

Variance reduction by importance sampling

a

Suppose we want to compute $\alpha \equiv \mathbb{E}(g(W))$ with g non-negative and $W \sim f_W$. Let X be a random variable $X \sim f_X$, then

$$\alpha = \mathbb{E} \left(g(X) \frac{f_W(X)}{f_X(X)} \right). \quad (1)$$

Proof:

$$\alpha = \int_{\mathbb{R}} g(x) f_W(x) dx \quad (2)$$

$$= \int_{\mathbb{R}} g(x) \frac{f_X(x)}{f_X(x)} f_W(x) dx \quad (3)$$

$$= \int_{\mathbb{R}} g(x) \frac{f_W(x)}{f_X(x)} f_X(x) dx \quad (4)$$

$$= \mathbb{E} \left(g(X) \frac{f_W(X)}{f_X(X)} \right). \quad (5)$$

b,c

The above result allows solving α by Monte Carlo (MC) methods. When we sample from distribution f_X to estimate α , the mean square error (MSE) of the MC estimator

$$\text{MSE} \propto \int_{\mathbb{R}} \left(\left(g(x) \frac{f_W(x)}{f_X(x)} \right)^2 - \alpha^2 \right) f_X(x) dx. \quad (6)$$

The integrand is zero by choosing $f_X(x) = f_X^*(x) = \frac{f_W(x)g(x)}{\alpha}$, and thus the variance of the MC estimator is minimised by choosing $f_X = f_X^*$. In practice, this result is of little significance, since if we knew the exact value of α , there would be no need for the MC estimator in the first place.

d

Suppose we wish to evaluate $\mathbb{P}(\mathcal{N}(0, 1) > 3.75)$. The naive MC estimator would be given by $g(x) = \mathbf{1}_{x > \frac{15}{4}}(x)$ and $f_W(x) = \sqrt{2\pi} \exp\left(-\frac{x^2}{2}\right)$. Alternatively, using an affine change of variables, we may choose for any $\delta \in \mathbf{R}$: $g_\delta(x) = \mathbf{1}_{x > \frac{15}{4} + \delta}$ and $f_W = \sqrt{2\pi} \exp\left(-\frac{(x-\delta)^2}{2}\right)$ and obtain the correct result.

In order to use the MC estimate, we need to be able to sample numbers, from the distribution $\mathcal{N}(\delta, 1)$. And evaluate g for each of those realisations. An example implementation of this is given in <https://github.com/>

M	$M = 10^4$	$M = 10^5$
Naive MC	[0.000000, 0.000433] (116 %)	[0.000055, 0.000165] (50 %)
Refined MC	[0.000084, 0.000090] (3%)	[0.000088, 0.000089] (1%)
Exponential	[0.000086, 0.000103] (9%)	[0.000085, 0.000091] (3%)

Table 1: Confidence intervals for the MC estimators and relative errors when sampling from $\mathcal{N}(0, 1)$ and from $\mathcal{N}(5.75, 1)$. Note that the reference value is $\Phi(5.75) = 0.000088$ so that there is almost 40% that none of the realisations of a random sample of $M = 10^4$ exceed 5.75, leading to an estimate of zero, and vanishing variance.

Virtakuono/SME_HW3_Example/blob/master/examples.py The MC estimator is given by:

$$\bar{\alpha} = \sum_{m=1}^M \frac{g_{\delta}(X)}{M}, \quad (7)$$

with $X_m \sim N(\delta, 1)$ i.i.d. In the example we set $\delta = 2$. Since $E(X_m) = \delta$, $E(\bar{\alpha}) = \alpha$. Central limit theorem guarantees that $\bar{\alpha}$ is approximately normally distributed. To estimate the confidence interval, we compute the sample variance as

$$\bar{\sigma}^2 = \sum_{m=1}^M \frac{(g_{\delta}(x) - \bar{\alpha})^2}{M-1}. \quad (8)$$

Let

$$\Phi_{\mu, \sigma}(z) = (2\pi\sigma^2)^{-1} \int_{-\infty}^z \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx. \quad (9)$$

Then, there are ready implementations for the inverse of $\Phi_{\mu, \sigma}$, that allow us to compute z^* such that $\Phi_{0,1}(z^*) = 0.95$. Using this together with the fact that our MC estimator is approximately normally distributed, we set $\eta = z^* \sqrt{\frac{\bar{\sigma}^2}{M}}$ and obtain the confidence interval $[\bar{\alpha} - \eta, \bar{\alpha} + \eta]$. For the precise results, see table 1.

e

We may, of course, use an exponential density too. To ensure that half of the samples generated exceed 5.75, we may set

$$\lambda = \frac{\ln 2}{5.75}. \quad (10)$$

. Further improvements can naturally be made, optimising over λ as well as making affine transformation of the random variable.

