Negative conclusion cases: a proposal for likelihood ratio evaluation

Bruno Cardinetti * Camillo Cammarota †

Abstract

The use of Bayesian approach in forensic science requires the evaluation of likelihood ratio related to the crime scene evidence event E and the suspect characteristic event C. This evaluation fails when the two events are disjoint, i.e. the evidence E is not compatible with the characteristic C of the suspect, and then the casework has a negative conclusion. This situation is very common, especially using continuous variables, e.g. height, refractive index, voice frequencies, etc. In particular in standard approach there is no difference between an evidence E close to C (for instance, heights with 1 centimeter of difference) and an evidence far from C.

We propose a method of calculation of the likelihood ratio, based on a bivariate representation of the database, supposed to be Gaussian, with a correlation coefficient $r > 0$. The likelihood ratio, calculated with this method, has larger values when both the events are in the tail of the distribution, as expected. Moreover, it reduces to the standard one when r tends to 1.

Application in the case of height is performed using Italian Carabinieri database.

Keywords: forensic science, Bayesian model, likelihood ratio, bivariate normal

2000 MSC: 62C10, 62P25, 91F99

*Reparto Carabinieri Investigazioni Scientifiche di Roma, viale Tor di Quinto 151, 00191 Rome, Italy

†Università degli Studi di Roma, La Sapienza, Dipartimento di Matematica, piazzale Aldo Moro 5, 00185 Rome, Italy

1 Introduction

The aim of forensic science is to support intelligence and the judge in making decision in Court by means of a scientific approach in evaluation of evidences. Accordingly, the forensic scientist has to present the results of his analysis in such appropriate way that no doubt of interpretation could arise.
The widespread accepted approach in the evaluation of evidences is the probabilistic one, based on the Bayesian model [1, 2, 3, 4, 5].
The theory allows to consider two complementary events C and $\bar{C}$ in very different cases. For example, the event C could be related to a blood group of a biological stain, or a refractive index (RI) of a glass fragment, or a height, and so on, concerning the suspect of the crime; so the event C is usually related to the guilty of a person, or his presence on the crime scene, or his responsibility in breaking a glass window, and so on.
In a suitable probability space with a probability P, the *a priori* odds $O(C)$ in favor of the event C is defined by:

$$O(C) = \frac{P(C)}{P(\bar{C})} \tag{1}$$

The *a priori* odds $O(C)$ represents how many times the probability of the event C is stronger with respect to the probability of the event $\bar{C}$: if $O(C) = 1$ the two probabilities have the same value, and no help in making decision in favor of one of them can arise; if $O(C) > 1$, the event C is more probable than $\bar{C}$; finally, if $O(C) < 1$, the event C is less likely than the event $\bar{C}$.
When information changes the scenario, new elements could be considered, and they could improve or not the evaluation of odds. If the event E represents the evidence, the judge must evaluate the *a posteriori* odds $O(C|E)$, which is defined by:

$$O(C|E) = \frac{P(C|E)}{P(\bar{C}|E)} \tag{2}$$

where the usual notation for the conditional probability of the event A with respect to the event B has been used, i.e.:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3}$$

For example, the evidence E represents information that arises from the crime scene analysis, and concerns the same variable of C: for instance, they are both intervals of heights.
According to Bayes' theorem, the *a posteriori* odds $O(C|E)$ can be written with respect to the *a priori* odds $O(C)$ as:

$$O(C|E) = LR(C, E) \cdot O(C) \tag{4}$$

where the quantity:

$$LR(C, E) = \frac{P(E|C)}{P(E|\bar{C})} = \frac{P(E \cap C)}{P(E \cap \bar{C})} \cdot \frac{P(\bar{C})}{P(C)} \tag{5}$$

is called *likelihood ratio*.
If $LR(C, E) > 1$, the evidence E favors the event C, and the greater the value of the likelihood ratio is, the more determinant the evidence is. On the contrary, if $LR(C, E) < 1$, the evidence E favors the event $\bar{C}$. Finally, $LR(C, E) = 1$ implies the absolutely unconcern of the evidence about the two events C and $\bar{C}$.
Note that, even if two complementary events C and $\bar{C}$ are considered in eq. (5), the generality of the method allows to consider the evaluation of the likelihood ratio for any two events.
Let us consider the following example. Let the probability space be the set of the real numbers, $\Re$, and the exclusive events C and $\bar{C}$ be the following:

