Sampling Distribution of ML Estimators: Cauchy Example

Jan Vrbik

We show how to use the Edgeworth series to construct an accurate approximation to the sampling distribution of the maximum likelihood estimator of a parameter of a Cauchy distribution. We then demonstrate the accuracy of this approximation, valid even for relatively small samples.

■ Introduction

Random variables in statistics have many different distributions; one of them is called Cauchy, and has the following probability density function:

$$f(x) = \frac{\sigma}{\pi} \cdot \frac{1}{\sigma^2 + (x - \mu)^2}, \tag{1}$$

where x can have any real value. The distribution has two parameters μ and σ, which represent its median (the "location" parameter) and semi-interquartile deviation (the "scale" parameter), respectively. This rather unusual distribution has no mean and infinite standard deviation.

The exact parameter values are usually not known, and need to be estimated by repeating the corresponding random experiment independently n times, and converting the information thus gathered into two respective estimates of μ and σ. The best way of doing this is by maximizing the corresponding likelihood function:

$$\prod_{i=1}^{n} \frac{\sigma}{\pi} \cdot \frac{1}{\sigma^2 + (X_i - \mu)^2}, \tag{2}$$

where the X_i are the individual observations, or, equivalently, its logarithm

$$n \ln \sigma - n \ln \pi - \sum_{i=1}^{n} \ln\left[\sigma^2 + (X_i - \mu)^2\right]. \tag{3}$$

The maximization is usually achieved by setting the first derivative (with respect to each parameter) of the previous expression to zero, thus obtaining the so-called normal equations, namely

$$\sum_{i=1}^{n} \frac{X_i - \hat{\mu}}{\hat{\sigma}^2 + (X_i - \hat{\mu})^2} = 0, \tag{4}$$

and

$$\frac{n}{\hat{\sigma}} = 2\,\hat{\sigma} \sum_{i=1}^{n} \frac{1}{\hat{\sigma}^2 + (X_i - \hat{\mu})^2}, \tag{5}$$

where the implicit solution (these equations cannot be solved explicitly for μ and σ) defines the maximum likelihood (ML) estimators of μ and σ, denoted $\hat{\mu}$ and $\hat{\sigma}$, respectively. Since these estimators are always implicit functions of the n random observations, they are both random variables, with their own distribution called the sampling distribution.

The normal equations can be rewritten more elegantly as

$$\overline{\frac{X - \hat{\mu}}{\hat{\sigma}^2 + (X - \hat{\mu})^2}} = 0 \tag{6}$$

and

$$\overline{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + (X - \hat{\mu})^2}} = \frac{1}{2}, \tag{7}$$

where each bar indicates taking the corresponding sample average—X now symbolically represents all n observations.

■ Approximating Sampling Distribution of $\hat{\sigma}$

To simplify our task, we assume that the value of μ is known and equal to zero. From now on, we are thus concerned with estimating only the scale parameter σ, based on the following simplified version of (7):

$$\overline{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + X^2}} = \frac{1}{2}. \tag{8}$$

Furthermore, we can introduce the following so-called pivot

$$\hat{\Gamma} \equiv \frac{\hat{\sigma}}{\sigma}, \tag{9}$$

which reduces (8) to

$$\frac{\overline{\hat{\Gamma}^2}}{\hat{\Gamma}^2 + \left(\frac{X}{\sigma}\right)^2} \equiv \frac{\overline{\hat{\Gamma}^2}}{\hat{\Gamma}^2 + U^2} = \frac{1}{2}, \tag{10}$$

where $U_1, U_2, \ldots, U_n$ is a random independent sample from the Cauchy distribution with $\mu = 0$ and the scale parameter equal to 1. It is thus obvious that $\hat{\Gamma}$ (and its sampling distribution) will be free of σ.

