Journal of Universal Computer Science, vol. 12, no. 9 (2006), 1332-1344
submitted: 31/12/05, accepted: 12/5/06, appeared: 28/9/06 © J.UCS

Parameter Estimation of the Cauchy Distribution in Information Theory Approach

Ferenc Nagy
(University of Miskolc, Hungary
matnf@gold.uni-miskolc.hu)

Abstract: As we know the Cauchy distribution plays an important role in Probability Theory and Statistics. In this paper, we investigate the estimation of the location and the scale parameter. Both the one-dimensional problem and the multidimensional problem are studied for large sample. In the one-dimensional case, we give two algorithms for the estimation. The first one is an iterative method for which we prove the convergence and we show that the rate of convergence is geometric. The second algorithm provides an exact solution to the problem. In the multidimensional case, we give an algorithm analogous to the one-dimensional case. Computer experiments show that the rate of convergence is similar to the one-dimensional iterative algorithm.

Keywords: Parameter estimation, discrimination information, inaccuracy, Cauchy distribution
Categories: G.3

1 Introduction

Let ξ, ξ_1 and ξ_2 be real random variables with density functions $f(x), f_1(x)$ and $f_2(x)$ respectively. Denote by $F(x), F_1(x)$ and $F_2(x)$ their corresponding distribution functions. The three expected values given below define respectively the *entropy*, *inaccuracy* and *discrimination information* [Mathai].

$$H(\xi) = \int_R \log \frac{1}{f(x)} dF(x) \tag{1}$$

$$T(\xi_1 \| \xi_2) = \int_R \log \frac{1}{f_2(x)} dF_1(x) \tag{2}$$

$$D(\xi_1 \| \xi_2) = \int_R \log \frac{f_1(x)}{f_2(x)} dF_1(x), \tag{3}$$

where $R=(-\infty,\infty)$ and log means the natural logarithm. The above definitions can be extended to m-dimensional case. We point out that the discrimination information is nonnegative and can assume zero in the only case when the two density functions coincide almost everywhere. It defines some kind of directed distance measure between two distributions. It can easily be seen that

$$D(\xi_1 \| \xi_2) = T(\xi_1 \| \xi_2) - H(\xi_1) \tag{4}$$

It follows from (4) that the minimum of $T(\xi_1 \| \xi_2)$ at fixed ξ_1 is $T(\xi_1 \| \xi_1) = H(\xi_1)$. Random variables ξ_1, ξ_2 used in formulas (2) and (3) are called accordingly *posterior* and *prior* variables. In the parameter estimation the supposed Cauchy distribution is treated as prior and the empirical distribution computed from the sample is used as posterior distribution. For the parameter estimation, we l accept the prior Cauchy distribution which is the closest to the posterior distribution in the sense of discrimination information. To this end, we minimize the discrimination information between the empirical distribution of the sample and the prior Cauchy distribution with respect to the parameters. More precisely: If the density function of the random variable ξ_p is $f_{\xi_p}(x)$ where the parameter *p* takes its values in a set *P*, and η_n is a random variable with distribution function $F_n(x)$ that is an empirical distribution function given by the realizations of ξ_p (sample of size n) then we accept that $\hat{p}$ value as an estimation of *p* for which

$$D(\eta_n \| \xi_{\hat{p}}) = \min_{p \in P} D(\eta_n \| \xi_p) \tag{5}$$

holds. As the entropy of η_n does not depend on *p* (namely *log n*) this is equivalent to the minimization of *T*:

$$T(\eta_n \| \xi_{\hat{p}}) = \min_{p \in P} T(\eta_n \| \xi_p) \tag{6}$$

The inaccuracy in this case is

$$T(\eta_n \| \xi_p) = \int_R \log \frac{1}{f_{\xi_p}(x)} dF_n(x) = \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{f_{\xi_p}(a_i)} \tag{7}$$

Here a_i denotes the i^{th} element of the sample. Formula (7) corresponds to the appropriate formula of the maximum likelihood method multiplied by *–1/n*. Therefore, the location of the minimum of (7) is assumed at the same place as the maximum in the maximum likelihood method of the parameter estimation. Our estimation has the properties of the maximum likelihood estimation [Rao], [Zakhs]. We need large samples in order to apply the Law of large numbers.