> C: the height in centimeters of the suspect **does** belong to the interval $[170, 171]$;
>
> $\bar{C}$: the height in centimeters of the suspect **does not** belong to the interval $[170, 171]$;

Let us suppose now that the evidence E is:

> E: the height in centimeters of the crime author measured from the videotape recorded on the crime scene belongs to the interval $[164, 169]$;

An analog situation could present in the analysis of the RI of glasses, for which we refer to [2], if information on RI on both the suspect and the crime scene were associated to intervals.
The trivial consideration about the possibility that C matches E is obviously naught ($C \cap E = \emptyset$), and so a negative conclusion follows about the fact that the measured height of the crime author from the recorded videotape can belong to the interval C.
The reasons for which this negative conclusion is unsatisfactory lie on the fact that an error of measurement could be done, or the evidence E could arise from non 100%-reliable information (see, for example, [5], where the information about the color of a taxi could be not 100%-reliable since it was detected in darkness time).
Attempts to overcome this match/non-match approach into a continuous one has been already done in case of interpretation of evidence for RI of glass

fragments [6, 7]. In Lindley [6], the evidence and the control are expressed by two groups of measures, whose means are naturally considered to be normal variables, say X and Y. These two variables are supposed to be dependent under the assumption of guilty, and a bivariate distribution is taken into account. In Walsh [7], the denominator and the numerator of the likelihood ratio are, respectively, the values that the probability density of the database and of the t-density evaluated with control take on the mean value of the evidence. In both these papers, the relevant feature is that a greater evidence of identity occurs when X and Y are close and near to an unusual index with respect the case of frequently occurring indices.

In this paper, we try to overcome the match/non-match approach in the case in which the measures are represented by intervals, maintaining the advantages of the previous ideas. A distinctive feature of our approach is that probabilities are computed by integrating the probability density on intervals and not from a particular value of the density.

For the above considerations, we give a definition of the likelihood ratio which is non-zero also in case that the events are disjoint. Our definition has the following main properties: it depends on a parameter r and reduces to the standard case as r reaches a given value (in our case, $r \to 1$); the likelihood ratio is greater if the two intervals are both in the same tail of the distribution with respect to the case that they are both close to the mean. The method is based on a bivariate representation of the probability P as explained in the next section.

2 The extended likelihood ratio

Measurements in a database can be modeled as a random variable, which is characterized by a probability distribution from which it is straight forward to determine the probabilities of intervals of interest. For instance, the database of the heights of a population is described by a random variable, say X, with probability distribution P_X. The probability of events like $\{X \in A\}$, where A is for instance an interval of the real line $\Re$, is denoted as $P_X(A)$.

Our method is based on the idea that there is another random variable Y, with distribution P_Y equal to P_X. This means that for any event A, one has $P_X(A) = P_Y(A)$. We remark that X and Y are distinct random variables: if one randomly picks an individual according to the first one and randomly picks an individual according to the second one, this generally gives different outcomes of the height. In probability theory language, these two variables are called 'equally distributed'; only if two variables are assumed to be 'equal', the two random outcomes are equal.

A graphical representation of the outcomes, as ordered pairs of numbers, is a set of points which form a so called 'scatter plot'. For instance, in figure 1 there is the result of the simulation of 500 outcomes of two variables (heights) equally distributed according to a common database.
An important ingredient of our approach is that the two variables are 'nearly equal', i.e. for any pair of outcomes the two values are 'nearly equal', in the sense that the scatter plot is concentrated along the main diagonal $\{x = y\}$. If one has, for instance, two intervals, A and B, the event $\{X \in A \cap B\}$ can be written as the intersection of the two events $\{X \in A\}$ and $\{X \in B\}$, and in our approach we replace the second one by $\{Y \in B\}$. In a graphical representation, the two events $\{X \in A\}$ and $\{Y \in B\}$ are two strips of the plane, and their intersection is a rectangle. Hence, we replace the intersection $A \cap B$ by the rectangle $A \times B$, which is not empty also if the first is (see figure 1). Consequently, the probability of the intersection is replaced by a bivariate probability of the rectangle, and so an extended likelihood ratio can be defined. We describe more precisely our approach below, while mathematical details are given in the Appendix.
We assume that the measurements in the database can be modeled by a normal probability distribution X with mean μ and variance σ^2.
We consider a jointly normal pair of variables (X, Y), with distribution P^r_{XY}, such that its marginal distributions coincide with the distribution of X. The strength of the dependence between the two variables is quantified by their correlation index r, and it is well known that in general one has:

$$-1 < r < 1 \tag{6}$$

and that the case $r = 0$ corresponds to the independence of X and Y. We are interested in the case r close to 1, i.e. when the bivariate density is concentrated along the straight line $\{x = y\}$, corresponding to a strong positive dependence between X and Y (see figure 1).
With reference to this figure, the bivariate probability of a rectangle can be numerically evaluated as the fraction of points contained in the rectangle over the total number of points.
Our method is based on the following rule:

to replace in the likelihood ratio the univariate probability of the intersection of two events with the bivariate probability of the associated rectangle.

More precisely, we make the replacement:

$$P_X(A \cap B) \quad \rightarrow \quad P^r_{XY}(A \times B) \tag{7}$$

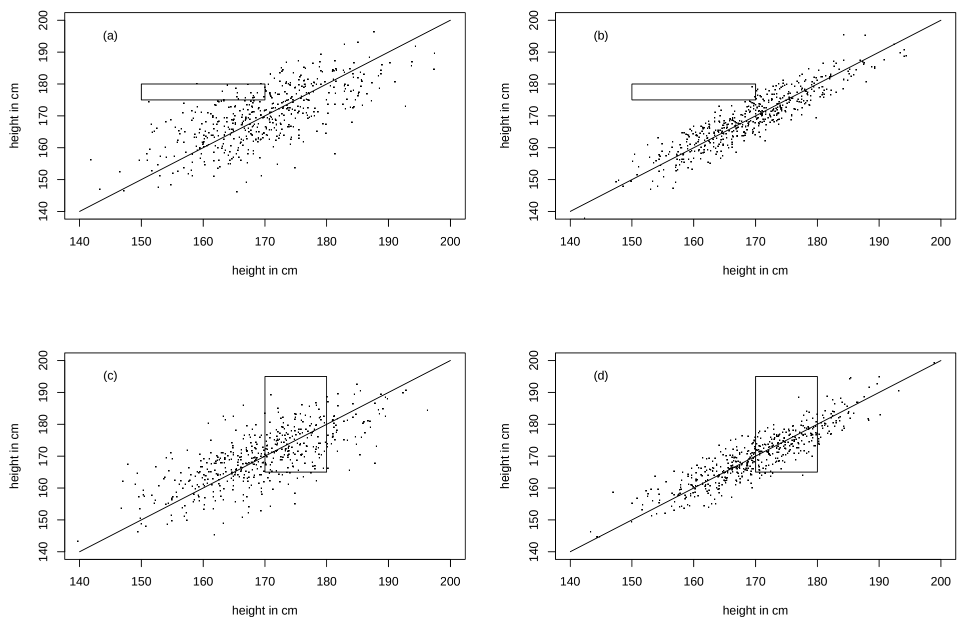

Figure 1: Simulation of a bivariate normal database with 500 points. (a) and (c): simulated points for correlation $r = 0.7$; (b) and (d): simulated points for correlation $r = 0.9$; the greater the correlation is, the more closer to the main diagonal the concentration of points is. (a) and (b): rectangle not intersecting the diagonal for $r = 0.7$ and $r = 0.9$ respectively; (c) and (d): rectangle intersecting the diagonal for $r = 0.7$ and $r = 0.9$ respectively.

We remark that the bivariate representation computed on rectangles enjoys the properties of the univariate probability of intersection, but if $A \cap B = \emptyset$, while $P_X(A \cap B) = 0$ in the classical case, in our approach one has $P^r_{XY}(A \times B) > 0$.
With these assumptions, we can provide a new version of the formula for computing the likelihood ratio: the extended likelihood ratio $LR^r(C, E)$ is:

$$LR^r(C, E) = \frac{P^r_{XY}(E \times C)}{P^r_{XY}(E \times \bar{C})} \cdot \frac{P_X(\bar{C})}{P_X(C)} \qquad (8)$$