Generate non-uniformly distributed random numbers given uniform

a

A mean zero unit variance random variable X has a Laplace distribution if its pdf is $f(x) = \frac{1}{2}e^{-|x|}$.

Algorithm to generate such random variable:

- $u \sim U(0, 1)$
- $X \sim F_U^{-1}(u)$, where

$$F_X(x) = \begin{cases} 1 - \frac{1}{2}e^{-x}, & x \geq 0 \\ \frac{1}{2}e^{-x}, & x < 0 \end{cases}$$

$$F_U^{-1}(u) = \begin{cases} \log(2u), & 0 < u \leq \frac{1}{2} \\ -\log(2(1-u)), & \frac{1}{2} \leq u < 1. \end{cases}$$

b

Algorithm to generate $Y \sim N(\mu, \sigma)$ random variables using the result above. Assume that there exists $\epsilon \in (0, 1]$ such that $\epsilon \frac{f_Y(X_k)}{f_X(X_k)} \leq 1$.

Algorithm (acceptance-rejection)

- Set $k=1$
- Sample two independent random variables X_k and $U_k \sim U(0, 1)$
- If $U_k \leq \epsilon \frac{f_Y(X_k)}{f_X(X_k)}$, then accept $Y = X_k$ as sample from $N(\mu, \sigma)$. Otherwise, reject X_k , increment k by 1 and go back to previous step.

c

Let U and V be two independent standard Gaussian random variables. Prove that the ratio $\frac{U}{V}$ is a Cauchy random variable.

Proof Let $Z = \frac{U}{V}$ then cdf of Z is given by

$$\begin{aligned} F_Z(z) &= P\left(\frac{U}{V} \leq z\right), \\ &= P(U \leq zV | V > 0) + P(U \geq zV | V < 0), \\ &= \int_0^\infty \left(\int_{-\infty}^{zv} f_U(u) \right) f_V(v) dv + \int_{-\infty}^0 \left(\int_{zv}^{-\infty} f_U(u) \right) f_V(v) dv. \end{aligned}$$

Then, the pdf of Z is given by

$$\begin{aligned}
f_Z(z) &= \frac{dF_Z(z)}{dz}, \\
&= \int_0^\infty v f_U(zv) f_V(v) dv + \int_{-\infty}^0 v f_U(zv) f_V(v) dv, \\
&= 2 \int_0^\infty v f_U(zv) f_V(v) dv = \frac{1}{\pi(1+z^2)}
\end{aligned}$$

Algorithm

- Generate samples from independent standard Gaussian random variables U and V .
- Compute the samples $Z = \frac{U}{V}$.

Kernel Density Estimator (KDE)

a

We propose to estimate a density function $\rho(y)$ based on a MC sample of realisations $y_i \in \mathbb{R}^d$ from the density ρ as follows

$$\hat{\rho}(y) = \frac{1}{Mh^{-d}} \sum_{m=1}^M K\left(\frac{y - y_i}{h}\right), \quad (11)$$

with

$$K(x) = 2^{-d} \mathbf{1}_{\|x\|_{L_2}}(x) \quad (12)$$

In this estimation, we commit two errors.

bias error

$$\rho(u) - \mathbb{E}(\hat{\rho}(y)) \quad (13)$$

$$= \rho(y) \int_{\mathbb{R}^d} h^{-d} K\left(\frac{y-z}{h}\right) \rho(z) dz \quad (14)$$

$$= \int_{\mathbb{R}^d} (\rho(y) - \rho(y-hz)) K(z) dz \quad (15)$$

$$\approx \int_{\mathbb{R}^d} \left(\rho(y) - \left(\rho(y) - h z_i \partial_{y_i} \rho(y) + \frac{1}{2} h^2 z_i z_j \partial_{z_i, z_j}^2 \rho(y) \right) \right) K(z) dz \quad (16)$$

