A Monte Carlo Method to Approximate Conditional Expectations based on a
Theorem of Besicovitch: Application to Equivariant Estimation
of the Parameters of the General Half-Normal Distribution

A.G. Nogales*, P. Pérez* and P. Monfort**

Dpto. de Matemáticas, Universidad de Extremadura
Avda. de Elvas, s/n, 06071–Badajoz, SPAIN.
e-mail: nogales@unex.es

Abstract.
A natural Monte Carlo method to approximate conditional expectations in a probabilistic
framework is justified by a general result inspired on the Besicovitch covering theorem on
differentiation of measures. The method is specially useful when densities are not available
or are not easy to compute. The method is illustrated by means of some examples and
can also be used in a statistical setting to approximate the conditional expectation given a
sufficient statistic, for instance. In fact, it is applied to evaluate the minimum risk equiv-
ariant estimator (MRE) of the location parameter of a general half-normal distribution
since this estimator is described in terms of a conditional expectation for known values
of the location and scale parameters. For the sake of completeness, an explicit expression
of the the minimum risk equivariant estimator of the scale parameter is given. For all we
know, these estimators have not been given before in the literature. Simulation studies are
realized to compare the behavior of these estimators with that of maximum likelihood and
unbiased estimators.

AMS Subject Class. (2010): Primary 65C05, 62F10, 62-04.

Key words and phrases: Monte Carlo simulation, conditional expectation, general half-normal distri-
bution, equivariant estimation.

(*) This author has been supported by the spanish Ministerio de Ciencia e Innovación under the
project MTM2010-16845 and the Junta de Extremadura Autonomous Government under the
grant GR10064.

(**) This author has been supported by the spanish Ministerio de Ciencia e Innovación under
the project MTM2010-16660.
1 Introduction

Let \((\Omega, \mathcal{A}, P)\) be a probability space, \(X : (\Omega, \mathcal{A}, P) \to \mathbb{R}^n\) be an \(n\)-dimensional random variable and \(Y : (\Omega, \mathcal{A}, P) \to \mathbb{R}\) a random variable with finite mean. The conditional expectation \(E(Y|X)\) is defined as a random variable on \(\mathbb{R}^n\) such that \(\int_{\Omega} f(Y) \, dP = \int_{\Omega} f(E(Y|X)) \, dP\) for all Borel set \(B\) in \(\mathbb{R}^n\), where \(P^X\) denotes the probability distribution of \(X\). Although the existence of the conditional expectation is guaranteed via the Radon-Nikodym theorem, its computation becomes, generally, a hard problem. When the joint density \(f\) of \(Y\) and \(X\) is known, \(E(Y|X = x)\) is the mean of the conditional distribution \(P^Y|X=x\) of \(Y\) given \(X = x\), whose density is \(f(x, y)/f_X(x)\), where \(f_X\) denotes the marginal distribution of \(X\). In this case the problem to compute a conditional expectation is reduced to that to “evaluate” a mean, and we have a lot of methods to do that, interpreting “evaluation” as “approximation” or “simulation” in a probabilistic context or “estimation” in a statistical framework.

Theorem 1 (Besicovitch (1945, 1946)). Let \(\lambda\) be a Radon measure on \(\mathbb{R}^n\), and \(f : \mathbb{R}^n \to \mathbb{R}\) a locally \(\lambda\)-integrable function. Then

\[
\lim_{\epsilon \downarrow 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f \, d\lambda = f(x)
\]

for \(\lambda\)-almost all \(x \in \mathbb{R}^n\), where \(B_r(x)\) denotes the ball of center \(x\) and radius \(r > 0\) for the norm \(\| \cdot \|_\infty\) on \(\mathbb{R}^n\).

Let now \((\Omega, \mathcal{A}, P)\) be a probability space, \(U : (\Omega, \mathcal{A}, P) \to \mathbb{R}^n\) be an \(n\)-dimensional random variable and \(f : (\Omega, \mathcal{A}, P) \to \mathbb{R}\) be a real random variable with finite mean. Then, for \(P^U\)-almost every \(u \in \mathbb{R}^n\),

\[
\lim_{\epsilon \downarrow 0} \frac{1}{P^U(B_r(u))} \int_{U^{-1}(B_r(u))} f(\omega) \, dP^U(\omega) = \lim_{\epsilon \downarrow 0} \frac{1}{P^U(B_r(u))} \int_{B_r(u)} E(f|U = u') \, dP^U(u') = E(f|U = u)
\]

2 A method to approximate conditional expectations

Let us recall a theorem of Besicovitch on differentiation of measures (see, for instance, Corollary 2.14 of Mattila (1995)):

Let \(\lambda\) be a Radon measure on \(\mathbb{R}^n\), and \(f : \mathbb{R}^n \to \mathbb{R}\) a locally \(\lambda\)-integrable function. Then

\[
\lim_{\epsilon \downarrow 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f \, d\lambda = f(x)
\]

for \(\lambda\)-almost all \(x \in \mathbb{R}^n\), where \(B_r(x)\) denotes the ball of center \(x\) and radius \(r > 0\) for the norm \(\| \cdot \|_\infty\) on \(\mathbb{R}^n\).