2 The One-dimensional Case

We remind that a random variable ξ possesses the Cauchy distribution with location *(-∞<c<∞)* and scale *(s>0)* parameters if its density function is:

$$f_\xi(x) = \frac{1}{\pi \cdot s} \cdot \frac{1}{1+\left(\frac{x-c}{s}\right)^2} \quad for\ x \in R \tag{8}$$

We denote this as $\xi \sim C(c,s)$. The inaccuracy (7) (briefly *T(c,s)*) is in this case:

$$T(c,s) = \sum_{i=1}^{n} \frac{1}{n} \log\left\{\pi \cdot s\left[1+\left(\frac{a_i-c}{s}\right)^2\right]\right\} \tag{9}$$

Let us introduce the following notations, where parameters *c* and *s* are unknown and fixed. Values c_k and s_k play the role of the estimations of the parameters *c* and *s* obtained in the k^{th} step of the iteration.

$$\begin{aligned}
&u_i = (a_i - c)/s,\ i = 1,2,\ldots,n && u_{ik} = (a_i - c_k)/s_k,\ i = 1,2,\ldots,n \\
&e_0 = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+u_i^2}, && e_{0k} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+u_{ik}^2} \\
&e_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{u_i}{1+u_i^2}, && e_{1k} = \frac{1}{n}\sum_{i=1}^{n} \frac{u_{ik}}{1+u_{ik}^2} \\
&e_2 = \frac{1}{n}\sum_{i=1}^{n} \frac{u_i^2}{1+u_i^2}, && e_{2k} = \frac{1}{n}\sum_{i=1}^{n} \frac{u_{ik}^2}{1+u_{ik}^2}
\end{aligned} \tag{10}$$

Lemma 1: The inaccuracy between two Cauchy distributions. IF $\xi_1 \sim C(c_1,s_1)$ and $\xi_2 \sim C(c_2,s_2)$ then

$$T(\xi_1 \| \xi_2) = \log \frac{\pi \cdot \left[(s_1+s_2)^2 + (c_1-c_2)^2\right]}{s_2} \tag{11}$$

Especially, if $\xi_1 = \xi_2 = \xi \sim C(c,s)$ then $H(\xi) = \log(4\pi s)$.

Proof: First we introduce an integral that is be used in the proof.

$$\int_R \frac{\log\left(a^2 - 2abx + x^2\right)}{1+x^2} dx = \pi \cdot \log\left[1 + a^2 + 2a\sqrt{1-b^2}\right], \; a \in R, \, |b| < 1 \quad (12)$$

The proof of it is in the appendix of the paper. The inaccuracy is:

$$T(\xi_1 \| \xi_2) = \int_R \frac{1}{\pi \cdot s_1} \cdot \frac{1}{1 + \left(\frac{x - c_1}{s_1}\right)^2} \log\left\{\pi s_2 \left[1 + \left(\frac{x - c_2}{s_2}\right)^2\right]\right\} dx \quad (13)$$

Substituting $y=(x-c_1)/s_1$ and applying $\int_R \frac{1}{1+y^2} dy = \pi$ we can transform (13) into

$$\begin{aligned} T(\xi_1 \| \xi_2) &= \log(\pi s_2) + \frac{1}{\pi}(2\pi \log s_1 + 2\pi \log s_2) + \\ &+ \frac{1}{\pi} \int_R \frac{1}{1+y^2} \log\left[\frac{s_2^2}{s_1^2} + \frac{(c_1 - c_2)^2}{s_1^2} + 2\frac{c_1 - c_2}{s_1} y + y^2\right] dy \end{aligned} \quad (14)$$

This integral can be computed using (12) by making the substitutions

$$a = \sqrt{\frac{s_2^2}{s_1^2} + \frac{(c_1 - c_2)^2}{s_1^2}}, \quad b = -\frac{c_1 - c_2}{a s_1} \quad (15)$$

By a simple rearrangement we obtain formula (11), which ends the proof.