$$= \frac{h^2}{2} \partial_{y_i y_j} \rho(y) \int_{\mathbb{R}^d} z_i z_j K(z) dz \propto h^2 \quad (17)$$

statistical error

The variance is bounded by

$$\mathbb{V} \left(K \left(\frac{y-z}{h} \right) \right) \leq \mathbb{E} \left(K^2 \left(\frac{y-z}{h} \right) \right) \quad (18)$$

and

$$\mathbb{E} \left(K^2 \left(\frac{y-z}{h} \right) \right) \approx \int_{\mathbb{R}^d} K^2(z) h^d \left(\rho(y) - h z_i \partial_{y_i} \rho(y) + \frac{h^2}{2} z_i z_j \partial_{y_i y_j}^2 \rho(y) \right) dz, \quad (19)$$

thus

$$\mathbb{V}(\hat{\rho}(y)) = M^{-1} h^{-2d} \mathbb{V} \left(K \left(\frac{y-z}{h} \right) \right) \quad (20)$$

$$\geq M^{-1} h^{-2d} h^d \rho(y) \int_{\mathbb{R}^d} K^2(z) dz \propto M^{-1} h^d \quad (21)$$

Total error

With undefined constants A and B we have that the total error squared (TSE) is given as

$$\text{TSE} \approx Ah^4 + \frac{B}{Mh^d}. \quad (22)$$

The critical point of the total square error is thus given by

$$h^{4+d} \propto M^{-1}. \quad (23)$$

This choice implies that the bias and sampling errors are of the same order

$$\sqrt{\text{TSE}} = \mathcal{O} \left(M^{-\frac{2}{d+4}} \right) \quad (24)$$

With the above analysis, we produce the set of three images (1) for varying M for a one-dimensional KDE and an example of two-dimensional gaussian example (2).

KDE for conditional expectation

a

Consider the Nadaraya-Watson estimator (cf. [1]) $\hat{g}(x)$ for $E[Y|X=x]$ which is derived as follows:

$$g(x) = E[Y|X=x] = \frac{\int y f(y, x) dy}{f(x)},$$

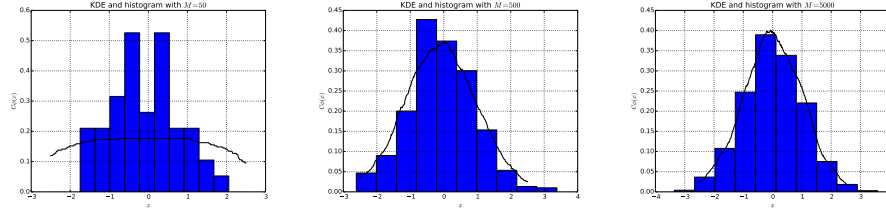


Figure 1: One-dimensional multivariate Gaussian and corresponding KDEs.

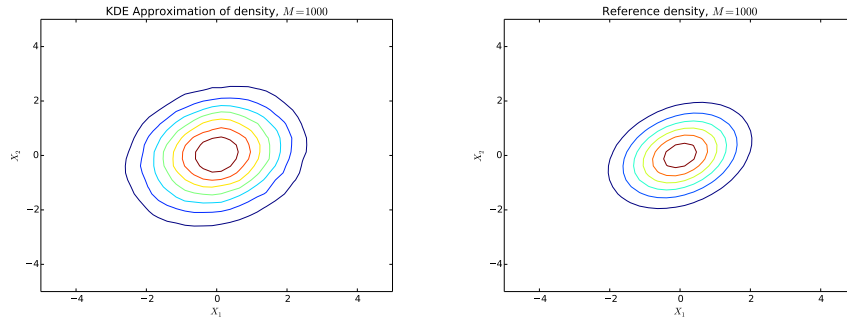


Figure 2: Two-dimensional multivariate Gaussian estimated (left) and exactly computed (right). $M = 100$ samples, $h = 0.5$

using the KDE for both $f(y, x)$ and $f(x)$

$$\begin{aligned}\hat{f}(y, x) &= \frac{1}{n} \sum_{i=1}^n \kappa_h(y - Y_i) \kappa_H(x - X_i), \\ \hat{f}(x) &= \frac{1}{n} \sum_{i=1}^n \kappa_H(x - X_i),\end{aligned}$$

and the fact that $\int z \kappa_h(z) dz = 0$, we obtain

$$\hat{g}(x) = \frac{\sum_{i=1}^n \kappa_H(x - X_i) Y_i}{\sum_{i=1}^n \kappa_H(x - X_i)}.$$

Optimal rate of convergence

Note we have

$$\begin{aligned}Y_i &= g(X_i) + \epsilon_i, \\ Y_i &= g(x) + (g(X_i) - g(x)) + \epsilon_i,\end{aligned}$$

where $E(\epsilon_i | X_i) = 0$ and $E(\epsilon_i^2 | X_i = x) = \sigma^2(x)$.