Let now \((\Omega, \mathcal{A}, P)\) be a probability space, \(U : (\Omega, \mathcal{A}, P) \to \mathbb{R}^n\) be an \(n\)-dimensional random variable and \(f : (\Omega, \mathcal{A}, P) \to \mathbb{R}\) be a real random variable with finite mean. Then, for \(P^U\)-almost every \(u \in \mathbb{R}^n\),

\[
\lim_{\epsilon \downarrow 0} \frac{1}{P^U(B_r(u))} \int_{U^{-1}(B_r(u))} f(\omega) \, dP^U(\omega) = \lim_{\epsilon \downarrow 0} \frac{1}{P^U(B_r(u))} \int_{B_r(u)} E(f|U = u') \, dP^U(u') = E(f|U = u)
\]
By the Strong Law of Large Numbers, for almost every sequence \((\omega_i)\) in \(\Omega\), we have

\[
P^U(B_u) = \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} I_{B_u}(U(\omega_i))
\]

and

\[
\int_{B_u} E(f|U = u') \, dP^U(u') = \lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} I_{B_u}(U(\omega_i)) f(\omega_i)
\]

Hence, we have proved the following result:

**Theorem 2.** Let \((\Omega, \mathcal{A}, P)\) be a probability space, \(U : (\Omega, \mathcal{A}, P) \to \mathbb{R}^n\) be an \(n\)-dimensional random variable and \(f : (\Omega, \mathcal{A}, P) \to \mathbb{R}\) be a real random variable with finite mean. Then, for \(P^U\)-almost every \(u \in \mathbb{R}^n\) and almost every sequence \((\omega_i)\) in \(\Omega\), we have

\[
E(f|U = u) = \lim_{\epsilon \to 0} \frac{\sum_{i=1}^{k} I_{B_u}(U(\omega_i)) f(\omega_i)}{\sum_{i=1}^{k} I_{B_u}(U(\omega_i))}
\]

This theorem yields a way to approximate the conditional expectation of \(f\) given \(U\). Let us give a simple example to illustrate the method.

**Example 1.** Let \((X, Y)\) be a bidimensional random variable normally distributed with null mean and covariance matrix

\[
\begin{pmatrix}
1 & 1/2 \\
1/2 & 1
\end{pmatrix}
\]

In this case, we don’t need any approximation of the conditional expectation of \(Y\) given \(X = x\) because it is \(x/2\). Notice that, in this simple example, the conditional distribution of \(Y\) given \(X = x\) is \(N(\frac{1}{2}x, \frac{1}{4}\sqrt{3})\). Nevertheless, if we want to apply the suggested method to calculate \(E(Y|X = 1)\), given a small \(\epsilon > 0\) small, we may choose a sample \((x_i, y_i)_{1 \leq i \leq k}\) of the joint distribution of \(X\) and \(Y\) and approximate \(E(Y|X = 1)\) by

\[
\frac{\sum_{i=1}^{k} I_{[1-\epsilon,1+\epsilon]}(x_i) \cdot y_i}{\sum_{i=1}^{k} I_{[1-\epsilon,1+\epsilon]}(x_i)}
\]

Taken \(\epsilon = 0.1\) and samples of the joint distribution of \(X\) and \(Y\) with sample sizes \(k\) enough to obtain \(m = \sum_{i=1}^{k} I_{[1-\epsilon,1+\epsilon]}(x_i) = 10, 20, 30, 50, 100\), and using the statistical software R, we have obtained the following approximations of \(E(Y|X = 1)\) and box-plots (dotted red line represents the mean) after 100 simulations:

| \(m\) | 10  | 20  | 30  | 50  | 100 |
|-------|-----|-----|-----|-----|-----|
| \(E(Y|X = 1)\) | 0.5007 | 0.5211 | 0.5037 | 0.5211 | 0.5114 |

Table 1. Approximation of \(E(Y|X = 1)\) (\(\epsilon = 0.1\), = 10, 20, 30, 50, 100, 100 simulations).

![Figure 1. Box plots of the approximations of \(E(Y|X = 1)\) (\(\epsilon = 0.1\), = 10, 20, 30, 50, 100, 100 simulations).](image)
A similar simulation study has been performed to approximate the conditional expectation $E(V|U = 0.5)$, where $V = \sin(X \cdot Y)$ and $U = \cos(X^2 + Y^2)$; the obtained results are:

| $m$ | 10   | 20   | 30   | 50   | 100  |
|-----|------|------|------|------|------|
| $E(V|U = 0.5)$ | 0.1235 | 0.1058 | 0.1309 | 0.1341 | 0.1281 |

Table 2. Approximation of $E(V|U = 0.5)$ ($\epsilon = 0.1$, $\epsilon = 10, 20, 30, 50, 100$, 100 simulations).