obtained by making the replacement (7) in the definition of the likelihood ratio in equation (5).
We shall prove that $P^r_{XY}(A \times B)$ tends to $P_X(A \cap B)$ as r tends to 1, i.e. when the correlation takes its maximum value.
We note that if $r = 0$, i.e. in the case of independence, the extended likelihood ratio as computed according to equation (8) is 1 for any pair of events. Hence $r = 0$ implies a total unconcern of the two events.
In this sense, the parameter r can be interpreted as the degree of uncertainness related to the measurement of both intervals, being $r = 0$ the case of total uncertainness, while $r = 1$ the case of measurements without uncertainness.
In applications, the value of r has to be chosen in the interval $(0, 1)$ depending on the specific case-work under consideration. The question of an optimal choice of r is an important problem which needs further investigation.

3 A case-work: evaluation of extended likelihood ratio for the height

We apply the proposed approach to a case-work in which the measurements of the height of the suspect is the interval $C = [x_1, x_2]$ and the evidence is the interval $E = [y_1, y_2]$, and these are disjoint but very close. We have used the database of about 200,000 people filed in the police archive of Italian Carabinieri. This distribution is very close to a Gaussian one with mean $\mu = 170.73cm$ and standard deviation $\sigma = 8.86cm$. In this application, we choose as an example the value $r = 0.95$, very close to the maximum $r = 1$, which means that we retain the measurements very reliable.
The bivariate normal distribution has been simulated using the free statistical software R [8], and the number of simulated points is 50,000. The probability of the rectangle $C \times E \subseteq \Re^2$ can be easily computed counting the points that fall inside the rectangle and dividing by their total number.

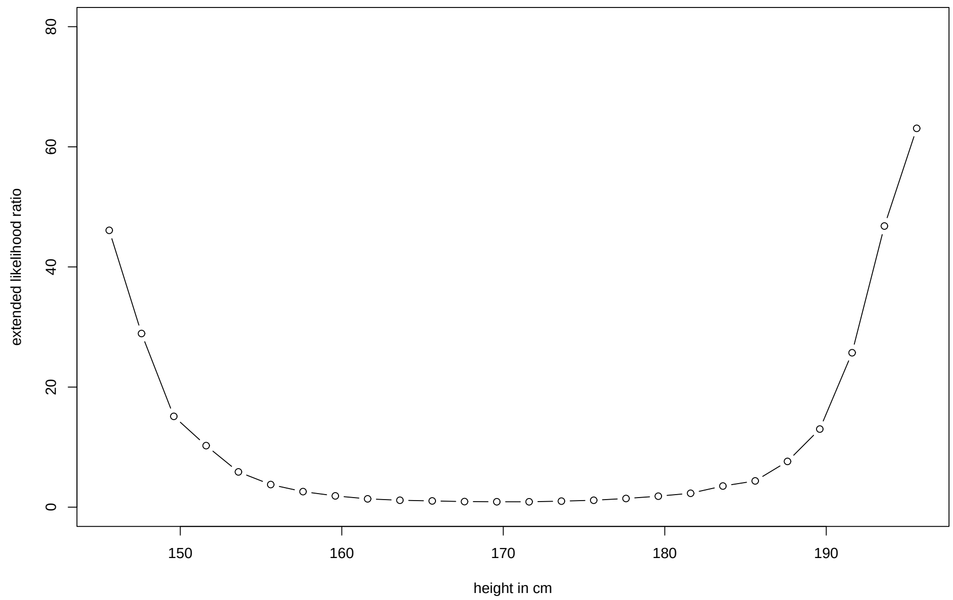

Figure 2: Extended likelihood ratio as a function of x for $r = 0.95$, for the intervals $E = [x-4, x]$ and $C = [x+1, x+5]$, computed for values of x from 145.6 to 195.6 at steps of 2 centimeters. The minimum is 0.90.

Putting $C = [x - 4, x]$, $E = [x + 1, x + 5]$, the likelihood ratio is a function of x, and it is investigated in order to validate our method. As result, this function has the following main property: its value is below 1 if x is close to the average of the heights, and it is very large if x is in the tails of the distribution. The plot of this function is shown in figure 2.