We first expand $\hat{\Gamma}$ as a power series in ε (a "smallness" parameter—ε^j indicates that its coefficient is proportional to $n^{-j/2}$), thus:

$$\hat{\Gamma} = 1 + \varepsilon\, g_1 + \varepsilon^2\, g_2 + \varepsilon^3\, g_3 + \ldots. \tag{11}$$

The first term is 1 since the corresponding expansion of $\hat{\sigma}$ must start with σ—this is a general property of a maximum likelihood estimator. We then similarly expand the expression under the bar in (10).

```
XP[q_, k_] := Series[q, {ε, 0, k}] // Normal // Expand;

lhs = XP[ Γ² / (Γ² + U²) /. {Γ → 1 + Sum[gᵢ εⁱ, {i, 3}]}, 3]
```

$$\frac{1}{1 + U^2} + \frac{2\,U^2\,\varepsilon\,g_1}{\left(1 + U^2\right)^2} - \frac{3\,U^2\,\varepsilon^2\,g_1^2}{\left(1 + U^2\right)^3} + \frac{U^4\,\varepsilon^2\,g_1^2}{\left(1 + U^2\right)^3} + \frac{4\,U^2\,\varepsilon^3\,g_1^3}{\left(1 + U^2\right)^4} -$$

$$\frac{4\,U^4\,\varepsilon^3\,g_1^3}{\left(1 + U^2\right)^4} + \frac{2\,U^2\,\varepsilon^2\,g_2}{\left(1 + U^2\right)^3} + \frac{2\,U^4\,\varepsilon^2\,g_2}{\left(1 + U^2\right)^3} - \frac{6\,U^2\,\varepsilon^3\,g_1\,g_2}{\left(1 + U^2\right)^4} -$$

$$\frac{4\,U^4\,\varepsilon^3\,g_1\,g_2}{\left(1 + U^2\right)^4} + \frac{2\,U^6\,\varepsilon^3\,g_1\,g_2}{\left(1 + U^2\right)^4} + \frac{2\,U^2\,\varepsilon^3\,g_3}{\left(1 + U^2\right)^4} + \frac{4\,U^4\,\varepsilon^3\,g_3}{\left(1 + U^2\right)^4} + \frac{2\,U^6\,\varepsilon^3\,g_3}{\left(1 + U^2\right)^4}$$

Placing a bar over each term and then replacing every $\overline{\dfrac{U^k}{\left(1+U^k\right)^m}}$ by

$\mathbb{E}\left[\dfrac{U^k}{\left(1+U^2\right)^m}\right] + \varepsilon\left(\overline{\dfrac{U^k}{\left(1+U^2\right)^m}} - \mathbb{E}\left[\dfrac{U^k}{\left(1+U^2\right)^m}\right]\right)$, where $\mathbb{E}$ stands for the corresponding expected value (a plain number), lets us solve for g_1, g_2, and g_3 [1]. Here $\Gamma 1$ stands for $\hat{\Gamma} - 1$.

```mathematica
L[q_] :=
   L[q] = Integrate[ q/(1 + U^2) , {U, -Infinity, Infinity}] / π;

Γ1 = Module[{lhs, eq, g},

   lhs = XP[ Γ^2/(Γ^2 + U^2)  /. {Γ → 1 + {g[1], g[2], g[3]}.{ε, ε^2, ε^3}}, 3];

   eq =
    XP[

     lhs /. U^k_ (1 + U^2)^m_ :>

        L[ U^k/(1 + U^2)^-m ] + ε ( U^k/(1 + U^2)^-m ‾ - L[ U^k/(1 + U^2)^-m ])

   /. (1 + U^2)^-1 :> L[ 1/(1 + U^2) ] + ε ( 1/(1 + U^2) ‾ - L[ 1/(1 + U^2) ]), 3];
   Do[
      g[i] =
       Expand[g[i] /. Solve[Coefficient[eq, ε^i] == 0, g[i]][[
          1]]], {i, 3}]; {g[1], g[2], g[3]}.{ε, ε^2, ε^3}
]
```

$$
-4\,\varepsilon\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}} + \varepsilon^2 \left(32\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}} + 8\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^2\right) +
$$

$$
\varepsilon^3 \left(-256\,\overline{-\frac{1}{16}+\frac{U^2}{\left(1+U^2\right)^3}}\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}} - \right.
$$

$$
256\,\overline{-\frac{1}{16}+\frac{U^4}{\left(1+U^2\right)^3}}\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}} +
$$

$$
128\,\overline{-\frac{1}{16}+\frac{U^2}{\left(1+U^2\right)^3}}\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^2 - 128\,\overline{-\frac{1}{16}+\frac{U^4}{\left(1+U^2\right)^3}}
$$

$$
\left. \overline{-\frac{1}{2}+\frac{1}{1+U^2}}^2 - 128\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^2 - 16\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^3\right)
$$