Figure 2. Box plots of the approximations of $E(V|U = 0.5)$ ($\epsilon = 0.1$, $\epsilon = 10, 20, 30, 50, 100$, 100 simulations)

Notice that in the example $k$ should be an enough great number to secure a good size $m$ of non-null terms in the denominator of this expression. Besides, the smaller $\epsilon$, greater has to be $k$. This may become a problem when this method is applied, especially when $X$ is a random vector of high dimension. Any additional information about the distribution of $X$ may be useful in some way to circumvent this problem, as indeed occur when determining the minimum risk equivariant estimator (MRE) of the location parameter $\xi$ of the general half-normal distribution in the next section.

3 Application to equivariant estimation of the location parameter of the general half-normal distribution

Let $Z$ be a real random variable (r.r.v.) with distribution $N(0,1)$. The distribution of the r.r.v. $X := |Z|$ is the so-called half-normal distribution. It will be denoted $HN(0,1)$ and its density function is

$$f_X(x) = \sqrt{\frac{2}{\pi}} \exp\left\{-\frac{1}{2} x^2 \right\} I_{[0,\infty]}(x).$$

A general half-normal distribution $HN(\xi, \eta)$ is obtained from $HN(0,1)$ by a location-scale transformation: $HN(\xi, \eta)$ is the distribution of $Y = \xi + \eta X$.

The classical paper Daniel (1959) introduces half-normal plots and the half-normal distribution. The half-normal distribution is a special case of the folded normal and truncated normal distribution (see Johnson et al. (1994)). Bland et al. (1999) and Bland (2005) propose a so-called half-normal method to deal with relationships between measurement error and magnitude, with applications in medicine. Pewsey (2002) uses the maximum likelihood principle to estimate the parameters, and contains a brief survey on the general half-normal distribution, its relations with other well-known distributions and its usefulness in the analysis of highly skew data; Pewsey (2004) proposes bias-corrected estimators of the estimators quoted before. Nogales et al. (2011) deals with the problem of unbiased estimation in the general half-normal distribution. This paper is mainly devoted to the problem of equivariant estimation of the location and scale parameters, $\xi$ and $\eta$, but first we do a brief review on the results about unbiased and maximum likelihood estimation appearing in the literature.

The density function of $HN(\xi, \eta)$ is
\[
    f_Y(y) = \frac{1}{\eta} f_X \left( \frac{y - \xi}{\eta} \right) = \frac{1}{\eta} \sqrt{\frac{2}{\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{y - \xi}{\eta} \right)^2 \right\} I_{[\xi, \infty)}(y).
\]

It is readily shown that

\[
    E(Y) = \xi + \eta \sqrt{\frac{2}{\pi}} \quad \text{and} \quad \text{Var}(Y) = \frac{\pi - 2}{\pi} \eta^2.
\]

Let \( Y_1, \ldots, Y_n \) be a sample of size \( n \) from a general half-normal distribution with unknown parameters, \( \xi \) and \( \eta \). \( Y_{1:n} \) denotes the minimum of \( Y_1, \ldots, Y_n \). From the factorization criterion, we obtain that \((\sum_{i=1}^n Y_i^2, \sum_{i=1}^n Y_i, Y_{1:n})\) is a sufficient statistic. Indeed, it is minimal sufficient, although not complete.

We write \( Y_i = \xi + \eta X_i \), where \( X_i = |Z_i| \), \( 1 \leq i \leq n \), \( Z_1, \ldots, Z_n \) being a sample of the standard normal distribution \( N(0, 1) \). Throughout this paper, we also write

\[
    c_n := E(X_{1:n})
\]

For \( n \geq 2 \), it is readily shown that \( 0 < c_n < \sqrt{\frac{\pi}{2}} \). In fact, the next lemma (Nogales el al. (2011)) yields an alternative expression and a refined bound for \( c_n \). We write \( \Phi \) for the standard normal cumulative distribution function.

**Lemma 1.** (i) \( c_n = \int_0^\infty (2 - 2\Phi(t))^n \, dt \).

(ii) For \( n \geq 1 \), \( c_n \leq \frac{1}{n} \sqrt{\frac{\pi}{2}} \leq \Phi^{-1} \left( \frac{1}{2} + \frac{1}{2n} \right) \).

Notice also that \( Y_{1:n} = \min_i Y_i = \xi + \eta X_{1:n} \) and \( E(Y_{1:n}) = \xi + \eta c_n \).

The next proposition (Nogales el al. (2011)) yields unbiased estimators of the location and scale parameters, \( \xi \) and \( \eta \). Both estimators are \( L \)-statistics and function of the minimal sufficient statistic cited.

**Proposition 1.** (i) \( \tilde{\xi} := \frac{\sqrt{\pi} Y_{1:n} - c_n \bar{Y}}{\sqrt{\pi} - c_n} \) is an unbiased estimator of the location parameter \( \xi \).

(ii) \( \tilde{\eta} := \frac{\bar{Y} - Y_{1:n}}{\sqrt{\pi} - c_n} \) is an unbiased estimator of the scale parameter \( \eta \) whose distribution does not depend on \( \xi \).