Lemma 2: The characterization of the location of the minimum. At the place of the minimum of $T(c,s)$ given by (9) there holds:

$$c = c + s \cdot \frac{e_1}{e_0}, \quad s^2 = s^2 \cdot \frac{e_2}{e_0} = s^2 \cdot \left(\frac{1}{e_0} - 1\right) \quad (16)$$

Proof: We can write the inaccuracy between the posterior and the prior distributions as follows:

$$T(c,s) = \sum_{i=1}^{n} \frac{1}{n} \log\left\{\pi s \left[1 + \left(\frac{a_i - c}{s}\right)^2\right]\right\} \quad (17)$$

By the standard way, we compute the partial derivatives with respect to c and s and we equate them to zero. Rearranging the terms after some steps, we arrive at our pair of formulas to be proved. This ends the proof.

By Lemma 2, we give the next pair of iterative formulas as the solution for the system of equations (16):

$$c_{k+1} = c_k + s_k \cdot \frac{e_{1k}}{e_{0k}}, \quad s_{k+1} = s_k \cdot \sqrt{\frac{1}{e_{0k}} - 1}, \quad for\ k = 0,1,2,\ldots \tag{18}$$

Here $-\infty < c_0 < \infty$ and $s_0 > 0$ are arbitrary real numbers.

Theorem 1: Theorem of the convergence. The sequences of pairs of real numbers defined in (18) converge to the theoretical fixed values of the parameters c and s in case of large sample. Moreover, the rate of convergence is geometric.

Proof: The large sample is needed for the possibility of the substitution of e_{0k} and e_{1k} with their expected values according to the Law of large numbers. Let us use notations $E(e_{0k})=v_{0k}$ and $E(e_{1k})=v_{1k}$ where E means the expected value. Let there be $\xi \sim C(c,s)$ and $\xi_k \sim C(c_k, s_k)$. Let us first introduce the following random variable $\xi_{k,p} = \frac{1}{p} \cdot \xi_k + \left(1 - \frac{1}{p}\right) \cdot c_k \sim C(c_k, s_k / p)$. Then we can write the inaccuracy as follows

$$T_p = T(\xi \| \xi_{k,p}) = \log \frac{\pi}{s_k} + \log p + \log \left[\left(s + \frac{s_k}{p} \right)^2 + (c - c_k)^2 \right] \tag{19}$$

We can easily check that $v_{0k} = \frac{1}{2} \cdot \left(1 - p \frac{dT_p}{dp} \right)\Big|_{p=1}$ and finally we obtain the expression for this expected value:

$$v_{0k} = \frac{s_k \cdot (s + s_k)}{(s + s_k)^2 + (c - c_k)^2} \tag{20}$$

In the same way we define another random variable $\xi_{k,p} = \xi_k + (p-1) \cdot c_k \sim C(pc_k, s_k)$. Computing again the inaccuracy we get:

$$T_p = T(\xi \| \xi_{k,p}) = \log \frac{\pi}{s_k} + \log\left[(s + s_k)^2 + (c - pc_k)^2\right] \tag{21}$$

Now we can check that $v_{1k} = -\frac{s_k}{2c_k} \cdot \left.\frac{dT_p}{dp}\right|_{p=1}$ from which we obtain:

$$v_{1k} = \frac{s_k \cdot (c - c_k)}{(s + s_k)^2 + (c - c_k)^2} \tag{22}$$

Making the substitutions in the iterative formulas (18) via the Law of large numbers we obtain that:

$$\begin{aligned} c_{k+1} &= c_k + s_k \cdot \frac{v_{1k}}{v_{0k}} = c_k + s_k \cdot \frac{c - c_k}{s + s_k}, \\ s_{k+1}^2 &= s_k^2 \cdot \left(\frac{1}{v_{0k}} - 1\right) = s_k \cdot \left[s + \frac{(c - c_k)^2}{s + s_k}\right] \end{aligned} \tag{23}$$