◻ **Cumulants**

The next step is to compute the expected value, variance, and the next two cumulants of $\hat{\Gamma}$. We recall the definition of cumulants as coefficients of $t^i / i!$ in the following expansion.

```
Series[Log[1 + Sum[μᵢ tⁱ / i!, {i, 2, 4}]], {t, 0, 4}]
```

$$\frac{\mu_2\, t^2}{2} + \frac{\mu_3\, t^3}{6} + \frac{1}{24}\left(-3\,\mu_2^2 + \mu_4\right) t^4 + O[t]^5$$

This tells us that the second and third cumulants are equal to μ_2 and μ_3, respectively, but the fourth cumulant is $\mu_4 - 3\,\mu_2^2$, where μ_i denotes the corresponding central moment, that is,

$$\mu_i \equiv \mathbb{E}\left[\left(\hat{\Gamma} - \kappa_1\right)^i\right], \tag{12}$$

and κ_1 is the expected value of $\hat{\Gamma}$ (which also happens to be the first cumulant).

To find these, we need a procedure for computing the expected value of expressions involving products and powers of all possible sample averages (our bars) in our previous expansion of $\Gamma 1$. This is trivial for a single bar (its expected value is always zero), and relatively easy for a product of two or three bars (including powers, as a special case), since one can show that

$$\mathbb{E}\left[\overline{f_1(U)} \cdot \overline{f_2(U)}\right] = \frac{\mathbb{E}[f_1(U) \cdot f_2(U)]}{n}, \tag{13}$$

and

$$\mathbb{E}\left[\overline{f_1(U)} \cdot \overline{f_2(U)} \cdot \overline{f_3(U)}\right] = \frac{\mathbb{E}[f_1(U) \cdot f_2(U) \cdot f_3(U)]}{n^2}. \tag{14}$$

For example,

$$\mathbb{E}\left[\overline{\left(-\frac{1}{8} + \frac{U^2}{\left(1+U^2\right)^2}\right)} \cdot \overline{\left(-\frac{1}{2} + \frac{1}{1+U^2}\right)^2}\right] =$$

$$\frac{1}{n\,\pi}\int_{-\infty}^{\infty}\frac{1}{1+U^2}\left(-\frac{1}{8} + \frac{U^2}{\left(1+U^2\right)^2}\right)\left(-\frac{1}{2} + \frac{1}{1+U^2}\right)^2 dU = -\frac{1}{128\,n}, \tag{15}$$

and so on.

Things get more complicated when the overall power is four or higher (we need to go up to six). In such a case, we first need to find all partitions of $f_1(U), f_2(U), f_3(U), \ldots$, where the individual subsets must have at least two elements each. Thus, for example, having four terms, we either take all of them as a single subset or partition them into two pairs,

$$\{f_1, f_2\}, \{f_3, f_4\} \qquad\qquad f_i$$

$$f_1(U), f_2(U), f_3(U), \ldots$$

such as $\{f_1, f_2\}, \{f_3, f_4\}$—there are three ways of doing this. Similarly, having six f_i, we can either take all of them or split them into two triplets (10 ways), a pair and a quadruplet (15 ways), or three pairs (15 ways).

The contribution of each of these partitions is then computed using the following scheme: we multiply the expected values of the product of the f_i in each subset, multiply the answer by $n(n-1)(n-2)\ldots(n+1-k)$, where k is the number of these subsets, and divide by n^k. Thus, for example the $\{f_1, f_2\}, \{f_3, f_4\}, \{f_5, f_6\}\}$ partition contributes

$$\frac{n(n-1)(n-2) \cdot \mathbb{E}[f_1\, f_2] \cdot \mathbb{E}[f_3\, f_4] \cdot \mathbb{E}[f_5\, f_6]}{n^3}. \tag{16}$$