**Remark.** We also have that the sample mean \( \bar{Y} \) is an unbiased estimator of the mean \( \xi + \eta \sqrt{\frac{\pi}{2}} \). Moreover, an unbiased estimator of \( \eta^2 \) is

\[
    \frac{\pi}{\pi - 2} S^2,
\]

where \( S^2 := \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \) is the sample variance; notice that its distribution does not depend on \( \xi \), \( \bar{Y} \) and \( S^2 \) also are functions of the sufficient statistic given above. The reader is referred to Nogales et al. (2011) for these and other results about unbiased estimation of the parameters of the general half-normal distribution. \( \square \)

**Remark.** Pewsey (2002) provides maximum likelihood estimates for each of the parameters \( \xi \) and \( \eta \):

\[
    \hat{\xi} := Y_{1:n}, \quad \hat{\eta} := \left( \frac{1}{n} \sum_{i=1}^n (Y_i - Y_{1:n})^2 \right)^{1/2}
\]

A large sample based bias-correction is used in Pewsey (2004) to improve the performance of the maximum likelihood estimators \( \hat{\xi} \) and \( \hat{\eta} \). \( \square \)

In this section we consider the problem of determining the minimum risk equivariant estimator of the position parameter \( \xi \) of the general half-normal distribution \( HN(\xi, \eta) \) when the scale parameter \( \eta \) is unknown. We cannot provide an explicit expression for this estimator, since it is described in terms
of two conditional expectations that had to be estimated by simulation. To achieve this goal, an R program has been developed based on the method of the previous section.

For the sake of completeness, we also give MRE estimators of the scale parameter, and of one of the parameters when the other is supposed to be known since, as far as we know, they have not been yet reported in the literature. The results are a consequence of the classical equivariant estimation theory, as it appears, for instance, in Lehmann (1983).

To estimate the location parameter $\xi$ when the scale parameter $\eta$ is unknown, we have the next result (a direct consequence of Lehmann (1986, p. 182)).

**Proposition 2.** When the loss function $W_2(x; \xi, \eta) = \eta^{-2}(x - \xi)^2$ is considered, the MRE estimator $\hat{\xi}$ of $\xi$ is

$$\hat{\xi} = T_\eta^* - (\rho \circ U) \cdot T_\xi^*$$

where

$$T_\eta^* = \bar{Y}, \quad T_\xi^* = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{Y}|$$

$$U = \left( \begin{array}{cccc} Y_1 - Y_n & \cdots & Y_{n-2} - Y_n & Y_{n-1} - Y_n \\ Y_{n-1} - Y_n & \cdots & Y_{n-2} - Y_n & Y_{n-1} - Y_n \end{array} \right)$$

$$\rho = \frac{E_{\xi=0,\eta=1}[T_\eta^* \cdot T_\xi^*|U]}{E_{\xi=0,\eta=1}[T_\xi^*|U]}$$

**Remark.** $T_\eta^*$ can be replaced by any other equivariant estimator of $\xi$, and $T_\xi^*$ can be replaced by any positive estimator of $\eta$ satisfying $T_\eta^* (a + by_1, \ldots, a + by_n) = bT_\eta^* (y_1, \ldots, y_n)$ for every $a \in \mathbb{R}$, $b > 0$. □

**Remark.** A simulation was realized to visualize the behavior of the minimum risk equivariant estimator $\hat{\xi}$. For this simulation we did 100 simulations with sample sizes $n = 10, 20, 30, 50, 100$ of a half-normal distribution $HN(10, 4)$ obtaining the next results:

| $n$ | 10   | 20   | 30   | 50   | 100  |
|-----|------|------|------|------|------|
| Mean| 9.8710 | 9.6038 | 9.4242 | 9.5351 | 9.6429 |
| MSE | 1.0684 | 0.9301 | 1.6223 | 0.9041 | 0.4050 |

Table 3. Simulated mean and MSE of the estimators $\hat{\xi}$.

![Figure 3. Box plots of the simulations.](image)

To compare the behavior of the unbiased estimator $\hat{\xi}$, the maximum likelihood estimator $\hat{\xi}$ and the minimum risk equivariant estimator $\hat{\xi}$, we did 100 simulations with sample sizes $n = 100$ of a half-normal distribution $HN(10, 4)$ obtaining the next results:
Table 4. Simulated mean and MSE of the estimators ˜ξ, ˆξ and ◦ξ.

| n  | Mean | MSE  | Mean | MSE  | Mean | MSE  |
|-----|------|------|------|------|------|------|
| 100 | 9.9964 | 0.0018 | 10.0457 | 0.0039 | 9.6429 | 0.4050 |

We can see the biased character of the maximum likelihood estimator ˆξ and the minimum risk equivariant estimator ◦ξ. Obviously, as it can be expected, the behavior of this approximation of the MRE estimator is worse than those of the unbiased estimator ˜ξ or the maximum likelihood estimator ˆξ. However, this method provides a way to proceed when other estimation methods are not available.