From this pair of formulas we can see that if $s_0 > 0$ then $s_k > 0$. In addition, we can see that if $s_k < s$ then $s_{k+1} > s_k$, because $s_{k+1}^2 - s_k^2 = s_k \cdot (s - s_k) + s_k \cdot \frac{(c - c_k)^2}{s + s_k} > 0$. Finally, we can see that if $s_k > s$ then $s_{k+1} > s$ because $s_{k+1}^2 - s^2 = s \cdot (s_k - s) + \frac{s_k}{s + s_k} \cdot (c - c_k)^2 > 0$. Consequently we can observe that $s_{k+1} \geq \min\{s_0, s\}$ and so $s_{k+1} > 0$. We are now in the position to prove the convergence. Let $q = \max\left\{\frac{1}{2}, \frac{s}{s + s_0}\right\}$. It is clear that $\frac{1}{2} \leq q < 1$. The next inequalities hold:

$$\begin{aligned}
\left|c-c_{k+1}\right| &= \left|c-c_k-s_k\cdot\frac{c-c_k}{s+s_k}\right| = \\
&= \left|\left(1-\frac{s_k}{s+s_k}\right)\cdot\left(c-c_k\right)\right| = \\
&= \frac{s}{s+s_k}\cdot\left(c-c_k\right) \le q\left|c-c_k\right| \le \ldots \\
&\le q^{k+1}\left|c-c_0\right| \xrightarrow[k\to\infty]{} 0
\end{aligned} \tag{24}$$

$$\begin{aligned}
\left|s^2-s_{k+1}^2\right| &= \left|s^2-s_k\cdot\left[s+\frac{\left(c-c_k\right)^2}{s+s_k}\right]\right| = \\
&= \left|s^2-ss_k+\frac{s_k}{s+s_k}\left(c-c_k\right)^2\right| = \\
&= \left|s\left(s-s_k\right)+\frac{s_k}{s+s_k}\left(c-c_k\right)^2\right| = \\
&= \left|\frac{s}{s+s_k}\left(s^2-s_k^2\right)+\frac{s_k}{s+s_k}\left(c-c_k\right)^2\right| \le \\
&\le \left|\frac{s}{s+s_k}\right|\left|s^2-s_k^2\right|+\left|\frac{s_k}{s+s_k}\right|\left|c-c_k\right|^2 \le \\
&\le q\left|s^2-s_k^2\right|+\left|c-c_k\right|^2 \le q\left|s^2-s_k^2\right|+q^{2k}\left|c-c_0\right|^2 \le \\
&\le q\left(q\left|s^2-s_{k-1}^2\right|+q^{2(k-1)}\left|c-c_0\right|^2\right)+q^{2k}\left|c-c_0\right|^2 \le \ldots \le \\
&\le q^{k+1}\left|s^2-s_0^2\right|+\left|c-c_0\right|^2\left(q^k+q^{k+1}+\ldots+q^{2k}\right) = \\
&= q^{k+1}\left|s^2-s_0^2\right|+q^k\frac{1-q^{k+1}}{1-q}\left|c-c_0\right|^2 \xrightarrow[k\to\infty]{} 0
\end{aligned} \tag{25}$$

This ends the proof.

Theorem 2: The uniqueness of the location of the minimum. In case of large sample, the location of the minimum is unique.

Proof: It can easily be shown that (7) is an unbiased estimation of the inaccuracy. For large samples (9) gives the inaccuracy between a Cauchy distributed random variable with parameters *(c,s)* and the sample distribution, which we suppose to be in the limit a Cauchy distribution with parameters (c^*,s^*). The place of the minimum of the inaccuracy between them is unique because of the properties of (4). It is an interesting fact, that the contour curves of the inaccuracy in this case are embedded circles in the upper semiplane of the coordinates (c^*,s^*) with centre $(c^*,(1+2r)s^*)$ and radius $2s^*(r(r+1))^{1/2}$ where r comes from $e^t=4\pi s^*(1+r)$ for the contour level t.

A similar problem was studied in [Gabrielsen] and [Copas].

Theorem 3: Exact solution. In the case of the large sample for any c_k and $s_k>0$ there hold:

$$c = c_k + s_k \cdot \frac{v_{1k}}{v_{0k}^2 + v_{1k}^2}, \quad s = s_k \cdot \left(\frac{v_{0k}}{v_{0k}^2 + v_{1k}^2} - 1 \right) \tag{26}$$

Proof: From formulas (20) and (22) we can simply express c and s as the solution of a system of equations with two unknowns. This ends the proof.