To find, for example, $\mathbb{E}\left[\overline{f_1} \cdot \overline{f_2} \cdot \overline{f_3} \cdot \overline{f_4} \cdot \overline{f_5} \cdot \overline{f_6}\right]$, all such contributions from all corresponding partitions must be added together. This is done by the following program.

```
EV[x_, io_] :=
  Quiet@CL[Distribute[H[Expand[x] /. ε^j_/;j>2 io → 0]] /. H → Q /.
     ε → 1, io];
CL[x_, k_] := Collect[x, n] /. n^i_/;i<-k → 0;
H[q_ w_] := q H[w] /; FreeQ[q, U];
H[q___, w_ t_, s___] := Apply[H, Sort[{q, w, t, s}]];
H[w___, q_^i_, t___] :=
  Apply[H, Sort[{w, Table[q, {i}], t} // Flatten]];
Quiet[Needs["Combinatorica`"]];
Q[q__] := Module[{k = Length[{q}], r, s},
   r = Select[SetPartitions[{q}],
 FreeQ[Map[Length, #], 1] &]; s = Apply[Fa, r, {2}];
 r = Apply[Times, s, {1}] Map[P, Map[Length, s]] / n^k;
    Total[CL[r, io]] /. Fa → F];
P[i_] := Product[n + 1 - j, {j, i}];
F[q__] := Module[{w = {q} /. a_ → a}, L[Apply[Times, w]]];
```

The second argument of EV specifies the accuracy to be achieved (up to $\frac{1}{n}$ when the argument is equal to 1, up to $\frac{1}{n^2}$ when it is equal to 2, etc.). At this point, ε becomes a redundant parameter and is eliminated by setting it to 1.

We can thus get the expected value, variance, and the next two cumulants of $\hat{\Gamma}$.

```
κ₁ = 1 + EV[Γ1, 1]
κ₂ = EV[Γ1², 2] - EV[Γ1, 1]²
κ₃ = EV[Γ1³, 2] - 3 EV[Γ1², 1] EV[Γ1, 1]
κ₄ = CL[EV[Γ1⁴, 3] - 4 EV[Γ1³, 2] EV[Γ1, 1] +
    6 EV[Γ1², 1] EV[Γ1, 1]² - 3 κ₂², 3]
```

$$1 + \frac{1}{n}$$

$$\frac{7}{n^2} + \frac{2}{n}$$

$$\frac{12}{n^2}$$

$$\frac{138}{n^3}$$

□ Edgeworth Expansion

One can show [2] that a good approximation to the probability density function of

$$Z \equiv \frac{\hat{\Gamma} - \kappa_1}{\sqrt{\kappa_2}} \tag{17}$$

is provided by

$$\frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} \left(1 + \frac{\gamma_3}{6} H_3(z) + \frac{\gamma_4}{24} H_4(z) + \frac{\gamma_3^2}{72} H_6(z) + \ldots\right), \tag{18}$$

where $\gamma_3 \equiv \frac{\kappa_3}{\kappa_2^{3/2}}$ and $\gamma_4 \equiv \frac{\kappa_4}{\kappa_2^2}$ are the normalized cumulants, and $H_i(z)$ are simple polynomials defined by

$$H_i(z) \equiv (-1)^i \exp\left(\frac{z^2}{2}\right) \cdot \frac{d^i \exp\left(-\frac{z^2}{2}\right)}{dz^i}. \tag{19}$$

These are closely related to Hermite polynomials; the exact correspondence is clear from the first command.

```
H_i_[z_] := HermiteH[i, z/√2] 2^(-i/2) // Expand;
```

$$\texttt{pst} = \frac{\text{Exp}\left[-z^2/2\right]}{\sqrt{2\,\pi}}\left(1 + \frac{\kappa_3}{6\,\kappa_2^{3/2}}\,H_3[z] + \frac{\kappa_4}{24\,\kappa_2^2}\,H_4[z] + \frac{\kappa_3^2}{72\,\kappa_2^3}\,H_6[z]\right);$$
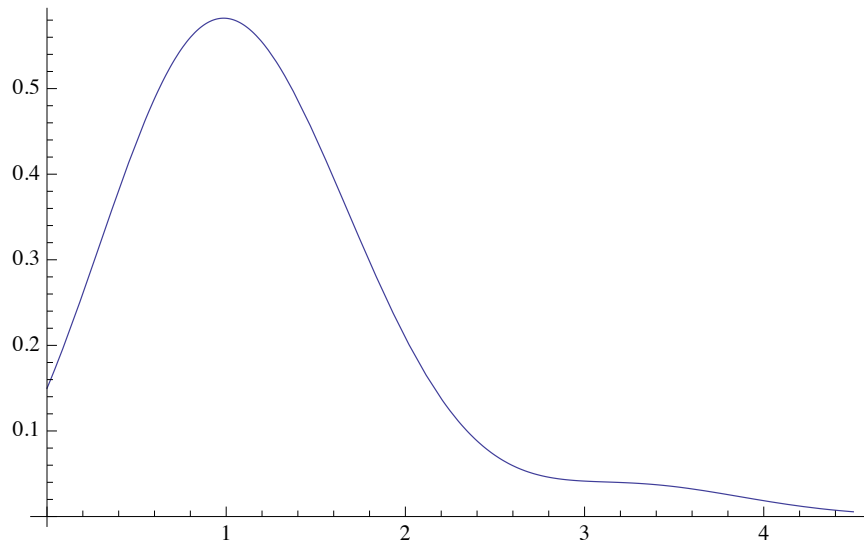
The resulting expression can be easily transformed into a probability density function of $\hat{\Gamma}$, which can then be plotted for $n = 5$, for example.

