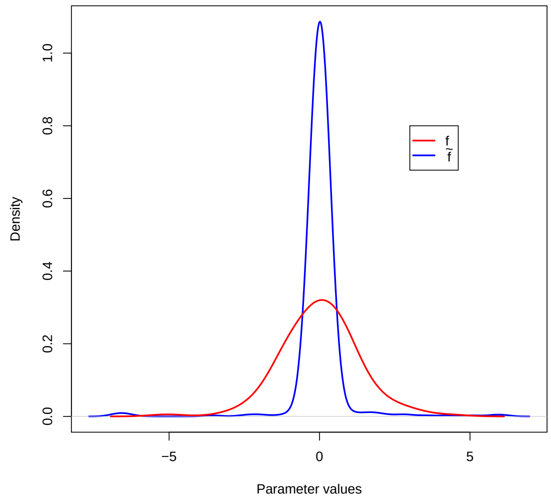


Figure 5: Student-T(5): kernel density estimation of the distribution of $f_1$ and $\tilde{f}_1$.

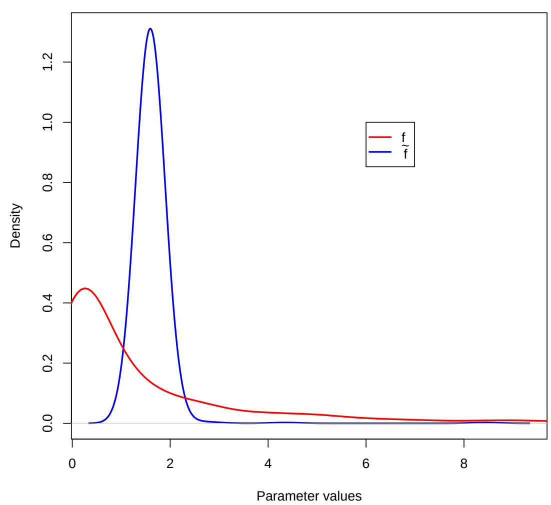


Figure 6: Student-T(5): Kernel density estimation of the distribution of $f_2$ and $\tilde{f}_2$.

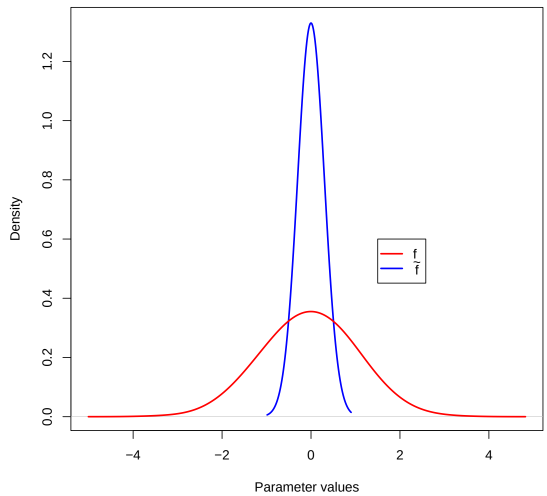


Figure 7: Bivariate normal: kernel density estimation of the distribution of $f_1$ and $\tilde{f}_1$.

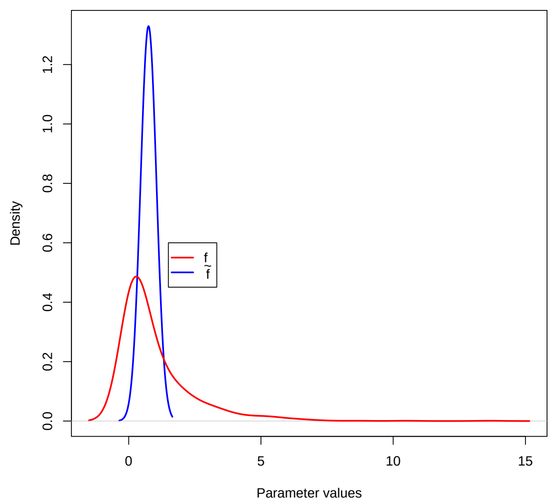


Figure 8: Bivariate normal: kernel density estimation of the distribution of $f_2$ and $\tilde{f}_2$.