Making use of Theorem 3 the value of the inaccuracy to be minimized can be written as:

$$T = \log \frac{\pi \cdot \left[(s + s_k)^2 + (c - c_k)^2 \right]}{s_k} = \log \frac{\pi s_k}{v_{0k}^2 + v_{1k}^2} \tag{27}$$

The given algorithms are related to the infinitely large samples. Their application to the finite case is heuristic. However, computer experiments give convergence even in the case of very small sample sizes. (If the sample size is two and the elements are different then the minimum of the accuracy is not unique, it takes place at a semicircle. The iterative algorithm practically finds this semicircle. The point where the method stops depends on the starting point.)

3 The Multidimensional Case

Let us now deal with the m-dimensional case. The density function of the Cauchy distribution is given by:

$$f_\xi(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\pi^{\frac{m+1}{2}} \cdot \sqrt{\det S}} \cdot \frac{1}{\left[1 + (x-c)^* S^{-1} (x-c)\right]^{\frac{m+1}{2}}} \quad for\ x \in R^m, \tag{28}$$

where c is the m-dimensional location parameter vector, S is an $m \times m$ positive definite symmetric real matrix that plays the role of the scale parameter, and the * denotes the transposed vector or matrix. We use the following notations corresponding to the one-dimensional ones in (10):

$$\begin{aligned} S &= T \cdot T^* & S_k &= T_k \cdot T_k^* \\ u_i &= T^{-1} \cdot (a_i - c), & u_{ik} &= T_k^{-1} \cdot (a_i - c_k), \quad i = 1,\ldots,n \\ e_0 &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+u_i^* u_i}, & e_{0k} &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+u_{ik}^* u_{ik}} \\ e_1 &= \frac{1}{n}\sum_{i=1}^{n} \frac{u_i}{1+u_i^* u_i}, & e_{1k} &= \frac{1}{n}\sum_{i=1}^{n} \frac{u_{ik}}{1+u_{ik}^* u_{ik}} \\ e_2 &= \frac{1}{n}\sum_{i=1}^{n} \frac{u_i u_i^*}{1+u_i^* u_i}, & e_{2k} &= \frac{1}{n}\sum_{i=1}^{n} \frac{u_{ik} u_{ik}^*}{1+u_{ik}^* u_{ik}} \end{aligned} \tag{29}$$

Here T is the matrix from the Cholesky decomposition of the matrix S. Observation (sample element) u_i is a vector, e_0 is a scalar, e_1 is a vector and e_2 is a matrix.

Lemma 3: The characterization of the location of the minimum for multidimensional case. At the location of the minimum of $T(c,S)$ formulas given below hold:

$$c = c + T \cdot \frac{e_1}{e_0}, \quad S = T \cdot \frac{e_2}{e_0} \cdot T^* \tag{30}$$

Now we can get a pair of iterative formulas for the solution of this system of equations in the next form:

$$c_{k+1} = c_k + T_k \cdot \frac{e_{1k}}{e_{0k}}, \quad S_k = T_k \cdot \frac{e_{2k}}{e_{0k}} \cdot T^*, \quad for\ k = 0,1,2,\ldots \tag{31}$$

In this iteration the starting vector c_0 can be chosen arbitrarily from R^m, the starting matrix S_0 can be any real symmetric positive definite matrix of size $m \times m$. Especially we may start for example from the zero vector and the unit matrix.

The proof of (30) is based on the usual differentiation of the inaccuracy with respect to the parameters c and S, but it contains some technical details. We omit it. Unfortunately, I have not succeeded in getting a formula analogous to (11).

4 A computer test

The algorithms described in (18), (26) are easily programmable for the computers. We demonstrate a one-dimensional case. A sample of size 1000 was generated from the Cauchy distribution using random number generator in the software environment MATLAB 6.1. The location and scale parameters were 5 and 3 respectively. We performed 20 iterative steps starting from the initial point 100, 200. The results of the iterations are in Table 1. The formula (26) was rewritten in an iterative form analogue to the formula (18).