$$\texttt{pdf} = \left(\texttt{pst} /. z \rightarrow \frac{s - \kappa_1}{\sqrt{\kappa_2}}\right)\Big/ \sqrt{\kappa_2};$$

```
Plot[pdf /. n → 5 // Simplify, {s, 0, 4.5}]
```

■ Comparison with Empirical Distribution

The best way of verifying the accuracy of this approximation is to generate a large collection of random independent samples of size five from a Cauchy distribution (with the location and scale parameters equal to 0 and 1, respectively), for each such sample compute the value of $\hat{\Gamma}$ by solving (10), and then display these $\hat{\Gamma}$-values in a histogram.

```
res = Table[0, {10 000}];
Do[u = RandomReal[CauchyDistribution[0, 1], 5];
 res[[i]] =
   Abs[Γ] /. FindRoot[Total[Γ² / (Γ² + u²)] == 5/2, {Γ, 1}],
 {i, 10 000}]
hist = Histogram[res];
```

We are now in a position to check how well our approximate probability density function, correspondingly rescaled, matches this histogram (which, for large total frequency, represents—up to small statistical fluctuations—the exact distribution of $\hat{\Gamma}$).

```
Show[hist, Plot[2000 pdf /. n → 5, {s, 0, 4.5}],
 PlotRange → {{0, 4.5}, {0, 1600}}]
```
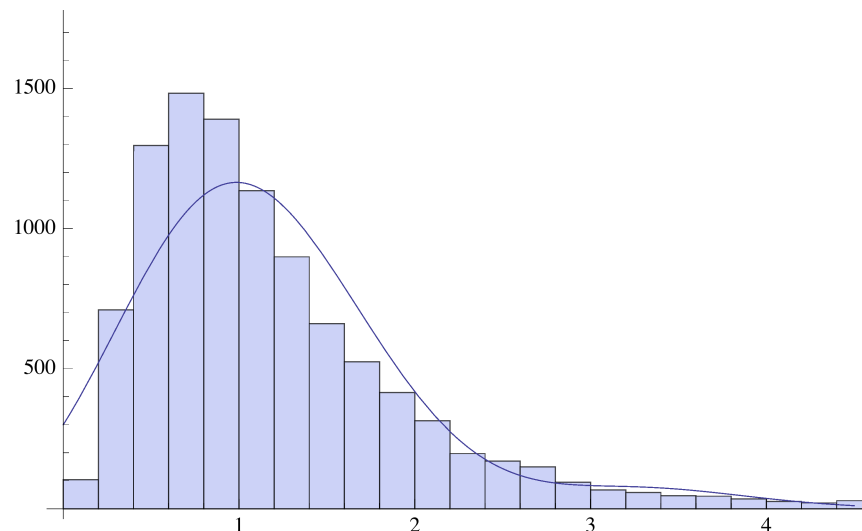

Even though this is an improvement over the usual normal approximation, it is still far from a reasonably good match (but one should remember that five is an extremely small sample size—things do improve, rather dramatically, as n increases).

Nevertheless, there is one particular technique one can employ to achieve a substantially better agreement, even with the current value of $n = 5$.