This simulation was performed with the statistical program R.

Let us summarize the idea used in this estimation: for a sample \( y = (y_1, \ldots, y_n) \), \( n = 10, 20, 30, 50, 100 \), of the distribution \( HN(10, 4) \), we have

\[
\rho(U(y)) = \lim_{\epsilon \to 0} \frac{N_\epsilon}{D_\epsilon}
\]

where

\[
N_\epsilon = \int_{A_\epsilon(y)} f(y')dy' \quad D_\epsilon = \int_{A_\epsilon(y)} g(y')dy'
\]

\[
f(y') = T_0^*(y') \cdot T_1^*(y') \cdot \exp\left\{ -\frac{1}{2} \| y' \|^2 \right\} \quad g(y') = T_1^*(y')^2 \cdot \exp\left\{ -\frac{1}{2} \| y' \|^2 \right\}
\]

\[A_\epsilon(y) = \{y' \in [0,10]^n : \max_{1 \leq i \leq n} |U_i(y') - U_i(y)| \leq \epsilon \}\]

Now, take a sample \( S \) of \( A_\epsilon(y) \) and approximate \( N_\epsilon \) and \( D_\epsilon \) by \( \frac{1}{\text{card}(S)} \sum_{y' \in S} f(y') \) and \( \frac{1}{\text{card}(S)} \sum_{y' \in S} g(y') \), resp. So, \( \rho(U(y)) \) can be estimated by

\[
C(y) := \frac{\sum_{y' \in S} f(y')}{\sum_{y' \in S} g(y')}
\]

and ◦ξ(y) is approximated by \( D(y) := T_0^*(y) - C(y) \cdot T_1^*(y) \).

To approximate \( C(y) \), a first idea would be to divide the interval \([0,10]\) in multiple subintervals of small length \( \epsilon > 0 \) and consider the grid in the interval \([0,10]^n\) formed by the \( n \)-power set of the ends of these subintervals (we have restricted ourselves to the interval \([0,10]\) because we have considered virtually nil the functions \( f(y) \) and \( g(y) \) when one of the coordinates of the vector \( y \) is greater than 10). Sample \( S \) would be formed by the grid nodes that are in \( A_\epsilon \). The main problem with this approach is that the size \( m \) of the sample \( S \) is very small (it becomes smaller when the greater is the dimension \( n \)). To secure a sample size \( m \) enough for \( S \) (given \( n \), we take \( m = 100 \cdot n \)), we have used the following algorithm, that benefits from the invariance of \( U \) under scale and location transformations:
• Given a sample \( y = (y_1, \ldots, y_n) \) of the distribution \( HN(10, 4) \), take \( w_{n-1}, w_n \) at random in [0, 10] such that \( w_{n-1} - w_n \) has the same sign than \( y_{n-1} - y_n \).

• For \( 1 \leq i \leq n - 2 \), let \( a_i := \frac{y_{i+1} - y_i}{y_{n-1} - y_n} \) and take \( 0 < \epsilon < \min\{0.1, \min_{1 \leq i \leq n-2} |a_i|\} \).

• For \( 1 \leq i \leq n - 2 \) take \( w_i \) at random on the interval determined by \( w_n + (w_{n-1} - w_n)(a_i - \epsilon) \) and \( w_n + (w_{n-1} - w_n)(a_i + \epsilon) \).

• The process is repeated until 100 \( n \) vectors \( w^{(j)} = (w_1^{(j)}, \ldots, w_n^{(j)}) \), \( 1 \leq j \leq 100 \cdot n \) are obtained.

• If \( w_i^{(j_0)} < 0 \) for some \( i_0, j_0 \), we replace \( w_i^{(j)}, 1 \leq i \leq n, 1 \leq j \leq 100 \cdot n \), by \( v + w_i^{(j)} \), where \( v \) is chosen at random between \( -\min_{1 \leq i \leq n, 1 \leq j \leq 100 \cdot n} w_i^{(j)} \) and \( 1 - \min_{1 \leq i \leq n, 1 \leq j \leq 100 \cdot n} w_i^{(j)} \).

• Each new \( w^{(j)} \) is divided by \( \max_{1 \leq i \leq n} w_i^{(j)} \) and multiplied by a random number choosen in [0, 10].

• Take \( S = \{w^{(j)} : 1 \leq j \leq 100 \cdot n\} \).

Finally, we choose \( k := 100 \) samples \( y^{(i)} \) of size \( n \) of the distribution \( HN(10, 4) \) and estimate the mean of \( \xi \) by

\[
\frac{1}{k} \sum_{i=1}^{k} D(y^{(i)})
\]

and the mean squared error \( \xi \) by

\[
\frac{1}{k} \sum_{i=1}^{k} (D(y^{(i)}) - 10)^2.
\]

\( \square \)

**Remark.** When the scale parameter \( \eta \) is supposed known (say \( \eta = \eta_0 \)), the joint density of \( Y_1, \ldots, Y_n \) is

\[
f_\xi(y_1, \ldots, y_n) = \frac{1}{\eta_0^{\frac{n}{2}}} \sqrt{\frac{2}{\pi}} \exp \left\{-\frac{1}{2\eta_0^2} \sum_{i=1}^{n} (y_i - \xi)^2 \right\} I_{\xi, \infty}(y_{1:n}),
\]

where \( y_{1:n} := \min\{y_1, \ldots, y_n\} \). This family remains invariant under translations of the form \( g_n(y_1, \ldots, y_n) = (y_1 - a, \ldots, y_n - a) \).