Therefore, the estimator can be written as

$$\hat{g}(x) = g(x) + \frac{\hat{m}_1(x)}{\hat{f}_X(x)} + \frac{\hat{m}_2(x)}{\hat{f}(x)},$$

where

$$\begin{aligned}\hat{m}_1(x) &= \frac{1}{n} \sum_{i=1}^n \kappa_H(x - X_i) (g(X_i) - g(x)), \\ \hat{m}_2(x) &= \frac{1}{n} \sum_{i=1}^n \kappa_H(x - X_i) \epsilon_i.\end{aligned}$$

If $d = 1$, we can show that

$$\begin{aligned}E(\hat{m}_1(x)) &= \frac{1}{h} \int k\left(\frac{x-u}{h}\right) (g(u) - g(x)) f(u) du \\ &= \int k(z) (g(x + hz) - g(x)) f(x + hz) dz \\ &\quad \text{(Taylor expansion)} \\ &= h^2 B(x) f(x) \int k(z) z^2 dz + o(h^2),\end{aligned}$$

where $B(x) = \frac{1}{2} g''(x) + \frac{g'(x)}{f(x)} f'(x)$.

Similarly, we can obtain $Var(\hat{m}_1(x)) = O(\frac{1}{nh})$.

$$\begin{aligned}
E(\hat{m}_2(x)) &= 0, \\
Var(\hat{m}_2(x)) &= \frac{1}{nh^2} \int k\left(\frac{x-u}{h}\right)^2 \sigma^2(u) f(u) du \\
&= \frac{1}{nh} \int k(z) \sigma^2(x+hz) f(x+hz) dz \\
&\quad \text{(Taylor expansion)} \\
&= \frac{\sigma^2(x) f(x)}{nh} \int k(z)^2 dz + o(h^2),
\end{aligned}$$

The asymptotic mean square error(AMSE) when $d = 1$ is

$$(h^2 B(x))^2 \left(\int k(z) z^2 dz \right)^2 + \frac{\sigma(x)^2 f_X(x)}{nh} \left(\int k(z)^2 dz \right).$$

In General, the asymptotic mean square error(AMSE) is given by

$$\left(\sum_{j=1}^d h_j^2 B_j(x) \right)^2 \left(\int k(z) z^2 dz \right)^2 + \frac{\sigma(x)^2 f_X(x)}{n|H|} \left(\int k(z)^2 dz \right)^d,$$

where $B_j(x) = \frac{1}{2} \partial_{x_j}^2 g(x) + f(x)^{-1} \partial_{x_j} g(x) \partial_{x_j} f(x)$ and the optimal value for h is proportional to $N^{-\frac{1}{d+4}}$.

b

Let us reuse the sample from the last exercise and show an illustration of the performance of the algorithm in figure 3

Monte Carlo and Variance Reduction

The task at hand is to evaluate an option on an underlying that is an arithmetic mean of a log-normal random variable. No closed form solution exists. However, we do have a closed form expression for the arithmetic mean of normal variables. Based on a visual inspection (cf. fig 4), the distribution of the underlying is, however, somewhat close to a Gaussian and especially the relevant part, the right tail of the distribution could be approximated with a Gaussian with a decent fidelity. We have

$$V_N = V_0 \prod_n^N R_n. \quad (25)$$

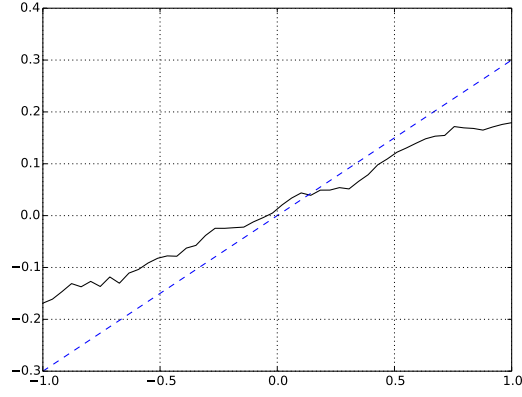


Figure 3: The conditional expectation of y , given x . The sample data is generated from $M = 400$ realisations of i.i.d. Gaussian random variables z_1 and z_2 . $x_2 = z_2$, $x_1 = z_1 + 0.3z_2$. The blue dashed line indicates the reference at slope 0.3. $h = 0.2$.