	Formula (18)			**Formula (26)**		
K	c_k	s_k	T_k	c_k	s_k	T_k
0	100.000000	200.000000	6.668564	100.000000	200.000000	6.668564
1	6.289809	97.805534	5.787873	4.783113	3.215649	3.684150
2	4.831633	17.841248	4.353839	4.975382	3.143052	3.683094
3	4.791041	7.561761	3.869480	4.976653	3.140030	3.683094
4	4.869049	4.890789	3.731749	4.976705	3.140012	3.683094
5	4.916448	3.923892	3.695515	4.976706	3.140011	3.683094
6	4.942617	3.511557	3.686236	4.976706	3.140011	3.683094
7	4.957335	3.321027	3.683886	4.976706	3.140011	3.683094
8	4.965734	3.229410	3.683293	4.976706	3.140011	3.683094
9	4.970536	3.184466	3.683144	4.976706	3.140011	3.683094
10	4.973266	3.162195	3.683107	4.976706	3.140011	3.683094
11	4.974804	3.151103	3.683097	4.976706	3.140011	3.683094
12	4.975663	3.145563	3.683095	4.976706	3.140011	3.683094
13	4.976138	3.142791	3.683094	4.976706	3.140011	3.683094
14	4.976398	3.141404	3.683094	4.976706	3.140011	3.683094
15	4.976540	3.140709	3.683094	4.976706	3.140011	3.683094
16	4.976617	3.140361	3.683094	4.976706	3.140011	3.683094
17	4.976658	3.140187	3.683094	4.976706	3.140011	3.683094
18	4.976680	3.140099	3.683094	4.976706	3.140011	3.683094
19	4.976692	3.140055	3.683094	4.976706	3.140011	3.683094
20	4.976699	3.140033	3.683094	4.976706	3.140011	3.683094

Table 1: One-dimensional case computer test

References

[Mathai] A. M. Mathai, P. N. Rathie: Basic Concepts in Information Theory and Statistics. Axiomatic Foundation and Application, Wiley Eastern limited, New Delhi, 1974.

[Rao] C. R. Rao: Linear Statistical Inference and its Applications, John Wiley and Sons, Inc. New York – London - Sydney

[Zakhs] S. Zakhs: The Theory of Statistical Inference, John Wiley and Sons, Inc. New York – London – Sydney – Toronto, 1971

[Gabrielsen] G. Gabrielsen: On the unimodality of the likelihood for the Cauchy distribution. Some comments. Biometrika 69, pp 677-8

[Copas] J. B. Copas: On the unimodality of the likelihood for the Cauchy distribution, Biometrika 62, pp 701-4

Appendix

To show formula (12), first we calculate another integral we need in the proof.

$$I_a(p) = \int_R \frac{\log\left(a^2 + p^2 x^2\right)}{1+x^2} dx = 2\pi \log|a+p| \quad for\ a, p \in R \tag{32}$$

In the special case, when $p=0$ formula (32) becomes

$$I_a(0) = \int_R \frac{\log a^2}{1+x^2} dx = 2\pi \log|a| \quad for\ a \in R \tag{33}$$

We calculate (32) as a parametric integral.

$$\begin{aligned} \frac{dI_a(p)}{dp} &= \int_R \frac{2px^2}{\left(1+x^2\right)\left(a^2+p^2x^2\right)} dx = \frac{2}{p}\int_R \frac{a^2+p^2x^2-a^2}{\left(1+x^2\right)\left(a^2+p^2x^2\right)} dx = \\ &= \frac{2}{p}\int_R \frac{dx}{1+x^2} - \frac{2a^2}{p}\int_R \frac{dx}{\left(1+x^2\right)\left(a^2+p^2x^2\right)} dx = \\ &= \frac{2\pi}{p} - \frac{2a^2}{p}\int_R \left(\frac{A}{1+x^2} + \frac{B}{a^2+p^2x^2}\right) dx \\ &where \quad A = \frac{1}{a^2-p^2},\ B = -\frac{p^2}{a^2-p^2} \end{aligned} \tag{34}$$

Calculating the elementary integrals in (34) we get:

$$\frac{dI_a(p)}{dp} = \frac{2\pi}{a+p} \tag{35}$$

This yields:

$$I_a(p) = 2\pi \log|a+p| + C \tag{36}$$

The value of the constant C can be obtained by substitution $p=0$ in (36). Comparing it with (33) we see that $C=0$. With this step the proof of (32) is complete.