The equivariant estimator of minimum mean squared error of the location parameter \( \xi \) is

\[
T_1 = \bar{Y} - \frac{\eta_0}{\sqrt{2\pi n}} \exp \left\{-\frac{1}{2\eta_0^2} (Y_{1:n} - \bar{Y})^2 \right\} \frac{1}{\sqrt{\eta_0}} \Phi \left( \frac{\sqrt{n}}{\eta_0} (Y_{1:n} - \bar{Y}) \right).
\]

In fact, for the loss function \( W_2^2(\xi, x) = (x - \xi)^2 \), the MRE estimator of the location parameter \( \xi \) is the Pitman estimator

\[
T_1(y_1, \ldots, y_n) = \frac{\int_{-\infty}^{+\infty} u f_0(y_1 - u, \ldots, y_n - u) du}{\int_{-\infty}^{+\infty} f_0(y_1 - u, \ldots, y_n - u) du}
\]

For \( y \in \mathbb{R}^n \), we write \( \bar{y} \) for the mean of \( y_1, \ldots, y_n \). After some algebraic manipulations, we obtain:

\[
\int_{-\infty}^{+\infty} u f_0(y_1 - u, \ldots, y_n - u) du = \left( \frac{\sqrt{2}}{\eta_0 \sqrt{\pi}} \right)^n \exp \left\{-\frac{1}{2\eta_0^2} \sum_{i=1}^{n} (y_i^2 - n\bar{y}^2) \right\} \frac{\eta_0}{\sqrt{n}} \cdot \left[ -\frac{\eta_0}{\sqrt{n}} \exp \left\{-\frac{n}{2\eta_0^2} (y_{1:n} - \bar{y})^2 \right\} + \bar{y} \sqrt{2\pi} \Phi \left( \frac{\sqrt{n}}{\eta_0} (y_{1:n} - \bar{y}) \right) \right]
\]
and
\[ \int_{-\infty}^{+\infty} f_0(y_1 - u, \ldots, y_n - u) du = \left( \frac{\sqrt{2}}{\eta_0 \sqrt{\pi}} \right)^n \exp \left\{ -\frac{1}{2\eta_0^2} \left( \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \right) \right\} \frac{\eta_0}{\sqrt{n}} \sqrt{2\pi} \Phi \left[ \frac{\sqrt{n}}{\eta_0} (y_{1:n} - \bar{y}) \right] \]
and the statement follows easily from these expressions. \( \square \)

Unlike what happens with the location parameter \( \xi \), for the scale parameter \( \eta \) an explicit expression for the MRE estimator is obtained.

We consider the scale-location family of densities
\[ f(\xi, \eta)(y_1, \ldots, y_n) = \frac{1}{\eta^n} f \left( \frac{y_1 - \xi}{\eta}, \ldots, \frac{y_n - \xi}{\eta} \right), \]
where
\[ f(y_1, \ldots, y_n) = \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} y_i^2 \right\} \cdot I_{[0, +\infty]}(y_{1:n}). \]
This family remains invariant under transformations of the form \( g_{a,b}(y_1, \ldots, y_n) = (a + by_1, \ldots, a + by_n) \), \( a \in \mathbb{R}, b > 0 \).

**Proposition 3.** When using the loss function \( W_1(x; \xi, \eta) = \eta^{-2} (x - \eta)^2 \), the MRE estimator \( \hat{\eta} \) of \( \eta \)
is
\[ \hat{\eta}(y) = \sqrt{\frac{n-1}{2}} \cdot \Gamma \left( \frac{n+1}{2} \right) \cdot \frac{t(n+1)}{t(n+2)} \left( \frac{C(n+1)}{\sqrt{n-1}} \right) \cdot S(y) \]
where \( t(n) \) denotes Student’s \( t \)-distribution with \( n \) degrees of freedom and \( S^2 \) is the sample variance.

**Proof.** The MRE estimator of the scale parameter \( \eta \), when using the loss function \( W_1 \), is
\[ \hat{\eta}(y) = \frac{\int_0^{+\infty} v^n f'(vy_1', \ldots, vy_{n-1}') dv}{\int_0^{+\infty} v^{n+1} f'(vy_1', \ldots, vy_{n-1}') dv}, \]
where \( f' \) is the joint density when \( \eta = 1 \) of \( Y_i' := Y_i - Y_n, 1 \leq i \leq n - 1 \), and \( y_i' := y_i - y_n, 1 \leq i \leq n - 1 \).