4 Conclusion

In this paper we have extended the notion of likelihood ratio to those cases in which the control and the evidence are expressed by disjoint events, and in particular by disjoint intervals.
Our proposal is based on the assumption that the database is normal and uses a bivariate normal distribution.
The value of the extended likelihood depends on a parameter r such that in the limit $r \to 1$ the standard definition is recovered.
We have applied our method to the database of heights filed in the police archive of Italian Carabinieri (200,000 people). We have considered the

events C and E, both of which are obtained from a indirect measurement. For this case, we have chosen $r = 0.95$.

We have computed the extended likelihood ratio in a case of two disjoint but nearby intervals of heights $[x-4, x]$ and $[x+1, x+5]$, and we have investigated the dependence on x. As expected, the likelihood ratio calculated with our method is very large when x is in the tails of the distribution.

In some applications, it may be useful to compare the likelihood ratio of E with respect to two events C_1 and C_2, related to two different suspects. In this case, in which a comparison is required, the choice if r is less critical.

The paper can be improved in several directions: to consider a non-normal database and a non-unidimensional database. The first case can be developed constructing a bivariate representation of the database, by using the theory of copulas. The second case is important, for instance, in analysis of voice frequencies. Finally, also the problem of dependence of the likelihood ratio on r, fixed a pair of intervals, should be investigated. For instance, the maximum or the mean value of the likelihood ratio could provide interesting quantities independent on r.

References

[1] Aitken C.G.G., *Statistics and evaluation of evidence for forensic scientists*, John Wiley & Sons, 1995

[2] Curran J.M., Hicks T.N., Buckleton J.S., *Forensic interpretation of glass evidence*, CRC Press, 2000

[3] Evett I.W., Weir B.S., *Interpreting DNA evidence*, Sinauer Associates, 1973

[4] Garbolino P., Taroni F., *Evaluation of scientific evidence using Bayesian networks*, Forensic Sci. Int. 125 (2002) 149-155

[5] Taroni F., Mangin P., *L'interprétation de la preuve scientifique: les juristes, les scientifiques et le probabilités*, Médicine et droit, 30 (1998) 6-15

[6] Lindley D.V., *A problem in forensic science*, Biometrika 64 (1977) 207-213

[7] Walsh K.A.J., Buckleton J.S., Triggs C.M., *A practical example of the interpretation of glass evidence*, Science & Justice 36 (1996) 213-218

[8] `http://www.R-project.org`

[9] Rohatgi V.K., *An introduction to Probability Theory and Mathematical Statistics*, New York, Wiley ed., 1976

Appendix

We assume that the database is modeled by a normal random variable X with mean μ and variance σ^2, whose probability density is:

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{9}$$

and the probability of the event $A \subseteq \Re$ is:

$$P_X(A) = \int_A ds\ p_X(s) \tag{10}$$

The probability density of the pair (X, Y) is a two variable function $p^r_{XY}(s,t)$, with $(s,t) \in \Re^2$, so that the probability of rectangles is:

$$P^r_{XY}(A \times B) = \int_A ds \int_B dt\ p^r_{XY}(s,t) \tag{11}$$

It is well known [9] that the joint probability density of a pair (X, Y) of normal variables is the following two variables function:

$$p^r_{XY}(s,t) = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\frac{1}{2\sigma^2(1-r^2)}[(s-\mu)^2 - 2r(s-\mu)(t-\mu) + (t-\mu)^2]} \tag{12}$$

where the parameter r represents the correlation index. In particular, the following property holds:

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{13}$$

where:

$$Cov(X,Y) = E((X-\mu)(Y-\mu)) \tag{14}$$

and E denotes the expectation.
We also note that the function $p^r_{XY}(s,t)$ is symmetric, i.e.:

$$p^r_{XY}(s,t) = p^r_{XY}(t,s) \tag{15}$$

and its marginals are the densities of X and Y, according to the following equations:

$$p_X(s) = \int_\Re dt\ p^r_{XY}(s,t) \tag{16}$$

$$p_Y(t) = \int_\Re ds\ p^r_{XY}(s,t) \tag{17}$$

In turn, the two densities are equal:

$$p_X(s) = p_Y(s), \quad s \in \Re \tag{18}$$

Similarly, the probability of rectangles satisfies the following properties: by symmetry, one has:

$$P^r_{XY}(A \times B) = P^r_{XY}(B \times A) \tag{19}$$

and taking marginals:

$$P^r_{XY}(A \times \Re) = P_X(A) \tag{20}$$

Proof of the limit $P^r_{XY}(A \times B) \to P_X(A \cap B)$ *as* $r \to 1$
In eq. (12), we make the substitutions:

$$\xi = x - \mu, \quad \eta = y - \mu \tag{21}$$

which is equivalent to consider centered variables, and we reduce the quadratic form into a diagonal one:

$$(x-\mu)^2 - 2r(x-\mu)(y-\mu) + (y-\mu)^2 = \xi^2 - 2r\xi\eta + \eta^2 = \tag{22}$$

$$= (1+r)(\frac{\xi - \eta}{\sqrt{2}})^2 + (1-r)(\frac{\xi + \eta}{\sqrt{2}})^2$$

Introducing the new variables:

$$u = \frac{\xi - \eta}{\sqrt{2}}, \quad v = \frac{\xi + \eta}{\sqrt{2}} \tag{23}$$

the joint density p^r_{XY} can be expressed as a product:

$$p^r_{XY}(x, y) = g^r_+(v) g^r_-(u) \tag{24}$$

where:

$$g^r_+(v) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{1+r}} e^{-\frac{v^2}{2\sigma^2(1+r)}} \tag{25}$$

and:

$$g^r_-(u) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{1-r}} e^{-\frac{u^2}{2\sigma^2(1-r)}} \tag{26}$$

The two functions g^r_+ and g^r_- are the densities of centered normal variables with variances given by $\sigma^2(1+r)$ and $\sigma^2(1-r)$, respectively. In particular, we shall use the following property of g^r_-: given any interval J, one has:

$$\lim_{r\to 1} \int_J du \; g^r_-(u) = \begin{cases} 0 & \text{if} \quad 0 \notin J \\ 1 & \text{if} \quad 0 \in J \end{cases} \tag{27}$$

This property, whose proof follows from standard arguments, is equivalent to say that the density g_-^r tends to the delta mass concentrated on 0. Furthermore, the density g_+^r tends to the one of a centered normal with variance $2\sigma^2$.
Since the Jacobian determinant of the transformation $(\xi, \eta) \to (v, u)$ is 1, denoting by $T(A \times B)$ the transformed rectangle, one has:

$$P_{XY}^r(A \times B) = \int\int_{T(A\times B)} dudv\ g_+^r(v) g_-^r(u) \tag{28}$$

The axes (v, u) are a pair of orthogonal axes centered at (μ, μ), the first one corresponding to the line $\{x = y\}$. The set $T(A \times B)$ is a normal domain with respect to the v axis, and so it can be represented by means of the collection of intervals $J(v)$, $v \in K$, of the u axis, where K is the projection of the domain on the v axis. The double integral is then reduced to:

$$P_{XY}^r(A \times B) = \int_K dv\ g_+^r(v) \int_{J(v)} du\ g_-^r(u) \tag{29}$$

We first consider the case $A \cap B = \emptyset$. We have already noticed that in this case the domain $T(A\times B)$ does not intersect the v axis, hence all the intervals $J(v)$ do not contain zero. Hence, from equation (27), first case:

$$\lim_{r\to 1} \int_{J(v)} du\ g_-^r(u) = 0 \tag{30}$$

and also the double integral tends to zero.
In the case $A \cap B = D \neq \emptyset$, the only non-vanishing contribution is:

$$P_{XY}^r(D \times D) = \int_K dv\ g_+^r(v) \int_{J(v)} du\ g_-^r(u) \tag{31}$$

where all the $J(v)$'s contain 0. In force of eq. (27), second case, the inner integral has limit 1 for any $v \in K$, and so:

$$\lim_{r\to 1} P_{XY}^r(D \times D) = \int_K dv\ g_+^r(v) \tag{32}$$

If c_1 and c_2 are the extremes of D, then the extremes of K are $\sqrt{2}(c_1 - \mu)$ and $\sqrt{2}(c_2 - \mu)$. For the computation of the integral, if we use the new variable $s = \mu + v/\sqrt{2}$, the new extremes are exactly c_1 and c_2, and the integrating function becomes $p_X(s)$. Hence:

$$\int_K dv g_+^r(v) = \int_D p_X(s) ds = P_X(D) \tag{33}$$

and this concludes the proof.