□ Transforming $\hat{\sigma}$

We can use the same approach and program to find an approximation for the distribution of $T(\hat{\sigma})$, where T is an arbitrary (preferably increasing) function of a non-negative argument. This can be achieved by defining $\mathtt{T}\sigma$ and computing its first four cumulants.

```
Tσ = Series[T[(1 + Γ1) σ] - T[σ], {ε, 0, 3}] // Normal
```

$$-4\,\epsilon\,\sigma\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}\;T'[\sigma]\,+$$

$$\epsilon^2\left(\left(32\,\sigma\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\;\overline{-\frac{1}{2}+\frac{1}{1+U^2}}+8\,\sigma\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,2}\right)T'[\sigma]\,+\right.$$

$$\left. 8\,\sigma^2\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,2}\;T''[\sigma]\right)\,+$$

$$\epsilon^3\left(\left(-256\,\sigma\,\overline{-\frac{1}{16}+\frac{U^2}{\left(1+U^2\right)^3}}\;\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\;\overline{-\frac{1}{2}+\frac{1}{1+U^2}}\,-\right.\right.$$

$$256\,\sigma\,\overline{-\frac{1}{16}+\frac{U^4}{\left(1+U^2\right)^3}}\;\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\;\overline{-\frac{1}{2}+\frac{1}{1+U^2}}\,+$$

$$128\,\sigma\,\overline{-\frac{1}{16}+\frac{U^2}{\left(1+U^2\right)^3}}\;\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,2}-128\,\sigma\,\overline{-\frac{1}{16}+\frac{U^4}{\left(1+U^2\right)^3}}$$

$$\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,2}-128\,\sigma\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\;\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,2}-$$

$$\left. 16\,\sigma\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,3}\right)T'[\sigma]-4\,\sigma\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}$$

$$\left(32\,\sigma\,\overline{-\frac{1}{8}+\frac{U^2}{\left(1+U^2\right)^2}}\;\overline{-\frac{1}{2}+\frac{1}{1+U^2}}+8\,\sigma\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,2}\right)T''[\sigma]\,-$$

$$\left.\frac{32}{3}\,\sigma^3\,\overline{-\frac{1}{2}+\frac{1}{1+U^2}}^{\,3}\;T^{(3)}[\sigma]\right)$$

```
κ₁ = T[σ] + EV[Tσ, 1]
κ₂ = Collect[EV[Tσ², 2] - EV[Tσ, 1]², n, Simplify]
κ₃ = EV[Tσ³, 2] - 3 EV[Tσ², 1] EV[Tσ, 1] // Simplify
κ₄ =
  CL[EV[Tσ⁴, 3] - 4 EV[Tσ³, 2] EV[Tσ, 1] + 6 EV[Tσ², 1] EV[Tσ, 1]² -
    3 κ₂², 3] // Simplify
```

$$T[\sigma] + \frac{\sigma\, T'[\sigma] + \sigma^2\, T''[\sigma]}{n}$$

$$\frac{2\, \sigma^2\, T'[\sigma]^2}{n} + \frac{1}{n^2}$$
$$\sigma^2 \left(7\, T'[\sigma]^2 + 2\, \sigma^2\, T''[\sigma]^2 + 4\, \sigma\, T'[\sigma]\, \left(4\, T''[\sigma] + \sigma\, T^{(3)}[\sigma]\right)\right)$$

$$\frac{12\, \sigma^3\, T'[\sigma]^2\, (T'[\sigma] + \sigma\, T''[\sigma])}{n^2}$$

$$\frac{1}{n^3}$$
$$2\, \sigma^4\, T'[\sigma]^2 \left(69\, T'[\sigma]^2 + 48\, \sigma^2\, T''[\sigma]^2 + 16\, \sigma\, T'[\sigma]\, \left(9\, T''[\sigma] + \sigma\, T^{(3)}[\sigma]\right)\right)$$

The third cumulant can be thus made equal to zero for any σ by solving the following differential equation.

```
DSolve[T'[σ] + σ T''[σ] == 0, T[σ], σ]
```

$$\{\{T[\sigma] \rightarrow C[2] + C[1]\, Log[\sigma]\}\}$$

The simplest nontrivial solution is clear.

```
T[σ_] := Log[σ]
```

Now, we have to recompute the first four cumulants of Tσ.

```
κ₁ = T[σ] + EV[Tσ, 1]
κ₂ = Collect[EV[Tσ², 2] - EV[Tσ, 1]², n, Simplify]
κ₃ = EV[Tσ³, 2] - 3 EV[Tσ², 1] EV[Tσ, 1] // Simplify
κ₄ =
  CL[EV[Tσ⁴, 3] - 4 EV[Tσ³, 2] EV[Tσ, 1] + 6 EV[Tσ², 1] EV[Tσ, 1]² -
    3 κ₂², 3] // Simplify
```

```
Log[σ]
```