Notice that
\[ f'(y_1', \ldots, y_{n-1}') = \int_{-\infty}^{+\infty} f(y_1 + t, \ldots, y_n + t) dt \]
\[ = \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} y_i^2 + \frac{n}{2} \bar{y}^2 \right\} \int_{-y_{1:n}}^{\infty} \exp \left\{ -\frac{n}{2} (t + \bar{y})^2 \right\} dt \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} (n-1)S^2(y) \right\} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} \exp \left\{ -\frac{1}{2} u^2 \right\} du \]
Hence, for \( k \in \mathbb{N} \), applying Fubini’s Theorem after a suitable change of variables in the inner integral,
\[ I_k(y) := \int_0^{\infty} v^k f'(vy_1', \ldots, vy_{n-1}') dv \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \int_0^{\infty} v^k \exp \left\{ -\frac{1}{2} (n-1)S^2(y) \right\} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} \exp \left\{ -\frac{1}{2} u^2 \right\} du dv \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} J_k(t, y) dt \]
\[ J_k(t, y) := \int_0^{\infty} v^k \exp \left\{ -\frac{1}{2} S^2(y) \right\} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} \exp \left\{ -\frac{1}{2} u^2 \right\} du dv \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} J_k(t, y) dt \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} J_k(t, y) dt \]
\[ = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \int_{\sqrt{n}(\bar{y} - y_{1:n})}^{\infty} J_k(t, y) dt \]
9
where  
\[ J_k(t,y) := \int_0^\infty v^{k+1} \exp \left\{ -\frac{1}{2} v^2 (t^2 + (n-1)S^2(y)) \right\} \, dv = \frac{2^{k/2} \Gamma \left( \frac{k+2}{2} \right)}{(t^2 + (n-1)S^2(y))^\frac{k+2}{2}} \]

where, for \( t \geq \sqrt{n}(\bar{y} - y_{1:n}) \), we have made the change of variables \( w = \frac{1}{2} v^2 (t^2 + (n-1)S^2(y)) \).

So
\[
I_k(y) = \frac{1}{\sqrt{n}} \left( \frac{2}{\pi} \right)^{\frac{k}{2}} 2^{k/2} \Gamma \left( \frac{k+2}{2} \right) \int_{\sqrt{n}(\bar{y} - y_{1:n})}^\infty \frac{dt}{\sqrt{n} \pi t^{k+2} (n-1)^{\frac{k+2}{2}} S(y)^{k+1}} \cdot t(k+1) \left( \left[ \frac{n(k+1)}{n-1} \right] \bar{y} - y_{1:n} \right) \left( S(y)^{k+1} \right) \cdot \frac{1}{S(y)}. 
\]

Finally
\[
\hat{\eta}(y) = \frac{I_n(y)}{I_{n+1}(y)} = \sqrt{\frac{n-1}{2}} \frac{\Gamma \left( \frac{n+1}{2} \right) t(n+1)}{\Gamma \left( \frac{n+2}{2} \right) t(n+2)} \left( \left[ \frac{n(n+1)}{n-1} \right] \bar{y} - y_{1:n} \right) \left( \frac{S(y)^{k+1}}{S(y)} \right)^{\frac{1}{2}}. 
\]

\[ \square \]

**Remark.** To compare the behavior of the unbiased estimator \( \hat{\eta} \), the maximum likelihood estimator \( \hat{\eta} \) and the MRE estimator \( \hat{\eta} \), we have made a simulation study for different sample sizes \((n = 10, 20, 30)\) of a general half-normal distribution \( HN(10,4) \); from 10000 values of the corresponding estimators we have simulated its mean and its mean squared error (MSE). The next table contains the results:

| \( n \) | Mean | MSE | \( \tilde{\eta} \) | MSE | \( \hat{\eta} \) | MSE | \( \hat{\eta} \) | MSE |
|---|---|---|---|---|---|---|---|---|
| 10 | 3.9977 | 1.1024 | 3.5220 | 1.0055 | 3.5698 | 0.9832 |
| 20 | 3.9940 | 0.4932 | 3.7595 | 0.4487 | 3.7942 | 0.4404 |
| 30 | 3.9971 | 0.3255 | 3.8339 | 0.2909 | 3.8600 | 0.2865 |

Table 5. Simulated mean and MSE of the estimators \( \hat{\eta} \), \( \tilde{\eta} \) and \( \hat{\eta} \).