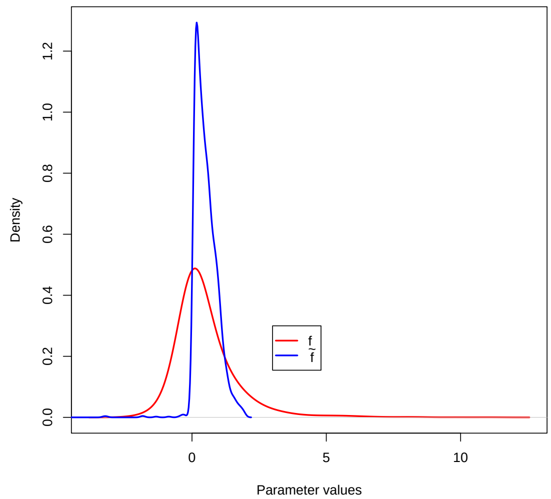


Figure 9: Bivariate normal: kernel density estimation of the distribution of $f_3$ and $\tilde{f}_3$.

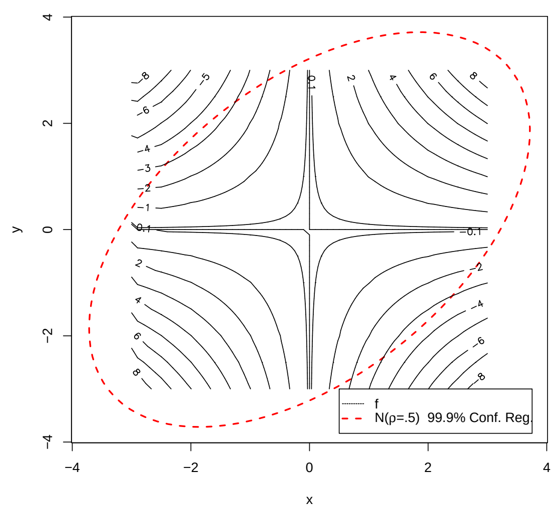


Figure 10: Bivariate normal 99.9 % confidence region and $f_3$ contour plot.

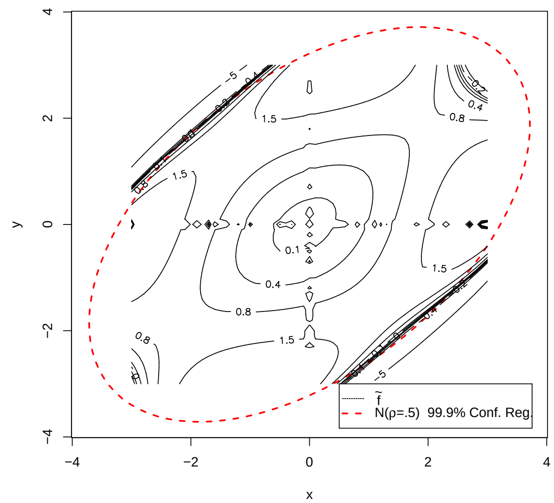


Figure 11: Bivariate normal 99.9 % confidence region and $\tilde{f}_3$ contour plot.

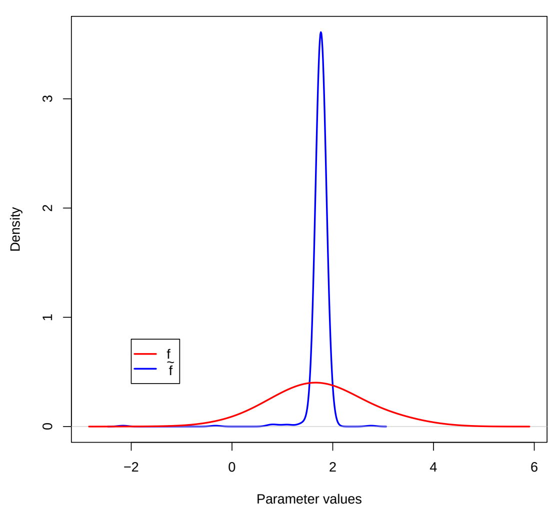


Figure 12: Bayesian example: kernel density estimation of the distribution of $f$ and $\tilde{f}$.