Now we start with the proof of (12). We use the method of parametric integral again. Let us denote the integral in (12) by *I(b)*.

$$\begin{aligned}\frac{dI(b)}{db} &= \int_R \frac{-2ax}{\left(1+x^2\right)\left(a^2-2abx+x^2\right)}dx = \\ &= \int_R \left(\frac{Ax+B}{1+x^2}+\frac{Cx+D}{a^2-2abx+x^2}\right)dx \\ &\text{where } A=-\frac{2a\left(a^2-1\right)}{q(b)},\ B=\frac{4a^2b}{q(b)},\ C=-A,\ D=-a^2B \\ &\text{and } q(b)=\left(a^2-1\right)^2+4a^2b^2\end{aligned} \tag{37}$$

Formula (37) can be written in the form

$$\begin{aligned}\frac{dI(b)}{db} &= \int_R \left(\frac{A}{2}\frac{2x}{1+x^2}+\frac{C}{2}\frac{2x-2ab}{a^2-2abx+x^2}\right)dx + \\ &+ \int_R \left(\frac{Cab+D}{a^2-2abx+x^2}+\frac{B}{1+x^2}\right)dx\end{aligned} \tag{38}$$

Having *C*=-*A* we see that the first integral is zero and the second is elementary:

$$\begin{aligned}\frac{dI(b)}{db} &= \int_R \frac{d}{dx}\left[\frac{A}{2}\left(\log\left(1+x^2\right)-\log\left(a^2-2abx+x^2\right)\right)\right]dx + \\ &+ B\pi+\left(D-Aab\right)\frac{\pi}{a\sqrt{1-b^2}}\end{aligned} \tag{39}$$

Substituting constants *B,D* and *A* by their values we get

$$\frac{dI(b)}{db}=\frac{4a^2b}{q(b)}\pi+\left(\frac{2a\left(a^2-1\right)}{q(b)}-a^2\frac{4a^2b}{q(b)}\right)\frac{\pi}{a\sqrt{1-b^2}} \tag{40}$$

After a little algebra this formula simplifies to

$$\frac{dI(b)}{db}=\frac{\pi}{2}\frac{8a^2b}{q(b)}-\pi\frac{2ab\left(a^2+1\right)}{q(b)\sqrt{1-b^2}} \tag{41}$$

Integrating (41) we get

$$I(b)=\frac{\pi}{2}\log q(b)-2\pi a\left(a^2+1\right)\int\frac{b}{q(b)\sqrt{1-b^2}}db+K\,, \tag{42}$$

Where K is a constant to be determined. We substitute the square root in (42) by *y*, which yields:

$$I(b)=\frac{\pi}{2}\log q(b)+2\pi a\left(a^2+1\right)\int\left[\frac{dy}{\left(a^2-1\right)^2+4a^2-4a^2y^2}\right]+K \tag{43}$$

Making the substitution $z=\frac{2a}{1+a^2}y$ we obtain

$$I(b)=\frac{\pi}{2}\log q(b)+\pi\int\frac{dz}{1-z^2}+K=\frac{\pi}{2}\log q(b)+\frac{\pi}{2}\log\left|\frac{1+z}{1-z}\right|+K \tag{44}$$

Returning back to variable *y* and then to *b* we get

$$I(b)=\frac{\pi}{2}\log\frac{\left|\left(1-a^2\right)^2+4a^2b^2\right|\left|1+a^2+2a\sqrt{1-b^2}\right|}{1+a^2-2a\sqrt{1-b^2}}+K \tag{45}$$

Eliminating the square root in the denominator we get

$$I(b)=\frac{\pi}{2}\log\frac{\left[\left(1-a^2\right)^2+4a^2b^2\right]\left[1+a^2+2a\sqrt{1-b^2}\right]^2}{\left(1+a^2\right)^2-4a^2\left(1-b^2\right)}+K \tag{46}$$

(46) simplifies to

$$I(b)=\pi\log\left|1+a^2+2a\sqrt{1-b^2}\right|+K \tag{47}$$

Let us choose now in (12) *b=0*. It leads to the special case of (32) for *p=1*, which gives $2\pi\log|a+1|$. Quantity (47) yields $\pi\log(1+a)^2+K$ for *b=0*. Comparing them we can see that *K=0*.
This finishes the proof of (12).