Obviously, the MRE estimator \( \hat{\eta} \) always exhibit the minimum squared error, as \( \tilde{\eta} \) and \( \hat{\eta} \) are equivariant estimators of \( \eta \). Notice also the biased character of the maximum likelihood and MRE estimators. \( \square \)

**Remark.** Although less interesting for the applications, let us consider now the problem of estimating the scale parameter \( \eta \) when the position parameter \( \xi \) is known, say \( \xi = \xi_0 \). After the shift \((y_1, \ldots, y_n) \mapsto (y_1 - \xi_0, \ldots, y_n - \xi_0)\), the statistical model remains invariant under the transformations (dilations) of the form \((y_1, \ldots, y_n) \mapsto (ay_1, \ldots, ay_n)\), for \( a > 0 \). For the loss function \( W_1(y, x) = (x - \eta)^2/\eta^2 \), the MRE estimator of the scale parameter \( \eta \) is
\[
T_2 = \frac{\Gamma \left( \frac{n+1}{2} \right)}{\sqrt{2} \Gamma \left( \frac{n+2}{2} \right)} \left\{ \sum_{i=1}^n (Y_i - \xi_0)^2 \right\} = \frac{B \left( \frac{n+1}{2}, \frac{1}{2} \right)}{\sqrt{2\pi}} \sqrt{\sum_{i=1}^n (Y_i - \xi_0)^2} 
\]

where \( \Gamma \) and \( B \) denote Euler’s Gamma and Beta functions. In fact, for the loss function \( W_1 \), the MRE estimator of \( \eta \) is
\[
T_2(y_1, \ldots, y_n) = \frac{\int_0^\infty v^n h_1(v(y_1 - \xi_0), \ldots, v(y_n - \xi_0)) \, dv}{\int_0^\infty v^{n+1} h_1(v(y_1 - \xi_0), \ldots, v(y_n - \xi_0)) \, dv },
\]

where
\[
h_1(y_1, \ldots, y_n) = \left( \frac{2}{\pi} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n y_i^2 \right\} \cdot I_{[0, +\infty]}(y_{1:n}).
\]
To facilitate the notation, we suppose without loss of generality that $\xi_0 = 0$. The change of variables $t = \frac{1}{2} \sum_{i=1}^{n} y_i^2 v^2$ shows that, for $k = n, n + 1$,

$$
\int_{0}^{\infty} v^k h_1(vy_1, ..., vy_n) dv = 2^{\frac{n+k-1}{2}} \pi^{-\frac{k}{2}} \left( \sum_{i=1}^{n} y_i^2 \right)^{-\frac{k+1}{2}} \Gamma \left( \frac{k+1}{2} \right) I_{[0, +\infty)}(y_1; n),
$$

and the assertion follows easily from this.

Note also that, when $\xi = \xi_0$,

$$
\frac{1}{n} \sum_{i=1}^{n} (Y_i - \xi_0)^2
$$

is the minimum variance unbiased estimator of $\eta^2$. This is a consequence of the Lehmann-Scheffé Theorem and the facts that $\sum_{i=1}^{n} (Y_i - \xi_0)^2$ is a sufficient and complete statistic and $\eta^{-2} \sum_{i=1}^{n} (Y_i - \xi_0)^2$ has distribution $\chi^2(n)$. A little more work shows that

$$
\frac{\Gamma \left( \frac{n}{2} \right)}{\sqrt{2} \pi^{\frac{n+1}{2}}} \left( \sum_{i=1}^{n} (Y_i - \xi_0)^2 \right) = \frac{B \left( \frac{n}{2}, \frac{1}{2} \right)}{\sqrt{2} \pi} \left( \sum_{i=1}^{n} (Y_i - \xi_0)^2 \right)
$$

is the minimum variance unbiased estimator of $\eta$. □

References

[1] Bland, J.M., The half-normal distribution method for measurement error: two case studies, Unpublished talk available on [http://www-users.york.ac.uk/~mb55/talks/halfnor.pdf](http://www-users.york.ac.uk/~mb55/talks/halfnor.pdf), 2005.

[2] Bland J.M., Altman D.G., Measuring agreement in method comparison studies, *Stat Methods Med Res.* 8, (1999), 135-160.

[3] Besicovitch, A.S. (1945), A general form of the covering principle and relative differentiation of additive functions, I, *Proceedings of the Cambridge Philosophical Society* 41, 103-110.

[4] Besicovitch, A.S. (1946), A general form of the covering principle and relative differentiation of additive functions, II, *Proceedings of the Cambridge Philosophical Society* 42, 205-235.

[5] Daniel, C., Use of half-normal plots in interpreting factorial two-level experiments, *Technometrics* 1, (1959), 311–341.

[6] Johnson, N.L., Kotz, S., Balakrishnan, N., *Continuous Univariate Distributions*, Vol. 1, 2nd Ed.; Wiley: New York, 1994.

[7] Lehmann, E.L., *Theory of Point Estimation*, Wiley, 1983.

[8] Lindqvist, B. H. and Taraldsen, G., Monte Carlo conditioning on a sufficient statistic, *Biometrika* bbf 92, (2005) 451-464.

[9] Mattila, P., *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge University Press, 1995.

[10] Nogales, A.G., Perez, P., Unbiased Estimation for the General Half-Normal Distribution, *Comm. Statist. Theory Methods* (2011), to appear.

[11] Pewsey, A., Large-sample inference for the general half-normal distribution, *Comm. Statist. Theory Methods* 31, (2002), 1045–1054.

[12] Pewsey, A., Improved likelihood based inference for the general half-normal distribution, *Comm. Statist. Theory Methods* 33, (2004), 197–204.