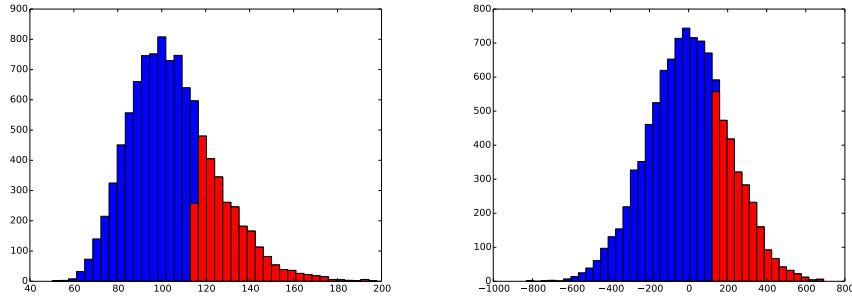


Figure 4: The distribution of the underlying of the Asian Option (left) and a control variate (right). The right tail of the distribution is coloured left to indicate realisations *in the money* with strikes K and \tilde{K} .

Empirically, we see that the probability P_{otm} of V_N being out of the money is approximately 72 per cent. Define

$$Q = \sum_{n=1}^N (50 - n) (\ln R_n - r\Delta t), \quad (26)$$

we note that both Q and V_N are increasing in all the random variables R_n . Furthermore, we know that Q entered and normally distributed with $\tilde{\sigma}^2 = \sum_{n=1}^N (50 - n)^2$. With the inverse cumulative distribution Φ^{-1} we can define

$$\tilde{K} = \tilde{\sigma}\Phi^{-1}(P_{otm}). \quad (27)$$

Then our control variate is

$$G = (Q - \tilde{K})^+. \quad (28)$$

We have

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\tilde{\sigma}^2}\right) (x - k)^+ = \frac{\tilde{\sigma}}{\sqrt{2\pi}} \exp\left(-\frac{k^2}{2\tilde{\sigma}^2}\right) + k(\Phi(k) - 1). \quad (29)$$

We have that the correlation coefficient ρ between G and $C = \left(\sum_{n=1}^N \frac{V_n}{K} - K, 0\right)^+$ exceeds 0.99, see fig 5.

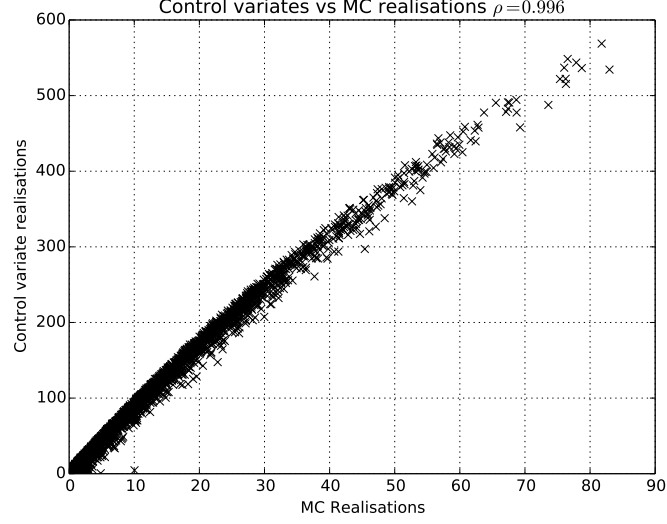


Figure 5: The MC realisations and corresponding control variates show a near-perfect correlation