$$\frac{1}{n^2} + \frac{2}{n}$$

```
0
```

$$\frac{10}{n^3}$$

Having reduced κ_3 to zero not only greatly simplifies the approximate probability density function of

$$\frac{\ln(\hat{\sigma}) - \ln(\sigma)}{\sqrt{\frac{2}{n} + \frac{1}{n^2}}}, \tag{20}$$

but it also makes the corresponding approximation substantially more accurate.

This is now the new approximate probability density function.

```
pst = Exp[-z²/2]/√2π  Series[1 + κ₄/(24 κ₂²) H₄[z], {n, Infinity, 1}] //

  Normal
```
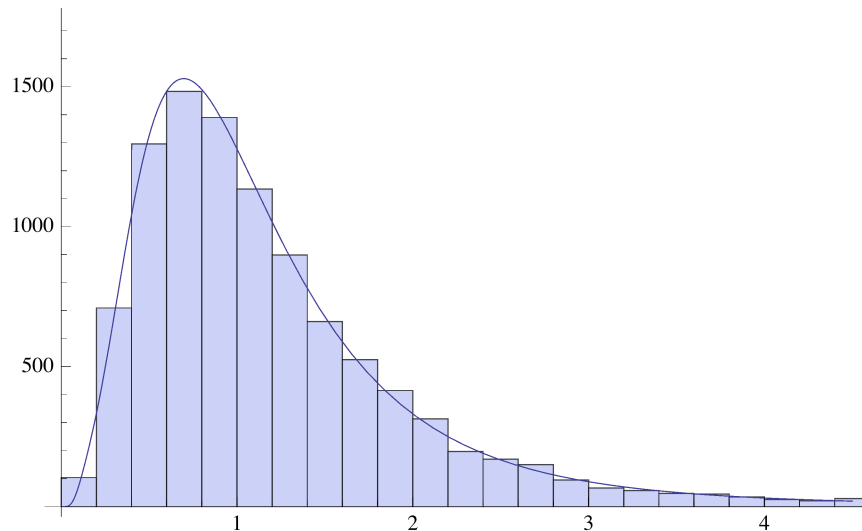
$$\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} + \frac{5 e^{-\frac{z^2}{2}} \left(3 - 6 z^2 + z^4\right)}{48 n \sqrt{2\pi}}$$

The improvement can be seen by converting the previous probability density function of (20) to that of $\hat{\Gamma}$.

```
pdf = (pst /. z → Log[s]/√κ₂) / √κ₂ / s;
```

Let us now compare the last function to the original histogram.

```
Show[hist, Plot[2000 pdf /. n → 5, {s, 0, 4.5}],
 PlotRange → {{0, 4.5}, {0, 1600}}]
```

This has now produced a rather respectable agreement between the new approximation and the exact distribution.

□ Potential Extensions

In a similar manner, one can construct an approximate joint probability density function for both $\hat{\mu}$ and $\hat{\sigma}$. This requires the corresponding generalization of cumulants, Hermite polynomials, and the Edgeworth expansion. Readers with enough statistical expertise should have no difficulty pursuing that direction on their own since, surprisingly, the required modifications of the program presented in this article would not be that substantial. The most difficult part, namely the routine for computing the expected value of products and powers of sample averages, needs no change at all.

The same technique can also be used to find a good approximation to a probability density function of almost any parameter estimator of a specific distribution.

■ References

[1] D. F. Andrews and J. E. Stafford, *Symbolic Computation for Statistical Inference*, New York: Oxford University Press, 2000.

[2] F. Y. Edgeworth, "The Law of Error," *Transactions of the Cambridge Philosophical Society*, **20**, 1905 pp. 36–66 and 113–141.
books.google.com/books/about/
Transactions_of_the _Cambridge _Philosophi.html?id=uTs8AAAAMAAJ.

About the Author

Jan Vrbik
Department of Mathematics, Brock University
500 Glenridge Ave., St. Catharines
Ontario, Canada, L2S 3A1
jvrbik@brocku.ca