We can form the estimators

$$\bar{C}_1 = \sum_{m=1}^M \frac{\left(\left(\sum_{n=1}^N V_0 \prod_{m=1}^n R_n \right) - K \right)^+}{M} \quad (30)$$

$$\bar{C}_2 = \sum_{m=1}^{2M} \frac{\left(\left(\sum_{n=1}^N V_0 \prod_{m=1}^n \tilde{R}_{n,m} \right) - K \right)^+}{M} + \sum_{m=1}^{2M} \frac{\left(\left(\sum_{n=1}^N V_0 \prod_{m=1}^n \tilde{R}_{n,m} \right) - K \right)^+}{M} \quad (31)$$

$$\bar{C}_3 = \sum_{m=1}^M \frac{\left(\left(\sum_{n=1}^N V_0 \prod_{m=1}^n R_n \right) - K \right)^+ + \beta (G_m - E(Q))}{M} \quad (32)$$

$$\bar{C}_4 = \sum_{m=1}^{2M} \frac{\left(\left(\sum_{n=1}^N V_0 \prod_{m=1}^n R_n \right) - K \right)^+ + \beta (G_m - E(Q))}{M} \\ \sum_{m=1}^{2M} \frac{\left(\left(\sum_{n=1}^N V_0 \prod_{m=1}^n R_n \right) - K \right)^+ + \beta (Q_m - E(Q))}{M}, \quad (33)$$

Plain vanilla	Antithetic	Control variate	Hybrid
0.15 %	0.10 %	0.012 %	0.0085 %

Table 2: Widths of 95 % confidence bands for different Monte Carlo estimators for $M = 1000$. Modified option price approximately 4.3.

for the plain vanilla, antithetic, control variate, and combination methods, respectively. The $R_{n,m}$ are understood to be independent realisations of the incrementation for n th time step for realisation m . The antithetic variables are defined as

$$\tilde{R}_{n,m} = \exp(-\log R_{n,m} + \Delta tr) \quad (34)$$

and

$$\tilde{Q}_m = \sum_{n=1}^N (50 - n) \left(\ln \tilde{R}_{nm} - r\Delta t \right), \quad (35)$$

$$G_m = \left(Q_m - \tilde{K} \right)^+ \quad (36)$$

$$\tilde{G}_m = \left(\tilde{Q}_m - \tilde{K} \right)^+. \quad (37)$$

In order to minimise the variance of the estimators, we set $\beta = \rho \sqrt{\frac{\sigma_C^2}{\sigma_Q^2}}$ with σ_C^2 being the variance of the plain vanilla MC realisations and σ_Q^2 the variance of the control variate realisations. With $M = 10000$, we obtain the results in table 2: The exact computational cost of each estimator depends on how computationally costly it is to generate time steps compared to the cost of generating random numbers. For a rough approximation, we may state that the computational cost of the antithetic and control variate estimators are twice that of the plain vanilla and the hybrid method requires four-fold computational effort.

Confidence interval, Bootstrapping

In order to compute the expected shortfall we reuse the MC sample from the previous exercise. Let our MC estimator \bar{C} be given as:

$$\bar{C} = \sum_{m=1}^M \frac{X_i}{M}. \quad (38)$$

Then let us draw $N \times M$, $N, M \in \mathbb{Z}$ random variables $k_i \in \{1, 2, 3, \dots, M\}$. Then define S_j such that

$$\#A_j = \# \{k_i : X_{k_i} > S_j, (j-1)M \leq i \leq jM - 1\} = \lfloor pM \rfloor \quad (39)$$

p	0.9	0.95	0.99
Confidence interval	[28.28, 30.01]	[36.51, 38.80]	[52.34, 57.68]

Table 3: Confidence intervals for the Expected Shortfall for problem 5 example for various quantiles. The width of the confidence interval tends to increase as a function of q .

for $j \in 1, 2, 3, 4, \dots, N$. and

$$Q_j = \sum_{i \in A_j} \frac{X_i \mathbf{1}_{x > S_j}(X_i)}{\#A_j} \quad (40)$$

then the estimated confidence interval can be noted as $[Q^-, Q^+]$ as

$$\# \{j : Q_j < Q^-\} = \left\lfloor \frac{qN}{2} \right\rfloor \quad (41)$$

$$\# \{j : Q_j < Q^+\} = \left\lfloor \frac{qN}{2} \right\rfloor. \quad (42)$$

Setting $N = 1000$, we obtain the results in the table 3

References

- [1] Nadaraya-Watson estimator - <http://www.maths.manchester.ac.uk/~peterf/MATH38011/NPR%20N-W%20Estimator.pdf>
- [2] Example code in python. https://github.com/Virtakuono/SME_HW3_Example/blob/master/examples.py